Dresden University of Technology

Faculty of Forestry, Geosciences and Hydrosciences

SYSTEM ANALYSIS IN THE WATER MANAGEMENT

Peter-Wolfgang Gräber

Summer Semester 2005

Author presentation

The lecture **water management system** analysis is based on foundation courses such as mathematics, physics and informatics. And the chapters, which play a role in water management task solutions, will be specially discussed in the lecture, such as vector analysis, solution of equation systems, matrix calculation, and solution of differential equations as well as numerical integration.

The following chapters are revisions of basic knowledge and will only be shown with corresponding key points. Self study is strongly recommended during the revision.

The further chapters go beyond basic knowledge and indicate mathematical methods, which are related to the water management practice.

The teaching contents of subject **water management system** analysis require an advanced mathematical knowledge, including abstraction ability. In the exercise and computer courses, some problem will be discussed combined with practically relevant cases in order to develop a deeper understanding of this lecture.

Content

Abbreviations and Formula symbols

Symbol	Unit	Meaning
A	m^2	area
a	$s \cdot m^{-2}$	geohydraulic time constant
a	m	groundwater layer
B	m	feeding factor
b	m	width
C	$g \cdot l^{-1}$; %	concentration
DGL		differential equation
D	m	through flow potency
d	m	layer thickness, spacing
E	V	electrode potential, voltage source
e^-	C	electron, elementary charge, $e^- = -1{,}60210 \cdot 10^{-19} C$
F	C	faraday constant, F = 96491,6C
F		FOURIER transpose
G	S	electric conductance
g	$m \cdot s^{-2}$	gravity acceleration, $g \approx 9{,}832 m \cdot s^{-2}$(pole), $g \approx 9{,}780 m \cdot s^{-2}$ (equator)
g		weighting function
GF		quality function

Symbol	Unit	Meaning
GW		groundwater
GWBR		groundwater observation well
GWL		aquifer
GWÜ		groundwater monitoring
GWPN		groundwater sampling
GWR		groundwater resource
H	m	water level, general
H	m^2	equivalent potential
h	m	groundwater level
h	m, s	step length
h		transition function
I	A	electric current
k	$m \cdot s^{-1}$	k_f-value, permeability coefficient
L	$-$	LAPLACE transform
LGS		system of linear equations
l	m	length
LF	$S \cdot cm^{-1}$	electric conductivity
M	$-$	frequency

Symbol	Unit	Meaning
ODE		ordinary differential equation
PDE		partial differential equation
p	$N \cdot m^{-2}$	pressure
Q	C	electric charge
Q	$m^3 \cdot s^{-1}$	volume flow
R	Ω	electric resistance
R		plant
Re	$-$	REYNOLD's number
REV		representative unit volume
S	$-$	storage coefficient
S		plant controlled system
s	m	distance
T	K	temperature, δ[ºC] = T[K] – 273,15K
T	$m^2 s^{-1}$	transmissibility
t	s	time
U	V	electric tension

Symbol	Unit	Meaning
V	m^3	volume
$\dot{V}$	$m^3 \cdot s^{-1}$	volume flow
v	m/s	velocity
$W(\sigma)$		THEIS well formula, power series
w	$-$	command variable
x	m	position coordinates
x	$-$	general signal variable
x	$-$	control variable, control factor
y	m	position coordinates
y	$-$	manipulated variable
Z	m^2	potential difference (groundwater)
z	m	position coordinates, altitude
z	$-$	disturbance variable

Symbol	Unit	Meaning
α	$Grad$	angle
γ	$-$	EULER's constant, $\gamma \approx 0{,}5772156649$
δ	^{o}C	temperature, δ / ºC = T / K – 273,15K
δ	$-$	DIRAC Impulse
ε	$F \cdot m^{-1}$	dielectric constant
ε_0	$F \cdot m^{-1}$	vacuum dielectric constant, $\varepsilon_0 = 8{,}855 \cdot 10^{-12} F \cdot \mathrm{m}^{-1}$
ε	$-$	general error
λ	m	wavelength
λ^*	m	effective boundary condition range
μ	$H \cdot m^{-1}$	permeability
μ_0	$H \cdot m^{-1}$	vacuum permeability, $\mu_0 = 1{,}2566 \cdot 10^{-6} H \cdot \mathrm{m}^{-1}$
ρ	$g \cdot m^{-3}$	density
ρ	$\Omega \cdot m^{-1}$	electrical resistivity
σ	$-$	argument of THEIS well formula W(σ)
τ	s	delay time
Φ	m^2	GIRINSKIJ potential (groundwater)
φ	V	electric potential
$\varkappa$	$S \cdot m^2 m^{-1}$	electric conductivity
$\varkappa$	$S \cdot cm^{-1}$	electric conductivity
$\mathbf{1}$		unit step

Part I

Mathematical fundamentals

The chapter **mathematical fundamentals** are directly based on the elementary learning of mathematics and actually show some solutions of selected problems, which are important in the water management subjects. After this review special, advanced topics will be discussed.

Chapter 1

1 Algebra Fundamentals

1.1 Exponential- and logarithmic expressions

The most important transformations of exponential expressions are:

$$x^a \cdot x^b = x^{a+b} \tag{1.1}$$

$$\frac{x^a}{x^b} = x^{a-b}$$

$$(x^a)^b = x^{a \cdot b}$$

$$\sqrt[b]{x^a} = x^{\frac{a}{b}}$$

$$x^0 = 1$$

In such cases, exponential expressions can be transformed by definition.

$$x = a^{\log_a(x)} \qquad \text{specially} \quad : x = e^{\ln(x)} \tag{1.2}$$

For logarithmic expressions one can use the following rules to transform:

$$\ln(a \cdot b) = \ln(a) + \ln(b) \tag{1.3}$$

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$$

$$\ln\left(x^b\right) = b \cdot \ln(x) \qquad \text{specially} \quad \ln\left(e^b\right) = b \cdot \ln(e) = b, \text{ da: } \ln(e) = 1$$

$$\ln\left(\sqrt[b]{x}\right) = \frac{1}{b}\ln(x)$$

$$\log_x(1) = 0$$

$$\log_x(x) = 1 \qquad \text{specially} \quad : \lg(10) = 1, \ \ln(e) = 1$$

Task:

Simplify the following expressions:

a) $\dfrac{(18a^2x)^4}{(27ax^2)^5} \cdot \dfrac{(15ax^2)^4}{(20a^3x)^2}$

b) $\dfrac{0,004 \cdot 10^2 \cdot 0,2^3}{0,2 \cdot 10^{-3} \cdot 16}$

c) $\dfrac{\sqrt{x}\sqrt[3]{x^2}\sqrt[10]{x}}{\sqrt[5]{x^3}}$

d) $2\log_{10} x^3 - 3\log_{10} y^2$

e) $(\ln u + 4\ln v)$

Transform the following formulas according to *t*:

a) $I = I_0\left(\exp\left(-\dfrac{t}{T}\right)\right)$

b) $I = I_0\left(1 + 9\log 9\dfrac{t}{T}\right)$

c) $I = I_0\left(e^{\mu t} - 1\right)$

1.2 Matrices

1.2.1 Fundamentals

In the following the most important calculation rules for matrixes are to be specified.

• **General Matrix**
A system of *m* times *n* elements (e.g. numbers, functions), which is arranged in *m* rows and *n* columns, is called **matrix** of type (m; n):

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \quad (1.4)$$

$\longleftarrow$ *i*-th row

$\uparrow$

j-th column

• **Square Matrix**
If the number of rows *m* and columns *n* of a matrix are same, i.e. $m = n$, the matrix is square. The elements $a_{11}, a_{22}, a_{33} \cdots a_{nn}$ form the **main diagonal** and $a_{1n}, a_{2n-1}, a_{3n-2} \cdots a_{n-1}$ are **secondary diagonal.**

• **Transpose of a matrix**
A resultant matrix A^T will be obtained by interchanging the rows and columns of matrix A, which is called **transpose** of A.
In general it can be written as:

$$\mathrm{A} = [a_{jk}] \quad (1.5)$$

$$\mathrm{A}^T = [a_{jk}]^T = [a_{kj}]$$

specially $[a_{jj}] = [a_{kk}]$

The transpose of a square matrix is its reflection through the main diagonal.
The transpose of a symmetrical matrix is the same like the original matrix.

$$\mathbf{A}_s = \mathbf{A}_s^T \tag{1.6}$$

In a skew symmetrical or antimetrical matrix:

$$\mathbf{A}_a = -\mathbf{A}_a^T \tag{1.7}$$

$$[a_{jk}] = -[a_{kj}]$$

$$[a_{jj}] = 0$$

Each square matrix (Aq) can be decomposed in form of the sum of a symmetrical (As) and an antimetrical (Aa) matrix.

$$\mathbf{A}_q = \mathbf{A}_s + \mathbf{A}_a \tag{1.8}$$

$$\mathbf{A}_s = \frac{1}{2} \cdot (\mathbf{A}_q + \mathbf{A}_q^T)$$

$$\mathbf{A}_a = \frac{1}{2} \cdot (\mathbf{A}_q - \mathbf{A}_q^T)$$

Examples of transpose matrix calculation:

1.

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 4 \\ -3 & 0 & 2 \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} 2 & -3 \\ 1 & 0 \\ 4 & 2 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 5 \\ 7 & 2 & 6 \\ 8 & 9 & 3 \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} 1 & 7 & 8 \\ 4 & 2 & 9 \\ 5 & 6 & 3 \end{bmatrix}$$

2.

$$\mathbf{A}_{sym} = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 6 \\ 5 & 6 & 3 \end{bmatrix} = \mathbf{A}_{sym}^T$$

3.

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 4 \\ 5 & 7 & 8 \\ -4 & 0 & 5 \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} 3 & 5 & -4 \\ -1 & 7 & 0 \\ 4 & 8 & 5 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{A}_s + \mathbf{A}_a$$

$$\mathbf{A}_s = \frac{1}{2} \cdot \left(\begin{bmatrix} 3 & -1 & 4 \\ 5 & 7 & 8 \\ -4 & 0 & 5 \end{bmatrix} + \begin{bmatrix} 3 & 5 & -4 \\ -1 & 7 & 0 \\ 4 & 8 & 5 \end{bmatrix} \right) = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 7 & 4 \\ 0 & 4 & 5 \end{bmatrix}$$

$$\mathbf{A}_a = \frac{1}{2} \cdot (\mathbf{A}_q - \mathbf{A}_q^T)$$

$$\mathbf{A}_s = \frac{1}{2} \cdot \left(\begin{bmatrix} 3 & -1 & 4 \\ 5 & 7 & 8 \\ -4 & 0 & 5 \end{bmatrix} - \begin{bmatrix} 3 & 5 & -4 \\ -1 & 7 & 0 \\ 4 & 8 & 5 \end{bmatrix} \right) = \begin{bmatrix} 0 & -3 & 4 \\ 3 & 0 & 4 \\ -4 & -4 & 0 \end{bmatrix}$$

• **Unit matrix**

A square matrix, in which all elements of the main diagonal are equal to one and all other elements are zero, is called unit matrix and designated as **E**.

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{1.9}$$

$$\mathbf{A} \cdot \mathbf{E} = \mathbf{E} \cdot \mathbf{A} = \mathbf{A} \tag{1.10}$$

$$\mathbf{E}^n = \mathbf{E}$$

• **Diagonal matrix:**
A matrix, in which all elements are zero except the elements of main diagonal, is called **diagonal matrix**.

$$[a_{ii}] = \vee \neq 0 \qquad (1.11)$$

The elements of the main diagonal [a_{ii}] can be same and unequal to zero. And the diagonal matrixes belong to symmetrical matrixes.

• **Tridiagonal matrix**
It is named **tridiagonal matrix** if a square matrix possesses such a characteristic that the elements of the main diagonal and both of neighbouring diagonals are same and unequal to zero.

$$\begin{aligned} [a_{ii}] &= \vee \neq 0 \qquad (1.12) \\ [a_{i-1j}] &= \vee \neq 0 \\ [a_{ij-1}] &= \vee \neq 0 \end{aligned}$$

Generally speaking, the tridiagonal matrices are also symmetrical.

• **Band matrix**

A band matrix contains a large number of zero elements. The diagonal and selected parallel diagonals contain elements, which are different from zero. The extension of neighbouring diagonal as many as desired (≤ n - 1) leads to **band matrix**.

$$\begin{aligned} [a_{ii}] &= \vee \neq 0 \\ [a_{i-1j}] &= [a_{ij-1}] = \vee \neq 0 \\ [a_{ki}] &= [a_{ik}] = \vee \neq 0 \end{aligned}$$

The elements of a band matrix are all arranged on a diagonal band.
The band width is characterized by the occupied band. The "furthest distance" of an element from the main diagonal is the width, and the main diagonal is counted as well.

Example of a band matrix:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdots & a_{1k} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & \cdots & 0 & a_{2k+1} & 0 \\ 0 & a_{32} & a_{33} & a_{34} & \cdots & 0 & a_{3k+2} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ a_{k1} & 0 & 0 & a_{ij-1} & a_{ij} & a_{ij+1} & 0 \\ 0 & a_{k+12} & 0 & 0 & \ddots & \ddots & a_{m-1n} \\ 0 & 0 & a_{k+23} & 0 & \cdots & a_{mn-1} & a_{mn} \end{bmatrix}$$

In this example the band width is *k*.

1.2.2 Calculation rules

• Matrix addition and subtraction

If A = $[a_{jk}]$ and B = $[b_{jk}]$ have the same orders, i.e. the same rows and columns, then:

$$\mathrm{A} \pm \mathrm{B} = [a_{jk}] \pm [b_{jk}] \tag{1.13}$$

Example of matrix addition:

$$\mathrm{A} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} 3 & -15 \\ -2 & 10 \end{bmatrix}$$

$$\mathrm{A} + \mathrm{B} = \begin{bmatrix} 4+3 & 6-15 \\ 6-2 & 9-10 \end{bmatrix} = \begin{bmatrix} 7 & -9 \\ -4 & -1 \end{bmatrix}$$

• **Matrix multiplication and division**

If A = [a_{jk}] is a $m \times p$ matrix and B = [b_{jk}] is $p \times n$ matrix, we define the product of A · B as matrix C = [c_{jk}], with order $m \times n$. It means that matrix C has the same number of rows with matrix A and the same number of columns with matrix B.

The product A · B is only defined, when the number of column of A (p of matrix $m \times p$) matches the number of row of B (p of matrix $p \times n$).

We speak about multiplication as ”rows time columns”. The rows of A will be multiplied by the columns of B. According to FALK’s scheme: C = [c_{jk}] = $\sum$ [a_{jl}] · [b_{lk}] with $j = 1 \cdots m$ and $k = 1 \cdots n$.

The multiplication B · A forms a matrix C with order of $p \times p$:

What should be noticed is that the **commutative law** is not applicable (A · B ≠ B · A).

In case of **square matrix,** A, B and C are the same order $m \times m$.

The **division** is reverse procedure of multiplication, as the inverse matrix (see page 10) is built.

Example of matrix multiplication:

Given: $\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ -1 & 0 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ -1 & 2 \end{bmatrix}$

Find: $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{B} \cdot \mathbf{A}$

$$\mathbf{A} \cdot \mathbf{B} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot 0 + (4 \cdot (-1)) & 1 \cdot 1 + 2 \cdot 3 + 4 \cdot 2 \\ (-1 \cdot 2) + 0 \cdot 0 + (3 \cdot (-1)) & (-1 \cdot 1) + 0 \cdot 3 + 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} -2 & 15 \\ -5 & 5 \end{bmatrix}$$

$$\mathbf{B} \cdot \mathbf{A} = \begin{bmatrix} 2 \cdot 1 + (1 \cdot (-1)) & 2 \cdot 2 + 1 \cdot 0 & 2 \cdot 4 + 1 \cdot 3 \\ 0 \cdot 1 + (3 \cdot (-1)) & 0 \cdot 2 + 3 \cdot 0 & 0 \cdot 4 + 3 \cdot 3 \\ ((-1) \cdot 1) + (2 \cdot (-1)) & ((-1) \cdot 2) + 2 \cdot 0 & ((-1) \cdot 4) + (2 \cdot 3) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 4 & 11 \\ -3 & 0 & 9 \\ -3 & -2 & 2 \end{bmatrix} \neq \mathbf{A} \cdot \mathbf{B}$$

Given: $\mathbf{A} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & -15 \\ -2 & 10 \end{bmatrix}$

Find: $\mathbf{A} \cdot \mathbf{B}$

$$\mathbf{A} \cdot \mathbf{B} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} \cdot \begin{bmatrix} 3 & -15 \\ -2 & 10 \end{bmatrix} = \begin{bmatrix} (3 \cdot 4 - 2 \cdot 6) & (4 \cdot (-15) + 6 \cdot 10) \\ (3 \cdot 6 - 2 \cdot 9) & (6 \cdot (-15) + 9 \cdot 10) \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Attention:
In contrast to the algebra of real numbers the commutative law is not applicable in the matrix multiplication. $A \cdot B \neq B \cdot A$
The **associative law** $A \cdot (B \cdot C) = (A \cdot B) \cdot C$ and the **distributive law** $A \cdot (B + C) = A \cdot B + A \cdot C$ on the other hand are available in the matrices.
A square matrix can be multiplied by itself. $\mathbf{A^2 = A \cdot A}$. Then we get the exponentiation of matrices.

The inverse Matrix

The inverse of a square matrix A is a matrix $\mathbf{A^{-1}}$ such that:

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{E} \qquad (1.14)$$

The inverse matrix $\mathbf{A^{-1}}$ can be exactly built if the determinant D = det A (Chapter 1.2.3 matrix determinant, page13) of matrix A is unequal to zero ($D \neq 0$). The inverse matrix $\mathbf{A^{-1}}$ is formed by the subdeterminant U_{ij} for the element a_{ij}. With the determinant U_{ji} pertaining to element a_{ji}:

$$a_{ij} = (-1)^{i+j} \frac{U_{ji}}{D} \qquad (1.15)$$

The change between row and column must be considered. The matrix of subdeterminat is transposed. Furthermore a change of sign takes place as a function of the distance to main diagonals a_{ii} of the matrix, i.e., when the sum i and j is odd number, the subdeterminant should be multiplied by -1.

The construction of the inverse matrix is demonstrated on the basis of a two- and a three-row matrix.

For a two-row matrix:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{1.16}$$

$$\mathrm{A}^{-1} = \begin{bmatrix} \frac{a_{22}}{D} & -\frac{a_{12}}{D} \\ -\frac{a_{21}}{D} & \frac{a_{11}}{D} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} a_{11} & -a_{12} \\ -a_{21} & a_{22} \end{bmatrix}$$

$$D = \det \mathrm{A} = a_{11} \cdot a_{22} - a_{21} \cdot a_{12}$$

For a three-row matrix:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1.17}$$

$$\mathrm{A}^{-1} = \begin{bmatrix} \frac{U_{11}}{D} & -\frac{U_{21}}{D} & \frac{U_{31}}{D} \\ -\frac{U_{12}}{D} & \frac{U_{22}}{D} & -\frac{U_{32}}{D} \\ \frac{U_{13}}{D} & -\frac{U_{23}}{D} & \frac{U_{33}}{D} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} U_{11} & -U_{21} & U_{31} \\ -U_{12} & U_{22} & -U_{32} \\ U_{13} & -U_{23} & U_{33} \end{bmatrix}$$

$$D = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}$$

Example of inverse matrix calculation:

1. Wenn $\mathrm{A} = \begin{bmatrix} 5 & 4 \\ 2 & 2 \end{bmatrix}$ ist, dann ist $D = 5 \cdot 2 - 2 \cdot 4 = 2$ und

$$\mathrm{A}^{-1} = \frac{1}{2} \begin{bmatrix} 2 & -4 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ -1 & 2,5 \end{bmatrix}$$

2. Find the inverse matrix of

$$\mathrm{A} = \begin{bmatrix} 2 & 1 & -1 \\ 5 & 2 & 0 \\ 1 & 1 & -2 \end{bmatrix}$$

The determinant A can be developed for example according to the second row:

$$D = -5\begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} + 2\begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix} - 0\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix}$$

$$= -5(-2+1) + 2(-4+1) = -1$$

The subdeterminants are:

$$U_{11} = \begin{vmatrix} 2 & 0 \\ 1 & -2 \end{vmatrix} = -4; \quad U_{12} = \begin{vmatrix} 5 & 0 \\ 1 & -2 \end{vmatrix} = -10; \quad U_{13} = \begin{vmatrix} 5 & 2 \\ 1 & 1 \end{vmatrix} = 3$$

$$U_{21} = \begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} = -1; \quad U_{22} = \begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix} = -3; \quad U_{23} = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = 1$$

$$U_{31} = \begin{vmatrix} 1 & -1 \\ 2 & 0 \end{vmatrix} = 2; \quad U_{32} = \begin{vmatrix} -2 & 1 \\ 5 & 0 \end{vmatrix} = 5; \quad U_{33} = \begin{vmatrix} 2 & 1 \\ 5 & 2 \end{vmatrix} = -1$$

Then

$$\mathrm{A}^{-1} = \begin{bmatrix} 4 & -1 & -2 \\ -10 & 3 & 5 \\ -3 & 1 & 1 \end{bmatrix}$$

1.2.3 Matrix determinant

A number can be assigned to a square matrix A, whose value is determinant D = det A. The *n*-th order determinant development can be defined recursively with the help of LAPLACE expansion theorem.

$$D = \det \mathrm{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{1.18}$$

The development can take place according to the elements of rows or columns.

1. Development according to the elements of *i*-th row

$$D = \det \mathrm{A} = \sum_{\nu=1}^{n} a_{i\nu} A_{i\nu} \qquad i \quad \text{fixed} \tag{1.19}$$

2. Development according to the elements of *k*-th column

$$D = \det \mathrm{A} = \sum_{\mu=1}^{n} a_{\mu k} A_{\mu k} \qquad k \quad \text{fixed} \tag{1.20}$$

Here A_{ik} means the **adjunct** belonging to the element a_{ik}, i.e. the **subdeterminant** U_{ik} of the element a_{ik} multiplied by factor $(-1)^{i+k}$. We get the **subdeterminant** U_{ik} of the element a_{ik} from *n*-th order determinant by deleting the *i*-th row and *k*-th column; it has order *n*-1.
Thus the rank of the subdeterminant is always one order lower than the associated determinant.
The **rank** of a matrix is determined by the highest order.

For a three-row matrix A:

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1.21}$$

then for example the subdeterminant for element a_{11}:

$$U_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32} \tag{1.22}$$

For three-row determinant the **SARRUS's rule** can be also applied. The first two columns are written fictitiously right beside the determinant and afterwards the sum of the diagonal products of the "minor diagonal" are drawn off from that of the "main diagonals".

$$A = \det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

Tips:

For the development of the determinants it is always favourable to select the row or column with most zero elements.

If the matrix consists of only one element A= [*a*], then: det A = *a*.

1.2.4 Task for calculation of matrices

1. Given $\mathrm{A} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$, $\mathrm{B} = \begin{bmatrix} -1 & 1 \\ 2 & -4 \end{bmatrix}$, $\mathrm{C} = \begin{bmatrix} 1 & 4 \\ -2 & -1 \end{bmatrix}$

Find a) $2\mathrm{A} + 3\mathrm{B}$ b) $\mathrm{A} - 2\mathrm{B} - 3\mathrm{C}$ c) $\mathrm{A} \cdot \mathrm{B}$

d) $(\mathrm{A} \cdot \mathrm{B})^T$ e) $\mathrm{B} \cdot \mathrm{A}$ f) $(\mathrm{A} \cdot \mathrm{B}) \cdot \mathrm{C}$

g) $(\mathrm{B} \cdot \mathrm{A}) \cdot \mathrm{C}$ h) $\mathrm{B}^T\mathrm{A}^T$ i) $\mathrm{A}^T + \mathrm{B}^T$

2. Find matrix $\mathrm{C} = A \cdot B^T$:

a) $\mathrm{A} = \begin{bmatrix} 2 & 3 & 1 & 1 \end{bmatrix}$ $\mathrm{B} = \begin{bmatrix} 2 & -1 & 1 & 3 \end{bmatrix}$

b) $\mathrm{A} = \begin{bmatrix} 1 & 3 \\ -2 & 1 \end{bmatrix}$ $\mathrm{B} = \begin{bmatrix} 1 & 0 \\ 2 & -3 \\ 3 & -8 \end{bmatrix}$

3. Find inverse matrix A^{-1}:

a) $\mathrm{A} = \begin{bmatrix} 3 & 5 \\ 2 & 7 \end{bmatrix}$

b) $\mathrm{A} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix}$

c) $\mathrm{A} = \begin{bmatrix} 2 & 2 & 3 \\ -4 & -2 & 3 \\ 4 & 3 & 2 \end{bmatrix}$

1.3 Linear Equation Systems (LGS)

Linear equation systems play an important role in the mathematical treatment of one-, two-, or three dimensional field problem as well as in hydrology and hydrogeology. Such equation systems with a large number of unknown quantities and equations, whose magnitudes of order may reach million, originate from quantization methods. The continuous field problems are decomposed in generally potential fields and discontinuous partial processes. And these can be described by linear, non-linear or linearized equations. The equation systems have the following figure:

$$\begin{array}{lllllll}
a_{11}x_1 & +a_{12}x_2 & +a_{13}x_3 & +\cdots a_{1j}x_j & \cdots + a_{1n}x_n & = r_1 & \\
a_{21}x_1 & +a_{22}x_2 & +a_{23}x_3 & +\cdots a_{2j}x_j & \cdots + a_{2n}x_n & = r_2 & \\
a_{31}x_1 & +a_{32}x_2 & +a_{33}x_3 & +\cdots a_{3j}x_j & \cdots + a_{3n}x_n & = r_3 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & (1.23) \\
a_{i1}x_1 & +a_{i2}x_2 & +a_{i3}x_3 & +\cdots a_{ij}x_j & \cdots + a_{in}x_n & = r_i & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
a_{m1}x_1 & +a_{m2}x_2 & +a_{m3}x_3 & +\cdots a_{mj}x_j & \cdots + a_{mn}x_n & = r_m &
\end{array}$$

In this equation system there are x_j, with j = 1.....n, n unknown quantities and r_i, with i = 1.....m, m known quantities at the right side. And a_{ij} with i = 1.....m and j = 1.....n are characterized as coefficients. So we have n unknown quantities and m equations.

If $n = m$, the equation system is definitely solvable, and it is **determined**. For $n < m$ it is so called **overdetermined equation system**, and mostly there are approximate solutions which fulfil all equations. In the case n > m several solutions exist, i.e. the equation system is not clearly solvable, and it is **undetermined**. By the introduction of the matrix notation we can write them for short and the rules of the matrix calculus can also be used at the same time.

Thus the above equation systems can be noted in the following form:

$$\mathrm{A} \cdot \mathrm{X} = \mathrm{R} \qquad (1.24)$$

A means coefficient matrix, X the solution vector and R on the right side as column vector:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad X = \begin{bmatrix} x1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

Different methods are used for the solution of such matrix equations or equation systems.

The relatively simple direct solution methods are mostly not treatable, since the construction of the inverse coefficient matrix proves itself complicated with higher rank. We differentiate the solution methods between the direct total step methods and iterative equation solutions. The methods are just as their names imply. By the total step methods the equation system is separated based on algebraic transformations until the equation with only one unknown quantity remaining. In the following some representative examples of both methods are indicated.

1.3.1 Total step method

1.3.1.1 GAUSS Elimination

In the **GAUSS elimination** we try to get an equation with one unknown by successive substitution.

The result of the elimination is the equation system:

$$A \cdot X = R$$

In an equation system with an upper triangle matrix A' and R'

$$A' \cdot X = R'$$

$$\mathrm{A}' = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ 0 & a'_{22} & a'_{23} & \cdots & a'_{2j} & \cdots & a'_{2n} \\ 0 & 0 & a'_{33} & a'_{34} & a'_{3j} & \cdots & a'_{3n} \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & a'_{ij} & a'_{ij+1} & a'_{in} \\ 0 & 0 & 0 & 0 & \ddots & \ddots & a'_{m-1n} \\ 0 & 0 & 0 & 0 & 0 & 0 & a'_{mn} \end{bmatrix} \qquad \mathrm{R}' = \begin{bmatrix} r_1 \\ r'_2 \\ r'_3 \\ \vdots \\ r'_j \\ r'_{n-1} \\ r'_n \end{bmatrix} \tag{1.25}$$

The toppest row, or the toppest equation remains unchanged. The (j - 1)-equations are multiplied by factor fak_j and subtracted from the j-equation. For the second row or the second equation:

$$fak_2 = \frac{a_{i2}}{a_{i1}} \tag{1.26}$$

or in general

$$fak_j = \frac{a'_{ij}}{a'_{ij-1}} \tag{1.27}$$

The lowest row or equation can be solved. This solution is inserted backwards into all other equations. This back substitution generally yields value x_i:

$$x_n = \frac{r'_n}{a'_{mn}} \tag{1.28}$$

$$x_i = \frac{1}{a'_{ii}} \left(r'_i - \sum_{j=i+1}^{n} a'_{ij} x_j \right)$$

Thus the computation of vector x can be achieved.

Example to application of GAUSS elimination:
To solve this system:

$$2x - 3y + 4z = 19$$

$$4x - 4y + 3z = 22$$

$$-6x - y + 5z = 7$$

The coefficients and absolute terms are shown in the following scheme:

x	y	z	1
2	-3	4	19
4	-4	3	22
-6	-1	5	7

The row serving for the elimination (in this example - first row) is marked by the letter *E*. A variable should be eliminated under the help of *E* i.e. how many multiples of *E* should be added on other rows. The added multiple of *E* can be marked beside the corresponding rows:

	x	y	z	1
E	2	-3	4	19
-2	$4-4$	$-4+6$	$3-8$	$22-38$
3	$-6+6$	$-1-9$	$5+12$	$7+57$

In such way the 2nd and 3rd row are similarly changed:

x	y	z	1
0	2	-5	-16
0	-10	17	64

	x	y	z	1
E	0	2	-5	-16
5	0	$-10+10$	$17-25$	$64-80$

The *E*-row and the last row contain the coefficients and absolute terms of the new system:

$$\begin{bmatrix} 2 & -3 & 4 \\ 0 & 2 & -5 \\ 0 & 0 & 8 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 19 \\ -16 \\ 16 \end{bmatrix}$$

$$2x - 3y + 4x = 19$$

$$2y - 5z = 16$$

$$8z = 16$$

The results of successive insertion

$$z = 2, \qquad y = -3, \qquad x = 1$$

1.3.1.2 CRAMER's Rule

According to the CRAMER's rule the solution of matrix equation is achieved by the determinant computations. The elements of the solution vector **X**:

$$x_i = \frac{D_i}{D} \tag{1.29}$$

D_i is for CRAMER's determinant. It is developed from the determinant D = det A of matrix A in consequence of replacing the i-th column by the right side, the vector R. This method only has practical meaning for small matrices or for the matrices which contain many zeros.

Example to application of CRAMER' rule:

Find solutions of linear equation systems (LGS) with CRAMER's rule:

$$2x + y - z = 0$$

$$5x + 2y = 8$$

$$x + y - 2z = -5$$

LGS is written in matrix form as follows:

$$\begin{bmatrix} 2 & 1 & -1 \\ 5 & 2 & 0 \\ 1 & 1 & -2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ -5 \end{bmatrix}$$

or

$$\mathrm{A} \cdot \mathrm{X} = \mathrm{R}$$

with

$$\mathrm{A} = \begin{bmatrix} 2 & 1 & -1 \\ 5 & 2 & 0 \\ 1 & 1 & -2 \end{bmatrix}$$

The determinant D = det A of matrix A can be developed according to second row:

$$D = \begin{vmatrix} 2 & 1 & -1 \\ 5 & 2 & 0 \\ 1 & 1 & -2 \end{vmatrix}$$

$$= -5\begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} + 2\begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix} - 0\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix}$$

$$= -5(-2+1) + 2(-4+1)$$

$$D = -1$$

The coefficient determinants are:

$$D_x = \begin{vmatrix} 0 & 1 & -1 \\ 8 & 2 & 0 \\ -5 & 1 & -2 \end{vmatrix} = -1\begin{vmatrix} 8 & 0 \\ -5 & -2 \end{vmatrix} - 1\begin{vmatrix} 8 & 2 \\ -5 & 1 \end{vmatrix} = -2$$

$$D_y = \begin{vmatrix} 2 & 0 & -1 \\ 5 & 8 & 0 \\ 1 & -5 & -2 \end{vmatrix} = 2\begin{vmatrix} 8 & 0 \\ -5 & -2 \end{vmatrix} - 1\begin{vmatrix} 5 & 8 \\ 1 & -5 \end{vmatrix} = 1$$

$$D_z = \begin{vmatrix} 2 & 1 & 0 \\ 5 & 2 & 8 \\ 1 & 1 & -5 \end{vmatrix} = 2\begin{vmatrix} 2 & 8 \\ 1 & -5 \end{vmatrix} - 1\begin{vmatrix} 5 & 8 \\ 1 & -5 \end{vmatrix} = -3$$

Thus the solution is:

$$x = \frac{D_x}{D} \qquad y = \frac{D_y}{D} \qquad z = \frac{D_z}{D}$$

$$x = 2 \qquad y = -1 \qquad z = 3$$

1.3.1.3 Construction of inverse matrix

The solution vector X of LGS is noted:

$$\begin{aligned} \mathrm{A}\cdot\mathrm{X} &= \mathrm{R} \\ (\mathrm{A}^{-1}\cdot\mathrm{A})\cdot\mathrm{X} &= (\mathrm{A}^{-1}\cdot\mathrm{R}) \\ \mathrm{X} &= (\mathrm{A}^{-1}\cdot\mathrm{R}) \qquad (1.30) \end{aligned}$$

with $(\mathrm{A}^{-1}\cdot\mathrm{A}) = \mathrm{E}$

Example to application of inverse coefficient matrix A^{-1}:

Find solution of the same LGS (see former example - LGS, page 21) with help of inverse matrix.

The inverse matrix of A is (see example - construction of inverse matrix, page 11):

$$\mathrm{A}^{-1} = \begin{bmatrix} 4 & -1 & -2 \\ -10 & 3 & 5 \\ -3 & 1 & 1 \end{bmatrix} \cdot \qquad (1.31)$$

then

$$\mathrm{X} = \begin{bmatrix} 4 & -1 & -2 \\ -10 & 3 & 5 \\ -3 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 8 \\ -5 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

$$x = 2, \qquad y = -1, \qquad z = 3$$

1.3.1.4 LU-Decomposition

The so called LU-decomposition method assumes that a symmetrical matrix A of equation system can be dismantled

$$\mathbf{A \cdot B = R}$$

as product of two matrices L and U:

$$\mathbf{L \cdot U = A} \qquad \textbf{(1.32)}$$

And L (lower) is lower triangle matrix and U (upper) is upper one.

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mn} \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & u_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Thus the matrix equation forms the following figure:

$$\mathrm{A \cdot X = (L \cdot U) \cdot X = L \cdot (U \cdot X) = L \cdot Y = R} \qquad (1.33)$$

So the resultant equation can be decomposed into two equations, which contain the simple solvable triangle matrices. First the equation about Y will be solved. This solution vector Y then serves on the right side for the confirmation of the original solution vector X.

$$\begin{aligned} \mathrm{L \cdot Y = R} \\ \mathrm{U \cdot X = Y} \end{aligned} \qquad (1.34)$$

For the first solution the forward substitution is used:

$$\begin{aligned} y_1 &= \frac{r_1}{l_{11}} \\ y_i &= \frac{1}{l_{ii}} \left(y_i - \sum_{j=1}^{i-1} l_{ij}\, y_j \right) \text{ mit } i = 2, 3, \ldots..n \end{aligned} \qquad (1.35)$$

We can get the second as well as vector X by backward substitution:

$$\begin{aligned} x_n &= \frac{y_n}{u_{mn}} \\ x_i &= \frac{1}{u_{ii}} \left(y_i - \sum_{j=i+1}^{n} u_{ij}\, x_j \right) \text{ mit } i = n-1, n-2, \ldots., 1 \end{aligned} \qquad (1.36)$$

Assume a definition equation for confirmation the elements of L- und U-matrix:

$$\mathbf{L} \cdot \mathbf{U} = \mathbf{A} \qquad \text{or} \tag{1.37}$$

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mn} \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & u_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.38}$$

If these two matrices are multiplied and an element comparison is made, a complicated equation system with $m \cdot n$ unknown quantities appears. The difficulty can be avoided if the main diagonal elements l_{ii} of L-matrix are set to one.

$$[l_{ii}] = 1 \tag{1.39}$$

Then we get the following simple calculation scheme for the elements:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & u_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.40}$$

$$\begin{bmatrix} 1 \cdot u_{11} & 1 \cdot u_{12} & \cdots & 1 \cdot u_{1n} \\ l_{21} \cdot u_{11} & l_{21} \cdot u_{12} + 1 \cdot u_{22} & \cdots & l_{21} \cdot u_{1n} + 1 \cdot u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} \cdot u_{11} & l_{m1} \cdot u_{12} + l_{m2} \cdot u_{22} & \cdots & l_{m1} \cdot u_{1n} + \cdots \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.41}$$

Example of solving an equation system with LU-decomposition method:

$$-3x_1 + 2x_2 - 3x_3 = 6$$

$$9x_1 - 2x_2 + 10x_3 = -10$$

$$6x_1 + 8x_2 + 14x_3 = 22$$

$$\Longrightarrow \qquad \mathrm{A} = \begin{bmatrix} -3 & 2 & -3 \\ 9 & -2 & 10 \\ 6 & 8 & 14 \end{bmatrix}$$

with $l_{ii} = 1$

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{21} & 1 & 0 \\ \alpha_{31} & \alpha_{32} & 1 \end{bmatrix} \qquad \mathbf{U} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ 0 & \beta_{22} & \beta_{23} \\ 0 & 0 & \beta_{33} \end{bmatrix}$$

$$\mathbf{L} \cdot \mathbf{U} = \mathbf{A}$$

We know from the element comparison:

$$1 \cdot \beta_{11} + 0 \cdot 0 + 0 \cdot 0 = -3 \quad \Longrightarrow \quad \beta_{11} = -3$$

$$\alpha_{21} \cdot \beta_{11} + 1 \cdot 0 + 0 \cdot 0 = 9 \quad \Longrightarrow \quad \alpha_{21} = -3$$

$$\alpha_{31} \cdot \beta_{11} + \alpha_{32} \cdot 0 + 1 \cdot 0 = 6 \quad \Longrightarrow \quad \alpha_{31} = -2$$

$$1 \cdot \beta_{12} + 0 \cdot \beta_{22} + 0 \cdot 0 = 2 \quad \Longrightarrow \quad \beta_{12} = 2$$

$$\alpha_{21} \cdot \beta_{12} + 1 \cdot \beta_{22} + 0 \cdot 0 = -2 \quad \Longrightarrow \quad \beta_{22} = 4$$

$$\alpha_{31} \cdot \beta_{12} + \alpha_{32} \cdot \beta_{22} + 1 \cdot 0 = 8 \quad \Longrightarrow \quad \alpha_{32} = 3$$

By the same way we find:

$$1 \cdot \beta_{13} + 0 \cdot \beta_{23} + 0 \cdot \beta_{33} = -3 \quad \Longrightarrow \quad \beta_{13} = -3$$

$$\alpha_{21} \cdot \beta_{13} + 1 \cdot \beta_{23} + 0 \cdot \beta_{33} = 10 \quad \Longrightarrow \quad \beta_{23} = 1$$

$$\underbrace{\alpha_{31} \cdot \beta_{13}}_{6} + \alpha_{32} \cdot \beta_{23} + 1 \cdot \beta_{33} = 14 \quad \Longrightarrow \quad \beta_{33} = 5$$

Then the triangle matrices with following figures:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & 3 & 1 \end{bmatrix} \qquad \mathbf{U} = \begin{bmatrix} -3 & 2 & -3 \\ 0 & 4 & 1 \\ 0 & 0 & 5 \end{bmatrix}$$

The solution vector X can be computed:
In general:

$$\mathbf{A} \cdot \mathbf{X} = \mathbf{R}$$

$$\mathbf{A} = \mathbf{L} \cdot \mathbf{U}$$

$$\mathbf{L} \cdot \underbrace{\mathbf{U} \cdot \mathbf{X}}_{Y} = \mathbf{R}$$

$$\mathbf{L} \cdot \mathbf{Y} = \mathbf{R}$$

$$1 \cdot y_1 + 0 \cdot y_2 + 0 \cdot y_3 = 6$$

$$-3 \cdot y_1 + 1 \cdot y_2 + 0 \cdot y_3 = -10$$

$$-2 \cdot y_1 + 3 \cdot y_2 + 1 \cdot y_3 = 22$$

$$y_1 = 6 \qquad \mathbf{Y} = \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix}$$

$$y_2 = 8$$

$$y_3 = 10$$

$$\mathbf{U} \cdot \mathbf{X} = \mathbf{Y}$$

$$-3 \cdot x_1 + 2 \cdot x_2 - 3 \cdot x_3 = 6$$

$$0 \cdot x_1 + 4 \cdot x_2 + 1 \cdot x_3 = 8$$

$$0 \cdot x_1 + 0 \cdot x_2 + 5 \cdot x_3 = 10$$

$$x_1 = -3$$

$$x_2 = \frac{3}{2}$$

$$x_3 = 2$$

$$\mathrm{X} = \begin{bmatrix} -3 \\ \frac{3}{2} \\ 2 \end{bmatrix}$$

1.3.1.5 CHOLESKY method

With the CHOLESKY method the solution of the matrix equation for the special case of the symmetrical coefficient matrix can be traced back by the solution of two subsystems, as the coefficient matrix is decomposed into an upper and a lower triangle matrix. This dismantling is also called decomposition.

The CHOLESKY method is not generally applicable, and it presupposes that the coefficient matrix A of equation system must be symmetrical i.e. $\mathbf{A} = \mathbf{A}^T$, and positive definite.

$$\mathbf{A} \cdot \mathbf{X} = \mathbf{R}$$

Positive definite means that all elements of the main diagonals must be greater than zero $a_{ii} > 0$, for example simulation of groundwater flow in the quantization methods (FDM, FEM or FVM).

In the CHOLESKY method the symmetrical matrix A of equation system is written as product of two matrices, a lower triangle matrix B and an upper B^T, which is equal to the transpose of the lower,

$$\mathrm{B} \cdot \mathrm{B}^{\mathrm{T}} = \mathrm{A} \qquad (1.42)$$

B is an upper triangle matrix, whose elements $b_{ik} = 0$ if i > k. The equation system:

$$\mathrm{A} \cdot \mathrm{X} = \mathrm{R}$$

$$\mathrm{B} \cdot \mathrm{B}^{T} \cdot \mathrm{X} = \mathrm{R}$$

We set

$$\mathrm{B}^{T} \cdot \mathrm{X} = \mathrm{Y} \qquad (1.43)$$

and determine the elements of B by element comparison:

$$\mathrm{B} \cdot \mathrm{B}^{T} = \mathrm{A}$$

so Y can be computed:

$$\mathrm{B} \cdot \mathrm{Y} = \mathrm{R} \qquad (1.44)$$

The solution of X results from the back calculation according to equation 1.43.

Generally the following algorithm can be indicated for the computation of the elements of B:

$$b_{kj} = \begin{cases} \left(a_{kj} - \sum_{l=1}^{k-1} b_{lj} b_{lk} \right) \frac{1}{b_{kk}} & \text{für } k+1 < j < n, \qquad j = 2 \text{ bis } n \\ 0 & \text{für } k > j \end{cases} \tag{1.45}$$

$$b_{jj} = \sqrt{a_{jj} - \sum_{l=1}^{j-1} b_{lj}^2} \qquad \text{für } j = 1 \text{ bis } n$$

$$y_j = \left(r_j - \sum_{l=1}^{j-1} b_{lj} y_l \right) \frac{1}{b_{jj}} \qquad \text{für } j = 1 \text{ bis } n$$

CHOLESKY method possesses some advantages compared with the Gauss procedure. Thus e.g. the method is characterised by the fact that it works numerically very stably, since the dominance of the main diagonals is strengthened by extraction of the square root from very small elements. If the coefficient matrix A possesses a band structure, this is also transferred to the triangle matrix. The algorithm is independent of the values on the right side. Thus the solution equation system can be repeated with small expenditure for different values on the right side (boundary and initial values), which makes variant calculations very effective.

For an equation system with three equations and three unknown quantities:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = r_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = r_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = r_3$$

$$\mathrm{A} \cdot \mathrm{X} = \mathrm{R}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

We must check whether the conditions, the symmetrical coefficient matrix ($\mathbf{A} = \mathbf{A}^T$) and positive definite ($a_{ii} > 0$), are given, before CHOLESKY method may be used.

The equation system can be written also in the following form.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

According to regulation for the CHOLESKY method the triangle matrix B is introduced and the pertinent transpose is formed:

$$\mathrm{B} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \tag{1.46}$$

$$\mathrm{B}^T = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix} \tag{1.47}$$

$$\mathrm{B} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{12} & b_{22} & 0 \\ b_{13} & b_{23} & b_{33} \end{bmatrix}$$

According to equation 1.42:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{12} & b_{22} & 0 \\ b_{13} & b_{23} & b_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix}$$

The determination of the elements of the matrix B takes place according to multiplication of $\mathbf{B} \cdot \mathbf{B}^T$ via an element comparison with the matrix A:

$$
\begin{aligned}
b_{11} \cdot b_{11} \quad & +0 \cdot 0 \quad & +0 \cdot 0 \quad & = a_{11} \\
b_{11} \cdot b_{12} \quad & +0 \cdot b_{12} \quad & +0 \cdot 0 \quad & = a_{12} \\
b_{11} \cdot b_{13} \quad & +0 \cdot b_{23} \quad & +0 \cdot b_{33} \quad & = a_{13} \\
b_{12} \cdot b_{11} \quad & +b_{22} \cdot 0 \quad & +0 \cdot 0 \quad & = a_{12} \\
b_{12} \cdot b_{12} \quad & +b_{22} \cdot b_{22} \quad & +0 \cdot 0 \quad & = a_{22} \\
b_{12} \cdot b_{13} \quad & +b_{22} \cdot b_{23} \quad & +0 \cdot b_{33} \quad & = a_{23} \\
b_{13} \cdot b_{11} \quad & +b_{23} \cdot 0 \quad & +b_{33} \cdot 0 \quad & = a_{13} \\
b_{23} \cdot b_{12} \quad & +b_{23} \cdot b_{22} \quad & +b_{33} \cdot 0 \quad & = a_{23} \\
b_{33} \cdot b_{13} \quad & +b_{23} \cdot b_{23} \quad & +b_{33} \cdot b_{33} \quad & = a_{33}
\end{aligned}
$$

We recognize that some equations are redundant in the developed equation system due to the symmetry characteristics of the coefficient matrix. Thus only six of these equations are needed for the determination of the matrix B.

$$
\begin{aligned}
& a_{11} = b_{11} \cdot b_{11} && \Longrightarrow && b_{11} = \sqrt{a_{11}} \\
& a_{12} = b_{11} \cdot b_{12} && \Longrightarrow && b_{12} = \frac{a_{12}}{b_{11}} = \frac{a_{12}}{\sqrt{a_{11}}} \\
& a_{13} = b_{11} \cdot b_{13} && \Longrightarrow && b_{13} = \frac{a_{13}}{b_{11}} = \frac{a_{13}}{\sqrt{a_{11}}} \\
& a_{22} = b_{12}^2 + b_{22}^2 && \Longrightarrow && b_{22} = \sqrt{a_{22} - b_{12}^2} = \sqrt{a_{22} - \frac{a_{12}^2}{a_{11}}} \\
& a_{23} = b_{12} \cdot b_{13} + b_{22} \cdot b_{23} && \Longrightarrow && b_{23} = \frac{a_{23} - b_{12} \cdot b_{13}}{b_{22}} = \frac{a_{23} - \frac{a_{12} \cdot a_{13}}{a_{11}}}{\sqrt{a_{22} - \frac{a_{12}^2}{a_{11}}}} \\
& a_{33} = b_{13}^2 + b_{23}^2 + b_{33}^2 && \Longrightarrow && \begin{cases} b_{33} = \sqrt{a_{33} - b_{13}^2 - b_{23}^2} \\ \quad = \sqrt{a_{33} - \frac{a_{13}^2}{a_{11}} - \frac{\left(a_{23} - \frac{a_{12} \cdot a_{13}}{a_{11}}\right)^2}{a_{22} - \frac{a_{12}^2}{a_{11}}}} \end{cases}
\end{aligned}
\qquad (1.48)
$$

After the matrix B and its transpose B^T were determined, the auxiliary matrix Y from the equation 1.44 can be computed:

$$\mathbf{B \cdot Y = R}$$

$$\begin{bmatrix} b_{11} & 0 & 0 \\ b_{12} & b_{22} & 0 \\ b_{13} & b_{23} & b_{33} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

$$\begin{aligned} b_{11} \cdot y_1 = r_1 \quad &\Longrightarrow \quad y_1 = \frac{r_1}{b_{11}} \\ b_{12} \cdot y_1 + b_{22} \cdot y_2 = r_2 \quad &\Longrightarrow \quad y_2 = \frac{r_2 - b_{12} \cdot y_1}{b_{22}} \\ b_{13} \cdot y_1 + b_{23} \cdot y_2 + b_{33} \cdot y_3 = r_3 \quad &\Longrightarrow \quad y_3 = \frac{r_3 - b_{13} \cdot y_1 - b_{23} \cdot y_2}{b_{33}} \end{aligned} \tag{1.49}$$

With known matrix B and the auxiliary matrix Y the solution of equation system X can be computed now by means of equation 1.43:

$$\mathbf{B}^T \cdot \mathbf{X} = \mathbf{Y}$$

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$\begin{aligned} b_{33} \cdot x_3 = y_3 \quad &\Longrightarrow \quad x_3 = \frac{y_3}{b_{33}} \\ b_{22} \cdot x_2 + b_{23} \cdot x_3 = y_2 \quad &\Longrightarrow \quad x_2 = \frac{y_2 - b_{23} \cdot x_3}{b_{22}} \\ b_{11} \cdot x_1 + b_{12} \cdot x_2 + b_{13} \cdot x_3 = y_1 \quad &\Longrightarrow \quad x_1 = \frac{y_1 - b_{12} \cdot x_2 - b_{13} \cdot x_3}{b_{11}} \end{aligned} \tag{1.50}$$

Attention:

The equations 1.48 to 1.50 can be accordingly applied for the determination of the elements of the matrix B, the auxiliary matrix Y, and the solution vector X in all equation systems with three rows and three unknown quantities, if they fulfil the conditions for the CHOLESKY method. For each case the elements of the coefficient matrix and those on the right side must be accordingly used. These algorithms can be extended easily to any size of equation system.

Examples of CHOLESKY method application:

1. The method is to be demonstrated exemplary in the equation system, which can be used in other cases:

$$2x + y - z = 0$$

$$5x + 2y = 8$$

$$x + y - 2z = -5$$

Here however the prerequisites for the application of the CHOLESKY method, positive definite ($a_{ii} > 0$) and the symmetry ($A = A^T$), are not given, thus this method is not applicable ($a_{33} = -2$) und $A \neq A^T$.

2. As the second example the following equation system is given:

$$9x_1 + 2x_2 + 3x_3 = 6$$

$$2x_1 + 8x_2 + 4x_3 = -10$$

$$3x_3 + 4x_2 + 10x_3 = 22$$

With $A \cdot X = R$

$$\begin{bmatrix} 9 & 2 & 3 \\ 2 & 8 & 4 \\ 3 & 4 & 10 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ -10 \\ 22 \end{bmatrix}$$

According to equation 1.42:

$$\begin{bmatrix} 9 & 2 & 3 \\ 2 & 8 & 4 \\ 3 & 4 & 10 \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{12} & b_{22} & 0 \\ b_{13} & b_{23} & b_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ 0 & 0 & b_{33} \end{bmatrix}$$

The determination of the elements of the matrix B takes place according to multiplication of $\mathbf{B} \cdot \mathbf{B}^T$ via an element comparison. Since only six unknown elements must be determined, only six of these equations are needed:

$$
\begin{aligned}
9 &= b_{11} \cdot b_{11} && \Longrightarrow && b_{11} = 3 \\
2 &= b_{11} \cdot b_{12} && \Longrightarrow && b_{12} = \frac{2}{3} \\
3 &= b_{11} \cdot b_{13} && \Longrightarrow && b_{13} = 1 \\
8 &= b_{12}^2 + b_{22}^2 && \Longrightarrow && b_{22} = \sqrt{8 - \frac{4}{9}} = \frac{\sqrt{60}}{3} \\
4 &= b_{12} \cdot b_{13} + b_{22} \cdot b_{23} && \Longrightarrow && b_{23} = \frac{4 - \frac{2}{3} \cdot 1}{\frac{\sqrt{60}}{3}} = \frac{10}{\sqrt{60}} \\
10 &= b_{13}^2 + b_{23}^2 + b_{33}^2 && \Longrightarrow && b_{33} = \sqrt{10 - 1 - \frac{100}{60}} = \frac{44}{6}
\end{aligned}
$$

After the matrix B and its Transpose B^T were determined, the auxiliary matrix Y can be computed according to equation 1.44:

$$\mathbf{B} \cdot \mathbf{Y} = \mathbf{R}$$

$$
\begin{bmatrix}
3 & 0 & 0 \\
\frac{2}{3} & \frac{\sqrt{60}}{3} & 0 \\
1 & \frac{10}{\sqrt{60}} & \frac{44}{6}
\end{bmatrix}
\cdot
\begin{bmatrix}
y_1 \\ y_2 \\ y_3
\end{bmatrix}
=
\begin{bmatrix}
6 \\ -10 \\ 22
\end{bmatrix}
$$

$$
\begin{aligned}
3 \cdot y_1 = 6 \quad &\Longrightarrow \quad y_1 = \frac{6}{3} = 2 \\
\frac{2}{3} \cdot y_1 + \frac{\sqrt{60}}{3} \cdot y_2 = -10 \quad &\Longrightarrow \quad y_2 = \frac{-10 - \frac{2}{3} \cdot 2}{\frac{\sqrt{60}}{3}} = \frac{-34}{\sqrt{60}} \\
1 \cdot y_1 + \frac{10}{\sqrt{60}} \cdot y_2 + \frac{44}{6} \cdot y_3 = 22 \quad &\Longrightarrow \quad y_3 = \frac{22 - 1 \cdot 2 - \frac{10}{\sqrt{60}} \cdot \frac{-34}{\sqrt{60}}}{\frac{44}{6}} = \frac{558}{440} = \frac{297}{220}
\end{aligned}
$$

With the known matrix B and the auxiliary matrix Y the solution of the equation system X can be computed now by means of equation 1.43:

$$\mathbf{B}^T \cdot \mathbf{X} = \mathbf{Y}$$

$$\begin{bmatrix} 3 & \frac{2}{3} & 1 \\ 0 & \frac{\sqrt{60}}{3} & \frac{10}{\sqrt{60}} \\ 0 & 0 & \frac{44}{6} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{-34}{\sqrt{60}} \\ \frac{297}{220} \end{bmatrix}$$

$$\begin{aligned} \frac{44}{6} \cdot x_3 = \frac{297}{220} \quad &\Longrightarrow \quad x_3 = \frac{\frac{297}{220}}{\frac{44}{6}} = \frac{837}{4840} \\ \frac{\sqrt{60}}{3} \cdot x_2 + \frac{10}{\sqrt{60}} \cdot x_3 = \frac{-34}{\sqrt{60}} \quad &\Longrightarrow \quad x_2 = \frac{\frac{-34}{\sqrt{60}} - \frac{10}{\sqrt{60}} \cdot \frac{837}{4840}}{\frac{\sqrt{60}}{3}} \\ 3 \cdot x_1 + \frac{2}{3} \cdot x_2 + 1 \cdot x_3 = 2 \quad &\Longrightarrow \quad x_1 = \frac{y_1 - b_{12} \cdot x_2 - b_{13} \cdot x_3}{b_{11}} \end{aligned}$$

1.3.1.6 Task of solving equation systems

Determine the solution vector X in five ways with the following equation system ($A \cdot X = R$)

- GAUSS elimination,
- CRAMER's rule
- $A^{-1} \cdot R$
- LU-decomposition
- CHOLESKY method

a)
$$\frac{x-2}{3} - \frac{y+2}{2} = \frac{x-2y}{5}$$
$$\frac{x-y}{6} + \frac{3y+2}{4} = \frac{x-2(y-1)}{3}$$

b)
$$3x + y - z = 2$$
$$2x - y + 4z = 0$$
$$x + 5y - 2z = 1$$

c)
$$\begin{bmatrix} 2 & 0 & -1 \\ 2 & 4 & -1 \\ -1 & 8 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

d)
$$x + y + z = 3$$
$$2x + 4y + 8z = 13$$
$$3x + 9y + 27z = 34$$

e)
$$2x - 4y + 9w = 25$$
$$3x - 3y - 3z + 11w = 27$$
$$4x + 6y - 15z + 5w = -5$$
$$3x + y - 4z + 12w = 32$$

1.3.2 Iterative method

In the iterative method firstly an approximate solution is assumed for equation system. This solution will be inserted into the system, and by means of optimisation method the components of solution vector are best adapted. After *n* iteration steps the approximation will approach to the accurate solution with a residue. The iterative methods play a dominating role for large equation systems, since they are usually substantially faster than the directly solving equation.

Further common applications are the **CG method (Conjugate gradient Method)** and the **multi-grid method**. These methods are particularly subject to further development in connection with applications of simulation in field problems. In the CG method the use of additional preconditioning became generally accepted in recent years, with which the search strategy of the iterative optimisation steps is specified. Generally the kind of the preconditioning substantially determines the optimisation speed or the number of optimisation steps.

We assume the general matrix notation of a linear equation system as below; X is designated as unknown solution vector,

$$\mathbf{A \cdot X = R} \qquad \mathbf{(1.51)}$$

So we can insert a known approximation solution X + δX into this equation. This yields an unknown δR on the right side, which deviates from the given value.

$$\mathrm{A} \cdot (\mathrm{X} + \delta\mathrm{X}) = \mathrm{R} + \delta\mathrm{R} \qquad (1.52)$$

If we subtract equation 1.51 from equation 1.52, then:

$$\mathrm{A} \cdot \delta\mathrm{X} = \delta\mathrm{R} \qquad (1.53)$$

Or with above equation 1.52:

$$\mathrm{A} \cdot \delta\mathrm{X} = \mathrm{A} \cdot (\mathrm{X} + \delta\mathrm{X}) - \mathrm{R} \qquad (1.54)$$

The right side of this equation is known, since X + δX is the approximation solution. The goal now is to make the right side equal to zero while a new δX will be found. This can be done via solving the matrices equation (see section 1.3.1, page 17) or by means of purposeful optimisation, e.g. with the CG method.

Generally **the CG method** can be well used for linear, square, symmetrical matrices (*m* = *n*). The basic idea is to minimize a function.

$$f(x) = \frac{1}{2}\mathrm{X} \cdot \mathrm{A} \cdot \mathrm{X} - \mathrm{b} \cdot \mathrm{X}$$

The function possesses a minimum if the gradient (see section 2.2) is equal to zero:

$$\nabla f(x) = \mathrm{A} \cdot \mathrm{X} - \mathrm{R} \Rightarrow 0 \qquad (1.55)$$

This minimum can be found, if we formulate a function $f(x_k + \alpha_k p_k)$ by means of a search direction p_k. The index k means the number of continuous search loop.

1.3.3 Overdetermined equation system (m > n)

In contrary to the methods described so far solutions for overdetermined equation system are to be demonstrated here. In this case there are more equations than unknown quantity, i.e., *m* is larger than *n* (*m* > *n*). This occurs when mathematical models are to be adapted to measured data. A typical case is thereby the application of the quantized (discrete) faltung integral (see section 12.3 faltung integral, page 355).

A usual method is to regard this task as optimisation, and we try to make that, the free parameters, i.e. the solution vector X adapting to the measured values. Thereby most methods differ in the conditioning of the optimisation problem and in the choice of the optimisation strategy.

The **SVD method (Singular Value Decomposition)** proceeds in the following way. The matrix equation is given:

$$\mathrm{A} \cdot \mathrm{X} = \mathrm{R} \qquad \text{bzw.} \qquad [a_{ij}] \cdot [x_i] = [b_i] \qquad m > n$$

In this case the coefficient matrix A can be decomposed into

$$\mathrm{A} = \mathrm{U} \cdot [w_{ii}] \cdot \mathrm{V}^T \qquad (1.56)$$

Whereby the matrix U has the same figure as A, w_{ii} a square diagonal matrix with rank n and V^T is a transpose with rank n. This decomposition employed on the equation above and solve according to solution vector, shows that:

$$\mathrm{X} = \mathrm{V} \cdot \left[\mathrm{diag} \left(\frac{1}{w_j} \right) \right] \cdot \mathrm{U}^T \cdot \mathrm{R} \qquad (1.57)$$

This equation can be solved with the HOUSEHOLDER routine.

Chapter 2

2 Vector algebra and analysis

On the basis of simple, well-known representations of vector calculus the basic rules of the vector algebra are specified. Subsequently, the rules of vector differentiation with descriptive examples are discussed.

2.1 Unit vector

Different unit vectors of vector representations are dependent on the use of coordinate system. Then the vector $\vec{a}$ can be expressed the sum of multiples of unit vector. The unit vectors possess the length (modulus) one | *e*| = 1, and are always parallel to the coordinate system axles.

For the practical work in water management three coordinate systems, the Cartesian, the cylindrical and the spherical, are generally used. The same unit vector $\vec{a}$ can be described in table 2.1 (also see figure 2.2 and 2.1).

Table 2.1: coordinate system of vector representation

Koordinaten-system	Einheits-vektoren	Vektor $\vec{a}$
Kartesisch	$\vec{i}, \vec{j}, \vec{k}$	$\vec{a} = a_x\vec{i} + a_y\vec{j} + a_z\vec{k}$
Zylindrisch	$\vec{r}, \vec{\phi}, \vec{z}$	$\vec{a} = a_r\vec{r} + a_\phi\vec{\phi} + a_z\vec{z}$
Sphärisch	$\vec{r}, \vec{\theta}, \vec{\phi}$	$\vec{a} = a_r\vec{r} + a_\theta\vec{\theta} + a_\delta\vec{\phi}$

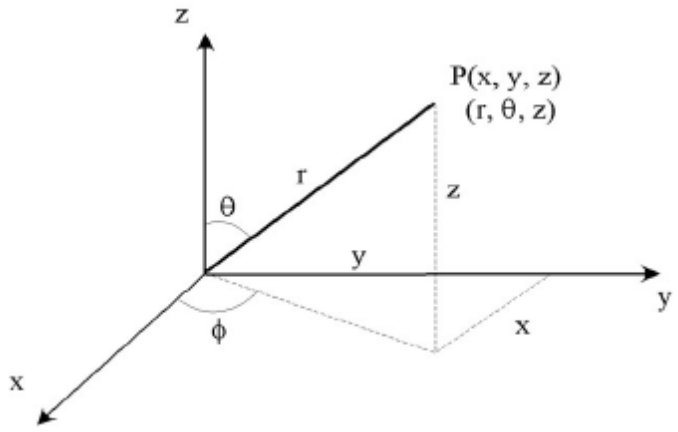


Figure 2.1: vector representation in Cartesian coordinate system

In two-dimensional space polar coordinate system will be used (see Figure 2.3).

Since the vector $\vec{a}$ is independent of the used coordinate system, the following conversion is applicable between the Cartesian and the polar coordinate system:

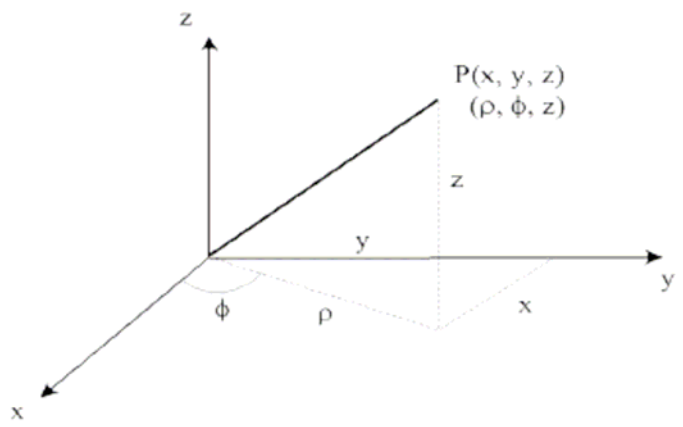


Figure 2.2: vector representation in spherical coordinate system

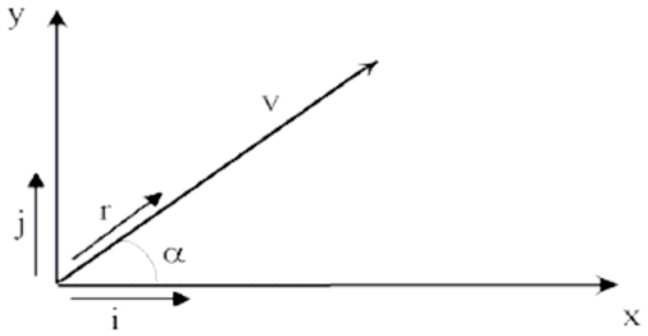


Figure 2.3: vector representation in two-dimensional space

$$\begin{aligned} a_r &= \sqrt{a_x^2 + a_y^2} = |\vec{a}| \\ a_\alpha &= \arctan\left(\frac{a_y}{a_x}\right) \\ a_x &= \tan(a_\alpha) \cdot a_y \\ a_x &= \cos(a_\alpha) \cdot |\vec{a}| \end{aligned} \tag{2.1}$$

2.2 Calculation rules

In the following some important basic arithmetic rules for vectors are to be demonstrated by examples in the Cartesian coordinate system.

• **Addition**
The arguments of the Cartesian unit vectors are respectively added in the vector addition:

$$\vec{a} + \vec{b} = (a_x + b_x)\,\vec{i} + (a_y + b_y)\,\vec{j} + (a_z + b_z)\,\vec{k} \qquad (2.2)$$

Notice:
This relationship applies **only** to the Cartesian coordinate system and can **not** be transferred to other coordinate systems.

In the vector algebra the following laws apply:

commutative law $$\vec{A} + \vec{B} = \vec{B} + \vec{A} \qquad (2.3)$$

distributive law $$m\left(n\vec{A}\right) = (mn)\,\vec{A} = n\left(m\vec{A}\right) \qquad (2.4)$$

distributive law $$(m+n)\,\vec{A} = m\vec{A} + n\vec{A} \qquad (2.5)$$

distributive law $$m\left(\vec{A} + \vec{B}\right) = m\vec{A} + m\vec{B} \qquad (2.6)$$

associative law $$\vec{A} + \left(\vec{B} + \vec{C}\right) = \left(\vec{A} + \vec{B}\right) + \vec{C} \qquad (2.7)$$

• **Modulus**

The modulus of a vector is equal to its length and thus a scalar, which is direction-independent:

$$|\vec{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2} \qquad (2.8)$$

In particular it applies that the modulus of the unit vectors is equal to one:

$$\left|\vec{i}\right| = \left|\vec{j}\right| = \left|\vec{k}\right| = |\vec{r}| = |\vec{\alpha}| = \left|\vec{\delta}\right| = 1 \qquad (2.9)$$

• **Product**

We differentiate two kinds of products with respect to the vector algebra, the scalar product (point product) and the vector product (cross product).

The **scalar product** between two vectors is defined:

$$\vec{a} \cdot \vec{b} = |\vec{a}| \cdot \left|\vec{b}\right| \cdot \cos\left(\vec{a}; \vec{b}\right) \tag{2.10}$$

Hence the scalar product between two vectors is equal to zero, if they stand perpendicularly to each other. In particular it applies that the scalar product of a vector with itself, i.e. the square, is equal to the square of the modulus:

$$\vec{a} \cdot \vec{b} = \begin{cases} 0 & \vec{a} \perp \vec{b} \\ |\vec{a}| \cdot \left|\vec{b}\right| & \vec{a} \uparrow\uparrow \vec{b} \\ -|\vec{a}| \cdot \left|\vec{b}\right| & \vec{a} \uparrow\downarrow \vec{b} \\ |\vec{a}| \cdot \left|\vec{b}\right| \cdot \cos\left(\vec{a}; \vec{b}\right) & \text{beliebig} \end{cases} \tag{2.11}$$

Particularly for the unit vectors:

$$\begin{aligned} &\vec{i} \cdot \vec{j} = 0; \quad \vec{i} \cdot \vec{k} = 0; \quad \vec{j} \cdot \vec{k} = 0; \quad \vec{r} \cdot \vec{\alpha} = 0; \quad \vec{r} \cdot \vec{z} = 0; \\ &\vec{i} \cdot \vec{i} = 1; \quad \vec{j} \cdot \vec{j} = 1; \quad \vec{k} \cdot \vec{k} = 1; \quad \vec{r} \cdot \vec{r} = 1; \quad \vec{\alpha} \cdot \vec{\alpha} = 1; \quad \vec{z} \cdot \vec{z} = 1; \end{aligned} \tag{2.12}$$

The formation of the scalar product in Cartesian coordinate in following way according to above rules:

$$\begin{aligned} \vec{a} \cdot \vec{b} &= \left(a_x \vec{i} + a_y \vec{j} + a_z \vec{k}\right) \cdot \left(b_x \vec{i} + b_y \vec{j} + b_z \vec{k}\right) \\ &= a_x b_x + a_y b_y + a_z b_z \end{aligned} \tag{2.13}$$

From this and the equation stated above the angle between two vectors:

$$\cos\left(\vec{a}; \vec{b}\right) = \frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\sqrt{b_x^2 + b_y^2 + b_z^2}} \tag{2.14}$$

The **vector product** between two vectors yields a vector:

$$\vec{a} \times \vec{b} = \vec{v} \tag{2.15}$$

its modulus is equal to stretching parallelogram by $\vec{a}$ and $\vec{b}$:

$$|\vec{v}| = \left|\vec{a} \times \vec{b}\right| = |\vec{a}| \cdot \left|\vec{b}\right| \cdot \sin\left(\vec{a}; \vec{b}\right)$$

and its direction stands perpendicularly to $\vec{a}$ and $\vec{b}$:

$$\vec{v} \perp \vec{a}$$

$$\vec{v} \perp \vec{b}$$

in general:

$$\left|\vec{a} \times \vec{b}\right| = \begin{cases} 0 & \vec{a} \| \vec{b} \\ |\vec{a}| \cdot \left|\vec{b}\right| & \vec{a} \perp \vec{b} \\ -|\vec{a}| \cdot \left|\vec{b}\right| & \vec{b} \perp \vec{a} \\ |\vec{a}| \cdot \left|\vec{b}\right| \cdot \sin\left(\vec{a}; \vec{b}\right) & \text{beliebig} \end{cases} \tag{2.16}$$

For the Cartesian coordinate system applies:

$$\overrightarrow{a} \times \overrightarrow{b} = \begin{bmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{bmatrix} \tag{2.17}$$

Especially for unit vector:

$$\begin{array}{lllll} \left|\vec{i} \times \vec{j}\right| = 1; & \left|\vec{i} \times \vec{k}\right| = 1; & \left|\vec{j} \times \vec{k}\right| = 1; & |\vec{r} \times \vec{\alpha}| = 1; & |\vec{r} \times \vec{z}| = 1; \\ \vec{i} \times \vec{j} = \vec{k} & \vec{i} \times \vec{k} = \vec{j} & \vec{j} \times \vec{k} = \vec{i} & \vec{r} \times \vec{\alpha} = \vec{z} & \vec{r} \times \vec{z} = \vec{\alpha} \\ \left|\vec{i} \times \vec{i}\right| = 0; & \left|\vec{j} \times \vec{j}\right| = 0; & \left|\vec{k} \times \vec{k}\right| = 0; & |\vec{r} \times \vec{r}| = 0; & |\vec{\alpha} \times \vec{\alpha}| = 0; \quad |\vec{z} \times \vec{z}| = 0; \end{array} \tag{2.18}$$

Attention:

For the **vector product commutative law is not applicable**, but:

$$\vec{a} \times \vec{b} = -\vec{b} \times \vec{a} \tag{2.19}$$

In contrast however the **distributive law** applies**.**

$$\vec{a} \times \left(\vec{b} + \overrightarrow{c}\right) = \vec{a} \times \vec{b} + \vec{a} \times \overrightarrow{c} \tag{2.20}$$

- **Differentiation**

In the vector analysis we speak of three different kinds of the differentiation, **gradient** (grad), **divergence** (div) and **rotation** (rot). For the all three methods a uniform differential vector, NABLA-**Operator** ∇ applies (see table 2.2). Table 2.3 shows the ways of writing of the different kinds of differentiation in the overview as a function of the used coordinate system. For further simplification the LAPLACE differential operator Δ can also be used as the way of writing. This is double application of the NABLA-**Operator**

$$\Delta = \nabla \cdot \nabla \tag{2.21}$$

Table 2.2: Description of NABLA-**Operator** in different coordinate systems

Koordinatensystem		
kartesisch	zylindrisch	sphärisch
$\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}$	$\nabla = \frac{\partial}{\partial r}\vec{r} + \frac{1}{r}\cdot\frac{\partial}{\partial \varphi}\vec{\varphi} + \frac{\partial}{\partial z}\vec{z}$	$\nabla = \frac{\partial}{\partial r}\vec{r} + \frac{1}{r\cdot\sin\theta}\cdot\frac{\partial}{\partial \varphi}\vec{\varphi} + \frac{1}{r}\cdot\frac{\partial}{\partial \theta}\vec{\theta}$

In the **gradient formation**

$$\nabla\varphi = \operatorname{grad}\varphi \tag{2.22}$$

$$Skalar\ \varphi \Longrightarrow Vektor\ \nabla\varphi$$

$$\nabla\varphi = \left(\frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}\right)\varphi = \frac{\partial\varphi}{\partial x}\vec{i} + \frac{\partial\varphi}{\partial y}\vec{j} + \frac{\partial\varphi}{\partial z}\vec{k}$$

the NABLA-operator is applied to a scalar potential field φ. The result of the gradient formation is a vector. The gradient formation can be regarded as the formal multiplication of the NABLA-operator with a scalar quantity. In the field of the hydrogeology this quantity can be groundwater level *h*, temperature fields *T*, concentration distributions *C*, evaporation or groundwater

regeneration rates v_N and others. These scalar quantities (potentials) are nondirectional and have thereby no vector character. However they are location dependent. The most important application of the gradient formation is the DARCY law for the computation of the groundwater flow velocity (see section 7.1, page 184).

$$\vec{v} = -k \ \text{grad} \ h \tag{2.23}$$

Example of gradient formation application:

The groundwater level of an aquifer is indicated by function:

$$h = 2xy - 3x + 2$$

We compute the groundwater flow speed, if the permeability coefficient of the aquifer is $k = 2 \cdot 10^{-3}\ \text{m} \cdot \text{s}^{-1}$

It applies:

$$\begin{aligned}\vec{v} &= -k \ \text{grad}\,(h) \\ &= -2 \cdot 10^{-3} \left(\frac{\partial\,(2xy - 3x + 2)}{\partial x} \vec{i} + \frac{\partial\,(2xy - 3x + 2)}{\partial y} \vec{j} + \frac{\partial\,(2xy - 3x + 2)}{\partial z} \vec{k} \right) \frac{m}{s} \\ &= (6 - 4y)\, 10^{-3} \frac{m}{s} \cdot \vec{i} - 4 \cdot x \cdot 10^{-3} \frac{m}{s} \cdot \vec{j}\end{aligned}$$

It is be recognized that

a) there is no vertical stream

b) the speed is dependent on the coordinates. The current in the aquifer is thus not constant.

We understand the application of NABLA-Operator on a vector by **divergence**:

$$\nabla \vec{v} = \text{div}\ \vec{v} \qquad \text{Vektor } \vec{v} \Longrightarrow \text{Skalar} \tag{2.24}$$

$$\nabla \vec{v} = \left(\frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k} \right) \left(v_x \vec{i} + v_y \vec{j} + v_z \vec{k} \right) = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}$$

The result of divergence formation is a scalar quantity. The divergence can be regarded as the formal application of the scalar product formation between the NABLA-Operator and a vector. According to the rule of scalar product formation the divergence of a vector is a scalar quantity. The divergence, also as productivity of an area G noted, indicates, that whether source or sink in this area. If the divergence of a vector field is equal to zero$(\nabla \vec{v} = \text{div}\ \vec{v} = 0)$, the area is neither source nor sink.

According to **GAUSS law** the entire source and sink activity of an area *G* can be computed by the volume integral of the divergence. At the same time it is known from the balance laws that the difference between the source and sink activities, i.e. the flow rates of the surface must:

$$\iiint_G \operatorname{div} \vec{v}\, dV = \oint_S \vec{v} \cdot \vec{n}\, dS \tag{2.25}$$

For the two-dimensional area similarly:

$$\iint_A \operatorname{div} \vec{v}\, dA = \oint_L \vec{v} \cdot \vec{n}\, dL \tag{2.26}$$

$\vec{n}$ is a normal (perpendicularly standing) unit vector to the surface or to the circumference. With **Gauss**' theorem volume integral can be converted into integral over the surface and area integral can be converted into integral over the bound. Also the divergence plays a fundamental role in the hydrogeology, since all processes must be balance in the mathematical description. In particular a large number of further derivatives is based on the following relation:

$$\operatorname{div}\ \vec{v} = \operatorname{div}\left(-k\ \operatorname{grad}\ h\right) = q \tag{2.27}$$

Example of divergence calculation:

We compute the divergence of the velocity vector $\overrightarrow{v}$ in the previous example:

$$\nabla \overrightarrow{v} = \frac{\partial}{\partial x}\left(3 - 2y\right) 10^{-3} + \frac{\partial}{\partial y}\left(-4 \cdot 10^{-3} x\right) = 0$$

This area is neither source nor sink.

In the **rotation formation** the NABLA-operator is linked by means of cross product with a vector:

$$\operatorname{rot}\ \vec{v} = \nabla\ \times\ \vec{v} \tag{2.28}$$

$$\operatorname{rot}\ \vec{v} = \begin{bmatrix} \overrightarrow{i} & \overrightarrow{j} & \overrightarrow{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ v_x & v_y & v_z \end{bmatrix} = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) \overrightarrow{i} - \left(\frac{\partial v_z}{\partial x} - \frac{\partial v_x}{\partial z}\right) \overrightarrow{j} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) \overrightarrow{k}$$

The result represents again a vector.
If $\operatorname{rot}\ \vec{v} = 0$, we speak of irrotational field. We can also deduce from it, that **rot grad φ = 0** is always applicable for irrotational potential field *j*.

Further arithmetic rules in connection with the vectorial differentiation yield as a result of application of other vector rules and the extended rules for the differentiation of products:

$$\nabla(\varphi_1 \cdot \varphi_2) = \varphi_1 \nabla\varphi_2 + \varphi_2 \nabla\varphi_1 = \varphi_1 \operatorname{grad}\varphi_2 + \varphi_2 \operatorname{grad}\varphi_1 \tag{2.29}$$

$$\nabla \cdot (\varphi(\vec{a})) = \varphi \nabla\vec{a} + \vec{a} \cdot \nabla\varphi = \varphi \operatorname{div}(\vec{a}) + \vec{a} \cdot \operatorname{grad}(\varphi) \tag{2.30}$$

$$\nabla \times (\varphi\vec{a}) = \varphi\nabla \times \vec{a} + \vec{a} \times \nabla\varphi = \varphi \operatorname{rot}(\vec{a}) + \vec{a} \times \operatorname{grad}(\varphi) \tag{2.31}$$

If we examine the source and sink activity of an aquifer, we can write the DARCY law as follows:

$$\begin{aligned} \operatorname{div}(\vec{v}) &= \operatorname{div}(-k \cdot \operatorname{grad} h) = q \\ \nabla(\vec{v}) &= \nabla(-k \cdot \nabla h) = q \\ \nabla(\vec{v}) &= \nabla(-k) \cdot \nabla h - k \cdot \nabla(\nabla h) = q \\ \nabla(\vec{v}) &= \nabla(-k) \cdot \nabla h - k \cdot \Delta h = q \\ \operatorname{div}(\vec{v}) &= \operatorname{grad}(-k) \cdot \operatorname{grad}(h) - k \cdot \operatorname{div}(\operatorname{grad}(h)) = q \end{aligned} \tag{2.32}$$

Only for the isotropic aquifer **(-k) = 0**, and homogeneous may be set **grad** equation:

$$\operatorname{div}(\vec{v}) = -k \operatorname{div}(\operatorname{grad} h) = q \tag{2.33}$$

Table 2.3: coordinate system

		Koordinatensystem	
Bezeichnung	Operation	kartesisch	zylindr
Nabla-Operator	∇	$\nabla = \frac{\partial}{\partial x}\vec{i} + \frac{\partial}{\partial y}\vec{j} + \frac{\partial}{\partial z}\vec{k}$	$\nabla = \frac{\partial}{\partial r}\vec{r} + \frac{1}{r}$
Gradient	$grad\, h = \nabla h$	$\nabla h = \frac{\partial h}{\partial x}\vec{i} + \frac{\partial h}{\partial y}\vec{j} + \frac{\partial h}{\partial z}\vec{k}$	$\nabla h = \frac{\partial h}{\partial r}\vec{r} + \frac{1}{r}$
Divergenz	$div\, \vec{v} = \nabla \cdot \vec{v}$	$\nabla \cdot v = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}$	$\nabla \cdot \vec{v} = \frac{1}{r}\frac{\partial (r\, v_r)}{\partial r}$
Laplace-Operator	$div\,(grad\, h) = \nabla^2 h = \Delta h$	$\Delta h = \frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} + \frac{\partial^2 h}{\partial z^2}$	$\Delta h = \frac{1}{r}\frac{\partial \left(r \frac{\partial h}{\partial r}\right)}{\partial r}$
Rotation	$rot\, \vec{v} = \nabla \times \vec{v}$	$\nabla\, x\, \vec{v} = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right)\vec{i} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right)\vec{j} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right)\vec{k}$	$\nabla\, x\, \vec{v} = \left(\frac{1}{r}\frac{\partial}{\partial}\right.$ $+ \left(\frac{\partial v_r}{\partial z}\right.$ $+ \left(\frac{1}{r}\frac{\partial (r\, v_\varphi)}{\partial r}\right.$
Volumen	dV	$dV = dx \cdot dy \cdot dz$	$dV = r \cdot dr$

2.3 Examples of Vector calculus

1. The filter velocity $\vec{v}$ consists of components $v_x = 3 \cdot 10^{-3} \cdot s^{-1}$, $v_z = -5 \cdot 10^{-4} \cdot s^{-1}$ and $v_y = 0$. Outline and compute the filter speed and indicate the modulus and the angle.
The modulus of a vector:

$$|\vec{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

With the given values we get the modulus:

$$|\vec{v}| = \sqrt{(3 \cdot 10^{-3})^2 + (-5 \cdot 10^{-4})^2} = 3,04 \cdot 10^{-3} \frac{m}{s}$$

The angle is calculated by the slope, which is equal to the tangent of the angle:

$$\begin{aligned} \alpha &= \arctan\left(\frac{v_z}{v_x}\right) = \arctan\left(\frac{-5 \cdot 10^{-4}}{3 \cdot 10^{-3}}\right) \\ &= \arctan\,(-0,166) = 350,52^o = 6,12\ rad \end{aligned}$$

Thus the vector of tasks in Cartesian and polar coordinates:

$$\vec{v} = 3 \cdot 10^{-3} \frac{m}{s} \cdot \vec{i} + 5 \cdot 10^{-4} \frac{m}{s} \cdot \vec{k} = 3,04 \cdot 10^{-3} \frac{m}{s} \cdot \vec{r} + 6,12 \frac{m}{s} \cdot \vec{\alpha}$$

2. A pollutant particle moves by convection ($\vec{v}_{konv}$) and by the hydrodynamic dispersion ($\vec{v}_{disp}$). Plot and compute the back way and the end point, if the particle is transported from the origin of the coordinates with the following part

$$\begin{aligned} \vec{v}_{konv} &= 1 \cdot 10^{-4} \frac{m}{s} \vec{i} + 10^{-3} \frac{m}{s} \vec{j} \\ \vec{v}_{disp} &= 3 \cdot 10^{-10} \frac{m}{s} \vec{r} + 0,785 \vec{\alpha} \end{aligned}$$

In the task two different coordinate representations are used. Since the natural processes are independent of the type of representation, the task can be solved with the use of the Cartesian coordinate representation or by means of polar coordinates. In both cases a conversion between the two systems is necessary.

