4. Setting up the Observation Equations

The purpose of the satellite missions CHAMP, GRACE, and GOCE is to determine the gravity field of the Earth as accurately as possible. Therefore, a relationship has to be established between the (unknown) gravity field parameters and the provided observations. This is obtained via a functional model that links the particular observations to the gravity field parameters. Combining the observation equations for the different observation times leads to a (linearized) system of equations that can be solved by a least-squares adjustment procedure. In the first part of this chapter, it will shortly be reviewed how the unknown parameters can be estimated from a given set of observation equations in general. Subsequently, different gravity field functionals will be expressed in terms of a modeling by radial basis functions. This is a necessary tool to set up the functional models for a regional analysis. These models will then be described in the following section, leading to the observation equations for the different types of satellite observations. In detail, these types of observations can be characterized as satellite-to-satellite tracking in the high-low and in the low-low mode and by the concept of satellite gravity gradiometry.

4.1 Least Squares Approximation

In the following, it will be described how the unknown gravity field parameters are estimated from a given set of observations. This estimation process represents a standard Gauss-Markoff-Model, as described, for example, by Koch (1997) or Niemeier (2002). All \( n \) observations can be arranged in a vector \( \vec{y} \) with the dimension \( n \times 1 \). The \( u \) unknown parameters can also be combined in a column vector \( \vec{x} \) with the dimension \( u \times 1 \). The observations can be linked to the unknown parameters via a functional model \( f(\vec{x}) \) and can then be formulated in terms of this model and an additional measurement noise \( \epsilon \) according to

\[
\vec{y} = f(\vec{x}) + \epsilon. \quad (4.1)
\]

Typically, the number of observations \( n = \dim \vec{y} \) is considerably larger than the number of unknown parameters \( u = \dim \vec{x} \). If the model is non-linear, a linearization becomes inevitable, therefore approximate values for the unknown parameters have to be introduced, and approximate observations can be calculated as functions of the approximate values. In case of a linear model, it is recommended to start with approximate values for the unknown parameters as well and to calculate their influence on the observations,

\[
y_0 = f(x_0). \quad (4.2)
\]

The linearization procedure of the model \( f \) with respect to the unknown parameters can be performed by a Taylor expansion truncated after the linear term,

\[
\vec{y} = y_0 + \left. \frac{\partial f(\vec{x})}{\partial \vec{x}} \right|_0 (\vec{x} - x_0) + ... \quad (4.3)
\]

The reduced observations and the corrections to the unknown parameters are then calculated according to

\[
y = \vec{y} - y_0 \quad \text{and} \quad x = \vec{x} - x_0. \quad (4.4)
\]

The partial derivatives of the linearization (4.3) can be combined in the design matrix \( A \). It is of dimension \( n \times u \), and its elements are defined by the partial derivatives of the function \( f_k(\vec{x}) \) with respect to \( x_i \),

\[
(A)_{ki} = \left. \frac{\partial f_k(\vec{x})}{\partial x_i} \right|_0. \quad (4.5)
\]

This leads to the linear system of equations

\[
y = Ax + \epsilon \quad \text{with} \quad C(\epsilon) = \sigma^2 P_\epsilon^{-1}. \quad (4.6)
\]
The standard Gauss-Markov model is based on the assumptions that the measurement errors have an expectation of zero and that the covariance matrix of the observations is known a-priori,

\[ E\{\epsilon\} = 0 \quad \text{and} \quad C(\epsilon) = C(y) = \sigma^2 P_{\epsilon}^{-1}. \quad (4.7) \]

Here \( \sigma \) stands for the unknown variance factor, and \( P_{\epsilon} \) denotes the weight matrix of the observations. An estimation by least squares adjustment corresponds to the minimization of the square sum of the residuals,

\[ \Omega = \frac{1}{\sigma^2} (y - Ax)^T P_{\epsilon} (y - Ax). \quad (4.8) \]