For the existing two-dimensional case the following relations are available:

$$\vec{a} = a_x\vec{i} + a_y\vec{j}$$
$$\vec{a} = a_r\vec{r} + a_\alpha\vec{\alpha}$$
$$a_r = \sqrt{a_x^2 + a_y^2} = |\vec{a}|$$
$$a_\alpha = \arctan\frac{a_y}{a_x}$$
$$a_x = \cos(a_\alpha)\cdot a_r$$
$$a_y = \sin(a_\alpha)\cdot a_r \qquad bzw.$$
$$a_x = \tan(a_\alpha)\cdot a_y$$

It is to be noted that a_α is usually indicated in radian measure and the following relation applies

$$\frac{\widehat{\alpha}}{2\,\pi} = \frac{\alpha^\circ}{360^\circ}$$

With the given numerical values we find:

$$\vec{v} = \vec{v}_{konv} + \vec{v}_{disp}$$

According to the above definition:

$$v_{r\,konv} = \sqrt{v_x^2 + v_y^2} = \sqrt{(10^{-4})^2 + (10^{-3})^2} = 10^{-7}\frac{m}{s}$$
$$v_{\alpha\,konv} = \arctan\left(\frac{v_y}{v_x}\right) = 93,65^o$$

3. Design and compute the end point of a pollutant particle after one day, if it moves from the point $x = 0m$; $y = 0m$ by a convection due to a potential gradient of Δh = 1m between the points $x = 0m$; $y = 0m$ and $x = 30m$; $y = 40m$ with a *k*-value $k = 5\cdot 10^{-4} m\cdot s^{-1}$.
As basis of convection the filter speed is set $\vec{v}$. The field velocity must be used in accurate way,

$$\vec{v}_a = \frac{\vec{v}}{n'}, \qquad \text{mit:} \quad \vec{v}\,\text{Filtergeschwindigkeit}, \qquad n'\,\text{durchströmtePorosität}$$
$$\vec{v} = -k\ \text{grad}\ h \qquad \text{(DARCY-Gesetz)}$$
$$v_r = k\frac{dh}{dr} \qquad \Rightarrow\ v_r \approx k\frac{\Delta h}{\Delta r}$$
$$\Delta r = \sqrt{(x_1 - x_2)^2\,(y_1 - y_2)^2} = \sqrt{(30\,m^2) + (40\,m^2)} = 50\,m$$
$$v_r = 5\cdot 10^{-4}\frac{m}{s}\cdot\frac{1\,m}{50\,m} = 10^{-5}\,\frac{m}{s}$$

however not in this task. The mean transit velocity $\vec{v}_a$ is equated thereby the pore velocity.

With: $\vec{v}$ filter velocity n' seep through porosity
(Darcy law)

Distance: $s = v_r \cdot t = 10^{-5}\frac{m}{s} \cdot 86400\,s = 0,864\,m$

Position:
$$\begin{aligned} s &= \sqrt{x^2 + y^2} \\ x^2 &= s^2 - y^2 \end{aligned}$$

From the equation of straight line: $y = mx + n$ or the two points equation of straight line results:

$$\frac{y_1 - y_0}{x_1 - x_0} = \frac{y - y_0}{x - x_0}$$

With $x_0 = y_0 = 0$, $x_1 = 30m$ and $y_1 = 40m$:

$$\frac{y}{x} = \frac{40}{30} = \frac{4}{3} \qquad \text{bzw. } y = \frac{4}{3}x = m \cdot x$$

$$y^2 = m^2 \cdot x^2 = m^2 \cdot \left(s^2 - y^2\right)$$

$$y^2 = \frac{m^2 \cdot s^2}{(1 + m^2)}$$

$$y = \sqrt{\frac{m^2 \cdot s^2}{(1 + m^2)}} = \sqrt{\frac{1,333^2\ (0,864\,m)^2}{(1 + 1,7778)}} = \sqrt{0,4777\,m} = 0,689\,m$$

$$x = \frac{y}{m} = 0,517\,m$$

We can insert y immediately into the equation of the length s:

$$\begin{aligned} y &= \frac{4}{3}x \\ s &= \sqrt{x^2 + y^2} \\ s &= \sqrt{x^2 + \frac{4}{3}x^2} \end{aligned}$$

With $s = 0.864m$ we get the value:

$$x = \frac{s}{\sqrt{(1 + \frac{4}{3})}} = \frac{0,864\,m}{1,5275} = 0,518\,m$$

$$y = \frac{4}{3}\,x = 0,69\,m$$

2.4 Task of vector calculus

1. The vectors $\vec{a}, \vec{b}, \vec{c}$ are given in the coordinates:

$a_x = 5 \quad b_x = 3 \quad c_x = -6$

$a_y = 7 \quad b_y = -4 \quad c_y = -9$

$a_z = 8 \quad b_z = 6 \quad c_z = -5$

Determine the length of vector $\vec{d} = \vec{a} + \vec{b} + \vec{c}$.

2. Given vectors $\vec{a} = 2\vec{i} - 3\vec{j} + 5\vec{k}$ and $\vec{b} = 3\vec{i} - w\vec{j} + 2\vec{k}$. Compute *w* such that the two vectors stand perpendicularly to each other.

3. Calculate for $\varphi = xy + yz + zx$ and $\vec{A} = x^2y\vec{i} + y^2z\vec{j} + z^2x\vec{k}$:

a) $\vec{A} \cdot \nabla\varphi$

b) $\varphi \cdot \left(\nabla \vec{A}\right)$ und

c) $(\nabla\varphi) \times \vec{A}$

4. A particle moves along a space curve in the coordinates $\boldsymbol{x = t^3 + 2t,\ y = -3e^{-2t},\ z = 2 \sin 5t.}$ Compute the velocity and the acceleration of the particle at any time *t*. Indicate the distances for *t* = 0 and *t* = 1.

5. Design and compute the end point of a pollutant particle after one day, if it moves from the point *x* = 0*m*; *y* = 0*m* by a convection due to a potential gradient of Δh = 1m between the points *x* = 0*m*; *y* = 0*m* and *x* =30*m*; *y* = 40*m* with a *k*-value $k = 5 \cdot 10^{-4}\ \mathrm{m \cdot s^{-1}}$.

6. The scalar potential field is given in a filter *h* = *xy* + *yz* + *xz*.

a) Determine the filter velocity (vector and modulus).
b) The activity is source or sink in the filter?
c) Is this irrotational flow in the filter?
Given $k = 10^{-4}\ \mathrm{ms^{-1}}$ and grad (-*k*) = 0.

7. A pollutant plume spreads underground. The distribution of the pollutant varies in the range of values x::= 0 to 10 and y::= 0 to 10 with the following figure:

$C(x, y) = 50 - ((x - 5)^2 + (y - 5)^2)$

a.) Outline the equipotential lines for the concentration values in range of *C* (*x, y*) = 0*mg* to 50*mg* with an increment Δ*C* (*x, y*) = 10.
b.) Compute the gradient at the point *P* (3, 4) and determine the modulus and the direction angle.

8. A pollutant plume spreads underground. Die The distribution of the pollutant varies in the range of values x::= 0 to 10 and y::= 0 to 10 with the following figure:

$$C(x,y) = 125 - \left((2x-10)^2 + (y-5)^2\right)$$

a.) Outline the equipotential lines for the concentration values in range of $C\,(x,\,y) = 0mg$ to $125mg$ with an increment $\Delta C\,(x,\,y) = 25$.
b.) Compute the gradient at the point $P\,(5,\,10)$ and determine the modulus and the direction angle.

9. The groundwater level of an aquifer which one side is limited by a barrier and a well are to be described by the following geometrical figure:

$$z_R = \frac{1}{2}\frac{(y-10)^2}{x}$$

a.) Outline the hydro isohypses in range of $z_R = 1m$ to $z_R = 5m$ with an increment $\Delta z_R = 1m$ for coordinate $0 \leq x \leq 10$.
b.) Compute the filter velocity with $k = 0.001 ms^{-1}$ at the point $P\,(5,\,5)$; determine the modulus and the direction angle.
c.) Is this field source or sink?

Chapter 3

3 Interpolation method

Problem:

Some measured values (dependent variable) are dependent on independent variables in one -, two -, three or four dimensional space measurement, and generally they are represented by the three space coordinates (depending upon coordinate system e.g. x_n, y_n, z_n or r_n, α_n, z_n or r_n, α_n, θ_n (see chapter 2 vector analysis, page 41)) and the time t_n. We have a discontinuous value tables in this case. For one dimensional case e.g.:

independent value	dependent value
x_0	$y_0 = f(x_0)$
x_1	$y_1 = f(x_1)$
$\vdots$	$\vdots$
x_n	$y_n = f(x_n)$

The places x_0, x_1...., x_n are so called **supporting places**, and the y_0, y_1....,y_n are **basic values**.

If function values, whose arguments lie within the range (x_0, x_n), we name it **interpolation**. In contrast the searched function values for independent variable outside of the range (x_0, x_n) will be called **extrapolation**. A continuous substitute function $w = p(x)$ is found by the interpolation or extrapolation, which reflects the original function as exactly as possible $y_n = f(x_n)$ (see figure 3,1). It is always assumed that the substitute function only matches the original function on the supporting places. The accuracy of intervals, i.e. the agreement of the both functions, depends on the number and the distribution of the supporting places. According to the sample theorem the quantization error increases proportionally to the rise of the function.

Attention:

No interpolation algorithm can be used as replacement for an enlargement of the measured value density. By means of the interpolation algorithms one receives in each case **approximate values.**

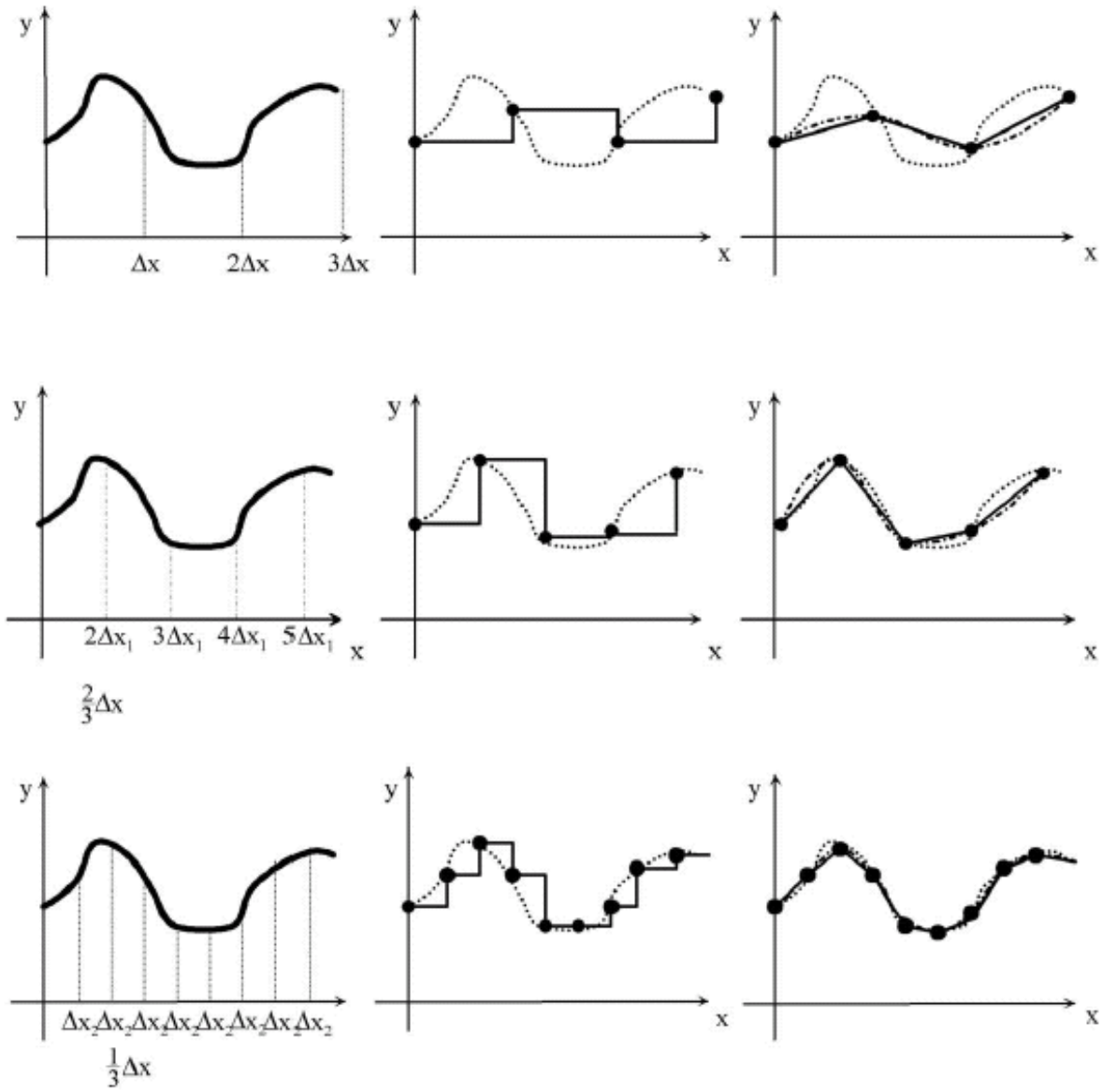


Figure 3.1: representation of the discontinuous measured data acquisition

Example for the application of interpolations:

The pollutant concentration *C(x)*, which runs out from a refuse dump, is measured at the points x_0, x_1, x_2 (see figure 3.2). The pollutant concentration at the point x_{Fl}, which flows into the river to cause danger, is to be estimated by interpolation. A conclusion is to be given whether this value exceeds the limiting value.

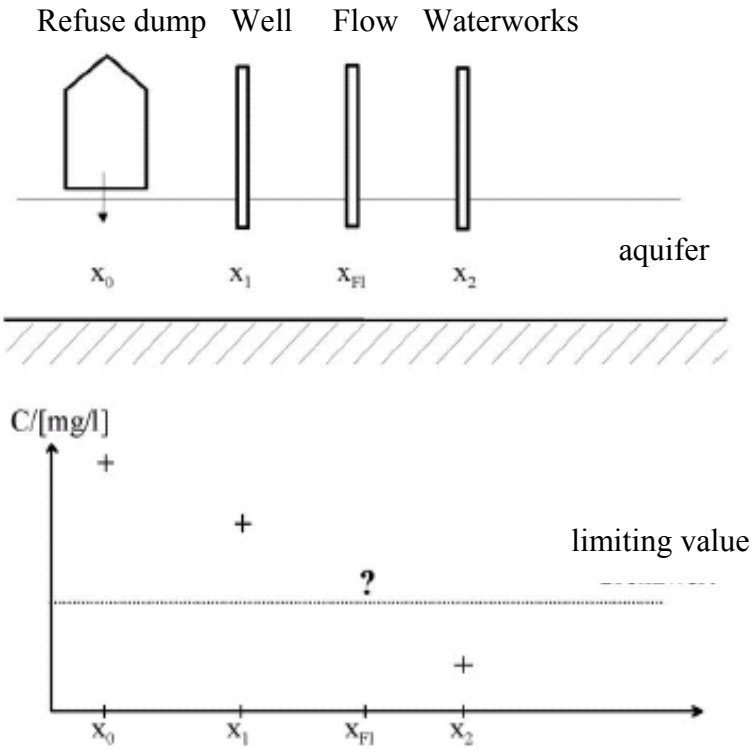


Figure 3.2: representation of an interpolation problem

x_0	$C_0 = f(x_0)$
x_1	$C_1 = f(x_1)$
x_{Fl}	?
x_2	$C_2 = f(x_2)$

For the solution of this problem an interpolation function $w = (p)$ is to seek for as "replacement" for the function $C_n = f(x_n)$. This function should fulfil the following condition:

$$w_i = p(x_i) = C_i \qquad (3.1)$$

i.e.

$$\begin{aligned} w_0 &= p(x_0) = C_0 \\ w_1 &= p(x_1) = C_1 \\ &\vdots \\ w_n &= p(x_n) = C_n \end{aligned} \qquad (3.2)$$

Then it is supposed that the intermediate values of the function $w = (p)$ are good approximation of the intermediate values of the function $C_n = f(x_n)$.

For the determination of the function $w = (p)$ different interpolation methods can be used. We differentiate thereby one- and multi-dimensional procedures. The multidimensional methods play an important role in connection with the geographical information systems (GIS) and are also often applied in connection with geostatistics.

In the following some methods will be introduced in connection with water economical questions.

- Polynomial interpolation

- Polynomial interpolation (spline)

- Kriging method

3.1 Polynomial interpolation

In this method $p(x)$ has the form of an algebraic polynomial of n order:

$$w = p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \tag{3.3}$$

This possesses such an advantage that intermediate values can be computed as easily as possible.

It is assumed a measured value table with n+1 pairs, and then maximally an n-th order polynomial can be exactly determined:

$$y := p_n(x) = \sum_{k=0}^{n} a_k \cdot x^k \tag{3.4}$$

with the character:

$$y(x_i) \approx p(x_i) = \sum_{k=0}^{n} a_k \cdot x_i^k = w_i \tag{3.5}$$

This polynomial is the interpolation polynomial to the given system of interpolation supporting places.

In the rules for low order polynomials ($n \leq 3$), the value pairs are sought at least piecewise to match:

$p(x) = a_0 + a_1 x$	linear interpolation
$p(x) = a_0 + a_1 x + a_2 x_2$	quadratic interpolation
$p(x) = a_0 + a_1 x + a_2 x_2 + a_3 x_3$	cubic interpolation

The application of polynomials with higher orders makes the arithmetic work more difficult and leads to very large fluctuations.

The different display formats for polynomials also yield the different interpolation procedures for the determination of the coefficients a_i of n-th polynomial. These different procedures all lead to the same polynomial. Thus interpolation formulas are differentiated according to:

- analytical power function
- LAGRANGE
- AIKEN
- NEWTON

3.1.1 Analytical power function

This method assumes that each supporting place of the polynomial $w = p(x)$ fulfils the condition $y(x_i) = p(x_i)$. In this case we get for the $n + 1$ supporting places $n + 1$ equations with the $n + 1$ unknown quantities a_0 to a_n.

$$\begin{aligned} a_0 + a_1 x_0 + a_2 x_0^2 + \ldots + a_n x_0^n &= y_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_n x_1^n &= y_1 \\ &\vdots \\ a_0 + a_1 x_n + a_2 x_n^2 + \ldots + a_n x_n^n &= y_n \end{aligned} \qquad (3.6)$$

This equation system can be written in accustomed way as matrix equation:

$$X \cdot A = Y$$

With

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \qquad A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad Y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

It is to be noted that the matrix **X** and **Y** on the right side represent well-known the coefficients, whereby the matrix **A** represents the searched solution vector. The LGS can be solved with all well-known methods (see section 1.3 solution of equation system, page 16).

The determinant of this linear equation system (LGS) is:

$$D = \begin{vmatrix} 1 & x_0 & \cdots & x_0^n \\ 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^n \end{vmatrix} = \begin{array}{r} (x_1 - x_0)\,(x_2 - x_0)\,(x_3 - x_0)\,\ldots\,(x_n - x_0) \\ (x_2 - x_1)\,(x_3 - x_1)\,\ldots\,(x_n - x_1) \\ \ldots\ldots \\ \ldots\ldots \\ (x_{n-1} - x_{n-2})\,(x_n - x_{n-2}) \\ (x_n - x_{n-1}) \end{array} \qquad (3.7)$$

and it is named as VANDERMOND determinant.

Since all supporting places are different with each other (must be), D ≠ 0 and the LGS is definitely solvable.

There is of an *n*-th order polynomial, which receives the values $y_i = f(x_i)$ and their coefficients: (cp. section 1.2.3 determinants, page 13 and the following):

$$a_0 = \frac{D_{a0}}{D}, \qquad a_1 = \frac{D_{a1}}{D}, \qquad \cdots \qquad a_n = \frac{D_{an}}{D} \tag{3.8}$$

From these coefficients we know the interpolation polynomial in demand:

$$y(x) \approx p(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_n \cdot x^n$$

Thereby the interpolation value in the place x_P results from:

$$y(x_P) \approx p(x_P) = a_0 + a_1 \cdot x_P + a_2 \cdot x_P^2 + \cdots + a_n \cdot x_P^n$$

Although the beginning of this procedure is very simple, the final determination of the interpolation polynomial requires a relative large computation, particularly if a great number of basic values are to be taken into account.

Example for the application of the polynomial interpolation:

Please find a quadratic polynomial by using the values of the following table and calculate the value *y = f (1/2)* at the place *x = 1/2*

x	0	1	2
y	0	1	0

Since only three supporting places are available, the polynomial can be only in a second order. A quadratic polynomial has the form:

$$p(x) = a_0 + a_1 x + a_2 x^2$$

It must be:

$$y_i = p(x_i)$$
$$y_i = a_0 + a_1 x_i + a_2 x_i^2$$

$$p(0) = 0 \implies a_0 + a_1 \cdot 0 + a_2 \cdot 0 = 0 \implies a_0 = 0$$

$$p(1) = 1 \implies a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 = 1 \implies a_1 + a_2 = 1$$

$$p(2) = 0 \implies a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 = 0 \implies 2a_1 + 4a_2 = 0$$

From the three equations:

$$a_0 = 0$$
$$a_1 = 2$$
$$a_2 = -1$$

Thus the interpolation polynomial:

$$p(x) = 2x - x^2$$

With this function the looked for function value in the place *x =1/2* can be computed:

$$f\left(\frac{1}{2}\right) \approx p\left(\frac{1}{2}\right)$$
$$p\left(\frac{1}{2}\right) = 2 \cdot \frac{1}{2} - \left(\frac{1}{2}\right)^2$$
$$f\left(\frac{1}{2}\right) \approx \frac{3}{4}$$

3.1.2 LAGRANGE interpolation formula

LAGRANGE wrote the interpolation function in the following form:

$$y(x_P) \approx p(x_P) = L_0(x_P) \cdot y_0 + L_1(x_P) \cdot y_1 + \ldots + L_n(x_P) \cdot y_n \qquad (3.9)$$

With the Lagrange interpolation no analytical functions are computed, but only individual values *p (xP)* for each interpolation place *(xP)*. *(Li (x) (i = 0, 1, ..., n))* are the coefficients of the basic values *yi* in the n-th order polynomial. These are computed from the supporting places *xi*. The Lagrange polynomial with n-th order has the following shape:

$$\begin{aligned}
L_0(x) &= \frac{(x_P - x_1)(x_P - x_2)\cdots(x_P - x_n)}{(x_0 - x_1)(x_0 - x_2)\cdots(x_0 - x_n)} \qquad (3.10)\\
L_1(x) &= \frac{(x_P - x_0)(x_P - x_2)\cdots(x_P - x_n)}{(x_1 - x_0)(x_1 - x_2)\cdots(x_1 - x_n)}\\
&\vdots\\
L_i(x) &= \frac{(x_P - x_0)(x_P - x_1)(x_P - x_2)\cdots(x_P - x_{i-1})(x_P - x_{i+1})\cdots(x_P - x_{n-1})}{(x_i - x_0)(x_i - x_1)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_{n-1})}\\
&\vdots\\
L_n(x) &= \frac{(x_P - x_0)(x_P - x_1)(x_P - x_2)\cdots(x_P - x_{n-1})}{(x_n - x_0)(x_n - x_1)\cdots(x_n - x_{n-1})}
\end{aligned}$$

Thus the LAGRANGE interpolation polynomial:

$$\begin{aligned}
y = f(x_P) \approx p(x_P) &= L_0(x_P)\, y_0 + L_1(x_P)\, y_1 + \ldots + L_n(x_P)\, y_n \qquad (3.11)\\
&= \frac{(x_P - x_1)(x_P - x_2)\cdots(x_P - x_n)}{(x_0 - x_1)(x_0 - x_2)\cdots(x_0 - x_n)} y_0\\
&+ \frac{(x_P - x_0)(x_P - x_2)\cdots(x_P - x_n)}{(x_1 - x_0)(x_1 - x_2)\cdots(x_1 - x_n)} y_1\\
&+ \ldots + \frac{(x_P - x_0)(x_P - x_1)(x_P - x_2)\cdots(x_P - x_{n-1})}{(x_n - x_0)(x_n - x_1)\cdots(x_n - x_{n-1})} y_n
\end{aligned}$$

If we insert the value of x_P choosing from $x_0, x_1 \ldots x_{n-1}, x_n$, there is always a factor which is equal to zero. Thus all Lagrange polynomials will become zero except the *i*-th item. The *i*-th item is one as the numerator is equal to the denominator. It proves:

$$y_i = f(x_i) \approx p(x_i) = 1 \cdot y_i$$

A disadvantage of the Lagrange method is that the computation of the Lagrange interpolation polynomials must be accomplished again when an increase of the supporting place number should be taken into account, which is identical with the increase of the order of the interpolation. This is to be clearly seen in the following example.

Attention:

· The weights (factors) *Li* (x_i) of LANGRANGE interpolation formula must be always again computed if the number of the supporting places changes itself.

· The sum of the weights always is equal to one (as a check of the results).

$$\sum L_i(x_i) = 1$$

Example for the application of LAGRANGE interpolation function:

For the function $y_n = f(x_n)$ the values in the equidistant places $x_n = x_0 + 2nh$, with n = -1, 0, 1, 2 (see table) are given:

n	-1	0	1	2
x_n	$x_0 - 2h$	x_0	$x_0 + 2h$	$x_0 + 4h$
$f(x_n)$	y_{-1}	y_0	y_1	y_2

Find an approximate value $w = f(x_0 + h)$ for $x = 1/2$

According to the rules of the polynomial interpolation maximally a polynomial with 3rd order can be developed in this case with four supporting places. It is also possible to accomplish a piecewise interpolation. This has the advantage that we can reduce computation work. The accuracy is however declined. In this case we try to find an optimum between the required accuracy and the cost of computation. The supporting places are used in the piecewise interpolation, which are next to the interpolation point.

1. Linear interpolation

The interpolation function in the place $x=1/2$ is written as follows with the help of the Lagrange interpolation formula (see equation 3.9):

$$w_{\frac{1}{2}} = L_0\left(x_{\frac{1}{2}}\right) y_0 + L_1\left(x_{\frac{1}{2}}\right) y_1$$

The supporting places values $x = 0$ and $x = 1$ are used, between which the value $x=1/2$ lies. The factors L_0 and L_1 are (see equation 3.10):

$$L_0\left(x_{\frac{1}{2}}\right) = \frac{x_{\frac{1}{2}} - x_1}{x_0 - x_1} = \frac{x_0 + h - x_0 - 2h}{x_0 - x_0 - 2h} = \frac{1}{2}$$
$$L_1\left(x_{\frac{1}{2}}\right) = \frac{x_{\frac{1}{2}} - x_0}{x_1 - x_0} = \frac{x_0 + h - x_0}{x_0 + 2h - x_0} = \frac{1}{2}$$

Then the searched value:

$$w_{\frac{1}{2}} = \frac{1}{2}(y_0 + y_1)$$

The result of the linear interpolation is thereby equal to the arithmetic means.

2. Quadratic interpolation
in this case (see equation 3.9):

$$w_{\frac{1}{2}} = L_0\left(x_{\frac{1}{2}}\right) y_0 + L_1\left(x_{\frac{1}{2}}\right) y_1 + L_2\left(x_{\frac{1}{2}}\right) y_2$$

The corresponding factors are (see equation 3.10):

$$\begin{aligned}
L_0\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_1\right)\left(x_{\frac{1}{2}} - x_2\right)}{(x_0 - x_1)(x_0 - x_2)} = \frac{3}{8} \\
L_1\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_2\right)}{(x_1 - x_0)(x_1 - x_2)} = \frac{3}{4} \\
L_2\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_1\right)}{(x_2 - x_0)(x_2 - x_1)} = -\frac{1}{8}
\end{aligned}$$

and the result is:

$$w_{\frac{1}{2}} = \frac{3}{8} y_0 + \frac{3}{4} y_1 - \frac{1}{8} y_2.$$

3. Cubic interpolation
In the same way (see equation 3.9):

$$w_{\frac{1}{2}} = L_{-1}\left(x_{\frac{1}{2}}\right) y_{-1} + L_0\left(x_{\frac{1}{2}}\right) y_0 + L_1\left(x_{\frac{1}{2}}\right) y_1 + L_2\left(x_{\frac{1}{2}}\right) y_2$$

We get the following LAGRANGE factors (see equation 3.10):

$$\begin{aligned}
L_{-1}\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_1\right)\left(x_{\frac{1}{2}} - x_2\right)}{(x_{-1} - x_0)(x_{-1} - x_1)(x_{-1} - x_2)} = -\frac{1}{16} \\
L_0\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_{-1}\right)\left(x_{\frac{1}{2}} - x_1\right)\left(x_{\frac{1}{2}} - x_2\right)}{(x_0 - x_{-1})(x_0 - x_1)(x_0 - x_2)} = \frac{9}{16} \\
L_1\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_{-1}\right)\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_2\right)}{(x_1 - x_{-1})(x_1 - x_0)(x_1 - x_2)} = \frac{9}{16} \\
L_2\left(x_{\frac{1}{2}}\right) &= \frac{\left(x_{\frac{1}{2}} - x_{-1}\right)\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_1\right)}{(x_2 - x_{-1})(x_2 - x_0)(x_2 - x_1)} = -\frac{1}{16}
\end{aligned}$$

Thus the result:

$$w_{\frac{1}{2}} = -\frac{1}{16} y_{-1} + \frac{9}{16} y_0 + \frac{9}{16} y_1 - \frac{1}{16} y_2$$

3.1.3 NEWTON interpolation formula

3.1.3.1 Arbitrary supporting places

The disadvantage of the Lagrange method is that the Lagrange polynomials must be computed again and again, which can be avoided in the Newton's method. With the Newton's method only one auxiliary item should be added when further supporting places are taken into account.

The method begins with following formula:

$$\begin{aligned} p(x) &= b_0 + b_1(x - x_0) \\ &\quad + b_2(x - x_0)(x - x_1) \\ &\quad + b_3(x - x_0)(x - x_1)(x - x_2) \\ &\quad \vdots \\ &\quad + b_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}) \end{aligned} \tag{3.12}$$

If we want to find a certain interpolation value $p\,(x_P)$, x will be replaced by x_P in the polynomial expression.

The coefficients are determined again in such a way that the polynomial accurately reflects the supporting places (x_n, y_n). If we respectively replace x_P with x_0, $x_1 \ldots x_n$ in the Newton's formula, gradually we get an equation system with n equations for n unknown quantities. Since in each case the corresponding factors ((x_P - x_i) = 0) are equal to zero, the polynomial items will be omitted. Then we know the basic value y_i from the polynomial value $p(x_i)$.

$$\begin{aligned} y_0 &= b_0 + b_1\underbrace{(x_0 - x_0)}_{=0} + \cdots \\ y_1 &= b_0 + b_1(x_1 - x_0) + b_2(x_1 - x_0)\underbrace{(x_1 - x_1)}_{=0} + \cdots \\ y_2 &= b_0 + b_1(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1) \\ &\vdots \\ y_n &= b_0 + b_1(x_n - x_0) + b_2(x_n - x_0)(x_n - x_1) + \cdots \\ &\quad + b_n(x_n - x_0)(x_n - x_1)\cdots(x_n - x_{n-1}) \end{aligned} \tag{3.13}$$

The equation system can be solved gradually with b_0, $b_1 \ldots b_n$. By inserting the first equation into second we get b_1. Once again inserting into the third equation it yields b_2. In $(n + 1)$-th equation the b_0, $b_1 \ldots b_{n-1}$ which are determined before are used to yield b_n.

$$
\begin{aligned}
b_0 &= y_0 \\
b_1 &= \frac{(y_1 - y_0)}{(x_1 - x_0)} = [x_1 x_0] \\
b_2 &= \frac{(y_2 - y_0) - \frac{(y_1 - y_0)}{(x_1 - x_0)}(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{(y_2 - y_1) + (y_1 - y_0) - [x_1 x_0]\,(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{(y_2 - y_1)}{(x_2 - x_0)(x_2 - x_1)} + \frac{(y_1 - y_0)}{(x_2 - x_0)(x_2 - x_1)} - \frac{[x_1 x_0]\,(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{[x_2 x_1]}{(x_2 - x_0)} + \frac{[x_1 x_0]\,(x_1 - x_0)}{(x_2 - x_0)(x_2 - x_1)} - \frac{[x_1 x_0]\,(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{[x_2 x_1]}{(x_2 - x_0)} + \frac{[x_1 x_0]\,(x_1 - x_0) - [x_1 x_0]\,(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{[x_2 x_1]}{(x_2 - x_0)} + \frac{-\,[x_1 x_0]\,(x_2 - x_1)}{(x_2 - x_0)(x_2 - x_1)} \\
b_2 &= \frac{[x_2 x_1] - [x_1 x_0]}{(x_2 - x_0)} = [x_2 x_1 x_0]
\end{aligned}
\tag{3.14}
$$

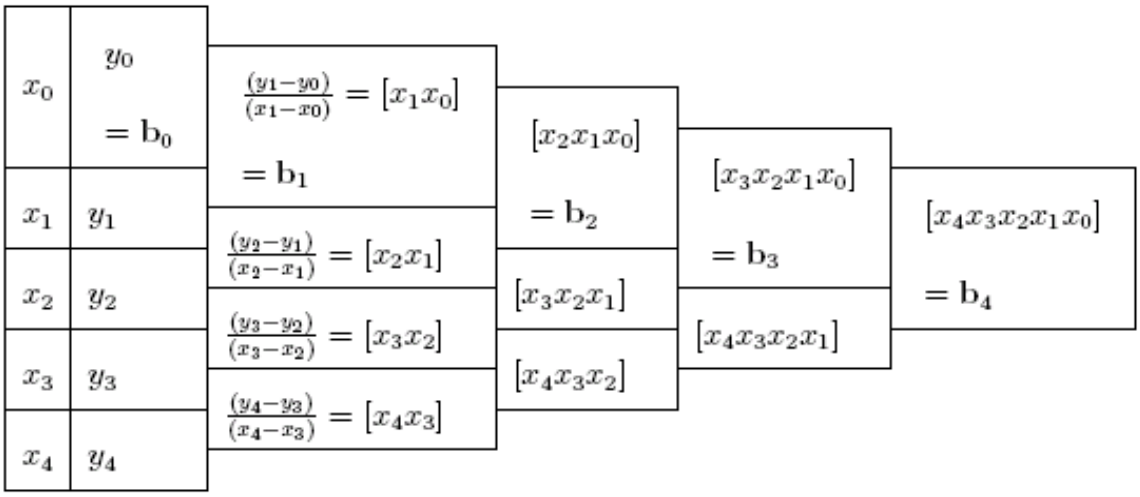
Generally the following notation is introduced for short, which are called **divided differences** of first and higher order:

$$
\begin{aligned}
[x_k x_i] &:= \frac{(y_k - y_i)}{(x_k - x_i)} \\
[x_l x_k x_i] &:= \frac{[x_l x_k] - [x_k x_i]}{(x_l - x_i)} \\
[x_m x_l x_k x_i] &:= \frac{[x_m x_l x_k] - [x_l x_k x_i]}{(x_m - x_i)} \\
&\vdots \\
[x_n x_{n-1} \cdots x_1 x_0] &:= \frac{[x_n x_{n-1} \cdots x_1] - [x_{n-1} \cdots x_1 x_0]}{(x_n - x_0)}
\end{aligned}
\tag{3.15}
$$

Thus the results of the coefficients:

$$
\begin{aligned}
b_0 &= y_0 \\
b_1 &= \frac{(y_1 - y_0)}{(x_1 - x_0)} = [x_1 x_0] \\
b_2 &= \frac{[x_2 x_1] - [x_1 x_0]}{(x_2 - x_0)} = [x_2 x_1 x_0] \\
b_3 &= \frac{[x_3 x_2 x_1] - [x_2 x_1 x_0]}{(x_3 - x_0)} = [x_3 x_2 x_1 x_0] \\
&\vdots \\
b_n &= \frac{[x_n x_{n-1} \cdots x_1] - [x_{n-1} \cdots x_1 x_0]}{(x_n - x_0)} = [x_n x_{n-1} \cdots x_1 x_0]
\end{aligned}
\tag{3.16}
$$

Particularly the coefficients can be determined conveniently according to the following **computation scheme** (example for 5 supporting places):



(3.17)

According to equation 3.12 the value *y* at the place *x* can be interpolated:

$$\begin{aligned} y(x) \approx p(x) &= b_0 + b_1(x - x_0) \\ &+ b_2(x - x_0)(x - x_1) \\ &+ b_3(x - x_0)(x - x_1)(x - x_2) \\ &+ b_4(x - x_0)(x - x_1)(x - x_2)(x - x_3) \end{aligned} \tag{3.18}$$

This equation also can be used, in order to compute the interpolation function $w = p(x)$ distribution and possibly to plot the function.

3.1.3.2 Equidistant supporting place distribution

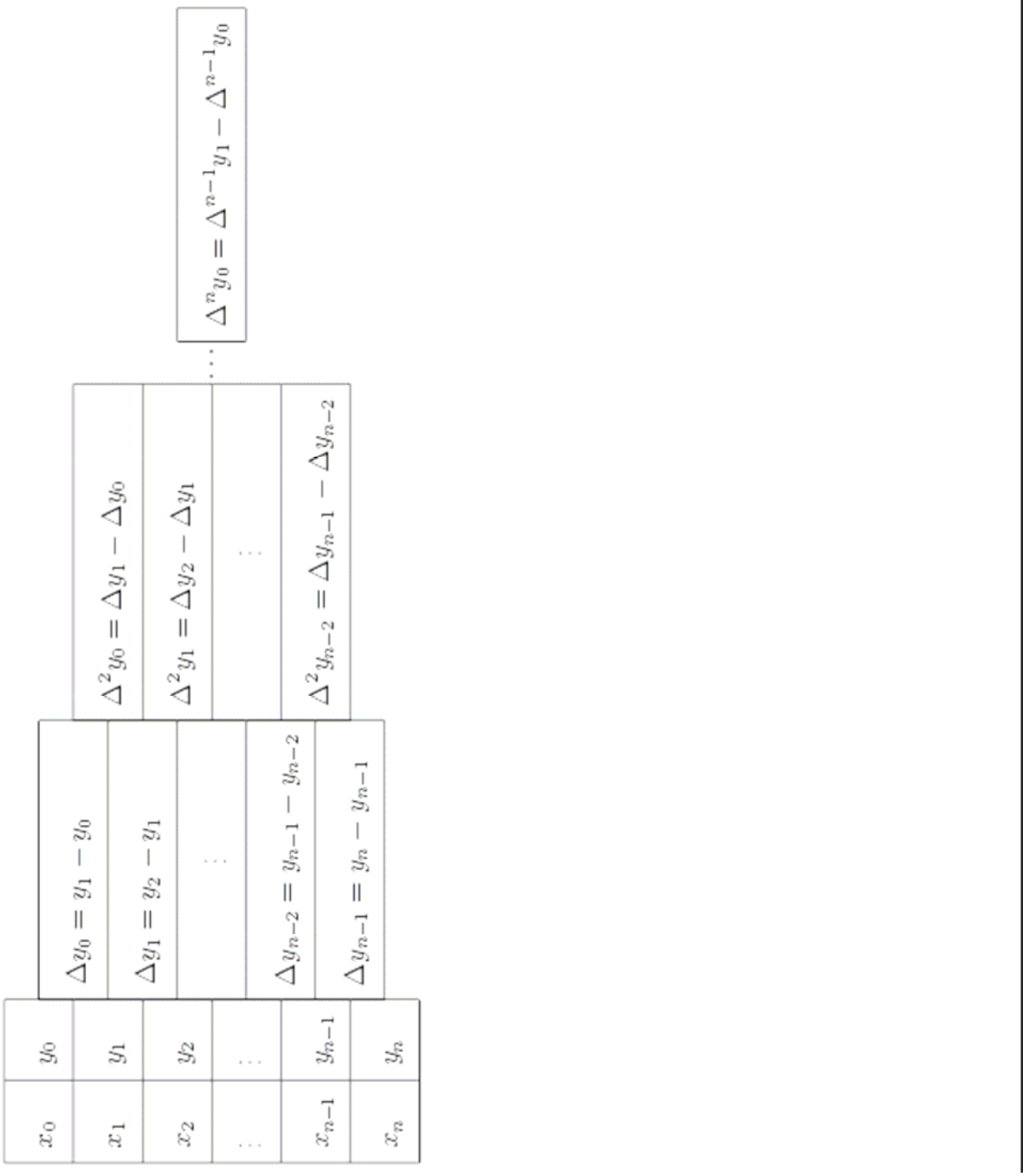
The **equidistant supporting place distribution** x_0, $x_1 = x_0 + h$, …, $x_n = x_0 + nh$ (h is the step length) are given, then the interpolation function by NEWTON:

$$p(x) = y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{2! \cdot h^2}(x - x_0)(x - x_1) + \ldots + \frac{\Delta^n y_0}{n! \cdot h^n}(x - x_0)\ldots(x - x_{n-1}) \tag{3.19}$$

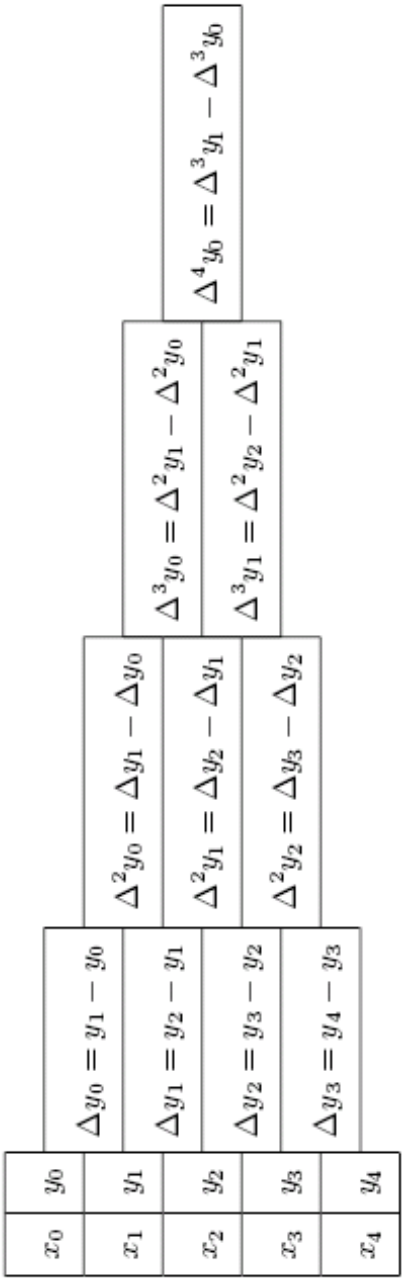
The elements Δy_0, $\Delta^2 y_0$, ..., $\Delta^n y_0$, are called **finite differences**. The exponent does not represent exponentiation, but gradual differences formation. We compare equation 3.19 with the equation 3.12 on page 70:

$$\begin{aligned} b_0 &\hat{=} y_0 \\ b_1 = \frac{y_1 - y_0}{x_1 - x_0} &\hat{=} \frac{\Delta y_0}{h} \\ b_2 = \frac{[x_2x_1] - [x_1x_0]}{(x_2 - x_0)} &\hat{=} \frac{\Delta^2 y_0}{2! \cdot h^2} \end{aligned} \tag{3.20}$$

These differences are computed according to the following scheme:



For example the scheme for $n = 4$:



By rear substitution we know that each finite difference is a combination of the y-values of the first column. e.g.:

$$\Delta^3 y_0 = y_3 - 3y_2 + 3y_1 - y_0, \tag{3.21}$$

3.1.3.3 Example for the application of Newton's method:

1. For the function $y_n = f(x_n)$ the values in the equidistant places are given $x_n = x_0 + 2nh$, $n = -1, 0, 1, 2$ (see table):

n	-1	0	1	2
x_n	$x_0 - 2h$	x_0	$x_0 + 2h$	$x_0 + 4h$
$f(x_n)$	y_{-1}	y_0	y_1	y_2

please find an approximate value $x = ½$ for $y_{1/2} = f(x_0 + h)$.
Solve this example with Newton's method and compare the results with those from LANGRANGE interpolation formula.

a) Linear interpolation:

$$\begin{aligned} p\left(x_{\frac{1}{2}}\right) &= y_0 + \frac{\Delta y_0}{h}\left(x_{\frac{1}{2}} - x_0\right) \\ &= y_0 + \frac{y_1 - y_0}{2h}\left(x_0 + h - x_0\right) \\ &= \frac{1}{2}\left(y_0 + y_1\right) \end{aligned}$$

b) Quadratic interpolation:

$$\begin{aligned} p\left(x_{\frac{1}{2}}\right) &= y_0 + \frac{\Delta y_0}{h}\left(x_{\frac{1}{2}} - x_0\right) + \frac{\Delta^2 y_0}{2!h^2}\left(x_{\frac{1}{2}} - x_0\right)\left(x_{\frac{1}{2}} - x_1\right) \\ &= \frac{1}{2}\left(y_0 + y_1\right) + \frac{y_2 - 2y_1 + y_0}{(2h)^2\, 2}\left(x_0 + h - x_0\right)\left(x_0 + h - x_0 - 2h\right) \\ &= \frac{3}{8}y_0 + \frac{3}{4}y_1 - \frac{1}{8}y_2 \end{aligned}$$

It applies:

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = y_2 - y_1 - (y_1 - y_0) = y_2 - 2y_1 + y_0$$

Remarks:

The advantage of the Newton's method is that the polynomial $L_i(x)$ does not change itself if the number of supporting places is changed, i.e. each time we only need calculate the additional part of the interpolation function.

2. The following measured values are given:

x	0	1	2	3	4	5
y	1	2	4	8	15	26

Determine the value y = f (2.5). Please select a polynomial with suitable order. How large is the deviation if the order of the polynomial is changed? Since the given supporting places are equidistant (*h = 1*), Newton's method is applicable to calculate the polynomials with different orders.

First the finite differences are computed:

$x_0 = 0$	$y_0 = 1$					
		$\Delta y_0 = 1$				
$x_1 = 1$	$y_1 = 2$		$\Delta^2 y_0 = 1$			
		$\Delta y_1 = 2$		$\Delta^3 y_0 = 1$		
$x_2 = 2$	$y_2 = 4$		$\Delta^2 y_1 = 2$		$\Delta^4 y_0 = 0$	
		$\Delta y_2 = 4$		$\Delta^3 y_1 = 1$		$\Delta^5 y_0 = 0$
$x_3 = 3$	$y_3 = 8$		$\Delta^2 y_2 = 3$		$\Delta^4 y_1 = 0$	
		$\Delta y_3 = 7$		$\Delta^3 y_2 = 1$		
$x_4 = 4$	$y_4 = 15$		$\Delta^2 y_3 = 4$			
		$\Delta y_4 = 11$				
$x_5 = 5$	$y_5 = 26$					

It is evident that the maximal interpolation polynomial order is third.

a) Linear interpolation

The searched value x = 2.5 lies between $x_2 = 2$ and $x_3 = 3$. Therefore the linear interpolation is accomplished only between this tow values

$$p(x) = y_2 + \frac{\Delta y_2}{h}(x - x_2)$$
$$= 4 + \frac{4}{1}(2,5 - 2)$$
$$p(2,5) = 6$$

b) Quadratic interpolation

Since the searched value is x = 2.5 the quadratic parabola interpolation can be stretched among x_1, x_2, und x_3.

$$\begin{aligned} p(x) &= y_2 + \frac{\Delta y_2}{h}(x - x_2) + \frac{\Delta^2 y_2}{2!h^2}(x - x_2)(x - x_3) \\ &= 4 + \frac{4}{1}(2,5 - 2) + \frac{3}{2 \cdot 1}(2,5 - 2)(2,5 - 3) \\ &= 4 + 2 - \frac{0,75}{2} \\ p(2,5) &= 5,625 \end{aligned}$$

c) Cubic interpolation

The cubic interpolation formula requires three supporting places. In this case both of the triple x1, x2, and x3 or the triple x2, x3, and x4 can be used. For the first case:

$$\begin{aligned} p(x) &= y_1 + \frac{\Delta y_1}{h}(x - x_1) + \frac{\Delta^2 y_1}{2!h^2}(x - x_1)(x - x_2) + \frac{\Delta^3 y_1}{3!h^3}(x - x_1)(x - x_2)(x - x_3) \\ &= 2 + \frac{2}{1}(2,5 - 1) + \frac{2}{2 \cdot 1}(2,5 - 1)(2,5 - 2) + \frac{1}{6 \cdot 1}(2,5 - 1)(2,5 - 2)(2,5 - 3) \\ &= 2 + 3 + 0,75 - 0,0625 \\ p(2,5) &= 5,6875 \end{aligned}$$

For the second triple:

$$\begin{aligned} p(x) &= y_2 + \frac{\Delta y_2}{h}(x - x_2) + \frac{\Delta^2 y_2}{2!h^2}(x - x_2)(x - x_3) + \frac{\Delta^3 y_2}{3!h^3}(x - x_2)(x - x_3)(x - x_4) \\ &= 4 + \frac{4}{1}(2,5 - 2) + \frac{3}{2 \cdot 1}(2,5 - 2)(2,5 - 3) + \frac{1}{6 \cdot 1}(2,5 - 2)(2,5 - 3)(2,5 - 4) \\ &= 4 + 2 - \frac{0,75}{2} + 0,0625 \\ p(2,5) &= 5,6875 \end{aligned}$$

The deviation between the linear and the quadratic result is:

$$\left| \frac{5,625 - 6}{5,625} \right| = 6,7\%,$$

While the deviation between the square and the cubic result is only:

$$\left| \frac{5,6875 - 5,625}{5,6875} \right| = 1,1\%$$

In order to estimate the results, the given points can be plotted (see figure 3.3). The diagram shows that:

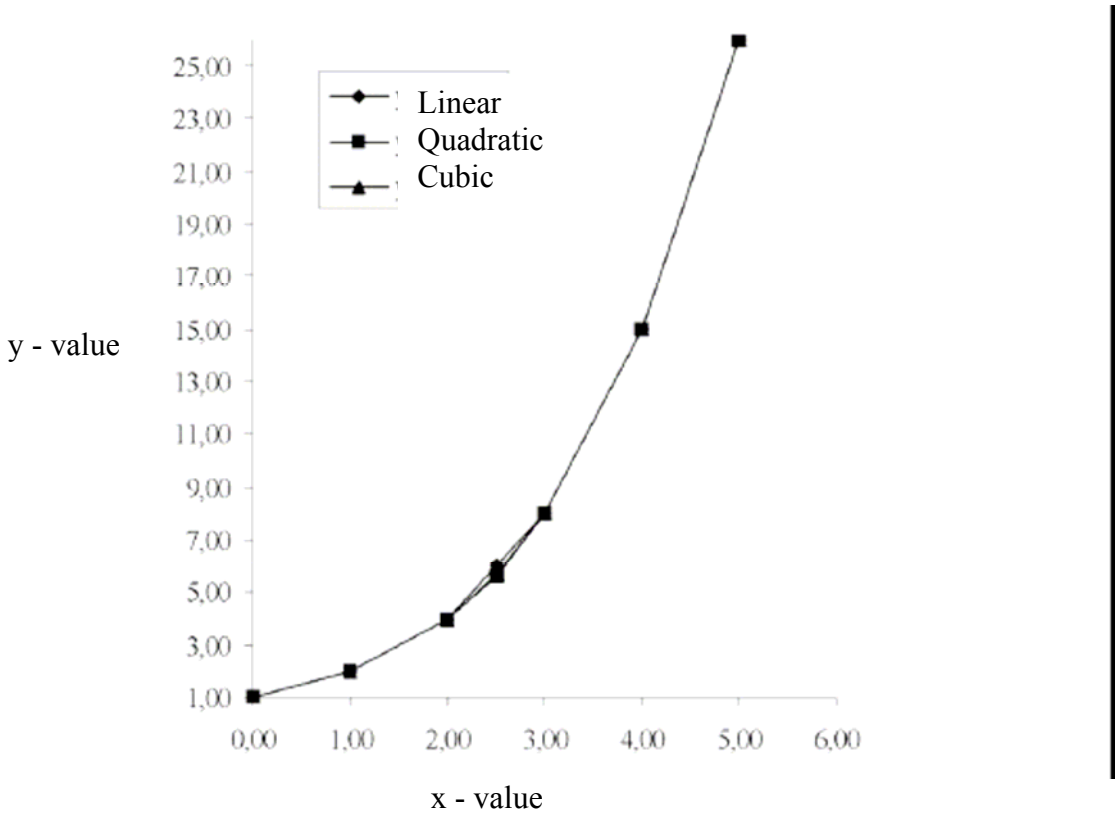


Figure 3.3: Representation of the measured interpolated values

Actually the value should lie between 5 and 6. Obviously the linear interpolation can not yield good results in this case. For this reason it is meaningful to plot given points and estimate the searched value. In a practical work it is important to have enough points in order to get a good approximation of the function. This can be ascertained that, the form of the function substantially does not change when additional points are taken into account.

3.2 Polynomial Interpolation (Spline)

To describe a given function in a certain interval we can link sections that consist of several lower degree polynomials together with only one polynomial with high degree. The classical examples are line segments in subintervals (seeing figure 3.4). It is assumed that the function between two supporting places is nearly linear. This can be applied, if the supporting places are narrow enough with each other.

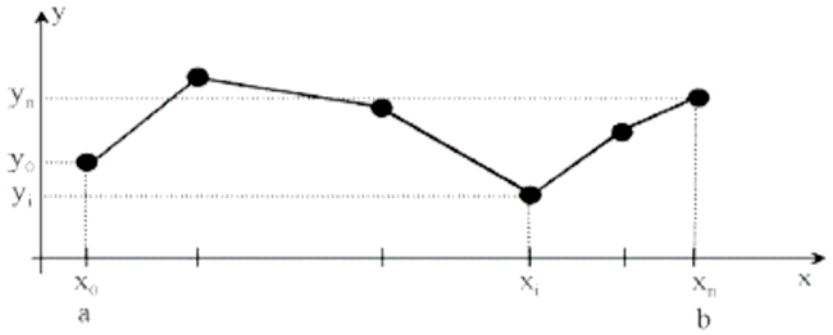


Figure 3.4: Representation of linear spline curves

Such approximations are continuous, however the first derivative is discontinuous, and i.e. vertex appears at the transition part from one interval to another. In the following spline interpolation method will be described, in which cubic parabola arches are built up such that the vertexes are rounded, then first and second derivatives of the approximation are constant. Polynomials with higher degree are in principle not used since they oscillate strongly.

A given interval of I = (a, b) is divided into n subintervals according to x-value $x_0 = a$, x_1, x_2 … $x_n = b$. The cubic parabola arches will adapt in each subinterval such that the given y-values y_i are fit at place x_i. The first and second derivatives must be agreement between left- and right side at the transition part of subintervals (see figure 3.5). The supporting places (x_i, y_i) are called the knots of the spline (the word "spline" originally designated a flexible curve template).

A cubic polynomial with third degree has four coefficients. Generally it can be written:

$$p_i(x) = c_{0i} + c_{1i}x + c_{2i}x^2 + c_{3i}x^3 \tag{3.22}$$

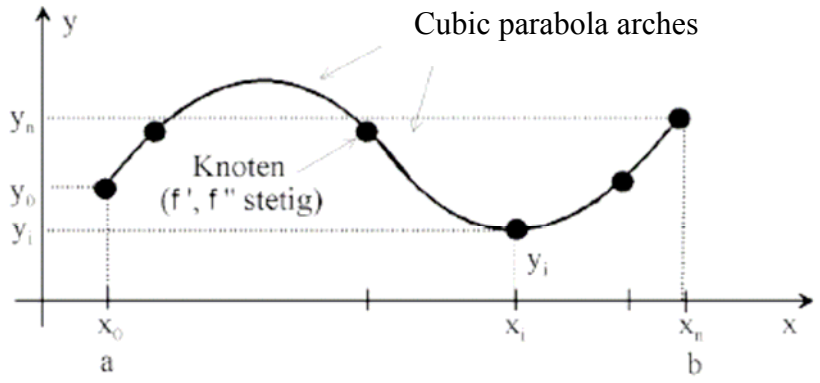


Figure 3.5 Representation of spline-curve for a cubic system

The spline function is defined as follows:

1. S(x) is twice continuously differentiable in the range [a, b].

2. In each interval $[x_i \ldots x_{i+1}]$ S(x) is given by a cubic polynomial p_i (x):

$$S(x) \equiv \sum p_i(x)$$
$$p_i(x) := a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \tag{3.23}$$

3. $S(x)$ fulfils the Interpolation constraints $S(x_i) = y_i$ for all i from $[1 \ldots n]$ in range [a, b].

4. Depending upon the form of connecting constraints we get different kinds from spline functions. The following is special **cubic spline functions**

.

Connecting conditions	Description	Comments
$S(x_0) = S''(x_0) = 0$ $S(x_n) = S''(x_n) = 0$	**natürlich**	$S(x_0)$ und $S(x_n)$ ist die Tangente an den Graphen von $S(x)$
$S''(x_0) = \alpha S''(x_n) = \beta$	**verallgemeinert**	
$S'(x_0) = \alpha S'(x_n) = \beta$	**vorgegeben**	erste Ableitung am Rand
$S'''(x_0) = \alpha S'''(x_n) = \beta$	**vorgegeben**	dritte Ableitung am Rand
$S(x_0) = S(x_n)$ $S'(x_0) = S'(x_n)$ $S''(x_0) = S''(x_n)$	**periodisch**	
$p_0(x) = p_l(x)$ $p_{n-2}(x) = p_{n-1}(x)$	**not-a-knot**	$S'''(x_l)$ und $S'''(x_n)$ sind stetig

In the case of n segments it yields 4n coefficients and 4n constraints for the 4n coefficient are expected. There are 4 constraints at each Knot (x_i, y_i) for $i = 1, 2 \ldots n - 1$ (y-value and agreement of the derivatives). This yields 4n-4 constraints. At the terminator points the y-value must be accepted, and thus sind4n ¡2 conditions found, i.e. the spline-curve is defined not completely; two degrees of freedom remain.

$$
\begin{aligned}
p_i(x_i) &= y_i && i = 0;1;\cdots n && \text{Interpolation constraints} \\
p_i(x_i) &= p_{i-1}(x_i) && && \\
p_i'(x_i) &= p_{i-1}'(x_i) && i = 0;1;\cdots n-1 && \text{Connecting constraints of polynomial } P_i \text{ at } P_{i-1} \\
p_i''(x_i) &= p_{i-1}''(x_i) && &&
\end{aligned}
\tag{3.24}
$$

We can set the second derivative at the terminator points zero and get a **natural spline curve**.

$$\begin{aligned} p_n(x_n) &= a_n && S(x_0) \text{ und } S(x_n) \text{ ist die Tangente} \\ p_n''(x_n) &= 2c_n && \text{an den Graphen von } S(x) \end{aligned} \qquad (3.25)$$

Alternatively the first derivative at the terminator points can be given, in order to approximate a function.
Thus it yields an equation system with 4n equations for 4n+2 unknown quantities. The two missing equations are covered by default of the boundary conditions.

$$\begin{aligned} p_0''(x_0) &= 0 \\ p_n''(x_n) &= 0 \end{aligned} \qquad \text{Boundary conditions} \qquad (3.26)$$

This equation system can be solved according to familiar methods. Usually the solution of this equation system is complex, so not only combination steps- but also iterative procedures (see section 1.3 solution methods of equation system, page 16) must be used. As is shown below, however a tridiagonal equation system can be generated by a certain scheme, then it can be solve with little operating expense.

Calculation scheme

N supporting places x_i with i = 0, 1 … n with the step length hi = x_{i+1}-x_i and the n basic values y_i with i = 0, 1 … n are given (e.g. as list of measurement readings), so the following calculation scheme (see equations 3.24 to 3.26) for interpolation by means of cubic spline functions can be applied with n-1 subfunctions for range $x_i \leq x \leq x_{i+1}$

$$S(x) \equiv \sum p_i(x)$$
$$p_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \qquad (3.27)$$

Schritt	Berechnung	Gültigkeitsbereich
1	$a_i = y_i$	$i = 0; 1; \cdots n$
2	$c_0 = c_n = 0$	
3	$h_{i-1}c_{i-1} + 2c_i(h_{i-1} + h_i) + h_i c_{i+1} = \frac{3}{h_i}(a_{i+1} - a_i) - \frac{3}{h_{i-1}}(a_i - a_{i-1})$	$i = 0; 1; \cdots n-1$
4	$b_i = \frac{1}{h_i}(a_{i+1} - a_i) - \frac{h_i}{3}(c_{i+1} - 2c_i)$	$i = 0; 1; \cdots n-1$
5	$d_i = \frac{1}{3h_i}(c_{i+1} - c_i)$	$i = 0; 1; \cdots n-1$

The equation in the third step of the table represents a linear equation system of n-1 equations for the unknown quantities c_1, c_2 …c_{n-1}. It can be written in the form of matrix:

$$\mathrm{A} \cdot \mathrm{C} = \mathrm{R} \qquad (3.28)$$

$$\mathrm{A} = \begin{bmatrix} 2(h_0 + h_1) & h_1 & & & & & \\ h_1 & 2(h_1 + h_2) & h_2 & & & & \\ & h_2 & 2(h_2 + h_3) & h_3 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & h_{n-3} & 2(h_{n-3} + h_{n-2}) & h_{n-2} & \\ & & & & h_{n-2} & 2(h_{n-2} + h_{n-1}) \end{bmatrix} \qquad (3.29)$$

$$\mathrm{R} = \begin{bmatrix} \frac{3}{h_1}(a_2 - a_1) - \frac{3}{h_0}(a_1 - a_0) \\ \frac{3}{h_2}(a_3 - a_2) - \frac{3}{h_1}(a_2 - a_1) \\ \frac{3}{h_3}(a_4 - a_3) - \frac{3}{h_2}(a_3 - a_2) \\ \vdots \\ \frac{3}{h_{n-2}}(a_{n-1} - a_{n-2}) - \frac{3}{h_{n-3}}(a_{n-2} - a_{n-3}) \\ \frac{3}{h_{n-1}}(a_n - a_{n-1}) - \frac{3}{h_{n-2}}(a_{n-1} - a_{n-2}) \end{bmatrix} \tag{3.30}$$

$$\mathrm{C} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{n-2} \\ c_{n-1} \end{bmatrix} \tag{3.31}$$

A is tridiagonally, symmetrically, diagonally dominant, positively definite matrix and possesses only positive elements. Thus this matrix is always invertable and definitely solvable. As solution method Gauss algorithm can be used for tridiagonal matrices (see section 1.3.1 solutions of equation system, Gauss algorithm, page 17).

Example for application of spline function:

The following measured values are given as supporting places and values.

i	0	1	2	3	4
x_i	-1	$-0,5$	0	$0,5$	1
y_i	$0,5$	$0,8$	1	$0,8$	$0,5$

For these 5 pairs a natural cubic spline function will be found. According to the definition of the spline functions, 4 subfunctions i = 1, 2, 3, 4 with respective ranges xi ≤ x ≤ xi+1 will be searched.

$$p_i(x) = a_i + b_i\,(x - x_i) + c_i\,(x - x_i)^2 + d_i\,(x - x_i)^3$$

Correspondently the computation schemes are implemented in five steps.

Schritt	Berechnung	Ergebnis
1	$a_i = y_i$	$a_0 = 0,5$ $a_1 = 0,8$ $a_2 = 1,0$ $a_3 = 0,8$ $a_4 = 0,5$
2	$c_0 = c_n = 0$	$c_0 = 0$ $c_4 = 0$

Schritt	Berechnung	Ergebnis
3	$\begin{bmatrix} 2(h_0+h_1) & h_1 & & \\ h_1 & 2(h_1+h_2) & h_2 & \\ & h_2 & 2(h_2+h_3) & h_3 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{h_1}(a_2-a_1) - \frac{3}{h_0}(a_1-a_0) \\ \frac{3}{h_2}(a_3-a_2) - \frac{3}{h_1}(a_2-a_1) \\ \frac{3}{h_3}(a_4-a_3) - \frac{3}{h_2}(a_3-a_2) \end{bmatrix}$ $\begin{bmatrix} 2(0,5+0,5) & 0,5 & & \\ 0,5 & 2(0,5+0,5) & 0,5 & \\ & 0,5 & 2(0,5+0,5) & 0,5 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{0,5}(1,0-0,8) - \frac{3}{0,5}(0,8-0,5) \\ \frac{3}{0,5}(0,8-1,0) - \frac{3}{0,5}(1,0-0,8) \\ \frac{3}{0,5}(0,5-0,8) - \frac{3}{0,5}(0,8-1,0) \end{bmatrix}$	$c_1 = 0$ $c_2 = 1,2$ $c_3 = 0$
4	$b_i = \frac{1}{h_i}(a_{i+1}-a_i) - \frac{h_i}{3}(c_{i+1}-2c_i)$	$b_0 = 0,6$ $b_1 = 0,6$ $b_2 = 0$ $b_3 = -0,6$
5	$d_i = \frac{1}{3h_i}(c_{i+1}-c_i)$	$d_0 = 0$ $d_1 = -0,8$ $d_2 = 0,8$ $d_3 = 0$

Thus it yields the following sub-spline-functions according to equation 3.23:

$$p_i(x) := a_i + b_i\,(x - x_i) + c_i\,(x - x_i)^2 + d_i\,(x - x_i)^3$$

Teil-Spline-Funktion	Geltungsbereich
$p_0(x) = 0,5 + 0,6(x+1)$	$-1 \leq x \leq -0,5$
$p_1(x) = 0,8 + 0,6(x+0,5) - 0,8(x+0,5)^3$	$-0,5 \leq x \leq 0$
$p_2(x) = 1,0 - 1,2x^2 + 0,8x^3$	$0 \leq x \leq 0,5$
$p_3(x) = 0,8 - 0,6(x-0,5)$	$0,5 \leq x \leq 1$

The diagram of the splines is shown in figure 3.6:

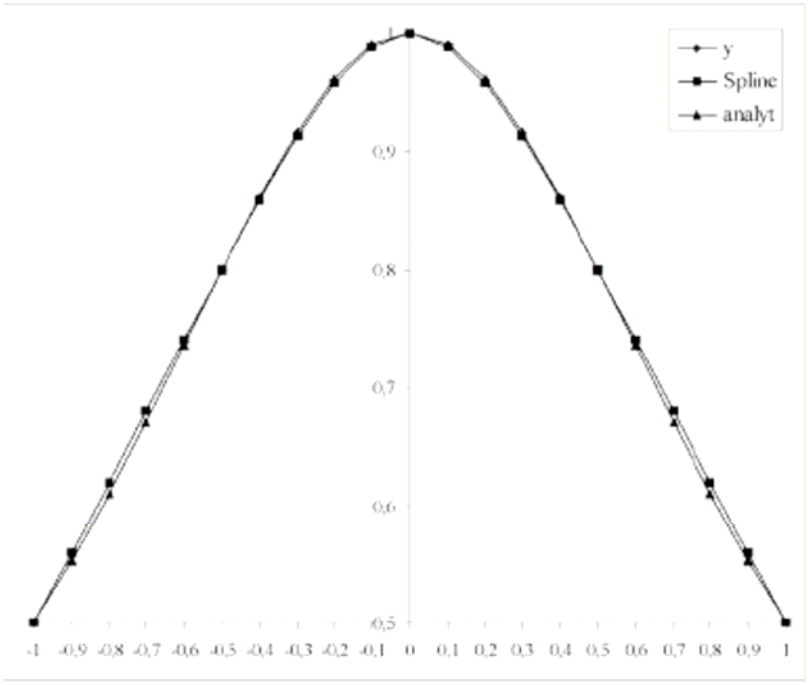


Figure 3.6: Spline interpolation function

We recognize that the spline simulates the original analytic function very well.

$$y = \frac{1}{x^2 + 1}$$

The maximum deviation of analytic solution amounts to 0.010244, which corresponds to 1.68%.

3.3 Kriging method

A family of special interpolation methods is marked with Kriging, which aims at the following problem:

The sampling at a place supplies information for certain space oriented points. However it is unknown which values are available for the measuring variables among these points. Kriging is a method, which makes possible, to compute the value of intermediate point or the average over an entire block. Different special methods are based on the creation of weighted average values of the space oriented variables. Block estimations are predominantly necessary in the mining industry, while estimated points are inserted for map, which is described in the following one.

The individual Kriging methods differ either in the kind of the goal sizes which can be estimated or in their methodical extension for the inclusion of additional information.

Additional information about the spatial behaviour of a location dependent variable exists in the cognition of other measurements, which relates to the observed variables. In hydrogeological practice for instance correlated dissolved matter or temporal repetition measurements of groundwater pressure head are common.

In a word Kriging methods are of following **advantages** compared to other interpolation procedures:

- Kriging yields the "best" estimated value

- Kriging involves the information of the spatial structure of the variable and the variogram into the estimation.

- The individual spatial arrangement of the measuring point net is considered with reference to the interpolation grid.

- The reliability of the results is indicated in form of Kriging error for each estimated point.

Attention:

In the Kriging method it must be also paid attention that no information gain can be achieved by the mathematical procedures. Only the information content of the measured values (basic values) is processed. Interpolation results might contradict physical laws (e.g. ground water contour line in receiving streams).

If we want to get physically correct interpolations, a fine quantized simulation by means of physical models (e.g. ground-water flow models) is necessary and meaningful. Therefore such simulation programs offer internal diagram routines for creation of isoline.

In order to understand the Kriging procedures, the following terms from the geostatistics must be known:

Mean value	$m =$	$\frac{1}{n}\sum_{a=1}^{n} Z_a$
Expected value	$E[Z] =$	$\int z \cdot p(z)\, dz = m,$ wobei $p(z)$ die Dichtefunktion ist
Variance	$\mathrm{var}(Z) =$	$\sigma^2 = E\left[(Z - E[Z])^2\right] = E\left[(Z-m)^2\right]$
Covariance of two random variables Z_i, Z_j	$\mathrm{cov}(Z_i, Z_j) =$	$E\left[(Z_i - m_i)(Z_j - m_j)\right] = \sigma_{ij}$
Correlations coefficient	$\rho_{ij} =$	$\frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$
variogram	$\gamma\left(\vec{h}\right) =$	$\frac{1}{2}E\left[\left(Z\left(\vec{x}+\vec{h}\right) - Z\left(\vec{x}\right)\right)^2\right]$

Z is a place dependent random variable with n measured values Z_a. The density function p(z) is probability that Z takes the value z_i. By computation of the inequality of two values, the variogram shows the variability of a random function, which correspond to points with distance to the vector $\vec{h}$.

Then the Kriging problem can be represented according to illustration 3.7:
We have a number of measured values Z ($\vec{x}_a$), whereby Z is a random variable and $\vec{x}_a$ is a measuring point of range D.
We assume then that Z ($\vec{x}_a$) is a subset of the random function Z ($\vec{x}$), which has the following characteristics:
It is a second order stationary function, i.e.:

1. The expected value is constant over the range D $\quad E\left[Z\left(\vec{x}+\vec{h}\right)\right] = E\left[Z\left(\vec{x}\right)\right]$

2. The covariance between two points depends only on the vector $\vec{h}$:
$\left[Z\left(\vec{x}+\vec{h}\right), Z\left(\vec{x}\right)\right] = C\left(\vec{h}\right)$

Due to these assumptions we want to compute a weighted mean, in order to get an estimated value for the place $\vec{x}_0$.

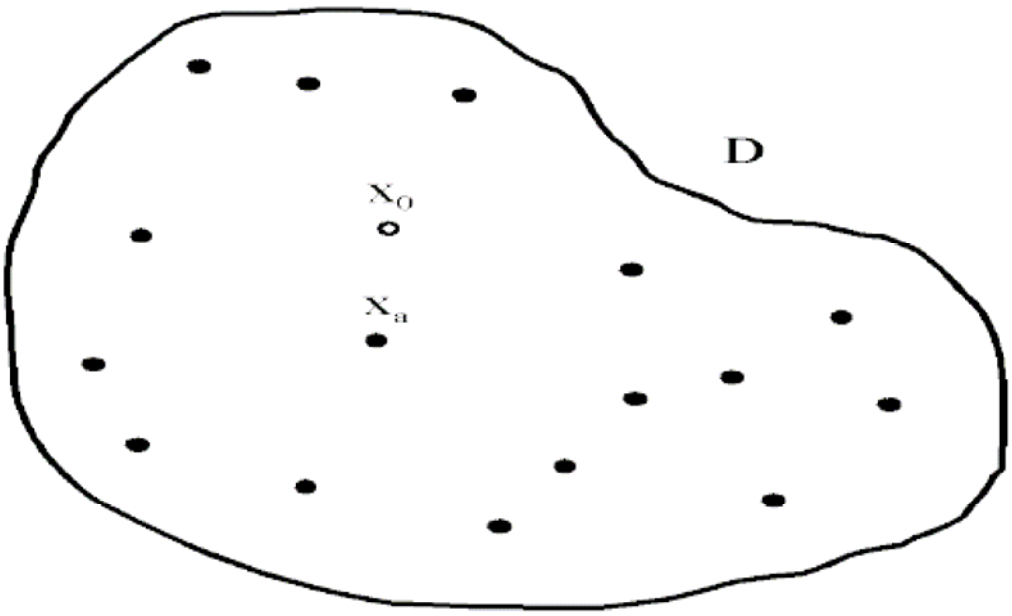


Figure 3.7: illustration of Kriging problem

The Kriging estimator $Z^*(\vec{x}_0)$ represents a linear combination of weighted sample values Z_i and n of neighbouring points:

$$Z^*(\vec{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\vec{x}_i) \tag{3.32}$$

The weights λ_i are determined in such a way that the estimated value $Z^*(\vec{x}_0)$ of the unknown true value fulfils the following conditions:

1. $Z^*(\vec{x}_0)$ is unbiased, i.e.: $E^*[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$
2. The mean square error $E[Z^*(\vec{x}_0) - Z(\vec{x}_0)]^2$ is minimal.

Under the assumption the stationarity is the expected value $E[Z(\vec{x}_i)] = m$ and $Z(\vec{x}_0) = m$. the condition 1 (unbiasedness) yields

$$E\left[\sum_{i=1}^{n} \lambda_i Z(\vec{x}_i) - Z(\vec{x}_0)\right] = \sum_{i=1}^{n} \lambda_i m - m = m\left(\sum_{i=1}^{n} \lambda_i - 1\right) = 0 \tag{3.33}$$

From this the sum of the weights must be one.

With the help of the variogram the expected value of the square error can be expressed:

$$\begin{aligned} E[Z^*(\vec{x}_0) - Z(\vec{x}_0)]^2 &= \text{var}\ (Z^*(\vec{x}_0) - Z(\vec{x}_0)) \\ &= 2\sum_{i=1}^{n} \lambda_i \gamma(\vec{x_i} - \vec{x_0}) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j (\vec{x_i} - \vec{x_j}) - \gamma(\vec{x_0} - \vec{x_0}) \end{aligned} \tag{3.34}$$

In order to minimize the error variance of the side condition 1 $\left(\sum_{i=1}^{n}\lambda_i = 1\right)$, Lagrange multiplier μ will be introduced. Then the following function is minimized:

$$\varphi = \text{var}\ \left(Z^*\left(\overrightarrow{x}_0\right) - Z\left(\overrightarrow{x}_0\right)\right) - 2\mu\left(\sum_{i=1}^{n}\lambda_i - 1\right)$$

We get the minimum by setting of the partial derivative zero $\frac{\partial\phi}{\partial\lambda_i}\ (i = 1, ..., n)$ and $\frac{\partial\phi}{\partial\mu}$.

These yield linear Kring equation system (KGS) with *n+1* equations:

$$\sum_{j=1}^{n}\lambda_j\gamma\left(\overrightarrow{x_i} - \overrightarrow{x_j}\right) + \mu = \gamma\left(\overrightarrow{x_i} - \overrightarrow{x_0}\right) \qquad \text{für } i = 1, 2, ..., n$$

$$\sum_{j=1}^{n}\lambda_j = 1$$

In matrix form the KGS is written as follows:

$$\begin{bmatrix} \gamma\left(\overrightarrow{x_1} - \overrightarrow{x_1}\right) & \gamma\left(\overrightarrow{x_1} - \overrightarrow{x_2}\right) & \dots & \gamma\left(\overrightarrow{x_1} - \overrightarrow{x_n}\right) & 1 \\ \gamma\left(\overrightarrow{x_2} - \overrightarrow{x_1}\right) & \gamma\left(\overrightarrow{x_2} - \overrightarrow{x_2}\right) & \dots & \gamma\left(\overrightarrow{x_2} - \overrightarrow{x_n}\right) & 1 \\ \vdots & \vdots & \ddots & & \vdots \\ \gamma\left(\overrightarrow{xn} - \overrightarrow{x_1}\right) & \gamma\left(\overrightarrow{x_n} - \overrightarrow{x_2}\right) & \dots & \gamma\left(\overrightarrow{x_n} - \overrightarrow{x_n}\right) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma\left(\overrightarrow{x_1} - \overrightarrow{x_0}\right) \\ \gamma\left(\overrightarrow{x_2} - \overrightarrow{x_0}\right) \\ \vdots \\ \gamma\left(\overrightarrow{x_n} - \overrightarrow{x_0}\right) \\ 1 \end{bmatrix}$$

In the case of point estimation $\gamma\left(\overrightarrow{x_i} - \overrightarrow{x_i}\right) = \gamma(0) = 0,$ i.e. the diagonal is occupied with zero.

Since in the steady case the relationship of $\gamma\left(\overrightarrow{h}\right) = C(0) - C\left(\overrightarrow{h}\right)$, $\gamma\left(\overrightarrow{h}\right)$ can be replaced by $C\left(\overrightarrow{h}\right)$ the covariance in the KGS.

Thus the diagonal of the matrix emerges large elements. In numeric aspect it is preferable therefore implemented in most programs.

The Kriging estimate variance σ^2_k for point estimation results from above equations:

$$\sigma_K^2 = \text{var}\ \left(Z^*\left(\overrightarrow{x}_0\right) - Z\left(\overrightarrow{x}_0\right)\right) = \mu + \sum_{i=1}^{n}\lambda_i\gamma\left(\overrightarrow{x_i} - \overrightarrow{x_0}\right) \qquad (3.35)$$

In a special case, in which no spatial dependence of the data exists, we get the weights $\lambda_i = \frac{1}{n}$. The Kriging estimator now the simple arithmetic means of the neighbouring samples. The following characteristics distinguish the Kriging estimator:

• The KGS is solvable only if the determinant of the matrix (γ_{ij}) = 0. Practically this means that a sample can not appear twice (i.e. with identical coordinates).

• Kriging yields an accurate interpolator.

• The KGS depends only on $\gamma\left(\vec{h}\right)$ or $C\left(\vec{h}\right)$, however not on the values of the variable Z in the points of sample x_i. With identical data configuration the KGS only need to be solved once.

• Confidential limits of the estimation can be indicated under the help of the estimation error σ_K.

In practice a series of Kriging procedures were developed and applied, which regard more complex situations, e.g. intermittent variable, space time dependence etc.

3.3.1 Task for application of interpolation method

Interpolate by means of

· Analytical power function

· LAGRANGE

· NEWTON

· Spline function

The following are measured value tables

1. For normal distribution function is tabulated $y(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$

x	1,00	1,20	1,40	1,60	1,80	2,00
y	0,2420	0,1942	0,1497	0,1109	0,0790	0,0540

and find out the value of y(1.50)

2. Please interpolate $\sqrt{1,03}$ and $\sqrt{1,26}$ on the basis of the table.

x	1,00	1,05	1,10	1,15	1,20	1,25	1,30
$y = \sqrt{x}$	1,00000	1,02470	1,04881	1,07238	1,09544	1,11803	1,14017

3. A rational function with degree as low as possible is supported by three points: (1, -2); (2, 3); (3, 1)? How does this interpolation function change, if another supporting point (4, 4) is taken into account?

Chapter 4

4 Optimisation problem

4.1 Analytical solution of extreme value problems

4.2 Iterative optimum search

4.3 Least squares methods (MKQ)

In water management practice the experimental process analysis (see section 11.1 model classification, page 284 and the following page) is used for the parameter determination of underground systems, e.g. k -, S- and T-values of soils, degradation rates, transportation parameters. The mathematical model structure is specified by a theoretical process analysis. we try to transfer this model structure into easily solvable representations. The parameters can be determined by the solving conditional equations or by a parameter approximation problem. So the task is, on the basis of structure knowledge or assuming such a model or such parameter sets, to develop

- the characteristics of the system which reflect reality as exact as necessary and
- eliminate the superposed influence of noise and errors to a large extent

For the fulfilment of these demands the comparison of the output value serves as function of the inputs or an independent variable (time or place). In the result a change of the parameters is to be made or the model change itself until the deviation reaches minimum. The changes can be carried out according to a certain strategy (search algorithms, optimisation programs), statistically (random number generator) or empirically. Also the visual comparison between the two diagrams (original and model output signal) is possible.

This task is also called parameter estimation. Particularly the procedure which is introduced here is classified as iterative estimation.

By the algorithmic model adjustment (see figure 4.1) we try to let the input vector and the manipulated vector *y* work in the process as well as in the model. With a first parameter set, the starting parameter, the output vector of the model x^1_M can be computed by first approximation. The deviation of this vector from the output vector *x* of the process (x_i - x_{Mi}) is named as quality of the adjustment of the model. In water management applications the square evaluation will be carried out. The aim of transformation of parameter is minimizing the value $Q = \sum (x_i - x_{Mi})^2 \Rightarrow$ Min.

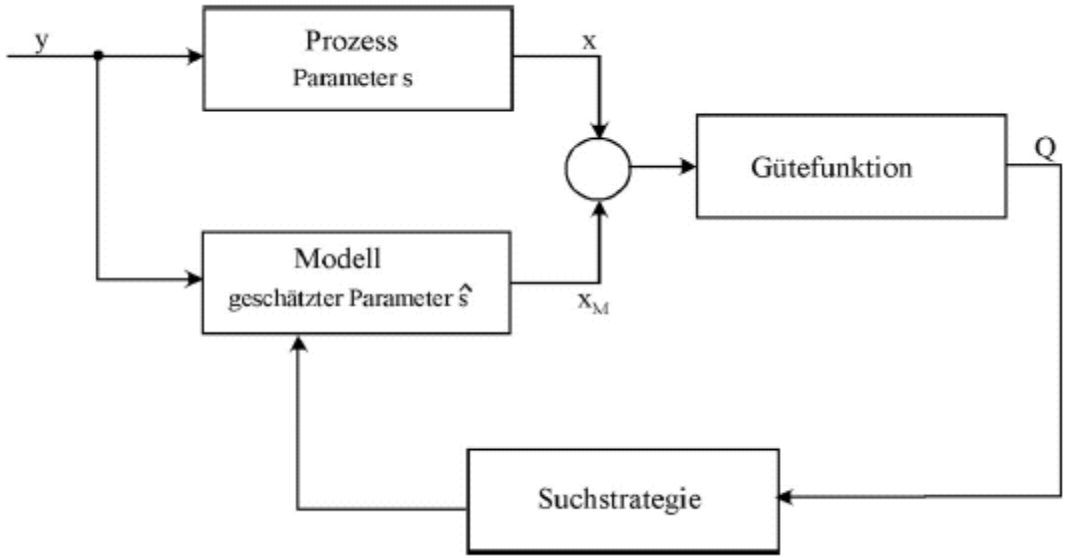


Figure 4.1: Iterative Model adjustment

4.4 Retrieval Strategy

In these optimisation tasks it is very crucial that at what processing time the minimum is found. The processing time T_v depends on the basic computing time T_n for the numeric analysis of the model and the number of iteration steps n. The number of solution procedure is mainly determined by four influences:

der Zahl der zu suchenden Parameter; sie entspricht der Anzahl der Suchrichtungen
und geht damit exponentiell ein,

- the number of the parameters which are looked for; it corresponds to the number of search directions and shrinks exponentially

- the formation of the quality mountains, i.e. the slope and the number of subminimum,

- the search strategy, whereby accuracy for the correct direction, the search step length and cognition of subminimum

- the starting parameters, which crucially prevent unnecessary search steps.

The formation of the quality mountains and the search strategy can not be regarded independently. Generally it must be noted that there is not a "best" search program, but for certain classes of quality mountains appropriate procedures are particularly suitable.
There is a series of procedures to solve of optimisation problems. We divide these search methods according to their search strategy into non-gradient-, gradient- and coincidence- search method.

In table 4.1 the most important procedures for iterative processes of estimation (WERNSTEDT) are compiled.

Table 4.1: iterative estimation method (WERNSTEDT)

Verfahren	Lösungsgleichung für $\hat{s}$ $\hat{s}^{i+1}=\hat{s}+\Delta\hat{s}^{i+1}$	Beispielsverfahren
Gradientenfreie Verfahren	$\Delta\hat{s}^{i+1}=a(i)\bullet\Delta\hat{s}$ wenn $a(i)\begin{cases}>\\=\\<\end{cases}0$ dann $Q(i)\begin{cases}<\\=\\>\end{cases}Q(i-1)$ $\Delta\hat{s}$ vorgegebener Vektor	POWELL ROSENBROCK FIBONACCI
Gradientenverfahren	$\Delta\hat{s}^{i+1}=-g(i+1)\bullet R(i+1)\nabla Q(i)$ $g(i)$ Schrittweitenvektor $R(i)$ Richtungsmatrix $\nabla Q(i)$ Gradient zum i-ten Schritt	steilster Abstieg NEWTON NEWTON RAPHSON MARQUART LEVENBERG POWELL JONES
Zufallssuchverfahren	$\Delta\hat{s}^{i+1}=b(i+1)\bullet R(i+1)\Delta\hat{s}$ $b(i+1)=\begin{cases}0\text{ für }Q(i+1)\geqq Q(i)\\1\text{ für }Q(i+1)<Q(i)\end{cases}$ $\Delta\hat{s}$ Zufallsvektor	BROOKS RASTRIGIN WHITE SCHWEFEL

4.4.1 JONES Spiral method

The methods of the nonlinear regression, which is best suitable for pumping test evaluations, is aimed at adapting model functions $x_M(\hat{s})$ by choosing the parameters $\hat{s}$ to given values (measured values) $x(s)$. The deviations between $x_M(\hat{s})$ and $x(s)$ are shown with weight factors the W. In the case we assume that, n samples (measured values) are available from the process and the model is determined by k independent parameters.

Starting from initial value $\hat{s}_0$ the goal function is to be minimized concerning parameters $\hat{s}_i$ $(I = 1,k)$

$$Q(\hat{s}) = \sum_{i=1}^{n} \left(W_i \left(x(s)_i - x_M(\hat{s})_i\right)^2\right) \tag{4.1}$$

This complies with the requirement of the least square error method.

Essentially iteration exists in the solution of the linear equation system:

$$\sum_{j=1}^{k} A_{i,j} \cdot T_j = G_i \tag{4.2}$$

with

$$G_i = \frac{dQ_i}{d\hat{s}_i} \qquad (i = 1, k) \tag{4.3}$$

$$G_j = \sum_{i=1}^{n} W_{ii} \cdot \left(x(s)_i - x_M(\hat{s}_0)_i \frac{dx_{Mi}}{d\hat{s}_j}\right) \tag{4.4}$$

and

$$A_{i,j} = \sum_{k=1}^{n} W_{kk} \cdot \frac{dx_{Mk}}{d\hat{s}_i} \cdot \frac{dx_{Mk}}{d\hat{s}_j} \tag{4.5}$$

T_j represents the change of j-th parameter. We can get the equation system by developing TAYLOR expansion of the objective function at the place S_0. If we set the partial derivatives of this function equal to zero, we get an equation system as above.
To check whether the linear approximation is adequate, the inequality must be fulfilled.

$$Q(\hat{s}_0 + T) < Q(\hat{s}_0)$$

It is not always like this case in practical. According to JONES we find a better goal function value by a vector manipulation between the TAYLOR direction T and the negative gradient direction G (see figure 4.2).

The minimum function value within the iteration is expected in the place $\hat{s}_0 + T$. On the other hand the goal function decreases in the direction of negative gradients. Thus it is sure that there is a better goal function value within the triangle $\hat{s}_0 - (\hat{s}_0 + T) - (\hat{s}_0 + G^*)$. G* has the direction of G and the modulus of T. we act on the assumption of TAYLOR step in this search.

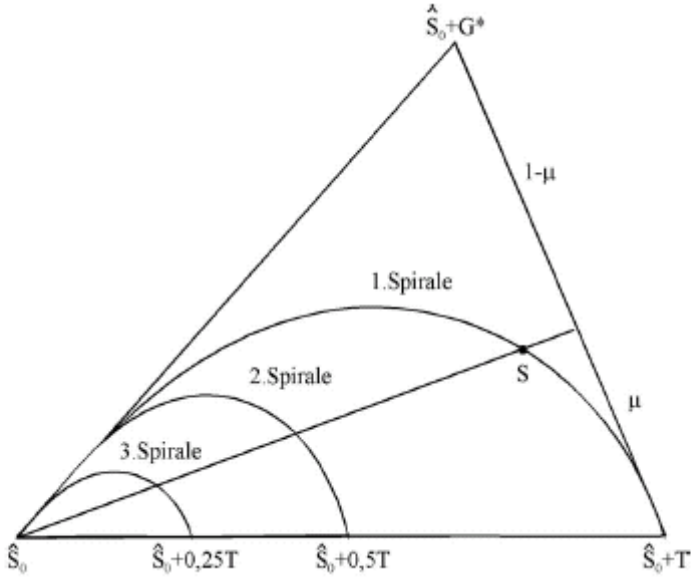


Figure 4.2: JONES spiral algorithm

Iteration is terminated, if a better goal function value were found. If the TAYLOR step is not successful, points will be calculated at the 1st spiral, which are shown:

$$S = A_1 \cdot (\mu G^* + (1 - \mu)T) \tag{4.6}$$

The different s-values are attained by change of the μ-value. μ begins with 0, 1 and is computed by the following relationship ($Z \geq 2$).

$$m_{n+1} = \frac{Z \cdot m_n}{1 + (Z - 1) \cdot m_n} \tag{4.7}$$

If μ >0.9 the search will stop on the current spiral. If $Q(\hat{s}_0 + S) \geq Q(\hat{s}_0)$ the vector T is halved and the next spiral will be searched.

The larger Z is, the fewer points on a spiral are computed. If possible we interpolate either on the spiral or in TAYLOR direction. If no better value is found even along the last spiral, search will be carried out in negative gradient direction with smaller steepening increment.

Chapter 5

5 Ordinary differential equation

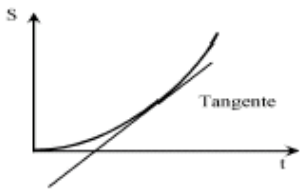
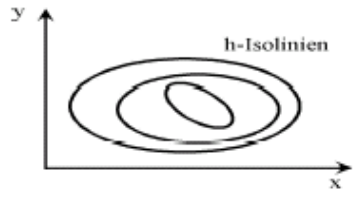
Ordinary differential equations are characterized by the fact that the searched function is dependent on a variable, while in the partial differential equations (PDE) more arguments and their appropriate derivatives appear as the following examples:

$\frac{dx}{dt} + t^2 \cdot x = 2t^2$ Ordinary Differential Equation (ODE)

$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} = 0$ Partial differential equation (PDE)

Their illustrations are evident in table 5.1.

Table 5.1: illustration of the differential equations

	Gewöhnliche DGL	Partielle DGL
Anzahl	eine Variable (unabhängige Veränderliche)	mehrere Variable (unabhängige Veränderliche)
Variablen	x, y, z oder t	x, y, z und/oder t
Beispiel	$v = \frac{ds}{dt}$	$\vec{v} = k \cdot \text{grad}\,(h)$ $\vec{v} = k \cdot \left(\frac{\partial h}{\partial x}\vec{i} + \frac{\partial h}{\partial y}\vec{j} + \frac{\partial h}{\partial z}\vec{k}\right)$
Grafik	 1-D-Funktion mit Tangente	 Spitzkegel mit beliebig gekrümmter Oberfläche

In the following derivations and examples with the ordinary differential equations it is assumed that "*x*" stands for the function value and "*t*" serves as argument. Of course all propositions can be also assigned to other arguments, and for dependent functions arbitrary variable names can be used. The particular use of the letter *x* as a symbol of variable name appears in many mathematic teaching materials and in the signal theory (See GRÄBER: Lehrmaterial zur Automatisierungstechnik bzw. Grundwassermesstechnik). In the partial differential equations x, y and z are used as independent local coordinates.

The general form of an ODE is:

$$F\left[t, x(t), \frac{dx(t)}{dt}, \frac{d^2x(t)}{dt^2}, \cdots\cdots, \frac{d^nx(t)}{dt^n}\right] = g(t) \tag{5.1}$$

These differential equations are identified according to table 5.2, if correspondent conditions fulfilled.

Table 5.2: identification of differential equations

Bezeichnung	Bedingung	Beispiel
Ordnung der DGL	n	$\frac{d^nx}{dt^n} + \frac{d^{n-1}x}{dt^{n-1}} + \frac{d^{n-2}x}{dt^{n-2}} = 0$
inhomogen	$g(t) \not\equiv 0$	$\frac{dx}{dt} + t^2 \cdot x = 2t^2$
homogen	$g(t) = 0$	$\frac{dx}{dt} + t^2 \cdot x = 0$
explizit	$\frac{d^nx}{dt^n} = 0$	$\frac{d^2x}{dt^2} = 0$
implizit	$F\left[t, x, \frac{dx}{dt}, \frac{d^nx}{dt^n}\right] = 0$	$\frac{dx}{dt} + t^2 \cdot x - 2t^2 = 0$
linear	$a_1(t) \not\equiv a_1(x,t)$	$t\frac{dx}{dt} + t^2x = 2t^2$
nichtlinear	$a_1(t) = a_1(x,t)$	$x\frac{dx}{dt} + t^2x = 2t^2$

5.1 Setting up equations

In the further sections solving differential equations is based on the mathematical description of natural processes. The derivative of mathematical equations as transformation of natural processes is noted as modelling and as the transformation of mathematical model. The development of such mathematical models is the subject of section 11.2.1 theoretical process analysis, page 291 as well as 11.2.2 experimental process analysis, page 292. The method described here is only how the mathematical models to be completed according to the physical or chemical basic laws and their effects. This way of the theoretical process analysis, also designated as mathematical modelling, is generally preferred by natural scientists.

In the theoretical process analysis the reciprocal effects of the process variables, state variables are formulated as mathematical model equations with their neighbourhoods. The most substantially reciprocal effects between the system and its neighbourhood are divided into causes and effects. The causes and the effects are called input and output variables. The description by means of the physical or chemical basic law is usually in formation of balance equation, particularly the formation of the energy and mass balance equations.

Most energy balance equations lead to force equilibrium law and flux laws. Generally we can speak of the transformation from potential to kinetic energy. Such energy transformations take place on so called flow resistances. A kinetic energy in form of material or mass flow results from different potential energies at the in- or outflow resistance (e.g. conduit, aquifer, electrical resistance), which act as driving force. We also say that flow resistances corresponding potential energy, also called as potential, is abolished, "drops" (e.g. pressure difference, voltage drop).

The mass balance equation assumes that mass is neither created nor destroyed within a regarded system (e.g. container, representative unit volume). The mass balance of a system can only be changed by outside sources or sinks. If dynamic systems are considered, the storage effect must be included likewise the mass balance. This means that all mass flows, which affect a system, must be zero in sum (junction law).

Mathematically this circumstance can be also described by the divergence of a flow vector, which must be zero in this case $(div\,\vec{v} = 0)$.

Examples of setting up differential equations:

Please find out the relationship of the flow rate V, which flows out from a pipe with free gradient, if this is attached at a container (see figure).

Task of setting up differential equation:

1. The padding to the remainder holes of the former brown coal open pit caused by the rise of groundwater under natural conditions will last too long time. Therefore external supply is introduced to the filling procedure for acceleration.

Set up the differential equation for the padding procedure $h_{(1,2)}(t)$, without consideration of the aquifer and contingent ground water regeneration rate.
Initial condition ($h_{t=0(1,2)} = 0$) is given for all cases.

a) Constant flow rate (see figure 5.1)

b) Variable flow rate (see figure 5.2)

c) Coupled storage cascade (see figure 5.3)

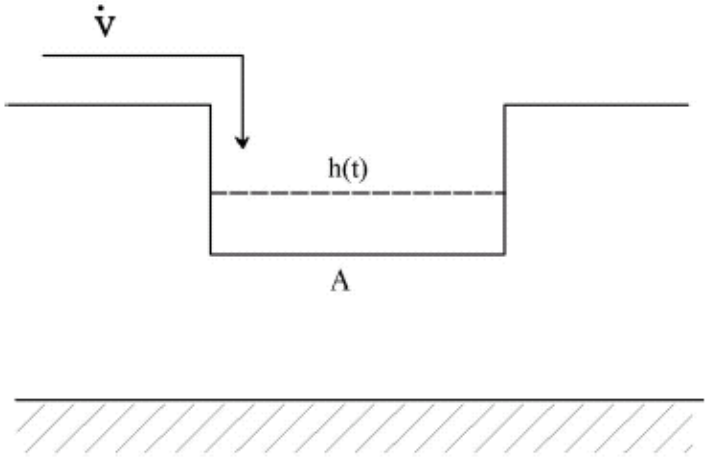


Figure 5.1: filling procedure of a remainder hole with constant flow rate

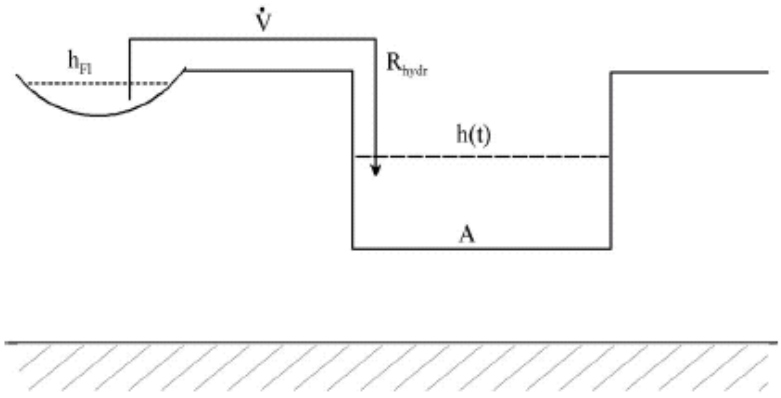


Figure 5.2: filling procedure of a remainder hole with variable flow rate

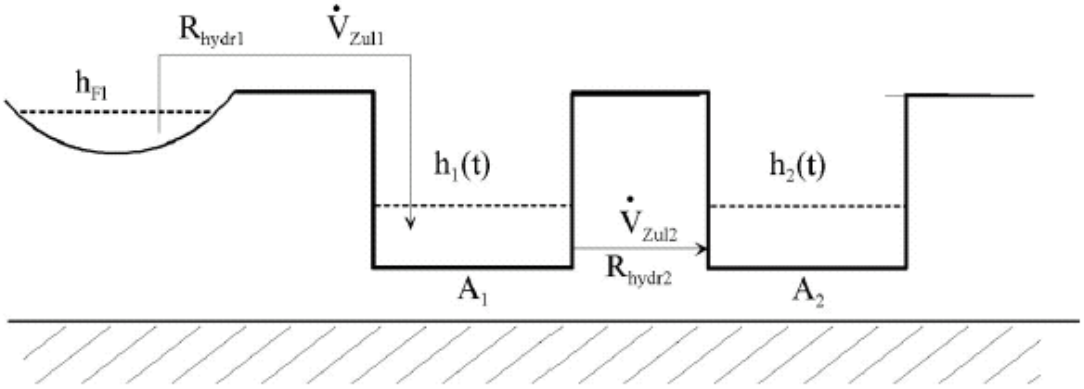


Figure 5.3: coupled storage cascade

2. Set up differential equation for the following hydraulic scheme (see figure 5.4) with associated block model.

Assume a linearized relationship and a homogeneous, isotropic aquifer with the following parameters $k = 5 \cdot 10^{-4}\ m/s$; $n_0 = 0.2$; $z_{Rmittel} = 20m$; $l = 50m$:

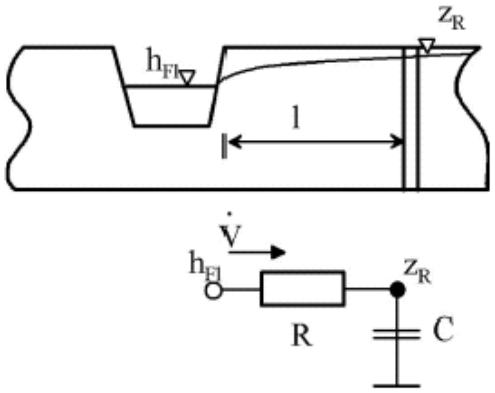


Figure 5.4: schematic illustration of the groundwater level with block diagram

3. A float control is used for the water level regulation of an irrigation ditch (see figure 5.5). Set up differential equations to calculate water level H. The surface of the container is A. The flow rate V is dependent on the water level H.

$$\dot{V} = K \cdot \dot{V}_{max} \cdot (H_{max} - H)$$

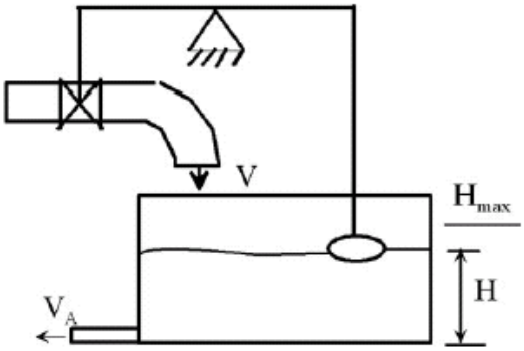


Figure 5.5: Water level control of an irrigation ditch

5.2 Analytical solution methods

5.2.1 First order Ordinary differential equations

One solution for the following inhomogenous first order ODE should be found

$$a_1(t) + a_0(t)x = g(t)$$

The following writing ways are often used for short:

$$\frac{dx}{dt} = \dot{x}$$
$$\frac{dy}{dx} = y'$$

Firstly it will be transferred into a homogeneous ODE in order to solve the inhomogenous one.

$$a_1(t)\frac{dx_h}{dt} + a_0(t)x_h = 0 \qquad (5.2)$$

There are several methods to solve homogeneous ODE, and the **separation of variables** and the **substitution method** are described here.

5.2.1.1 Solution of homogeneous differential equation

First order ODE:

$$a_1(t)\frac{dx_h}{dt} + a_0(t)x_h = 0 \qquad (5.3)$$

For simplification the functions a_0 and a_1 are regarded as constants.

• Separation of variables

The method of variable separation is aimed at rearranging the ODE algebraically such that, there is a total differential on each side of the equation which is conveniently integrable.

$$a_1 \frac{dx_h}{dt} + a_0 x_h = 0 \tag{5.4}$$

$$\frac{dx_h}{dt} = -\frac{a_0}{a_1} x_h$$

$$\frac{dx_h}{x_h} = -\frac{a_0}{a_1} dt$$

$$\int \frac{1}{x_h} dx_h = -\frac{a_0}{a_1} \int dt$$

$$\ln x_h + C_1 = -\frac{a_0}{a_1} \cdot t \qquad \text{oder} \qquad \ln x_h - \ln C_2 = -\frac{a_0}{a_1} \cdot t$$

$$\ln x_h = -\left(\frac{a_0}{a_1} \cdot t + C_1\right) \qquad \ln x_h = \ln C_2 - \left(\frac{a_0}{a_1} \cdot t\right)$$

$$x_h = e^{-\left(\frac{a_0}{a_1} \cdot t + C_1\right)} \qquad x_h = C_2 \cdot e^{-\left(\frac{a_0}{a_1} \cdot t\right)}$$

Both solutions are applicable and transferable with each other based on logarithm laws (see section 1.1, page 2).

By equating both equations we get:

$$C_2 = e^{C_1} \text{ bzw. } C_1 = -\ln C_2 \tag{5.5}$$

Since the integration constants C_1 and C_2 are still indefinite as well as the logarithm and the exponential function, both two solutions are equivalent. The constants can be determined from at the initial or final conditions, e.g. C_1 and C_2 can be determined at the point t = 0 with the known initial condition x_{h0}:

$$C_1 = -\ln x_{h0} \tag{5.6}$$

$$C_2 = x_{h0}$$

The solution of the homogeneous ODE:

$$x_h = x_{h0} \cdot e^{-\left(\frac{a_0}{a_1} t\right)} \tag{5.7}$$

Application of Separation of variables:

Solve the ODE:

$$\frac{dx}{dt} + t^2 \cdot x = 0, \qquad \text{wobei gilt: } x_{t=0} = 3$$

According to the algorithm we try to separate the total differentials (dx and dt) respectively on each side of the equation.

$$\begin{aligned}
\frac{dx}{dt} &= -t^2 x \\
\frac{1}{x} dx &= -t^2 dt \\
\int \frac{1}{x} dx &= \int -t^2 dt \\
\ln x &= -\frac{1}{3} t^3 + C_1 \\
x &= e^{-\frac{1}{3}t^3 + C_1} \\
x &= C_2 e^{-\frac{1}{3}t^3}
\end{aligned}$$

We insert $x_{t=0} = 3$ in the general solution, it yields $C_2 = 3$, so the answer is $x = 3e^{-\frac{1}{3}t^3}$.

• Substitution method

Basic idea of the substitution method is to find a possible solution by means of insertion, which have been proved with experiences. The most well known substitutions are combinations of exponential functions or sine functions as well as general power series. The advantage is no implementation of difficult integral operations, only the insertion to be differentiated, which is often more simply to realize:

Differential equation: $$a_1 \cdot \frac{dx_h}{dt} + a_0 x_h = 0 \tag{5.8}$$

substitution: $$x_h = K \cdot e^{\lambda t}$$

derived from: $$\frac{dx_h}{dt} = \lambda \cdot K \cdot e^{\lambda t}$$

inserting in homogenous differential equation: $$a_1 \cdot \lambda \cdot K \cdot e^{\lambda t} + a_0 \cdot K \cdot e^{\lambda t} = 0$$

$$\lambda = -\frac{a_0}{a_1}$$

The reciprocal value of this constant λ, which is in unit of time in the case, is also often called time constant T or τ.

$$x_h = K \cdot e^{-\frac{a_0}{a_1}t} = K \cdot e^{-\frac{t}{T}} = K \cdot e^{-\frac{t}{\tau}} \tag{5.9}$$

As in the method of separation of variables, the constants are determined from initial or final conditions. I.e. K can be determined at the point t = 0 with x_{h0}:

$$K = x_{h0} \tag{5.10}$$

It yields:

$$x_h = x_{h0} \cdot e^{-\left(\frac{a_0}{a_1}t\right)} \tag{5.11}$$

the same solution for ODE is obtained like other methods.

5.2.1.2 Solution of the inhomogenous differential equation

A general inhomogenous differential equation can be written in the form:

$$a_1 \frac{dx}{dt} + a_0 x = g(t) \tag{5.12}$$

its general solution results from adding the homogeneous solution $x_h(t)$ to a particular solution $x_p(t)$, i.e.

$$x(t) = x_p(t) + x_h(t) \tag{5.13}$$

We can get the particular solution of differential equation for example by variation of constants method, which assumes the homogeneous solution and takes the existing constant, here the time, as function of the arguments.

• Variation of constants method

The solution will be carried out in four steps:

Differential equation: $a_1 \frac{dx}{dt} + a_0 x = g(t)$

1st step: dismember: $a_1 \frac{dx_h}{dt} + a_0 x_h = 0$

2nd step: variation separation: $\frac{dx}{x_h} = -\frac{a_0}{a_1} dt$

solution of homogenous differential equation: $x_h = C e^{-\frac{a_0}{a_1} t}$

3rd step variation of constant: $x_p = C(t)\, e^{-\frac{a_0}{a_1} t}$

According to the rule of product differentiation:

$$\frac{dx_p}{dt} = \frac{dC}{dt} \cdot e^{-\frac{a_0}{a_1} t} + C \cdot \left(-\frac{a_0}{a_1}\right) \cdot e^{-\frac{a_0}{a_1} t} \tag{5.14}$$

4th step insertion in differential equation: $\frac{dx_p}{dt} + a_0 x_p = g(t)$

$$a_1 \cdot \left(\frac{dC}{dt} \cdot e^{-\frac{a_0}{a_1} t} + C \cdot \left(-\frac{a_0}{a_1}\right) \cdot e^{-\frac{a_0}{a_1} t}\right) + a_0 \cdot C \cdot e^{-\frac{a_0}{a_1} t} = g(t)$$

$$a_1 \cdot \frac{dC}{dt} \cdot e^{-\frac{a_0}{a_1} t} + a_1 \cdot C \cdot \left(-\frac{a_0}{a_1}\right) \cdot e^{-\frac{a_0}{a_1} t} + a_0 \cdot C \cdot e^{-\frac{a_0}{a_1} t} = g(t)$$

$$a_1 \cdot \frac{dC}{dt} \cdot e^{-\frac{a_0}{a_1} t} = g(t)$$

$$\frac{dC}{dt} = \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} \tag{5.15}$$

This differential equation can be treated according to the above methods for solution of homogeneous differential equation, e.g. by means of separation of variables:

$$dC = \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt \tag{5.16}$$

$$\int dC = \int \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt$$

$$C(t) = \int \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt$$

Thus the general solution of the differential equation:

$$x(t) = x_p(t) + x_h(t) \tag{5.17}$$

$$x = Ce^{-\frac{a_0}{a_1}t} + \int \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt \, e^{-\frac{a_0}{a_1}t} \tag{5.18}$$

The constants are determined by initial or final conditions, as in the method of variables separation, e.g. C can be determined at the point t = 0 with x_{h0}:

$$C = x_0 - \left(\int \frac{g(t)}{a_1} \cdot e^{\frac{a_0 t}{a_1}} dt \right)_{t=0} \tag{5.19}$$

Then it yields:

$$x(t) = \left(x_0 - \int \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt \right)_{t=0} \cdot e^{-\frac{a_0}{a_1}t} + \left(\int \frac{g(t)}{a_1} \cdot e^{+\frac{a_0}{a_1}t} dt \right) \cdot e^{-\frac{a_0}{a_1}t} \tag{5.20}$$

The solution possibility thereby depends on the integrability of the perturbation function g(t).

Tips:

The integration of two functions product is only possible in some functions. Particularly, if a function is the primitive function or the derivative of the others, the following substitution can be introduced:

$$\int u \cdot dv = u \cdot v - \int v \cdot du \tag{5.21}$$

That what we should keep in mind is not final formula, but the way of:

1. Transfer into a homogeneous differential equation (mutilating)
2. Separation of variables
3. Variation of the constants or substitution method
4. Insertion in the differential equation

Remarks on the method of the variation of the constants:

1. It can be only used for linear differential equations.
2. The general solution is linearly dependent on the constants.
3. The general solution has a member of free constants which are received from the particular solution of the inhomogenous differential equations.
4. It frequently occurs that a nonlinear differential equation is transferred into a linear one by a simple substitution.

Examples of solution of inhomogeneous differential equation:

1. Find out the solution of

$$t\dot{x} - x = t^2 \cos t$$

Solution:

Differential equation:

1. Dismembering: $t\dot{x} - x = t^2 \cos t$

2. variables separation: $t\dot{x} - x = 0$

$$\frac{dx}{x} = \frac{dt}{t}$$

3. variation of constants: $x = Ct$

$$x = C(t)\,t$$

4. Insertion: $\dot{x} = \dot{C}t + C$

$$t\left(\dot{C}t + C\right) - Cx = t^2 \cos t$$

$$t^2\dot{C} = t^2 \cos t$$

$$\dot{C} = \cos t$$

General solution: $C = \sin t + C_1$

$$x = (C_1 + \sin t)\,t = t\sin t + C_1 t$$

2. Find out the solution of

$$2tx \cdot \dot{x} + t^2 - x^2 + 1 = 0$$

Solution:

By the substitution $z = x^2$ and $z = 2\,x \cdot x$ the equation becomes a linear differential equation of unknown function z. It can be solved according to the substitution method as follows:

Differential equation: $2tx\dot{x} + t^2 - x^2 + 1 = 0$

Substitution: $z = x^2$

Substituted equation: $t\dot{z} - z = -1 - t^2$

1. Dismembering: $t\dot{z} - z = 0$

2. variables separation: $\frac{dz}{z} = \frac{dt}{t}$

$z = Ct$

3. variation of constants: $z = C(t)t$

$\dot{z} = \dot{C}t + C$

4. Insertion: $t\left(\dot{C}t + C\right) - Ct = -1 - t^2$

$t^2\dot{C} = -1 - t^2$

$\dot{C} = -\frac{1}{t^2} - 1$

$C = \frac{1}{t} - t + C_1$

General solution for z: $z = 1 - t^2 + C_1 t$

Back substitution: $x^2 = 1 - t^2 + C_1 t$

General solution for x: $x = \sqrt{1 - t^2 + C_1 t}$

5.2.1.3 Task of solving first order differential equation

1. Give the general solution of the following differential equation:

a) $y' = (y-3)\cos x$

b) $y' = e^{x+y}$

c) $y' \sin x = y \ln y$

d) $2xy' + \frac{y^2}{x} = 0$

e) $y' + y + e^x = 0$

f) $y' + \frac{y}{x} = \sin x$

g) $\frac{dx}{dt} + t^2 \cdot x = 2t^2$

h) $y' = -xy^2$ mit $y(0) = 2$

i) $\frac{dx}{dt} + t^2 x = 0$ mit $x(0) = 3$

j) $t\frac{dx}{dt} - x = t^2 \cos t$ mit $x(\pi/2) = \pi$

2. Differential equation $Tx_a + x_a = Kx_e$ for a system with simple memory effect (x_a output value, x_e input value, T time constant, K proportional transfer function).
How the output value x_a changes dependent on time t if x_e = ct (C = const.)?

3. Determine in each case the general and the special solution by specified initial conditions:

a) $y' = xy + 2x$ mit $y(0) = 2$

b) $y' + x^2 y = x^2$ mit $y(2) = 1$

4. Differential equation applies to the hydraulic scheme (see figure 5.6) with associated block diagram:

$$h_{Fl} = R \cdot C\frac{dz_R}{dt} + z_R$$

It is assumed a linearized relationship and a homogeneous, isotropic aquifer with the following parameters $k = 5 \cdot 10^{-4}\frac{m}{s}; n_0 = 0,2; z_{Rmittel} = 20m; l = 50m$. Compute the change of the water level, if the river surface changes as a first approximation as follows:

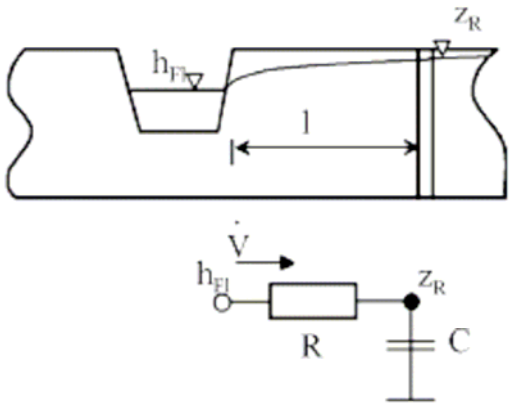


Figure 5.6: Schematic representation of the groundwater level

a) Erratic ($h_{Fl} = h_{Fl} \cdot l(t)$) and
b) sinusoidal ($h_{Fl} = h_{Flm} sin(\omega \cdot t) + h_{Fl0,}$ *with* $\omega = 2\pi\Gamma$ *and* $\Gamma = 7$ *days*)

5. The following differential equation applies to the concentration C in sorption of pollutants at the soil matrix:

$$T_1 C + C = K$$

T_1 is time constant and K is a constant. $T_1 = 1d^{-1}$, K = 100. The concentration should C(0) = 0 at time $t = 0$.

a) Solve the differential equation by means of the analytic methods and compute the concentration change for the time $t = 1d$
b) Outline the time process of concentration change.

6. The padding to the remainder holes of the former brown coal open pit caused by the rise of groundwater under natural conditions will last too long time. Therefore external supply is introduced to the filling procedure ($h_{t=0} = 0$) for acceleration. (see figure 5.7)

$$A\frac{dh}{dt} = \dot{V}_{Zustr}$$

Solve the differential equation by means of the analytic methods.

7. Solve the following differential equation by means of the analytic methods:

$$\frac{dh}{dt} + k \cdot h = g \quad \text{mit} \qquad h_{t=0} = 0$$

$g = 0,015m \cdot s^{-1}$ und $k = 0,01s^{-1}$

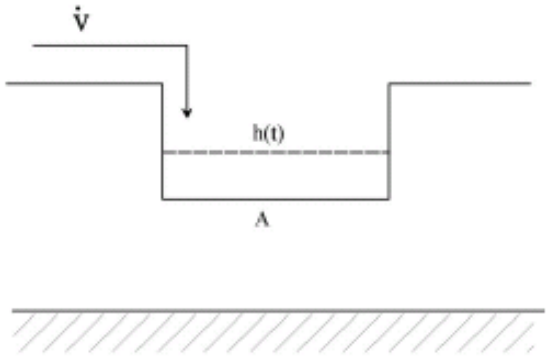


figure 5.7: filling procedure of a remainder hole

5.2.2 Ordinary differential equations of higher order

A general solution of a differential equation with *n*-th order has *n* constants and represents geometrically a *n*-parametric curve family. For determination of a single solution from this crowd we need *n* initial- or boundary conditions.

Example of a 2. order differential equation:

$y^2y' + y^2 - 1 = 0$ is given for the movement of a particle. The general solution of these differential equation is $(x - C)^2 + y^2 = 1$. This equation stands for all circles of the radius r = 1 with the centre on the x axis. According to initial condition y (0) = 1 yields C = 0; then the single solution is $x^2 + y^2 = 1$. The particle moves around the circle with radius r = 1 whose centre is on the origin of the coordinate system.

Different types of higher order differential equation can be solved with different methods:

5.2.2.1 Differential equation of type a

$$\lambda \cdot \frac{d^2y}{dt^2} + \frac{dy}{dt} = 0 \qquad (5.22)$$

The degrees of higher order differential equation can be reduced by means of the following substitution:

$$z = \frac{dy}{dt} \qquad (5.23)$$

Then the derivative:

$$\frac{dz}{dt} = \frac{d^2y}{dt^2} \qquad (5.24)$$

These two equations are inserted into the differential equation:

$$\lambda \cdot \frac{dz}{dt} + z = 0 \qquad (5.25)$$

According to the rules for homogeneous 1. Order differential equation(see section 5.2.1.1, page 111):

$$z = k_1 e^{-\frac{t}{\lambda}} \qquad (5.26)$$

or

$$\frac{dz}{dt} = -\frac{k_1}{\lambda} e^{-\frac{t}{\lambda}} \qquad (5.27)$$

Due to the substitution condition we get:

$$z = \frac{dy}{dt} \qquad (5.28)$$
$$\frac{dy}{dt} = k_1 e^{-\frac{t}{\lambda}}$$

This differential equation can be solved again with the method of the separation of the variables:

$$\frac{dy}{dt} = k_1 e^{-\frac{t}{\lambda}} \qquad (5.29)$$
$$\int dy = \int k_1 e^{-\frac{t}{\lambda}} dt$$
$$y = -\lambda \cdot k_1 \cdot e^{-\frac{t}{\lambda}} + k_2$$

Since here two constants exist, two condition equations must be found. t = 0 and t = 1 are supplied to e^{-t} functions, then the exponential function simple values (1 and 0) yields:

$$k_2 = y(\infty)$$
$$k_1 = -\frac{y(0) - y(\infty)}{\lambda}$$

And the solution:

$$y(t) = (y(0) - y(\infty)) \cdot e^{-\frac{t}{\lambda}} + y(\infty) \qquad (5.30)$$

Remarks:

This solution method can be applied also for the differential equation

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} = g(t) \qquad (5.31)$$

The substitution $z = dy/dt$ leads to a linear differential equation, which can be solved by method variation of the constants.

5.2.2.2 Differential equation of type b

$$a\frac{d^2y}{dt^2} + b\frac{dy}{dt} + cy = 0 \qquad (5.32)$$

This differential equation is to be solved according to **substitution method**. Sense and purpose of the substitution method are to avoid complicated operations of the integration of the differential equation and only carry out substantially simple operations of the deviation implement. A popular substitution is used here, which looks promising from the experience. At the beginning all derivatives are developed, which appear in the differential equation.

Following substitution and derivatives:

$$y = C \cdot e^{\lambda t} \qquad (5.33)$$
$$\frac{dy}{dt} = C \cdot \lambda \cdot e^{\lambda t}$$
$$\frac{d^2y}{dt^2} = C \cdot \lambda^2 \cdot e^{\lambda t}$$

These are inserted to the differential equation:

$$\left(a\lambda^2 + b\lambda + c\right) \cdot C \cdot e^{\lambda t} = 0 \qquad (5.34)$$

For t ≠ -∞ we can divide $e^{\lambda t}$:

$$a\lambda^2 + b\lambda + c = 0 \qquad (5.35)$$

If we introduce this to the standard format of quadratic equation, then it yields new constants $d = b/a$ and $f = c/a$:

$$\lambda^2 + \frac{b}{a}\lambda + \frac{c}{a} = 0 \qquad \text{bzw.} \qquad \lambda^2 + d\lambda + f = 0 \qquad (5.36)$$

This equation is also designated as **characteristic equation of differential equation**. And the general solution of this characteristic equation is:

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{\frac{d^2}{4} - f} \qquad (5.37)$$

Depending upon the coefficients d and f there are three different cases:

1. **1st case**, when $d^2/4 - f > 0 \Rightarrow d^2/4 > f$, *or* $b^2 > 2 \cdot c \cdot a$, then $\lambda_1 \neq \lambda_2$ and real number

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{\frac{d^2}{4} - f} \qquad (5.38)$$

The solution:

$$y = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} \qquad (5.39)$$

This case results in an asymptotic curve, which approaches a final steady state. This lies in the real number range if λ_1 and λ_2 take negative values.

2. **2nd case**, when $d^2/4 - f < 0 \Rightarrow d^2/4 < f$, *or* $b^2 < 2 \cdot c \cdot a$, then $\lambda_{1,2}$ will be displayed by complex number, as the radian is negative and the square root from (-1) yields complex number j.

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{\frac{d^2}{4} - f} \qquad (5.40)$$

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{(-1)\left(f - \frac{d^2}{4}\right)}$$

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{(-1)} \cdot \sqrt{\left(f - \frac{d^2}{4}\right)}$$

$$\lambda_{1,2} = -\frac{d}{2} \pm j \cdot \beta$$

Inserting this solution of the characteristic equation into the substitution function:

$$y = C_1 \cdot e^{\left(-\frac{d}{2} + j \cdot \beta\right) \cdot t} + C_2 \cdot e^{\left(-\frac{d}{2} - j \cdot \beta\right) \cdot t} \qquad (5.41)$$

$$y = e^{-\frac{d}{2} \cdot t} \left(C_1 \cdot e^{+j \cdot \beta \cdot t} + C_2 \cdot e^{-j \cdot \beta \cdot t}\right)$$

According to the law of exponential calculation the sum of the exponents could be decomposed into product of two exponential functions. At the same time we can consider that the exponential functions with imaginary exponent can be transformed into trigonometric functions.

Thus the solution:

$$y = e^{-\frac{d}{2} \cdot t} \left(C_1 \cos \beta t + C_2 \sin \beta t\right) \qquad (5.42)$$

This function represents the general form of the oscillation equation. For special cases we get sinusoidal oscillations. This is the case, if c1 or C2 are identically equal to zero. With d = 0 we get an undamped oscillation, i.e. the amplitude is constant, if d < 0 a damped, with which the amplitude goes to zero, and if d > 0 a swing oscillation.

3. **3rd case**, in the case the radian is equal to zero, and we get two identical solutions:

$$\lambda = \lambda_1 = \lambda_2 = -\frac{d}{2} = -\frac{b}{2a} \tag{5.43}$$

Thus the solution is no longer unique! We have two different functions, which satisfy the differential equation as solution:

$$\begin{aligned} y_1 &= Ce^{\lambda t} \\ y_2 &= C_1 t e^{\lambda t} + C_2 e^{\lambda t} \end{aligned} \tag{5.44}$$

Example of 2nd order differential equation:

1. find out the solution: $y'' - y = 0.$

	$y'' - y = 0$
Differential equation:	$y = e^{\lambda t}$
substitution:	$y' = \lambda e^{\lambda t}$
	$y'' = \lambda^2 e^{\lambda t}$
	$\left(\lambda^2 - 1\right) e^{\lambda t} = 0$
insertion:	$\lambda^2 - 1 = 0$
characteristic equation:	$\lambda_{1,2} = \pm 1$
solution of characteristic equation:	$y = C_1 e^{t} + C_2 e^{-t}$
general solution:	

2. find out the solution: $\ddot{y} + y = 0.$

Differential equation:	$\ddot{y} + y = 0\dot{y}$
substitution:	$y = e^{\lambda t}$
	$\dot{y} = \lambda e^{\lambda t}$
	$\ddot{y} = \lambda^2 e^{\lambda t}$
insertion:	$\left(\lambda^2 + 1\right) e^{\lambda t} = 0$
characteristic equation:	$\lambda^2 + 1 = 0$
solution of characteristic equation:	$\lambda_{1,2} = \pm j$
	$e^{\pm jt} = \cos t \pm j \sin t$
general solution:	$y = C_1 \cos t + C_2 \sin t$

3. find out the solution: $\ddot{y} + 2\dot{y} + y = 0.$

Differential equation:	$\ddot{y} + 2y' + y = 0$
substitution:	$y = e^{\lambda t}$
	$\dot{y} = \lambda e^{\lambda t}$
	$\ddot{y} = \lambda^2 e^{\lambda t}$
insertion:	$\left(\lambda^2 + 2\lambda + 1\right) e^{\lambda t} = 0$
characteristic equation:	$\lambda^2 + 2\lambda + 1 = 0$
solution of characteristic equation:	$\lambda_{1,2} = -1$
general solution:	$y = C_1 t e^{-t} + C_2 e^{-t}$

Remarks:

This solution method can be likewise used for differential equation of higher order (n≥3) with the appropriate substitution of higher order algebraic equations.

Example of 3rd order:

find out the solution: $\dddot{y} + \dot{y} = 0.$

Differential equation: $\dddot{y} + \dot{y} = 0\dot{y}$

substitution:
$$y = e^{\lambda t}$$
$$\dot{y} = \lambda e^{\lambda t}$$
$$\ddot{y} = \lambda^2 e^{\lambda t}$$
$$\dddot{y} = \lambda^3 e^{\lambda t}$$

insertion: $\left(\lambda^3 + \lambda\right) e^{\lambda t} = 0$

characteristic equation: $\lambda^3 + \lambda = \lambda\left(\lambda^2 + 1\right) = 0$

solution of characteristic equation: $\lambda_1 = +j, \qquad \lambda_2 = -j, \qquad \lambda_3 = 0$

general solution: $y = C_1 + C_2 \cos x + C_3 \sin x$

5.2.2.3 Differential equation of type c

$$\frac{d^2y}{dt^2} + \frac{1}{t}\frac{dy}{dt} + y = 0 \qquad (5.45)$$

This differential equation is again to be solved according to the substitution method. The solution with the pertinent derivatives:

$$\begin{aligned} y &= 1 + a_2t^2 + a_3t^3 + a_4t^4 \\ \frac{dy}{dt} &= 2a_2t + 3a_3t^2 + 4a_4t^3... \\ \frac{d^2y}{dt^2} &= 2a_2 + 2\cdot 3a_3t + 4\cdot 3a_4t^2.. \end{aligned}$$

If these equations are inserted into the differential equation and if the equation is arranged according to powers of t:

$$\begin{aligned} &(1 + 2\cdot 2a_2)\cdot t^0 + 3\cdot 3a_3\cdot t^1 + (a_2 + 4\cdot 4a_4)\cdot t^2 + (a_3 + 5\cdot 5a_5)\cdot t^3 \\ &+..... + (a_n + (n+2)^2\cdot a_{n+2})\cdot t^n = 0 \end{aligned} \qquad (5.46)$$

A solution of this equation, which applies to all t-values, is that the factors of the power series members are equal to zero.
In this case:

$$\begin{aligned} a_3 &= a_5 = a_7 = ...a_{2n+1}..... = 0 \\ a_2 &= -\frac{1}{2^2} \\ a_4 &= -\frac{a_2}{4^2} = \frac{1}{2^24^2} \\ a_6 &= -\frac{a_4}{6^2} = -\frac{1}{2^24^26^2} \\ a_{2n} &= -\frac{a_{2n-2}}{(2n)^2} = (-1)^{\frac{2n}{2}}\frac{1}{2^2\cdot 4^2\cdot 6^2\cdot ....\cdot (2n)^2} = (-1)^{\frac{2n}{2}}\frac{1}{\Pi_{k=1}^{n}(2k)^2} \end{aligned}$$

If we set these coefficients into the solution, then we receive the solution of differential equation, which are called zero order Bessel function:

$$y = 1 - \frac{t^2}{2^2} + \frac{t^4}{2^24^2} - \frac{t^6}{2^24^26^2} + ... + (-1)^{\frac{2n}{2}}\frac{t^{2n}}{\Pi_{k=1}^{n}(2k)^2} + ... = I_0(t) \qquad (5.47)$$

5.2.2.4 Tasks for the solution of higher order differential equation

The following differential equations are to be solved:

a) $yy'' = y'^2$

b) $y'' - y' = e^x$

c) $y'' + 4y' + a_0 y = 0 \qquad$ für $a_0 = 3, 4, 5$

5.3 Integral transform

5.3.1 Time- and Frequency domain

Integral transform is a method over a detour to solve differential equation. We distinguish two ranges in the transformations:

· the original or time domain and

· complex variable or frequency domain.

The integration according to the arguments within the original range is transformed into a multiplication in complex variable domain. The difficult integration procedures can be bypassed. The relations between the ranges and their special characteristics are represented in the following scheme (see figure 5.8).

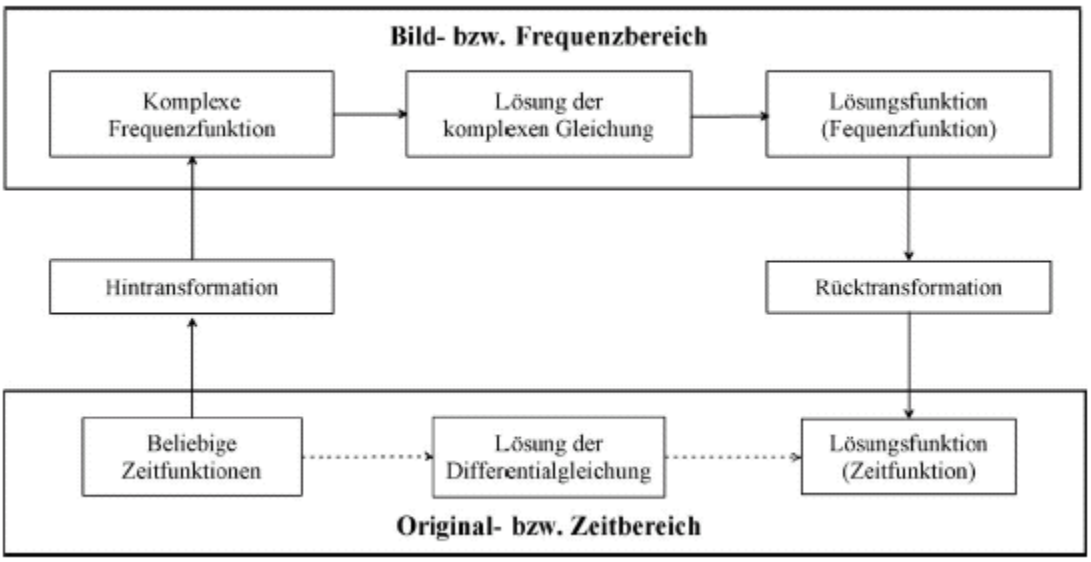


Figure 5.8: Connection between original and complex variable domain

The most well-known transformations are the LAPLACE -, the LAPLACE CARSON -, the FOURIER, LAURENT and the Z-transform. The theories of most these transformations can be gleaned in the multifaceted literature. Therefore here we only deal with the substantial criteria and disadvantages, which are against general application.

The following transformations are represented on the basis of time as argument, since these are most frequent applications of engineers, albeit the transformations are applicable to all arguments, i.e. also to space variables.

The group of the integral transforms can be divided into the continuous and discrete transformations. The continuous integral transforms can be generally written:

$$F(f(t)) = \int_{1t}^{2t} k(t, f(t)) f(t) dt \qquad (5.48)$$

Whereby *k(t, f(t))* is designated as core of the transformation. To simplify matters only an argument (e.g. *t*) is regarded.

As special cases the relations specified in table 5.3.

Table 5.3: Special cases in continuous integral transforms

Transformationskern $k(t, f(t))$	**untere Integrationsgrenze** t_1	**obere Integrationsgrenze** t_2	**Bezeichnung**
e^{-pt}	$0(-\infty)$	∞	LAPLACE-Transformation
pe^{-pt}	$0(-\infty)$	∞	LAPLACE-CARSON-Transformation
$e^{-j\omega t}$	$0(-\infty)$	∞	FOURIER-Transformation

The connection between the three numerated integral transforms can be represented in the following form descriptive. According to definition it will be characterized as complex frequency

$$p = \sigma + j\omega \qquad \text{mit } s = \text{Realteil und } j = \text{Imaginärteil} \qquad (5.49)$$

If the real part of the complex frequency p approaches to zero, the LAPLACE transformation changes into the FOURIER transformation. It means that arbitrary (theoretical) time procedure can be treated by means of the LAPLACE transformation, and only sinusoidal one by means of the FOURIER transformation. The LAPLACE transformation is particularly suitable for application to deadbeat procedures, like e.g. bar signals. Nevertheless the FOURIER transformation has a large advantage as it is simpler in practice. Each periodic or periodization function can be decomposed into a sum of sinusoidal oscillations by the Fourier series analysis. This decomposition of the excitation functions and overlay of the response functions are certainly only permitted in linear systems. The well-known complex computing methods of electro-

technology for sinusoidal alternating current results from the Fourier transformation. In the Fourier transformation the density of such oscillations, the so called spectrum will be analysed and treated by the rule of alternating current theory with only one sinusoidal oscillation, i.e. only one frequency. The **discrete transformations** are in contrast represented by a sum formula:

$$F(f(t)) = \sum_n k(t_n, f(t_n)) \tag{5.50}$$

We can get certain $k(t_n, f(t_n))$ for some special cases (see table 5.4).

Table 5.4: Special cases for discrete transformations

Transformationskern $k(t_n, f(t_n))$	**Summationsgrenzen**	**Bezeichnung**
e^{-pn}	$0 \leq n < \infty$	diskrete LAPLACE-Transformation
$\frac{1}{z^n}$	$-\infty < n < \infty$	LAURENT-Transformation
$\frac{1}{z^n}$	$0 \leq n < \infty$	Z-Transformation

We can also explain the connection between LAPLACE and Z-transform in the following way. Replace in the LAPLACE integral for continuous functions the function *f(t)* by the function value series *f(nT),*

$$F(p) = \int_0^\infty f(t)e^{-pt}dt \tag{5.51}$$

$$p = \sigma + j\omega$$

The integral correspondingly by an infinite sum and e^{-pt} by e^{-pnt}:

$$F(p) = T\sum_{n=0}^{\infty} f(nT)e^{-npT} \tag{5.52}$$

with $e^{pt} = z$

$$F_T(z) = T\sum_{n=0}^{\infty} f(nT)z^{-n} = T \cdot F(z) \tag{5.53}$$

5.3.2 LAPLACE Transformation

Forward transformation

The following symbols are used in LAPLACE transformation:

$$
\begin{aligned}
L\{f(t)\} &= F(p) && \text{LAPLACE-Transformierte der Funktion } f(t)\\
L^{-1}\{F(p)\} &= L^{-1}\{L\{f(t)\}\} && \text{LAPLACE-Rücktransformierte}\\
&= f(t)
\end{aligned}
$$

The transformation of time or original level into the LAPLACE level takes place by means of integral relationship stated above.

$$F(p) = L\{f(t)\} = \int_0^\infty f(t)e^{-pt}dt \tag{5.54}$$

$$p = \sigma + j\omega$$

Examples for the application of transformation to functions:

Example1:

$$f(t) = 0 \tag{5.55}$$

$$F(p) = L\{0\} = \int_0^\infty 0 \cdot e^{-pt}dt = 0$$

Example 2:

$$f(t) = 1 \tag{5.56}$$

$$
\begin{aligned}
F(p) = L\{1\} &= \int_0^\infty 1 \cdot e^{-pt}dt\\
&= -\frac{1}{p} \cdot \left[e^{-pt}\right]_0^\infty = -\frac{1}{p} \cdot (-1)\\
&= \frac{1}{p}
\end{aligned}
$$

Example 3:

$$f(t) = t \tag{5.57}$$

$$\begin{aligned} F(p) = L\{t\} &= \int_0^\infty t \cdot e^{-pt} dt \\ &= -\left[\frac{e^{-pt}}{p^2}(pt+1)\right]_0^\infty \\ &= \frac{1}{p^2} \end{aligned}$$

There are tabular compositions of LAPLACE transforming for further basic functions (see table 5.5, page 139).

5.3.2.1 Important calculation rules

- **Addition Theorem**

$$L\{f_1(t) + f_2(t)\} = L\{f_1(t)\} + L\{f_2(t)\}$$

This addition theorem can exemplarily prove another arithmetic rules, that according to transformation rule LAPLACE transformation is calculable as integral of product of exponential functions:

$$\begin{aligned} L\{f_1(t) + f_2(t)\} &= \int_0^\infty e^{-pt}(f_1(t) + f_2(t))dt \\ &= \int_0^\infty (e^{-pt} f_1(t) + e^{-pt} f_2(t))dt \\ &= \int_0^\infty e^{-pt} f_1(t)dt + \int_0^\infty e^{-pt} f_2(t)dt \end{aligned} \tag{5.58}$$

According to the definition of LAPLACE transformation:

$$L\{f_1(t) + f_2(t)\} = L\{f_1(t)\} + L\{f_2(t)\}$$

General form of the addition theorem

$$L\{\lambda_1 f_1(t) + ... + \lambda_n f_n(t)\} = \lambda_1 F_1(p) + ... + \lambda_n F_n(p) \tag{5.59}$$

- **Similarity theorem**

$$L\{f(at)\} = \frac{1}{a} F\left(\frac{p}{a}\right) \tag{5.60}$$

• **Theorem for damping**

$$L\left\{e^{-at}f(t)\right\} = F(p+a) \tag{5.61}$$

• **Shift theorem**

$$\begin{aligned} &L\{f(t-a)\} = e^{-ap}F(p) && \left\{\begin{array}{l}\text{nach rechts,}\\ \text{positiv in die Zukunft}\end{array}\right.\\ &L\{f(t+a)\} = e^{ap}\left[F(p) - \int_0^a e^{-pt}f(t)dt\right] && \left\{\begin{array}{l}\text{nach links,}\\ \text{negativ in die Vergangenheit}\end{array}\right.\\ &\text{mit} \quad F(p) = L\{f(t)\} && \left\{\begin{array}{l}\text{LAPLACE-Transformation}\\ \text{ohne Verschiebung}\end{array}\right. \end{aligned} \tag{5.62}$$

• **Differentiation**

The deviation rules form the basic application of LAPLACE transformation to differential equations and their solution.

$$\begin{aligned} L\{f'(t)\} &= pF(p) - f(0)\\ L\{f''(t)\} &= p^2F(p) - f(0)p - f'(0)\\ L\{f^n(t)\} &= p^nF(p) - f(0)p^{n-1} - f'(0)p^{n-2} - \cdots\\ &\quad \cdots\cdots - f^{(n-2)}(0)p - f^{(n-1)}(0) \end{aligned} \tag{5.63}$$

• **Integration**

$$L\left\{\int_0^t f(\tau)d\tau\right\} = \frac{1}{p}F(p) \tag{5.64}$$

• **Faltung theorem**

The faltung operation plays a role in transmission system analysis (see section 12.3, page 355 and the following page)

$$L\left\{\int_0^t f_1(t-\tau)f_2(\tau)d\tau\right\} = L\{f_1(t)\}\cdot L\{f_2(\tau)\} = F_1(p)\cdot F_2(p) \tag{5.65}$$

Inverse transformation

For the inverse transformation we use the so called L^{-1}-Transformation.

$$L^{-1}\{F(p)\} = L^{-1}\{L\{f(t)\}\} = f(t) \qquad \text{LAPLACE-Rücktransformierte} \qquad (5.66)$$

In principle the following are possible to be applied:

- **Integral formula**

$$f(t) = \frac{1}{2\pi j}\int_{a-j\infty}^{a+j\infty} L\{f(t)\}\, e^{pt} dp \qquad (5.67)$$

- **Residue formula** (expansion into partial fractions)

$$f(t) = \sum_{n=1}^{\infty} Res_{p=p_n}\left\{L\{f(t)\}\, e^{pt}\right\} \qquad (5.68)$$

p_n is the singular places on the left, complex half planes, and (p - p_n) yields the corresponding pole places.

- **Series development**

$$f(t) = \sum_{n=0}^{\infty} a_n B_n(\sigma_0 t) \qquad \text{mit } B_n(\sigma_o t) = e^{-\sigma_0 t} L_n(2\sigma_0 t) \qquad (5.69)$$

Because of this possibility the residue formula is always applicable and easy to handle for the technical problems.

5.3.2.2 Correspondence table

Since these integrals are relatively complicated and different functions are very often repeated, arithmetic rules and **correspondence tables** are set up, from which the forward transformation and their inverse transformations are easy for reading (see table 5.5).

Table 5.5: Correspondence table

Nr.	$\mathrm{F(p)} = \mathbf{L}\{\mathrm{f(t)}\}$	$\mathrm{f(t)} = \mathbf{L}^{-1}\{\mathrm{F(p)}\}$
1	0	0
2	$\frac{1}{p}$	1
3	$\frac{1}{p^n}$	$\frac{t^{n-1}}{(n-1)!}$
4	$\frac{1}{(p-\alpha)^n}$	$\frac{t^{n-1}}{(n-1)!}e^{\alpha t}$
5	$\frac{1}{(p-\alpha)(p-\beta)}$	$\frac{e^{\beta t}-e^{\alpha t}}{\beta-\alpha}$
6	$\frac{p}{(p-\alpha)(p-\beta)}$	$\beta\frac{e^{\beta t}-\alpha e^{\alpha t}}{\beta-\alpha}$
7	$\frac{\alpha}{p^2+\alpha^2}$	$\sin\alpha t$
8	$\frac{\alpha\cos\beta+p\sin\beta}{p^2+\alpha^2}$	$\sin(\alpha t+\beta)$
9	$\frac{p}{p^2+\alpha^2}$	$\cos\alpha t$
10	$\frac{p\cos\beta-\alpha\sin\beta}{p^2+\alpha^2}$	$\cos(\alpha t+\beta)$
11	$\frac{\alpha}{p^2-\alpha^2}$	$\sinh\alpha t$
12	$\frac{p}{p^2-\alpha^2}$	$\cosh\alpha t$
13	$\frac{p^2+2\alpha^2}{p(p^2+4\alpha^2)}$	$\cos^2\alpha t$
14	$\frac{2\alpha^2}{p(p^2+4\alpha^2)}$	$\sin^2\alpha t$
15	$\frac{p^2-2\alpha^2}{p(p^2-4\alpha^2)}$	$\sinh^2\alpha t$

Table 5.6: Correspondence table - continuation

Nr.	$\mathbf{F(p) = L\{f(t)\}}$	$\mathbf{f(t) = L^{-1}\{F(p)\}}$
16	$\frac{2\alpha^2 p}{p^4+4\alpha^4}$	$\sin\alpha t \sinh\alpha t$
17	$\frac{\alpha(p^2+2\alpha^2)}{p^4+4\alpha^4}$	$\sin\alpha t \cosh\alpha t$
18	$\frac{2\alpha p}{(p^2+\alpha^2)}$	$t\sin\alpha t$
19	$\frac{p^2-\alpha^2}{(p^2+\alpha^2)^2}$	$t\cos\alpha t$
20	$\frac{2\alpha p}{(p^2-\alpha^2)^2}$	$t\sinh\alpha t$
21	$\frac{1}{\sqrt{p}}$	$\frac{1}{\sqrt{\pi t}}$
22	$\frac{1}{p\sqrt{p}}$	$2\frac{1}{\sqrt{\frac{t}{\pi}}}$
23	$\frac{1}{\sqrt{p^2+\alpha^2}}$	$J_0(\alpha t)$ (BESSEL-Funktion der Ordnung 0)
24	$\frac{1}{\sqrt{p^2-\alpha^2}}$	$I_0(\alpha t)$ (modifizierte BESSEL-Funktion der Ordnung 0)
25	$\arctan\frac{\alpha}{p}$	$\frac{\sin(\alpha t)}{t}$
26	$\arctan\frac{2\alpha p}{p^2-\alpha^2+\beta^2}$	$\frac{2}{t}\sin(\alpha t)\cdot\cos(\beta t)$

5.3.3 Solution of differential equations by means of LAPLACE transformation

5.3.3.1 solution method

This solution method consists of three sub steps:

• Application of the LAPLACE transformation to differential equation (or differential equation system) with consideration of initial conditions

• Solution of the resulting algebraic equation (or equation system) with *F (p)* as unknown quantity

• Inverse transformation of *F (p)* and determination of the searched function, i.e. solution function of the differential equation.

5.3.3.2 Examples

1. Find out solution of $\ddot{y}(t)+y(t)=1$ with the initial conditions y (0) =1 and ŷ (0) = 0:

• application of LAPLACE transformation:

Differential equation: $\ddot{y}(t)+y(t)=1$

LAPLACE transformation: $L\left\{\ddot{y}(t)+y(t)\right\}=L\{1\}$

Addition Theorem: $L\left\{\ddot{y}(t)\right\}+L\{y(t)\}=L\{1\}$

Transforming: $p^2F(p)-f(0)p-f'(0)+F(p)=\frac{1}{p}$

Initial conditions: $p^2F(p)-p+F(p)=\frac{1}{p}$

• Solution of the resulting algebraic equation with *F* (*p*):

$$(p^2+1)F(p)=\frac{1}{p}+p$$
$$(p^2+1)F(p)=\frac{p^2+1}{p}$$
$$F(p)=\frac{1}{p}$$

• Inverse transformation und determination of y(*t*) by means of correspondence table (see table 5.5, Page 139, row 2):

$$y(t) = L^{-1}\{F(p)\} = L^{-1}\left\{\frac{1}{p}\right\} = 1$$

2. Solving by means of LAPLACE-transformation

$$\ddot{y} - 3\dot{y} + 2y = 2e^{-t}$$

The initial conditions are $y(0) = 2$ and $\hat{y}(0) = -1$

• application of LAPLACE transformation:

Differential equation:	$\ddot{y} - 3\dot{y} + 2y = 2e^{-t}$
LAPLACE transformation:	$L\left\{\ddot{y} - 3\dot{y} + 2y\right\} = L\left\{2e^{-t}\right\}$
Addition Theorem:	$L\left\{\ddot{y}(t)\right\} + L\left\{-3\dot{y}\right\} + L\{2y(t)\} = L\left\{2e^{-t}\right\}$
Transforming:	$p^2F(p) - f(0)p - f'(0)$
	$-3pF(p) + 3f(0) + 2F(p) = \frac{2}{p+1}$
Initial conditions:	$p^2F(p) - 2p + 1 - 3pF(p) + 6 + 2F(p) = \frac{2}{p+1}$

• Solving algebraic equation according to $F(p)$:

$$\left(p^2 - 3p + 2\right)F(p) = \frac{2}{p+1} + 2p - 7$$

$$F(p) = \frac{2 + (2p-7)(p+1)}{(p^2 - 3p + 2)(p+1)}$$

• Inverse transform and determination of *y(t)*:

The inverse transform is achieved in this case via expansion into partial fractions. The zero positions of the denominator polynomial are searched and expressed as sum product. In the case under consideration the denominator is equal to zero, if:

$$p + 1 = 0 \Longrightarrow p_1 = -1$$

$$p^2 - 3p + 2 = 0 \Longrightarrow p_{2,3} = -\left(\frac{-3}{2}\right) \pm \sqrt{\left(\frac{-3}{2}\right)^2 - 2}$$

$$p_2 = +1, \qquad p_3 = +2$$

The equation for $F(p)$ can be written:

$$F(p) = \frac{2p^2 - 5p - 5}{(p+1)(p-1)(p-2)}$$

The expansion into partial fractions:

$$\frac{2p^2 - 5p - 5}{(p+1)(p-1)(p-2)} = \frac{A}{p+1} + \frac{B}{p-1} + \frac{C}{p-2}$$

A common denominator *(p + 1) (p - 1) (p - 2)* should be used to determine the factors *A, B* and *C*:

$$\frac{2p^2 - 5p - 5}{(p+1)(p-1)(p-2)} = \frac{A(p-1)(p-2) + B(p+1)(p-2) + C(p+1)(p-1)}{(p+1)(p-1)(p-2)}$$

This equation is fulfilled only when apart from the denominators the numerators are also same, i.e.:

$$2p^2 - 5p - 5 = A(p-1)(p-2) + B(p+1)(p-2) + C(p+1)(p-1)$$
$$2p^2 - 5p - 5 = (A + B + C)p^2 + (-3A - B)p + (2A - 2B - C)$$

It must be an identity and valid for all p values. That means the coefficients power series of p are identical in each case. And we get follows:

$$p^2 \qquad 2 = A + B + C$$
$$p^1 \qquad -5 = -3A - B$$
$$p^0 \qquad -5 = 2A - 2B - C$$

This LGS can be solved by the known methods, so the solution is:

$$A = \frac{1}{3}, \qquad B = 4, \qquad C = -\frac{7}{3}$$

Then *F(p)* can be written in the following way:

$$F(p) = \frac{1}{3} \cdot \frac{1}{p+1} + 4 \cdot \frac{1}{p-1} - \frac{7}{3} \cdot \frac{1}{p-2}$$

The inverse transform is to be read from the correspondence table (see table 5.5, page 139, line 4)
It is:

$$L^{-1}\left\{\frac{1}{(p-a)^n}\right\} = \frac{t^{n-1}}{(n-1)!}e^{at} \qquad \text{für } n = 1 \text{ gilt}$$
$$L^{-1}\left\{\frac{1}{(p-a)^1}\right\} = e^{at}$$
$$\text{für } a_1 = -1, a_2 = 1, a_3 = 2 \text{ gilt}$$
$$L^{-1}\{F(p)\} = y(t) = \frac{1}{3}\cdot e^{a_1 t} + 4\cdot e^{a_2 t} - \frac{7}{3}\cdot e^{a_3 t}$$
$$y(t) = \frac{1}{3}e^{-t} + 4e^{t} - \frac{7}{3}e^{2t}$$

Remark:
The same procedure can also be used for the solution of linear DGL systems with constant coefficients.

5.3.3.3 Example of DGL system

Find out the solution of system:

$$\ddot{x} = -\dot{y}$$
$$\ddot{y} = \dot{x}$$

with initial condition $x(0) = 0, \dot{x}(0) = 1, y(0) = 0, \dot{y}(0) = 0.$

• application of LAPLACE transformation:

With F (p) = L { x (t)} and G(p) = L { y (t)} the application of LAPLACE transformation to the system yields (under consideration of the initial conditions)

$$p^2F(p) - 1 = -pG(p)$$
$$p^2G(p) = pF(p)$$

• Solving linear equation according to *F*(*p*), G(p):
According to the known rules or simple transformation, e.g. from the 2nd. equation:

$$G(p) = \frac{1}{p}F(p)$$
$$\text{Einsetzen in obere Gleichung} : \quad p^2F(p) - 1 = -F(p)$$
$$\text{Auflösen nach } F(p) : \quad F(p) = \frac{1}{p^2+1}$$
$$G(p) = \frac{1}{p}F(p) = \frac{1}{p(p^2+1)}$$

• Inverse transform and determination of x (t), y (t) by means of correspondence table (see table 5.5, page 139, line 3 and 7)

$$x(t) = L^{-1}\{F(p)\} = L^{-1}\left\{\frac{1}{p^2+1}\right\} = \sin t$$

$$y(t) = L^{-1}\{G(p)\} = L^{-1}\left\{\frac{1}{p(p^2+1)}\right\} = L^{-1}\left\{\frac{1}{p} - \frac{p}{(p^2+1)}\right\} = 1 - \cos t$$

5.3.3.4 Tasks for the application of LAPLACE transformation

1. Solve the following differential equation by means of LAPLACE transformation

a) $y'(x) + y = 0$

b) $y''(t) - 3y'(t) + 2y(t) = 4$ mit $y(0) = 1$, $y'(0) = 0$

c) $y''(t) + 16y(t) = 32t$ mit $y(0) = 3$, $y'(0) = -2$

d) $y''(t) + 4y'(t) + 4y(t) = 6e^{-2t}$ mit $y(0) = 3$, $y'(0) = 8$

e) $y'''(t) + y'(t) = t + 1$ mit $y(0) =$, $y'(0) =$, $y''(0) =$

2. Solve the following equation system by means of LAPLACE transformation

a)
$$y'(t) + x(t) = 0$$
$$x'(t) + y(t) = 1$$
mit $x(0) = 0$, $y(0) = 0$

b)
$$2x(t) - y(t) - y'(t) = 4(1 - e^{-t})$$
$$2x'(t) + y(t) = 2(1 + 3e^{-2t})$$
mit $x(0) = 0$, $y(0) = 0$

3. Differential equation applies to the hydraulic scheme (see figure 5.9) with associated block diagram:

$$h_{Fl} = R \cdot C\frac{dz_R}{dt} + z_R$$

It is assumed a linearized relationship and a homogeneous, isotropic aquifer with the following parameters $k = 5 \cdot 10^{-4}\frac{m}{s}; n_0 = 0,2; z_{Rmittel} = 20m; l = 50m$.
Compute the change of the water level by means of LAPLACE transformation, if the river surface changes as a first approximation as follows:

c) Erratic ($h_{Fl} = h_{Fl} \cdot l(t)$) and
d) sinusoidal ($h_{Fl} = h_{Flm} \sin(\omega \cdot t) + h_{Fl0}$, with $\omega = 2\pi\Gamma$ and $\Gamma = 7$ days)

4. The following differential equation applies to the concentration C in sorption of pollutants at the soil matrix:

$$T_1 C + C = K$$

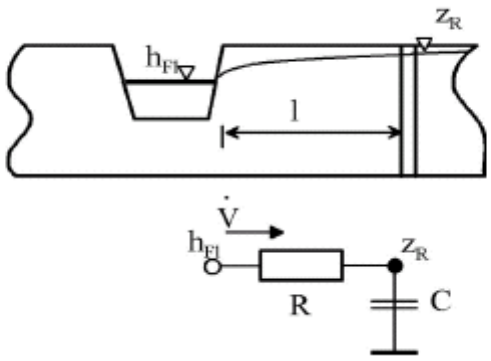


Figure 5.9: Schematic representation of the groundwater level

T_1 is time constant and K is a constant. T_1 = 1d^{-1}, K = 100. The concentration should C(0) = 0 at time t = 0.

a) Solve the differential equation by means of LAPLACE transformation and compute the concentration change for the time $t = 1d$
b) Outline the time process of concentration change.

5. The padding to the remainder holes of the former brown coal open pit caused by the rise of groundwater under natural conditions will last too long time. Therefore external supply is introduced to the filling procedure ($h_{t=0}$ = 0) for acceleration. (see figure 5.10) The corresponding differential equation:

$$A\frac{dh}{dt} = \dot{V}_{Zustr}$$

Transfer this differential equation by means of LAPLACE transformation in the image plane and solve this equation.

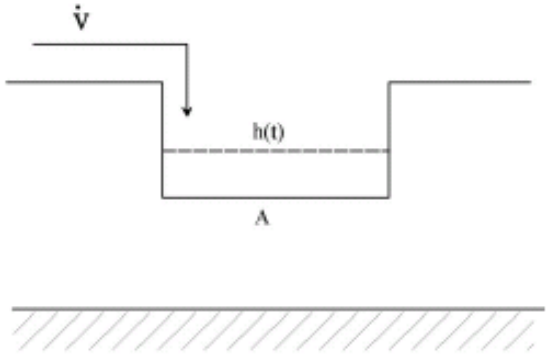


figure 5.10: filling procedure of a remainder hole

6. Solve the following differential equation by means of LAPLACE transformation

$$\frac{dh}{dt} + k \cdot h = g \quad \text{mit} \qquad h_{t=0} = 0$$

$g = 0,015m \cdot s^{-1}$ und $k = 0,01s^{-1}$

5.4 Methods for Numerical Integration

5.4.1 Integration

The numerical integration always yields the result of a certain integral between an upper and a lower limit. In a variable is inserted at the upper limit of the integral, the certain integral changes into a function, which is determined by the lower limit and the variable at upper border. The integral of a function, which can be also displayed as the area between the upper and lower limit and the abscissa, can be approximated by a simplified area computation. The numerical integration procedures differ with each other in the method of area computation. In most procedures it is assumed that total area between the upper and lower limit is divided into individual subarea and the summation of these subareas yields the integral. The accuracy strongly depends on the method of subarea creation and the quantization width of the abscissa. The approximation by summation of subareas will be worst with rectangle method and will be best with Predictor Corrector procedure or with higher order RUNGE KUTTA procedure with the same quantization increment. The advantage of the trivial procedures exists in the simple, fast and stable computation of the subareas also in complicated, e.g. discontinuous function.

5.4.1.1 Rectangle rule

The rectangle rule as the simplest method assumes the creation of rectangles as subarea (see figure 5.11). The area of the rectangles results from the multiplication of the function value (y_n) on the left supporting place (x_n) with the quantization increment $\Delta x = | x_n - x_{n+1}|$. These rectangles yield too small values with convex function, too large values with concave function. A substantial advantage of the rectangle method is no equidistant quantization necessary for the abscissa:

$$F_{links} = \int_a^b y(x)dx \approx \sum_{n=0}^{m} (|x_n - x_{n-1}|)\, y_n \qquad \text{(für } m \text{ Teilintervalle)} \qquad (5.70)$$

We can use function value on the right side instead of on the left.

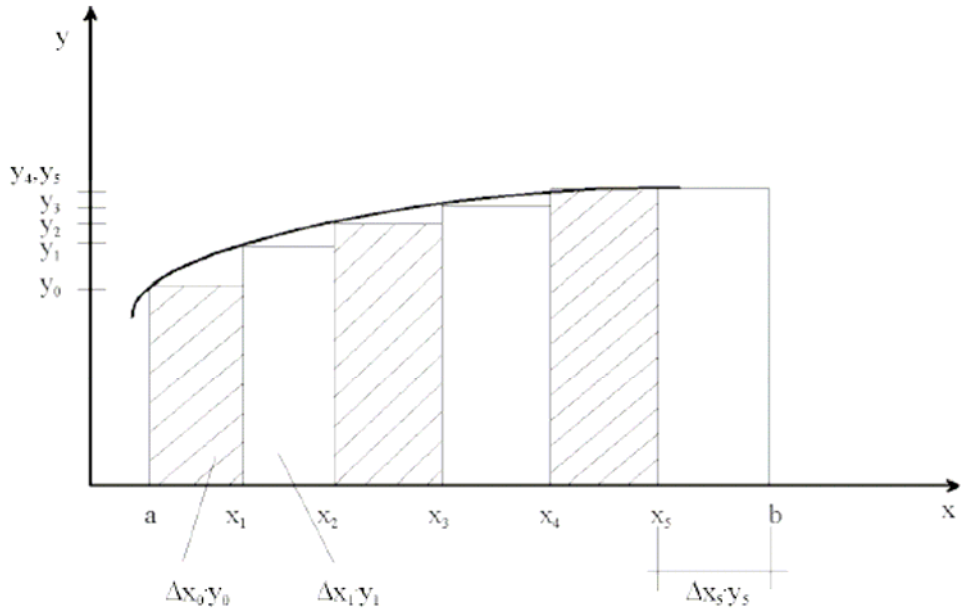


Figure 5.11: creation from rectangles to the numerical integration

In this case the rectangles yield too large values with convex function, and too small with concave function. The computation of the areas:

$$F_{rechts} = \int_a^b y(x)dx \approx \sum_{n=0}^{m} (|x_{n+1} - x_n|)\, y_{n+1} \qquad (5.71)$$

The correct value of the integral must lie between F_{links} and F_{rechts}.

Examples for application of rectangle rule:

1. We calculate the following integral according to the table by using rectangle rule (left and right), then compare the results with the analytical value

$$\int_1^2 \frac{1}{x} dx = \ln 2 = 0,693$$

x	1,0	1,2	1,4	1,6	1,8	2,0
$\frac{1}{x}$	1,000	0,833	0,714	0,625	0,556	0,500

The increment is regarded as constant, h = Δx = 0.2.

$$F_{links} = 0,2\,(1,000 + 0,833 + 0,714 + 0,625 + 0,556) = 0,746$$
$$F_{rechts} = 0,2\,(0,833 + 0,714 + 0,625 + 0,556 + 0,500) = 0,646$$

f (x) = 1/x is a concave function, then $F_{links} > F_{anal} > F_{rechts}$. The average value of the two results is:

$$F_{mittel} = \frac{0,746 + 0,646}{2} = 0,696$$
$$F_{anal} = 0,693$$

This value approaches to the actual analytical value.

Remark:

The increment plays an important role for exact determination of the integral. The smaller it is, the more approaches the numerical value to the analytical, i.e. the numerical value converges. This is not only valid for rectangle rule, but for all numerical methods.

2. We calculate the integral with an increment h = 0.1 in table and compare the results with example 1.

x	$\frac{1}{x}$	F_{links}	F_{rechts}
1,0	1,000	1,000	
1,1	0,909	0,909	0,909
1,2	0,833	0,833	0,833
1,3	0,769	0,769	0,769
1,4	0,714	0,714	0,714
1,5	0,667	0,667	0,667
1,6	0,625	0,625	0,625
1,7	0,588	0,588	0,588
1,8	0,556	0,556	0,556
1,9	0,526	0,526	0,526
2,0	0,500		0,500
	Teilsumme	7,187	6,669

$$F_{links} = 0,1 \cdot (7,187) = 0,719$$
$$F_{rechts} = 0,1 \cdot (6,669) = 0,667$$
$$F_{mittel} = 0,694$$

It is clear to notice that all the three values, which are calculated with increment 0.1, lie nearer to the analytical value F_{anal} = 0.693, than in example 1 with an increment of 0.2.

5.4.1.2 Trapezoidal rule

The approximation by polynomials plays a role in a multitude of procedures. The basic idea is that, if p (x) is an approximation for y (x), $\int_a^b p(x)\,dx \approx \int_a^b y(x)\,dx$.

Different situations are dependent on the selected approximation.

With the trapezoidal rule the function between the supporting places x_n and x_{n+1} is linear interpolated (see figure 5.12). Thus the wanted area is divided into trapezoid areas, which are calculated geometrically:

$$I = h\frac{a+b}{2} \qquad \text{oder} \qquad I = |x_n - x_{n-1}|\,\frac{y_n + y_{n-1}}{2} \tag{5.72}$$

By summation of the subareas:

$$F = \int_a^b y\,(x)\,dx \approx \sum_{n=0}^{m} (|x_n - x_{n+1}|)\left(\frac{y_n + y_{n+1}}{2}\right) \qquad (\text{für } m \text{ Teilintervalle}) \tag{5.73}$$

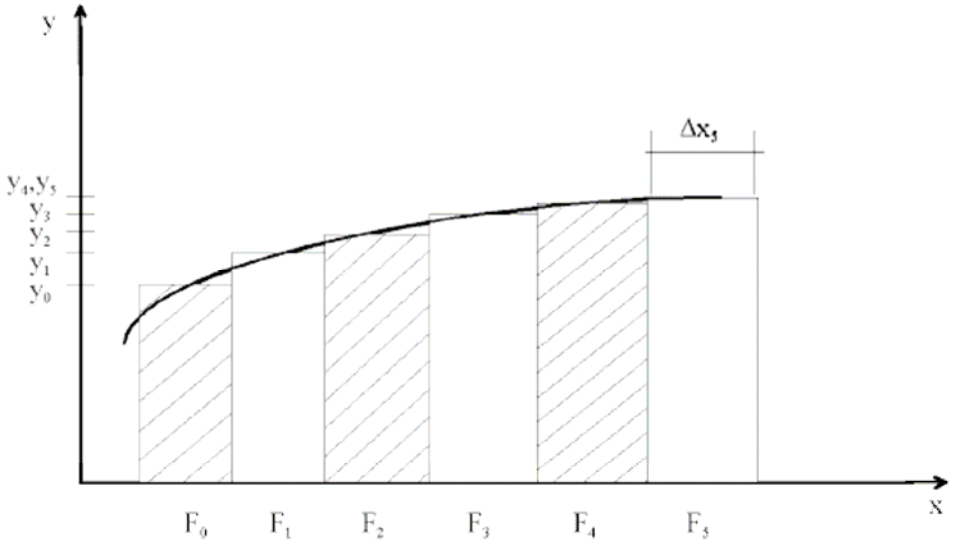


figure 5.12: Numerical integration by means of trapezoidal rule

In the case of equidistant division the computation of Δx can be simplified:

$$F = \int_a^b y(x)\,dx \approx \sum_{n=0}^{m} F_n = \Delta x \cdot \sum_{n=0}^{m-1} (y_{n+1}) + \frac{\Delta x}{2}(y_o + y_m) \qquad (5.74)$$

In the trapezoidal rule we have another simple possibility irregular increment, i.e. not equidistant quantization.

5.4.1.3 Simpson's Rule

The Simpson's rule is:

$$F = \int_{x_0}^{x_{2k}} f(x)dx = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + ... + 2y_{2k-2} + 4y_{2k-1} + y_{2k}) \qquad (5.75)$$

It is likewise a compound formula as parabolic arcs are used instead of y (x).

Pay attention:
· the supporting places must be equidistant (constant increment h).
· the number of supporting places x_n must be odd ($n = 0 \ldots 2k$):

5.4.1.4 Newton's Formula

In this method the Newton's interpolation function (also see section 3.1.3, page 70) will be integrated with following results

Gleichung	Stützstellenanz.	Interpolationsart
$\int_{x_0}^{x_1} p(x)\,dx = \frac{h}{2}(y_0 + y_1)$	2	linear
$\int_{x_0}^{x_2} p(x)\,dx = \frac{h}{3}(y_0 + 4y_1 + y_2)$	3	quadratisch
$\int_{x_0}^{x_3} p(x)\,dx = \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3)$	4	kubisch

5.4.1.5 examples for application of numerical integration

We use trapezoidal rule and Simpson's rule in order to determine the integral $\int_0^{\pi/2} sin\, x \cdot dx$ from the following table. And compare the results with the analytical value $I_{anal} = 1$.

x	0	$\frac{\pi}{12}$	$\frac{2\pi}{12}$	$\frac{3\pi}{12}$	$\frac{4\pi}{12}$	$\frac{5\pi}{12}$	$\frac{\pi}{2}$
sin x	0	0,259	0,500	0,707	0,866	0,966	1,000

Trapezoidal rule

$$I_{tr} = \frac{\pi}{12}(0 + 0,259 + 0,5 + 0,707 + 0,866 + 0,966 + 0,5) = 0,994$$

Simpson's rule

$$I_s = \frac{\pi}{3 \cdot 12}(0 + 4 \cdot 0,259 + 2 \cdot 0,5 + 4 \cdot 0,707 + 2 \cdot 0,866 + 4 \cdot 0,966 + 1) = 1,000$$

Obviously the adjustment of quadratic polynomials yields one up to three decimal places exact result.

Newton's interpolation function

a) linear

$$I_{Nl} = \frac{\pi}{2 \cdot 12}\begin{pmatrix} (0 + 0,259) + \\ (0,259 + 0,500) + \\ (0,500 + 0,707) + \\ (0,707 + 0,866) + \\ (0,866 + 0,966) + \\ (0,966 + 1,000) \end{pmatrix}$$

$$= 0,994$$

b) quadratic

$$I_{Nq} = \frac{\pi}{3 \cdot 12} \left(\begin{array}{c} (0 + 4 * 0,259 + 0,500) + \\ (0,500 + 4 * 0,707 + 0,866) + \\ (0,866 + 4 * 0,966 + 1,000) \end{array} \right)$$
$$= 1,00003$$

c) cubic

$$I_{Nk} = \frac{3 \cdot \pi}{8} \left(\begin{array}{c} (0 + 3 * 0,259 + 3 * 0,500 + 0,707) + \\ (0,707 + 3 * 0,866 + 3 * 0,966 + 1,000) \end{array} \right)$$
$$= 1,00006$$

Apparently the adjustment of quadratic polynomials yields one up to three decimal places exact result.

Remark:

The accuracy of numerical methods must be always relating to the significant number of computed and represented places. If we solve e.g. the same problem with seven digitals of significant number, then the Simpson's rule yields $I = 1.000003$, which does not match the analytical value of $I_{anal} = 1$.

5.4.1.6 Tasks of application to numerical integration

1. Compute the integral $I = \int_0^1 \frac{dx}{1+x}$ by using trapezoidal rule with increment $h = 0.1$

2. Calculate the following integrals. Use at least two numerical methods and two different increments then compare the results:

a) $\int_{-1}^{1} e^{-x^2 dx}$ und

b) $\int_1^2 \frac{e^x}{x} dx$

3. Calculate the integral $\int_1^{1,3} \sqrt{x} \cdot dx$ by means of three Newton's formulae and compare the results.

4. Calculate the integral $I = \int_1^{10} \frac{1}{x} dx$ by approximation.
Choose h=1.
Apply Simpson's rule for the interval [1, 9] and trapezoidal rule for the interval [9, 10].

5. A measurement series of Al_2O_3 specific heat C are listed in the table as a function of the temperature T.

Determine the amount of heat $Q = \int_{-200}^{1000} c(T) \cdot dT$, which must be supplied to a gram Al_2O_3, in order to warm it up from -200°C to 1000°C .

The integration is to be accomplished numerically according to

a) trapezoidal rule
b) Simpson's rule with an increment h = 200°C.

$\mathbf{T}\ [^{o}C]$	$\mathbf{c}\ [c/(g \cdot K)]$
-260	0
-200	$0,04$
-100	$0,012$
0	$0,18$
100	$0,22$
200	$0,24$
300	$0,25$
400	$0,26$
600	$0,27$
800	$0,275$
1000	$0,28$

6. In a pumping test the groundwater level were measured (see figure 5.13). Calculate the water deficit (volume) of the sinking funnel, if the aquifer is of following characteristic values.
$h_n = 16m$, $M = 10m$, $k = 0.001m \cdot s^{-1}$, $S_0 = 0.0001$, $n_0 = 0.20$

Apply methods of numerical integration.

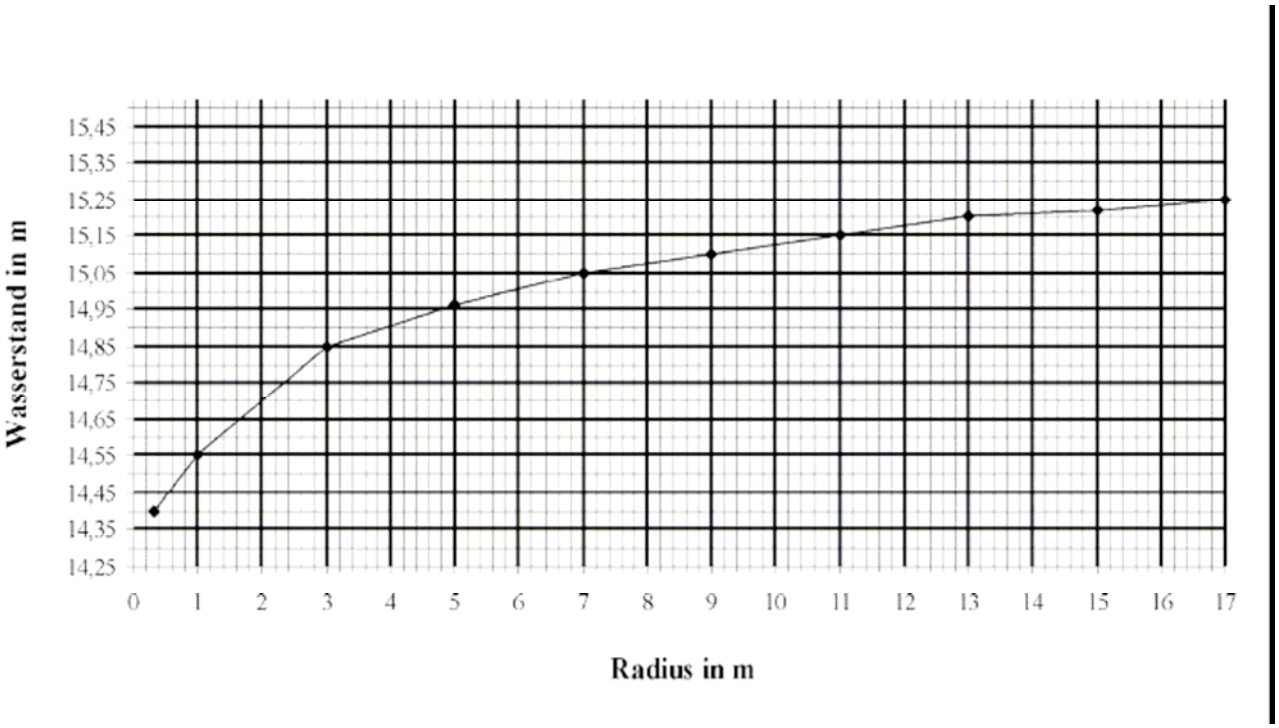


Figure 5.13: groundwater level as a function of the radius

5.4.2 Solution of Differential equations

While the solutions of definite integrals are in the foreground in the former sections, the Euler's method, RUNGE KUTTA method and Predictor Corrector method are showing how to solve ordinary differential equations. In contrast to analytical methods numerical method always assumes boundary conditions, i.e. the initial- and boundary conditions. Particularly in 1st. order differential equation the initial values are supposed, which leads to the concept of **initial value task**.

In the 1st. order differential equation

$$\frac{dy}{dx} = f(x, y) \tag{5.76}$$

with the initial condition beginning point $x = a$ and the function value $y_{(x=a)} = y_a$ yield the integration in range $x = a$ to $x = b$:

$$\int_a^b \frac{dy}{dx} dx = \int_a^b f(x, y) dx \tag{5.77}$$

$$[y]_a^b = \int_a^b f(x, y) dx$$

$$y_b = y_a + \int_a^b f(x, y) dx$$

Thus we obtain the wanted function value y_b in the place $x = b$ from the function value at the beginning point plus the definite integral of function y (see figure 5.14). The problem now is the function y is unknown. For this reason approximation solutions must be again used for the integral as described in the former section. The following methods differ from the application of approximation methods.

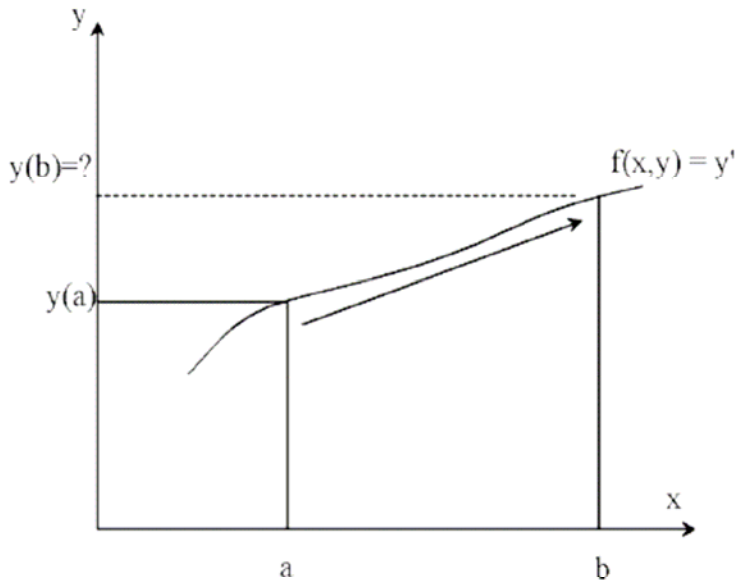


Figure 5.14: Computation of the function value y(b) from the initial value y(a)

These methods can be improved if this approximation is only applied in sections and then iteratively expanded to the whole integration interval (see figure 5.15).

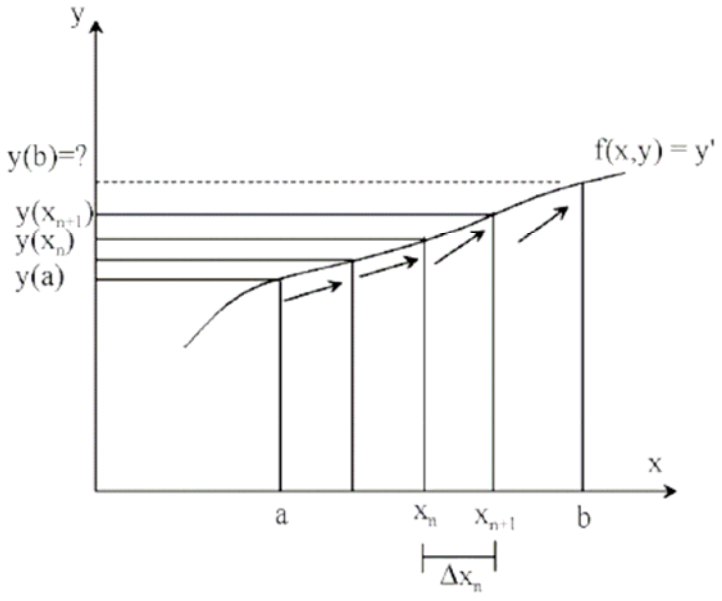


Figure 5.15: iterative solution of the differential equation

$$y_1 \approx y_a + \int_a^{x_n} f(x,y)dx \tag{5.78}$$

$$y_{n+1} \approx y_n + \int_{x_n}^{x_{n+1}} f(x,y)dx \tag{5.79}$$

$$y_b \approx y_m + \int_{x_m}^{b} f(x,y)dx \tag{5.80}$$

We recognize that the writing ways of integration limits could be synonymous:
lower limit: $x = a$ or $x = x_n$
upper limit: $x = b$ or $x = x_{n+1}$
The subscript is usually used for the intermediate intervals and the advantage is that it can be easily converted into programming language.

5.4.2.1 EULER method

The Euler method is the simplest method and actually the integral is approximately determined by means of the rectangle formula. The greater the distance between a and b, i.e. the increment h or Δx, the worse the approximation is.

$$h = b - a$$
$$\Delta x_n = x_{n+1} - x_n$$

Thus the solution with following shape:

$$y_b = y_a + \int_a^b f(x,y)dx \approx y_a + f(a,\, y_a) \cdot (b-a) \tag{5.81}$$

$$\approx y_a + f(a,\, y_a) \cdot h \tag{5.82}$$

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(x,y)dx \approx y_n + f(x_n,\, y_n) \cdot (x_{n+1} - x_n) \tag{5.83}$$

$$\approx y_n + f(x_n,\, y_n) \cdot \Delta x_n \tag{5.84}$$

In this method it is possible to work with different increments, i.e. with a not equidistant division. It is also suitable for an automatic increment control, because the error, which results from the approximation, is dependent on the slope of the function y and on the increment (b - a) = Δx. Inserting it into equation above, we get:

$$y_b = y_a + \int_a^b f(x,y)dx \approx y_a + \left|\frac{dy}{dx}\right|_{x=a} \cdot (b-a) \qquad (5.85)$$

Example for application of the Euler's method (also see figure 5.16):

An approximation solution for the differential equation $y' = xy^{1/3}$ with $y(1) = 1$ is looked for.

The EULER formula can be also written in the form:

$$y_{n+1} = y_n + h \cdot y'_n$$
$$y_{n+1} = y_n + (x_{n+1} - x_n) \cdot x_{n+1} \cdot y_n^{1/3}$$

The stop error $O(h^2)$, which is produced in the interval x_n to x_{n+1}, is rather large in Euler method (i.e. proportional to h^2), so that for a high accuracy very small increments h are necessary. E.g. for $h = 0.01$:

$$y_1 \approx 1 + (0,01)\,1 = 1,0100$$
$$y_2 \approx 1,0100 + (0,01)\,(1,01)\,(1,0033) \approx 1,0201$$
$$y_3 \approx 1,0201 + (0,01)\,(1,02)\,(1,0067) \approx 1,0304$$

The stop error in each interval is about 0.00007. The fourth decimal place should be considered with caution. If we want a higher accuracy, a smaller increment h is necessary. The analytical values are

$$y_1 = 1,01007$$
$$y_2 = 1,02027$$
$$y_3 = 1,03060$$

E.g. the fourth decimal place was actually inaccurate.

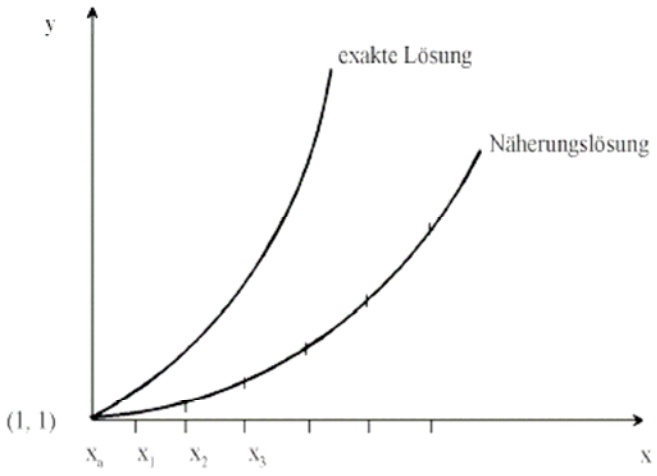


Figure 5.16: result development with the Euler method

5.4.2.2 RUNGE KUTTA method

The RUNGE KUTTA method assumes the same approach, the approximation of the integral by area calculation, as Euler method. The difference lies in the degree of approximation function for area calculation, which is linear in Euler method. Here with the RUNGE KUTTA method a higher order polynomial is used according to TAYLOR series expansion.

$$y_b = y_a + y_a^{'} \cdot h + y_a^{''} \cdot \frac{h^2}{2!} + y_a^{'''} \cdot \frac{h^3}{3!} + y_a^{''''} \cdot \frac{h^4}{4!} + \dots \tag{5.86}$$

with: $h = |b - a|$

Depending upon degrees of the considered derivative in the TAYLOR series we distinguish RUNGE KUTTA method in different n-th orders.

In the following subscript way of writing is used, as the entire integration interval (a to b) is mostly decomposed into subintervals and additionally this way will be converted in programming technique in practice.

$$y_{n+1} = y_n + k_n \tag{5.87}$$

The RUNGE KUTTA methods differ in the way of k_n determination. In this classification Euler method can be arranged:

$$k_n = \Delta x_n \cdot f\left(x_n, y_n\right)$$
$$\Delta x_n = |x_{n+1} - x_n|$$

The simplest procedure, which differs from Euler method in respect of accuracy, is 2nd order RUNGE KUTTA method:

$$y_{n+1} = y_n + k_2 \tag{5.88}$$

with:

$$k_1 = \Delta x_n \cdot f\left(x_n, y_n\right) \tag{5.89}$$

$$k_2 = \Delta x_n \cdot f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right) \tag{5.90}$$

$$\Delta x_n = |x_{n+1} - x_n|$$

The error this method grows proportionally with h power 3 ($0(h^3)$) and is better one power than Euler method ($0(h^2)$).
The 4th order RUNGE KUTTA method is frequently used, which is a good compromise between accuracy and numerical expenditure. For the general form:

$$y_b = y_a + k \tag{5.91}$$

We write:

$$
\begin{aligned}
k &= \frac{1}{6} \cdot (k_1 + 2\,k_2 + 2\,k_3 + k_4) \\
k_1 &= h \cdot f\,(a,\, y_a) \\
k_2 &= h \cdot f\left(a + \frac{h}{2},\, y_a + \frac{k_1}{2}\right) \\
k_3 &= h \cdot f\left(a + \frac{h}{2},\, y_a + \frac{k_2}{2}\right) \\
k_4 &= h \cdot f\,(a + h,\, y_a + k_3)
\end{aligned}
\qquad (5.92)
$$

with: $h = |b - a|$

The error of this procedure is 5th.order ($0(h^5)$). Here also for improvement of the accuracy we can divide the total interval of *a* to *b* into subintervals x_n with y_n and iteratively solve y_b. Since we cannot change the increment within the subintervals, it is possible to control increment as a function of gradients:

$$
\begin{aligned}
y_2 &= y_a + \frac{1}{6}\,(k_{1,1} + 2k_{2,1} + 2k_{3,1} + k_{4,1}) \\
&\vdots \\
y_{n+1} &= y_n + \frac{1}{6}\,(k_{1,n} + 2k_{2,n} + 2k_{3,n} + k_{4,n}) \\
&\vdots \\
y_b &= y_{b-\Delta x_n} + \frac{1}{6}\,(k_{1,b-\Delta x_n} + 2k_{2,b-\Delta x_n} + 2k_{3,b-\Delta x_n} + k_{4,b-\Delta x_n})
\end{aligned}
\qquad (5.93)
$$

When $n = 1$, it yields:

$$\Delta x_n = |x_{n+1} - x_n|$$

$$
\begin{aligned}
k_{1,n} &= \Delta x_n \cdot f\,(x_n,\, y_n) \\
k_{2,n} &= \Delta x_n \cdot f\left(x_n + \frac{\Delta x_n}{2}, y_n + \frac{k_1}{2}\right) \\
k_{3,n} &= \Delta x_n \cdot f\left(x_n + \frac{\Delta x_n}{2}, y_n + \frac{k_2}{2}\right) \\
k_{4,n} &= \Delta x_n \cdot f\,(x_n + \Delta x_n, y_n + k_3)
\end{aligned}
\qquad (5.94)
$$

Example for application of the RUNGE-KUTTA method:

An approximation solution for the differential equation $y' = xy^{1/3}$ with $y(1) = 1$ is looked for.

With $x0 = 1$ und $h = 0.1$ we get 4th order according to above RUNGE-KUTTA formula (see equation 5.94):

$$k_1 = 0,1 \cdot f(1,1) = 0,1$$
$$k_2 = 0,1 \cdot f(1,05; 1,05) = 0,10672$$
$$k_3 = 0,1 \cdot f(1,1; 1,105336) = 0,10684$$
$$k_4 = 0,1 \cdot f(1,1; 1,10684) = 0,11378$$

We calculate:

$$y_1 = 1 + \frac{1}{6}(0,1 + 0,21344 + 0,21368 + 0,11378) = 1,10682$$

The analytical value is y = 1.10326. The correspondent value with EULER method is y = 1.10000, i.e. the RUNGE KUTTA method yields better result. However the increment must be likewise smaller selected if a higher accuracy is demanded.

5.4.2.3 Predictor-Corrector method

The Predictor Corrector method is a two-step procedure. In the first step an auxiliary value y^*_b is computed and then y_b. Thus an increased numerical expenditure develops, but the accuracy rises substantially compared to one-step method. Besides RUNGE KUTTA method Predictor Corrector method represents the most substantial integration procedure. Predictor step in the simplest form, like in Euler method, a rectangle formula for the computation of the integral is used:

$$y^*_b = y_a + \int_a^b f(x,y)dx \approx y_a + f(a, y_a) \cdot (b-a) \qquad (5.95)$$

We can also write:

$$y^*_b = y_a + \int_a^b f(x,y)dx \approx y_a + \left.\frac{dy}{dx}\right|_{x=a} \cdot (b-a) = y_a + y'_a(b-a) \qquad (5.96)$$

The difference is that in Predictor step wanted value y_b will not be computed, but as the first approximation of this value y^*_b is regarded. As the second step, Corrector step, the integral will be calculated by the trapezoid formula, while the value y^*_b is used as upper value in the trapezoid formula:

$$y_b = y_a + \frac{(b-a)}{2} \cdot (f(a, y_a) + f(b, y^*_b)) \qquad (5.97)$$

Similar to the Predictor step here the basis of output differential equation can be formulated here:

$$y_b = y_a + \frac{(b-a)}{2} \cdot \left(y_a^{'} + y_b^{*'} \right) \tag{5.98}$$

Also this procedure can be expanded to n subintervals of the range a to b and computed iteratively. Then the Predictor- and the Corrector- step for the n+1 th interval:

$$\begin{aligned} y_{n+1}^{*} &= y_n + \Delta x \cdot (f(x_n, y_n)) = y_n + \Delta x \cdot y_n^{'} \\ y_{n+1} &= y_n + \frac{\Delta x}{2} \cdot \left(f(x_n, y_n) + f(x_{n+1}, y_{n+1}^{*}) \right) = y_n + \frac{\Delta x}{2} \cdot \left(y_n^{'} + y_{n+1}^{*'} \right) \end{aligned} \tag{5.99}$$

with: $\Delta x = |x_{n+1} - x_n|$

A series of procedures were developed. The above procedure possesses the disadvantage that there is a relative large residue, i.e. a residual value error, which grows proportionally with Δx^2 ($O(\Delta x^2)$). The advantage lies in a relatively simple increment control, only the value f (x_{n+1}; y^*_{n+1}) or the derivative of y_{n+1} is to be computed. A very widespread Predictor Corrector method is the Adam BASHFORTH MOULTON scheme. This method is very stable. In contrast to the simple Predictor Corrector method several supporting places of integration steps are needed here. Thus the approximation area will not be made a rectangle any longer, but polyline is used for boundary. The frequently used 3rd. order ADAM BASHFORTH:

For Predictor step:

$$\begin{aligned} y_{n+1}^{*'} &= y_n + \frac{\Delta x}{12} \left(5\, f(x_{n-2}, y_{n-2}) - 16 f(x_{n-1}, y_{n-1}) + 23\, f(x_n, y_n) \right) \\ &= y_n + \frac{\Delta x}{12} \left(5 y_{n-2}^{'} - 16 y_{n-1}^{'} + 23 y_n^{'} \right) \end{aligned} \tag{5.100}$$

Then for Corrector step:

$$\begin{aligned} y_{n+1} &= y_n + \frac{\Delta x}{12} \left(-f(x_{n-1}, y_{n-1}) + 8 f(x_n, y_n) + 5 f(x_{n+1}, y_{n+1}^{*'}) \right) \\ &= y_n + \frac{\Delta x}{12} \left(-y_{n-1}^{'} + 8\, y_n^{'} + 5 y_{n+1}^{*'} \right) \end{aligned} \tag{5.101}$$

This method yields a residue, grows proportionally with 4th power of Δx ($\sim \Delta x^4$), i.e. $O(\Delta x^4)$. The disadvantage is that the intervals n-1 and n-2 must be calculated again if changing increment for interval n. So it is necessary to calculate the intervals n, n-1 and n-2 with the same increment Δx. This can lead to an increased numerical expenditure when strong gradient oscillation.

Example for application of the Predictor-Corrector method:

An approximation solution for the differential equation $y' = xy^{1/3}$ with $y(1) = 1$ is looked for. The accuracy $\varepsilon \leq 10^{-5}$

For each forward step the simple Euler formula is used as a Predictor. It presupposes a first estimation of y_{n+1}. Here $x_0 = 1$ and h=0.05

$$y(1,05) \approx 1 + 0,05 \cdot 1 = 1,05$$

The differential equation:

$$y'(1,05) = 1,05 \cdot 1,05^{\frac{1}{3}} = 1,0661$$

The Euler formula will be modified for Corrector (according to trapezoidal rule):

$$y_{n+1} = y_n + \frac{1}{2}h\left(y'_n + y'_{n+1}\right)$$

It yields:

$$y(1,05) \approx 1 + 0,025\,(1 + 1,0661) = 1,05165$$

With this new value of the differential equation y' (1.05) will be corrected to 1.0678; afterwards the Corrector is used again and yields the result:

$$y(1,05) \approx 1 + 0,025\,(1 + 1,0678) = 1,0517$$

Further calculations confirm these four decimal places, so that the desired accuracy is reached. It is noticed that the same accuracy can be achieved with increment h = 0.01 in simple Euler formula.

Generally we iterate until it converges if it exists. Afterwards we can continue with the next interval in order to start again with a simple Predictor formula.

5.4.2.4 Tasks for the numeric solution of differential equation

1. Apply the simple Euler's method to the differential equation, until x = 1 with intervals, e.g. 0.5 0.2 and 0.1

$$y' = -xy^2 \qquad \text{with:} \qquad y(0) = 2$$

Do the results converge the accurate solution value $y(1) = 1$?

2. Apply the RUNGE-KUTTA method 4th order and a Predictor Corrector method on the problem specified above and compare the results.

3. The following differential equation applies to the concentration *C* [*mg*] in sorption of pollutants at the soil matrix:

$$T_1 C + C = K$$

T_1 is time constant and K is a constant. $T_1 = 1d^{-1}$, K = 100. The concentration should C(0) = 0 at time $t = 0$.
a) Solve the differential equation by means of Euler's method (Rectangle rule with $h = 0, 1d$) and compute the concentration change for the time $t = 1d$
b) Outline the time process of concentration change.

4. The padding to the remainder holes of the former brown coal open pit caused by the rise of groundwater under natural conditions will last too long time. Therefore external supply is introduced to the filling procedure ($h_{t=0} = 0$) for acceleration. (see figure 5.17)

Set up the differential equation for the filling up procedure *h(t)*, without consideration of the aquifer and contingent groundwater formation rate. Describe the solution by means of numerical methods.

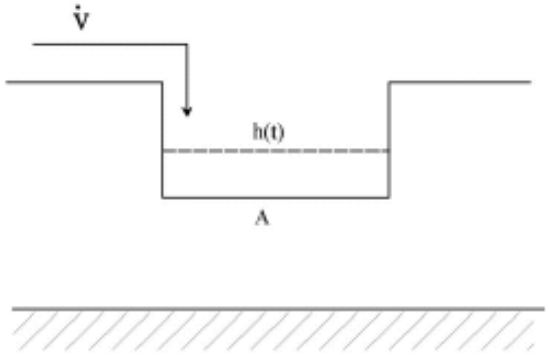


Figure 5.17: filling procedure of a remainder hole

5. Solve the following differential equation by means of numerical methods

$$\frac{dh}{dt} + k \cdot h = g \qquad \text{with} \qquad h_{t=0} = 0$$

$g = 0,015m \cdot s^{-1}$ und $k = 0,01s^{-1}$

Part II

Partial differential equations of underground processes

Chapter 6

6 Overview

There are no generally valid solutions for partial differential equations (**PDE**), which are characterised by consideration of functional dependency on more arguments.
Because of this reason the following selected PDE, which plays an important role in hydrogeology, will be discussed. The groundwater flow equation and convection dispersion equation predominantly stand in the foreground.
Different mathematical methods, such as **analytical** or **numerical** solution will be introduced according to the complexity of the equation, the number of independent parameters and the consideration of inhomogeneity, anisotropy, as well as nonlinearity.

The physical processes are divided into so called quantity flow- and material transportation. The application of energy- and mass conservation law lead to following coupled partial differential equation, for material process only transportation is displayed:

- The **dynamic basic equation** of **flow processes**:

$$\vec{v} = k \text{ grad } h \tag{6.1}$$

- The **balance equation** of **flow processes**:

$$\text{div } \vec{v} = S_0 \frac{\delta h}{\delta t} - w \tag{6.2}$$

- The **boundary condition** of **flow processes**:

initial- and boundary condition 1. 2. and 3. type

This equation system must be set up in the modelling of material and energy transportation for each substance contained in water or in immiscible material processes for each group of materials and for each phase in the multi-phase system (liquid (water, oils), solid (stone matrix), gaseous (air, gases)). Balance equations must be defined for each subsystem, which consist of the following parts:

- The **dynamic basic equation** for **transportation processes**:

Transport durch Dispersion: $$\vec{g}_1 = \vec{D} \text{ grad } P \tag{6.3}$$

Transport durch Konvektion: $$\vec{g}_2 = \vec{v} P \tag{6.4}$$

- The **balance equation** for **transportation processes**:

$$\text{div } \vec{g} = (n_0 + \alpha) \frac{\delta P}{\delta t} - w_g \tag{6.5}$$

• The **boundary condition** for **transportation processes**:

initial- and boundary condition 1. 2. and 3. type

The connection of the equations within each subsystem is given by the exchange terms. In the subsystems it is made via internal reaction terms. The following balance equation applies to a balance area, which is also designated as representative elementary volume (REV):

Transport = internal reaction + storage + exchange + external sources

The chemical reaction equations (material change processes) and biological growth processes can be added to these basic equations. The mathematical model thereby consists of a system of ordinary or partial differential equations and algebraic equations, whose coefficients are usually a function of place, time and potential. Thus the system is nonlinear and local- and time variant. The processes in the soil and groundwater range are characterized by a high complexity, a bad condition, a large range of time constants and a very uncertainty of initial parameters.

The basic equations can be summarized in each case for the flow and the material process, and yield two nonlinear partial differential equations with second order:

• The conduction equation for the flow process (parabolic PDE):

$$\text{div}\,(k_{(x,y,z)}\text{grad}\,h) = S_0\frac{\delta h}{\delta t} - w \qquad (6.6)$$

• The convection diffusion equation for the transportation process (hyperbolic PDE):

$$\text{div}\,\left(\vec{D}\text{grad}\;P\;-\vec{v}\,P\right) = (n_0+\alpha)\frac{\delta P}{\delta t} -\;w_g \qquad (6.7)$$

Depending upon the relationship of dispersion portion $\left(\vec{D}\,\text{grad}\,P\right)$ to convection $\left(\vec{v}\,P\right)$ in total transportation process the property of these PDE varies among predominantly parabolic, hyperbolic or first order PDE. If convection approaches to zero $\left(\vec{v}\,P \longrightarrow 0\right)$, the PDE is parabolic type, and we get first order PDE with $\left(\vec{D}\,\text{grad}\,P \longrightarrow 0\right)$.

The connection of the quantity and of the material flow is characterised by the critical values of the water property (temperature T, material concentration C, kinematical viscosity *v* and density ρ) and by the critical values of the underground flow processes (filter velocity $\left|\vec{v}\right|$, change of memory contents $C{\cdot}\delta p/\delta t$ as well as internal flow source and -sink *w*).

This complex form of the system description is often approximated by simplified forms, in which one or more processes are neglected or the dependence on one or other arguments is ignored. A

fundamental simplification results from the decoupled approach of flow and transportation processes and chemical kinetics. Substantial simplification can be also achieved by reduction of the multidimensional area to a local coordinate or time variable.

This procedure will be demonstrated exemplary by often used and significant engineering examples in the following chapters.

6.1 One dimensional flow equation

Under the prerequisite of simplified flow conditions, the view in the cylindrical coordinate space as well as the integration over the height of *z* by a transformation, e.g. the so called GIRINSKIJ potential ϕ, we get the following equations for the **rotational symmetric flow field**:

steady flow:
$$\frac{d^2\Phi}{dr^2} + \frac{1}{r}\frac{d\Phi}{dr} + \frac{w}{k} = 0 \qquad (6.8)$$

"leakier" flow (leaky aquifer):
$$\frac{d^2Z}{dr^2} + \frac{1}{r}\frac{dZ}{dr} - \frac{Z}{B^2} = 0 \qquad (6.9)$$

non steady flow:
$$\frac{\partial^2 Z}{\partial r^2} + \frac{1}{r}\frac{\partial Z}{\partial r} - \frac{Z}{B^2} = a\frac{\partial Z}{\partial t} \qquad (6.10)$$

These equations and other analytic solutions (see to section 8.1 THEIS well equation, page 196) were found by THEIS. The importance of these equations is that they supply useful results with a local character (approx. 200 m expansion, e.g. foundation pit) for many engineering investigations, which fulfils the hydraulic geological conditions. In addition they form the basic procedures for the indirect parameter investigation, e.g. for the so called pumping test evaluations (see section 14.1 pumping test evaluation, page 380).

Similar conditions occur at **parallel ditch flow**, and the PDE has the following form:

parallel ditch incident flow:
$$\frac{\partial^2 Z}{\partial x^2} - \frac{w}{k} = a\frac{\partial Z}{\partial t} \qquad (6.11)$$

The one dimensional processes are also meaningful for the investigation of transportation procedures in stream tube connected with pollution.

6.2 Horizontal plane groundwater flow equation

The horizontal plane groundwater flow equation represents a fundamental principle for the flow processes apart from the well equation (see equations 6.8 to 6.10). A simplified aquifer characterized by means of the DUPUIT assumption (see to section 7.1 DUPUIT assumption and balance equation, page 184), and an integral transform for the description of the profile permeability, the transmissibility *T*:

$$\text{div}\,(T_{(x,y)}\text{grad}\,z_R) = S\frac{\partial z_R}{\partial t} - w_N \qquad (6.12)$$

This equation can be built separately for each groundwater story and the coupling between the aquifer can be achieved by hydraulic windows. This equation forms the basics of most hydraulic geological region models, also for the mining districts of the Central Germany and the Lusatia area.

6.3 One dimensional material transfer

For the material transfer the modeling of the one dimensional processes also plays important role since on the one hand it is partly analytically solvable and on the other hand it is basics for the model measuring (e.g. the so called column test). Also it is backwards indirect Parameter estimation (e.g. tracer test). As example equations can be derived:

- **Heat transport due to precipitation in the unsaturated soil zone:**

$$\text{Wasser (Index } w\text{):} \quad \frac{\partial}{\partial z}(k_w \frac{\partial}{\partial z}(\frac{p_w}{p_w} + z)) = \frac{dn_w}{dt} - w_w \tag{6.13}$$

$$\text{Luft (Index } L\text{):} \quad \frac{\partial}{\partial z}(k_L \frac{\partial}{\partial z}(\frac{p_L}{p_L} + z)) = \frac{dn_L}{dt} - w_L \tag{6.14}$$

- **One dimensional Transport:**

$$\text{Konvektion:} \quad \varepsilon \frac{\partial C}{\partial t} = -q \frac{\partial C}{\partial x} \tag{6.15}$$

$$\text{Dispersion:} \quad \frac{\partial^2 C}{\partial z^2} = a \frac{\partial C}{\partial t} \tag{6.16}$$

$$\text{Dispersion und Konvektion:} \quad M D_1 \frac{\partial^2 C}{\partial x^2} - q \frac{\partial C}{\partial x} = \varepsilon \frac{\partial C}{\partial t} + \lambda C - w \tag{6.17}$$

The three cases are differentiated, with whether λ or ω are equal or unequal to zero.

6.4 Multiphase flow

Simultaneous effects of several phases in the porous medium, soil or aquifer are considered in the modeling of the multiphase flow. In the literature the relations of three-phase system are illustrated. According to the equation 6.7 on page 177 with neglecting the dispersion portion:

$$\text{div } (p_\alpha \vec{v}_\alpha) + \frac{\partial(\Phi p_\alpha S_\alpha)}{\partial t} = p_\alpha \qquad (6.18)$$

In this case α represents a general fluid phase. Within the three-phase system water (α = ω), NAPL (n) and air (a) will be considered. NAPL is the abbreviation for petroleum products (Non Aqueous phase liquid - NAPL).

DARCY law can be extended to the multiphase system by neglecting of the theorem of momentum between the fluid phases:

$$\vec{v}_\alpha = -\frac{k \cdot k_{r\alpha}}{\mu_\alpha}(\text{grad } p_\alpha + p_\alpha \cdot \vec{g}) \qquad (6.19)$$

The reciprocal effects between the individual phases will be described by additional equations, secondary conditions:

$s_w + s_n + s_a = 1$ (the pore area is filled out by the sum of the three phases)

$p_n \cdot p_w = P_{Cnw}(S_w, S_a)$ (Capillary pressure saturation relationship)

$p_a \cdot p_n = P_{Can}(S_w, S_a)$

$k_{ra} = k_{ra}(S_w, S_a)$ (relative permeability saturation relationship)

In many practical applications the nonlinear equation system is limited by the assumption that air is infinitely mobile in each case a movement phase for the water phase and for the NAPL phase.

Analytical solution from BUCKLEY and LEVERETT will be used for one dimensional displacement procedure of oil through water, which describes the transient procedures.
To describe the relative permeability saturation curve the COREY function can be set:

$$k_{rw} = S^{*4} \qquad (6.20)$$

$$k_{rn} = (1 - S^*)^2 \cdot (1 - S^{*2}) \qquad (6.21)$$

$$\text{mit : } S^* = \frac{(S_w - S_{wr})}{(1 - S_{wr} - S_{nr})}; \quad S_{wr} = S_{nr} = 0,2$$

For two-dimensional case and the investigation of three-phase system air/NAPL/water the following beginnings are examined:
For the **capillary pressure saturation relationship** PARKER formula:

$$P_{Cnw} = p_n - p_w = \frac{1}{\alpha_{vG} \cdot -\beta_{nw}} (S_e \frac{n_{vG}}{(1 - n_{vG})} - 1)^{\frac{1}{n_{vG}}} \tag{6.22}$$

$$P_{Can} = p_a - p_n = \frac{1}{\alpha_{vG} \cdot -\beta_{an}} ((\frac{S_n + S_w - S_{wr}}{1 - S_{wr}})^{\frac{n_{vG}}{(1-n_{vG})}} - 1)^{\frac{1}{n_{vG}}} \tag{6.23}$$

$$\text{mit} : S_e = \frac{S_w - S_{wr}}{1 - S_{wr}}; \beta_{nw} = \frac{\sigma_{aw}}{\sigma_{nw}}; \beta_{an} = \frac{\sigma_{aw}}{\sigma_{an}}$$

For the **relative permeability saturation relationship** for the Non Aqueous phase liquid(NAPL) a model by STONE will be used :

$$k_{rn} = \frac{S_n(1 - S_{wr}) k_{rnw} \cdot k_{ran}}{(1 - S_w)(S_n + S_w - S_{wr})} \tag{6.24}$$

Here k_{rnw} and k_{ran} stand for the relative permeability saturation relationship in NAPL phase of a two-phase system (water/NAPL) and (air/NAPL). The parameters S_{nr} and k_{rncw} were occupied with the values "0" and "1" in the original form of the STONE model. For the water phase the relationship PARKER formula is used:

$$k_{rw} = \sqrt{S_e} \left(1 - \left(1 - S_e \frac{n_{vG}}{(n_{vG} - 1)} \right)^{\frac{(n_{vG}-1)}{n_{vG}}} \right)^2 \tag{6.25}$$

$$\text{mit:} \qquad S_e = \frac{S_w - S_{wr}}{(1 - S_{wr})}$$

Chapter 7

7 Horizontal plane Groundwater flow equation

7.1 DUPUIT assumption and balance equation

The description of the rotationally symmetric groundwater flow field is based on horizontal planes flow processes, in which the vertical flow vector is neglected. The transfer of the three-dimensional flow regime into a two-dimensional mathematical description takes place with consideration of

DUPUIT assumption:

• The potential lines $h = const$ run parallelly to z-axis. This means that the vertical component of groundwater flow ($v_z \rightarrow 0$) is equal to zero. This can be realized by an infinitely large vertical flow resistance (specific permeability coefficient in z-direction ($k_z \rightarrow \infty$)) or by a no gradient gauge level:

$$\frac{\partial h}{\partial z} = 0$$

• The horizontal speed is constant during the entire through flow height of aquifer. It means the vertical gradients of the horizontal flow components are equal to zero.

$$\frac{\partial v_x}{\partial z} = \frac{\partial v_y}{\partial z} = 0 \qquad (7.1)$$

• The horizontal speed is proportional to the decline gradient of free surface according to the DARCY law:

$$v_y = -k_y \cdot \frac{\partial h}{\partial y} \qquad (7.2)$$

$$v_x = -k_x \cdot \frac{\partial h}{\partial x} \qquad (7.3)$$

The **force equilibrium law** is set up under the condition that only pressure force, gravity force, capillary force and internal friction are effective. Inertia forces, adhesive force, turbulent friction forces and others are small enough to be negligible. Since the groundwater movement is regarded as saturated filter flow, we know following law from DARCY:

$$\vec{v} = -k \cdot \text{grad } h \qquad (7.4)$$

The DARCY law is only valid when the precondition its derivative existing fulfils. Thus it loses its validity when the above neglected forces increase. In practical groundwater flow procedures the validity of DARCY law can be however accepted with sufficient accuracy. Only directly in

the proximity of well a breach of this law can occur with large filter velocity. With the DUPUIT assumption the balance equation for horizontal plane groundwater flow is built up. The specific flow rate, refer to flow field width of 1m, can be calculated:

$$\vec{q} = \int_{z=a}^{D} \vec{v}\, dz \tag{7.5}$$

D through flow thickness

$$D = \left\{ \begin{array}{ll} M & \text{Aquifer thickness in confined} \\ z_R & \text{Position of free groundwater surface in unconfined} \end{array} \right\} \text{ aquifer}$$

And the balance equation:

$$\operatorname{div} \vec{q} = \left(n_0 + \int_{z=a}^{D} S_0\, dz \right) \cdot \frac{\partial h}{\partial t} - w \tag{7.6}$$

w sources/sinks

n_0 storage coefficient at the free groundwater surface due to gravimetric effects

S_0 elastic storage coefficient, which works within the aquifer

The summary expression for the storage capability is designated with S as **general storage coefficient:**

$$S = \left(n_0 + \int_{z=a}^{D} S_0\, dz \right) \tag{7.7}$$

$$S = \left\{ \begin{array}{ll} n_0 + \int_{z=0}^{z_R} S_0\, dz & \text{confined} \\ \int_{z=0}^{M} S_0\, dz & \text{unconfined} \end{array} \right\} \text{ aquifer} \tag{7.8}$$

If the gravimetric storage coefficient is substantially larger than the sum of all elastic effects in vertical direction:

$$n_0 >> \int_{z=a}^{z_R} S_0\, dz \tag{7.9}$$

It results in that the storage coefficient is only dependent on gravimetric coefficient in the case of a free groundwater surface and a small through flow thickness aquifer (D <<100m). The storage coefficient S can take the following value:

$$S = \left\{ \begin{array}{ll} n_0 \approx n_a \approx n_e & \text{confined} \\ \int_{z=0}^{M} S_0 dz & \text{unconfined} \end{array} \right\} \text{ aquifer} \qquad (7.10)$$

For the effect water height h:

$$h = \left\{ \begin{array}{ll} h & \text{confined} \\ z_R & \text{unconfined} \end{array} \right\} \text{ aquifer}$$

Thus the balance equation, also as continuity equation, is written in the form:

$$\text{div}\, \vec{q} = S \cdot \frac{\partial h}{\partial t} - w \qquad (7.11)$$

With the equations 7.5 and 7.6 we get the horizontal plane groundwater flow equation in the following form:

$$\text{div} \left(\int_{z=a}^{D} k \, dz \text{grad}\, h \right) = S \cdot \frac{\partial h}{\partial t} - w \qquad (7.12)$$

According to definition h, *grad h* is independent on z thus it can be pulled out of integral. For further writing simplification the integral of permeability coefficient, the term **transmissibility** T is introduced:

$$T = \int_{z=a}^{D} k \, dz \qquad (7.13)$$

This integral of transmissibility will be analysed numerically poorly as the permeability coefficient is only expressed as step function and not continuous function.

Thus the **horizontal plane groundwater flow equation** in the representation of water height:

$$\left. \begin{array}{ll} \text{div}\,(T\ \text{grad}\, h) = S \cdot \frac{\partial h}{\partial t} - w & \text{confined} \\ \text{div}\,(T\ \text{grad}\, z_R) = S \cdot \frac{\partial z_R}{\partial t} - w & \text{unconfined} \end{array} \right\} \text{ aquifer} \qquad (7.14)$$

7.2 Potential illustration

An integral transform was used for solving partial differential equation of underground flow processes in the former chapter, which yields the value of transmissibility. Now in this section another integral transform is applied, the so called **GIRINSKIJ potential** Φ also a relatively simple solution, thus the horizontal plane groundwater flow equation is illustrated in potential expression.

The **GIRINSKIJ potential** Φ is defined as:

$$\Phi(x,y) = \int_{z=a}^{D} g(z) \cdot (h(x,y,z) - z)\, dz \qquad (7.15)$$

In this equation the function *g(z)* characterizes the dependence of permeability coefficient *k* on height *z*.

$$k(x,y,z) = k(x,y) \cdot g(z) \qquad (7.16)$$

For the following considered unstratified aquifer:

$$g(z) = 1$$

Here with the validity of DUPUIT assumption $(\frac{\partial h}{\partial z} = 0;\ h \neq f(z))$ and the assumption of lower bound of aquifer *a* equal to zero (*a* = 0), the integral yields two solutions:

$$\Phi(x,y) = \int_{z=0}^{D} (h - z)\, dz = \left[h \cdot z - \frac{z^2}{2} \right]_0^D \qquad (7.17)$$

$$= \left\{ \begin{array}{ll} M \cdot h - \dfrac{M^2}{2} & \text{confined} \\ \dfrac{z_R^2}{2} & \text{unconfined} \end{array} \right\} \text{ aquifer} \qquad (7.18)$$

With consideration of DARCY law the specific volume flow:

$$\vec{q} = \int_{z=0}^{D} \vec{v} dz \qquad (7.19)$$

And with $\vec{v} = -k \text{ grad } h$

$$\vec{q} = \int_{z=0}^{D} (-k \cdot \text{grad } h) dz \qquad (7.20)$$

Since k and h are not functions of z, we can write k before the integral and exchange the consequence of the deviation (gradient) and the integration.

We get:

$$\vec{q} = -k \cdot \operatorname{grad} \int_{z=0}^{D} h \, dz \tag{7.21}$$

$$\vec{q} = -k \cdot \operatorname{grad} \Phi \tag{7.22}$$

The horizontal plane groundwater flow equation in potential form:

$$\left.\begin{array}{ll} \operatorname{div}(k \ \operatorname{grad} \ \Phi) = S\dfrac{\partial h}{\partial t} - w & \text{confined} \\ \operatorname{div}(k \ \operatorname{grad} \ \Phi) = S\dfrac{\partial z_R}{\partial t} - w & \text{unconfined} \end{array}\right\} \text{ aquifer} \tag{7.23}$$

This PDE can be transferred into a uniform potential writing way, if the definition for the GIRINSKIJ potential is separately introduced according to confined and unconfined conditions:

$$S\frac{\partial h}{\partial t} = S \cdot \frac{\partial h}{\partial \Phi} \cdot \frac{\partial \Phi}{\partial t} \tag{7.24}$$

$$= \left\{\begin{array}{lll} S \cdot \dfrac{1}{M} \cdot \dfrac{\partial \Phi}{\partial t} = \dfrac{S}{M}\dfrac{\partial \Phi}{\partial t} & \text{mit } \Phi = Mh - \dfrac{M^2}{2} & \text{confined} \\ S \cdot \dfrac{1}{\sqrt{2\Phi}} \cdot \dfrac{\partial \Phi}{\partial t} = \dfrac{S}{z_R}\dfrac{\partial \Phi}{\partial t} & \text{mit } \Phi = \dfrac{z_R^2}{2} & \text{unconfined} \end{array}\right\} \text{ aquifer} \tag{7.25}$$

And:

$$h = \frac{\Phi}{M} + \frac{M}{2} \tag{7.26}$$

$$\frac{\partial h}{\partial \Phi} = \frac{1}{M} \tag{7.27}$$

Or:

$$z_R = \sqrt{2\Phi} \tag{7.28}$$

$$\frac{\partial z_R}{\partial \Phi} = \frac{1}{\sqrt{2\Phi}} \tag{7.29}$$

Assuming a homogeneous, isotropic aquifer, i.e. $k = const.$, then k can be calculated from the divergence and the division of the right side. With introduction of the transmissibility and the geohydraulic time constant we get:

$$\operatorname{div}(\operatorname{grad} \Phi) = a\frac{\partial \Phi}{\partial t} - \frac{w}{k}$$

With:

$$a = \frac{S}{T} \tag{7.30}$$

$$T = \int_{z=0}^{D} k\,dz = \left\{ \begin{array}{ll} k \cdot M & \text{confined} \\ k \cdot z_R & \text{unconfined} \end{array} \right\} \text{ aquifer} \tag{7.31}$$

Now we have found a universally valid PDE, which is linear and analytically solvable.

However it must be noted that the linearity is not exact under free groundwater surface conditions, i.e. with unconfined aquifer, since the geohydraulic time constant a is a function of z_R. In this case a temporal average value will be taken for T and also for a. The following approximation has been well proved for a:

$$\tilde{a} \approx \frac{a_{t=0} - 2a_t}{3} \tag{7.32}$$

For extreme drawdown ratios over 10% of groundwater level this equation is only valid approximately. The water level of standpipe often changes, i.e. drawdown, of note in GIRINSKIJ potential. Therefore the potential difference between the output potential Φ_0 and the current potential Φ is used. In unstratified aquifer, i.e. with $k(z)$ = const. and thus $g(z) = 0$:

$$Z = \Phi_{t=0} - \Phi = \int_{z=0}^{D} (h_{t=0} - z)\,dz - \int_{z=0}^{D} (h_t - z)\,dz \tag{7.33}$$

$$= \left\{ \begin{array}{ll} M(h_{t=0} - h) & \text{confined} \\ \dfrac{(z^2_{Rt=0} - z^2_R)}{2} & \text{unconfined} \end{array} \right\} \text{ aquifer}$$

The subscript 0 stands for conditions to time point $t = 0$, *i.e.* $\Phi_0 = \Phi_{t=0}$, $h_0 = h_{t=0}$, $z_{R0} = z_{Rt=0}$. In some citation the subscript n (Φ_n, h_n, z_{Rn}) is also used for it.

Inserting this into the PDE:

$$\text{div (grad } Z) = a\frac{\partial Z}{\partial t} + \frac{w'}{k} \tag{7.34}$$

By definition w' is the supply quantity caused by the change of potential Z, is by definition, while w represents the supply quantity, which affects from the outside of aquifer, e.g. the natural groundwater replenishment:

$$\frac{w'}{k} = \frac{w}{k} - \left(\frac{w}{k}\right)_{t=0} \tag{7.35}$$

Practically the following two cases are interested:

- $w' = 0$,
  i.e. supply conditions won't change when the regarded groundwater level varies, and

- $\frac{w'}{k} = \frac{Z}{B^2}$,
  i.e. the difference of potential Z causes an additional proportional supply (see section 8.1.3 supply from neighbouring layers, page 221)

If we a supply factor B add in all cases, which approaches to infinite for the first case($B \Rightarrow \infty$), we get the general form, the **standard form** of horizontal planes groundwater flow equation:

$$\text{div } (\text{grad } Z) = a\frac{\partial Z}{\partial t} + \frac{Z}{B^2} \qquad (7.36)$$

To solve this PDE it is necessary to transfer the general vectorial differential way of writing into a coordinate related way of writing (see section 2.2 arithmetic rules of the vector algebra, page 47).

During introduction of the cartesian coordinates we get a PDE, and we can inspect the groundwater flow processes in connection with the ditch flow (see to section 6.1 one dimensional flow equation, page 178). The cylindrical coordinates lead to a demonstration, which is very useful for rotationally symmetrical problems (see to section 8.1 THEIS well equation, page 196).

7.3 Marginal conditions

Each flow process takes place in a locally and temporally defined area, i.e. it represents a closed system, which is connected with its environment under certain conditions. Information, energy and matter can be exchanged through such couple conditions. They are called **marginal conditions**. While the system is described by the PDE and generally valid for all conditions, a unique solution will be achieved by marginal conditions. The effect of the marginal conditions is identical to determination of the integration constant by solving differential equation. Marginal conditions are impressed to the regarded system from the outside and influence independently of state variables.

It is differentiated between **boundary conditions** (marginal conditions at certain local points) and initial conditions (marginal conditions of reference time point). Besides force equilibrium and mass conservation, the initial- and boundary condition serve explicit mathematical description of the original process, the flow process. They are regarded as a part of the mathematical model.

7.3.1 Initial conditions

In dynamic systems relative time will be discussed. The absolute time point, from which the system behaviour is changing from static into movement, is regarded as starting point with relative time t = 0. The initial conditions serve to definition of the dynamic system state at this point. Since the state variable of horizontal planes groundwater flow is the piezometric head h or the situation of free surface, the initial condition is a matter of potentials within the systems, i.e. the groundwater height or the pertinent transform potential. In the model this will be a function of h, z_R or Φ dependent on location. Other conditions are also necessary for the maintenance of steady state at the border. These are from the same type like the boundary conditions. The difference is they are valid for $t < 0$.

7.3.2 Boundary conditions

Three different kinds of boundary conditions differ in physical action modes in the groundwater flow (see figure 7.1):

1st type (DIRICHLET condition)

2nd type (NEUMANN condition)

3rd type (CAUCHY condition)

Boundary conditions are general functions of place and time. We differentiate boundary conditions between influence inside of flow field (e.g. well, lakes, rivers, precipitation, evaporation), and outside effect at the edge (e.g. delimitation of flow field by rivers or barriers). It is characteristic for boundary conditions that its effect is independent on the flow conditions (e.g. Groundwater level) of the investigation area. Generally it is nearly impossible to find a complete analytical expression for geohydraulic boundary conditions.

• **1st type boundary condition (DIRICHLET condition)** works,
if the hydraulic potential (e.g. h, z_R, Z, Φ) on the boundary is known as a function of the time t and independent on the potential, i.e. the system variables of the investigation area. This appears e.g. in rivers, lakes or drainage:

$$\varphi = \varphi(x, y, t) \qquad (7.37)$$

• **2nd type boundary condition (Neumann condition)** works,
if the source intensity distribution and thus the hydraulic potential gradient on the bound are known as a function of time t. This may be aroused for example by wells with constant flow rate, supply due to groundwater regeneration, sealing of sheet pile wall or underground structures:

$$\text{grad } \varphi = \text{grad } \varphi(x, y, t) \qquad (7.38)$$

• **3rd type boundary condition (CAUCHY condition)** works,
if in general a temporally constant flow resistance exists between a surface with known potential distribution and the boundary of flow field. Such boundary conditions work in rivers with colmation bottom layer as well as flow resistance of lift wells:

$$\varphi + A \text{ grad } \varphi = B \text{ (A and B are definite constants)} \qquad (7.39)$$

In the figure 7.1 the effect of boundary conditions on an aquifer is demonstrated. We recognize that the flow rates of 1st and 3rd boundary conditions dependent on the difference between effect potential (water level h) in the aquifer and the boundary conditions. Therefore the flow rate can vary in amount and direction

In 2nd boundary condition the potential of the boundary condition (possibly regarded as negative or positive pressure) can vary accordingly with the potential of aquifer.

With the numerical models (see to section 9.1.1 numerical methods, e.g. finite differences methods, page 237) additional boundary conditions arise in the course of the definition of the model borders. In contrast to the original procedure, which possesses an infinite spatial expansion, the numerical models are spatially limited due to the finite computing capacity (memory space, computing speed). Thus a considerable error arises, which must be reduced or eliminated by suitable measures (see section 9.1.1 finite differences method, page 237).

Another problem related to boundary conditions arises in the interaction investigation of surface and aquifer systems. A volume flow appears between the surface and the aquifer or the unsaturated soil zone.
Depending upon potential conditions an ex- or infiltration of surface water can come out or into the aquifer. If this filtration stream is substantially smaller than the volume flow within water, or the filtration amount is substantially smaller than the entire storage volume of surface water, the surface water has an effect of boundary condition on the groundwater. In the other case, if the potential of surface water varies due to the filtration phenomenon, the surface water may be not considered as boundary condition, but components of the system and are coupled to the aquifer model according to suitable mathematical relations.

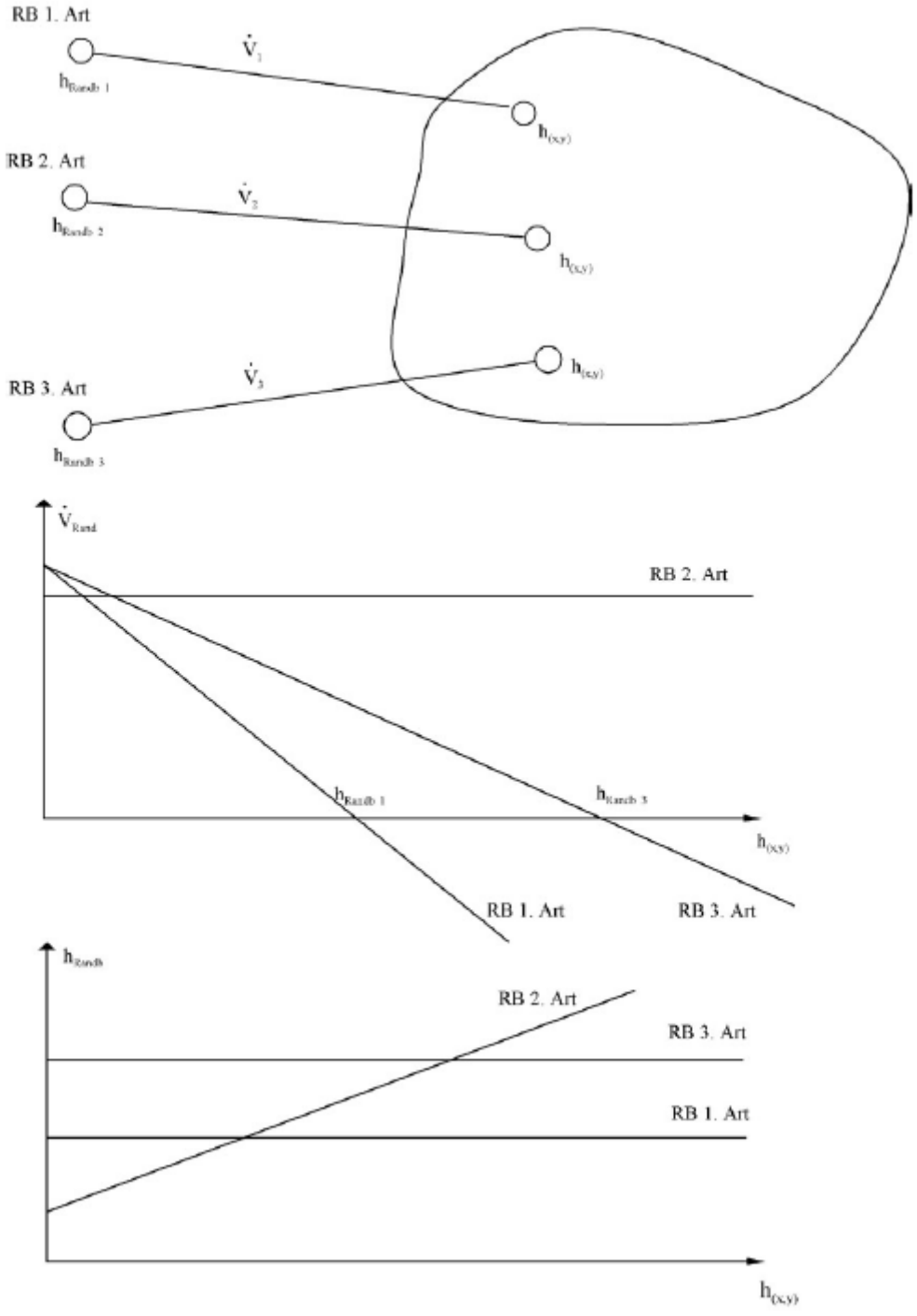


figure 7.1: Effect of boundary conditions on an aquifer

Chapter 8

8 Analytical Solution

8.1 THEIS well equation (Rotationally symmetrical flow)

The computation of rotationally symmetric flow field, i.e. the solution of partial differential equation, represents primary task of geohydraulics.

Such procedures can be described by means of horizontal planes groundwater flow equation **standard form**, which is deduced in the section 7.2 potential illustration, page 187:

$$\mathrm{div}(\mathrm{grad}\ Z) = a\frac{\partial Z}{\partial t} + \frac{Z}{B^2}$$

Changes of groundwater flow conditions with employment of vertical filter wells were first calculated by THEIS in the year 1935 and replenished by several other authors (e.g. THIEM, JACOB, COOPER, NEUMANN, HANTUSH and others). Due to importance of this task numerous publications and text books refer to this topic (for instance BUSCH/LUCKNER/TIEMER, Geohydraulik, and others).

8.1.1 General solution

On the basis of partial differential equation we are looking for a solution for the standard form of horizontal planes groundwater flow equation, i.e. the drawdown of the groundwater level as a function of place and time, if a change of boundary condition takes place at relative time point t = 0.

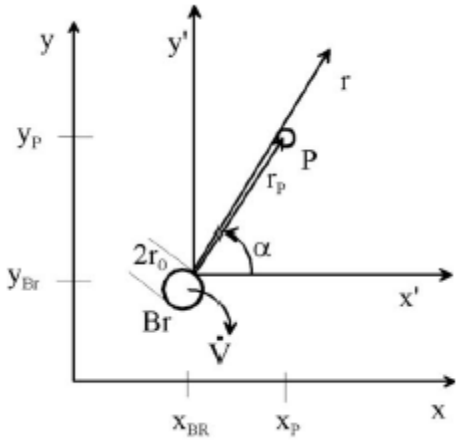


Figure 8.1 Coordinate system for rotationally symmetrical well

According to vectorial transfer in coordinate binding differential operators, here in cylindrical coordinates (see figure 8.1, see section 2.2 arithmetic rules of vector algebra, page 47), we know that the flow field is rotationally symmetric ($Z(\alpha) = 0$) and no dependent on the local coordinate z ($Z(z) = 0$):

$$\text{div}(\text{grad } Z) = a\frac{\partial Z}{\partial t} + \frac{Z}{B^2} \tag{8.1}$$

$$\frac{\partial^2 Z}{\partial r^2} + \frac{1}{r}\frac{\partial Z}{\partial r} = a\frac{\partial Z}{\partial t} + \frac{Z}{B^2} \tag{8.2}$$

Thus the solution of this partial differential equation, the drawdown potential Z depends, only on the time t and the radius r (distance between well and calculation point). Under above prerequisite the parameters of the aquifer, the permeability coefficient k and the storage coefficient S change neither with the height of z nor with the angle α in the regarded area.

The simplest solution can be achieved if following initial- and boundary conditions are considered and the aquifer is regarded as infinitely expanded, homogeneous and isotropic flow field (see figure 8.2).

Initial condition: $$Z_{(r,t=0)} = 0 \tag{8.3}$$

External boundary condition: $$\lim_{r\to\infty} Z_{(r,t)} = 0 \tag{8.4}$$

Internal boundary condition: $$\lim_{r\to 0} 2\pi r k\frac{\partial Z}{\partial r} = -\dot{V} = \text{const} \tag{8.5}$$

The internal boundary condition can be technically realized, if a vertical filter well is arranged at the origin of the cylindrical coordinates (r = 0) , which conveys a constant flow rate V starting from time point $t = 0$ (see figure 8.2).

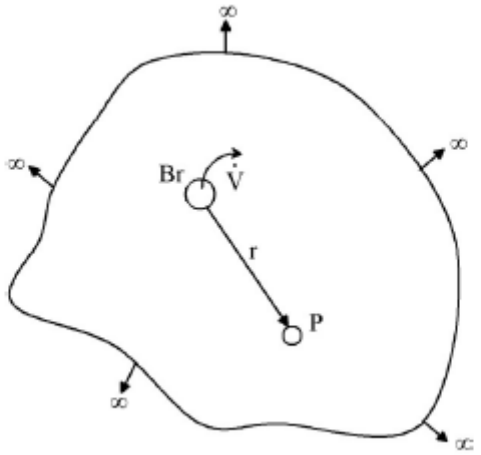


Figure 8.2 infinitively expanded aquifer

The solution of corresponding simplified partial differential equation was found by THEIS under aforementioned conditions:

$$\frac{\partial^2 Z}{\partial r^2} + \frac{1}{r}\frac{\partial Z}{\partial r} = a\frac{\partial Z}{\partial t} \tag{8.6}$$

$$Z(r,t) = \frac{\dot{V}}{4\pi k}W(\sigma), \qquad \text{mit } \sigma = \frac{ar^2}{4t} \quad \text{und } a = \frac{S}{T} \tag{8.7}$$

Neither groundwater generation rates nor supply from neighbouring layers are taken into account here. Remarks in addition are in the section 8.1.3 supply from neighbouring layers, page 221.

The so called **well function** $W(\sigma)$ is specified as the integral of an exponential function, which is known as **exponential integral** Ei(x) in the analysis and defined as follows:

$$\mathrm{Ei}(x) = \int_1^\infty \frac{e^{-xt}}{t} dt \tag{8.8}$$

$$\mathrm{Ei}(x) = \gamma + \ln x + \sum_{n=1}^{\infty} \frac{x^n}{n \cdot n!}, \tag{8.9}$$

γ stands for **Euler's constant** and is equal to:

$$\gamma = \lim_{n\to\infty} \left(\sum_{n=1}^{\infty} \left(\frac{1}{n}\right) - \ln(n) \right) \tag{8.10}$$

$$\gamma \approx 0,5772156649$$

with x = - σ :

$$W(\sigma) = -\mathrm{Ei}(-\sigma) = \int_1^\infty \frac{e^{\sigma}}{\sigma} d\sigma \tag{8.11}$$

This integral is not elementarily solvable, but expressed as an infinite series.

$$W(\sigma) = -\gamma - \ln(\sigma) + \sigma - \frac{\sigma^2}{2 \cdot 2!} + \frac{\sigma^3}{3 \cdot 3!} - \frac{\sigma^4}{4 \cdot 4!} + \cdots (-1)^{n+1} \frac{\sigma^n}{n \cdot n!} \tag{8.12}$$

$$W(\sigma) = -\ln(C\sigma) + \sigma - \frac{\sigma^2}{2 \cdot 2!} + \frac{\sigma^3}{3 \cdot 3!} - \frac{\sigma^4}{4 \cdot 4!} + \cdots (-1)^{n+1} \frac{\sigma^n}{n \cdot n!}$$

$$W(\sigma) = -\ln(C\sigma) + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sigma^n}{n \cdot n!}$$

with $C = e^{\gamma}$

$$\gamma \approx 0,5772156649$$

$$C \approx 1,7810724$$

To solve equation 8.6 series development and substitution method can be applied and similar methodology such as Bessel function (see section 5.2.2.3 differential equation of type C, page 130).
Table 8.1 contains the values of well function $W(\sigma)$ for range: $1 \cdot 10^{-12} \leq \sigma \leq 9$

Table 8.1: well function $W(\sigma)$ for range: $1 \cdot 10^{-12} \leq \sigma \leq 9$

σ Mantis.	1	2	3	4	5	6	7	8	9
Exponent									
$1 \cdot 10^{+00}$	0,2194	0,0496	0,0143	0,00378	0,00115	$3,60 \cdot 10^{-4}$	$1,16 \cdot 10^{-4}$	$3,77 \cdot 10^{-5}$	$1,24 \cdot 10^{-5}$
$1 \cdot 10^{-01}$	1,8229	1,2227	0,9057	0,7024	0,5598	0,4544	0,3738	0,3106	0,2602
$1 \cdot 10^{-02}$	4,0379	3,3547	2,9591	2,6813	2,4679	2,2953	2,1508	2,0269	1,9187
$1 \cdot 10^{-03}$	6,3315	5,6394	5,2349	4,9482	4,7261	4,5448	4,3916	4,2591	4,1423
$1 \cdot 10^{-04}$	8,6332	7,9402	7,5348	7,2472	7,0242	6,8420	6,6879	6,5545	6,4368
$1 \cdot 10^{-05}$	10,9357	10,2426	9,8371	9,5495	9,3263	9,1440	8,9899	8,8563	8,7386
$1 \cdot 10^{-06}$	13,2383	12,5451	12,1397	11,8520	11,6289	11,4465	11,2924	11,1589	11,0411
$1 \cdot 10^{-07}$	15,5409	14,8477	14,4423	14,1546	13,9314	13,7491	13,5950	13,4614	13,3437
$1 \cdot 10^{-08}$	17,8435	17,1503	16,7449	16,4572	16,2340	16,0517	15,8976	15,7640	15,6462
$1 \cdot 10^{-09}$	20,1460	19,4529	19,0474	18,7598	18,5366	18,3543	18,2001	18,0666	17,9488
$1 \cdot 10^{-10}$	22,4486	21,7555	21,3500	21,0623	20,8392	20,6569	20,5027	20,3692	20,2514
$1 \cdot 10^{-11}$	24,7512	24,0581	23,6526	23,3649	23,1418	22,9595	22,8053	22,6718	22,5540
$1 \cdot 10^{-12}$	27,0538	26,3607	25,9552	25,6675	25,4444	25,2620	25,1079	24,9744	24,8566

From the definition of drawdown or GIRINSKIJ potential the inverse transform for physical dimension water level h or z_R and the drawdown s can be accomplished:

$$Z = \Phi_n - \Phi = \frac{\dot{V}}{4\pi k} W(\sigma) \tag{8.13}$$

$$\Phi = \int_{z=a}^{D} (h - z) dz = \left\{ \begin{array}{ll} Mh - \frac{M^2}{2} & \text{confined} \\ \frac{z_R^2}{2} & \text{unconfined} \end{array} \right\} \text{ aquifer} \tag{8.14}$$

and

$$s = \left\{ \begin{array}{ll} h_n - h & \text{confined} \\ z_{Rn} - z_R & \text{unconfined} \end{array} \right\} \text{ aquifer} \tag{8.15}$$

For **confined** flow condition:

$$\begin{aligned} Z &= \Phi_n - \Phi \\ &= \left(Mh_n - \frac{M^2}{2} \right) - \left(Mh - \frac{M^2}{2} \right) \\ &= M \left(h_n - h \right) \end{aligned}$$

$$s_{gesp.} = \frac{Z}{M} = \frac{\dot{V}}{4\pi T} W(\sigma) \qquad \text{mit } T = k \cdot M \tag{8.16}$$

or

$$h = h_n - \frac{Z}{M} = h_n - \frac{\dot{V}}{4\pi T} W(\sigma) \tag{8.17}$$

For **unconfined** flow condition:

$$\begin{aligned} Z &= \Phi_n - \Phi \\ &= \frac{z_{Rn}^2}{2} - \frac{z_R^2}{2} \end{aligned}$$

$$s_{ungesp.} = z_{Rn} - \sqrt{z_{Rn}^2 - 2Z} = z_{Rn} - \sqrt{z_{Rn}^2 - \frac{\dot{V}}{2\pi k} W(\sigma)} \tag{8.18}$$

or

$$z_R = \sqrt{z_{Rn}^2 - 2Z} = \sqrt{z_{Rn}^2 - \frac{\dot{V}}{2\pi k} W(\sigma)} \tag{8.19}$$

This sharp separation between confined and unconfined groundwater conditions and modelling by means of the THEIS solution are not consistently implemented in all literature. This can also be applied to graphic methods for pumping test evaluation. In the case of very thick aquifer, for instance in North Germany the drawdown only amounts to a few percentage of thickness, it can be possibly calculated with the formula for confined aquifer.

From above derived formulas of the drawdown for different aquifers we know that under unconfined groundwater conditions the position change of free surface represents a strongly nonlinear process.

Please note that the well formula is not valid in the proximity of well with $r \rightarrow r_0$. The reason is that it does not fulfil the prerequisites, which were the derivation of rotationally symmetrical flow equation. So the vertical flow component which should be $v_z = 0$ can not be applied in the proximity of well. Also the effective well radius r_0^* is mostly not exactly confirmed. The storage effects and the flow resistances in the well area are hardly predictable (see section 14.2 pumping test simulator, page 388). In spite of some restrictions the well formula is a fundamental calculation formula in geohydraulics for computation of groundwater level height drawdown according to a volume stream.

The potential series $W(\sigma)$ (see equation 8.12, page 198) already strongly converge from the value $\sigma < 0.03$, so that terms with higher order of σ are small enough to be negligible. Thus $W(\sigma)$ can be computed in the case of an error < 1% only with logarithmic function. This approximation was advanced by COOPER & JACOB in 1946:

$$W(\sigma) \approx -\ln(C \cdot \sigma) \tag{8.20}$$

$$W(\sigma) \approx \ln\left(\frac{2,246\, T \cdot t}{S \cdot r^2}\right) \tag{8.21}$$

This simplified formula has great importance for many practical applications. In particular all graphic procedures of pumping test evaluation (see section 14.1 pumping test evaluation, page 380) are based on this formula (also see table 8.2).

The relative error, which results from the approximation, can be computed as follows:

$$\varepsilon = \left|\frac{W(\sigma)_{C\&J} - W(\sigma)}{W(\sigma)}\right| = \left|\frac{\sum_{n=1}^{\infty} \frac{(-1)^n \sigma^n}{n \cdot n!}}{\left(-\ln(C\sigma) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \sigma^n}{n \cdot n!}\right)}\right| \tag{8.22}$$

Since the series for $\sigma < 1$ converge very fast, it is only necessary to include the first term of the series to error estimation. All further terms will be substantially smaller than the linear term and thereby can be neglected. The second term only contributes a portion, which is squarely smaller than the first one.

$$\varepsilon = \frac{\sigma}{-\ln(C\sigma) + \sigma}$$

In table 8.2 the error for approximation by COOPER & JACOB is dependent on σ.

Table 8.2 Error of COOPER and JACOB formula as a function of σ

σ	$\varepsilon = \frac{\sigma}{(\sigma - \ln(C \cdot \sigma))}$	**Minimale Zeit** t
$C = 1,7811$		
$0,25000$	$23,61\%$	$1,00 \cdot r^2 \cdot a$
$0,20000$	$16,23\%$	$1,25 \cdot r^2 \cdot a$
$0,15000$	$10,20\%$	$1,67 \cdot r^2 \cdot a$
$0,10000$	$5,48\%$	$2,50 \cdot r^2 \cdot a$
$0,07500$	$3,59\%$	$3,33 \cdot r^2 \cdot a$
$0,05000$	$2,03\%$	$5,00 \cdot r^2 \cdot a$
$0,03000$	$1,01\%$	$8,33 \cdot r^2 \cdot a$
$0,02500$	$0,80\%$	$10,00 \cdot r^2 \cdot a$
$0,01000$	$0,25\%$	$25,00 \cdot r^2 \cdot a$
$0,00750$	$0,17\%$	$33,33 \cdot r^2 \cdot a$
$0,00500$	$0,11\%$	$50,00 \cdot r^2 \cdot a$
$0,00250$	$0,05\%$	$100,00 \cdot r^2 \cdot a$
$0,00100$	$0,02\%$	$250,00 \cdot r^2 \cdot a$
$0,00075$	$0,01\%$	$333,33 \cdot r^2 \cdot a$

Simultaneously we get an estimation for the relation of computation place (*r*), computation time (*t*) and approximation errors (*ε*). It is also recognized that the geohydraulic time constant ($a = S/T$) representatively determines the stop accuracy. Thus the more computation time point approaches to steady state, the more exact the computation is. For the unsteady transition region the approximation of COOPER &JACOB is not well applicable and large approximation errors yield.

The computational evaluation of well function *W* (σ) can be substantially simplified by a recursive expression of sum formula.

$$W(\sigma) = -\ln(C\sigma) + \sum_{n=1}^{a_n \le \varepsilon} a_n \tag{8.23}$$

With $a_n = a_{n-1} \frac{(-1)\,\sigma\,(n-1)}{n^2}$

And $a_1 = \sigma$

Thus it yields:

• Only certain terms must be calculated for a given accuracy (e.g. between two sum terms).

• The computation of each sum term requires only a multiplication.

8.1.2 Consideration of special effects

The general solution of well equation according to THEIS only applies to a very ideal aquifer. So it is assumed e.g. as homogeneous and isotropic. Further more an infinite expansion is supposed. Only one well is considered in the solution, which conveys from time t_0 with constant stream and is arranged as singularity with a radius of r_{Br} = 0m at the coordinate origin. These idealizations can not be found in practice with real aquifers. For some practice-relevant conditions however results can be obtained based on THEIS solution, if appropriate auxiliary computations and substitutions are accomplished.
Such are for example the consideration of technical well radii, imperfection of the wells and boundary conditions, as well as the laminated aquifers and supply from neighbouring layers. These special effects are regarded as additional potential in well equation, which are established and dismantled.

8.1.2.1 imperfect well

Imperfect boundary conditions, particular wells appear, if the boundary conditions or wells do not act on the entire thickness of the aquifer. In the case of wells it happens if the working filter pipe length is smaller than the thickness of the aquifer through flow. With the imperfect wells it is assumed a potential loss results from through flow thickness near the imperfect well smaller than the actual aquifer. Besides we can suppose that the average flow path via redirecting is longer than the geometrical distance r. Figure 8.3 shows conditions with different filter installation.

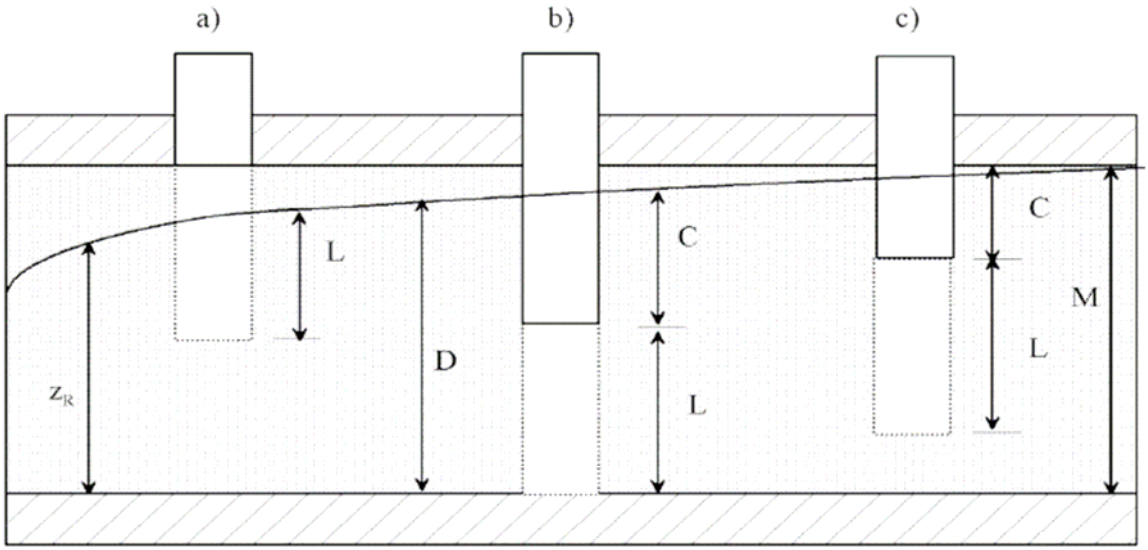


Figure 8.3: Imperfect wells with filter in the a) upper, b) lower, c) middle part of the aquifer

C Length of the full pipe

D through flow thickness

L Length of the filter pipe within through flow thickness

M thickness of the aquifer

Z_R Position of the groundwater free surface

φ_V Filter losses

We can describe the potential loss as follows:

$$Z(r,t) = \frac{\dot{V}}{4\pi k}\left(W(\sigma) + \varphi_V\right) \tag{8.24}$$

$$\varphi_V = 2\left[\frac{D}{L}\ln\frac{\alpha L}{r_0} - \ln\frac{\alpha D}{r_0}\right] - \sqrt{1-\frac{L}{D}} \tag{8.25}$$

$$\alpha = 0,735\left(1 + (H-1)^4\right) \tag{8.26}$$

$$H = \frac{2C}{D-L} \tag{8.27}$$

We recognize that a stronger sink occurs due to the imperfection, comparing with the case of a perfect well.

8.1.2.2 Multi-well plants

In practical multi-well plants are meaningful. In seldom cases e.g. foundation pit drainage or a ground water works only one well is operated. The computation for such multi-well plants is possible based on principle of superposition. The solutions, i.e. the partial drawdown potentials, which apply to the individual wells, will be superposed, i.e. overlaid together (see figures 8.4 and 8.5). The principle of superposition can be only applied in linear systems. Relating to THEIS solution this means that the superposition can be only used in the potential expression. When the superposed potential is formed, the inverse transform of physical dimension drawdown or water level can be accomplished. Since the connection between potential and drawdown for the confined aquifer is linear, the superposition in this case could be also exceptionally applied to the drawdown.

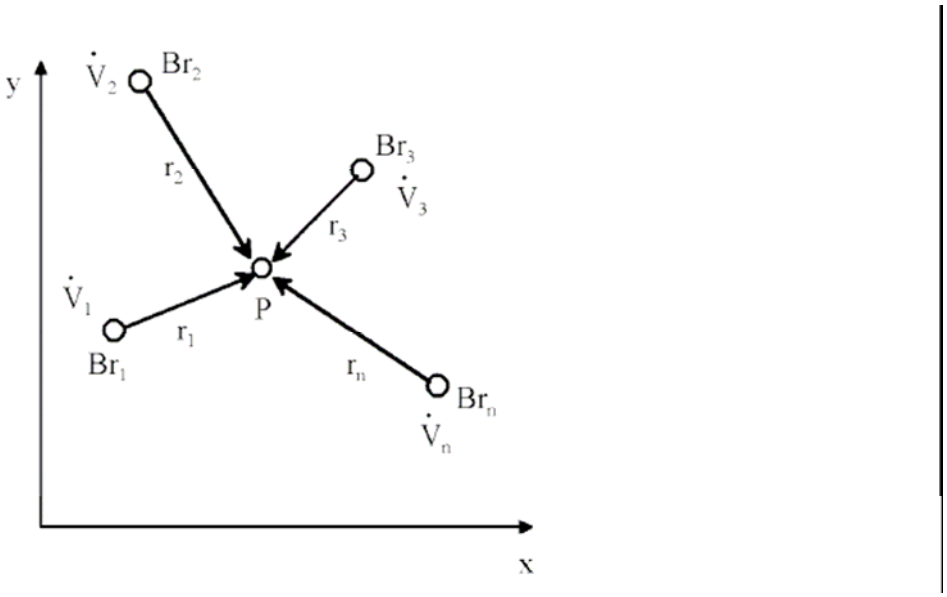


Figure 8.4 Multi-well plant

In principle:

$$
\begin{aligned}
Z_{Ges.x,y,t} &= \sum_{i=1}^{n} Z\left(r_i, t\right) \\
&= \sum_{i=1}^{n} \left(\frac{\dot{V}_i}{4\pi k} W_i\left(\sigma_i\right) \right) \\
&= \frac{1}{4\pi k} \sum_{i=1}^{n} \left(\dot{V}_i \cdot W_i \left(\frac{r_i^2 a}{4t} \right) \right)
\end{aligned}
$$

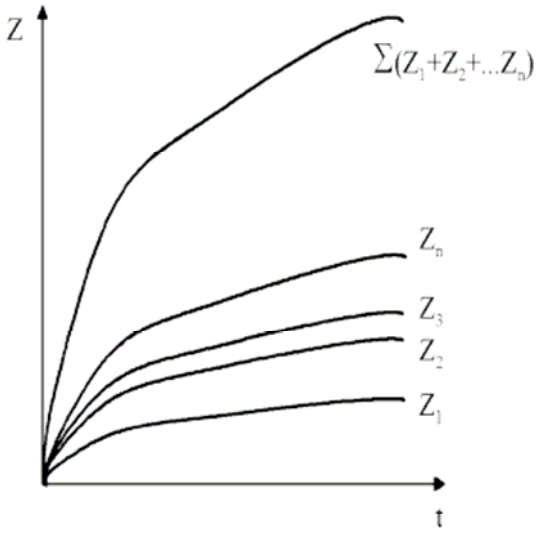


figure 8.5 Superposition of the drawdown potentials

r_i distances between the individual well and the computation point $P_{x,y}$
a geohydraulic time constant for the entire area a = const.

First we calculate the individual drawdown portions of *Z(ri, t)* due to the well effects *Vi* and then sum them up. Afterwards the conversion in the total drawdown takes place according to groundwater conditions (confined or unconfined) and equations 8.16 and 8.18 (see page 201).

8.1.2.3 Variable conveying curve of wells

The THEIS solution assumes the internal boundary condition that, the flow rate affects on the aquifer from time point $t = 0$. In practice it often happens that, this condition is not fulfilled. Because of technical/technological criteria a variation of pump capacity is often required. This problem play an important role, if the groundwater level after switching off the pump is in the so called rising phase.

Also here a solution based on THEIS formula can be obtained by means of superposition principle. The basic idea consists of the fact that time-dependent conveyor capacity is set as summation of temporally transfer step functions. Figuratively we can imagine *n* fictitious pumps, which are put on the same well successively according to the conveying stages and switched on (see figure 8.6 and 8.7).

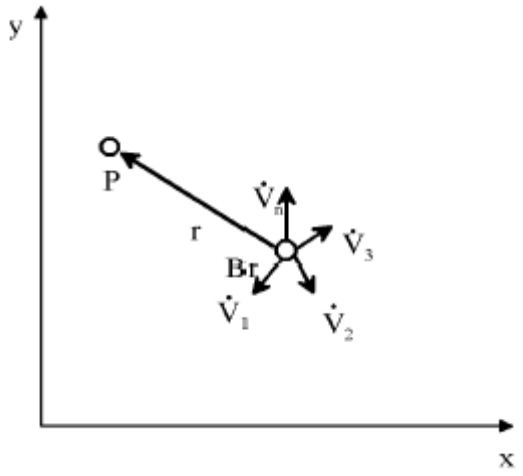


figure 8.6 Virtual conveying flows with time-dependent conveying curve

Subsequently we check the individual fictitious partial conveying capacities in the total drawdown potential and add these accordingly (see figure 8.8):

$$\begin{aligned} Z_{Ges,r,t} &= \sum_{i=1}^{m} Z(r, t_i) \\ &= \sum_{i=1}^{m} \left(\frac{\dot{V}_i}{4\pi k} W_i(\sigma_i) \right) \\ &= \frac{1}{4\pi k} \sum_{i=1}^{m} \dot{V}_i \cdot W_i \left(\frac{r^2 a}{4 t_i} \right) \end{aligned} \qquad (8.28)$$

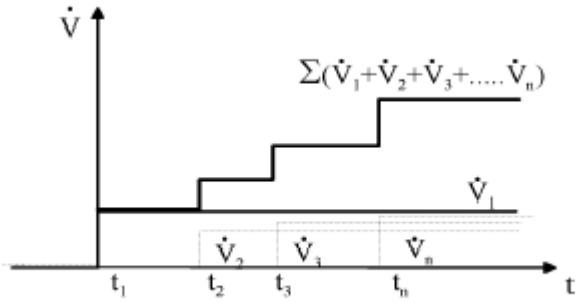


figure 8.7 composite conveying curve

If we introduce the real time *t* and the starting times τ_i of conveying capacity, we get:

$$Z_{Ges,r,t} = \sum_{i=1}^{m} Z(r, t-\tau_i)$$
$$= \frac{1}{4\pi k}\sum_{i=1}^{m} \dot{V}_i \cdot W_i\left(\frac{r^2 a}{4(t-\tau_i)}\right) \qquad (8.29)$$

We can also calculate the partial conveying capacities V*i* from the real conveying capacity at time *t* of $V_{real,i,t}$, by subtracting those time stages from this:

$$\dot{V}_{i,t} = \dot{V}_{real,i,t} - \dot{V}_{real,i-1,t-\tau_i} \qquad (8.30)$$

$$Z_{Ges,r,t} = \frac{1}{4\pi k}\sum_{i=1}^{m}\left(\dot{V}_{real,i,t} - \dot{V}_{real,i-1,t-\tau_i}\right) W_i\left(\frac{r^2 a}{4(t-\tau_i)}\right) \qquad (8.31)$$

Observing this formula it is recognized that the rising phase can be computed. In this case the last partial conveying amount is negative (see figure 8.9). The sign reversal means that this component current is not exfiltration but as treated as infiltration and thus not leads to a drawdown, but an increase of the groundwater level compared with the foregoing time.

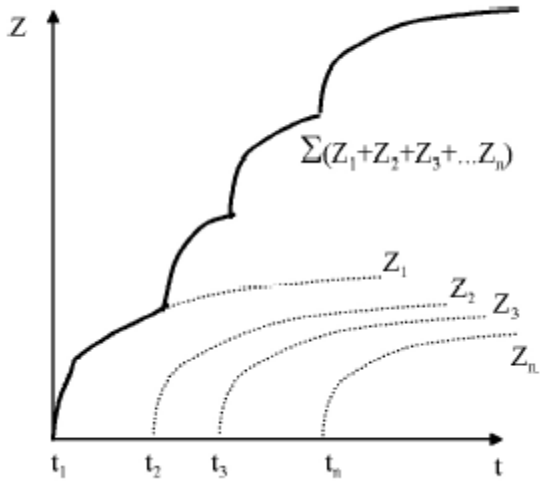


figure 8.8 drawdown potential

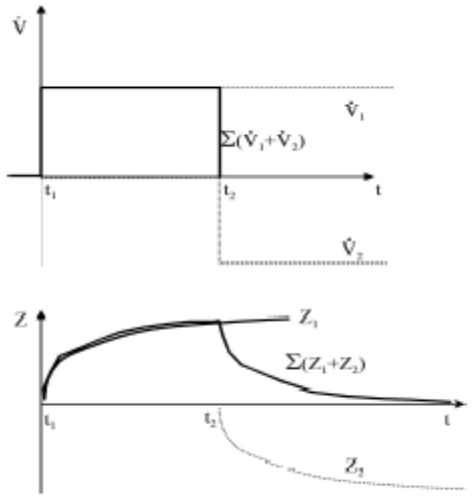


figure 8.9 groundwater rising

The described method for computation of drawdown potentials with temporally conveying curves can be also used, if the conveying curves are not step functions, but as continuous, concave or convex functions. In this case the function will be approximated by a step function (see figure 8.10), whereby step height and width may be not constant. We do not have to assume an equidistant quantization (see section 11.3.5 approximation of signals, page 304). The decision between necessary accuracy and expenditure here is of importance for editors.

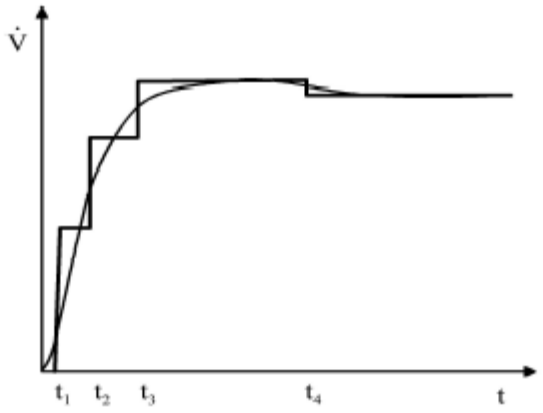


figure 8.10: Approximation of a continuous conveying curve

The methods for computation of multi-well plants and variable conveying curves can be also summarized, so we get a solution for the superposition of both effects:

$$Z_{Ges.r,t} = \frac{1}{4\pi k} \sum_{t=1}^{n} \left(\sum_{j=1}^{m} \dot{V}_{j,t} W_{j,t} \left(\frac{r_i^2 a}{4\,(t - \tau_{j,t})} \right) \right) \qquad (8.32)$$

The drawdown potentials are to be superposed first over all conveying stages of one well and afterwards over all wells.

8.1.2.4 Limitation

- **1st and 2nd type boundary conditions**

The solution of well equation according to THEIS is derived for the infinitely expanded aquifer. For special limitations of aquifer the boundary conditions of THEIS well equation can be modified into a geometrically simple form by means of superposition principle in order to find out a solution.
Basic idea is to arrange a virtual well with drawdown potential in the overlay in such a way that the same hydraulic effect as real well can be exactly obtained like 1st or 2nd boundary condition, i.e. a constant change of potential or a constant influx at the flow limitation. These are special cases, which occur very often in practice.
This method of arranging virtual sources or sinks in a potential field for model building of special boundary conditions is designated as **reflection method** in general potential theory, which applies to many different potential fields (e.g. thermal conduction, electrostatic and magnetic fields). The realization of different kinds of boundary conditions is under consideration of different volume flow directions (ex- or infiltration).

With a limitation in **1st type boundary conditions**, we try to keep the drawdown potential value at zero by means of a virtual infiltration well with same vertical distance l between real well and boundary condition ($Z_{Rand} = 0$) (see figures 8.11 and 8.12).

With 2nd boundary conditions the flow rate at the boundary will be by definition remained constant, here at value zero ($d_{ZRand}/dr = 0$). We can model this with a virtual conveying well, which is axially symmetric and is pressurized with the same conveying capacity (see figures 8.13 and 8.14).

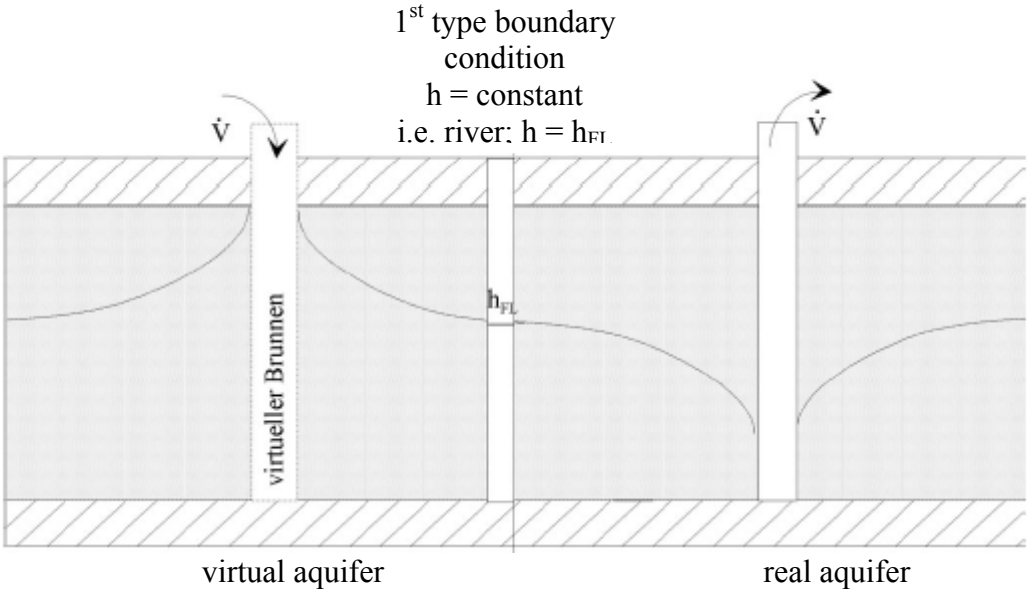


Figure 8.11: 1st type boundary condition modelling by a virtual well

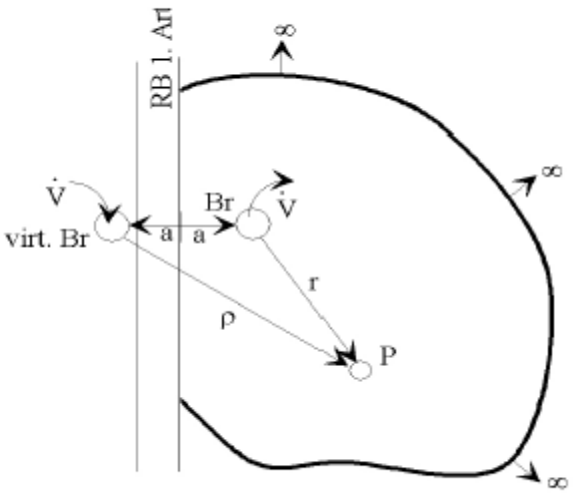


Figure 8.12: consideration of one-sided 1st type boundary condition

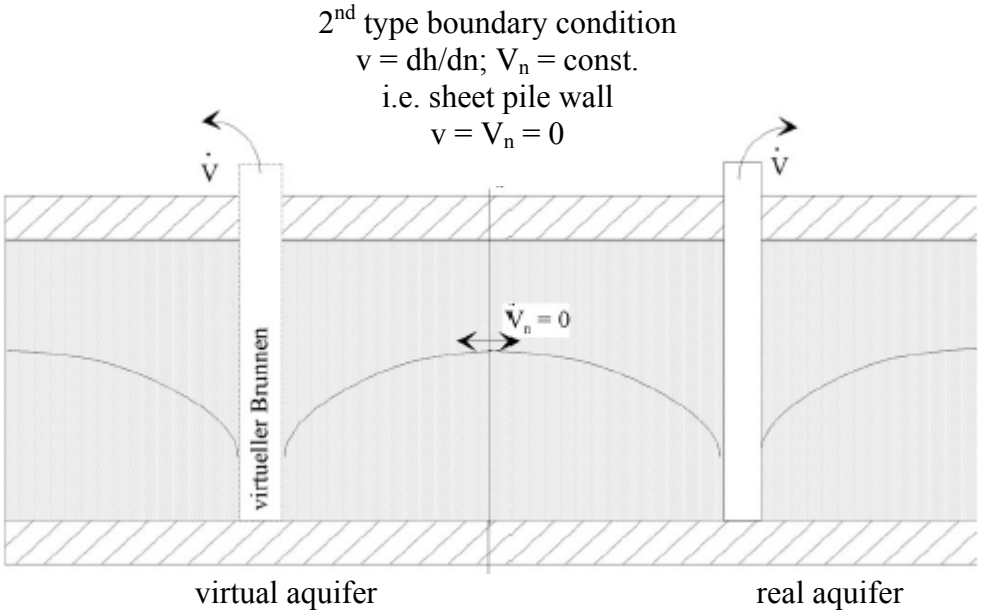


Figure 8.13: 2nd type boundary condition modelling by a virtual well

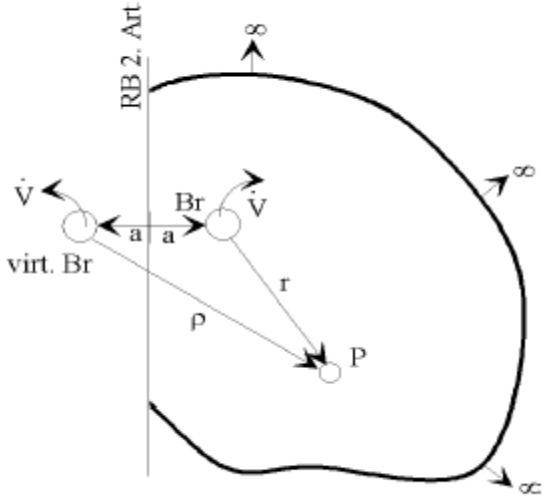


Figure 8.14: consideration of one-sided 2nd type boundary condition

One-sided, linear aquifer limitation and a conveying well at the point $B_r(x,y)$ yield the drawdown potential Z at the computation point $P(x,y)$:

$$Z(r,t) = \frac{\dot{V}}{4\pi k}\left(W_{\text{real}}(r,t) + (-1)^m W_{virt}(\rho,t)\right) \tag{8.33}$$

$$m = \begin{cases} 1^{\text{st}} \text{ type boundary condition} \\ 2^{\text{nd}} \text{ type boundary condition} \end{cases}$$

and the distance from well to computation point:

$$r^2 = (x_{Br} - x_P)^2 + (y_{Br} - y_P)^2 \tag{8.34}$$

or the distance from virtual well to computation point:

$$\rho^2 = (x_{Br\,virt} - x_P)^2 + (y_{Br\,virt} - y_P)^2 \tag{8.35}$$

The superposition must be carried on accordingly in multi-well arrangements or variable the conveying curves. The transformation of drawdown potentials into real drawdown is accomplished according to the arithmetic rules for confined or unconfined aquifer and equations 8.16 and 8.18 (page 201) (see section 7.2 potential illustration, page 187).

Under 1st type boundary conditions a method for computation of the final steady state can be derived from above equation with consideration of COOPER & JACOB approximation:

$$\begin{aligned} Z_{(r,t)} &= \frac{\dot{V}}{4\pi k}\left(W_{real}(r,t) - W_{virtuell}(\rho,t)\right) \\ Z_{stat}(r, t\to\infty) &= \frac{\dot{V}}{4\pi k}\left(-\ln(C\sigma_r) + \ln(C\sigma_\rho)\right) \\ Z_{stat}(r) &= \frac{\dot{V}}{4\pi k}\left(\ln\frac{\sigma_\rho}{\sigma_r}\right) \\ Z_{stat}(r) &= \frac{\dot{V}}{4\pi k}\ln\frac{\rho}{r} \end{aligned} \tag{8.36}$$

We recognize that the final steady state of drawdown potential of the aquifer with one side limited by 1st. type boundary condition proportionally depends on the ratio of conveying capacity to permeability and proportionally on the logarithm of distance.

Taking the same considerations on the steady case of a aquifer with one side limited by 2nd type boundary condition, then we get that the drawdown potential approaches infinitely. This is technically impossible. In the practical operation this means, it takes an infinite time to drain the aquifer circumscribed by sheet pile wall completely.

- **3rd type boundary conditions**

Real boundary conditions are characterized by the fact that due to their effects the definitions and conditions for derivation of mathematical model are not fulfilled. These are imperfection of 1st type boundary condition and additional flow resistances between 1st type boundary condition and the aquifer. Such flow resistances are e.g. colmation layers of water surface. Both effects cause an additional potential decrease between the boundary condition and the drawdown potential point *P*. These effects correspond to 3rd type boundary condition (see section 7.3.2 boundary conditions, page 192). The additional potential decrease depends on the flow quantity, which flows between 1st type boundary condition and the aquifer.

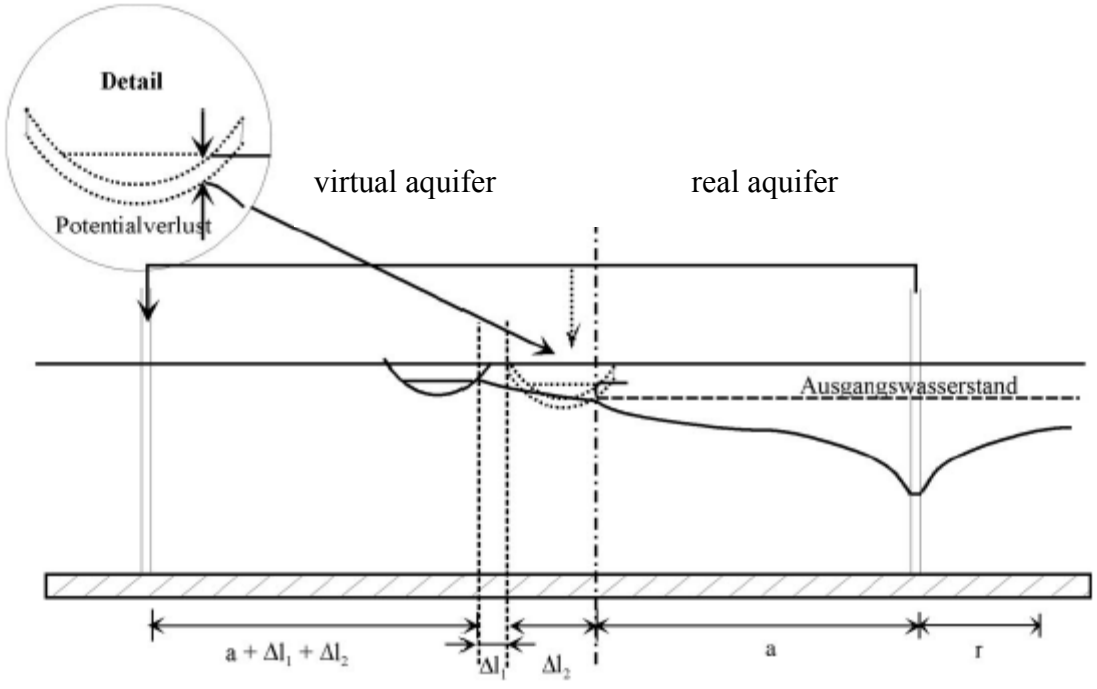


Figure 8.15: consideration of one-sided 3rd type boundary condition

In connection with analytical solution of well equation 3rd type boundary conditions can be solved in such a way, that we find out the equivalent flow resistance of a piece of aquifer, which causes the same potential decrease in ideal 1st type boundary conditions. In the model we shift the real boundary condition a virtual auxiliary length ΔL away from the well (see figure 8.15). Thus the influence of the boundary condition on the drawdown potential is reduced.

We differentiate two kinds of extra lengths by imperfection or by colmation bottom layer.

In the first case of imperfection, i.e. the boundary condition does not extend over the entire through flow thickness, the extra length is determined by the following diagramm (see figure 8.16).

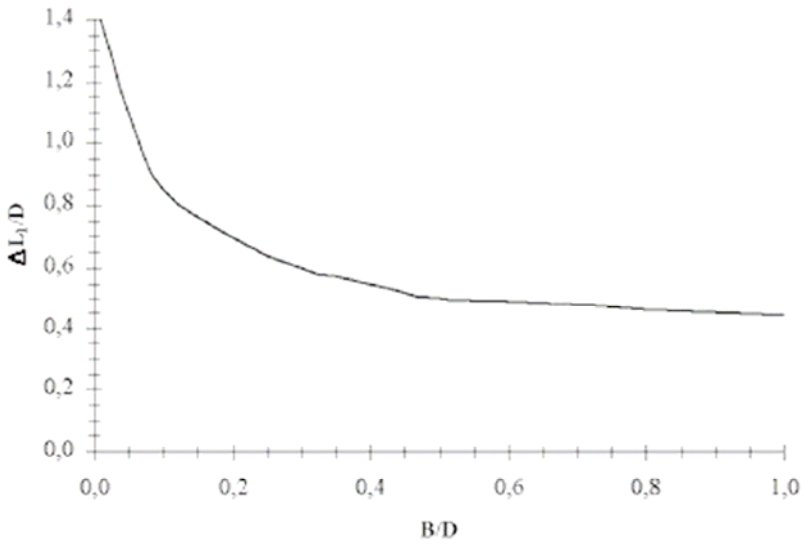


figure 8.16: Dependence of the auxiliary length on the standardized river width

In lakes with 3rd type boundary conditions it can be always assumed that the extra length amounts to:

$$\Delta L_1 = 0,43 \cdot D \tag{8.37}$$

In the colmation bottom layers the length of equivalent aquifer, the same decrease potential caused like the colmation layer can be computed in such way that hydraulic resistances are equated:

$$R_{hydrGWL} = \frac{1}{k} \cdot \frac{\Delta L_2}{D \cdot b} \tag{8.38}$$

$$R_{hydrKOL} = \frac{1}{k'} \cdot \frac{M'}{\Delta L_2 \cdot b} \tag{8.39}$$

According to equation of these two hydraulic resistances the extra length is:

$$\Delta L_2 = \sqrt{\frac{k \cdot D \cdot M'}{k'}} \tag{8.40}$$

with:

D Through flow thickness of aquifer

k' permeability coefficient of colmation layer

k permeability coefficient of aquifer

M' Thickness of colmation layer

8.1.2.5 Multilateral boundary

Besides the linear one-sided boundaries through investigated area until infinite described in the preceding sections, which fulfil the correspondent conditions of THEIS well equation solution, many practical cases are characterized by the fact that the boundary conditions do not have linear process or multilateral boundary conditions rise at the same time. In these cases it is a matter of multilateral limited aquifer systems. These nonlinear boundary condition processes, e.g. the confluence of several receiving streams, will be approximated piecewise by linear boundary conditions. To be noticed that the linear boundary conditions are to be considered with an infinite length.

On the basis of superposition the effect of boundary condition can be computed as additive overlays of the different linear processes. The reflecting method can be here again applied (see figure 8.17). However it must be considered that the virtual wells (reflecting well) at each linear boundary are also to be reflected and lead to further virtual wells. And the combination of different boundary conditions (1^{st} and 2^{nd} type) is possible.

In principle arbitrary angle arrangements between the multilateral boundaries can be considered mathematically based on analytical geometry. Practically however borders are set and only with comfortable computer programs realized (e.g. CAE Groundwater/THEIS). The restriction of perpendicular or parallel standing bounds is for simpler applications. With the perpendicular or parallel bounds the number of repeated reflection will be estimated, i.e. how large the influence of n-*th* reflection well is. Since the drawdown potential is proportional to the $W(\sigma)$ function and $W(\sigma)$ decreases strongly with large σ nascence, the distance between the n-*th* reflection well and the computation point plays a dominative role. It is pointed out that σ grows quadratically with the distance r.

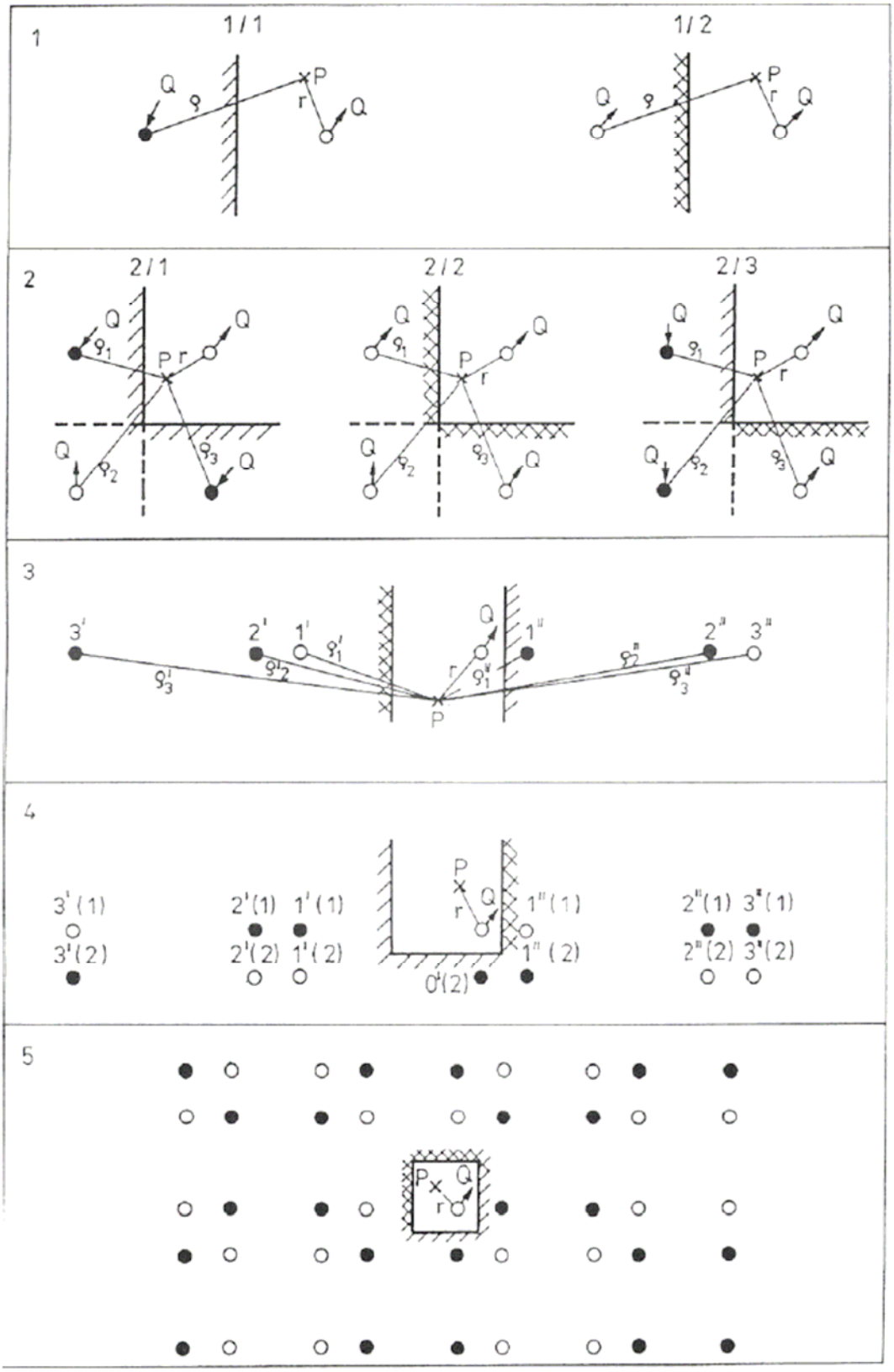


figure 8.17: Repeated boundary conditions with corresponding virtual wells

8.1.3 Supply from neighbouring layers

In former sections a homogeneous aquifer was presupposed. This is however in the rarest cases justifiable. The question, whether a vertical soil profile is to be regarded as homogeneous, laminated or as impervious layer, depends on the variation of the aquifer parameters, the permeability coefficient k and the storage coefficient S (n_0 or $S_0 \cdot D$). In practice the followings are generally accepted. If we consider two abutting soil relationship with the permeability coefficients k_1 and k_2, then the following classification can be carried out according to real precision demand:

$$\frac{k_1}{k_2} = \left\{ \begin{array}{ll} \leq 20 & \text{Homogeneous} \\ \geq 50 \text{ und} \leq 100 & \text{Laminated} \\ \geq 150 & \text{Vertically limited} \end{array} \right\} \text{ aquifer}$$

The first case, the homogeneous aquifer, leads to THEIS well equation solution, the third, the vertically limited, to groundwater storey. In the second case, the laminated aquifer, a supply of the better permeable comes from more badly conducting layer due to larger capillary forces in the layer with larger permeability coefficient (blotting paper effect). This supply was considered by the supply factor B in general well equation. The laminated aquifer is also called **Leakage Aquifer** and the supply factor B is **Leakage factor**. All three cases are transferable into one another and represent only simplified computation possibilities. Furthermore the borders between the computation possibilities are not rigid, but cross over into each other.

This supply factor describes the portion of groundwater regeneration, which results in a potential change in the aquifer and originates from semipermeable layer. These supply rates are thereby calculated under the quasi-stable potential conditions in the semipermeable layer.

Thereby in three cases the spatial arrangements of good and semipermeable layers differ in: the supply above (from hanging), the supply below (from lying) and the combination of both. The aquifer lies between two semipermeable layers. The groundwater level is understood as piezometer head h for the semipermeable layer in all three cases.

The supply factor are computed for the three forms as follows; the aquifer is expressed by thickness M and the permeability coefficient k, the semipermeable layer by thickness M_n and the permeability k_n:

$$B = \begin{cases} \sqrt{\frac{k}{K_o}} M & \text{Supply from layer above} \\ \sqrt{\frac{k}{K_u}} \frac{(h_n + h)}{2} & \text{Supply from layer below} \\ \sqrt{\frac{k}{(K_o + K_u)}} M & \text{Supply from above and below} \end{cases} \qquad (8.41)$$

With: $K_n = \frac{k_n}{M_n}$

The general well equation in polar coordinates:

$$\frac{\partial^2 Z}{\partial r^2} + \frac{1}{r} \cdot \frac{\partial Z}{\partial r} = a \frac{\partial Z}{\partial t} + \frac{Z}{B^2} \qquad (8.42)$$

HANTUSH has found the solution of the drawdown potential:

$$Z(r,t) = \frac{\dot{V}}{4\pi k} W\left(\sigma, \tfrac{r}{B}\right) \quad \sigma = \frac{a r^2}{4t} \qquad (8.43)$$

It is designated also as well function of a semipermeable aquifer, **Leaky aquifer**. Also here the inverse transform from the potential plane for physical dimension water level or drawdown still must be carried out.

The function W(*σ, r/B*) is again a notation short for the exponential integral Ei, here with an extended argument. The derivation of this solution is topic of section 8.1.1 general solution of well equations, page 196.

$$W\left(\sigma, \frac{r}{B}\right) = \int_{\sigma}^{\infty} e^{\left(-\sigma - \frac{r^2}{4B^2\sigma}\right)} \frac{1}{\sigma} d\sigma \qquad (8.44)$$

This function exists as diagram (see figure 8.18). There are simplifications for different ranges of parameters *σ* or *r/B* , so that in practical tasks this complicated formula does not have to be applied:

$$Z(r,t) = \begin{cases} \frac{\dot{V}}{4\pi k} W_{\left(\sigma, \frac{r}{B}\right)} & \text{Generally valid} \\ \frac{\dot{V}}{4\pi k} W_{(\sigma)} & \sigma > 2r / B \\ \frac{\dot{V}}{2\pi k} K_{o\left(\frac{r}{B}\right)} & \text{long time} \\ \frac{\dot{V}}{2\pi k} \ln\left(1, 12\frac{B}{r}\right) & \text{long time and } r < 0.03\ B \end{cases} \qquad (8.45)$$

In the application of this solution for the Leaky aquifer it must be noted that the supply is set as constant value and thus steady groundwater flow conditions in less permeable aquifer, semipermeable aquifer are presupposed. The storage capacities of these layers are neglected.

The error is not too large for the lying, since in such cases confined groundwater conditions are predominant there, which possess smaller storage effects. In hanging also only a small error occurs under confined conditions. Larger errors can come up if free groundwater surface exists in the lying. Particularly in the evaluation of pumping tests this simplified assumption emerges as not acceptable. In this case the supply factor must be increased empirically.

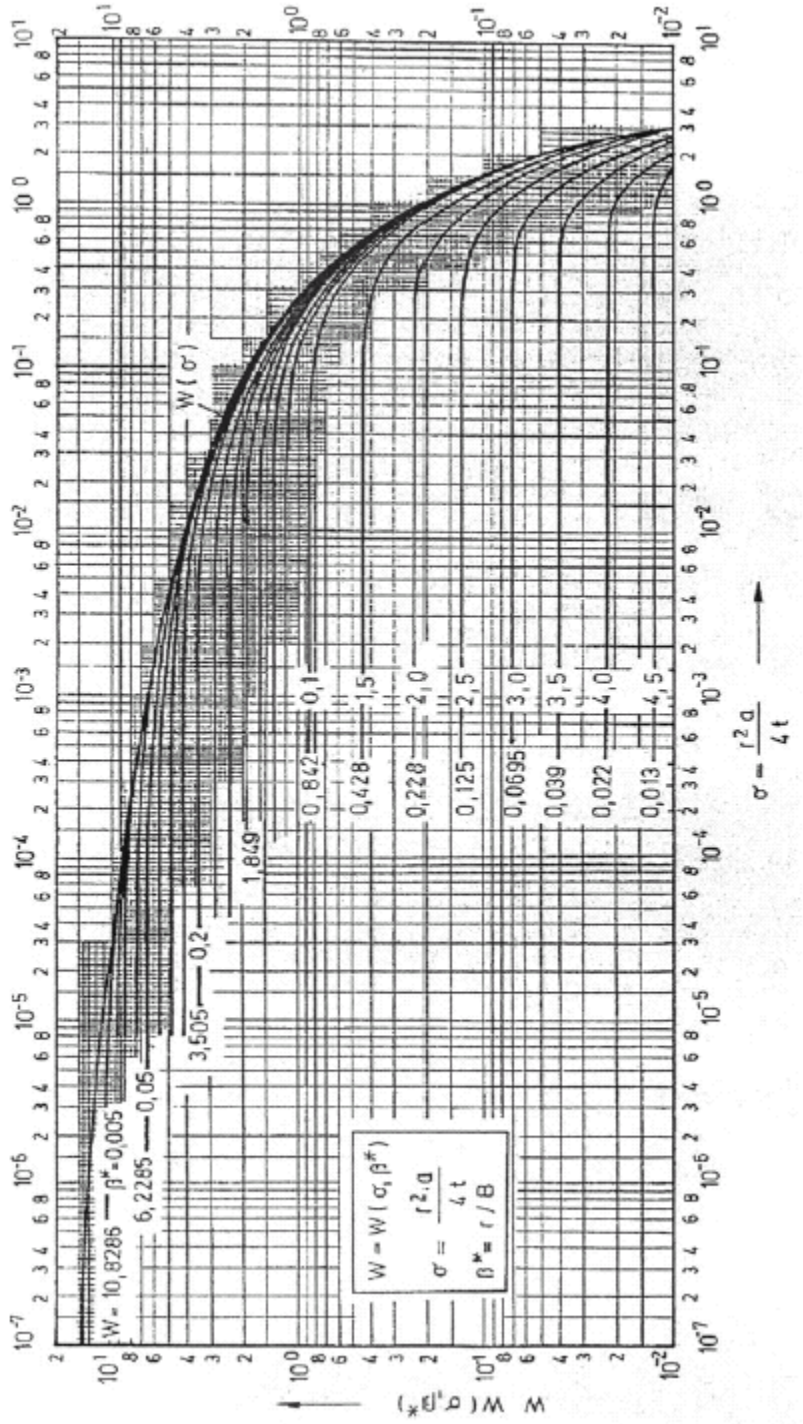


Figure 8.18: function r/B dependent on σ

8.2 Tasks of analytical calculation

1. Compute the drawdown *s* for the groundwater observation tubes (GWOT) with distance r and at time t, which results from water conveying in the well for consecutively infinitely expanded aquifer (see figure 8.19) and state the result graphically.

$k = 1 \cdot 10^{-3} \frac{m}{s}$; $M = 10m$; $S = 0,001$; $a = \frac{S}{T} = 0,1 \frac{s}{m^2}$; $r_0 = 0,25m$; $\dot{V} = 0,015 \frac{m^3}{s}$;
$h_n = 16m$;
$r = 5m$; $10m$; $20m$; $50m$
$t = 1min$; $2min$; $5min$; $10min$; $20min$; $30min$; $45min$; $60min$; $90min$; $120min$

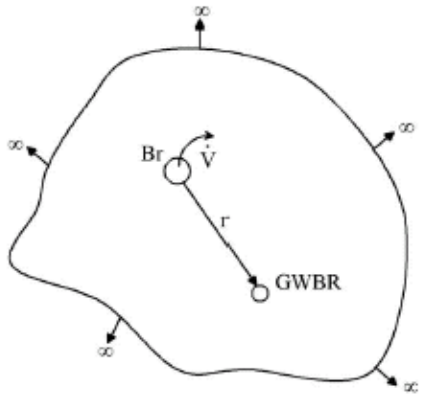


Figure 8.19: infinitely expanded aquifer with well and GWOT

2. Compute the drawdown in the GWOT (r = 10m) for the aquifer from task 1 (see figure 8.19) every 10 minutes until maximally 100 minutes, if that flow rate of conveying well is subject to following stagger time. And plot the solution.

Volumenstrom $\left[\frac{m^3}{s}\right]$	0,005	0,010	0,015	0,020	0,025	0,030	0,000
Förderbeginn [min]	0	10	20	30	40	50	60

3. A foundation pit should be lowered in an aquifer near a river. The centre of the foundation pit is 100m far away from the river; the drainage well is 80m. Three wells are arranged parallel to the river, which are 25m distant from each other. The diameter of wells r_0 = 0.3m and conveying capacity is V = 0.015m^3/s
The width of the river is B = 20m and a colmation layer k' = 3 · 10^{-6}m/s; M' = 1m. (see figure 8.20)

The properties of the aquifer:

$$k = 5 \cdot 10^{-4} \frac{m}{s}; n_0 = 0,20; h_n = 15m; M = 20m.$$

Will a drawdown of 2.5m be achieved in 10 days in the centre of the foundation pit?

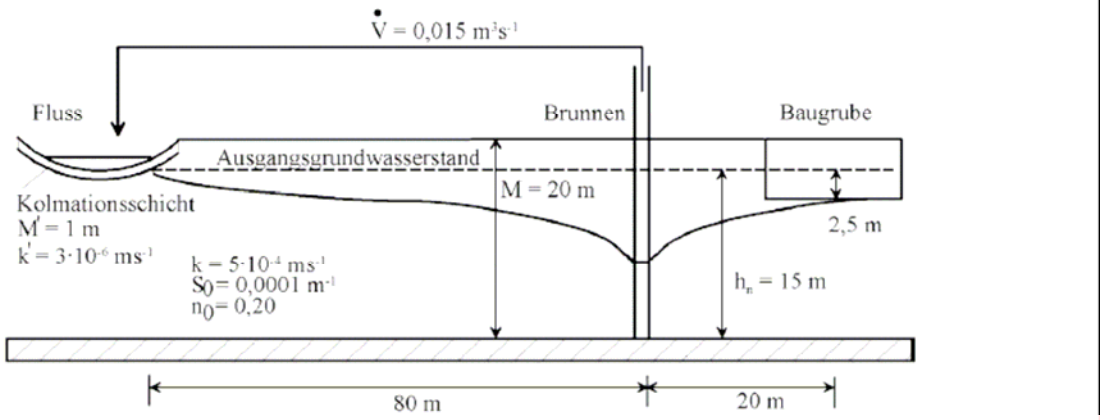


figure 8.20: aquifer with imperfect river, well and foundation pit

4. Please apply analytical method of well flow to check whether the centre of the foundation pit is drained after 7 days with conveying capacity of V = 0.01m^3/s, r_0 = 0.30m and a security of 0.5m (see figure 8.21).

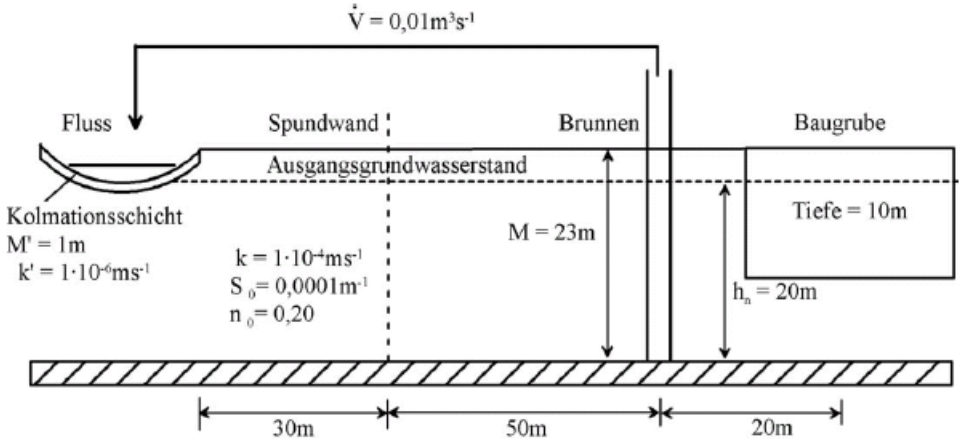


figure 8.21: aquifer with well and foundation pit

5. In a pumping test in an infinitely expanded aquifer the following water levels are measured as a function of the distance to the well after pumping 120min (see figure 8.22).

Compute the water deficit (volume) of the drawdown funnel, if the aquifer are of following characteristic values:

$$h_n = 16m,\ M = 10m,\ k = 0,001\frac{m}{s},\ S_0 = 0,0001m^{-1},\ n_0 = 0,20.$$

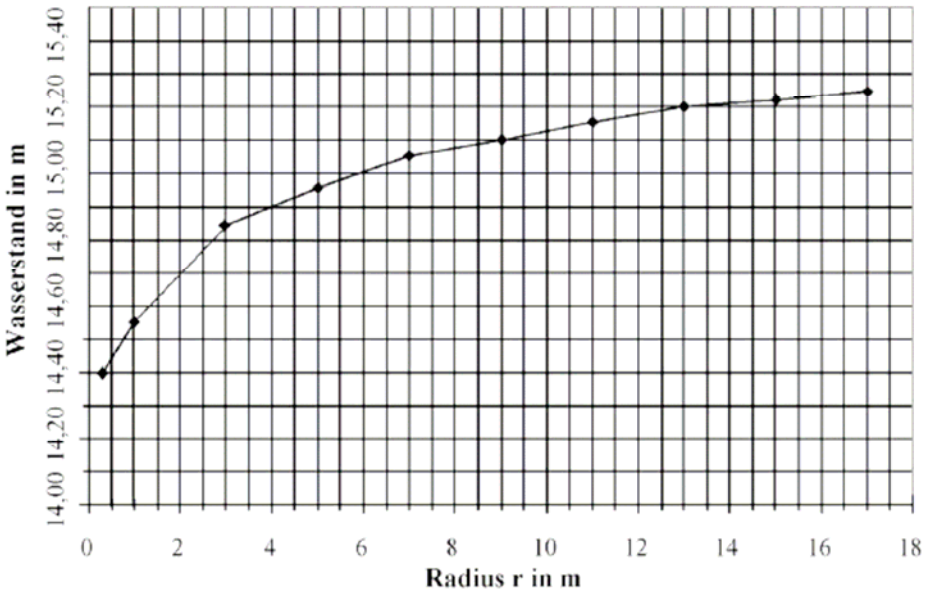


figure 8.22: groundwater level dependent on radius

6. A constant flow rate of 25 l/s is conveyed from a well, which connects an ideal river (x = 0; -∞ < y < + ∞) ($Br_{(100m,500m)}$). The well has a radius of r_0 = 0.35m. The aquifer is characterized by the following parameters:

$$h_n = 15m,\ M = 17m,\ k = 1 \cdot 10^{-3}\frac{m}{s},\ S_0 = 0,0002m^{-1},\ n_0 = 0,25.$$

a) Calculate the final steady state (the portion of temporal functionality should be smaller than 0.001) for the point ($P_{(200m,600m)}$) and
b) the time point, from when to calculate
   Tips: Work as long as possible with general symbols.

7. A simulation system is to be developed for a induced recharge water works (see figure 8.23) with parallel flow regime. The river should be considered as idealized boundary condition. Compute the final steady state under these hydraulic conditions based on the analytical well equation solution.

$$k = 1 \cdot 10^{-3}\frac{m}{s};\ h_{Fl} = 15m;\ z_{R0} = 15m;\ S = 0,25;\ \dot{V} = 50\frac{l}{s};\ q = 0,001\frac{l}{s \cdot m^2}l;$$
$$b = 100m;\ k_{Kolm} = 5 \cdot 10^{-5}\frac{m}{s};\ M_{Kolm} = 1m;$$

Establish the solution

a) with idealized river
b) with consideration of real river (colmation and imperfection)

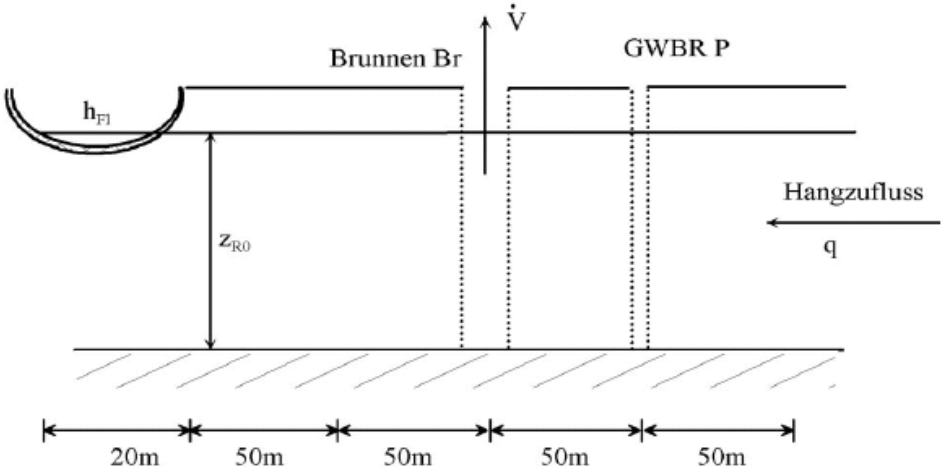


figure 8.23: aquifer with river, well and influx

8. In geohydraulics pumping tests are used for the determination of the aquifer parameters. Under certain conditions the drawdown can be determined according to the THEISS/JAKOB/COOPER formula.

$$s = \frac{\dot{V}}{4 \cdot \pi T} \ln\left(\frac{2,25 \cdot T \cdot t}{r^2 \cdot S}\right)$$

By using this formula deduct an equation to determine k-value for a local point P, which is with a distance r away from well. The determination of the k-value is based on the use of drawdown value s_1 at the time t_1 and s_2 at the time t_2. The relation of measurement period t_1: t_2 amounts to 1: 10.

9. Compute the drawdown in the GWOT (see figure 8.24) for the time point = 10h, if a flow rate of 0.2m^3/s is conveyed for 5h in the well and afterwards the pumps were switched off.

$$h_{t=0} = 10m,\ M = 15m,\ k = 0,0001\tfrac{m}{s},\ S_0 = 0,0001m^{-1},\ n_0 = 0,25$$

10. Compute the drawdown at point P after one year based on induced recharge for a groundwater recovery plant.

Conveying rate of each well: 25 l·s^{-1}

$$k = 2 \cdot 10^{-3} m \cdot s^{-1};\ h_n = 15m;\ S = 0,25$$
$$k' = 1 \cdot 10^{-5} m \cdot s^{-1};\ M' = 1m;\ B = 25m;\ r_1 = 250m;\ r_2 = 500m;\ r_P = 375m$$

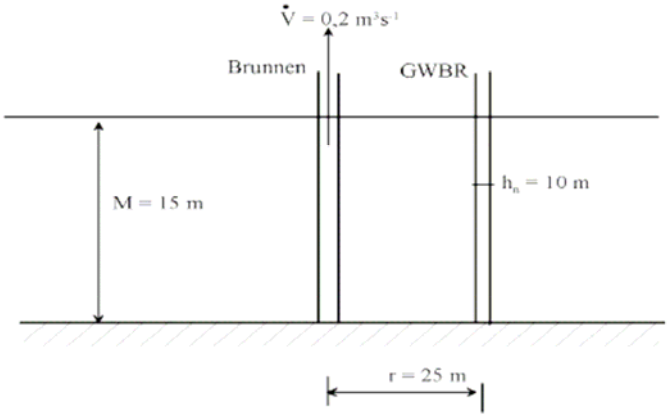


figure 8.24: Infinitely expanded aquifer with a conveying well

11. From two wells to a river (without colmation and perfect) a constant flow 25 l·s^{-1} is conveyed. Compute the drawdown at the point P after one month and the final steady state.

Coordinates:

Well 1: x = 750m y = 100m
Well 2: x = 700m y = 400m
Point P: x = 1000m y = 500m

h_n = 15m, n_0 = 0.25, k = 10^{-3}m·s^{-1}

12. Calculate the change of groundwater level after 10 days for the following schematic groundwater plant by means of THEIS well equation (see figure 8.25). Given:

$$\dot{V} = 0,001m^3 \cdot s^{-1},\ S_0 = 0,001m^{-1},\ n_0 = 0,25,\ k = 0,001m \cdot s^{-1}$$

13. Calculate the change of groundwater condition after 10 days for the following schematic groundwater plant by means of THEIS well equation (see figure 8.26). Given:

$$\dot{V} = 0,001m^3 \cdot s^{-1},\ S_0 = 0,001m^{-1},\ n_0 = 0,25,\ k = 0,001m \cdot s^{-1}$$

14. Calculate the drawdown at the gauge for time point t = 15h, if a flow rate of 0.1m^3/s is conveyed for 10h in the well and afterwards the pumps were switched off (see figure 8.27). Given:

$$h_{t=0} = 40m,\ M = 50m,\ k = 0,0001ms^{-1},\ S_0 = 0,0001m^{-1},\ n_0 = 0,25$$

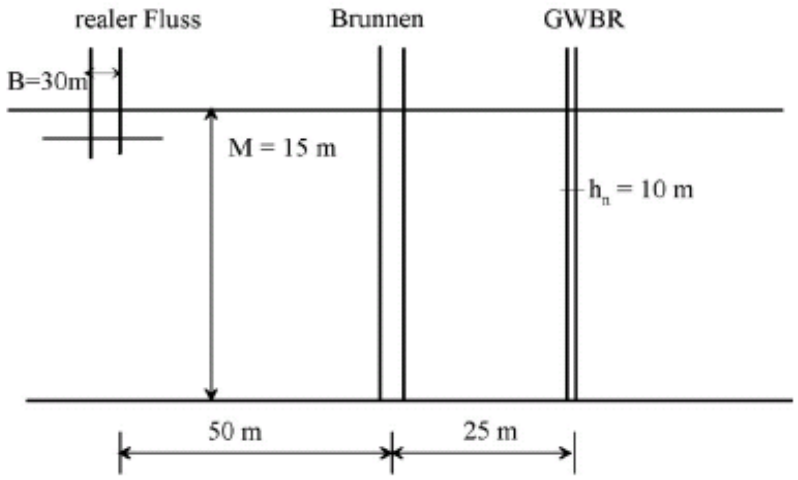


Figure 8.25: real river with pumping well (artificial groundwater recharge plant)

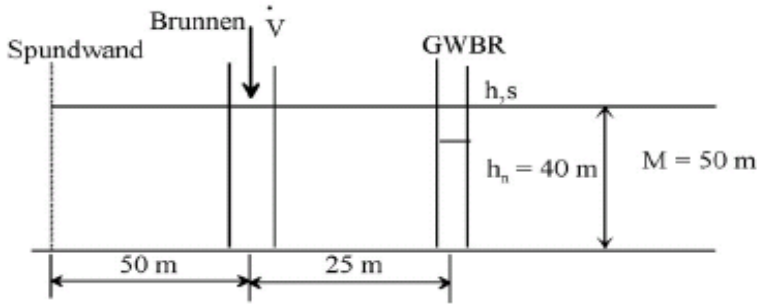


Figure 8.26: groundwater plant with well and sheet pile wall

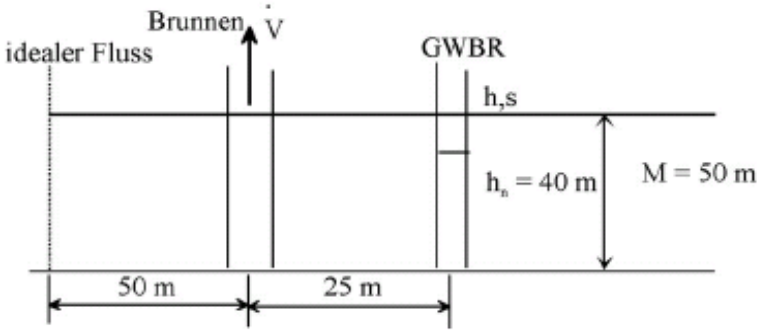


figure 8.27: induced recharge plant with well and river

Chapter 9

9 Numerical method

The horizontal plane groundwater flow equation in general form of a nonlinear partial differential equation is not completely solvable. By local and the time coordinates quantization a numerically solvable discontinuous model can be set up. In the literature the term **discretisation** is also often used instead of quantization, and the discontinuous model is called **discrete model**. This misuse of the words in the literature is caused by careless separation of independent and dependent variables. The **quantization of an independent variable** leads to the discontinuous, while the **quantization of a dependent variable** leads to the discrete model.

The discontinuous models can be simply adapted to the structures of data models, which are also existent in the form of samples generally.

The quantization of local variables should first be regarded independent of time variable. This is justifiable in any case of steady processes, but in unsteady procedures these classes of the variables can be also regarded as independent of each other. Only in special interpolation scheme some connections of these two quantization procedures come up and must be treated separately.

We assume the continuous function of the piezometer head or the position of free groundwater surface in the original ($z_R(x,y,t)$), then we get a discontinuous function ($z_R(xi,yi,ti)$) by the discontinuous simulation. Subsequently, the problem editor will try again to approximate a continuous function from it. Following demands result from setting up tasks for the execution of quantization:

- No information loss appears in the quantization of function, since otherwise the continuous cannot be retrieved one to one from the discontinuous function.

- No redundant data processing is caused by quantization, i.e. the distance of the tactile point is not too small to select.

Further demands concerning quantization result from the used simulator and the existing input data:

- the quantized field is designed in such a way that the hydro geological and technical/ technological conditions of the original can be clearly, physically descriptively and with high accuracy taken into consideration.

- quantization must allow a simple simulation.

The demands are contradictory to some extent. Above all the demands of the theoretical information side contradict practical application. Thus a search of an optimal organization of tactile points will be also carried out in quantization.

9.1 Methods of local quantization

The quantized values of the x- and y- axis yield tactile points in the x-y plane, which can be distributed arbitrarily. These points are connected with straight lines, and we get a reticular structure. The tactile point function is from nodes. The function values at the tactile points or the nodes are supposed to be known as possible solution of the simulation so that the values can be interpolated along the net lines and within the mesh. If the function value (e.g. water level, temperature, concentration) is represented as Z-axis, then we get a three-dimensional plane, which is supported by the given function values at the knots. With respect to practical simulation we differentiate various network configurations, which can be divided according to the following criteria (see table 9.1 and figure 9.1).

Table 9.1: introduction of network configurations

Coordinate reference	**coordinate true**	**coordinate independent**
Quantization	equidistant	arbitrary quantized
Topology	regular	irregular
Geometry	orthogonal	triangular

The different network configurations possess advantages and disadvantage concerning the fulfilment of initial demands. The regular network configurations allow a relatively simple subsequent treatment by means of simulators. However they poorly fulfil the demand that hydro geological conditions should be considered under minimum simulation expenditure. In addition a redundant data processing cannot be avoided due to small increments. The irregular network configurations, particularly the coordinate-independent, can not be well simulated. The network configuration can be however well adapted to the hydro geological and technical/technological conditions by the arbitrary distribution of the nodes. The substantial advantage in the arbitrary node density distribution is that, they can lie as at close quarters or as far away as required. Thus a minimized redundancy of data processing is realized with minimum number of nodes. The disadvantage of the irregular network configurations is a complicated execution of the simulation. We can make a compromise if we transfer the arbitrary network configurations into topologically regular, but geometrically irregular triangle nets. the topological organization of the net is crucial for the execution of the simulation. There are always six connections from a net point to neighbouring points in the topologically regular triangle net. Finally lots of connections exist to neighbouring knots in a arbitrary triangle net.

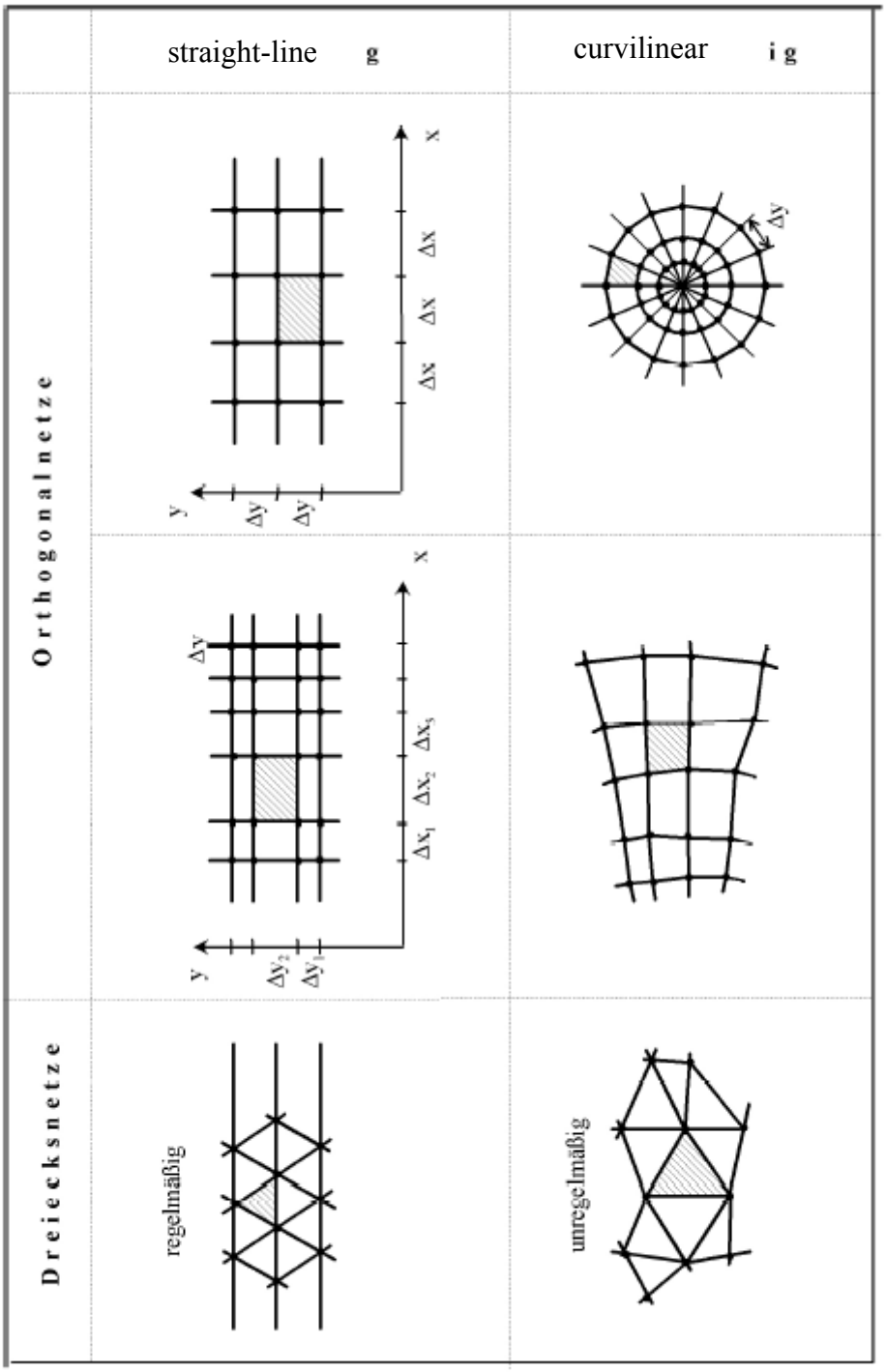


figure 9.1: network configurations

Independent of the organization of network configuration a further effect appears in the modelling of hydraulic areas by means of numerical models. The generally infinite flow conditions in aquifer are artificially limited by the finite expansion of mathematical models. The edges of model belong to 2^{nd} type boundary conditions ($\partial h/\partial n$ = const.). Thus at the edge of the model a volume flow $V_{Rand} = 0$ is outwards forced. And the barriers of model edges are set to be equal. This is often contrast to the hydraulic conditions of original process. This effect is recognizable from the fact that the equipotential lines, e.g. water level, the so called isolines, which in principle stand perpendicularly at the model edge. We can minimize this error in order to apply natural 1^{st} and 3^{rd} boundary conditions to the model boundary such as rivers and lakes, as well as considering possible colmation effects. With know hanging influx or phreatic divide, in particular with groundwater basin borders, we can input these at the model border as 2^{nd} boundary condition.

There are different ways to describe the geohydraulic behaviour within the mesh. The most known ones are **method of finite differences (FDM)**, **method of finite volumes (FVM)** and **method of finite elements (FEM)**. In FDM it is assumed that the exchange takes place along the mesh borders and in the nodes. In contrast the basic idea of FEM consists in the fact that the network mesh is regarded as continuum and the reciprocal effect with the neighbour elements is achieved by energy-, impulse- and mass exchange perpendicularly through mesh bound. In the following FDM will be described in detail. The other quantization procedures such as FEM and FVM are only roughly represented.

9.1.1 Finite Difference Method

As previously mentioned, no time-dependent procedures are considered in local discretisation. Therefore the derivation of local quantizations is independent of whether an unsteady or steady flow field. On this account the following assertions are derived on the basis of steady flow equation. They can be also applied to the unsteady flow regime under consideration of the mathematical conditions (see section 9.2 time quantization, page 258).

9.1.1.1 Balance equation

The differential equation of the horizontal planes groundwater flow in steady case:

$$\operatorname{div}(T \cdot \operatorname{grad} h) = 0 \tag{9.1}$$

$$\text{mit} \quad h = \left\{ \begin{array}{ll} h & \text{Confined} \\ z_R & \text{Unconfined} \end{array} \right\} \text{aquifer}$$

$$\text{und} \quad T = \left\{ \begin{array}{ll} k(x,y,z)M & \text{Confined} \\ k(x,y,z)z_R & \text{Unconfine} \end{array} \right. = \int_{z=a}^{D} k(x,y,z)dz$$

The steady flow is characterized by the fact that no storage procedures appear. This partial differential equation can be written in cartesian coordinates as follows:

$$\frac{\partial}{\partial x}\left(T\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial h}{\partial y}\right) = 0 \tag{9.2}$$

If the finite difference method is applied to this equation with reference to a coordinate, topologically regular grid, then this is equated with transfer of derivatives in difference quotients:

$$\frac{\Delta}{\Delta x}\left(T\frac{\Delta h}{\Delta x}\right) + \frac{\Delta}{\Delta y}\left(T\frac{\Delta h}{\Delta y}\right) = 0 \tag{9.3}$$

The quantization error, which occurs during this transition, has a quadratic order ($O(x, y) \sim \Delta x^2$, Δy^2). We can also describe this error as deviation of secant, which is used in the difference quotient, and the tangent, which is defined by the derivative. This deviation is dependent on the gradients, i.e. the slope of the tangent.

Apart from this mathematically justifiable error still the following error phenomenon can be enumerated. In the quantization the geometrical position at surface will be changed by arbitrarily arranged boundary conditions, since it can be only arranged at the net points, the crossings between row and column in discontinuous net (see figure 9.2).

initial state – continuum discontinuous model

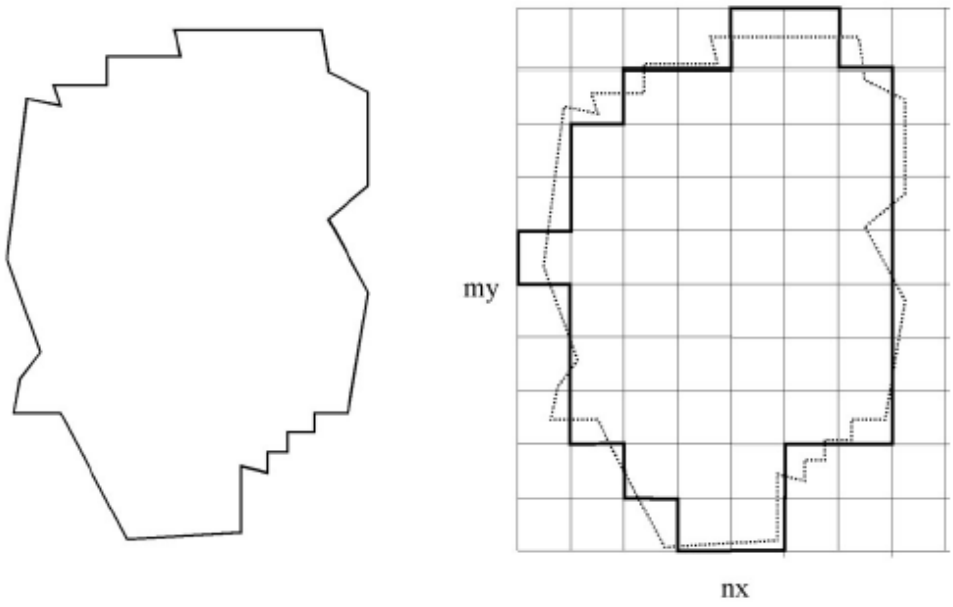


figure 9.2: transition between continuum and resistance network

Also the geometrical size of the boundary conditions is changed. In the continuum each boundary condition has an arbitrary finite geometrical expansion. In the discontinuous field each boundary condition can accept only one expansion, which is an integral multiple of the quantization increment Δx and Δy. Generally this means that a substantially larger effect area is arranged in the discontinuous simulated network and the original of this case is the boundary condition. In this relation the effect of the finite expanded net is again pointed out. The net edges lead to a limitation of flow field, since no flow rate flows at the borders ($V_{Rand} = 0$). In original however it does not have to be so. For this reason the edge of model should agree with the position of the hydraulic boundary conditions if possible.

Apart from the mathematical derivative the quantization procedure can be also physically justified. For this purpose the continuum is covered with an appropriate mesh. At the individual nodes the balance equations, here the mass balance equations, are set up. The connections are represented by flow resistances between the net knots. The mass balance equation at the point $P_{n,m}$ results from the sum of water quantities, which flow along the net lines (see figure 9.3). The sum must be equal to zero according to definition (under steady condition steps no storage effect). Thus the balance equation:

$$\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} = 0 \qquad (9.4)$$

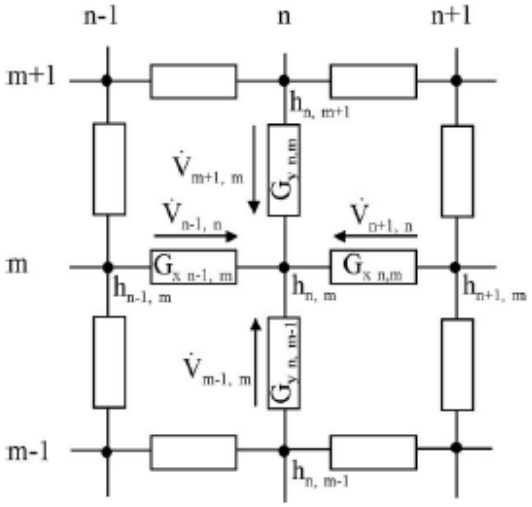


figure 9.3: balance for knot n, m

The flow rates can also be represented as quotient of differences of the water level or pressure and hydraulic flow resistances or as product of differences of the water level or pressure and hydraulic conductivity (general DARCY law, also see theory of pipe- and channel hydraulics), and we get:

$$\begin{aligned} &G_{x\,n-1,m}(h_{n,m} - h_{n-1,m}) + G_{x\,n,m}(h_{n,m} - h_{n+1,m}) + \\ &G_{y\,n,m-1}(h_{n,m} - h_{n,m-1}) + G_{y\,n,m}(h_{n,m} - h_{n,m+1}) = 0 \end{aligned} \qquad (9.5)$$

or arranged according to water height:

$$\begin{aligned} &h_{n-1,m}(-G_{x\,n-1,m}) + \\ &h_{n,m-1}(-G_{y\,n,m-1}) + \\ &h_{n,m}(G_{x\,n-1,m} + G_{x\,n,m} + G_{y\,n,m-1} + G_{y\,n,m}) + \\ &h_{n,m+1}(-G_{y\,n,m}) + \\ &h_{n+1,m}(-G_{x\,n,m}) = 0 \end{aligned} \qquad (9.6)$$

If we go through the grid net in columns, i.e. the knots of row1 to my are processed within the columns 1 to nx, then it yields an equation system with nx · my rows and with nx · my unknown quantities. This high dimensional equation system, in practical until several hundred thousand or

millions, can be solved according to GAUSS total step iteration, GAUSS SEIDEL or by means of iterative methods, in particular the method of Preconditioned Conjugate Gradients (PCG) (see section 1.3 linear equation system, page17 and the following).

If we arrange these equations according to the elements $h_{n,m}$ we get an equation system (see table 9.3), which has the characteristic diagonal shape. If we represent this as matrix equation, we get the coefficient matrix G with nx · my columns and just as many as rows. This coefficient matrix is only located in certain places, namely in the diagonals, the two directly bordering second diagonals and two further second diagonals, which is nx distance from main diagonal. All other elements of the matrix are equal to zero. Since this matrix is still symmetrical, the actual significant value space reduces to 3 · nx · my elements (see section 1.2.1 band matrix, page 8). The solution function, i.e. the water levels $h_{n,m}$ at the knots, is represented by a column vector nx · my. In contrast on the right side a column vector nx · my stands likewise.

An example of the quantization of a two-dimensional aquifer by means of a net with four columns and four rows (see figure 9.4) as well as the equation system will be demonstrated.

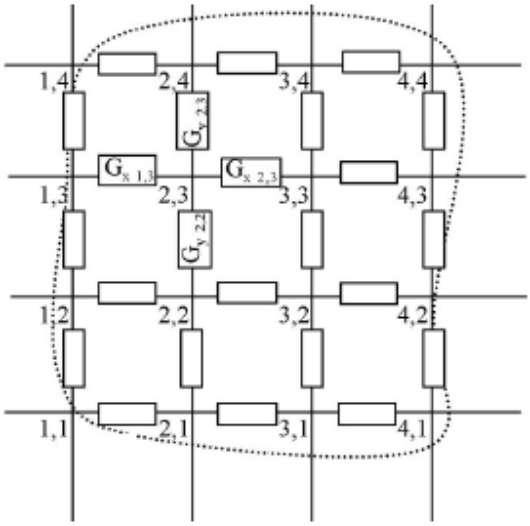


figure 9.4: 4 x 4 grid net

This example yields following 16 equations (see table 9.2)

Table 9.2: equation system for two dimensional aquifer quantization

Spalte	Nr.	Knotenpunktgleichung				
n	**n,m**	$G_{x\,n-1,m}(h_{n,m}-h_{n-1,m})+$	$G_{y\,n,m-1}(h_{n,m}-h_{n,m-1})+$	$G_{x\,n,m}(h_{n,m}-h_{n+1,m})+$	$G_{y\,n,m}(h_{n,m}-h_{n,m+1})$	$=0$
	1,1			$G_{x\,1,1}(h_{1,1}-h_{2,1})+$	$G_{y\,1,1}(h_{1,1}-h_{1,2})$	$=0$
1	**1,2**		$G_{y1,1}(h_{1,2}-h_{1,1})+$	$G_{x\,1,2}(h_{1,2}-h_{2,2})+$	$G_{y\,1,2}(h_{1,2}-h_{1,3})$	$=0$
	1,3		$G_{y\,1,2}(h_{1,3}-h_{1,2})+$	$G_{x\,1,3}(h_{1,3}-h_{2,3})+$	$G_{y\,1,3}(h_{1,3}-h)$	$=0$
	1,4		$G_{y\,1,3}(h_{1,4}-h_{1,3})+$	$G_{x\,1,4}(h_{1,4}-h_{2,4})$		$=0$
	2,1	$G_{x\,1,1}(h_{2,1}-h_{1,1})+$		$G_{x\,2,1}(h_{2,1}-h_{3,1})+$	$G_{y\,2,1}(h_{2,1}-h_{2,2})$	$=0$
2	**2,2**	$G_{x\,1,2}(h_{2,2}-h_{1,2})+$	$G_{y2,1}(h_{2,2}-h_{2,1})+$	$G_{x\,2,2}(h_{2,2}-h_{3,2})+$	$G_{y\,2,2}(h_{2,2}-h_{2,3})$	$=0$
	2,3	$G_{x\,1,3}(h_{2,3}-h_{1,3})+$	$G_{y\,2,2}(h_{2,3}-h_{2,2})+$	$G_{x\,2,3}(h_{2,3}-h_{3,3})+$	$G_{y\,2,3}(h_{2,3}-h_{2,4})$	$=0$
	2,4	$G_{x\,1,4}(h_{2,4}-h_{1,4})+$	$G_{y\,2,3}(h_{2,4}-h_{2,3})+$	$G_{x\,2,4}(h_{2,4}-h_{3,4})+$		$=0$
	3,1	$G_{x\,2,1}(h_{3,1}-h_{2,1})+$		$G_{x\,3,1}(h_{3,1}-h_{4,1})+$	$G_{y\,3,1}(h_{3,1}-h_{3,2})$	$=0$
3	**3,2**	$G_{x\,2,2}(h_{3,2}-h_{2,2})+$	$G_{y\,3,1}(h_{3,2}-h_{3,1})+$	$G_{x\,3,2}(h_{3,2}-h_{4,2})+$	$G_{y\,3,2}(h_{3,2}-h_{3,3})$	$=0$
	3,3	$G_{x\,2,3}(h_{3,3}-h_{2,3})+$	$G_{y\,3,2}(h_{3,3}-h_{3,2})+$	$G_{x\,3,3}(h_{3,3}-h_{4,3})+$	$G_{y\,3,3}(h_{3,3}-h_{3,4})$	$=0$
	3,4	$G_{x\,2,4}(h_{3,4}-h_{2,4})+$	$G_{y\,3,3}(h_{3,4}-h_{3,3})+$	$G_{x\,3,4}(h_{3,4}-h_{4,4})+$		$=0$
	4,1	$G_{x\,3,1}(h_{4,1}-h_{3,1})+$			$G_{y\,4,1}(h_{4,1}-h_{4,2})$	$=0$
4	**4,2**	$G_{x\,3,2}(h_{4,2}-h_{3,2})+$	$G_{y\,4,1}(h_{42}-h_{4,1})+$		$G_{y\,42}(h_{4,2}-h_{4,3})$	$=0$
	4,3	$G_{x\,3,3}(h_{4,3}-h_{3,3})+$	$G_{y4,2}(h_{4,3}-h_{4,2})+$		$G_{y\,4,3}(h_{4,3}-h_{4,4}$	$=0$
	4,4	$G_{x\,3,4}(h_{4,4}-h_{3,4})+$	$G_{y\,4,3}(h_{4,4}-h_{4,3})+$			$=0$

Table 9.3: equation system in diagonal shape

		Zeile 1			
Spalte	**Nr.**	$h_{1,1}$	$h_{1,2}$	$h_{1,3}$	$h_{1,4}$
1	**1,1**	$\begin{pmatrix} G_{x\,1,1} \\ +G_{y\,1,1} \end{pmatrix}$	$-G_{y\,1,1}$		
	1,2	$-G_{y\,1,1}$	$\begin{pmatrix} G_{y\,1,1} \\ +G_{x\,1,2} \\ +G_{y\,1,2} \end{pmatrix}$	$-G_{y\,1,2}$	
	1,3		$-G_{y\,1,2}$	$\begin{pmatrix} G_{y\,1,2} \\ +G_{x\,1,3} \\ +G_{y\,1,3} \end{pmatrix}$	$-G_{y\,1,3}$
	1,4			$-G_{y\,1,3}$	$\begin{pmatrix} G_{y\,1,3} \\ +G_{x\,1,4} \end{pmatrix}$
2	**2,1**	$-G_{x\,1,1}$			
	2,2		$-G_{x\,1,2}$		
	2,3			$-G_{y\,1,3}$	
	2,4				$-G_{y\,1,4}$
3	**3,1**				
	3,2				
	3,3				
	3,4				
4	**4,1**				
	4,2				
	4,3				
	4,4				

Table 9.4: continuation 1

Spalte	Nr.	Zeile 2			
		$h_{2,1}$.	$h_{2,2}$.	$h_{2,3}$.	$h_{2,4}$.
1	1,1	$-G_{x\,1,1}$			
	1,2		$-G_{x\,1,2}$		
	1,3			$-G_{x\,1,3}$	
	1,4				$-G_{x\,1,4}$
2	2,1	$\begin{pmatrix} G_{x\,1,1} \\ +G_{x\,2,1} \\ +G_{y\,2,1} \end{pmatrix}$	$G_{y\,2,1}$		
	2,2	$-G_{y\,2,1}$	$\begin{pmatrix} G_{x\,1,2} \\ +G_{y\,2,1} \\ +G_{x\,2,2} \\ +G_{y\,2,2} \end{pmatrix}$	$G_{y\,2,2}$	
	2,3		$-G_{y\,2,2}$	$\begin{pmatrix} G_{x\,1,3} \\ +G_{y\,2,2} \\ +G_{x\,2,3} \\ +G_{y\,2,3} \end{pmatrix}$	$G_{y\,2,3}$
	2,4			$-G_{y\,2,3}$	$\begin{pmatrix} G_{x\,1,4} \\ +G_{y\,2,3} \\ +G_{x\,2,4} \end{pmatrix}$
3	3,1	$-G_{x\,2,1}$			
	3,2		$-G_{x\,2,2}$		
	3,3			$-G_{x\,2,3}$	
	3,4				$-G_{x\,2,4}$
4	4,1				
	4,2				
	4,3				
	4,4				

Table 9.5: continuation 2

Spalte	Nr.	Zeile 3			
		$h_{3,1}$.	$h_{3,2}$.	$h_{3,3}$.	$h_{3,4}$.
1	1,1				
	1,2				
	1,3				
	1,4				
2	2,1	$-G_{x\,2,1}$			
	2,2		$-G_{x\,2,2}$		
	2,3			$-G_{x\,2,3}$	
	2,4				$-G_{x\,2,4}$
3	3,1	$\begin{pmatrix} G_{x\,2,1} \\ +G_{x\,3,1} \\ +G_{y\,3,1} \end{pmatrix}$	$-G_{y\,3,1}$		
	3,2	$G_{y\,3,1}$	$\begin{pmatrix} G_{x\,2,2} \\ +G_{y\,3,1} \\ +G_{x\,3,2} \\ +G_{y\,3,2} \end{pmatrix}$	$-G_{y\,3,2}$	
	3,3		$-G_{y\,3,2}$	$\begin{pmatrix} G_{x\,2,3} \\ +G_{y\,3,2} \\ +G_{x\,3,3} \\ +G_{y\,3,3} \end{pmatrix}$	$-G_{y\,3,3}$
	3,4			$-G_{y\,3,3}$	$\begin{pmatrix} G_{x\,2,4} \\ +G_{y\,3,3} \\ +G_{x\,3,4} \end{pmatrix}$
4	4,1	$-G_{x\,3,1}$			
	4,2		$-G_{x\,3,2}$		
	4,3			$-G_{x\,3,3}$	
	4,4				$-G_{x\,3,4}$

Table 9.6: continuation 3

		Zeile 4				RS
Spalte	**Nr.**	$h_{4,1}$	$h_{4,2}$	$h_{4,3}$	$h_{4,4}$	
1	**1,1**					$= 0$
	1,2					$= 0$
	1,3					$= 0$
	1,4					$= 0$
2	**2,1**					$= 0$
	2,2					$= 0$
	2,3					$= 0$
	2,4					$= 0$
3	**3,1**	$-G_{x\,3,1}$				$= 0$
	3,2		$-G_{x\,3,2}$			$= 0$
	3,3			$-G_{x\,3,3}$		$= 0$
	3,4				$-G_{x\,3,4}$	$= 0$
4	**4,1**	$\begin{pmatrix} G_{x\,3,1} \\ +G_{y\,4,1} \end{pmatrix}$	$-G_{y\,4,1}$			$= 0$
	4,2	$-G_{y\,4,1}$	$\begin{pmatrix} G_{x\,3,2} \\ +G_{y\,4,1} \\ +G_{y\,4,2} \end{pmatrix}$	$-G_{y\,4,2}$		$= 0$
	4,3		$-G_{y\,4,2}$	$\begin{pmatrix} G_{x\,3,3} \\ +G_{y\,4,2} \\ +G_{y\,4,3} \end{pmatrix}$	$-G_{y\,4,3}$	$= 0$
	4,4			$-G_{y\,4,3}$	$\begin{pmatrix} G_{x\,3,4} \\ +G_{y\,4,3} \end{pmatrix}$	$= 0$

The calculation of conductivity can be achieved according to following scheme (see figure 9.5)

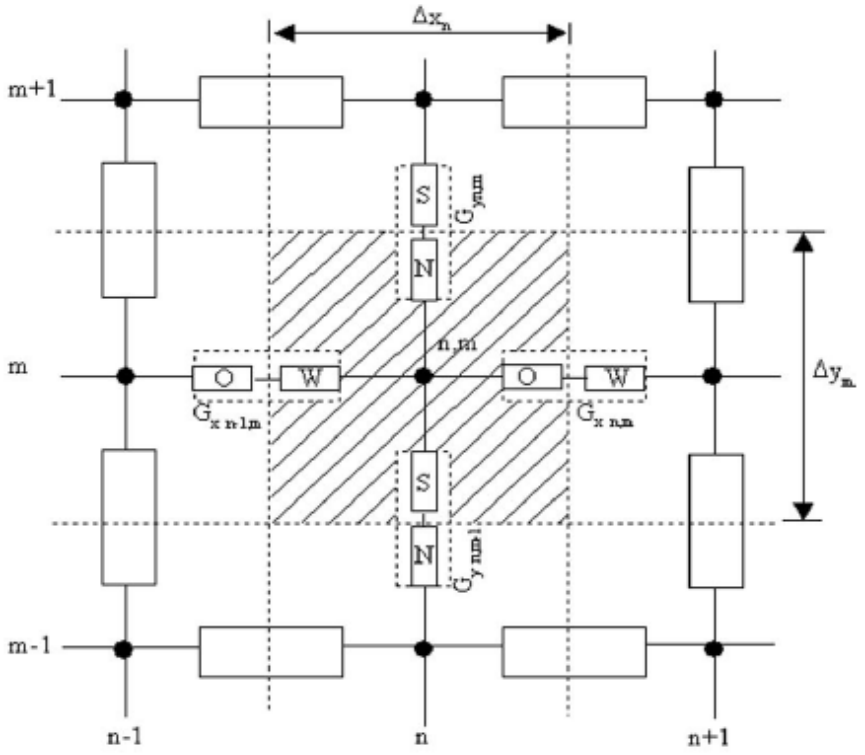


figure 9.5: allocation of hydraulic parameter to compensatory conductivity

The smallest quantization area is the area, which results from around the regarded point $P_{n,m}$ with the edge lengths Δx and Δy. According to the quantization regulations this area will be uniformly parameterised, i.e. the parameters of the aquifer such as *k*-value, storage coefficient *S*, position of the aquiclude *a*, through flow thickness D and the current piezometer head or the situation of the free groundwater surface z_R are considered as independent of location x, y within this planning element. A change of these values can take place only at the borders of the planning element. There however large jumps may arise. Thus the parameters and state variables of the aquifer are reflected by means of FDM with discontinuous functions in the quantized net. These usually contradict continuous functions in original process. For the flow processes this means this planning element is regarded as homogeneous aquifer with horizontal bed and horizontal groundwater level. According to the quantization step an interpolation between the nodes is not allowed. This corresponds to the same statements, which apply to quantized signals (see GRÄBER: Scripte zu den Vorlesungen Automatisierungstechnik bzw. Grundwassermesstechnik). If we consider these premises, the individual hydraulic conductivity can be defined.

Generally:

$$G = k \cdot \frac{A}{l} \tag{9.7}$$

whereby A is the perpendicularly through flow area and *l* is the parallel through flow length. A results from the flow width b and through flow thickness D, which is equal to the thickness of the confined aquifer or the free groundwater height z_R. The differential conductivity of a streamline:

$$dG = k \cdot \frac{dz \cdot b}{l} \tag{9.8}$$

or the total conductivity of a perpendicularly through flow area results from parallel connection of the individual streamline conductivity. The parallel connection is regarded as summation or integration of the differential conductivity:

$$G = \int_{z=a}^{D} k \frac{b}{l} dz$$

$$G = \frac{b}{l} \int_{z=a}^{D} k dz$$

$$G = \frac{T \cdot b}{l}$$

with:

$$T = \int_{z=a}^{D} k dz$$

$$D = \left\{ \begin{array}{ll} M & \text{confined} \\ z_R & \text{unconfined} \end{array} \right\} \text{ aquifer}$$

This integral expression of the transmissibility will be evaluated numerically poorly, since the permeability coefficient is a step function and can not be represented as continuous function. Thus the transmissibility is always a piecewise linear function of the variable z and thereby can be written as sum formula:

$$T = \sum_{l=0}^{n} k_l \left((z_{l+1} - z_l)\Gamma_1 + (z_R - z_l)\Gamma_2 \right) \tag{9.9}$$

with:

$$\Gamma_1 = \begin{cases} 0 & z_R \leq z_{l+1} \\ 1 & z_R > z_{l+1} \end{cases}$$

$$\Gamma_2 = \begin{cases} 0 & z_R > z_{l+1} \\ 1 & z_l \leq z_R \leq z_{l+1} \end{cases}$$

z_l is the absolute heights of the different soil layers and k_l is the pertinent permeability coefficients. The single conductivity (indices O, W, S, N) is computed as follows:

$$\begin{aligned} G_{x\,n,m\,O} &= T_{n,m}\frac{\Delta y}{\frac{\Delta x}{2}} = 2 \cdot T_{n,m}\frac{\Delta y}{\Delta x} \\ G_{x\,n,m\,W} &= G_{x\,n,m\,O} \end{aligned} \tag{9.10}$$

Since it is presupposed that in the planning element n,m all parameters, including though flow thickness, are constant, each pair of conductivities can be assumed equal:

$$\begin{aligned} G_{y\,n,m\,S} &= T_{n,m}\frac{\Delta x}{\frac{\Delta y}{2}} = 2 \cdot T_{n,m}\frac{\Delta x}{\Delta y} \\ G_{y\,n,m\,N} &= G_{x\,n,m\,S} \end{aligned} \tag{9.11}$$

The interconnection of two partial conductivities, e.g. the $G_{x\,n,m\,O}$ and the $G_{x\,n+1,m\,W}$, results in the conductivity between two knots. It is know from the fluid engineering or electro-technology that the series connection of two resistances is their sum:

$$R = R_1 + R_2 \tag{9.12}$$

$$\begin{aligned} G = \frac{1}{R} &= \frac{1}{R_1 + R_2} \\ &= \frac{1}{\frac{1}{G_1} + \frac{1}{G_2}} \\ &= \frac{G_1 \cdot G_2}{G_1 + G_2} \end{aligned} \tag{9.13}$$

with

$$\begin{aligned} G_{xn,m} &= \frac{G_{xn,mO} \cdot G_{xn+1,mW}}{G_{xn,mO} + G_{xn+1,mW}} \\ &= \frac{2T_{n,m}\frac{\Delta y_m}{\Delta x_n} \cdot 2T_{n+1,m}\frac{\Delta y_m}{\Delta x_{n+1}}}{2T_{n,m}\frac{\Delta y_m}{\Delta x_n} + 2T_{n+1,m}\frac{\Delta y_m}{\Delta x_{n+1}}} \\ &= \frac{2T_{n,m}\frac{1}{\Delta x_n} \cdot T_{n+1,m}\frac{\Delta y_m}{\Delta x_{n+1}}}{T_{n,m}\frac{1}{\Delta x_n} + T_{n+1,m}\frac{1}{\Delta x_{n+1}}} \\ &= \frac{2T_{n,m} \cdot T_{n+1,m}\Delta y_m}{T_{n,m}\Delta x_{n+1} + T_{n+1,m}\Delta x_n} \end{aligned} \tag{9.14}$$

in the equidistant part, i.e.

$$\Delta x_n = \Delta x_{n+1} = \Delta x \qquad \text{bzw.}$$
$$\Delta y_m = \Delta y_{m+1} = \Delta y$$

we get:

$$G_{xn,m} = \frac{2T_{n,m} \cdot T_{n+1,m}}{(T_{n,m} + T_{n+1,m})} \frac{\Delta y}{\Delta x}$$

generally:

$$G_{xn,m} = T \frac{\Delta y_m}{\Delta x_n}$$

it yields:

$$T = 2 \frac{T_{n,m} \cdot T_{n+1,m}}{T_{n,m} + T_{n+1,m}} \qquad (9.15)$$

This harmonious averaging also corresponds to the hydraulic real condition very well, in which the smaller permeability coefficient or the smaller transmissibility dominantly intersperses. The computation of transmissibility is complicated in the steady case with unconfined aquifer, since here it is a matter of a value, which depends on the solution of the equation system. In the example of an unstratified aquifer the transmissibility is:

$$T_{n,m} = k_{n,m} \cdot z_{Rn,m} \qquad (9.16)$$

In this case it is necessary to compute the equation system iteratively. And we start with an estimated value $z^{(1)}_{R\,n,m}$. The better this estimated value to the true solution $z_{R\,n,m}$ approaches, the fewer iteration steps have to be implemented. The first approximation for the transmissibility $T^{(1)}_{n,m}$ can be computed by means of the estimated value and the coefficient matrix with the appropriate conductivity can be developed. It leads to solution $z^{(2)}_{R\,n,m}$, an improved approximation of exact solution $z_{R\,n,m}$. This is again used for the computation of improved transmissibility $T^{(2)}_{n,m}$. The procedure is continued, until the deviation of two approximations is smaller than a certain limit ε.

$$\left| \frac{z^{(t)}_{R\,n,m} - z^{(t+1)}_{R\,n,m}}{z^{(t)}_{R\,n,m}} \right| < \varepsilon$$

9.1.1.2 Consideration of boundary condition

The preceding accomplishment for the quantization of continuum applies to the case of uninfluenced groundwater flow. For a unique simulation boundary conditions must work as in original procedure, and no groundwater flow comes about without it. With the discontinuous models on the basis of finite differences method the boundary conditions affect in principle on the knots. Above all the consideration of 1st type boundary conditions come across difficulties.

During the numerical simulation at least one 1st type boundary condition must work on a knot. Models, which are endued with 2nd or 3rd boundary conditions, yield no unique simulation results. The case of the infinitely expanded aquifer, which can be calculated by means of THEIS analytical solution, does not exist in discontinuous models.

In the following the realization possibilities will be indicated for the different kinds of boundary conditions. The **2nd type boundary condition**, which affects on a knot (see figure 9.6), can be considered as follows based on balance equation:

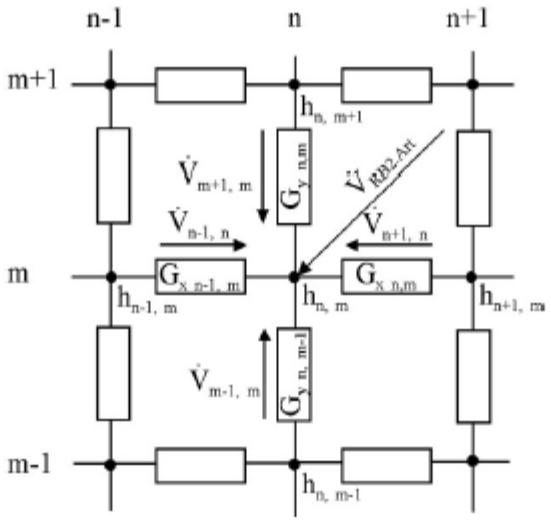


figure 9.6: consideration of 2nd type boundary condition

$$\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} + \dot{V}_{RB\,2.Art\,n,m} = 0$$

$$\underbrace{\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m}}_{\text{unknown quantities}} = \underbrace{-\dot{V}_{RB\,2.Art\,n,m}}_{\text{known quantities}} \qquad (9.17)$$

Thus the 2nd type boundary condition can be directly written on the right side of the equation. With introduction of the potential differences at volume flow place this equation turns into to the accustomed matrix equation for grid network, whereby for the knots with 2nd type boundary condition the right side of is different from zero.

With **3rd type boundary condition** (see figure 9.7) we proceed directly in principle. The balance equation is built for the knot, on which the boundary condition affects. The boundary condition 3. Kind is by definition a potential decrease of a flow resistance, which is nothing else but a variable flow rate in turn. With 3rd type boundary condition the potential difference between a given potential (e.g. water level of a receiving stream or another surface water) and the water level at the point of aquifer, where the boundary condition affects, will be built:

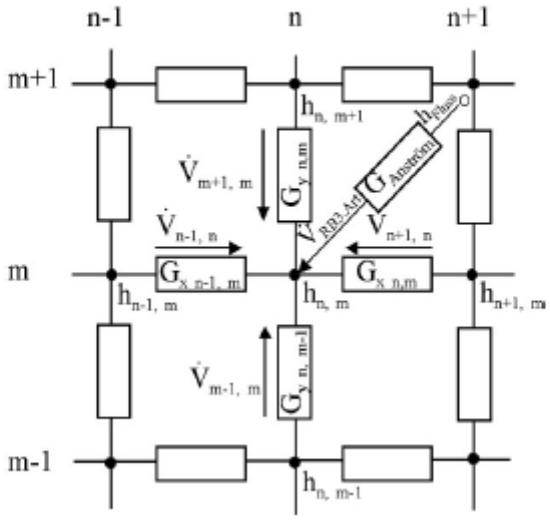


figure 9.7: consideration of a 3rd type boundary condition

$$\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} + \dot{V}_{RB3.Art\,n,m} = 0$$
$$\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} + G_{anstrm\,n,m}\,(h_{n,m} - h_{Flu}) = 0$$
$$\underbrace{\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} + G_{anstrm\,n,m} \cdot h_{n,m}}_{\text{unknown quantities}} = \underbrace{G_{anstrm\,n,m} \cdot h_{Flu}}_{\text{known quantities}} \qquad (9.18)$$

With introduction of potential difference for all volume flows:

$$\begin{aligned} & G_{x\,n-1,m}(h_{n,m} - h_{n-1,m}) + \\ & G_{y\,n,m-1}(h_{n,m} - h_{n,m-1}) + \\ & G_{x\,n,m}(h_{n,m} - h_{n+1,m}) + \\ & G_{anström\,n,m} h_{n,m} \\ & G_{y\,n,m}(h_{n,m} - h_{n,m+1}) = G_{anström\,n,m} h_{Fluss\,n,m} \end{aligned} \tag{9.19}$$

or arranged according to *h:*

$$\begin{aligned} & h_{n-1,m}(-G_{x\,n-1,m}) + \\ & h_{n,m-1}(-G_{y\,n,m-1}) + \\ & h_{n,m}(G_{x\,n-1,m} + G_{x\,n,m} + G_{y\,n,m-1} + G_{y\,n,m} + G_{anström\,n,m}) + \\ & h_{n,m+1}(-G_{y\,n,m}) + \\ & h_{n+1,m}(-G_{x\,n,m}) = G_{anström\,n,m} h_{Fluss\,n,m} \end{aligned} \tag{9.20}$$

Thus we recognize that with existence of a 3rd type boundary condition the main diagonal also has further addends besides the right side different from zero.

1st type boundary conditions will be treated in the discontinuous model as a special case of 3rd type boundary condition with evanescent flow resistance. Since the balance equations of the knots orient flow rates, the potential of the boundary condition won't arise explicitly. Hydraulically this mathematical step is quite meaningfully interpretable, since the 3rd type boundary conditions can be also regarded as combination of a 1st type boundary condition and a flow resistance. If the flow resistance approaches to zero, i.e. the potential loss between groundwater level and surface water disappears, this is equated with the influence of a 1st type boundary condition. Based on this conclusion the derivation can be taken over according to the above statements of 3rd type boundary condition:

$$
\begin{aligned}
&h_{n-1,m}(-G_{x\,n-1,m})+\\
&\quad h_{n,m-1}(-G_{yn,m-1})+\\
&\quad\quad h_{n,m}(G_{x\,n-1,m}+G_{x\,n,m}+G_{y\,n,m-1}+G_{y\,n,m}+G_{RB\,1.Art\,n,m})+\\
&\quad\quad\quad h_{n,m+1}(-G_{y\,n,m})+\\
&\quad\quad\quad\quad h_{n+1,m}(-G_{x\,n,m}) = G_{RB\,1.Art\,n,m}h_{Fluss\,n,m}
\end{aligned}
\tag{9.21}
$$

In this equation $G_{anström\,n,m} = G_{RB\,1.Art\,n,m}$ is set since it is a matter of mathematical shaping of 1st type boundary condition. An evanescent resistance, i.e. $R_{anström} \rightarrow 0$, means a conductivity $G_{RB1.Art\,n,m} \rightarrow \infty$. This is not numerically realizable here. Since $G_{RB1.Art\,n,m}$ has influent conductivity in additive linkage to the others at the knot n,m, it is sufficient that the conductivity $G_{RB1.Art\,n,m}$ dominates the summation of conductivity. This is given, if the following inequation is fulfilled:

$$G_{RB\,1.Art\,n,m} >> G_{x\,n-1,m} + G_{x\,n,m} + G_{y\,n,m-1} + G_{y\,n,m}$$

This inequation can be regarded as fulfilled if:

$$G_{RB\,1.Art\,n,m} = 100 \cdot (G_{x\,n-1,m} + G_{x\,n,m} + G_{y\,n,m-1} + G_{y\,n,m})$$

Due to of numerical instabilities within the equation solution $G_{RB1.Art\,n,m}$ should not be select too large.

In some simulation programs 1st type boundary conditions are also interpreted as infinitely large storage effect. This however only works in the application of unsteady flow regime.

If the boundary conditions are located outside the nodes, which is in the majority of cases, and not all the field element boundary condition properties are arranged, then the boundary condition can be connected with four of resistances of neighbouring knots (see figure 9.8). The computation of resistances and the associated allocation of boundary condition effect on the neighbouring knots can be achieved according to geometrical conditions, i.e. according to the distance between boundary condition and knots and the appropriate effect range. The effect range results from the gravity centre of the representative area between the connecting lines of gravity centres and the adjacent knots. And the gravity centres should have coordinates $x_{Mn,m}$; $y_{Mn,m}$.

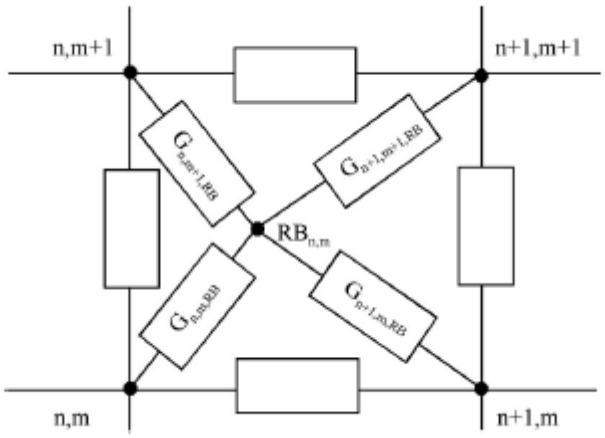


figure 9.8: outside knots lying boundary conditions

$$G_{n,m,RB} = T_{n,m}\frac{b}{l} \qquad \text{mit} \tag{9.22}$$

$$b = \sqrt{(x_{Mn,m} - x_{Mn+1,m})^2 + (y_{Mn,m} - y_{Mn+1,m})^2} \tag{9.23}$$

$$l = \sqrt{(x_{n,m} - x_{RBn,m})^2 + (y_{n,m} - y_{RBn,m})^2} \tag{9.24}$$

The mathematical formulation of different kinds of boundary conditions can be achieved according to above forms for 1st 2nd 3rd type boundary conditions.

9.1.2 Finite Element Method

The finite element method (FEM) represents a further kind of continuum transfer into a quantized representation. It differs from the FDM (see section 9.1.1 finite differences method, page 237) in the determination of planning elements and parameters. While in the FDM the balance is essentially computed on the basis of the knot equation, in the FEM the balance takes sides with planning elements.

Arbitrary compounds in the form of polyhedrons can be selected as planning elements. In the two-dimensional level the planning elements become planes, which are formed by arbitrary closed polygons. The simplest form is formation of triangle elements. Each higher order planning elements can be formed by the summary of more triangles.

In the planning elements potential functions, e.g. water level, piezometer head, concentration distribution, temperature or others, are presupposed as linear, homogeneous conditions. Thus the distribution can be computed analytically within the planning elements. This computation can be achieved by means of variation calculation or according to GALERKIN method.

According to the principle of variation calculation an approximation function $P^*_{(x,y)}$ of the quantized continuums is looked for the potential distribution $P_{(x,y)}$ (e.g. h, z_R, Φ) in the entire regarded area G. Since by definition linear system status dominates within the planning element, for a triangle element:

$$P^*_{(x,y)} = a + b \cdot x + c \cdot y \qquad (9.25)$$

In any cases this equation must fulfil the potential distribution at the supporting place, the triangle points *i, j, k*:

$$\begin{aligned} P^*_{i\,(x,y)} &= a + b \cdot x_i + c \cdot y_i \\ P^*_{j\,(x,y)} &= a + b \cdot x_j + c \cdot y_j \\ P^*_{k\,(x,y)} &= a + b \cdot x_k + c \cdot y_k \end{aligned}$$

Then we get three equations with three unknown quantities *a, b,* and *c,* which can be solved. As matrix equation:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{bmatrix}^{-1} \cdot \begin{bmatrix} P^*_i \\ P^*_j \\ P^*_k \end{bmatrix} \qquad (9.26)$$

To insert the solution into the equation about $P^*_{(x,y)}$ yields the solution of searched function. For the planning element *m* the potential distribution can be expressed as following:

$$P^*_m(x,y) = W_i(x,y) \cdot P^*_i + W_j(x,y) \cdot P^*_j + W_k(x,y) \cdot P^*_k$$

the weight functions W(x, y) are linear in x and y direction in the region G and subject to orthogonal conditions, i.e.:

$$W_n(x_m, y_m) = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

$$m, n \in i, j, k$$

Individually:

$$W_i(x,y) = \frac{1}{2\Delta}\left[(x_j \cdot y_k - x_k \cdot y_j) + (y_j - y_k) \cdot x + (x_k - x_j) \cdot y\right]$$

$$W_j(x,y) = \frac{1}{2\Delta}\left[(x_k \cdot y_i - x_i \cdot y_k) + (y_k - y_i) \cdot x + (x_i - x_k) \cdot y\right]$$

$$W_k(x,y) = \frac{1}{2\Delta}\left[(x_i \cdot y_j - x_j \cdot y_i) + (y_i - y_j) \cdot x + (x_j - x_i) \cdot y\right]$$

$$\text{mit}: \Delta = \frac{1}{2} \cdot \left[x_i\,(y_j - y_k) + x_j\,(y_k - y_i) + x_k\,(y_i - y_j)\right]$$

Δ is the surface area of the planning triangle, and:

$$W_i(x,y) + W_j(x,y) + W_k(x,y) = 1 \qquad (9.27)$$

P^* is continuous in the entire range G. If the basic values are known at the knots, the function is representable in the whole area. In the following it is to be demonstrated how to determine the basic values p_i of the continuum such that the potential values P^*_i of the quantized area are adapted in best way. This is achieved according to GALERKIN method.
The set up differential equations apply to the continuum exactly. In the case of the horizontal planes groundwater flow equation (see equation 7.14, page 186):

$$\frac{\partial}{\partial x}\left(T\frac{\partial z_R}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial z_R}{\partial y}\right) = 0$$

In quantized system in contrast only following approximation is valid:

$$\frac{\partial}{\partial x}\left(T\frac{\partial z_R^*}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial z_R^*}{\partial y}\right) = r(x,y) \neq 0$$

r(x, y) is designated as residue. The better the quantization area to the continuum adapts, the smaller the **residue** is. The approximation solution $P^*_{i,}$ here concretely $z^*_{Ri,}$ converges for the continuum solution P or z_R or other expressions in case of infinitely small planning elements or infinitely large number of supporting places, knots and planning elements. And the residue becomes zero.
According to the weighted residues method the approximation solution will be searched in such a way that the residue disappears at the weighted mean. This is accomplished with the following expression for each planning element:

$$\iint_G W_i(x,y) \cdot r(x,y)\, dG = 0 \qquad (9.28)$$

Under constant transmissibility condition in individual planning elements, the residue can be written by definition:

$$\iint_G W_i(x,y) \cdot T \cdot \left(\frac{\partial^2 z_R^*}{\partial x^2} + \frac{\partial^2 z_R^*}{\partial y^2} \right) dx \cdot dy = 0 \qquad (9.29)$$

It was still considered that the time dependence should be identically zero. This applies to the steady processes.

But also the unsteady processes can be treated in such way, since the time dependence is regarded independent of local quantization (see section 9.2 time quantization method, page 258).
This area integral must be solved for each supporting point or net point. We get n equations weighting functions with n unknowns. For the approximation solution z^*_R at the net point a linear substitution was selected, which however has no second derivative. On this account the integral must be transformed with **1st Green's formula**. Generally:

$$\iint_G \left[u \cdot \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} \right) \right] dxdy = - \iint_G \left[\frac{\partial w}{\partial x} \cdot \frac{\partial u}{\partial x} + \frac{\partial w}{\partial y} \cdot \frac{\partial u}{\partial y} \right] dxdy + \int_B u \frac{\partial w}{\partial n} dB \qquad (9.30)$$

Thereby B means the edge of the area and n is unit normal, which stands perpendicularly on the edge and to is directed towards outside. w(x, y) and u(x, y) are arbitrary scalar potential functions. This equation applies to residue:

$$- \iint_G \left[T \cdot \left(\frac{\partial z_R^*}{\partial x} \cdot \frac{\partial W_i}{\partial x} + \frac{\partial z_R^*}{\partial y} \cdot \frac{\partial W_i}{\partial y} \right) \right] dx\, dy + \int_B T \cdot W_i \frac{\partial z_R^*}{\partial n}\, dB = 0 \qquad (9.31)$$

The line integral of the edge B describes the potential-independent flow over the edge and can be regarded as boundary condition for the planning element n:

$$\int_B T \cdot W_i \frac{\partial z_R^*}{\partial n}\, dB = \int_B W_i \cdot q'_n\, dB = \dot{V}$$

whereby q'_n means the specific flow rate per unit length and V is the influx to the knot n. In the same way internal boundary conditions e.g. wells can be also considered.

Similar to the finite differences method a high-dimensional equation system also develops here in the FEM. In contrast to FDM the FEM does not produce diagonal band matrices, but an irregular matrix structure develops as a function of the number of affecting planning elements. This leads to a increased numerical expenditure in the solution equation system.

9.2 Time quantization method

$$\operatorname{div}(T \operatorname{grad} h) = -S\frac{\partial h}{\partial t} \quad \text{mit} \quad h = \left\{ \begin{array}{ll} h & \text{confined} \\ z_R & \text{unconfined} \end{array} \right\} \text{aquifer}$$

On the basis of general form of horizontal planes groundwater flow equation the independent of treatment on the left side, local functionality, on the right side, the time dependence of a quantization must be undergone. Quantization of place dependence can be achieved by the described method in the section 9.1. The temporal derivative must be transferred into a difference quotient, since otherwise there are no possible simple numerical treatment. The construction of an appropriate equation system is only again possible by this transfer. The transfer from the derivative into a difference quotient should be visualised by first backward difference method as implicit procedure.

$$\frac{\partial h}{\partial t} \approx \frac{h_t - h_{t-\Delta t}}{\Delta t}$$

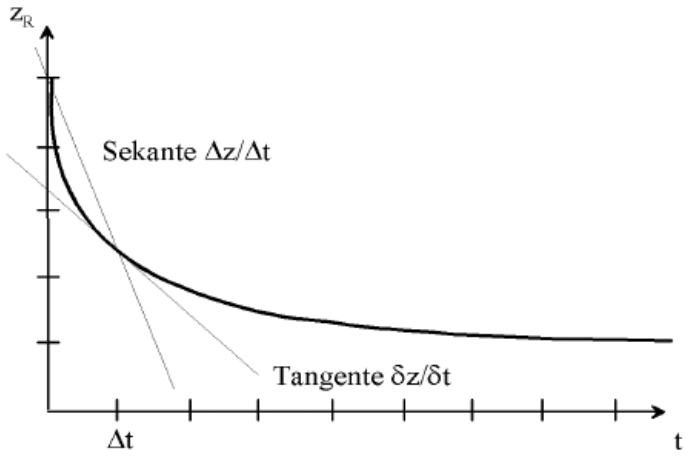


figure 9.9: Relationship from tangent to secant with a typical drawdown procedure

Due to the introduction of the temporal difference quotient, a time point must be also arranged at the convection part, i.e. the left side of the equation. Different methods for time quantization are differentiated for allocating the time to the left side of the differential equation, i.e. the local flow process (see figure 9.9). The most substantial distinctions lie between one- and multi-step procedures as well as between the explicit and implicit procedures. With the explicit one step method the equation system can be solved directly, since the parameters are known from the

preceding time step. The explicit method, also designated as forward difference, however has large disadvantage that for mathematical stability reasons only very small time steps (in minute order of magnitude) are realizable and an extremely large number of time steps (total simulation time divide by the time increments) must be worked out for a test run. The implicit one step method is also called backward difference method. It also yields stable solutions for large time steps and thus represents the standard method for numerical simulation systems. A series procedures are developed to decrease the quantization error (to multi-step method, Predictor Corrector method, higher order method, see figure 9.10), and also a special extrapolation procedure by GRÄBER.

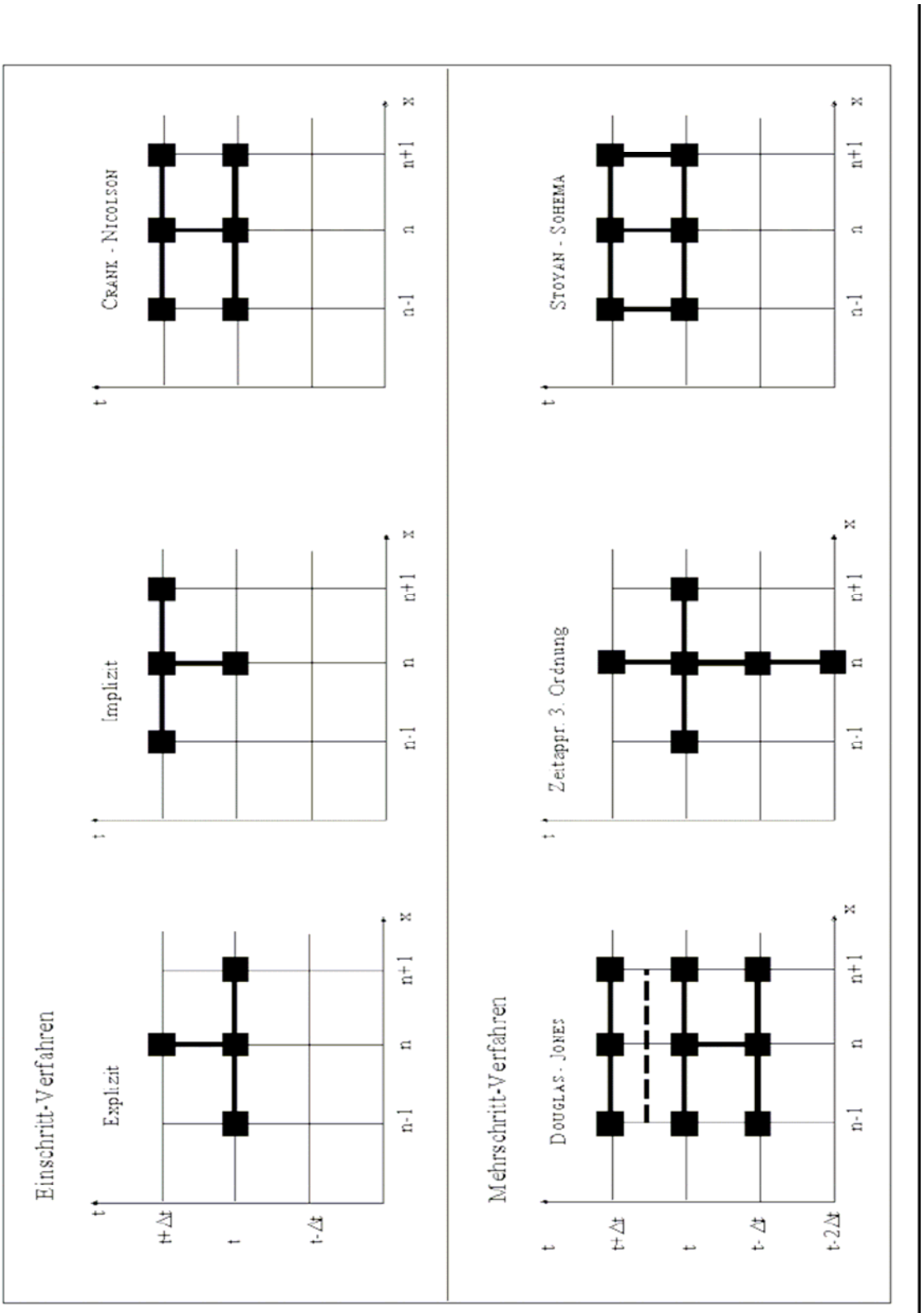


figure 9.10: Argument association in time quantization of a 1-D-field problem

9.2.1 Backward difference - Implicit method

In the implicit one step method, also designated as backward difference or LIEMANN method, the local flow part of the horizontal planes groundwater flow equation will be considered to time point *t*:

$$\operatorname{div}(T\ \operatorname{grad}\ h) = -S\frac{\partial h}{\partial t} \qquad \text{mit } h = \left\{ \begin{matrix} h & \text{confined} \\ z_R & \text{unconfined} \end{matrix} \right\} \text{ aquifer} \qquad (9.32)$$

After the transfer from the temporal derivative into the difference quotients:

$$\operatorname{div}(T\ \operatorname{grad}\ h)_{|t} \approx -S\frac{h_t - h_{t-\Delta t}}{\Delta t} \qquad (9.33)$$

Instead of the squiggly equal sign the equals sign is mostly used, which actually is not exact. If we implement quantization again also on the left side and insert the (local-) conductivity according to physical FDM method, then balance equation at knot n, m arises (see figure 9.11):

$$\begin{aligned} \left[\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m} + \dot{V}_{Zeit\,n,m}\right]_t &= 0 \\ \left[\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m}\right]_t + G_{Zeit\,t,n,m}\,(h_{t,n,m} - h_{\Delta t,n,m}) &= 0 \\ \underbrace{\left[\dot{V}_{n-1,n} + \dot{V}_{n+1,n} + \dot{V}_{m-1,m} + \dot{V}_{m+1,m}\right]_t + G_{Zeit\,t,n,m} \cdot h_{t,n,m}}_{\text{unknown quantities}} &= \underbrace{G_{Zeit\,t,n,m} \cdot h_{\Delta t,n,m}}_{\text{known quantities}} \end{aligned} \qquad (9.34)$$

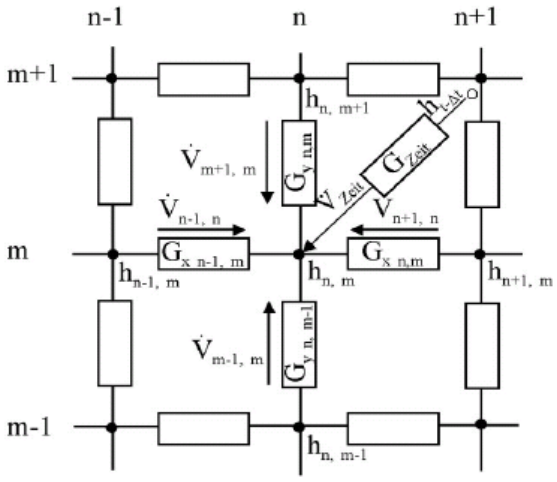


figure 9.11: consideration of time quantization

$$\left|\begin{array}{l}\left[\begin{array}{l} h_{n-1,m}(-G_{x\,n-1,m})+ \\ \quad h_{n,m-1}(-G_{y\,n,m-1})+ \\ \quad\quad h_{n,m}(G_{x\,n-1,m}+G_{x\,n,m}+G_{y\,n,m-1}+G_{y\,n,m})+ \\ \quad\quad h_{n,m+1}(-G_{y\,n,m})+ \\ \quad\quad h_{n+1,m}(-G_{x\,n,m}) \end{array}\right] \\ \qquad = -S\,\Delta x_n \Delta y_m (h_{n,m\,t} - h_{n,m\,t-\Delta t}) \cdot \dfrac{1}{\Delta t} \end{array}\right|_t \qquad (9.35)$$

The expression

$$S\,\Delta x_n \Delta y_m = C_{n,m} \qquad (9.36)$$

will be designated as hydraulic storage effect or capacity. The so called time conductivity is introduced For the quotient from capacity and time step.

$$\frac{S\,\Delta x_n \Delta y_m}{\Delta t} = G_{z\,n,m} = \frac{C_{n,m}}{\Delta t} \qquad (9.37)$$

This represents flow rate in connection with the temporal potential, which is released or received from aquifer due to storage effect within the time step Δt .

$$\dot{V}_{Zeit\,n,m} = G_{Zeit\,t,n,m}\,(h_{t,n,m} - h_{\Delta t,n,m}) \qquad (9.38)$$

If this is inserted into the upper equation system and within a row we arrange the known and unknown variables, then we get the following system, in which only known quantities stand on the right side:

$$\left|\begin{array}{l}\left[\begin{array}{l} h_{n-1,m}(-G_{x\,n-1,m})+ \\ \quad h_{n,m-1}(-G_{y\,n,m-1})+ \\ \quad\quad h_{n,m}(G_{x\,n-1,m}+G_{x\,n,m}+G_{y\,n,m-1}+G_{y\,n,m}+G_{z\,n,m})+ \\ \quad\quad h_{n,m+1}(-G_{y\,n,m})+ \\ \quad\quad h_{n+1,m}(-G_{x\,n,m}) \end{array}\right] \\ \qquad = h_{n,m\,t-\Delta t}(G_{z\,n,m}) \end{array}\right|_t \qquad (9.39)$$

Thus the pentadiagonal equation system remains. A known quantity from the preceding time step was added on the right side. The main diagonal was also extended by the addends $G_{z\,n,m}$. The developed matrix equation is not explicitly solvable due to the potential dependence of

conductivity, in particular the transmissibility with unconfined aquifer. Therefore iteration must be carried out. In the first step the conductivity $G^{(1)}$ is computed according to the initial water levels ($z_{Rt} \rightarrow \Delta t$). For groundwater drawdown procedures this means that the transmissibility and the conductivity are assumed too large. The matrix can be solved with these values, and we get water levels $z_{Rt}^{(1)}$, which are too low compared with real situation. Improved conductivity $G^{(2)}$ can be computed with this first approximation $z_{Rt}^{(1)}$, which leads to the second approximation of water height $z_{Rt}^{(2)}$. This exceeds the true solution, since the conductivity was assumed too small and too little discharge was realized. The solutions approach to true value in form of a damped oscillation. This iteration process will be continued until the changes between two iterations do not exceed an error bound any longer. Then we get the solution for the time step Δt. Nevertheless the iteration within the time step remains a quantization error, which grows proportionally with Δt, since the secant between the points t and t - Δt is calculated instead of tangent at the point *t*. Since the groundwater flow processes accordingly approach to an abating exponential function asymptotic steady final state, the time quantization error is not only dependent on the increment Δt, but also dependent on the dynamics of process. In the figures 9.12 and 9.13 the results are displayed to defined time for the example of one dimensional ditch flow with different time increments. The results with the increment Δt /24 are assumed accurate. We clearly recognize the strong dependence of time quantization error on temporal gradients. In figure 9.14 the convergence of the solution as a function of time increment is observed. At the same time we recognize the small quantization errors of extrapolation.

drawdown procedure t/24 - - - - - LIEBMANN
$Z_R = Z_{R(x)}$ parameter: t Δt ———LIEBMANN
Δt .....GRÄBER

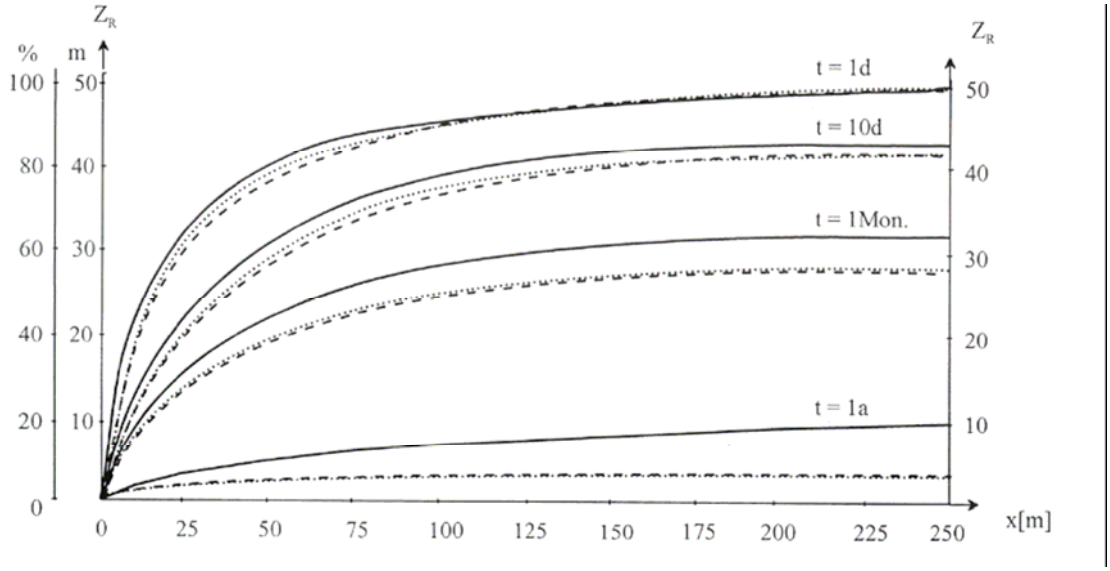


figure 9.12: Time quantized computed drawdown of a ditch flow

arising t/24 - - - - - LIEBMANN
$Z_R = Z_{R(x)}$ Δt ———LIEBMANN
parameter: t Δt .....GRÄBER

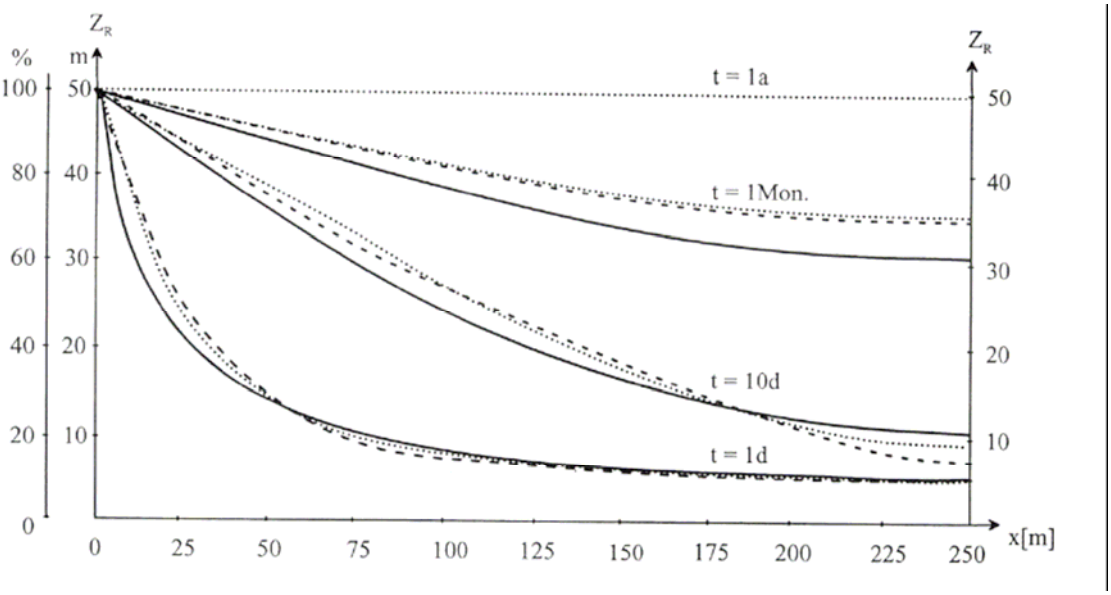


figure 9.13: Time quantized computed groundwater rise of a ditch flow

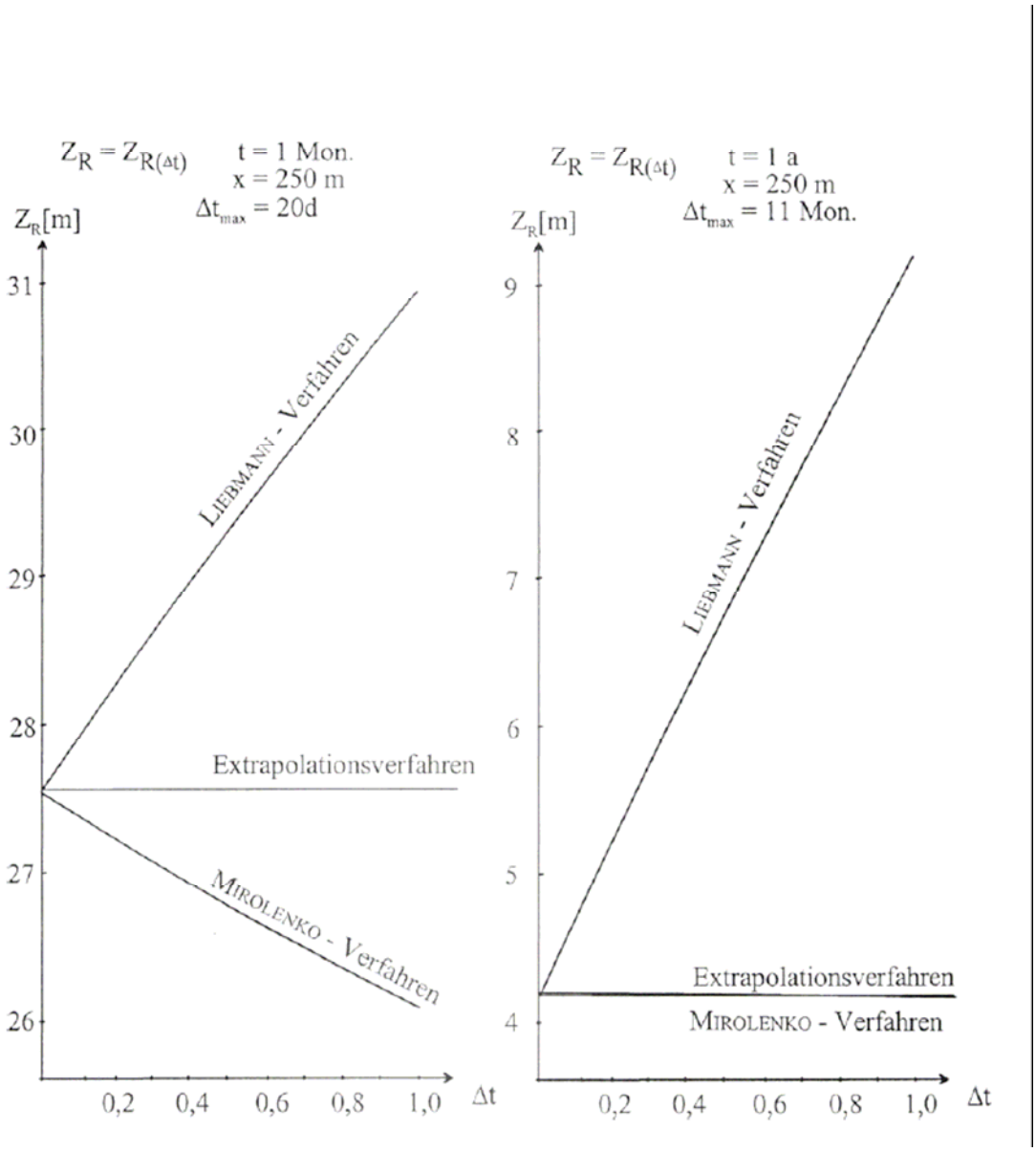


figure 9.14: Dependence of the time quantization error on the time increment

9.2.2 Mixed methods

The backward difference is often applied for simplifying time quantization like already described. That means the parameters will be determined for the time point *t*. This leads to difficulties with nonlinear parameters, like under unconfined groundwater conditions , since they must be adjusted iteratively. There is therefore a series attempts, by weighted allocation of local partial derivative as well as the parameter to quantized time, in order to achieve an error decrease. Generally we can write:

$$\begin{aligned}&(1-\gamma)\ \mathrm{div}\,(T\ \mathrm{grad}\ h)_{|t-\Delta t}+\gamma\ \mathrm{div}\,(T\ \mathrm{grad}\ h)_{|t}\\&=\left((1-\gamma)\,S_{|t-\Delta t}+\gamma S_{|t}\right)\frac{(h_t-h_{t-\Delta t})}{\Delta t}\end{aligned}\qquad(9.40)$$

Depending upon method selection (see figure 9.10, page 260) we get the following γ-values:

	0	explicit method
	1/2	CRANK-NICOLSON scheme
γ =		
	2/3	GALERKIN-weighting
	1	implicit method

Apart from the one step method (also with iteration cycles), in which only the time point *t* and t - Δt play a role, multi-step procedures are frequently used for the simulation to decrease the time quantization error. The Predictor Corrector method is common. In the Douglas JONES method, a two-step method, a half step Δt/2 will be attended according to implicit solution scheme (λ = 1), and all parameters are adjusted to time t - Δt and $h_{t-\Delta t/2}$ (Predictor step) in the substitution. The CRANK NICOLSON Scheme (λ = 1/2) realizes a total step Δt (Corrector step), whereby all parameters are set to time point t - Δt/2 (see section 5.4.1 numerical integration, page 149). Very high approximation accuracies can be also obtained that not only the time derivatives at local quantization point $P_{n,m}$, but also at neighboring knots are taken into consideration. In simplest form according to the Simpson's rule (by an example of one dimensional case):

$$\frac{dh}{dt}\approx\frac{1}{6}\left(\frac{dh}{dt}_{|n-1}+\frac{dh}{dt}_{|n}+\frac{dh}{dt}_{|n+1}\right)\qquad(9.41)$$

A special scheme is suggested by STOYAN, with which all partial derivatives are subject to a controlled weighting. Thus a very stable and exact numerical solution is obtained, which changes into analytical solution for the case of net convection. The disadvantage of this method consists of the fact that we try to reduce the time quantization error effects by the manipulation of the differential equation remainder, usually the parameter of local convection term (the right side of the differential equation). Thus the cause of error remains untouched. It generally leads to no satisfactory solution and numerical instabilities possibly.

9.2.3 Extrapolation method

a very effective method to decrease time quantization error is indicated by GRÄBER in extrapolation method. For the balancing processes, in which the horizontal planes groundwater flow arises, the function $h = h_{(t)}$ is always continuous monotonically increasing or decreasing between the time points t - Δt and t. It yields that the secant of backward difference is always smaller than the tangent at the point t according to amount. Thus the following inequation:

$$\left|\frac{dh}{dt}\right| \leq \left|\frac{h_t - h_{t-\Delta t}}{\Delta t}\right|$$

The inequation can be changed into equation by introducing a correction factor:

$$\frac{dh}{dt} = \frac{h_t - h_{t-\Delta t}}{K \cdot \Delta t} \qquad (9.42)$$

Thus the time quantization error is reduced and can converge to zero by proper selection of K. K is ≥ 1 by definition. Approximately the local point $P_{n,m}$ can be represented by the following equivalent circuit diagram (see figure 9.15). H stands for equivalent potential (1st type boundary condition) and R is an equivalent resistance, which summarizes the hydraulic characteristics of aquifer between the neighbour knots and the regarded knots. It could be e.g. the entire quantization network or a 1st type boundary condition. C_t stands for the effective storage effect of the aquifer in this time step.

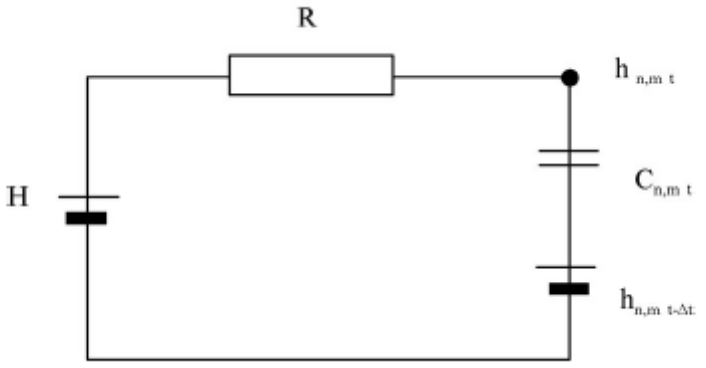


figure 9.15: equivalent circuit diagram of local point $P_{n,m}$

For this equivalent circuit we get (see sections 5.2.1 first order ordinary differential equations, page 111 and 12.1 transmission behaviour with first order delay, page 334):

$$h_{n,m\,t} = (H - h_{n,m\,t-\Delta t})\left(1 - e^{-\frac{\Delta t}{\tau}}\right) \qquad (9.43)$$

whereby τ is designated as time constant $\tau = R\,C$. In place of the capacity Ct an equivalent resistance R_z can be computed now, which engenders the same groundwater flow as the capacity C for time t:

$$R_{z\,n,m\,t} = \frac{h_{n,m\,t}}{\dot{V}} \qquad \text{mit:} \qquad \dot{V} = C_t \frac{dh_{n,m\,t}}{dt} \tag{9.44}$$

$$= \frac{\tau\left(1 - e^{-\frac{\Delta t}{\tau}}\right)}{C_{n,m\,t}\, e^{-\frac{\Delta t}{\tau}}}$$

The difficulty consists of determining time constant τ. An analytical expression cannot be found. Therefore the quotient $\Delta t/\tau$ is determined from the system behaviour during the offset procedure. We simulate the system according to backward difference in first step, i.e. with the time resistance $R_{z\,n,mt} = \Delta t/C_{n,m\,t}$. The change between the potentials $h_{n,m\,t}$ and $h_{n,m\,t-\Delta t}$ serves as basis for the time constant computation. For the case of the drawdown ($h_{n,m\,t} \geq h_{n,m\,t-\Delta t}$):

$$R_{z\,n,m\,t} = \frac{\Delta t}{C_{n,m\,t}} \frac{h_{n,m\,t-\Delta t}}{h_{n,m\,t}} \tag{9.45}$$

bzw.

$$G_{z\,n,m\,t} = \frac{C_{n,m\,t}}{\Delta t} \frac{h_{n,m\,t}}{h_{n,m\,t-\Delta t}}$$

and for a groundwater rising process:

$$R_{z\,n,m\,t} = \frac{\Delta t}{C_{n,m\,t}} \frac{(h_{\max} - h_{n,m\,t-\Delta t})}{(h_{\max} - h_{n,m\,t})} \tag{9.46}$$

bzw.

$$G_{z\,n,m\,t} = \frac{C_{n,m\,t}}{\Delta t} \frac{(h_{\max} - h_{n,m\,t})}{(h_{\max} - h_{n,m\,t-\Delta t})}$$

Here two relatively simple expressions are developed for the corrected time resistance. These solutions are numerically stable and are characterised by a good convergence behaviour and a very small quantization error. Accordingly a 24 times smaller time step is developed (see figure 9.14, page 265) with application of the backward difference. Using larger time steps means a substantial economisation of computing time.

9.3 Tasks of numerical calculation

1. By means of one dimensional steady ditch flow calculate the position of free surface as a function of x, the outflow from head water and the inflow to bottom water (see figure 9.16). Use five quantization elements.

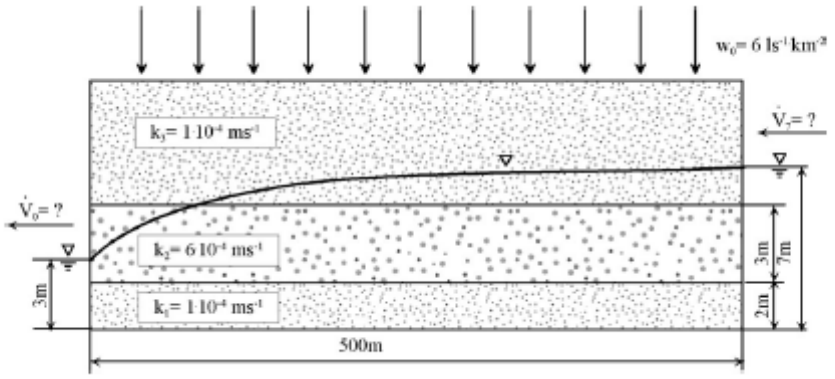


figure 9.16: stratified aquifer with steady flow regime

2. By means of one dimensional steady ditch flow calculate the position of free surface as a function of x and *t* (0 to 2*d*), the outflow from head water and the inflow to bottom water (see figure 9.17).
   Use five quantization elements and five time steps.
   Select the time step according to expected gradients.

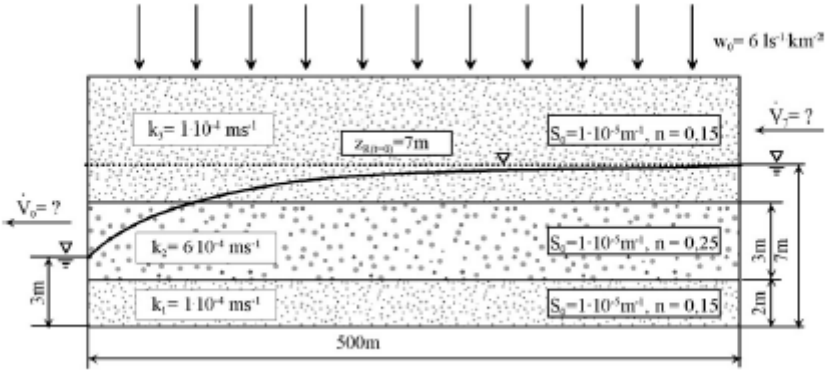


figure 9.17: stratified aquifer with steady flow regime

3. In a aquifer a tunnel (underground) will be built parallel to a river (see figure 9.18). Compute how does the groundwater condition for steady case change caused by this construction.
   Select a suitable rough quantization scheme.

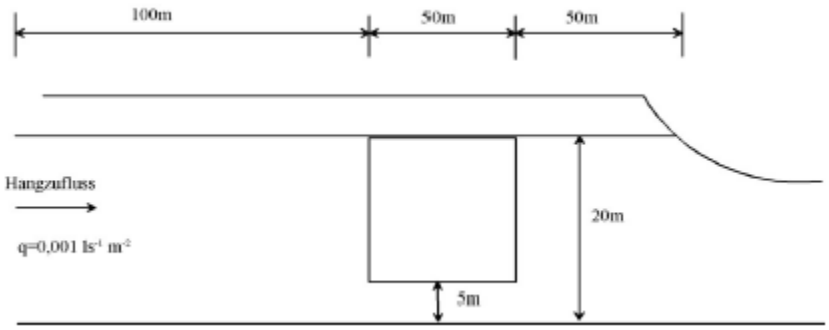


figure 9.18: tunnel construction in an aquifer

4. In a plain tract the polder area is to be protected against floods by means of a dyke construction (see figure 9.19) (according to simplified scheme).
   Dyke: $k = 10^{-4m}/_{s}$, $n_o = 0.15$, $S_0 = 0.002m^{-1}$;
   Sealing material: $k = 10^{-5m}/_{s}$, $n_o = 0.05$, $S_0 = 0.001m^{-1}$

   a) develop the simple discrete scheme to estimate the groundwater flow processes.
   b) How much water flows into the polder area per meter dyke length?

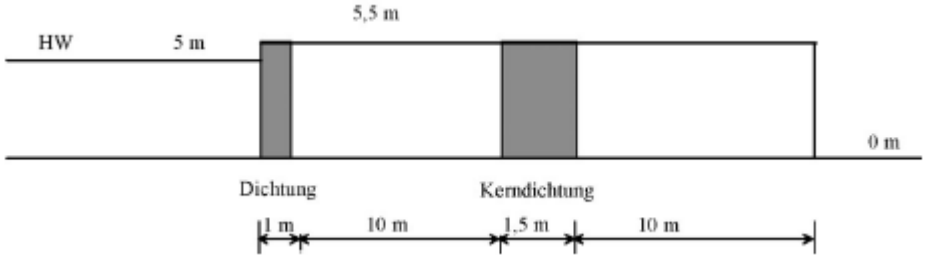


figure 9.19: dyke construction and core seal

5. Expected groundwater flow conditions will be simulated for an induced recharge waterworks (see figure 9.20).

Gegeben: $\dot{V} = 0,001 m^3 s^{-1}$, $S_0 = 0,0001 m^{-1}$, $n_0 = 0,25$, $k = 0,001 m s^{-1}$,
$z_{R(t=0)} = 10m$,
$h_{Fl} = 10m$, $M = 15m$, $b = 1m$

a) develop the simple quantization scheme with three knots to estimate the groundwater flow processes according to the given geometry.

b) calculate the water level $z_{R(t)}$ in the GWOT for time point t =1d.

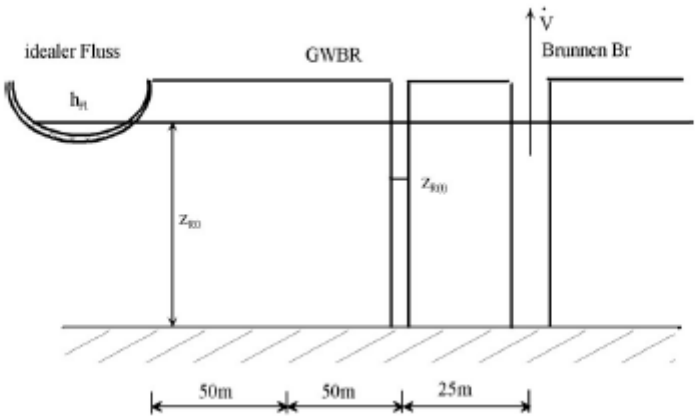


figure 9.20: aquifer with river and well

6. A numerical groundwater flow model is built for an induce recharge waterworks (see figure 9.21) with parallel flow regime. The river is to be considered as idealized boundary condition.

$k = 10^{-3}\frac{m}{s}; h_{Fl} = 15m; z_{R0} = 15m; S = 0,25; \dot{V} = 50\frac{l}{s}; q = 0,001\frac{l}{sm^2}; b = 100m;$
$k_{Kolm} = 5 \cdot 10^{-5}\frac{m}{s}; M_{Kolm} = 1m; B = 6,357m$

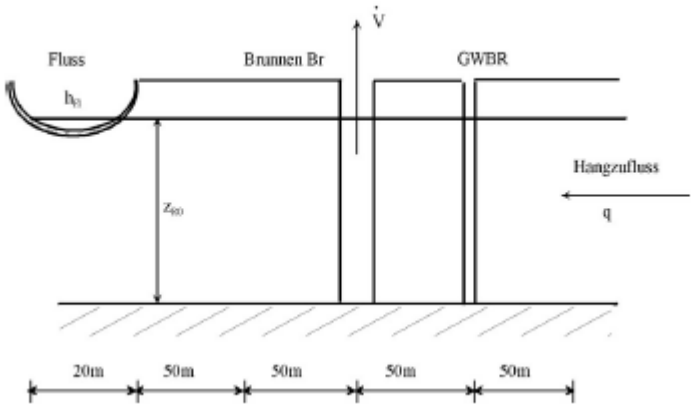


Figure 9.21: influence of river and hanging inflow on the aquifer

a) select a suitable simple quantization scheme with maximal five elements, in order to calculate the water level at gauge P in steady case most possible.
b) Formulate the balance equations at the centre of the elements and demonstrate it in matrix form.
c) calculate the hydraulic conductivity for the flow.
d) How does the equation system and the result change, if the river is not idealized, but imperfection a colmation are considered?
Outline the solution and roughly estimate the result.

Chapter 10

10 Simulation programme system ASM

10.1 Tasks

1. Simulate the drawdown for the given points with distance r and at time t, which results from water conveying V in the well for following aquifer and state the result graphically.

$$k = 1 \cdot 10^{-3}\frac{m}{s};\ M = 10m;\ S = 0,001;\ a = \frac{S}{T} = 0,1\frac{s}{m^2};\ r_0 = 0,25m;\ \dot{V} = 0,015\frac{m^3}{s;}$$

$$h_n = 16m;$$

$$r = 5m;\ 10m;\ 20m;\ 50m$$

$$t = 1min;\ 2min;\ 5min;\ 10min;\ 20min;\ 30min;\ 45min;\ 60min;\ 90min;\ 120min$$

2. Simulate the drawdown at point (r = 10m) for the above mentioned aquifer every 10 minutes until maximally 100 minutes, if that flow rate of conveying well is subject to following stagger time. And plot the solution.

Volumenstrom $\left[\frac{m^3}{s}\right]$	0,005	0,010	0,015	0,020	0,025	0,030	0,000
Förderbeginn [min]	0	10	20	30	40	50	60

3. A foundation pit should be lowered in an aquifer near a river. The centre of the foundation pit is 100m far away from the river; the drainage well is 80m. Three wells are arranged parallel to the river, which are 25m distant from each other. The diameter of wells r_0 = 0.3m and conveying capacity is V = 0.015m^3/s
The width of the river is B = 20m and a colmation layer
k' = 3 · 10^{-6}m/s; M' = 1m.
The properties of the aquifer:

$$k = 5 \cdot 10^{-4}\frac{m}{s};\ n_0 = 0,20;\ h_n = 15m;\ M = 20m.$$

Will a drawdown of 2.5m be achieved in 10 days in the centre of the foundation pit?

4. Please apply simulation programme ASM to check whether the centre of the foundation pit is drained after 7 days with conveying capacity of V = 0.01m^3/s, r_0 = 0.30m and a security of 0.5m (see figure 10.1).

5. A constant flow rate of 25 l/s is conveyed from a well, which connects an ideal river ($Br_{(100m,500m)}$). The well has a radius of r_0 = 0.35m. The aquifer is characterized by the following parameters:

$$h_n = 15m,\ M = 17m,\ k = 1 \cdot 10^{-3}\frac{m}{s},\ S_0 = 0,0002m^{-1},\ n_0 = 0,25.$$

Simulate the final steady state (the portion of temporal functionality should be smaller than 0.001) for the point ($P_{(200m,600m)}$). From when to calculate?

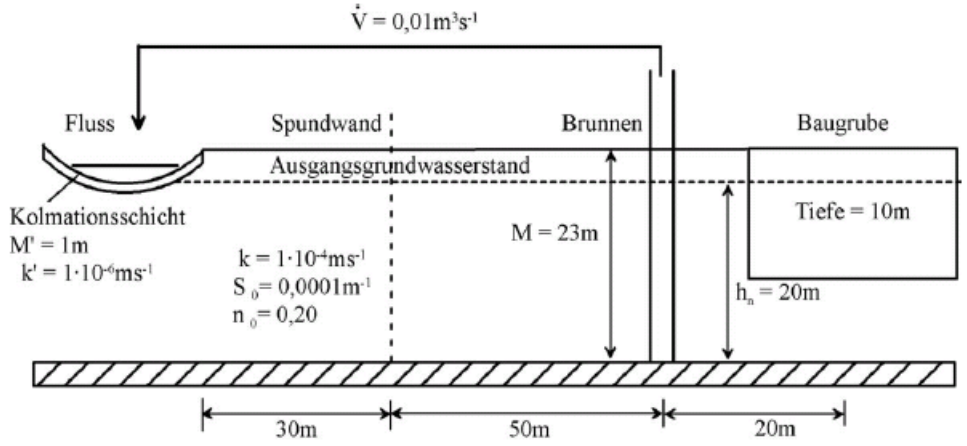


figure 10.1: aquifer with well and foundation pit

6. Simulate the following one dimensional groundwater flow:
a) By means of one dimensional steady ditch flow (see figure 10.2) simulate the position of free surface as a function of x and investigate the outflow from head water and the inflow to bottom water. Use five quantization elements.

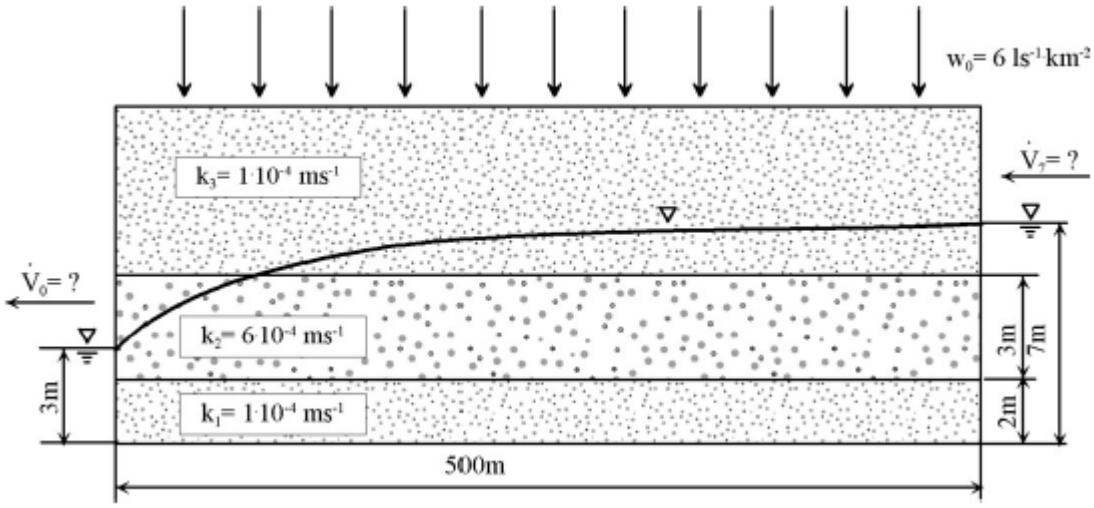


figure 10.2: stratified aquifer with steady flow regime

b) By means of one dimensional unsteady ditch flow (see figure 10.3) simulate the position of free surface as a function of x and time *t*.

7. In a aquifer a tunnel (underground) will be built parallel to a river (see figure 10.4). Simulate how does the groundwater condition for steady case change caused by this construction. Select a suitable rough quantization scheme.

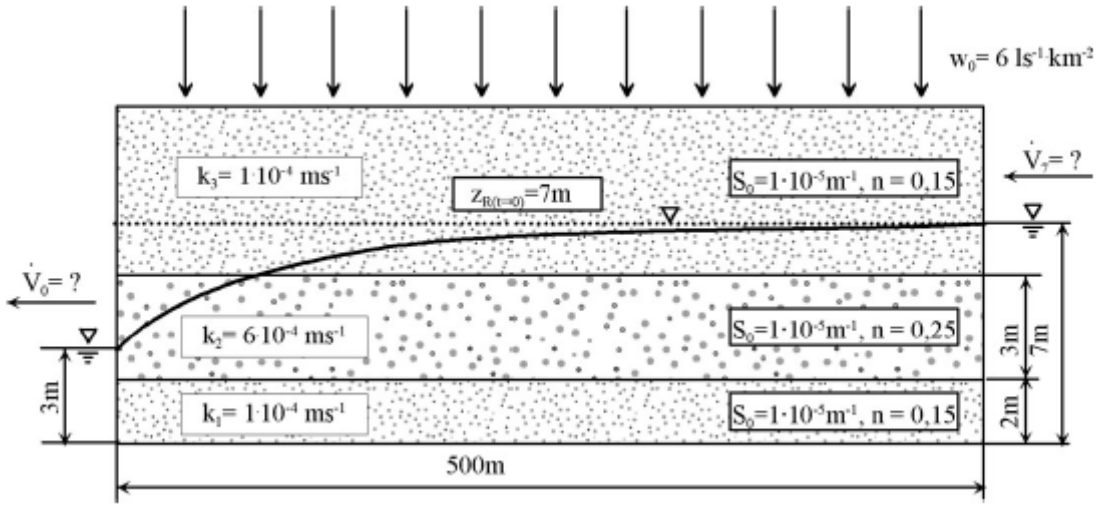


figure 10.3: stratified aquifer with unsteady flow regime

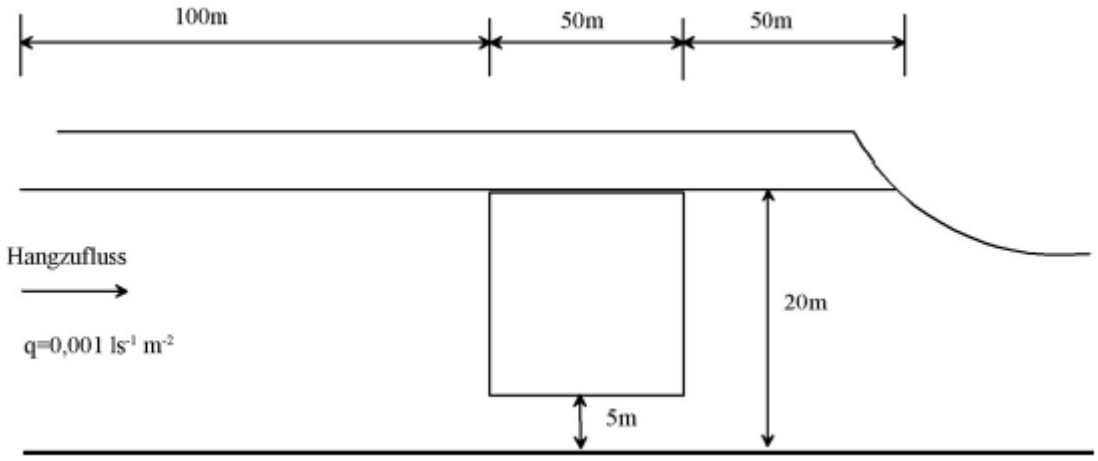


figure 10.4: tunnel construction in an aquifer

8. In a plain tract the polder area is to be protected against floods by means of a dyke construction according to simple scheme in figure 10.5.

a) Determine the time, when a steady flow regime appears, if the flood stands 5m over normal for long time.

b) How much water flows into the polder area per meter dyke length?

Dyke: $k = 10^{-4m}/_{s}$, $n_o = 0.15$, $S_0 = 0.002m^{-1}$;
Sealing material: $k = 10^{-5m}/_{s}$, $n_o = 0.05$, $S_0 = 0.001m^{-1}$

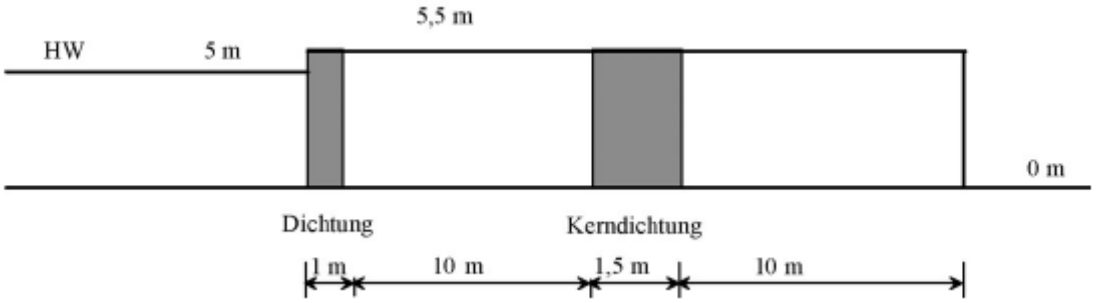


figure 10.5: dyke construction with core seal

9. Model the following horizontal aquifer by means of program system ASM, which is limited on the right and left side by two perfect complete receiving streams with a water height of 50m. The aquifer possesses a thickness of 20m, a transmissibility of $T = 0.01m^2/s$, a storage coefficient of $S = 0.001$ and a porosity of 0.1. A well with a surveying capacity of $V = 0.05m^3/s$ lies in the centre of the model area.

c) Simulate the water level distribution (contour line) after one day well surveying.

d) Graphically place the water level hydrograph curves at the well every 200m (parallel and perpendicular to the receiving stream).

e) Compute the water balance for the model area after one-day surveying, as well as the inflow from left receiving stream

f) Check the hydraulic system of the task of c) the influence of the local and time quantization increments and solution methods. First plot the hydro contour line after one day and compare.

Part III

System theory and Modelling

Chapter 11

11 Fundamentals

The system theory makes up the theoretical frame, with which the control- and feedback control engineering can be scientifically investigated. The application of this theory in the water management is particularly important for the process analysis, i.e. for the modelling, as well as for the data extraction, transmission and processing. The system theory yields the fundamentals for the terms **information** and **system**. The introduction of information concepts for physical or chemical data enables the possibility to apply different methods of computer science, cybernetics, mathematics and electro technology in planning and realization of water management monitoring-control- and automation systems.

11.1 Model classification

The modelling or the process analysis can be carried out in theoretical or experimental way.

The physicochemical processes are analysed and mathematically formulated under the help of the scientific laws in the **theoretical modelling** and **process analysis**. In this way the model structures, if possible, the model parameters of the internal influence mechanism are determinable. Analysing the objects takes place from inside to outside. The mathematical models are scientifically justified.

The input- and output signals of objects are measured and evaluated in the experimental modelling and process analysis. Natural or artificial test signals are applied. The analysis of objects takes place from the outside.

The disadvantages of theoretical modelling and process analysis are unreliability with insufficient process knowledge and high expenditure with complex processes. The disadvantages of experimental modelling and process analysis consist of the only selective model validity in contrast to the necessity in the real experimental process and the difficulty of the scientific interpretation.

Usually it is favourable to combine both methods and to a large extent the model structure are theoretically and the model parameters are experimentally determined.

The justified models play a dominant role for the migration processes. The experimental modelling above all gains importance lately. The results of the experimental process analysis, the transfer functions, are usually difficult to interpret or physically imagine with the real processes. Therefore this method often comes across baseless scepticism.

Like migration processes models the computers, hard- and software, can be also assumed as models (see figure 11.1). An important problem, which arises during processing of migration problems on computers, is the coupling of migration models to models on computers. In this connection the computer is regarded as simulators for the migration processes. However the coupling is impossible trouble free, if the models of simulators are not identical. Such differences arise e.g. in the consideration of the arguments (continuous, discontinuous) or the allocation of parameters and variables of state. In these cases an approximation must be accomplished between the two models.

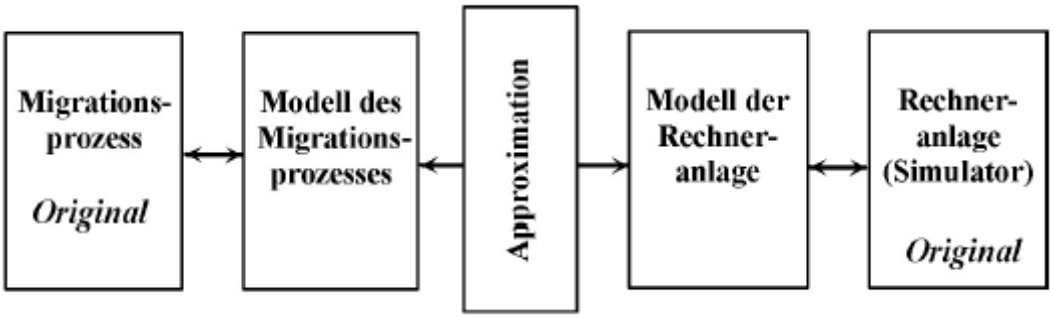


figure 11.1: coupling of migration- and simulator model

In investigation area the model basis contains different mathematical models of dynamic procedures, chemical processes and biological phenomena, i.e. the underground flow processes with the coupled material -, energy -, exchange -, and transformation processes as well as the migration processes in soil- and groundwater region. Thereby basis is generally nonlinear, and coupled material circulation. Simplified, aggregate models can be found for special cases.

KRUG classifies the mathematical models into the justified and describable models (see figure11.2) according to the model development background.

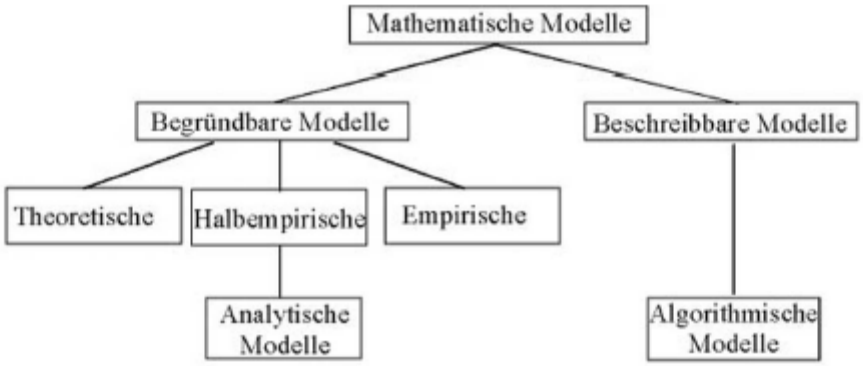


Figure 11.2: model classification by KRUG

The describable models are used e.g. for ecological systems and population problems. For the modelling of technical systems the class of the justified models is meaningful.

These are classified in:

● linear - non-linear

● continuous - discontinuous and

● dynamic - static

The question of modelling is in connection with the process control extremely important. The achievable quality of the control- and feedback control task solution depends particularly on whether sufficient qualitative and quantitative knowledge of plant controlled system, here the soil and groundwater region, are available. The concept of modelling must be considered closely connected with process analysis. TÖPFER/BESCH suggests following classification principles for model application in the automation technical investigation area.

Model Classification according to:

● **Model extraction method**

Theoretical model extraction/process analysis (laws of nature)
Experimental model extraction/process analysis (experiments)

● **Model purpose of use**

Construction-, calculation-, behaviour model
handling-, function model

● **Model notation**

Mathematical models in equation form/parametric models (equations)
Mathematical models in graphic form (signal flow chart), non parametric models (curve, pair of variates)
Physical models (analogy model, graphic model)

● **Model conclusion**

Static models
Dynamic models

● **Model adaptability**

Prediction models
Adaptive Models
Adapted Models

● **Relationship of variables**

Deterministic/stochastic models
Linear/non-linear models

● **Model validity**

Type models (for classes of Objects)
Special models (for concrete object)

The term model is usually not used uniformly. We can understand it as a triangle relation between **model**, **object** and **subject**. The model is characterized not only by, about what the model is, but also by for what it is. The object is also denoted by model original. The relationship between model and original is always a kind of image relation. The quality of a model is measured by

● relative compatibility in view to the subject

● relation fidelity of the image and the behaviour

● application simplicity

The status of process analysis, model and simulation within the scope of control and regulation is shown in the following figures 11.3 and 11.4. Thereby the model formulation of the soil- and groundwater processes should be understood in further process analysis. This can be also called procedure modelling. In contrary the progress of original process is reproduced in the simulation. The necessary parameters and variables of state are communicated to models and the process cycle starts.

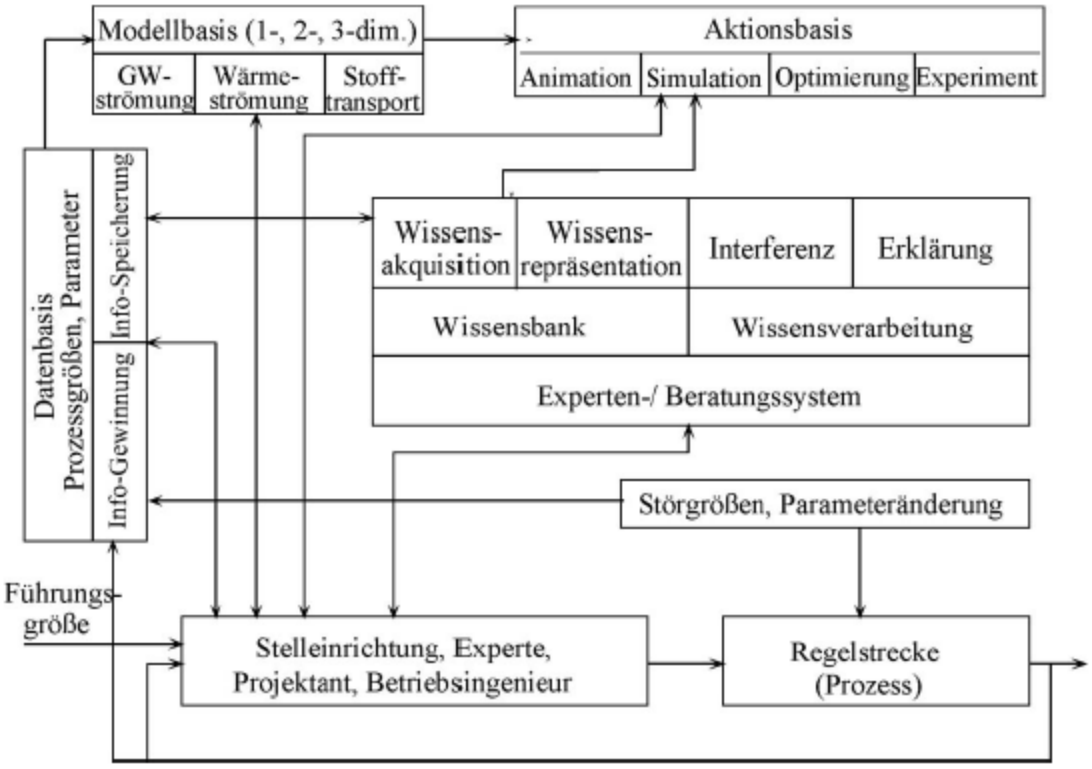


figure 11.3: devices- and technical programming realization

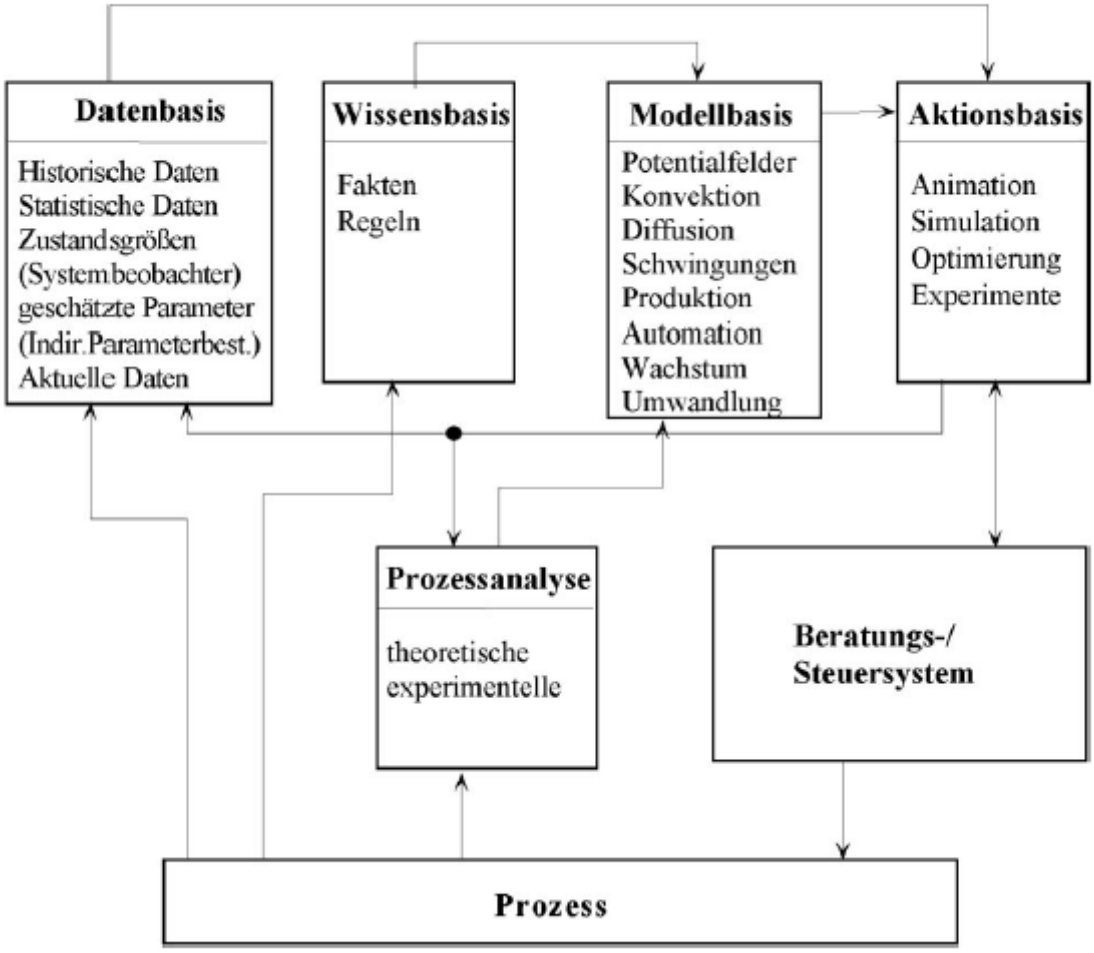


figure 11.4: process character of modelling and simulation

11.2 Methods of process analysis

By means of process analysis methods a mathematical description of the behaviour of transfer elements is to be extracted. This process is also denote by modelling. These models can be attained in two different ways, on one hand by means of theoretical, on the other hand by means of experimental Process analysis. While the theoretical process analysis yields the model structure, the related parameters must be determined or improved by the experimental analysis. In contrast to the theoretical proceeding an investigation of the output signals takes place as system response to input signals in the experimental process analysis. Both methods form a unity and complement each other, because an experimental analysis without theoretical advance information and a theoretical analysis without experimental supporting are hardly feasible.

11.2.1 Theoretical Process analysis

With the **theoretical process analysis** based on internal structure the transfer elements are tempted in order to find the mathematical descriptions (models) between in- and output variables. Transfer functions, which are formed on the basis of the theoretical process analysis, are always justified by natural laws. They always possess physical or chemical bases in water management application.

The characteristics of the theoretical process analysis consist of the fact that:

- the model can be already carried out before the practical realization,
- the analysis results with same process type are transferable,
- the connections between technological and constructional data remain,
- the process determinant variables are identified in the system and
- important statements about the model structure are attained.

The difficulties of this method consist of the fact that:

- the expenditure is very high and the models are complicated,
- the necessary process parameters can be achieved often very hard and only inaccurately,

- the proceeding is poor with respect to algorithm and

- the physicochemical process must be acquainted sufficiently.

For setting up mathematical models within the scope of theoretical process analysis a fact has been proved that, large systems by division into subsystems and then by individual balance equations (mass-, energy and momentum conservation law as well as source- and sink activities) are analysable.

11.2.2 Experimental process analysis

The **experimental process analysis** in contrast to the theoretical assumes the investigation of system in- and output signals. The systems are thereby regarded as transfer elements. Artificial experiments are in progress at the original system, and special attention must be dedicated to the choice of input signals. If the execution of experiments is not possible, also natural phenomena (e.g. flood waves) can be used as database. We also mention that the experimental process analysis checks the systems from the outside.

The methods of experimental process analysis are also known as black box method in cybernetics.

Both methods form a unity and complement each other, because an experimental analysis without theoretical advance information and a theoretical analysis without experimental supporting are hardly feasible.

In table 11.1 some selected characteristics of the two kinds of process analysis are compared.

Table 11.1: comparison of theoretical and experimental process analysis

Modelleigenschaften	Theoretische Prozessanalyse	Experimentelle Prozessanalyse
Struktur	Zustandsmodelle mit innerer Struktur des realen Prozesses	Ein-/Ausgangs- und Signalmodelle vorwiegend ohne Struktur des realen Prozesses
Parameter der Modelle	Zusammenhang mit konstruktiven technologischen Daten	Kein Zusammenhang zwischen den Modell-parametern und konstruktiven, technologischen und anderen Daten
Modellvereinfachung	schwierig	leicht möglich
Gültigkeit der Modelle	in großen Bereichen und für verschiedene Prozesse	im untersuchten Bereich und für den speziellen Prozess
Aufwand	Modelle häufig zu aufwändig für Lösung der Diagnose-, Überwachungs-, Steuerungs- und Vorhersageaufgabe	Modelle sehr günstig für die Lösung der Diagnose-, Überwachungs-, Steuerungs- und Vorhersageaufgabe
Zeitpunkt	Modell bereits in der Entwurfs-/ Projektphase erstellbar	Modell erst mit Existenz/ Inbetriebnahme des Prozeses erstellbar
Beeinflussung des Prozesses	Keine "Störung" des Prozesses bei der Modellerstellung	Häufig "Störung" des Prozesses erfor-derlich (aktives Experiment)

11.3 Signal representation

Processes can be characterized by means of signals. In case of water management processes it means that these can be described by in- and output variables (e.g. water level, flow, chemical concentrations, temperature).

We call a function carried by physical variables **signal**, if it has a parameter, the image function is a variable of the physicotechnical space. In principle a signal is represented by a four dimensional function *x = f(x, y, z, t)* mathematically. In the mathematical description of signals we consider the double meaning of symbol "*x*". It acts as general signal note and as character of local coordinate. Often it is more favourable to use as abbreviation of each physical variable than signal characteristics. Signal Parameters are called **information parameters**, and physical variables are **signal carriers**, by which the signal is carried. Examples are specified in table 11.2.

Table 11.2: classification of information parameter and signal carrier

Anwendung	Informationsparam.	Signalträger
Elektrizität	$220V$	elektrischer Strom
	$1A$	elektrischer Strom
Hydraulik	$10m$	Flusswasserstand
	$1m^3$	Wasserstrom
Thermik	$273K$	Wärmepotential
	$1kW$	Wärmestrom

In communications technology such signals are used e.g. in the form of voltage states, current and power changes. According to above definition signal and information concept can be also applied in other technical systems, like here in water technical processes. For example the water level, the flow rate and the temperature also appear as signal carriers analogously to current and voltage. Therefore the water level is expressed as three dimensional signal *x = f(x, y, t)*. In addition, chemical material concentrations are conceivable as signal.

For more simple representation usually only time is mentioned as independent variable in the following consideration. This should be without loss of generality. The implementations also apply exactly to the dependence concerning local coordinates.

The description of signals can be done with a diagram and by means of mathematical functions. The signals are usually defined with a start time $t = 0$. It is a matter of relative time to an event. In the mathematical description the original procedure in the so called time domain is usually distinguished from the transformed procedure in complex variable domain. Common transformations for signals are FOURIER- and LAPLACE transformation.
Different representations have proved their worth for the mathematical description of technical signals. The basic signals are of great importance (see section 11.3.1 basic signal forms, page 295), since they are the basis of all arbitrary signal forms. By means of basic signals arbitrary

signals can be generated (see section 11.3.3 signal synthesis, page 301). Likewise arbitrary signal forms can be decomposed into these basic signals (see section 11.3.4 signal analysis, page 301).

11.3.1 Basic signal forms

The most common basic signal forms (see figure 11.5) are the **identical magnitude**, **sine function**, **step function** (unit step) and the **DIRAC impulse**.

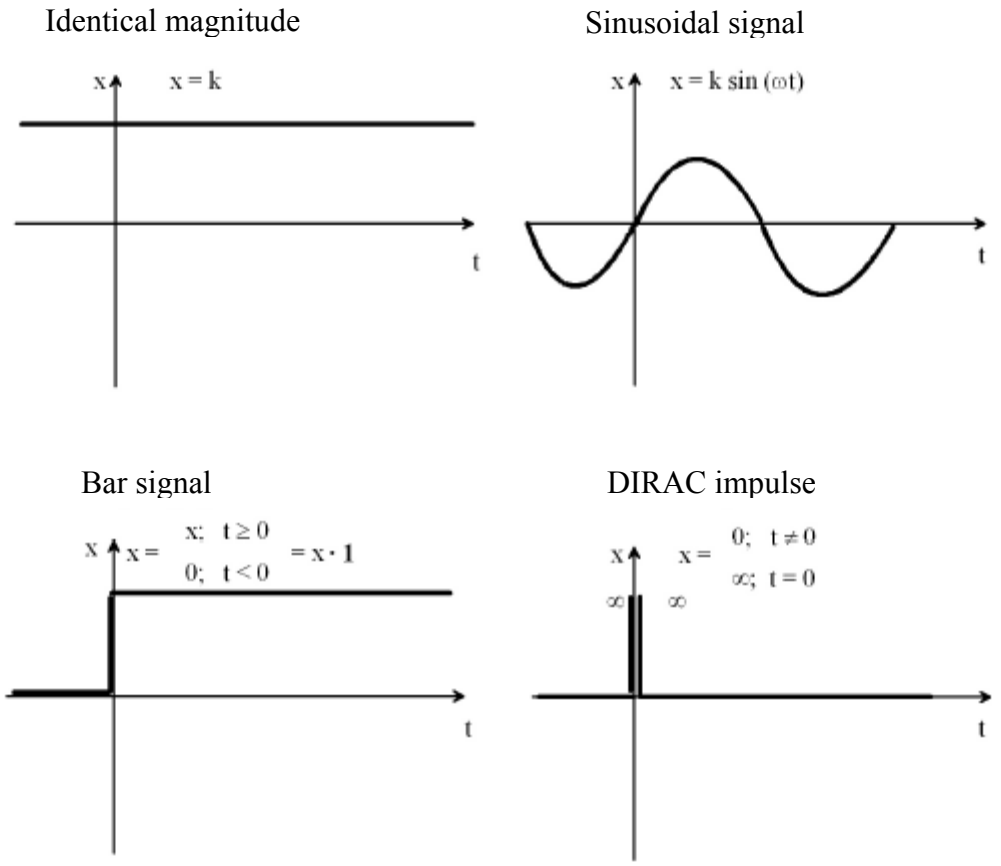


figure 11.5: basic signal form

The unit step is a normalised signal with step height (step height = 1) and is represented by 1(*t*). The DIRAC impulse only affects at $t = 0$ and has an infinite step height there. Unit step and DIRAC impulse are connected with each other mathematically by the integration or deviation.

11.3.2 Application of selected test signals

Test signals can be effectively used for the execution of the experimental process analysis (see figure 11.6). With these a special experiment must be accomplished to the real object. It is possible under different technical or technological conditions that only special test signals can be used. Since the same system description develops independently of test signal type and the different descriptive models are transferable, there are no restrictions in the experimental process analysis method.

Apart from the application of these special test signals the system reaction of natural events, i.e. natural signals, like flood waves, precipitation events etc. can be also drawn on for the determination of transient characteristic. This will be realized by the application of faltung operation (see to section 12.3 arbitrary transient characteristic, page 355).

There are following special definitions for the test signals according to TÖPFER, whereby the auxiliary term "**unit -**" always contains a normalisation to the value one.

11.3.2.1 Impulse function

The impulse function is defined as:

$$x_e(t) = \left\{ \begin{array}{lll} 0 & \text{für} & t < 0 \\ \dfrac{A}{\Delta t} & \text{für} & 0 \le t < \Delta t \\ 0 & \text{für} & t > \Delta t \end{array} \right\}, \qquad (11.1)$$

whereby Δt is the impulse width. The area of impulses amount to:

$$A = \int_0^{\Delta t} x_e(t)dt \qquad (11.2)$$

The area for a impulse with constant height and a definite impulse duration:

$$A = x_e \cdot \Delta t \qquad (11.3)$$

The impulse area embodies an appropriate effect in form mass- or energy deposit. A finite pulse width will be always available with technical impulses. If the pulse width is smaller than a tenth of the smallest time constant ($\Delta T < 0.1\text{г}$) (see section 12.2 second order transient characteristic, page 340), then it can be regarded as an approximately ideal impulse.

If we presuppose the fact that the area remains constant, and for $\Delta t \rightarrow 0$ the pulse amplitude must approach to infinite $x_e \rightarrow \infty$ (cp. also see section 12.2.3 DIRAC Impulse as input signal, page 349),

$$\lim_{\substack{\Delta t \longrightarrow 0 \\ x_e \longrightarrow \infty}} x_e \cdot \Delta t = A \tag{11.4}$$

$$x_e(t) = \left\{ \begin{array}{lll} 0 & \text{für} & t < 0 \\ \infty & \text{für} & t = 0 \\ 0 & \text{für} & t > 0 \end{array} \right\},$$

whereby the impulse area has a definite value:

$$A = \int_{-0}^{+0} x_e(t)dt \tag{11.5}$$

the so called **DIRAC impulse**, also designated as **unit impulse**, arises when the impulse area is normalised to value one:

$$\delta(t) = \frac{x_e(t)}{A} = \left\{ \begin{array}{lll} 0 & \text{für} & t < 0 \\ \infty & \text{für} & t = 0 \\ 0 & \text{für} & t > 0 \end{array} \right\}$$

$$A_{\delta(t)} = \int_{-\infty}^{+\infty} \delta(t)dt = 1 \tag{11.6}$$

11.3.2.2 step function

The step function is defined as:

$$x_e(t) = \left\{ \begin{array}{ll} 0 & \text{für} \quad t < 0 \\ x_{e_0} & \text{für} \quad t \geq 0 \end{array} \right\}$$

If the step height is normalised, we get the **unit step**:

$$1(t) = \frac{x_e(t)}{x_{e_0}} = \left\{ \begin{array}{ll} 0 & \text{für} \quad t < 0 \\ 1 & \text{für} \quad t \geq 0 \end{array} \right\} \qquad (11.7)$$

Please note that the bar signal already takes value x_{e0} at the time $t = 0$.

11.3.2.3 ramp function

The ramp function, also designated as slope function, is defined as:

$$x_e(t) = \left\{ \begin{array}{ll} 0 & \text{für} \quad t < 0 \\ c \cdot t & \text{für} \quad t \geq 0 \end{array} \right\}$$

A unit function can be also generated here by normalization. The **unit ramp function**:

$$\alpha(t) = \frac{x_e(t)}{c} = \left\{ \begin{array}{ll} 0 & \text{für} \quad t < 0 \\ t & \text{für} \quad t \geq 0 \end{array} \right\} \qquad (11.8)$$

Among the different illustrated test signals, in particular the unit signals, such connection exists that they are transferable with each other by integration or deviation (see table 11.3).

Table 11.3: relationship between different basic signals

	Einheitsimpuls DIRAC-Impuls $\delta(t)$	**Einheitssprungsignal** $1(t)$	**Einheitsrampensignal** $\alpha(t)$
Einheitsimpuls DIRAC-Impuls $\delta(t)$		$\delta(t) = \frac{d1(t)}{dt}$	$\delta(t) = \frac{d1(t)}{dt} = \frac{d^2\alpha(t)}{dt^2}$
Einheitssprungsignal $1(t)$	$1(t) = \int \delta(t)\,dt$		$1(t) = \frac{d\alpha(t)}{dt}$
Einheitsrampensignal $\alpha(t)$	$\alpha(t) = \int 1(t)\,dt$ $= \iint \delta(t)\,dt$	$\alpha(t) = \int 1(t)\,dt$	

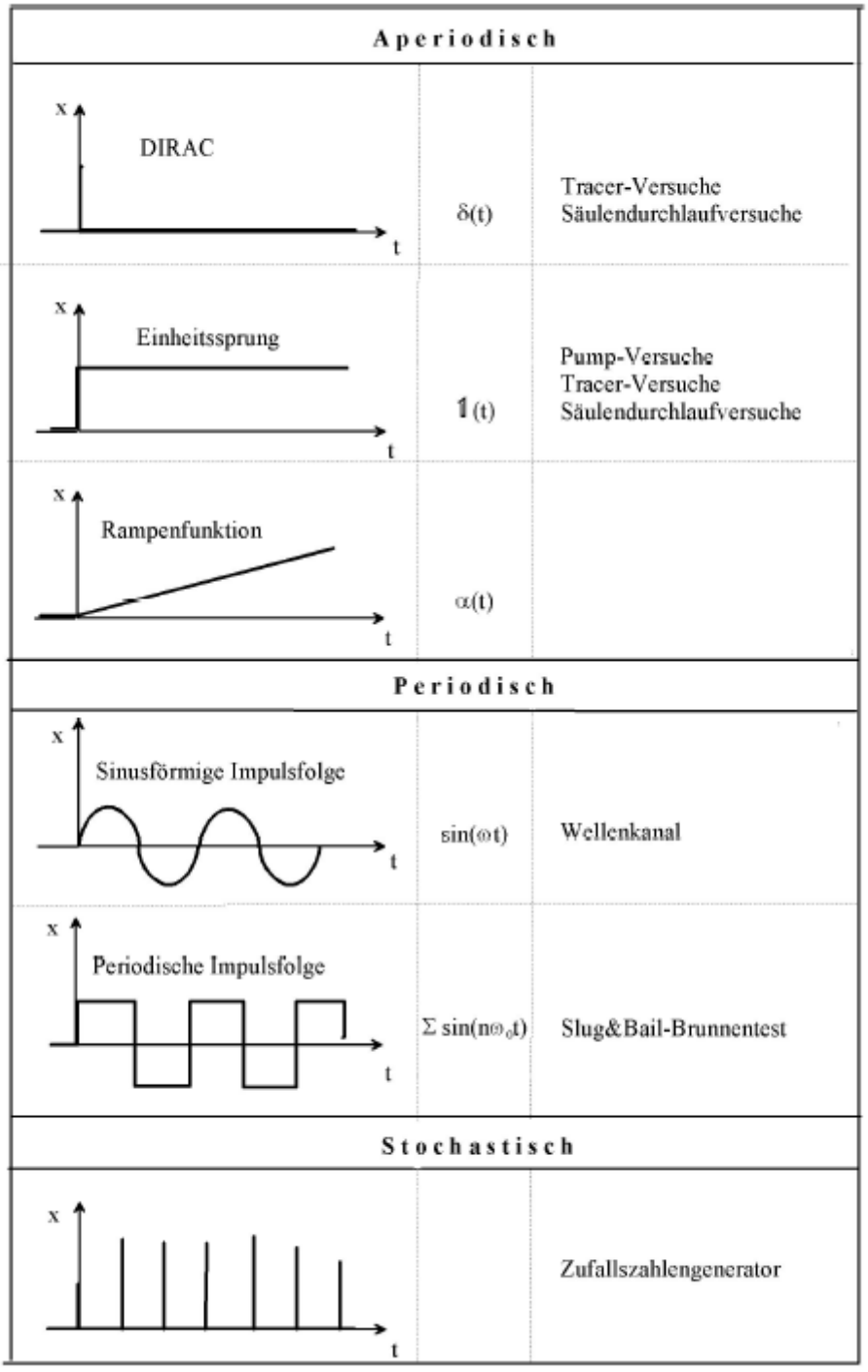


figure 11.6: test signals

11.3.3 Signal syntheses

Several input signals x_e are additively merged to an output signal x_a at a **mixing place** in the signal **synthesis** (overlay).

$$x_a = \sum_{t=1}^{n} x_{et} \qquad (11.9)$$

The output signal can be determined based on mathematical equation of the mixing place by a signed addition or by means of graphic methods (see figure 11.7).

11.3.4 Signal analysis

The **signal analysis** of arbitrary signal forms can be achieved via a decomposition in basic signal forms. The most common method is Fourier series decomposition, with which periodic signal sequences are approximated by different frequencies sinusoidal oscillations. A practically well manageable method with for deadbeat signals, i.e. one time expiring signals, is the approximation through temporally staggered signals. In the following the graphic method should be described, since in contrast to mathematical one it is substantially more simply manageable and more descriptive.

The first step with signal analysis by means of bar signals consists of the fact that arbitrary time function is approximated by a echelon form signals (see figure 11.8). The time of echelon flank and the step height should be selected in such a way that the smallest mean error occurs.
It has to be proved the integrals of the original curve and the step function draw near. It means that the same areas must be represented by both curves (see script for lecture groundwater measuring technique, section error calculation).

Arbitrary signals can be also decomposed into a sum of impulses, particularly into an infinite sum of DIRAC impulses. This leads to faltung integral method (see section 12.3 arbitrary transient characteristic, page 355).

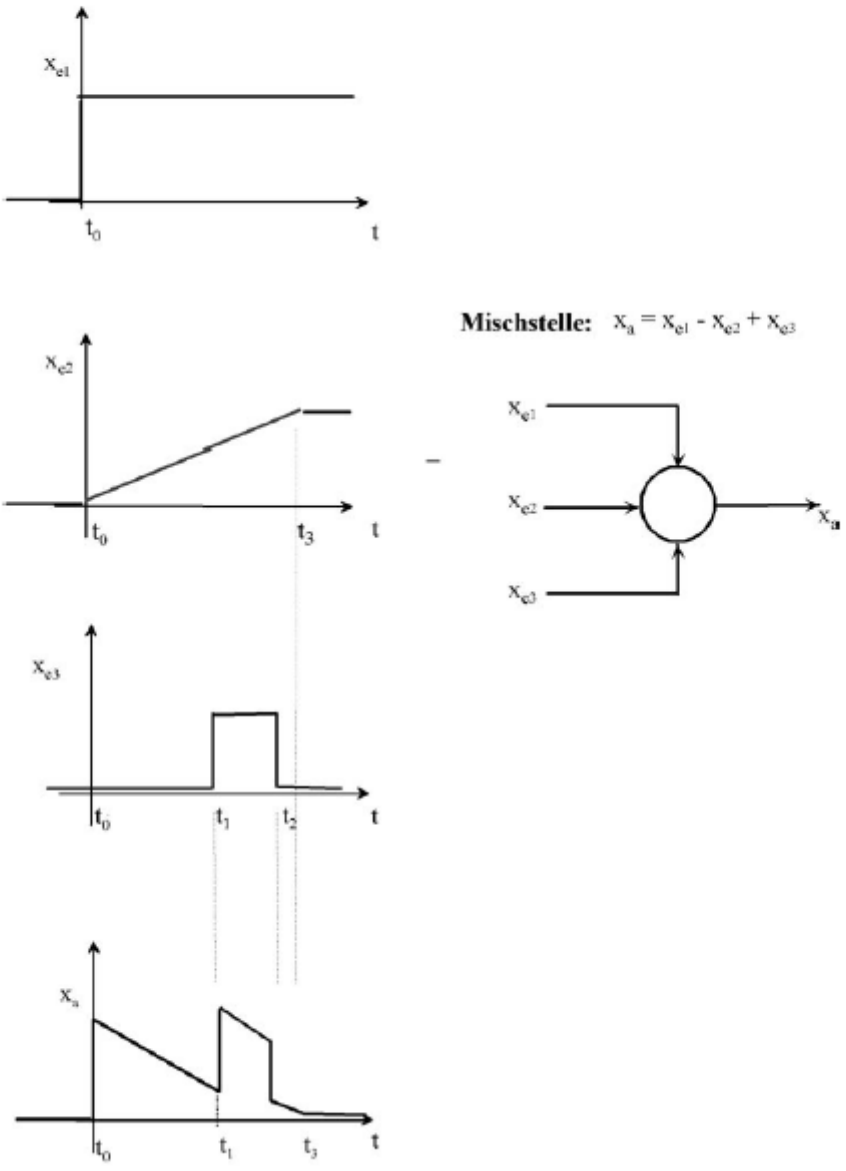


figure 11.7: signal syntheses

Example:

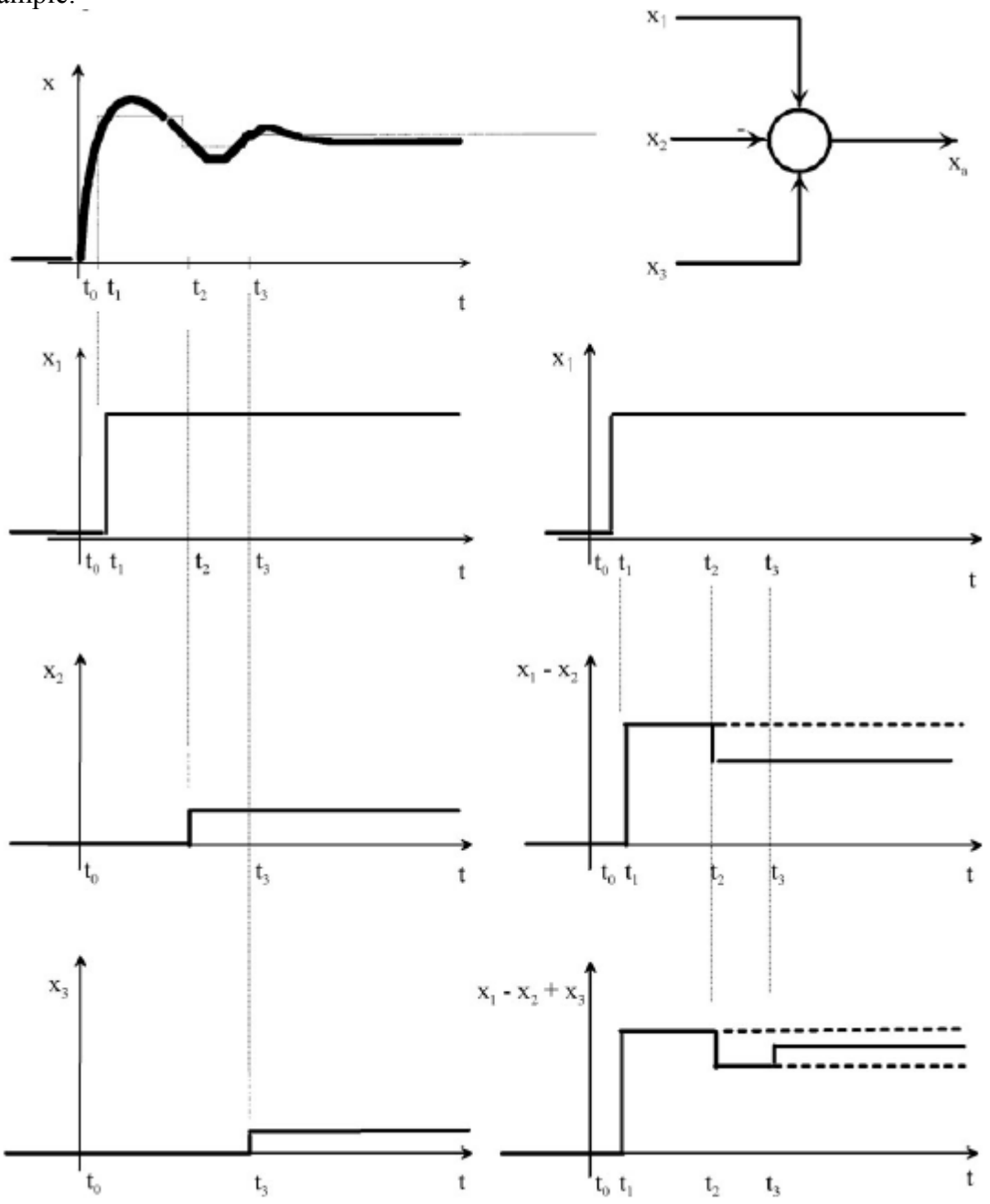


figure 11.8: approximation arbitrary signals by bar signals

Here it should be pointed out again that the exemplary signal representation is applicable to all arguments as time function. The special place dependence in x_i and y_i direction plays a important role in soil and groundwater range processes as well as in contaminated site treatment.

11.3.5 Quantization

The display format of signals can be classified according to different criteria. With respect to technology we differentiate signal classifications according to **information parameter quantization** and according to **independent variable quantization**.

In the classification according to **information parameter quantization** we get:

- **analogy signals**, whose information parameter can be assumed any intermediate value in a metric magnitude and

- **discrete signals**, whose information parameter can be assumed only definite (finitely many) values within certain limits.

The two mentioned classification principles can be also combined and we get the display format shown in the figure 11.9.

The following subdivisions can be made for the class of **independent variable quantization**:

- **continuous** signals, for those the information parameter to any value
and

- **discontinuous** signals, for those the information parameter can be only indicated for finitely many values of arguments.

Some measuring instruments and methods are specified in table 11.4 as examples of appearances of different signal forms.

Quatization of independent variables

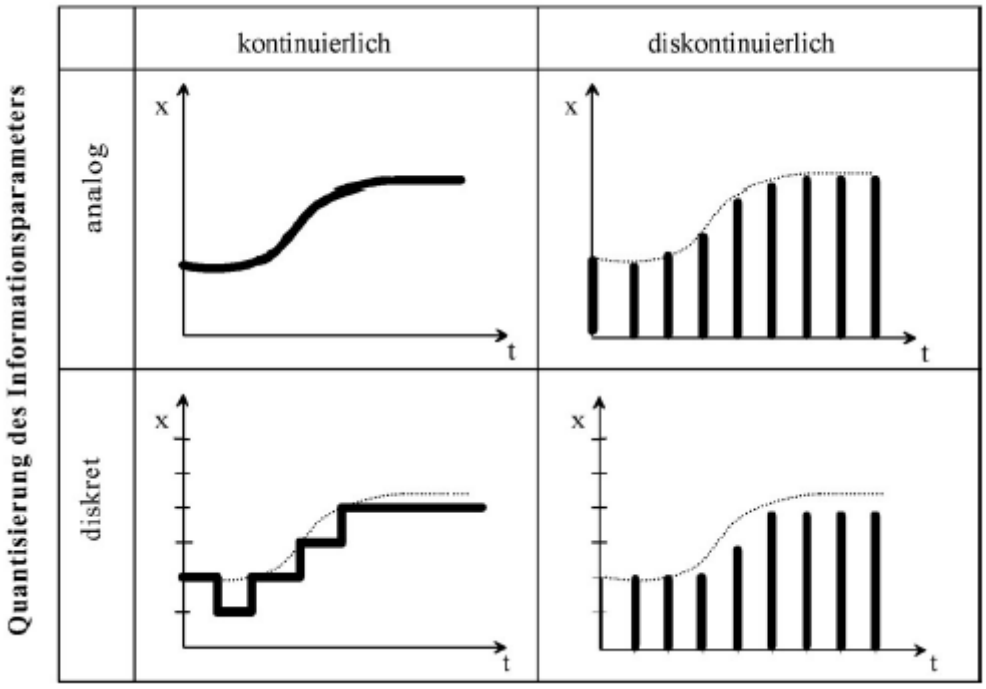


figure 11.9: representation forms of signals

It is still pointed out that the use of term cannot be always kept exactly as the German industry standards specified. These will be after all conditional due to different application of some terms in the foreign language literature. Thus there are no clear separation between the terms "discrete" and "discontinous". The word "discretisation" is often used for "independent variable quantization". Also the term "digital" (digital signals) is often used for discrete measured values, if they are indicated by means of number tablets. **Digital** measured values are rightly measured values from an information parameter quantization, whereby only the quantization stages "0" and "1" (or "0" and "L") are allowed.

Table 11.4: Measuring instruments and methods and signal forms

Signalform	Elektrotechnik	Geohydraulik
analog-kontinuierlich	X-Y-Schreiber Plotter	Schreibpegel
analog-diskontinuierlich	Punktschreiber Fallbügelschreiber	Tiefenlot
diskret-kontinuierlich	Digitalvoltmeter (Stufenverschlüsselung)	Widerstandsmesskette
diskret-diskontinuierlich	Digitalvoltmeter (Zeitverschlüsselung)	Brunnenpfeife

11.4 Transmission systems

The following methods serve the determination of dynamic behaviour of undisturbed systems. Since this is only technically idealized possibility, it must be required that the input signals substantially dominate compared to the disturbing signals.

Further important prerequisites for application of the methods are:

- the system is **undisturbed** (disturb << wanted signal).
- the system behaviour is **linear**.
- the system behaviour is **time-invariant.**
- the system has only **one** input- and **one** output signal.
- the system behaviour is describable by **concentrated** parameters.

In the system theory, particularly in connection with the experimental process analysis, each process can be represented as so called "black box", which is only characterized by the relation of in- and output variables. This "black box" is then designated as system.
A system is always identified by the boundary to its environment and coupled information exchange (see figure 11.10). With respect to the system theory we differentiate between concrete and mathematical systems. A **concrete system** is a spatially delimited part of reality, including some selected connections in its internal structure and its environment. The **mathematical systems** contain variables, equations or operators.

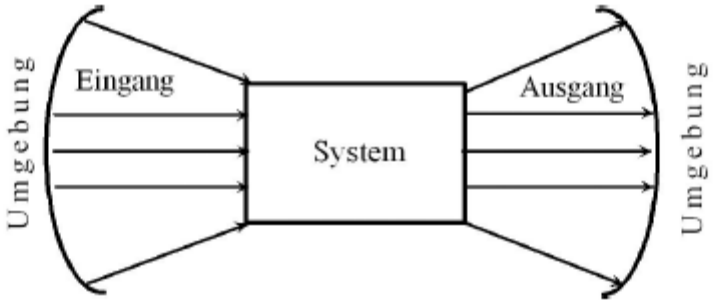


figure 11.10: system with its connection to environment

11.4.1 Mathematical description

Different methods emerge in the description of systems. The approach in connection with technical systems as **transmission systems** is most widely used. Each system is identified by a input quantity, which is a quantity dedicated by output variables (see figure 11.11). This functional connection between the out- and input variables is called **transient characteristic**. In the following for each case only the relation of a input- or an output variable is regarded. System with several in- and output variables, so called multi signal systems, can be treated with coupled equation system method.

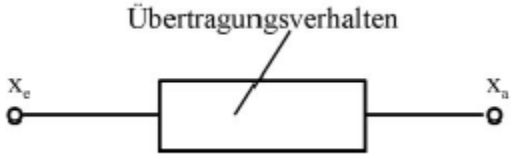


figure 11.11: transfer element

The description of the systems by the influence in- and output signals can take place in diverse forms. The most common is the mathematical equations and the time response diagram of step response function. Three kinds of definition equations are used for the description of transient characteristic. According to the basic signal relation, which are applied at input, we get the transit-, weight- and transfer function:

- The **transit function h(t)**
  we get if a unit bar signal is applied at input. This is also designated as step response function:

$$h(t) = x_a \mid_{x_e = 1(t)} \tag{11.10}$$

- The **weight function g(t)**
  is yielded if a DIRAC impulse act on input, and is also designated as impulse response function:

$$g(t) = x_a \mid_{x_e = \delta(t)} \tag{11.11}$$

- The **transfer function G(p)**

IS YIELDED FROM LAPLACE TRANSFORMATION DESCRIPTION OF OUTPUT SIGNAL IF THE INPUT SIGNAL IS DIRAC IMPULSE:

$$G(p) = X_a \mid_{X_e = L\{\delta(t)\}} \tag{11.12}$$

We differentiate between the time- (original-) domain and the complex variable domain in the description of the transient characteristic (see figure 11.12), with signals, and the transfer function, a transformation is subordinated. The most common integral transformations are FOURIER and LAPLACE transformations (see section 5.3.2 LAPLACE transformation, page 135). The advantage of the application of transformations consists of the fact that complicated arithmetic operations can be usually simplified with transfer functions in the complex variable domain based on four basic arithmetic operations. The disadvantage includes the poor descriptiveness of the complex variable domain as well as the expenditure to transform signals and mathematical models into complex variable domain and back again after solving the transfer function (inverse transformation). While prefabricated correspondences usually exist for the forward direction, the inverse transformation often proves more complex.

The designations of signals and transfer functions are lowercase letters in the time domain, in contrast capital letters in complex variable domain.

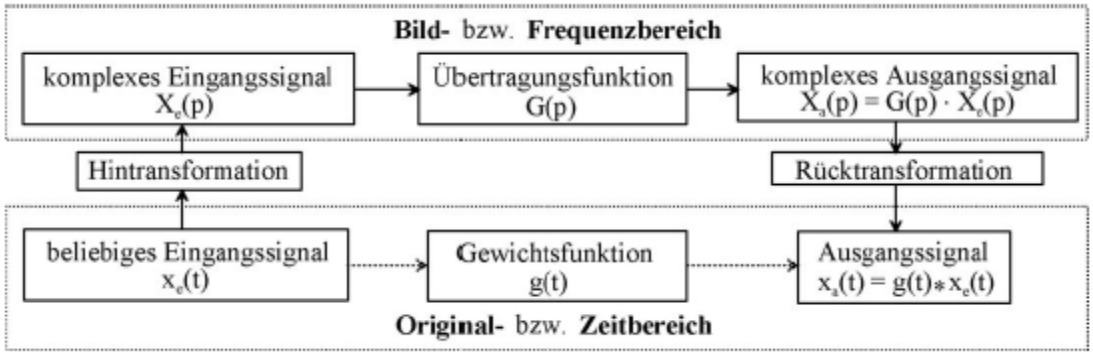


figure 11.12: relationship between time- and complex variable domain

The **FOURIER transformation** and its inverse transformation are defined as:

$$\text{FOURIER-Transformation} \qquad X(j\omega) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \tag{11.13}$$

$$\text{Inverse transformation} \qquad x(t) = F^{-1}\{X(j\omega\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t}d\omega; \tag{11.14}$$

The **LAPLACE transformation** and its inverse transformation are defined as:

$$\text{LAPLACE-Transformation} \qquad X(p) = L\{x(t)\} = \int_{0}^{\infty} x(t)e^{pt}dt \tag{11.15}$$

$$\text{Inverse transformation} \qquad x(t) = L^{-1}\{X(p\} = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} X(p)e^{pt}dp \tag{11.16}$$

While the FOURIER transformation is favourable for periodic or periodised signals, the LAPLACE transformation has been proved its worth for application to bar signals (e.g. switching operation) and impulses (DIRAC impulse). Particularly the LAPLACE transformation is also used for the computation of transient characteristic of arbitrary input signals (see section 12.3 faltung operation, page 355).

These three types, transition -, weight- and transfer function, the mathematical representation of transient characteristic, are equivalent, since they are only different mathematical computation forms for the same physical or chemical technical processes. They are therefore transferable with each other by means of mathematical connection (see table 11.5). According to definition, an integral or differential connection exists between unit bar signal and DIRAC impulse (also see table 11.3, page 299), and the arithmetic rules for LAPLACE transformation (see section 5.3.2 LAPLACE transformation, page 135), yield the connections between different description types.

Table 11.5: relationship between different functions of transient characteristic

	h(t)	**g(t)**	**G(p)**
h(t)		$h(t) = \int_0^t g(\tau) d\tau$	$h(t) = L^{-1}\left\{\frac{1}{p} G(p)\right\}$
g(t)	$g(t) = \frac{dh(t)}{dt}$		$g(t) = L^{-1}\{G(p)\}$
G(p)	$G(p) = p\, L\{h(t)\}$	$G(p) = L\{g(t)\}$	

The methods of experimental process analysis introduced in the following can be only applied under definite conditions. Usually these conditions can be met in the water management processes investigation. Advanced methods are subject to special literature or are still the research subject nowadays.

11.4.2 Basic transient characteristic

The **basic forms** of technical system transient characteristic can be described in **proportional**, **integral** and **differential** forms as well as **delay-** and **duration behaviour.**

The system designations concerning the transient characteristic are done by capitalized initial letters of the characteristic or by the system response pictogram on a bar signal at input, i.e. via step response. Examples of the particular transient characteristic are summarized in table 11.6.

Table 11.6: transient characteristic

Übertragungsverhalten	Bezeichnung	Beispiel
proportional	**P**	Hebelanordnungen
integral	**I**	Füllung eines Behälters $(V = const)$
differential	**D**	Druckstoß in Rohrleitungen
Verzögerungsverhalten 1.Ordnung	$\mathbf{T}_1$	Füllung eines Behälters $(V = f(t))$ Strömungsprozesse (Rohrleitung, Grundwasser)
Verzögerungsverhalten 2.Ordnung	$\mathbf{T}_2$	Füllung von Kaskadenbehälter Transportprozesse (Rohrleitung, Grundwasser)
Laufzeitverhalten	$\mathbf{T}_L$	Förderband, Mischrohr

These basic forms will be described exemplary based on some examples. More detailed descriptions for 1st and 2nd delay elements are given in chapter 12 model regulation, page 333.

11.4.2.1 Proportional characteristic => **P**-element

Example: first class lever

Law of lever:

$$l_a \cdot F_a = -l_b \cdot F_b \tag{11.17}$$

l length of the fulcrum
F force at the fulcrum

If we set the forces as input- and the others as output signals, we get proportional transient characteristic:

$$F_a = -\frac{l_b}{l_a} \cdot F_b \tag{11.18}$$

$$x_a = K \cdot x_e \tag{11.19}$$

The **transfer factor K** can be determined either in way of theoretical process analysis from geometrical conditions of the lever arms:

$$K = -\frac{l_b}{l_a} \tag{11.20}$$

or by an experiment, the experimental process analysis, by means of known input signals, e.g. the test signal unit step 1(t), and measured output signal. In this case:

$$K = \frac{x_a}{x_e} \qquad \text{bzw.} \tag{11.21}$$

$$K = x_a \,|_{x_e = 1(t)} \tag{11.22}$$

The transient characteristics are determined by inserting the appropriate input signals into the equation 11.19 as follows:

- **transit function h(t)**

  According to definition $x_e = 1(t)$ will be inserted. Thus:

$$h(t) = x_a \,|_{x_e = 1(t)}$$

$$= K \cdot 1(t)$$

$$h(t) = K \tag{11.23}$$

● **weight function g(t)**
In this case the DIRAC impulse will be used as input signal:

$$g(t) = x_a \mid_{x_e=\delta(t)}$$

$$g(t) = K \cdot \delta(t) \qquad (11.24)$$

● **transfer function G(p)**
The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function:

$$G(p) = X_a \mid_{X_e=L\{\delta(t)\}}$$

$$= L\{K \cdot \delta(t)\}$$

$$= K \cdot L\{\delta(t)\}$$

$$G(p) = K \qquad (11.25)$$

Since only linear systems are regarded according to prerequisite, K = is const. Otherwise $L\{\delta(t)\} = 1$.

11.4.2.2 Integral characteristic => **I**-element

Example: Filling procedure with constant flow rate
The filling of a container with surface area *A* by a constant flow rate V leads to a rise of water height *H* in this container (see figure 11.13).

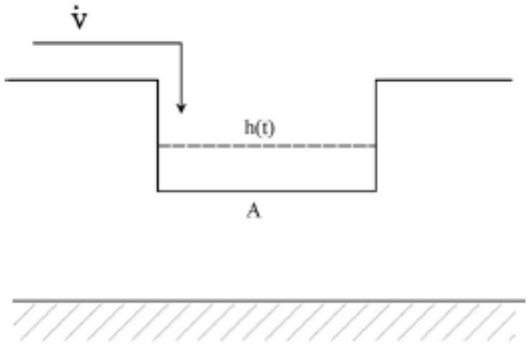


figure 11.13: Filling procedure with constant flow rate

If we check the dependence of the rising of water height H on the influent flow rate V, then we find out the following relation:

$$V = A \cdot H \tag{11.26}$$

$$V = \int_0^t \dot{V}\, dt \tag{11.27}$$

If we equate these two volumes and solve the equation respective to H, then we get:

$$H = \frac{1}{A} \int_0^t \dot{V}\, dt \tag{11.28}$$

For $V = const$.

$$H = \frac{1}{A} \dot{V} \cdot t \tag{11.29}$$

If we consider again the system as transfer element with x_e and x_a, then

$$x_a = K \int_0^t x_e\, dt \tag{11.30}$$

This equation represents an integral transient characteristic with a proportionality factor K. This can be also decomposed into a series connection (see section 11.4.3 combined transient characteristic, page 326) of a pure p-element and a pure I-part.

For $x_e = const$.

$$x_a = K \cdot x_e \cdot t \tag{11.31}$$

Also here the transmission constant K can be determined in two ways, by means of theoretical or experimental process analysis. In the first case, equation 11.28, which was derived on the basis of physical laws, is definite:

$$K = \frac{1}{A} \tag{11.32}$$

In the second case, the experimental process analysis based on equation 11.30 yields under the condition $x_e = const.$:

$$x_a = K \cdot x_e \cdot t \tag{11.33}$$

$$K \cdot x_e = \frac{x_{a1} - x_{a0}}{t_1 - t_0} \tag{11.34}$$

Particularly if the unit step $x_e = 1(t)$ is used as x_e, K can be determined by the straight line slope $x_a = K \cdot t \mid_{xe = 1(t)}$:

$$K = \frac{x_{a1} - x_{a0}}{t_1 - t_0} \mid_{x_e = 1(t)} \tag{11.35}$$

For $t = 0$ and $x_{a0} = 0$,

$$K = \frac{x_{a1}}{t_1} \mid_{x_e = 1(t)} \tag{11.36}$$

The transient characteristics are determined by inserting corresponding input signals into the equation 11.30 as follows:

- **transit function h(t)**

According to definition $x_e = 1(t)$ will be inserted. Thus:

$$h(t) = x_a \mid_{x_e = 1(t)}$$

$$= K \int_0^t 1(t)\, dt$$

$$h(t) = K \cdot t \tag{11.37}$$

- **weight function g(t)**

In this case the DIRAC impulse $x_e = \delta(t)$ will be used as input signal:

$$g(t) = x_a \mid_{x_e = \delta(t)}$$

$$= K \int_0^t \delta(t)\, dt$$

$$g(t) = K \tag{11.38}$$

according to definition:

$$\int_0^t \delta(t) dt = 1$$

● **transfer function G(p)**

The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function:

$$\begin{aligned} G(p) &= X_a \mid_{X_e = L\{\delta(t)\}} \\ &= L\{K \int_0^t \delta(t)\, dt\} \\ &= K \cdot L\{\int_0^t \delta(t)\, dt\} = K \cdot L\{1\} \end{aligned}$$

$$G(p) = K \cdot \frac{1}{p} \tag{11.39}$$

since here applies:

$$\int_0^t \delta(t)\, dt = 1 \tag{11.40}$$

$$L\{1\} = \frac{1}{p} \tag{11.41}$$

11.4.2.3 Differential characteristic => **D**-element

Example:

Transfer elements with differential behaviour come up in electro-technology and serve in control practice to affect processes with a certain mass- or energy deposit in a definite objective. In water management practice it appears in connection with oscillation phenomena such as water hammer in pipes. Differential transient characteristic is characterized by equation 11.42.

$$x_a = K \cdot \frac{dx_e}{dt} \tag{11.42}$$

● **transit function h(t)**

According to definition $x_e = 1(t)$ will be inserted. Thus:

$$\begin{aligned} h(t) &= x_a \mid_{x_e = \mathbf{1}(t)} \\ &= K \frac{d1(t)}{dt} \end{aligned}$$

$$h(t) = K \cdot \delta(t) \tag{11.43}$$

● **weight function g(t)**
In this case the DIRAC impulse $x_e = \delta(t)$ will be used as input signal:

$$g(t) = x_a \mid_{x_e=\delta(t)}$$
$$= K \cdot \frac{d\delta(t)}{dt}$$
$$g(t) = n.d. \qquad (11.44)$$

● **transfer function G(p)**
The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function:

$$G(p) = X_a \mid_{X_e=L\{\delta(t)\}}$$
$$= L\{K\frac{d\delta(t)}{dt}\}$$
$$= K \cdot L\{\frac{d\delta(t)}{dt}\} = K \cdot (pL\{\delta(t)\} - \delta(0))$$
$$G(p) = K \cdot p \qquad (11.45)$$

11.4.2.4 First order delay => **PT_1**-element

Example: Filling procedure with variable discharge
If a container is connected with a receiving stream through a hydraulic resistance R_{hydr} (e.g. gate valve, pipe), the container will have the same water level as in the receiving stream after infinitely long time. The container has a surface area A, and the water level in receiving stream and in the container are H_{Fl} and H respectively. The time dependence of water level H is wanted, if the water level H_{FL} has the value H_{max} over the entire period. H should be equal to zero at time $t = 0$ (see figure 11.14). Thus the following equations apply:

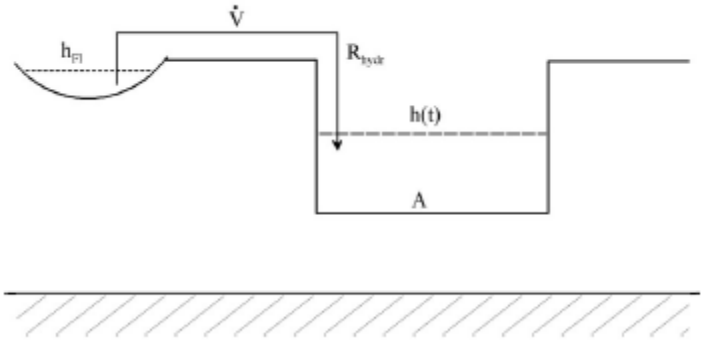


figure 11.14: Filling procedure with variable flow rate

$$V = H \cdot A \tag{11.46}$$

$$H = \frac{1}{A}\int_0^t \dot{V}\, dt \qquad \text{bzw.} \tag{11.47}$$

$$\dot{V} = A \cdot \frac{dH}{dt} \tag{11.48}$$

$$\dot{V} = \frac{(H_{max} - H)}{R_{hydr}} \tag{11.49}$$

$$A \cdot \frac{dH}{dt} = \frac{(H_{max} - H)}{R_{hydr}}$$

$$A \cdot \frac{dH}{dt} + \frac{H}{R_{hydr}} = \frac{H_{max}}{R_{hydr}}$$

$$A_2 \cdot R_{hydr}\frac{dH}{dt} + H = H_{max} \tag{11.50}$$

If we consider again the system as transfer element with x_e and x_a, and then the equation is, $T = A_2 \cdot R_{hydr}$:

$$T \cdot \frac{dx_a}{dt} + x_a = Kx_e \tag{11.51}$$

This differential equation has the solution (see section 5.2.1 solution of differential equation, page 111) for the case $x_e = const$:

$$x_a = Kx_e\left(1 - e^{-\frac{t}{T}}\right) \tag{11.52}$$

Methods for the determination of the parameters K and T are described in detail in section 12.1 model regulation, page 334.

The transient characteristics are determined by inserting corresponding input signals into the equation 11.51 as follows:

● **transit function h(t)**
According to definition $x_e = 1(t)$ will be inserted. Thus:

$$h(t) = x_a \mid_{x_e = \mathbf{1}(t)}$$

$$= \left(1 - e^{-\frac{t}{T}}\right) \cdot K \cdot 1(t)$$

$$h(t) = K\left(1 - e^{-\frac{t}{T}}\right) \tag{11.53}$$

● **weight function g(t)**
In this case the DIRAC impulse $x_e = \delta(t)$ will be used as input signal. The weight function is expressed as derivative of transit function:

$$g(t) = x_a \mid_{x_e = \delta(t)} = \frac{dh(t)}{dt}$$

$$= \frac{d\left(K\left(1 - e^{-\frac{t}{T}}\right)\right)}{dt}$$

$$g(t) = \frac{K}{T} e^{-\frac{t}{T}} \tag{11.54}$$

● **transfer function G(p)**
The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function. Here we assume the differential equation of transient characteristic (see equation 11.52):

$$G(p) = X_a \mid_{X_e = L\{\delta(t)\}}$$

$$T\frac{dx_a}{dt} + x_a = Kx_e \tag{11.55}$$

the LAPLACE transformed form (see section 5.3.3 solution of differential equation with LAPLACE transformation, page 141)

$$T \cdot p \cdot X_a - x_{a0} + X_a = L\{K \cdot \delta(t)\} \tag{11.56}$$

$$\text{mit} \qquad L\{\delta(t)\} = 1 \qquad x_{a0} = 0$$

$$G(p) = X_a \mid_{X_e = L\{\delta(t)\}} = \frac{K}{1 + T \cdot p} \tag{11.57}$$

11.4.2.5 Second order delay => **PT_2**-element

Example: filling procedure of two cascaded containers with variable discharge
Two containers with the surface areas A_2 and A_3 as well as the water levels H_2 and H_3 are cascaded connected one after another. The first one is arranged as the preceding 1st order delay behaviour example and connected receiving stream through the hydraulic resistance $R_{hydr.1}$. Second is coupled to the first container through hydraulic resistance $R_{hydr.2}$. The water level in the receiving stream remains constant value H_{max} (see figure 11.15).

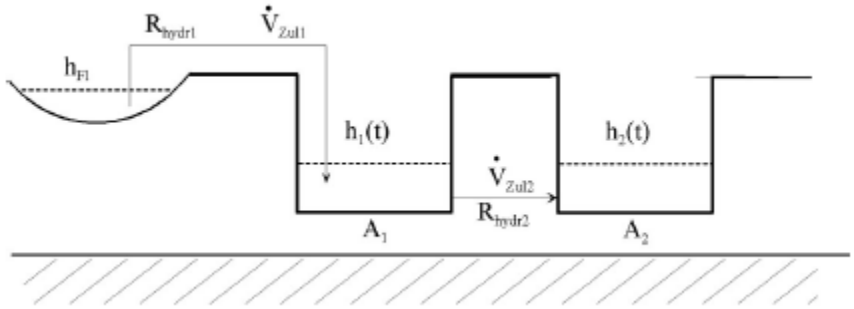


figure 11.15: coupled storage cascade

The water level in 1st container results from the derivation according to equation 11.50:

$$A_2 \cdot R_{hydr.1}\frac{dH_2}{dt} + H_2 = H_{max} \tag{11.58}$$

or with $T_1 = A_2 \cdot R_{hydr.1}$:

$$T_1\frac{dH_2}{dt} + H_2 = H_{max} \tag{11.59}$$

Similarly the water level in 2nd container:

$$A_3 \cdot R_{hydr.2}\frac{dH_3}{dt} + H_3 = H_2 \tag{11.60}$$

or with $T_2 = A_3 \cdot R_{hydr.2}$:

$$T_2\frac{dH_3}{dt} + H_3 = H_2 \tag{11.61}$$

Inserting equation 11.61 into equation 11.59:

$$T_1\frac{d\left(T_2\frac{dH_3}{dt} + H_3\right)}{dt} + T_2\frac{dH_3}{dt} + H_3 = H_{max} \tag{11.62}$$

$$T_1T_2\frac{d^2H_3}{dt^2} + (T_1 + T_2)\frac{dH_3}{dt} + H_3 = H_{max} \tag{11.63}$$

If we consider again the system as transfer element with x_e and x_a, and then the equation:

$$T_1 T_2 \frac{d^2 x_a}{dt^2} + (T_1 + T_2)\frac{dx_a}{dt} + x_a = K x_e \qquad (11.64)$$

This differential equation has solution (see section 5.2.2.2 solution of differential equation, page 125) in case of x_e = *const*. The determination of the Parameter K, T_1 and T_2 as well as the solution steps are described in detail in the section 12.2 transient characteristic with 2nd order delay.

$$x_a = K x_e \left(1 - e^{-\frac{t}{T_1}}\right)\left(1 - e^{-\frac{t}{T_2}}\right) \qquad (11.65)$$

The transient characteristics are determined by inserting corresponding input signals into the equation 11.64 as follows:

● **transit function h(t)**
According to definition $x_e = 1(t)$ will be inserted. Thus:

$$\begin{aligned} h(t) &= x_a \mid_{x_e = \mathbf{1}(t)} \\ &= K \cdot 1(t) \cdot \left(1 - e^{-\frac{t}{T_1}}\right)\left(1 - e^{-\frac{t}{T_2}}\right) \\ h(t) &= K \left(1 - e^{-\frac{t}{T_1}}\right)\left(1 - e^{-\frac{t}{T_2}}\right) \qquad (11.66) \end{aligned}$$

● **weight function g(t)**
In this case the DIRAC impulse $x_e = \delta(t)$ will be used as input signal. The weight function is expressed as derivative of transit function:

$$\begin{aligned} g(t) &= x_a \mid_{x_e = \delta(t)} = \frac{dh(t)}{dt} \\ &= \frac{d\left(K\left(1 - e^{-\frac{t}{T_1}}\right)\left(1 - e^{-\frac{t}{T_2}}\right)\right)}{dt} \\ g(t) &= K\left(\frac{e^{-\frac{t}{T_1}}}{T_1} + \frac{e^{-\frac{t}{T_2}}}{T_2}\right) \qquad (11.67) \end{aligned}$$

● **transfer function G(p)**
The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function. Here we assume the differential equation of transient characteristic (see equation 11.64):

$$G(p) = X_a \mid_{X_e = L\{\delta(t)\}}$$

$$T_1 T_2 \frac{d^2 x_a}{dt^2} + (T_1 + T_2) \frac{dx_a}{dt} + x_a = K \cdot x_e \qquad (11.68)$$

the LAPLACE transformed form (see section 5.3.3 solution of differential equation with LAPLACE transformation, page 141)

$$T_1 T_2 \cdot p^2 \cdot X_a - p \cdot x_{a0} - x_{a0} + (T_1 + T_2) \cdot p \cdot X_a - x_{a0} + X_a = L\{K \cdot \delta(t)\} \qquad (11.69)$$

with $L\{\delta(t)\} = 1$ and $x_{a0} = 0$

$$T_1 T_2 \cdot p^2 \cdot X_a + (T_1 + T_2) \cdot p \cdot X_a + X_a = K \mid_{X_e = L\{\delta(t)\}} \qquad (11.70)$$

$$\frac{K}{T_1 T_2 \cdot p^2 + (T_1 + T_2) \cdot p + 1} = X_a \mid_{X_e = L\{\delta(t)\}}$$

$$G(p) = \frac{K}{(1 + T_1 \cdot p) \cdot (1 + T_2 \cdot p)} \qquad (11.71)$$

11.4.2.6 Duration behaviour => $\mathbf{PT_L}$-element

Example:
The duration behaviour arises in transportation processes, but a change in the balance occurs, i.e. in the case of pure transport without accumulation effect. Thus this process can be also described by a coordinate transformation. This behaviour also plays an important role, if processes should be considered together with different starting points. In these cases different starting points can be convinced different durations.

The equation for this duration behaviour:

$$x_a(t) = K \cdot x_e(t - T_L) \qquad (11.72)$$

The transient characteristics are determined by inserting corresponding input signals into the equation 11.72 as follows:

- **transit function h(t)**

According to definition $x_e = 1(t)$ will be inserted. Thus:

$$h(t) = x_a \mid_{x_e = \mathbf{1}(t)}$$

$$= K \cdot 1(t - T_L)$$

$$h(t) = K \cdot 1(t - T_L) \qquad (11.73)$$

● **weight function g(t)**
In this case the DIRAC impulse $x_e = \delta(t)$ will be used as input signal:

$$g(t) = x_a \mid_{x_e = \delta(t)}$$

$$= K \cdot \delta(t - T_L)$$

$$g(t) = K \cdot \delta(t - T_L) \qquad (11.74)$$

according to definition:

$$\int_0^t \delta(t)dt = 1$$

● **transfer function G(p)**
The transfer equation of time domain must be transformed in complex variable domain by means of LAPLACE operation in this transfer function:

$$G(p) = X_a \mid_{x_e = L\{\delta(t)\}}$$

$$= L\{K \cdot \delta(t - T_L)\}$$

$$= K \cdot L\{\delta(t - T_L)\}$$

$$= K \cdot e^{-T_L p} L\{\delta(t)\}$$

$$G(p) = K \cdot e^{-T_L p} \qquad (11.75)$$

the LAPLACE transformed form (see section 5.3.2 LAPLACE transformation, page 135):

$$L\{f(t - a)\} = e^{-ap} L\{f(t)\} \qquad (11.76)$$

and $L\{\delta(t)\} = 1$.

11.4.2.7 Overview of basic transient characteristic

An overview of different basic form of the transient characteristic step response functions is summarized in figure 11.16.

The different kinds of the mathematical representation of transient characteristic, transition-, weight- and transfer function, for the basic transfer elements are displayed in table 11.7.

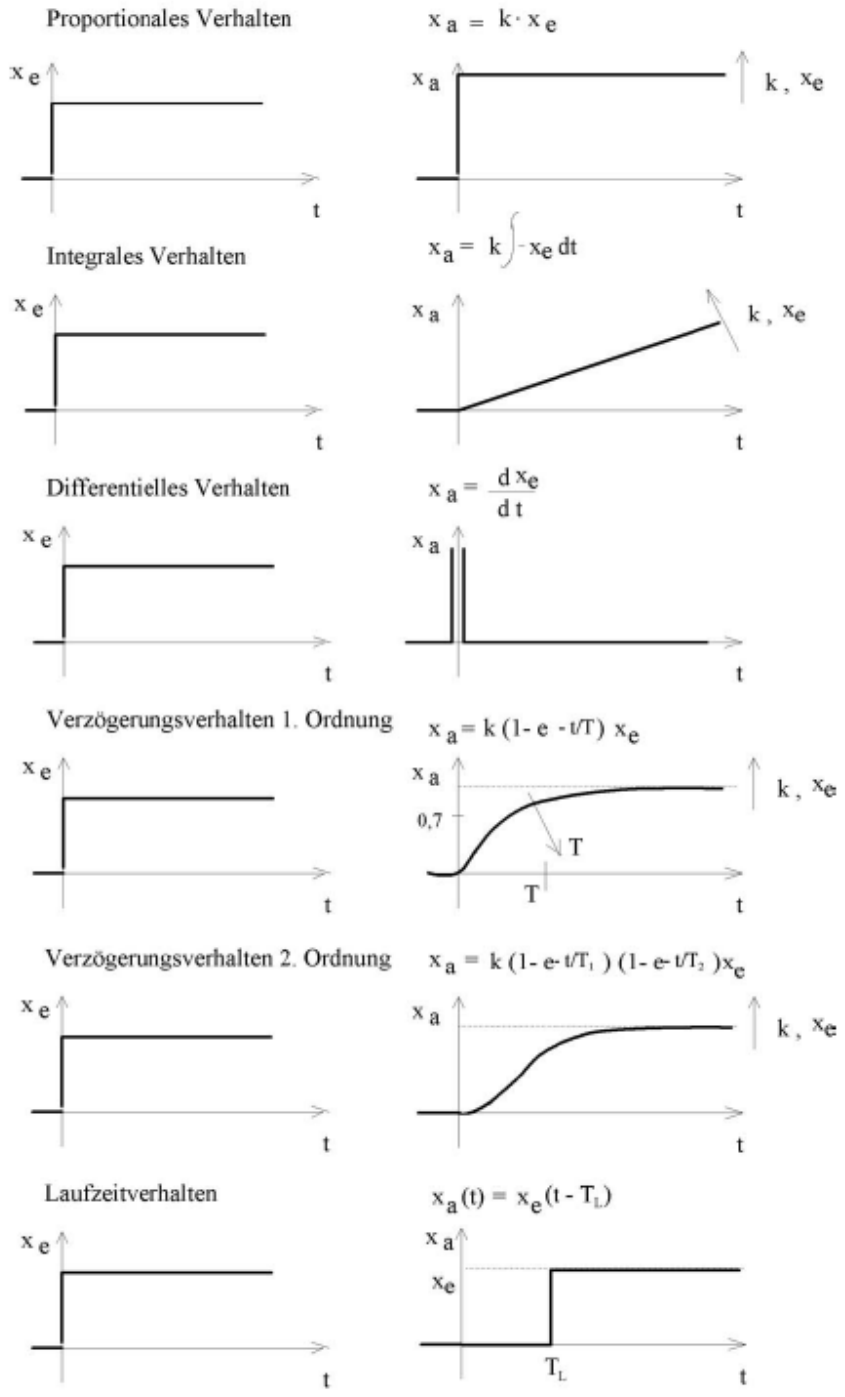


figure 11.16: basic forms of transient characteristic

table 11.7: basic transient characteristic with mathematical description

Verhalten		Bezeichnung: Übergangsfunktion $h(t) = x_a(t) \mid_{xe=1(t)}$	Gewichtsfunktion $g(t) = x_a(t) \mid_{xe=\delta(t)}$	Übertragungsfunktion $G(p) = X_a(p) \mid_{L\{xe=\delta(t)\}}$
	Anreg. $x_e(t)$	**Einheitssprung** $1(t)$	DIRAC-**Impuls** $\delta(t)$	DIRAC-**Impuls** $L\{\delta(t)\} = 1$
P	**proportional**	K	$K\,\delta(t)$	K
I	**integral**	$K \cdot t$	K	$\frac{K}{p}$
D	**differential**	$K \cdot \delta(t)$		$K \cdot p$
PT_1	**Verzöger. 1. Ordn.**	$K\left(1 - e^{-\frac{t}{T}}\right)$	$\frac{K}{T} e^{-\frac{t}{T}}$	$\frac{K}{1+pT}$
PT_2	**Verzöger. 2. Ordn.**	$K\left(1 - e^{-\frac{t}{T_1}}\right)\left(1 - e^{-\frac{t}{T_2}}\right)$	$K\left(\frac{e^{-\frac{t}{T_1}}}{T_1} + \frac{e^{-\frac{t}{T_2}}}{T_2}\right)$	$\frac{K}{(1+pT_1)(1+pT_2)}$
PT_L	**Laufzeit**	$K \cdot 1(t - T_L)$	$K \cdot \delta(t - T_L)$	$K \cdot e^{-T_L p}$

11.4.3 Combined transient characteristic

The interconnection of linear transfer elements can be ascribed to three basic types,

- the **series connection**,
- the **parallel connection** and
- the **circle circuit** (also designated as feedback or back coupling)

Transfer functions are shown in figure 11.7. And mathematical description is table 11.8.

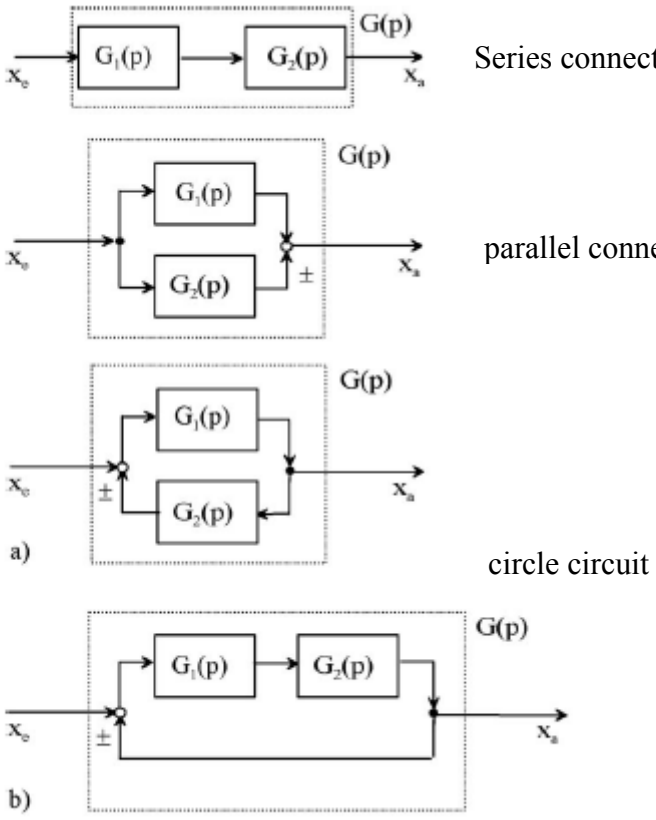


figure 11.7: interconnection of linear transfer elements

Table 11.8: composite transfer elements

Schaltungsart	Differentialgleichung	Gewichtsfunktion	Übertragungsfunktion
Reihe	$x_a(t) = \frac{B_1(D)B_2(D)}{A_1(D)A_2(D)} x_e(t)$	$g(t) = g_1(t) * g_2(t)$	$G(p) = G_1(p) \cdot G_2(p)$
Parallel	$x_a(t) = \frac{[B_1(D)A_2(D) \pm B_2(D)A_1(D)]}{A_1(D)A_2(D)} x_e(t)$	$g(t) = g_1(t) \pm g_2(t)$	$G(p) = G_1(p) \pm G_2(p)$
Kreis			$G(p) = \frac{G_1(p)}{1 \pm G_1(p) \cdot G_2(p)}$

According to the formation law of in **series connected transfer elements** the transfer function of a 2nd order delay element can be calculated as two in series connected 1st order delay elements (see figure 11.18):

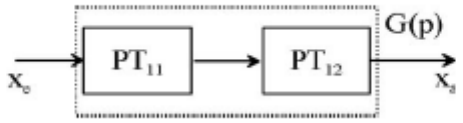


figure 11.18: two series connected PT_1-elements

$$G_1(p) = \frac{K_1}{1 + pT_{11}}$$

$$G_2(p) = \frac{K_2}{1 + pT_{12}}$$

$$G(p) = G_1(p) \cdot G_2(p) \tag{11.77}$$

$$= \frac{K_1}{1 + pT_{11}} \cdot \frac{K_2}{1 + pT_{12}}$$

$$G(p) = \frac{K}{(1 + pT_{11})(1 + pT_{12})} \Longrightarrow \quad \text{PT}_2\text{-Glied} \tag{11.78}$$

The combined transient characteristic of parallel connected transfer elements can be generated from the addition of individual transfer elements mixture in linear transfer elements. Therefore basic transfer elements are occupied with the same input signal and the outputs are added at mixing place, i.e. the elements are parallel connected (see figures 11.19 and 11.20).

For PI-element:

$$G_1(p) = G_P(p) = K_1$$

$$G_2(p) = G_I(p) = \frac{K_2}{p}$$

$$G(p) = G_1(p) + G_2(p) \tag{11.79}$$

$$G(p) = G_{PI}(p) = K_1 + \frac{K_2}{p} \tag{11.80}$$

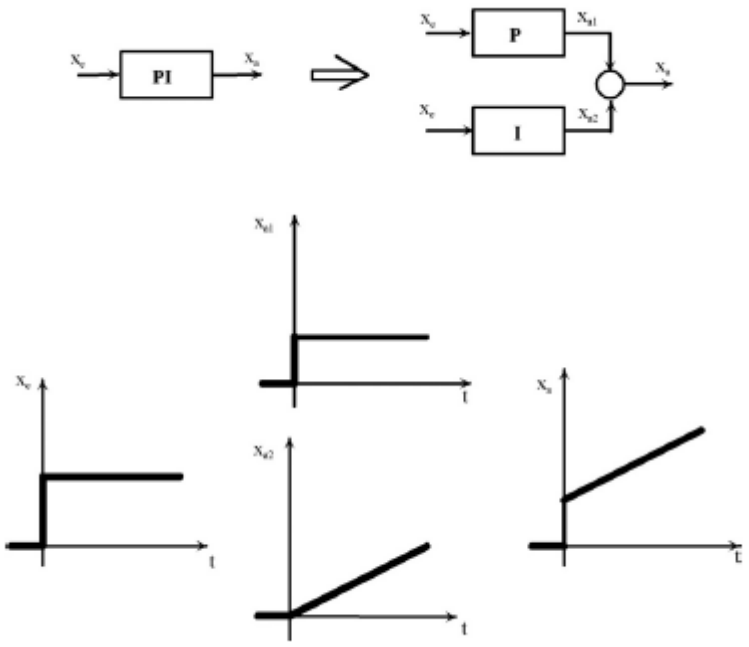


figure 11.19: realisation of a PI-element

The transfer function for PI-element:

$$G_1(p) = G_P(p) = K_1$$

$$G_2(p) = G_I(p) = \frac{K_2}{p}$$

$$G_3(p) = G_D(p) = K_3 \cdot p$$

$$G(p) = G_1(p) + G_2(p) + G_3(p)$$

$$G(p) = G_{PID}(p) = K_1 + \frac{K_2}{p} + K_3 \cdot p \qquad (11.81)$$

The **circle circuit** (feedback systems), which appears in closed control process, is to be explained based on filling level control. This regulation process (seeing figure 11.21) is avowed in GRÄBER "groundwater measuring technique". We recognize that the forward directional transfer element, filling of the container, has integral transient characteristic; the feedback according to technical construction, float with attached lever and gate valve, has proportional behaviour. So it results in the computation of total transient characteristic according to figure 11.17 (circle circuit, layout b) and table 11.8, page 327:

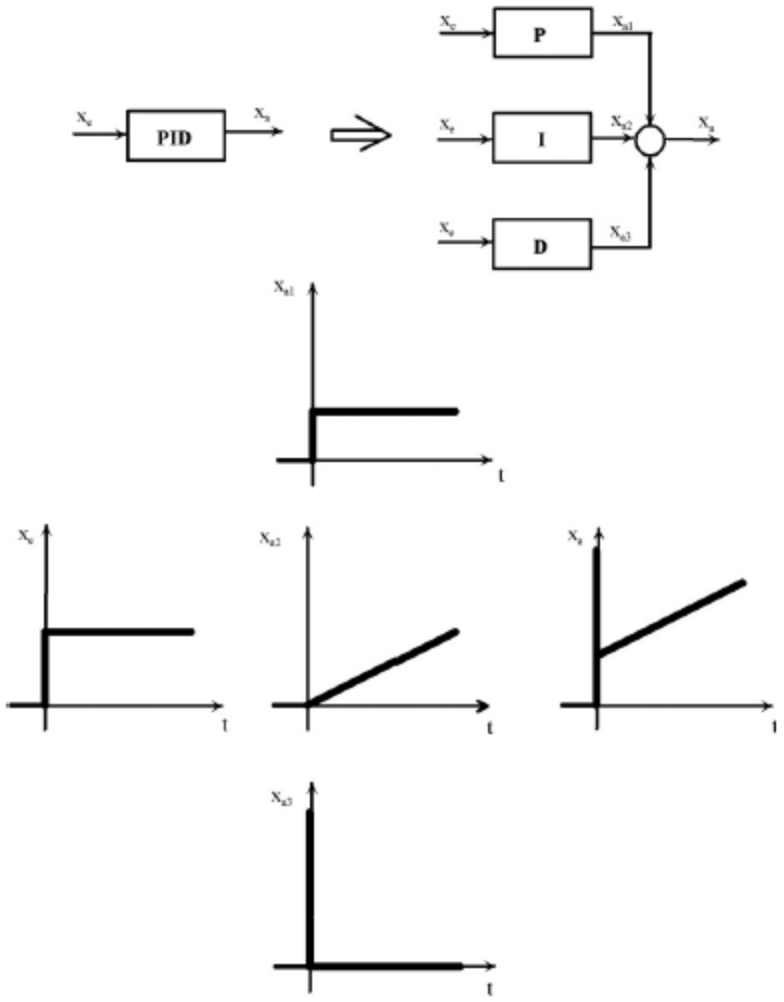


figure 11.20: realisation of a PID-element

$$G_1(p) = G_P(p) = K_1$$

$$G_2(p) = G_I(p) = \frac{K_2}{p} \tag{11.82}$$

$$G(p) = \frac{G_2(p)}{1 + G_1(p) \cdot G_2(p)} \tag{11.83}$$

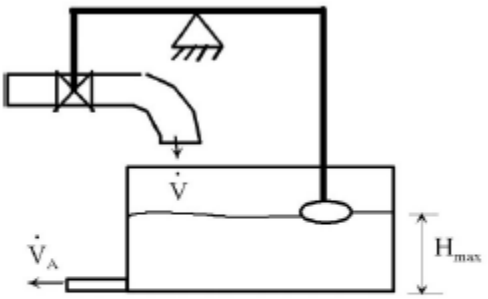
$$= \frac{\frac{K_2}{p}}{1 + \frac{K_2}{p} \cdot K_2}$$

$$= \frac{K_2}{p + K_1 \cdot K_2} = \frac{\frac{1}{K_1}}{\frac{1}{K_1 K_2} p + 1}$$

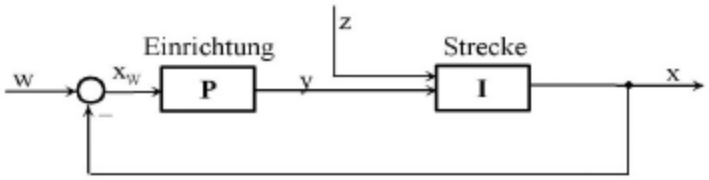
$$G(p) = \frac{K}{1 + Tp} \qquad \text{mit } T = \frac{1}{K_1 K_2} \tag{11.84}$$

This transfer function corresponds to a behaviour of 1st order delay element.

How to attain transient characteristic in experimental away is shown in figure 11.21, and it corresponds likewise to delay characteristic.



Route: filling procedure
Setting: lever and gate valve
x: actual water level H
w: reference water level H_{max}
z: outflow V_A
y: inflow V
x_w: difference H_{max} - H



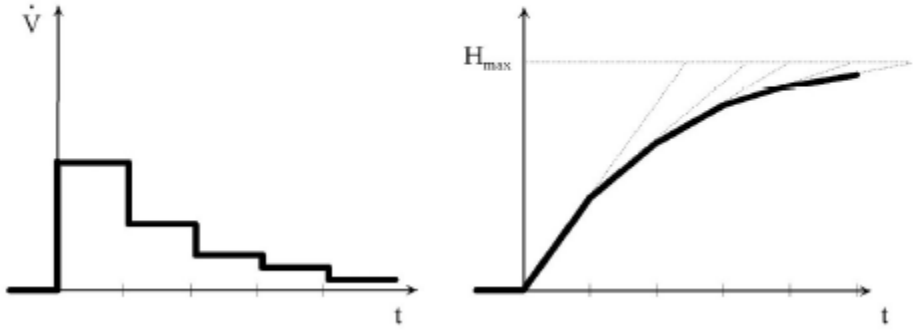


figure 11.21: feedback system for water level control

Chapter 12

12 Model regulation based on parameter

Concrete models and the determination of their characteristics and parameters will be described in the following sections. 1st and 2nd order delay elements play a special role for water management systems. We find this behaviour in filling- and transportation procedure not only in geohydraulics, in surface waters, as also in pipes.

Descriptive procedures are described in this section, which can be also partly graphically solved. Computational procedures, which are based on method i.e. the adjustment of measured values in regression functions, are treated in part IV indirect parameter identification, page 373. With these procedures optimisation will be applied, e.g. least squares (MKQ).

12.1 transient characteristic with 1st order delay

12.1.1 Mathematical description

The behaviour of water management systems corresponding to a 1st order delay can be found in all filling procedures with storage effect in connection with flow resistance (see figure 12.1). For determination of the required hydraulic parameters e.g. the so called pumping tests are used as filling attempts, which can be evaluated by means of the following described methods.

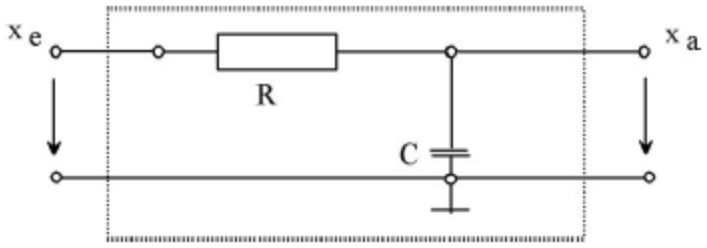


figure 12.1: equivalent circuit diagram of a transfer element with 1st order delay

According to section 5.2.1 solution methods of ordinary differential equations, page 111, the systems, which consist of a flow resistance and a storage capacity (see figure 12.1) can be described by the following differential equation:

$$RC\frac{dx_a}{dt} + x_a = Kx_e(t) \qquad (12.1)$$

$$T\frac{dx_a}{dt} + x_a = Kx_e(t) \qquad \text{mit } T = RC$$

The solution of this differential equation with the boundary condition, the input signals (see figure 12.2) ($x_{e\,t=0} = x_{e0} \cdot 1(t)$): (see section 5.2.1 first order ordinary differential equations, page 111):

$$x_a = K \cdot x_{e0} \cdot \left(1 - e^{-\frac{t}{T}}\right) \tag{12.2}$$

$$h(t) = x_a \mid_{x_e = 1(t)} = K \cdot \left(1 - e^{-\frac{t}{T}}\right)$$

Considering the impulse response g(t) is equal to the differential of step response:

$$g(t) = \frac{dh(t)}{dt} = \frac{K}{T} \cdot e^{-\frac{t}{T}} \tag{12.3}$$

The differential equation can be also solved by means of LAPLACE transformation (see section 5.3.2 LAPLACE transformation, page 135):

$$L\left\{\frac{dx_a}{dt} + \frac{1}{T}x_a\right\} = L\left\{\frac{Kx_e}{T}\right\}$$

$$L\left\{\frac{dx_a}{dt}\right\} + L\left\{\frac{1}{T}x_a\right\} = L\left\{\frac{Kx_e}{T}\right\}$$

$$pL\{x_a\} - x_{at=0} + \frac{1}{T}L\{x_a\} = \frac{K}{T}L\{x_e\} \tag{12.4}$$

bzw. mit $x_{at=0} = 0$

$$L\{x_a\} = \frac{K}{T} \cdot \frac{L\{x_e\}}{\left(p + \frac{1}{T}\right)}$$

The transfer function G(p) can be determined from this equation, while according to definition we consider $L\{\delta(t)\} = 1$ as LAPLACE transformed DIRAC impulse input signal.

$$G(p) = L\{x_a\} \mid_{x_e = L\{\delta(t)\}} = X_a \mid_{x_e = L\{\delta(t)\}} = \frac{K}{(1 + pT)} \tag{12.5}$$

These results are already contained in table 11.7, page 325 and have been verified.

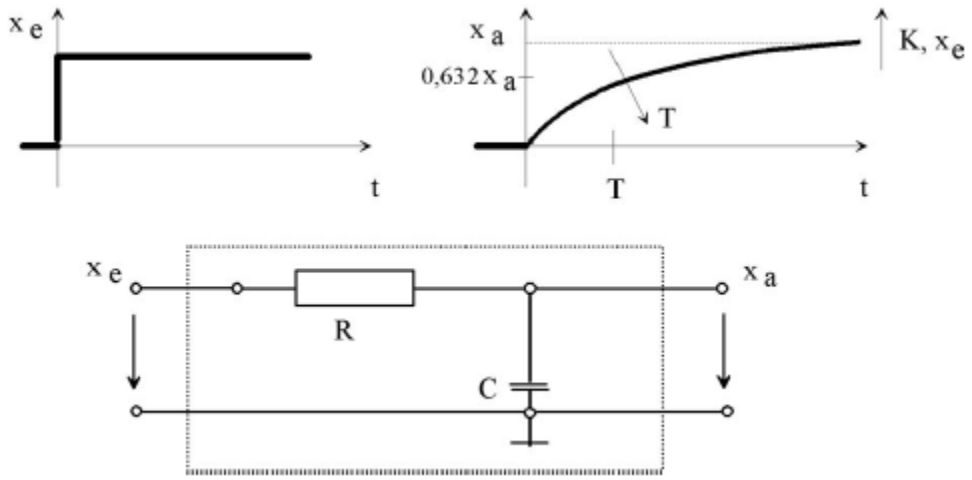


figure 12.2: transient characteristic and circuit of a PT_1-element

If such a RC circuit (see figure 12.2) shows a 1st order delay behaviour, then we can conclude backwards uniquely that the 1st order delay behaviour of any system can be alternatively reproduced by such RC circuit. The task of the experimental process analysis is to determine these substitute parameters from the transient characteristic. This does not have to be absolutely physically interpretable. These are parameters, which show the same behaviour in the equivalent network as the original system. The equivalent network is a model original procedure.

The determination of the transfer factor *K* and the time constants *T* is necessary for clear description of this behaviour. The basic approach of experimental process analysis enables it possible based on a step response function, i.e. the clear regulation of these constants is possible by original reaction on a bar signal.

The **transfer factor K** can be determined from the transfer element behaviour with 1st order delay for infinite time:

$$x_a = K \cdot x_{e0} \cdot \left(1 - e^{-\frac{t}{T}}\right) \tag{12.6}$$

with $t \rightarrow \infty$:

$$K = \frac{x_{a\infty}}{x_{e\infty}} \tag{12.7}$$

for a bar signal:

$$x_{e\,t=0} = x_{e\,t=\infty} \tag{12.8}$$

Thus the step response can be also written in following form:

$$x_a = x_{a\infty} \cdot \left(1 - e^{-\frac{t}{T}}\right) \qquad (12.9)$$

From the comparison between the input curve and the output curve (see figure 12.2) or also from above equation we recognize that a proportional behaviour exists in infinite.

There are several ways to determine time constant T:

- determination of time, in which an integer multiple of time constant available
- determination of slope at zero point.

12.1.2 Time constant from integer multiples

For the case that the time is an integral multiples of time constant:

$$x_{at=nT} = x_{a\infty}\left(1 - e^{-n}\right) \quad \text{mit } n = \frac{t}{T} = 0; 1; 2; 3; ..... \tag{12.10}$$

For bar signal as input function applies by definition $x_{e0} = x_{e\infty}$ and the following table:

$n = \frac{t}{T}$	0	1	1,2	2	3	4
$\frac{x_a}{x_{a\infty}}$	0	0,632	0,699	0,865	0,950	0,982

with

$$\frac{x_{at=nT}}{x_{a\infty}} = \left(1 - e^{-n}\right) \tag{12.11}$$

This table can be evaluated in such a way that we look for the point on the ordinate, where the ratio $x_a / x_{a\infty}$ is a certain value. According to the table a definite ratio between time *t* and time constant *T* belongs to this point on the curve (see figure 12.3).

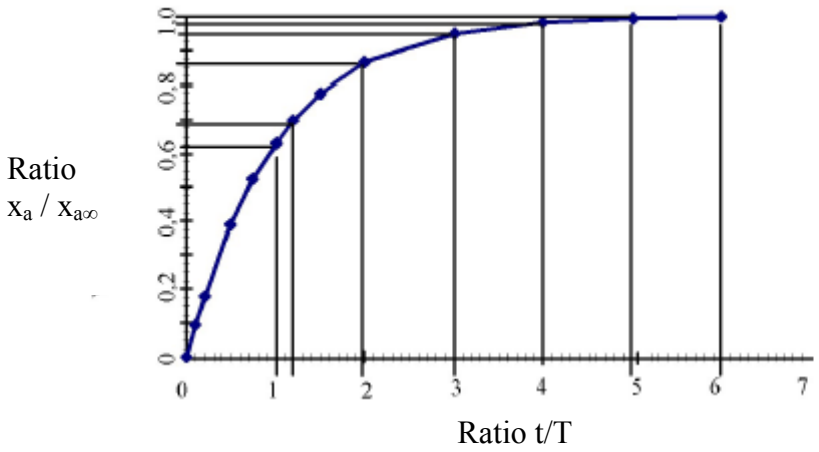


figure 12.3: time constant determination from its multiple

12.1.3 time constant from slope

The time constant can be also calculated from the tangent at any point t . According to equation 12.9:

$$x_a = x_{a\infty}\left(1 - e^{-\frac{t}{T}}\right)$$

$$\frac{dx_a}{dt} = \frac{1}{T} x_{a\infty} e^{-\frac{t}{T}} \qquad \text{für } t = 0 \qquad (12.12)$$

$$\frac{dx_a}{dt}\,|_{t=0} = \frac{x_{a\infty}}{T}$$

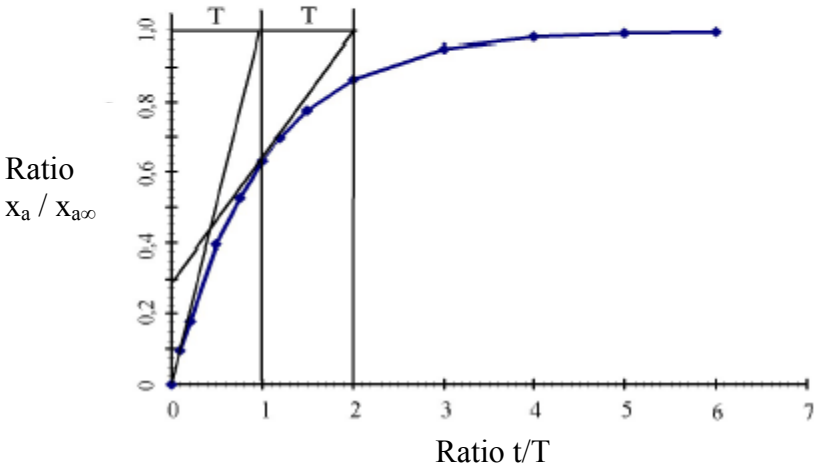


Figure 12.4: tangent intersection and time constant

We can see an intersection with asymptote of step response function $x_{a\infty}$ during setting up the straight line equation for tangent. This intersection has a distance $t_{Schn} = T$.

$$x_{Tang} = \frac{x_{a\infty}}{T} t_{Schn} \equiv x_{a\infty}$$

$$t_{Schn} = T$$

Since the measuring errors are largest at the beginning of measurement series, i.e. at time zero, we can also set the tangent at any place. Time difference between tangent point and intersection with the asymptote of step response is then equal to the time constant, because the exponential function possesses a constant slope (see figure12.4).

12.2 transient characteristic with 2nd order delay

12.2.1 Mathematical description

The behaviour of water management systems corresponding to a 2nd order delay can be found in all transportation procedures with storage effect in connection with flow resistance (see figure 12.5). The tracer tests are implemented to determine the associated hydraulic transportation parameters, which can be evaluated by means of following described methods. Also here we can assume the solution of an appropriate differential equation. According to the equivalent circuit diagram (see figure 12.5) we can set up the following differential equation:

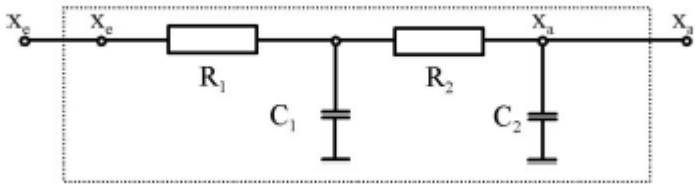


figure 12.5: equivalent circuit diagram of a transfer element with 2nd order delay

$$R_1 C_1 \frac{dx_{a1}}{dt} + x_{a1} = K_1 x_{e1} \tag{12.13}$$

$$R_2 C_2 \frac{dx_{a2}}{dt} + x_{a2} = K_2 x_{e2} \tag{12.14}$$

with coupled conditions:

$$x_{e_2} = x_{a_1} \tag{12.15}$$

$$x_e = x_{e_1}$$

$$x_a = x_{a_2}$$

and the time constant:

$$T_1 = R_1 C_1 \tag{12.16}$$

$$T_2 = R_2 C_2$$

we get

$$T_1 \frac{dx_{a1}}{dt} + x_{a1} = K_1 x_e$$

$$\frac{\left(T_2 \frac{dx_a}{dt} + x_a\right)}{K_2} = x_{a1}$$

$$T_1 \frac{d\frac{\left(T_2 \frac{dx_a}{dt} + x_a\right)}{K_2}}{dt} + \frac{\left(T_2 \frac{dx_a}{dt} + x_a\right)}{K_2} = K_1 x_e$$

$$T_1 T_2 \frac{d^2 x_a}{dt^2} + T_1 \frac{dx_a}{dt} + T_2 \frac{dx_a}{dt} + x_a = K_1 K_2 x_e$$

$$T_1 T_2 \frac{d^2 x_a}{dt^2} + \frac{dx_a}{dt}(T_1 + T_2) + x_a = K\, x_e \qquad (12.17)$$

We get the solution of this differential equation e.g. through the substitution method (see section 5.2.2.2 differential equation of type b, page 125). The characteristic equation of the homogeneous differential equation:

$$a\lambda^2 + b\lambda + c = 0 \qquad (12.18)$$

wobei $a = T_1 T_2$,

$b = T_1 + T_2$ und

$c = 1$ ist.

with new constants d = b/a and f =c/a:

$$\lambda^2 + d\lambda + f = 0 \qquad (12.19)$$

$$\lambda_{1,2} = -\frac{d}{2} \pm \sqrt{\frac{d^2}{4} - f}$$

With the solution of this quadratic equation we differentiate three cases depending upon radian value. For the regarded technical systems here only the positive case, different from zero radian plays a role.

$$\frac{d^2}{4} > f, \text{ bzw } .b^2 > 2 \cdot c \cdot a$$

$$(T_1 + T_2)^2 > 2 \cdot (T_1 \cdot T_2)$$

Thus we get:

$$
\begin{aligned}
\lambda_{1,2} &= -\frac{d}{2} \pm \sqrt{\frac{d^2}{4} - f} \\
&= -\frac{T_1 + T_2}{2 \cdot T_1 \cdot T_2} \pm \sqrt{\left(\frac{T_1 + T_2}{2 \cdot T_1 \cdot T_2}\right)^2 - \frac{1}{T_1 \cdot T_2}} \\
&= -\frac{T_1 + T_2}{2 \cdot T_1 \cdot T_2} \pm \frac{1}{2 \cdot T_1 \cdot T_2}\sqrt{(T_1 + T_2)^2 - 2 \cdot T_1 \cdot T_2} \\
&= \frac{1}{2 \cdot T_1 \cdot T_2}\left(-(T_1 + T_2) \pm \sqrt{(T_1 - T_2)^2}\right) \\
&= \frac{1}{2 \cdot T_1 \cdot T_2}\left(-T_1 - T_2 \pm (T_1 - T_2)\right) \\
\lambda_1 &= -\frac{1}{T_1} \\
\lambda_2 &= -\frac{1}{T_2}
\end{aligned}
$$

This yields the solution of differential equation:

$$
x_a = x_e(t)\left(K_1 e^{-\frac{t}{T_1}} + K_2 e^{-\frac{t}{T_2}}\right) \tag{12.20}
$$

The constants K_1 and K_2 can be determined base on concrete initial- or boundary conditions.

12.2.2 Unit Step as input signal (transfer function *h*(*t*))

For above transfer element the model parameters can be determined as the behaviour excited by a jump, the step response, in the 1st order delay elements (see figure 12.6).

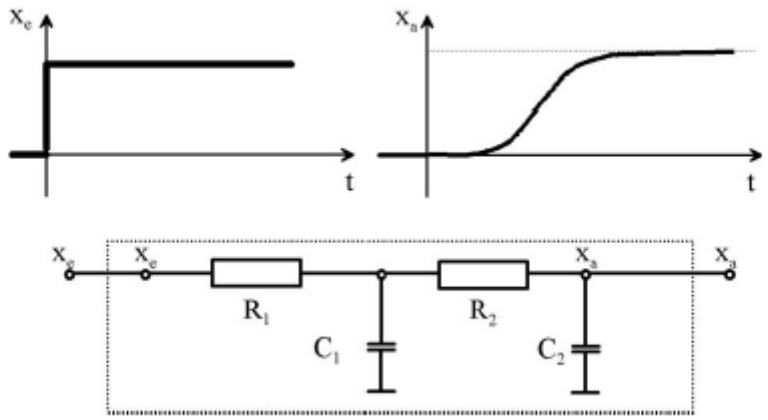


figure 12.6: Step response function and equivalent circuit diagram of a PT_2T_L-element

The behaviour can be described according to the derivation and under the consideration of excitement of a bar signal as follows. 2nd order delay elements, e.g. transportation processes, are bonded by their convective portion at delay characteristics. Therefore generally another delay T_L, i.e. a time lag, should be considered.

Assuming general solution of differential equation (see equation 12.20, page 342) under special condition of a bar signal $x_e(t) = x_{e0} \cdot 1(t)$:

$$x_a = x_e(t)\left(K_1 e^{-\frac{t-T_L}{T_1}} + K_2 e^{-\frac{t-T_L}{T_2}}\right) \qquad (12.21)$$

$$x_a = K \cdot x_{e0} \cdot \left(1 - e^{-\frac{t-T_L}{T_1}}\right)\left(1 - e^{-\frac{t-T_L}{T_2}}\right) \qquad (12.22)$$

The proportional transfer factor K, both time constants T_1 and T_2 as well as the delay T_L must be determined on the basis of complicated structure parameter here. Again selected values of the step response function will be used. The transfer function G(p) can be assumed following shape for the 2nd order delay elements (PT_2T_L):

$$G(p) = \frac{Ke^{-pT_L}}{(1+pT_1)(1+pT_2)} \quad \text{für } T_1 \neq T_2 \quad \text{So called model I} \qquad (12.23)$$

$$G(p) = \frac{Ke^{-pT_L}}{(1+pT)^2} \quad \text{für } T_1 = T_2 \quad \text{So called model II} \qquad (12.24)$$

The distinction, which type of model deals with appropriate measurement series of characterized transfer element, is achieved by STREJC in table 12.1.

Figure 12.1: classification of model type according STREJC

Modelltyp	$\frac{x_{a_W}}{x_{a\infty}}$	$\frac{T_u}{x_{a\infty}}$
I	$\leq 0,264$	≤ 0104
II	> 0264	> 0104

The transfer constant *K* again results from the behaviour in infinite:

$$x_{at\to\infty} = K \cdot x_{e0}, \qquad \text{da } e^{-\infty} = 0$$
$$K = \frac{x_{a\infty}}{x_{e0}}$$

and thus:

$$x_a = x_{a\infty}\left(1 - e^{-\frac{t-T_L}{T_1}}\right)\left(1 - e^{-\frac{t-T_L}{T_2}}\right) \tag{12.25}$$

The delay T_L can be read directly from the diagram of the step response (see figure 12.7). The occurrence of a delay must be considered as shift of time axis. The appropriate variables (x_{aW}, $x_{a\infty}$, Γ_u, $x_{a\infty}$) can be taken from figure 12.7.

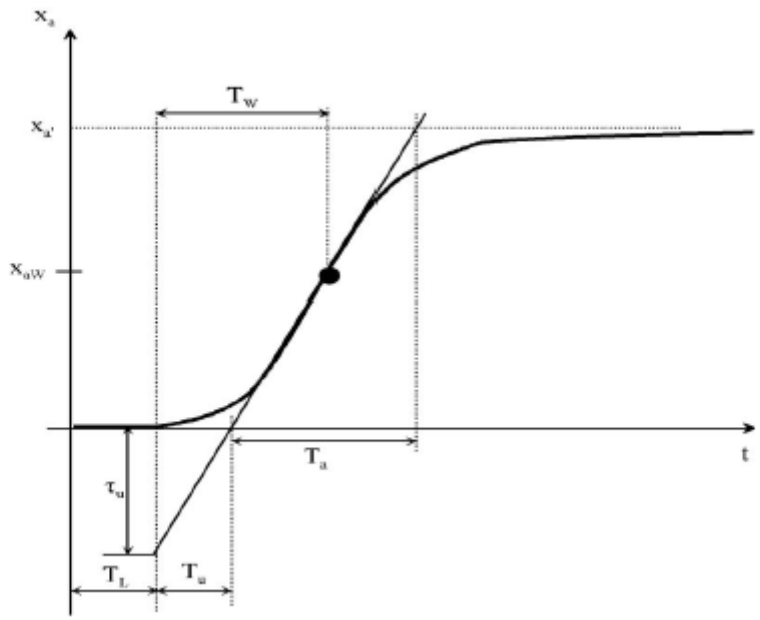


figure 12.7: parameter of step response

12.2.2.1 Model type I

In this type of model, the case of different time constants $T_1 \neq T_2$, the conditional equations must be found for both two time constants. According to the literature [STREJC] it is known that the value $x_{a\,0.7} = 0.7\, x_{a\infty}$ is nearly independent of the ratio of two time constants, but is strongly dependent on the sum of two time constants. With an error smaller than 1.7% we can apply:

$$t_{0,7x_{a\infty}} = 1,2\,(T_1 + T_2)$$
$$T_1 = \frac{t_{0,7x_{a\infty}}}{1,2\left(1 + \frac{T_2}{T_1}\right)} \qquad (12.26)$$

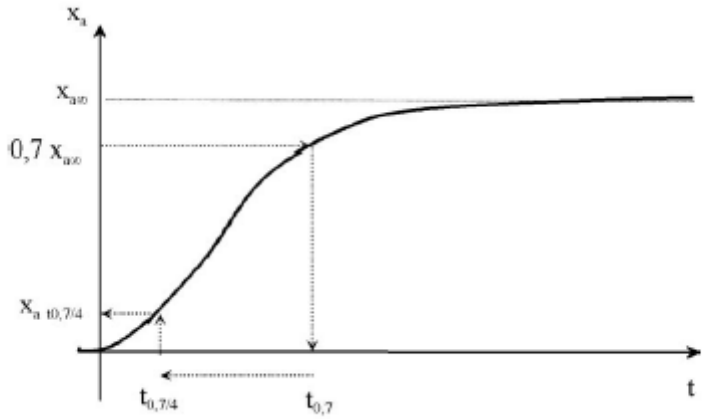


figure 12.8: determination of parameters T_1 and T_2

On the other hand we can assume that the function value $x_{a0.7/4} = x_{a(t(0.7xa\infty)/4)}$ according to figure 12.8 only depends on the ratio T_2/T_1. The ratio T_2/T_1 will be determined from table 12.2.

Thus two equations are available for determination of the time constants and the task is uniquely solvable.

It is still to be noted that these transfer elements for large time ($t >> T_W$) approximately behave as 1st order transfer elements. Particularly with large difference of time constants the later process is dominated by process with time constant T_2, since the processes with time constant T_1 already faded away.

Table 12.2: time constant ratio dependent on step response

$\frac{x_{a\,0,7/4}}{x_{a_\infty}}$	$\frac{T_2}{T_1}$
$0,260$	$0,00$
$0,200$	$0,10$
$0,174$	$0,20$
$0,150$	$0,33$
$0,135$	$0,40$
$0,131$	$0,50$
$0,126$	$0,60$
$0,125$	$0,70$
$0,124$	$0,80$
$0,123$	$0,90$
$0,123$	$1,00$

12.2.2.2 Model type II

The type of model II is shown in table 12.1 (T1 = T2, i.e. only one time constant), so we can determine the necessary parameters by table 12.3. In this case the transfer function can be even extended to arbitrary integer exponents *n*:

$$G(p) = \frac{K\ e^{pT_L}}{(1-pT)^n} \qquad \text{für Modelltyp II } (T_1 = T_2) \qquad (12.27)$$

The determination of K and T_L is independent of model type and can be achieved as described in model type I (see page 344).

Table 12.3: parameter estimation for model type II

n	$\frac{\tau_u}{x_{a_\infty}}$	x_{a_W}/x_{a_s}	$\frac{\mathbf{T}_W}{\mathbf{T}}$	$\frac{\mathbf{T}_u}{\mathbf{T}}$	$\frac{\mathbf{T}_a}{\mathbf{T}}$
1	0	0	0	0	1
2	0,104	0,264	1	0,282	2,718
3	0,218	0,323	2	0,805	3,695
4	0,319	0,352	3	1,425	4,463
5	0,410	0,371	4	2,100	5,119
6	0,493	0,384	5	2,811	5,699
7	0,570	0,394	6	3,549	6,226
8	0,642	0,401	7	4,307	6,711

Base on table 12.3 we have the possibility to determine the time constant T in different ways. By averaging these values we can obtain a value with a smaller error. This is important since measuring errors of experiment also completely shrink in the parameter estimation.

12.2.3 DIRAC impulse as input signal (Weight function *g*(*t*))

Apart from the possibility for the determination of transfer parameter treated in the preceding section, a further way is described here. The parameters of systems, i.e. time constants, transfer factor etc., should be invariant compared to input signal, since linear systems are considered here. This circumstance allows to use different test signals for description of the systems, which however always lead to the same transfer function. In technical test practice usually one or other input signals will be more feasible. In DIRAC impulse the impressing of a very large impulse (energy- or mass deposit) in a very short time interval $\Delta T << T$ (proportional to smallest appeared time constant) is observed. The introduced method here yields expedient value up to a pulse width of $\Delta T \leq 0.1$ T. Thus the circumstances are displayed in figure 12.9.

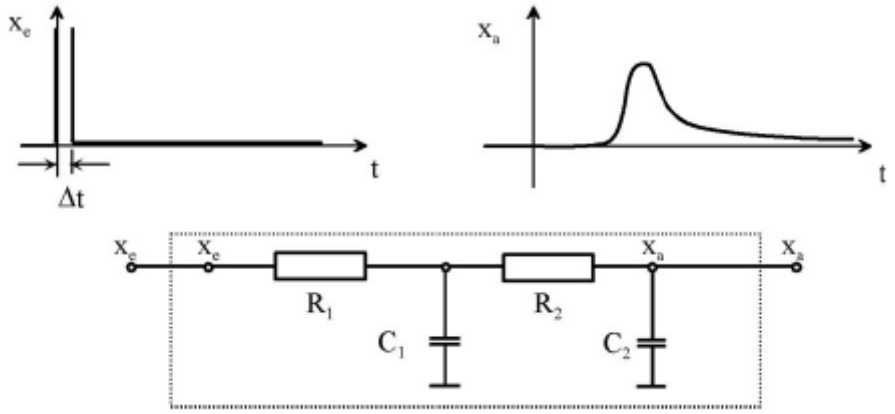


figure 12.9: 2nd order transfer element (PT_2T_L)

In contrast to the preceding section here only transfer elements with same time constant are considered (see equation 12.28). This is designated as model type II in the section 12.2.2.2.

$$G(p) = \frac{Ke^{-pT_L}}{(1+pT)^n} \qquad (12.28)$$

According to the relationship of the two output values $x_a(T_m)/x_a(T_m/2)$ (see figure 12.10) the parameters n (number of coupled RC elements = exponent of the denominator polynomial) and T (time constant) will be determined based on table 12.4. The time T_M is the point which the impulse response function, the weighting function g(t) reaches maximally (see figure 12.10). A possibly appeared delay is also mentioned here. $T_m/2$ stands for half time value to the maximum. T_m, and $T_m/2$ refer to time axis with T_L shifted.

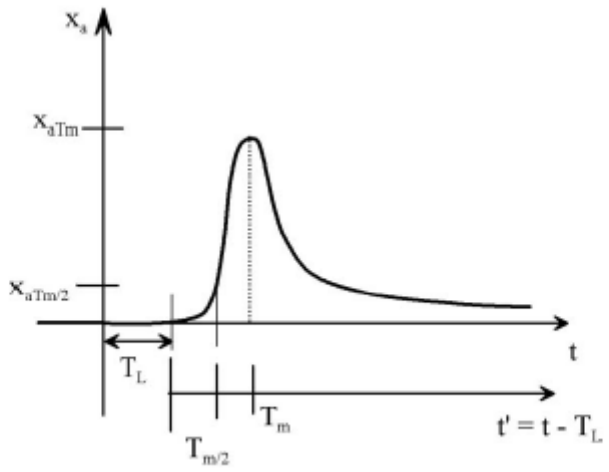


figure 12.10: impulse response function g(t) for a 2nd order delay element

table 12.4: variables of impulse response function for 2nd order delay

$\frac{x_a(T_m)}{x_a(T_m/2)}$	n	$\frac{T_m}{T}$	$\frac{(T\, x_a(T_m))}{(A\, K)}$
1,213	2	1	0,368
1,471	3	2	0,271
1,785	4	3	0,224
2,165	5	4	0,196
2,623	6	5	0,175
3,185	7	6	0,159

The transfer constant K can be determined by the fourth column in table 12.4. Base on variable A, the impulse area ($A = x_e \cdot \Delta t$), technically realizable impulses can also be evaluated with real $\Delta T \leq 0.1\ T$.

12.2.4 Tasks of experimental process analysis

1. Compute the drawdown curve for a conveyor capacity V = 0.005m^3/s by means of transfer element method with a flow rate of V = 0.015m^3/s, if a pumping test yields following values (see table). Depict the result graphically.

Zeit $[s]$	Absenkung $[m]$
320	0,63
426	0,69
564	0,73
743	0,77
976	0,81
1279	0,85
1673	0,88

Zeit $[s]$	Absenkung $[m]$
2185	0,92
2850	0,96
3715	0,99
4839	1,03
6302	1,06
8202	1,10
10.000	1,14

2. Please determine the transfer function including parameters for the following measurement series, which is originated from a supply function:

t $[min]$	0	1	2	4	8	15
$\dot{V}\ \left[\frac{m^3}{s}\right]$	0	0,1	0,1	0,1	0,1	0,1
s $[m]$	0	0,1	0,08	0,13	0,19	0,25

3. The following dependence between flow rate V and groundwater drawdown *s* is found for groundwater position in a pumping test:

$\dot{V}\ \left[\frac{m^3}{s}\right]$	0	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05
s $[cm]$	0	0	3	8	20	30	35	37	38
t $[min]$	−1	0	1	2	4	10	20	40	100

Calculate the drawdown process with a flow rate of V = 0.15 m^3/s. Apply the method of transfer functions. Plot the measured value and the result.

4. In pumping test two different position P1 and P2 are far away from the infiltration well with a distance of r_1 = 350m and r_2 = 1000m respectively, and following concentrations *C* of positioning tracer are measured. A steady concentration $C_{(0,t)} = 10g/m^3$ is added in the infiltration well.

$t\,[10^7 s]$	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5
$C_{r1}\,[\frac{g}{m^3}]$	2,4	2,7	3,05	3,55	4,80	6,50	7,95	8,70	9,10	9,20	9,25
$C_{r2}\,[\frac{g}{m^3}]$	2,4	2,7	3,04	3,32	3,57	3,81	4,01	4,20	4,37	4,53	4,67

Calculate the transfer functions for this system.

5. In a tracer test 50kg concentrated NaCl solution infiltrates 5min long into the soil at the well.

Calculate process of a possible pollutant dispersal, if average 1000kg solution had arrived into the soil. Place the measured values and prognosticated values graphically.

$t\,[min]$	24	30	35	40	42	50	60	70	80	90	100	120
$C\,[\frac{mg}{l}]$	0	2,0	7,0	9,7	9,8	7,5	5,0	3,5	1,5	0,5	0,3	0

6. In a column flow test the following impulse response function of a pollutant with concentration 30mg/l was measured (see figure 12.11).

   a) determine the weighting function and the transfer function for these measured values.
   b) prognosticate the concentration after 160min, if the input concentration is of following characteristic:

$t\,[min]$	0	20	40	60	80	100	120	140
$C\,[\frac{mg}{l}]$	30	50	80	60	100	50	10	0

7. Following concentrations were measured according to tracer test in a groundwater observation tube. 50kg concentrated NaCl solution infiltrated in this tracer test within 5 hours.

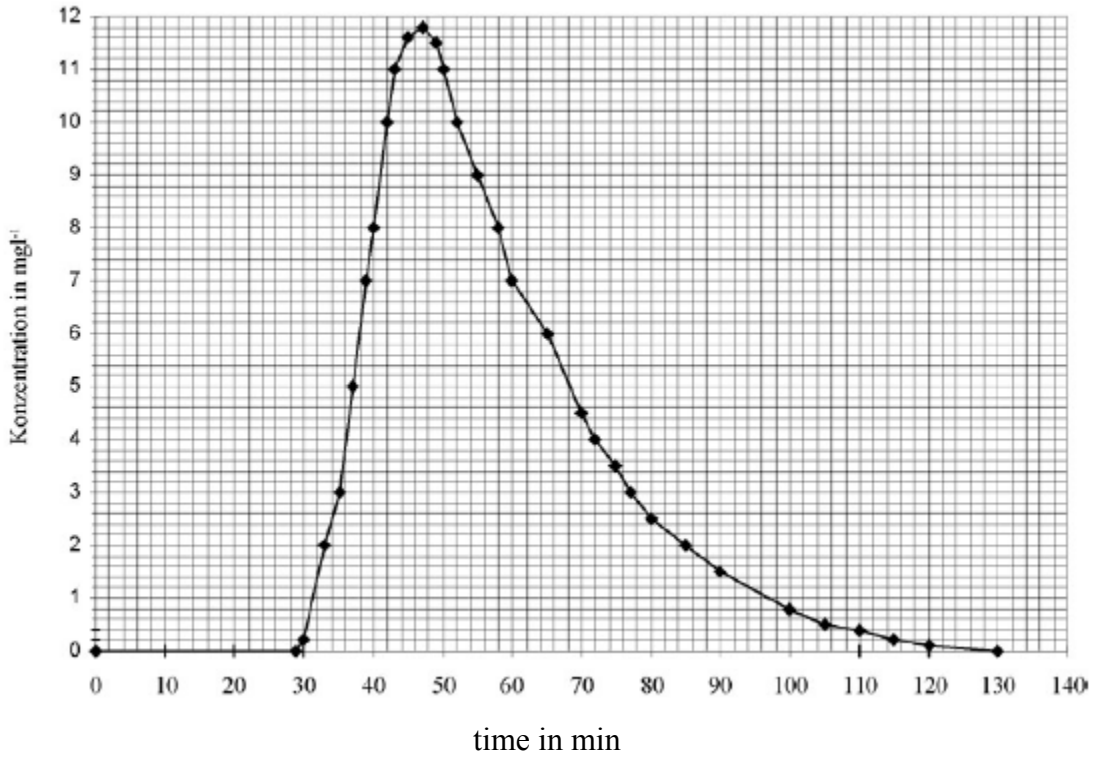


figure 12.11: impulse response of a column flow test

a) calculate process of total salt transport in the observation well, if the following individual measured values were obtained.
b) places the measuring curve and the computed function graphically.

t $[d]$	0	1	2	4	5	7	9
$C_{NaCl}\left[\frac{mg}{l}\right]$	0	0	1	2	1,5	1	0

c) calculate expected breakthrough curve by means of transfer function method and plot if an infiltration of 100kg worked within 2.5 hours.

8. The following groundwater levels were measured in a pumping test (see figure 12.12)

a) calculate water deficit (volume) of drawdown funnel, if the aquifer has the following characteristic values:
h_n = 16m, M = 10m, k = 0.001m · s^{-1}, S_0 = 0.0001m^{-1}, n_0 = 0.20
b) Compute by means of transfer element method and with a) founded value for the flow rate V the drawdown curve for conveyor capacity of 0.005m^3 · s^{-1}.

Set up the first four equations of faltung integral for the model until observation time point t = 1d.

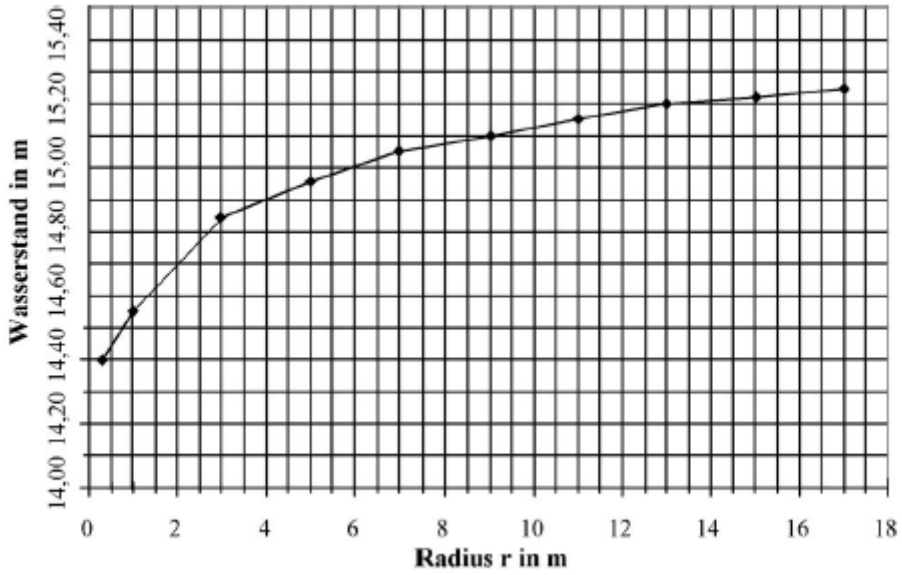


figure 12.12: groundwater level dependent on radius

12.3 Arbitrary transient characteristic and arbitrary input signals

12.3.1 Introductory

Most natural processes take place one time and are not reproducible. In rarest cases it is also possible to impress arbitrary test signals on natural ecological processes happened only once, in order to determine the type and the parameter of the transient characteristic by means of experimental process analysis. Very often the task consists of one-time natural processes, e.g. flood waves, precipitation discharge events, groundwater formation rates or pollutant disposals, by means of mathematical method to derive the relationship between the input- and output behaviour according to experimental process analysis, i.e. to determine the transient characteristic. For this reason other methods had to be developed. One of them is the application of faltung integral / DUHAMEL integral. The basic idea of this method is the decomposition of arbitrary input signal into a sum of impulses, which then possess a special transient characteristic individually. The faltung integral is in particular used for single deadbeat events. Afterwards the portions of the each transferred impulses will be again overlaid. Due to superposition law this method can be only applied in linear or in piecewise linearized systems. The application of FOURIER series analysis or syntheses is quoted in periodic functions.

The books can be consulted as literature for this section:

DYCK, S: Grundlagen der Hydrologie
LUCKNER, L.; SCHESTAKOV, W. A.: Migrationsprozesse
WERNSTEDT, J.: Experimentelle Prozessanalyse

Furthermore all books can be recommended, in which applications of faltung integrals on technical processes are described. The different notations or the different symbols and abbreviations must be paid attention in a comparative literature study. Following abbreviations according to international standard in system technology (see table 12.5) will be used.

Table 12.5 comparison of applied abbreviations

Bezeichnung	**Abkürzung bei intern. Standard**	**Abkürzung bei** DYCK	**Abkürzung bei** LUCKNER
DIRAC-**Impuls**	$\delta(t)$	$p_n(t)$	
Impulsantwort = **Gewichtsfunktion**	$g(t)$	$h(t)$	$h(t)$
Sprungantwort = **Übergangsfunktion**	$h(t)$	$S(t)$	$S(t)$
Eingangsfunktion	$x_e(t)$	$p(t)$	$R(\tau)$
Ausgangsfunktion	$x_a(t)$	$q(t)$	$P(t_0)$
Verzögerungszeit	τ	Δt	τ

12.3.2 Decomposition of arbitrary input function (Signal analysis)

While in the preceding sections selected input signals (e.g. step function, DIRAC impulse) are discussed, here arbitrary input signals are considered. This is very necessary for many tasks of water management, hydrology and geohydraulic. Always, if artificial test signals cannot be used, but natural events must be exploited to experimental systems analysis, only the faltung integral method described in the following can be applied. Time should be considered as independent variables. An application of faltung integral on local variables is also conceivable.

The basic idea of the signal analysis consists of the fact that arbitrary time response of a function can be represented as an infinite sum of selected single signals (see section 11.3.4 signal analysis, page 301). In principle the different signals can be used. The sinusoidal signals are of special meaning, which can be found in well known FOURIER series analysis application. The bar signals and the impulses lead to LAPLACE transformation. Therefore periodic and periodization functions are analysed by means of FOURIER analysis and unique, deadbeat procedure by means of LAPLACE transformation.

The arbitrary input signal is decomposed into a sum time shifted impulses in the application of the faltung integral (see figure 12.13).

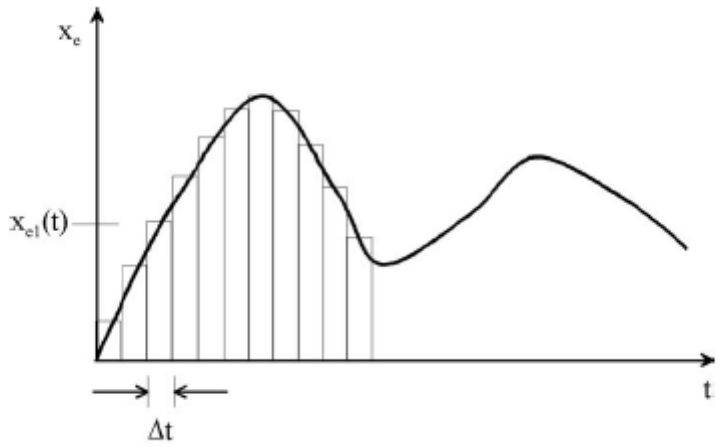


figure 12.13: approximation of a function by impulse

The effect of a signal on a system is usually characterised by energy- or mass flow. It is defined by the respective signal variable and the effect duration, i.e. by the function integral of time. With the approximation of input signal by a sum of individual square pulse, the integral is approximately described by a sum of the products of pulse amplitude $x_{ei}(\Gamma_i)$ and –width Δt:

$$\int_0^{t_e} x_e(t)\,dt \approx \sum_{t=1}^{n=\frac{t_e}{\Delta t}} x_{et}(\tau_t)\,\Delta t \qquad (12.29)$$

The individual impulse of time point Γ_i, which affects as input signal of transmission system, produces individual impulse response functions at system output (see figure 12.14), the weighting functions $g_t(t - \Gamma_i)$. These are superposed and yield system response to the input signal $x_e(t)$. It should be noted that the superposition can be only applied for linear systems.

For pulse width $\Delta t \rightarrow 0$ the technical impulse approaches DIRAC impulse and finite sum in integral representation, whereby an infinite number of impulses is considered.

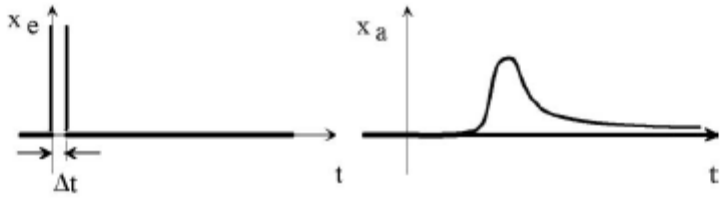


figure 12.14: impulse response function g(t) for a 2nd order delay element

12.3.3 composition of output function (Signal syntheses)

As already mentioned the output signal by the overlay of sum of individual impulse response functions, results in weighting functions $g_i(t - \Gamma_i)$ at time *t*. For computation we must note that all preceding impulses in time interval 0 to t contribute, since the weighting functions are not yet faded away at time *t*.

As recognized from figure 12.15, the output signal $x_a(t)$ to time t_0 is composed from the time shifted weighting functions portions for the time t_0:

$$x_a(t_0) \approx \sum_{t=1}^{n=\frac{t_0}{\Delta\tau}} x_{at}(t_0) = \sum_{t=1}^{n=\frac{t_0}{\Delta\tau}} (x_{et}(\tau_t) \cdot g(t_0 - \tau_t)) \qquad (12.30)$$

$\Delta\Gamma$ is the time lag of DIRAC impulse, the so called aperture time. If we arrange the border crossing to infinitesimal aperture time, the sum changes into integral form, which can be also designated as **faltung integral** or **DUHAMEL integral**:

$$x_a(t) = \int_0^t x_e(\tau) \cdot g(t - \tau) d\tau$$

$$= g(t) * x_e(t) \qquad (12.31)$$

In this case * - operation stands for **Faltung operation.**

We can also interpret faltung integral in such a way that, all impulses of input signal $x_e(t)$ in time interval $0 \leq \Gamma \leq t$ contribute to the value to output signal at time point *t*, which are weighted according to aperture time $(t - \Gamma)$ with the factor $g(t - \Gamma)$ in each case.

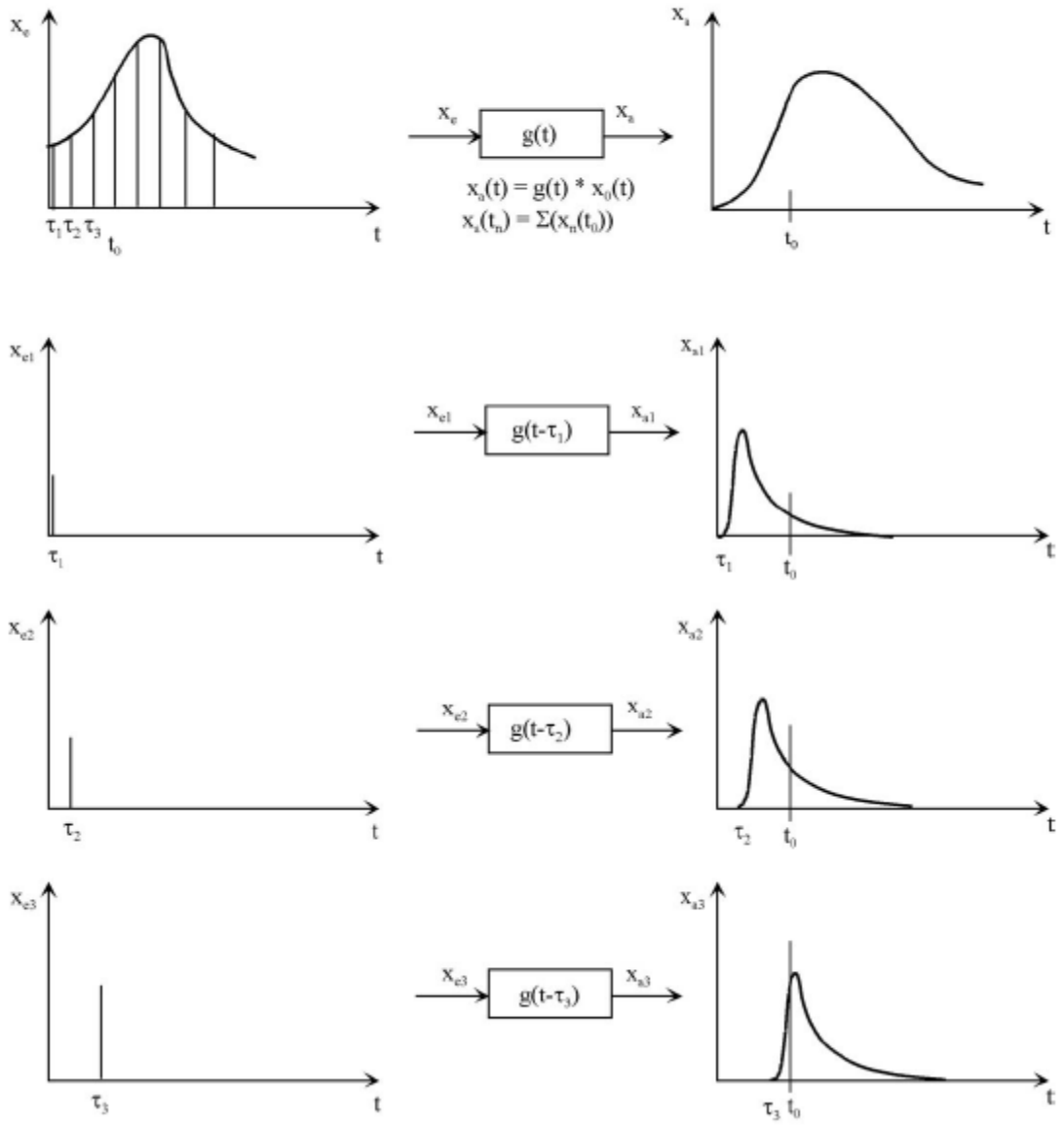


figure 12.15: overlay of individual step response function

Considering the connection between weight- and transfer function we can also carry out following identical transformation:

$$\begin{aligned}\frac{dh(t-\tau)}{dt} &= g(t-\tau) \\ x_a(t) &= \int_0^t \frac{d}{dt}(h(t-\tau))x_e(\tau)d\tau \\ &= \frac{d}{dt}\int_0^t h(t-\tau)x_e(\tau)d\tau \\ &= \frac{d}{dt}\left[h(t) * x_e(t)\right] \end{aligned} \tag{12.32}$$

With application of LAPLACE transformation the faltung operation changes into multiplication (see section 5.3.2 LAPLACE transformation, page 135):

$$\begin{aligned} L\left\{x_a(t)\right\} &= L\left\{g(t) * x_e(t)\right\} \\ &= L\left\{\int_0^t g(t-\tau)x_e(\tau)d\tau\right\} \\ &= G(p)\cdot X_e(p) \end{aligned} \tag{12.33}$$

In practice the numerical execution of faltung operation must be accomplished in a time quantization as its derivation. For a process, which begins from time t = 0, can be described in sum form introduced above as follows:

$$\begin{aligned} x_a(t_1) &= \Delta\tau g(\tau_1)x_e(\tau_1) \\ x_a(t_2) &= \Delta\tau g(\tau_2)x_e(\tau_1) + \Delta\tau g(\tau_1)x_e(\tau_2) \\ x_a(t_3) &= \Delta\tau g(\tau_3)x_e(\tau_1) + \Delta\tau g(\tau_2)x_e(\tau_2) + \Delta\tau g(\tau_1)x_e(\tau_3) \\ &\vdots \\ x_a(t_k) &= \Delta\tau \sum_{i=1}^{k} g(\tau_i)x_e(\tau_{k-i+1}) \end{aligned} \tag{12.34}$$

This equation system can be transformed in matrix equation:

$$\begin{bmatrix} x_a(t_1) \\ x_a(t_2) \\ x_a(t_3) \\ \vdots \\ x_a(t_k) \end{bmatrix} = \Delta\tau \begin{bmatrix} g(\tau_1) & 0 & 0 & \cdots & 0 \\ g(\tau_2) & g(\tau_1) & 0 & \cdots & 0 \\ g(\tau_3) & g(\tau_2) & g(\tau_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ g(\tau_k) & g(\tau_{k-1}) & g(\tau_{k-2}) & \cdots & g(\tau_1) \end{bmatrix} \cdot \begin{bmatrix} x_e(\tau_1) \\ x_e(\tau_2) \\ x_e(\tau_3) \\ \vdots \\ x_e(\tau_k) \end{bmatrix} \tag{12.35}$$

With different notations $t_k = k$ and $\Gamma_k = i$, and the introduction of aperture time $T = \Delta\Gamma$:

$$\begin{bmatrix} x_a(1) \\ x_a(2) \\ x_a(3) \\ \vdots \\ x_a(k) \end{bmatrix} = T \cdot \begin{bmatrix} g(1) & 0 & 0 & \cdots & 0 \\ g(2) & g(1) & 0 & \cdots & 0 \\ g(3) & g(2) & g(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ g(i) & g(i-1) & g(i-2) & \cdots & g(1) \end{bmatrix} \cdot \begin{bmatrix} x_e(1) \\ x_e(2) \\ x_e(3) \\ \vdots \\ x_e(i) \end{bmatrix} \tag{12.36}$$

The matrix equation can be written for short:

$$\mathbf{X}_a = T \cdot \mathbf{G} \cdot \mathbf{X}_e \tag{12.37}$$

or

$$\mathbf{X}_e = T^{-1} \cdot \mathbf{G}^{-1} \cdot \mathbf{X}_a \tag{12.38}$$

Thus G^{-1} is designated as inverse matrix of G.

12.3.4 Determination of weighting function g(t) for general case

The weighting function g(t) will be determined as follows:

$$g(t) = x_a(t) \mid_{x_e(t)=\delta(t)} \tag{12.39}$$

The experimental determination of g(t) were treated in the preceding sections. The determination of the weighting function can be achieved by a test attempt with an impulse as input function (see section 12.3.2 signal analysis, page 357). If a step function is used as input function, then weighting function must be obtained by appropriate differentiation (see section 11.4 transmission systems, table 11.5, page 310).

Experiments on real object will not always be accomplished for regulation of weighting function. Only in case the real input- and output signals can be obtained for computation of g(t). The matrix equation for calculation of output signal (see section 12.3.3 signal synthesis, page 359) can be used for regulation of g(t) or matrix G.

$$\begin{bmatrix} x_a(1) \\ x_a(2) \\ x_a(3) \\ \vdots \\ x_a(k) \end{bmatrix} = T \cdot \begin{bmatrix} g(1) & 0 & 0 & \cdots & 0 \\ g(2) & g(1) & 0 & \cdots & 0 \\ g(3) & g(2) & g(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ g(i) & g(i-1) & g(i-2) & \cdots & g(1) \end{bmatrix} \cdot \begin{bmatrix} x_e(1) \\ x_e(2) \\ x_e(3) \\ \vdots \\ x_e(i) \end{bmatrix} \tag{12.40}$$

If both the input- and output function are known for one observation period, the following matrix equation can be developed from the above equation system:

$$\begin{bmatrix} x_a(1) \\ x_a(2) \\ x_a(3) \\ \vdots \\ x_a(k) \end{bmatrix} = T \cdot \begin{bmatrix} x_e(1) & 0 & 0 & \cdots & 0 \\ x_e(2) & x_e(1) & 0 & \cdots & 0 \\ x_e(3) & x_e(2) & x_e(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ x_e(i) & x_e(i-1) & x_e(i-2) & \cdots & x_e(1) \end{bmatrix} \cdot \begin{bmatrix} g(1) \\ g(2) \\ g(3) \\ \vdots \\ g(i) \end{bmatrix} \tag{12.41}$$

The corresponding matrix equation:

$$\mathbf{X}_a = \mathbf{T} \cdot \mathbf{X}_e \cdot \mathbf{G} \tag{12.42}$$

or

$$\mathbf{G} = \mathbf{T}^{-1} \cdot \mathbf{X}_a \cdot \mathbf{X}_e^{-1} \tag{12.43}$$

Thus the weighting function g(t) can be described by a value sequence.

In metrological practice the regulation usually looks somewhat different. In above derivation it was presupposed that the process for t < 0 does not exist and started with the first input impulse. Real processes run off independently of observations. Therefore we must begin at arbitrary time point with the observation of in- and output signals. The determination accuracy of the weighting function g(t) must be specified by the administrator. The questions are the measuring expenditure and of the main dynamic process criteria for specification of the sample width T and the number of scanning values n. This is the same problem as discontinuous measuring signals treatment and error description (see GRÄBER: Lehrbrief Automatisierungstechnik).

If we always specify that *m* values are considered for the conditional equations, then 2 · m equations are to be set up, in order to determine *m* supporting places of the weighting function g(t). Hence we must already observe the process before explicit prognostication about the duration of 2 · *m* sampling intervals *T*, i.e. a period of 2 · *m* · T (see figure 12.16). According to this scheme following equation system can be set up:

$$x_a(t_4) = T\,(g\,(\tau_4)\,x_e\,(t_1) + g\,(\tau_3)\,x_e\,(t_2) + g\,(\tau_2)\,x_e\,(t_3) + g\,(\tau_1)\,x_e\,(t_4)) \tag{12.44}$$

$$x_a(t_5) = T\,(g\,(\tau_4)\,x_e\,(t_2) + g\,(\tau_3)\,x_e\,(t_3) + g\,(\tau_2)\,x_e\,(t_4) + g\,(\tau_1)\,x_e\,(t_5))$$

$$x_a(t_6) = T\,(g\,(\tau_4)\,x_e\,(t_3) + g\,(\tau_3)\,x_e\,(t_4) + g\,(\tau_2)\,x_e\,(t_5) + g\,(\tau_1)\,x_e\,(t_6))$$

$$x_a(t_7) = T\,(g\,(\tau_4)\,x_e\,(t_4) + g\,(\tau_3)\,x_e\,(t_5) + g\,(\tau_2)\,x_e\,(t_6) + g\,(\tau_1)\,x_e\,(t_7))$$

Thus we have four equations with four unknown weighting function portions, whereby the equation system is uniquely solvable. Since the observed values, generally measured values of the input signals as well as the output signals are erroneous, in practice more equations are built, which leads to an overdetermined equation system. They will be solved by means of special iterative methods, e.g. HOUSEHOLDER method. The solution is then the value range of weighting function portions, which fulfils the equation system with the smallest sum of square deviation.

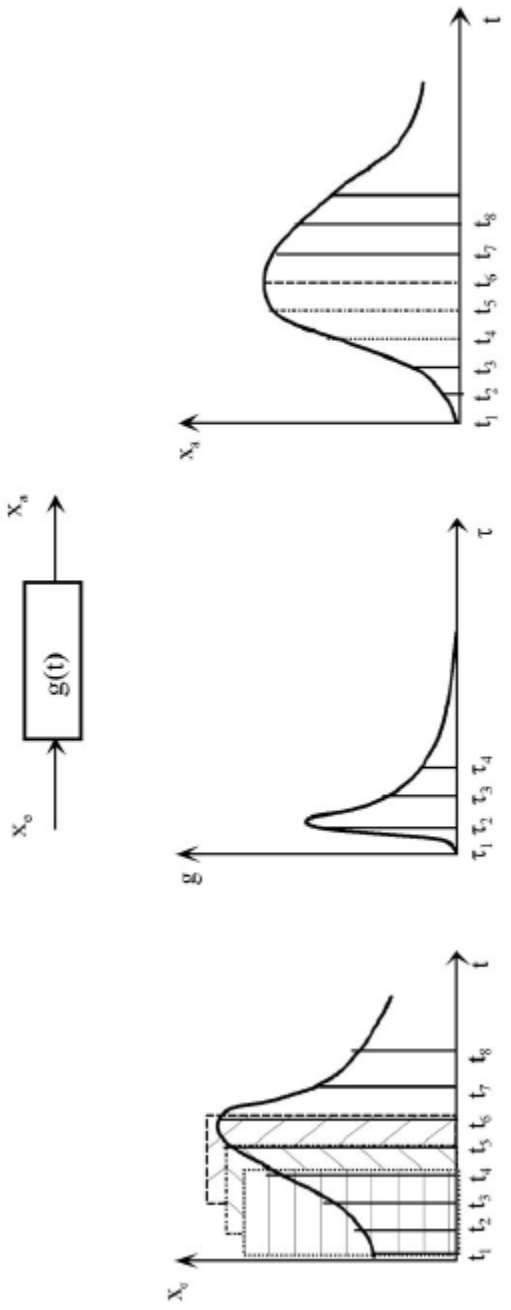


figure 12.16: formation of discontinuous response signals from measured values

12.3.5 Forecast models

In water management using models to prognosticate also shows a main field of application of faltung operation. The procedure of prognosis contains the algorithm as follows. A precondition of application faltung integral to prognosticate is that the process was already observed in a lead time, i.e. before a prognosis the in- and output signals of regarding system must be collected by means of suitable measurement. We also speak of the model learning curve in this time. These measured values serve for the determination of weighting function g(t), exactly g(i) or G. How long does the learning curve last, depends on the historical data base, the precision requirement and the desired numerical expenditure.

In the following examples the manipulation will be demonstrated (see figure 12.16, page 365). In this case measured input signal values exist for a range of seven sampling intervals before the forecast horizon. The output signal was measured from the fourth interval. Based on these values following four equations with unknown quantities g1 to g4 can be formulated. Therefore only four supporting places of the weighting function will be proceeded in this example. For real practical tasks this is quantized too roughly:

$$x_{a4} = T \cdot (g_4 x_{e1} + g_3 x_{e2} + g_2 x_{e3} + g_1 x_{e4}) \qquad (12.45)$$

$$x_{a5} = T \cdot (g_4 x_{e2} + g_3 x_{e3} + g_2 x_{e4} + g_1 x_{e5})$$

$$x_{a6} = T \cdot (g_4 x_{e3} + g_3 x_{e4} + g_2 x_{e5} + g_1 x_{e6})$$

$$x_{a7} = T \cdot (g_4 x_{e4} + g_3 x_{e5} + g_2 x_{e6} + g_1 x_{e7})$$

By means of suitable methods for the solution of equation system we get the weighting function portions g1 to g4. These are used into a conditional equation for the first prognosis time step (x_{a8}). Thus the prognosis value can be computed explicitly:

$$x_{a8Prog} = T \cdot (g_4 x_{e5} + g_3 x_{e6} + g_2 x_{e7} + g_1 x_{e8}) \qquad (12.46)$$

Parallel to prognosis the process should be further supervised metrologically. In this case we get a new value pair x_{e8Prog} and x_{a8gem} at time point 8. This can be used to calculate new weighting function portions. With retention quantity of weighting function portions the input signal x_{e1} will not be incorporated into calculation any longer. The first equation with the input signal x_{e1} can however remain in the computation, and then five equations are available for the determination of four weighting function portions. This overdetermined equation system is then solved iteratively with an appropriate method, e.g. HOUSEHOLDER method. That founded values represent the transient characteristic of the regarding system is possibly better than those with the definite system. This comparison between prognosis and real process is also called constant learning system.

12.3.6 Tasks of application of faltung integral:

1. The following groundwater levels were measured in a pumping test (see figure 12.17)

a) calculate water deficit (volume) of drawdown funnel, if the aquifer has the following characteristic values:
$h_n = 16m$, $M = 10m$, $k = 0.001m \cdot s^{-1}$, $S_0 = 0.0001m^{-1}$, $n_0 = 0.20$
b) Compute by means of transfer element method and with a) founded value for the flow rate V the drawdown curve for conveyor capacity of $0.005m^3 \cdot s^{-1}$.

Set up the first four equations of faltung integral for the model until observation time point t = 1d.

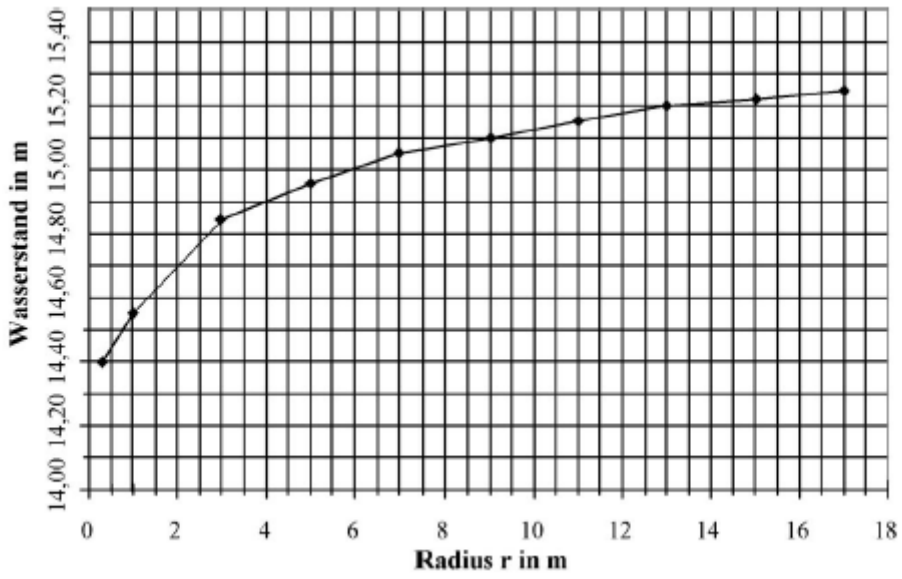


figure 12.17: groundwater level dependent on radius

2. Following groundwater level are measured in a pumping test:

$\dot{V}\left[\frac{dm^3}{s}\right]$	0	15	15	15	15	15	15	15	15	15
s $[cm]$	16,00	15,25	15,12	15,07	15,01	14,96	14,95	14,94	14,94	14,938
t $[min]$	0	15	30	45	60	75	90	105	120	135

The aquifer has following parameters:
h_n = 16m, M = 10m, k = 0.001m · s^{-1}, S_0 = 0.0001, n_0 = 0.20
Set up the first four equations of faltung integral for the model until observation time point t = 135min.

3. Prognosticate the temperature pattern in a bank filtration frame with application of faltung integral, if the following measured values are known:

$\delta^o_{Fluss}\,[°C]$	14,2	16,0	17,7	19,4	17,2	16,0	17,6	18,6	14,8	12,0	13,7
$\delta^o_{Fass.gemes.}\,[°C]$	8,5	10,0	11,4	14,0	14,1	14,7	15,4	15,8	15,6	14,9	14,1
$\delta_{Fass.progn.}\,[°C]$											
Zeit $[d]$	0	15	30	45	60	75	90	105	120	135	150

Calculate the temperature pattern from the sixth time step with application of three faltung integral equations in each case.
Compare the calculated temperature in the frame to the measured and correct the weighting function with consideration of measured values.

Part IV

Indirect Parameter identification

The indirect parameter identification is treated here as method for parameter estimation. It stands in contrast to physical and chemical methods for the direct Parameter estimation which is described in GRÄBER " Grundwassermesstechnik". The methods of indirect parameter identification are mathematical, which according to experimental process analysis determines parameter for a transmission system. The transient characteristic can be found by means of experimental or theoretical process analysis. Accordingly the identified parameters are more or less physically/chemically interpretable. On all accounts parameters can be found, which well reflect the system behaviour within validity scope of experiment.

Chapter 13

13 Estimation procedure

In water management practice the experimental process analysis is primarily used for parameter estimation. The model structure is specified by a theoretical process analysis. We try to transfer this model structure into simple mathematical representation. The parameters can be determined by solving conditional equations or by solving parameter approximation problems. Thus there is task, on the basis of structure cognition or -assumption, to determine such models, that

- reflects the characteristics of the system so exactly as required and
- eliminates the overlaid influences of noise and errors.

In order to satisfy these demands the comparison of output values of original function as input function or an independent variable (time or place) with the model is accomplished. In the result a change of model parameters is to be made or the model changes until the deviation reaches minimum. The changes can be achieved according to a certain strategy (search algorithms, optimisation programs), statistically (random number generator) or empirically. The visual comparison between two diagrams (original- and model output signal) is also possible.

This task is also designated as parameter estimation. In particular the following introduced approaches will be classified as **iterative estimation method**.

In the algorithm or iterative model adjustment (see figure 13.1) we try to let the same input vector, the manipulated vector y affect on the process and model. With a first parameter substitution, the initial parameter, the model output vector x^1_M can be calculated as first approximation. The deviation of these process output vectors is designated as quality of the model adaptation. In water management applications the quadratic evaluation will be carried out. The goal of changing parameters is to minimize the Q value.

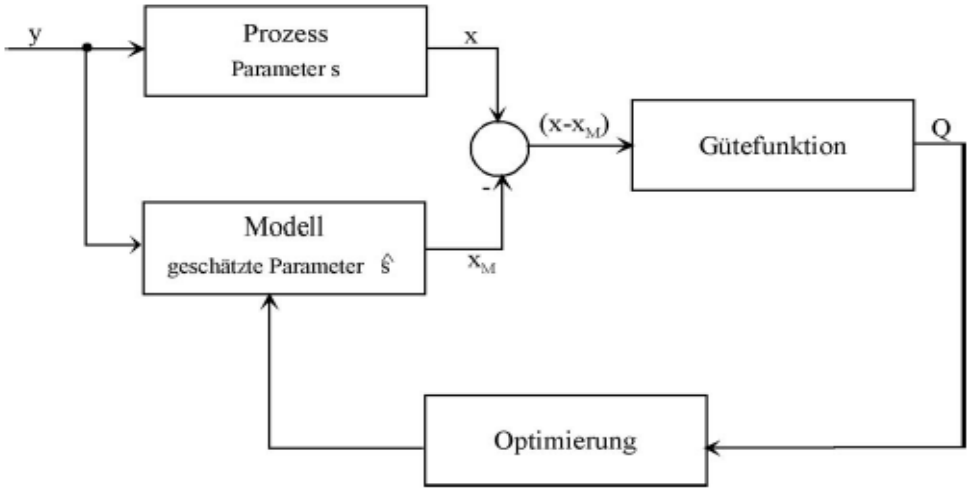


figure 13.1: iterative model adjustment

Chapter 14

14 Flow parameters

14.1 Pumping test evaluation

14.1.1 Fundamentals

The evaluation of pumping test, with which e.g. water is conveyed from a well and the drawdown is recorded as a discontinuous function of place and time, is very significant and representative method to determine geohydraulic parameter.

Compared to laboratory procedures it has advantages that:

- is accomplished in undisturbed aquifer and
- advance normally integral statements of the aquifer in regarded flow field section.

The parameter estimation of aquifer occurs in the direct **laboratory experiment** of soil samples, which can be achieved by means of Stechzylinder or cuttings. The disadvantages consist the fact that only a punctiform parameter estimation can be obtained in very inhomogeneous aquifer with this method. Besides the granular structure of soil is destroyed by the sampling and thus another is evaluated in lab. A third difference is that in lab the entire water content is determined, while in nature and with pumping test only the drainable pore volume affects. The representation by means of definite parameter method increases due to the integral character of the pumping tests (see table 14.1).

Table 14.1: difference between pumping test evaluation and laboratory method

Charaktereigenschaft	Labormethode	Pumpversuchsauswertung
Örtliche Ausdehnung	punktuell	integral
Repräsentanz	klein	groß
Korngerüst	zerstört	ungestört
Speicherkapazität	Gesamtwasservolumen	entwässerbares Volumen
Aufwand	relativ gering	sehr hoch

On the other hand the pumping tests are substantially more complex and expensive than laboratory test. Therefore the experimental design, execution and special worthy analysis must be attached. Also usually only one time test execution is possible.

For the evaluation of such pumping tests in practice particularly two methods are used:

- the graphic method; in water management practice designated as straight line method and typical curve method and
- the search method or optimisation method.

The well incident flow is used as model in the pumping tests. As in the section 8.1 THEIS well equation, page 196 deduced, the partial differential equation solution according to THEIS is suitable for the computation of drawdown processes due to water extraction from well. This model of course can be only assumed approximately for practical flow conditions. The application of model is prohibited to better reflect the original due to too large number of free parameters to adapt. The most substantial restriction of analytical models is the consideration of only one aquifer. The constant of parameter transmissibility *T* and storage coefficient *S* as well as the horizontal bed situation can be presupposed in the little spatial expansion of pumping tests as given. Of course it must be also noted that the transmissibility change during drawdown procedure remains negligibly small (linearization around operating point).

The quality function *GF* in the pumping test evaluation is defined as sum of the square deviation of the measured values at the original process and model results on different local- and time points (see figure 14.1):

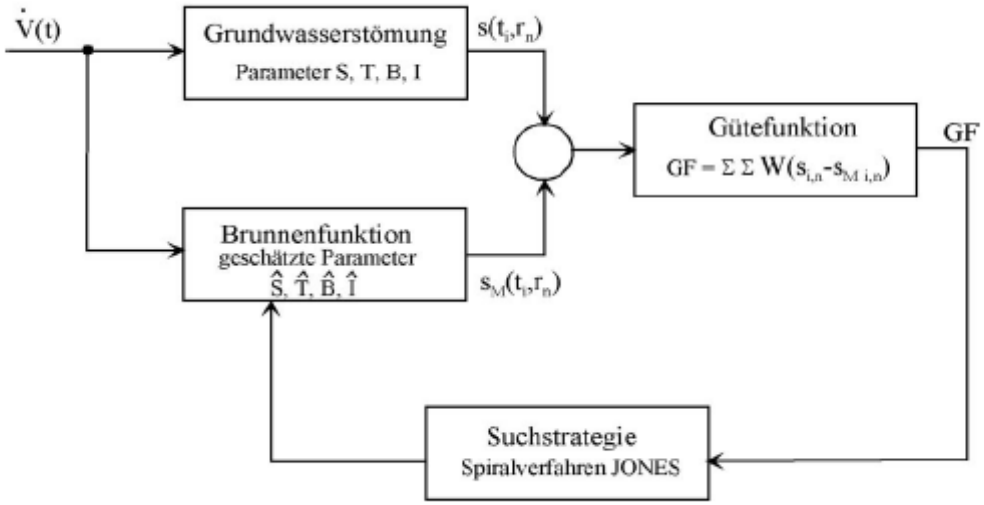


figure 14.1: iterative model adjustment in a pumping test

$$GF = \sum_{i=1}^{n} \sum_{j=1}^{m} W_{i,j} \cdot (s_{i,j} - s_{Mi,j})^2 \qquad (14.1)$$

With:

$W_{i,j}$ weighting factor

s original drawdown value

s_M model drawdown value

m maximal number of time steps

n maximal number of local observation points

14.1.2 Practical realisation

The method of least squares (MKQ) is used for parameter adjustment to the measured values. The goal is to minimize the quality function *GF*. General correspond effective methods were shown in section 4.3 least square method, page 96. It shows that the search strategy on that basis of the nonlinear regression with the utilization of gradients is best suitable for pumping test evaluation. In contrast to ROSENBROCK search algorithm the number of search steps will be drastically (factor 10) reduced by JONES (DAMMERT) spiral method in pumping test evaluation. This procedure presupposes that the quality function is constant and differentiable. Both are given in the analytical solution of well function according to THEIS.

The program system PSU (Pumping test evaluation) was developed by BEIMS/GRÄBER for practical realization. This program system (see figure 14.2) has a modular structure, which allow arbitrary model creations of quality function and search algorithm coupled with appropriate main programs.

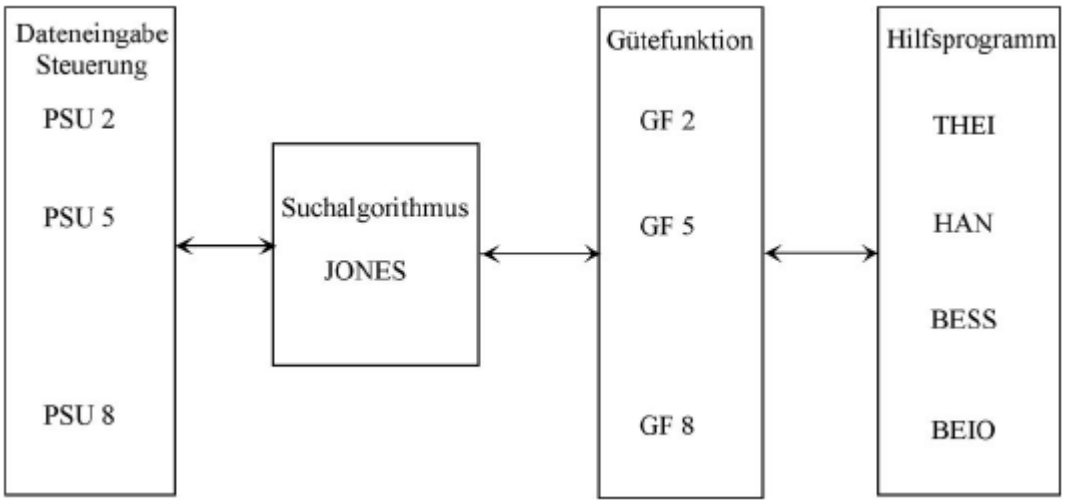


figure 14.2: programme system for pumping test evaluation according to BEIMS/GRÄBER

All the most substantial practical pumping tests can be evaluated with the following specified versions PSU2, PSU5 and PSU8 (see table 14.2).

Table 14.2: Realised programme versions with geohydraulic scheme

Programm	Geohydraulisches Schema	Ergebnisse
PSU2	Unendlich ausgedehnter Grundwasserleiter ohne Speisung	T, S
PSU5	Unendlich ausgedehnter Grundwasserleiter mit Speisung	T, S, B
PSU8	Einseitig begrenzter Grundwasserleiter ohne Speisung	T, S, λ^*

In table 14.2:

$T\ [\frac{m^2}{s}]$	Transmissibility, profile permeability
$S\ [-]$	Storage coefficient
$B\ [m]$	Supply factor
$\lambda^*\ [m]$	Effective boundary condition distance

The programs PSUX are written in the form of main program and realise data in- and output as well as the search algorithm control. The fitted values of each search step or only the parameters could serve as output by inserting appropriate control variables, which yield the adjustment according to given error bound. Furthermore a graphic comparison between the measured values

of pumping test and the adapted drawdown curve is possible. The search program was realized according to JONES spiral method. It acquires the minimum of the quality function on the basis of nonlinear regression with employment of differential (see figure 14.3 and 14.4). The quality function is to be selected according to the geohydraulic conditions. It is defined as the sum of square deviations between the measured value and the theoretical drawdown curve. In the module the auxiliary programmes are combined with important subprograms to solve well flow equation, e.g. the well functions $W(\sigma)$ according to THEIS and $W(\sigma, B)$ according to HANTUSCH as well as BESSEL function $K_0(x)$ and $I_0(x)$.

A special problem exists in the search of parameters B (supply factor) and λ^* (effective boundary condition distance), since they are not independent of T and S. The search algorithm is however then applied for more parameters if they are independent of each other. In these cases it is a trick to exclude the region in drawdown curve, which depend on different parameters dominantly. So physically it can be justified that, the drawdown strongly depends on the transmissibility T and the storage coefficient S of well vicinity area in the initial phase of a pumping test. In the quasi steady phase the supply factor B and/or the boundary condition (effective boundary condition distance λ^*) work as further influence variables. A so called stage search is accomplished based on this drawdown curve classification in different phases. The parameters T and S will be searched in phase 1. The measured values of the phase 2 serve to the estimation of supply factor and/or the effective boundary condition distance. The parameters T and S will be applied as known quantity (determined from phase 1) during this phase and thus are not included in search process. The measured values from arising process are combined in a phase 3. With them an improvement of adapted values can be obtained. In this phase the parameters of phase 2 will be as known, i.e. as not adjustable, regarded and again only one search for the two values T and S is carried out.

The pumping test evaluation with the programs PSUX can be of course only as good as measured values; the drawdown values from pumping test are as good as the well incident flow equation model reflects natural processes. For complex geohydraulic conditions we must resort to others, e.g. the pumping test simulator.

The expressiveness of pumping tests or experimental process analysis method generally also depends on the used test signal. In classical pumping test this is a step function with the step height V, the conveyed water quantity. The best results can be achieved by using a DIRAC impulse (theoretical impulse with a infinite height and a length of time, which approaches to zero). This is technically not realizable. As compromise the impulse function, the step function carry periodic signals and stochastic signal sequences. The step function is favourable for the determination of final steady state, the so called static behaviour.
A combination of different test signals in the Variants

- step function - impulse function,
- step function - periodic signals or
- step function - stochastic signal sequences

leads to effective determinations of dynamic transition- and static final state.

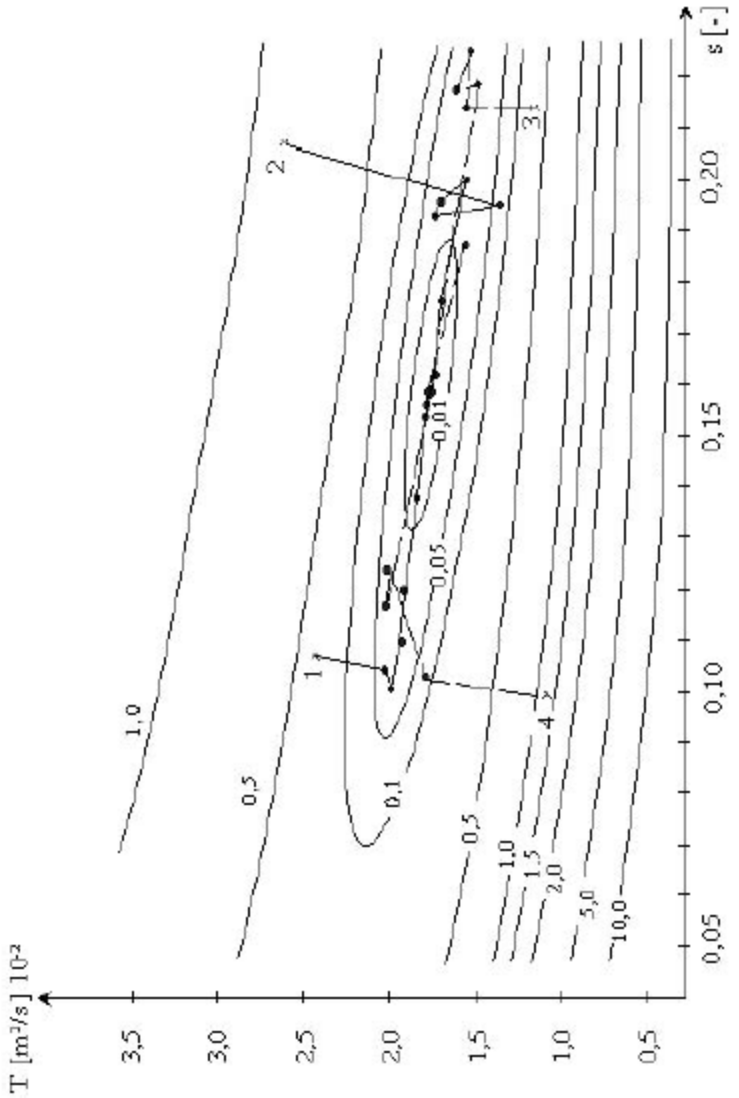


figure 14.3: quality mountain in a pumping test evaluation with search procedure of different start points

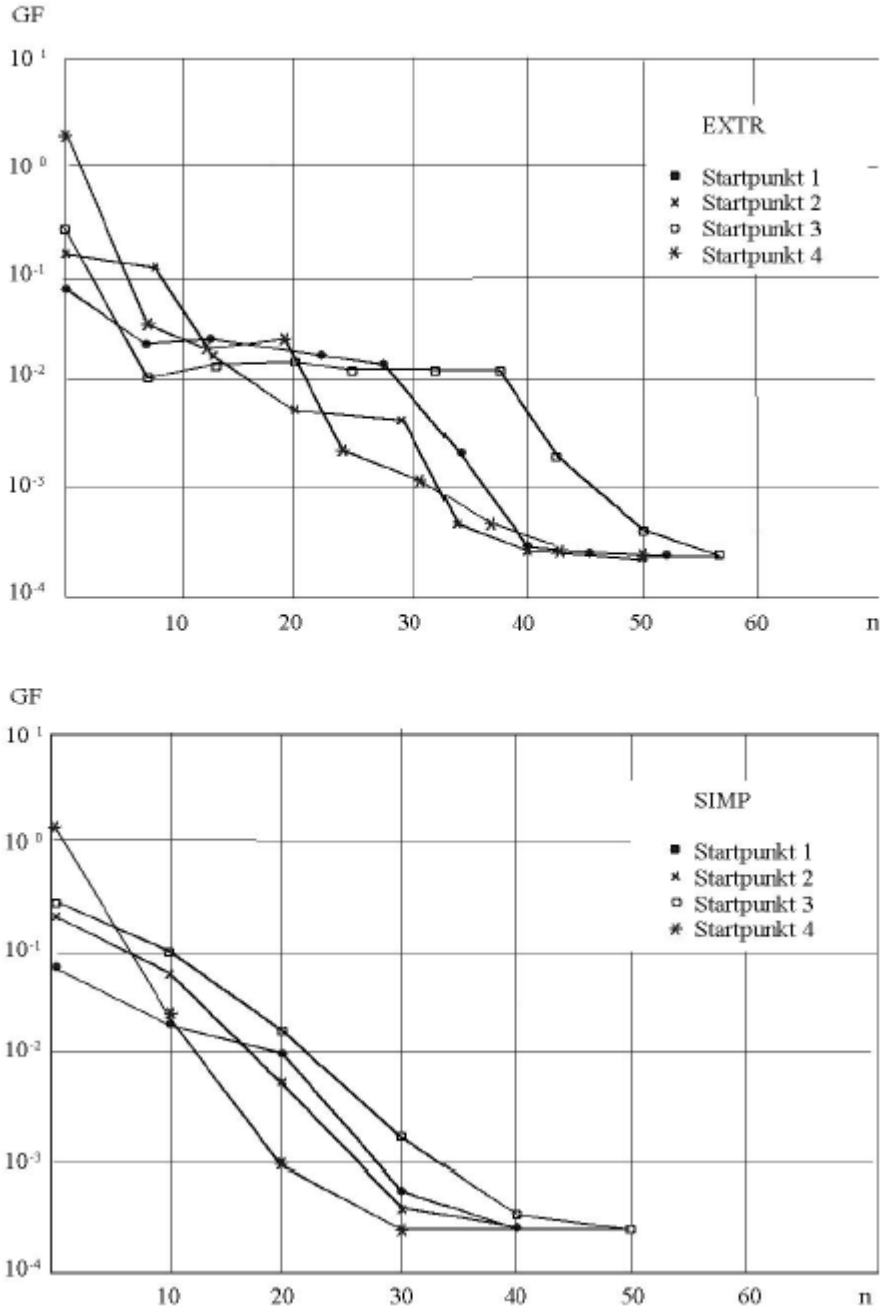


figure 14.4: Iteration performance in the pumping test evaluation

14.2 Pumping test simulator

The program packet PSUX for the evaluation of pumping tests with corresponding analytical solutions on the basis of THEIS well function or HAUTUSCH demonstrated above shows considerable restrictions. So not all characteristics of the aquifer such as anisotropy, stratification, heterogeneity or capillary space and not all kinds of characteristics of well design such as imperfectness, well diameter, well bottom inflow, filter losses can be considered with this system. Either with application of natural signals for parameter identification, or with artificially produced test signals like in the pumping tests, it has to be assumed that normally it is a matter of one time procedure, which is not reproducible or changeable due to cost and other reasons. Therefore it is necessary to consider such field tests intensively with the experimental design.

A numerical model was described by MUCHA/PAULIKOVA, with which and pumping test can be simulated and interactively evaluated. This model is based on a vertical plane rotation symmetrically quantized flow model, which does not consider simplified assumption. This was converted into a program system WELL, the so called pumping test simulator by BEIMS/GRÄBER. With it the effect of a pumping test can be demonstrated and optimised on the basis of hypothetical assumption of regarding area. With this model besides the inhomogeneity the existence of multiple aquifers and also key elements well vicinity area can be considered. The transmissibility can be considered in horizontal and vertical inhomogeneity. Furthermore the specific elastic as well as the gravimetric storage coefficient can be processed. The model takes free groundwater flow conditions as a compressible system and the free surface as a mobile border. The transmissibility is computed according to the concrete position of free surface. On the last radius point r_n the system is regarded as impermeable, i.e. the discretisation must be chosen in such a way that practically no drawdown appears there (see figure 14.5, page 392).

On the basis a graphic display the effect of different input signals can be demonstrated and at the same time the optimal measuring time points dependent on the distance and the drawdown gradients for real pumping test can be determined. The local situation of the level observation tubes is usually default due to technological conditions.

The pumping test simulator WELL will be applied for following fields:

- simplification and assumption analysis, which underlie different analytical solutions.
- determination of flow- and speed relationship in the proximity of well.
- calculation of typical curves for special well- and aquifer conditions.
- interactive pumping test evaluation.

The employment of pumping test simulator WELL represents a substantial complement of pumping test evaluation and -interpretation. The exertion as model in the indirect parameter identification can be only realised by means of specially large expenditure.

The pumping test simulator is a vertical finite difference model, whose quantization can be obtained in vertical direction according to geological stratification and well geometry. In horizontal direction quantization will be carried out logarithmically ($r_i = r_{i+1} \cdot 10^{0.25}$). The junctions lie in the centre of gravity of the reticule, i.e. they are in contrast to the geometrical centre around Δr_i outwards shifted:

$$\Delta r_i = \frac{(r_{i+1} - r_i)^2}{6(r_{i+1} + r_i)} \tag{14.2}$$

The permeability values are defined as hydraulic conductance between the knots. The conductance in horizontal direction, for example between the knots 5/4 and 6/4, under the DUPUIT THIEM equation assumption for groundwater flow to a well is:

$$TF_{5/4} \rightarrow = \frac{2\pi k_{h,4} b_4}{ln\left(\frac{r_6}{r_5}\right)} \tag{14.3}$$

With

$k_{h,4}$ horizontal permeability coefficient of 4th discrete layer

b_4 thickness of 4th layer

r_5, r_6 radii of knots 5 and 6

The conductance in vertical direction, for example between the knots 6/2 and 6/3

$$TF_{6/2} \uparrow = \pi(r_{7^2} - r_{6^2}) \left(\left(\frac{2k_{v,2}}{b_2}\right) + \left(\frac{2k_{v,3}}{b_3}\right)\right) \tag{14.4}$$

$$TF_{6/2} \uparrow = \frac{2\pi(r_{7^2} - r_{6^2})\,(k_{v,2}b_3 + k_{v,3}b_2)}{b_2 b_3} \tag{14.5}$$

With

$k_{v,2}$, $k_{v,3}$ vertical permeability coefficient of 2nd or 3rd layer

b_2, b_3 thickness of 2nd or 3rd layer

r_6, r_7 radii of element border 6 and 7

The conductance then yields flowing water quantity, for example between the knots5/4 and 6/4:

$$\dot{V}_{5,4,6,4,} = TF_{\not 34} \rightarrow (H_{5/4} - H_{6/4}) \tag{14.6}$$

whereby $H_{5/4}$ and $H_{6/4}$ are the piezometer head in knots 5/4 and 6/4. The storage coefficient *S* designates the relationship of the volume in unit water, which becomes 1m empty during gauge level change, to the total volume of this unit. So the storage factor in the knot 6/5 is e. g.:

$$SF_{6/5} = S_{s,5} b_5 \pi (r_7^2 - r_6^2) \tag{14.7}$$

$S_{s,5}$ specific elastic storage coefficient of 5th layer

b_5 thickness of 5th layer

r_6, r_7 radii of corresponding elements

The storage coefficient for the free waster surface e.g. at knot 5/1 is:

$$SF_{5,1} = S_y \pi (r_6^2 - r_5^2) \tag{14.8}$$

S_y gravitation storage coefficient

The released water volume for knot 5/1:

$$V_{5,1} = SF_{5,1} (H_{5,1,t} - H_{5,1,t-\Delta t}) \tag{14.9}$$

$H_{5,1,t}$ and $H_{5,1,t-\Delta t}$ are potentials at knot 5/1 to time point t and t - Δt, whereby Δt is time interval. The storage factor for well is expressed by knot 1/1:

$$SF_{1,1} = \pi r_{1,j}^2$$

$r_{1,j}$ effective well radius of *j*-th layer

The linear well loss is contained in coefficient φ:

$$\text{Well loss} \quad = \frac{\dot{V}}{4\pi T}\varphi$$

The flow from 4th layer in the well is expressed by the factor:

$$TF_{1/4} = \frac{2\pi k_{h/4} \cdot b_4 c/a}{ln\left(\frac{r_2}{r_1}\right) + \frac{\varphi}{2}}$$

Whereby c/a is the relative position of junction 1/4 between water level in well and water level in the knot 2/1. The flow in the well can take place through filter or well bottom.

The time discretisation begins with a small increment Δt and will automatically increase according to rules.

$$\Delta t_{i+1} = \Delta t_i \cdot 10^{0,1} \qquad (14.10)$$

If the discharge flow is not constant, the input of time steps and flow rate are achieved on the basis of each calculation period.

The resulting band matrix with the five diagonal elements will be solve according to a direct method (GAUSS method).

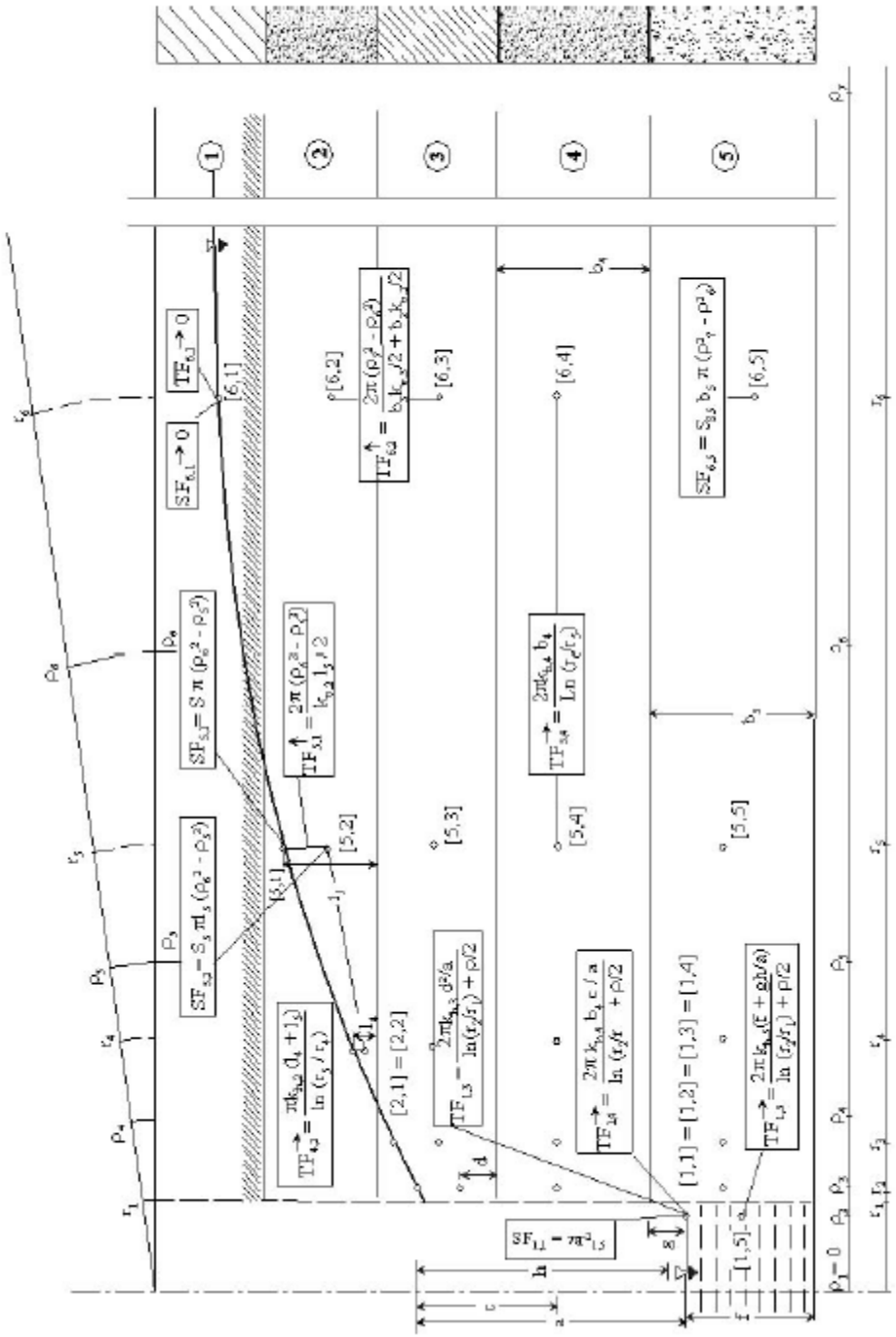


figure 14.5: structure scheme of pumping test simulator (BEIMS/MUCHA/GRÄBER)

Chapter 15

15 Suction power distribution

Another application field of the parameter identification methods in soil system exists in laboratory scale determination of suction power saturation distribution (SSV) within the unsaturated region. The measuring method for this soil behaviour is specified in GRÄBER " Grundwassermesstechnik" and the mathematical model is deducted in the section 6 differential equations, page 175 groundwater process.

Typical suction power saturation behaviour is connection similar to hysteresis between the suction power p_k and the saturation n_b. As mathematical model these SSV curves were indicated by an equation from LUCKNER/SCHESTAKOV, whereby A, B, C, and D are four constants, which must be separately determined from the experimentally technically gained upper (drainage) and lower (irrigation) limiting curves of hysteresis branches:

$$\Theta = \frac{\Theta - A}{n - A - B} = \left[\frac{1}{1 + C_{pc}^{D}}\right]^{(1-1/D)} \qquad (15.1)$$

Determination of these parameters comes into the indirect parameter identification task range, whereby in this case it is a matter of static characteristic curve approximation. On this account no conclusions about the test signal type have to be made. The dynamic behaviour corresponding to the partial differential equation of this process is not yet evaluated at present.

The estimation of the four parameters thereby can be achieved according to the empirical graphic method (method of typical curves) or the mathematical search algorithm. The first approach is used in order to obtain the initial value for the search strategy in the experimental design phase and the other is to reach optimal test conditions (measuring point selection). The mathematical search algorithm is then used to approximate the mathematical model at founded measured values pairs (p_k, n_b) as accurately as possible. Also the adjustment is made here by the square quality factor, which represents the deviations between n_0 and n_m in all adjusted operating points (pressure ranges). FIBONACCI method is applied for the minimum search. This has the advantage that it is relatively " robust ", but stagnates with a very rough approximation. This behaviour also depends on the relative flatness of the quality mountain. It is therefore suggested applying POWELL method. It shows good convergence behaviour. Since the quality mountain, due to its abstract formulation, is constructed in such a way that, minima exist in the range, which do not allow physically meaningful interpretation (e.g. negative saturation), and special weighting function will be introduced. As soon as physically conditional limits for the saturation are reached (air or water restsaturation), the quality function acquires appropriate maximal value. The gradient method of the nonlinear regression, as favourably used in pumping test evaluation, conks out here, since the quality mountain runs very flatly and shows kinks at physically conditional edges, which contradicts the required differentiability.

BIBLIOGRAPHY

[Bau63] BAULE, B.: Die Mathematik Des Naturforschers und Ingenieurs, I - IV. S. Hirzel Verlag, Leipzig, 1963.

[Bea79] BEAR, J.: Hydraulics of Groundwater. McGraw-Hill, 1979.

[Bey73] BEYER, O. U.A.: Mathematik Für Ingenieure, Naturwissenschaftler, Ökonomen und Landwirte. BSG B. G. Teubner Verlagsgesellschaft, Leipzig, 1973.

[Bus72] BUSCH, K.-F.; LUCKNER, L.: Geohydraulik. Deutscher Verlag für Grundstoffindustrie, Leipzig, 1972.

[Bus93] BUSCH, K.-F., LUCKNER L. TIEMER K.: Geohydraulik. Lehrbuch der Hydrogeologie, Bd. Gebrüder Bornträger, Berlin/Stuttgart, 1993.

[Dom90] DOMENICO, P. A.; SCHWARTZ, F. W.: Physical and Chemical Hydrogeology. John Wiley Sons, 1990.

[Dyc89] DYCK, S.; PESCHKE, G.: Grundlagen der Hydrologie. Berlin, 1989.

[Fet] FETTER, C. W.: Applied Hydrology. Prentice Hall, Englewood Cliffs, NJ 07632.

[Fre79] FREEZE, R. A.; CHERRY, J. A.: Groundwater. Prentice Hall, Englewood Cliffs, NJ 07632, 1979.

[Fri88] FRITZSCH, W.; KLOSE, J.: Mathematische Modellierung, Simulation und Optimierung. Taschenbuch Maschinenbau, Bd. 4, Abschn. Verlag Technik, Berlin, 1988.

[Gel68] GELLERT, W. U.A.: Kleine Enzyklopädie Mathematik. Bibliographisches Institut, Leipzig, 1968.

[Häf92] HÄFNER, F.; SAMES, D.; VOIGT H.-D.: Wärme- und Stofftransport. Springer-Verlag, Berlin/Heidelberg/New York, 1992.

[Hea88] HEATH, R. C.: Einführung in Die Grundwasserhydrologie. R. Oldenburg Verlag, München, Wien, 1988.

[Höl92] HÖLTING, B.: Hydrogeologie. Stuttgart, 1992.

[Kin] KINZELBACH, W.: Numerische Methoden Zur Modellierung Des Transportes Von Schadstoffen im Grundwasser.

[Kin95] KINZELBACH, W.; RAUSCH, R.: Grundwassermodellierung. Gebrüder Bornträger, Berlin/Stuttgart, 1995.

[Kru73] KRUSEMANN, G. P.; RIDDER DE, N. A.: Untersuchungen und Anwendungen Von Pumpversuchsdaten. Verlagsgesellsch. R. Müller, Köln-Braunsfeld, 1973.

[Kru89] KRUG, W.: Simulation Für Ingenieure in CAD/CAM-Systemen. Verlag Technik, Berlin, 1989.

[Lan] LANGGUTH, ; VOIGT, :: Hydrogeologische Methoden. Springer Verlag, Berlin/Heidelberg.

[Luc86] LUCKNER, L.; SCHESTAKOW, W. M.: Migrationsprozesse im Boden- und Grundwasserbereich. Deutscher Verlag für Grundstoffindustrie, Leipzig, 1986.

[Mat93] MATHESS, G.: Lehrbuch der Hydrogeologie, 1 - 3. Gebrüder Bornträger, Berlin/Stuttgart, 1993.

[Md86] MARSILY DE, G.: Quantitative Hydrogeology. Academic Press, Orlando, 1986.

[NZ94] NE-ZHENG, SUN: Inverse Problems in Groundwater Modeling. Kluwer Academic Publisher, London, 1994.

[Phi76] PHILIPPOW, E.: Taschenbuch Elektrotechnik, 1. Verlag Technik, Berlin, 1976.

[Pin77] PINDER, G. F.; GRAY, W. G.: Finite Element Simulation in Surface and Subsurface Hydrology. Academic Press, New York, 1977.

[Rem71] REMSON, I.; HORNBERGER, G. M.; MOLZ F. J.: Numericals Methods in Subsurfaces Hydrology. WIley-Interscience, New York, 1971.

[Töp87] TÖPFER, H. ; BESCH, P.: Grundlagen der Automaitisierungstechnik. Verlag Technik, Berlin, 1987.

[Ver70] VERUJIT, A.: Theory of Groundwater Flow. Macmillian, London, 1970.

[Wan82] WANG, H. F.; ANDERSON, M. P.: Introduction to Groundwater Modeling, Finite Difference and Finite Elements Methods. W. H. Freemann and Company, San Francisco, 1982.

INDEX