The minimum condition can be obtained by differentiation according to

\[ \frac{\partial \Omega}{\partial \hat{x}} = 2 A^T P_{\epsilon} Ax - 2 A^T P_{\epsilon} y = 0. \quad (4.9) \]

This leads to the following system of normal equations,

\[ Nx = n \quad \text{with} \quad N = A^T P_{\epsilon} A x \quad \text{and} \quad n = A^T P_{\epsilon} y. \quad (4.10) \]

The solution of the normal equations yields the estimation of the unknown parameters by means of least squares adjustment,

\[ \hat{x} = (A^T P_{\epsilon} A)^{-1} A^T P_{\epsilon} y = N^{-1} n. \quad (4.11) \]

This solution corresponds to the best linear unbiased estimate and results in the maximization of the likelihood function (Koch 1997). The covariance matrix of the unknown parameters \( C(\hat{x}) \) can be derived by applying the law of covariance propagation to (4.11) under consideration of (4.6),

\[ C(\hat{x}) = \sigma^2 N^{-1}. \quad (4.12) \]

The unknown variance factor can be estimated by

\[ \hat{\sigma} = \frac{1}{n - u} (y - A\hat{x})^T P_{\epsilon} (y - A\hat{x}), \quad (4.13) \]

which leads to the estimated covariance matrix of the unknown parameters,

\[ \hat{C}(\hat{x}) = \hat{\sigma}^2 N^{-1}. \quad (4.14) \]

### 4.2 The Gravity Field and its Functionals in Terms of Splines

In this section, the basis functions, as introduced in Section 3.4.2, will be adopted according to the specific task of parameterization and determination of the gravity field with its different functionals. The mathematical considerations above have primarily been performed for functions given on the surface of the (unit) sphere. When dealing with satellite data, on the contrary, the observation points are located in the exterior of the Earth. Therefore, the outward continued basis functions (3.75) have to be applied, and a reference sphere \( \Omega_R \) with radius \( R \) approximating the semi-major axis of the Earth is introduced. The spline kernels are located at nodal points \( r_i \) on \( \Omega_R \) with

\[ |r_i| = R. \quad (4.15) \]

Arbitrary points in the exterior of this sphere, for example at the satellite’s positions, are denoted by \( r \) with

\[ |r| = r. \quad (4.16) \]

For simplification reasons, the spline kernel located at a specific nodal point \( r_i \) will be denoted as \( \Phi_i(r) \) in the following according to

\[ \Phi_i(r) := \Phi(r, r_i). \quad (4.17) \]
The spline kernel evaluated at the point \( r \) reads
\[
\Phi_i(r) = \frac{GM}{R} \sum_{n=2}^{\infty} k_n \left( \frac{R}{r} \right)^{n+1} P_n(t),
\]  
(4.18)
with
\[
t = \cos(\psi) = \frac{\mathbf{r} \cdot \mathbf{r}_i}{\mathbf{R}}.
\]  
(4.19)

The factor \( GM/R \) is introduced into the basis functions, so that the spline coefficients \( a_i \) in (3.74) become dimensionless values. In the following sections, the representation of the different gravity field functionals in terms of basis functions defined by (4.18) will be specified. These gravity field functionals are the gravitational potential, gravity, and the gravity gradient.

### 4.2.1 Gravitational Potential

The gravitational potential in terms of spline functions (4.18) can be represented as follows,
\[
V(r) = \sum_{i=1}^{I} a_i \Phi_i(r).
\]  
(4.20)

The potential as well as gravity and the gravity gradient are linear functionals of the unknown parameters \( a_i \), thus they can be phrased as a matrix-vector product. Therefore, the unknown parameters are arranged in the vector \( x \),
\[
x = (a_0, ..., a_I)^T,
\]  
(4.21)
and the gravitational potential at \( N \) positions of the satellite’s orbit can then be expressed by
\[
\begin{pmatrix}
V(r_1) \\
\vdots \\
V(r_N)
\end{pmatrix} = Vx,
\]  
(4.22)
with the matrix \( V \) consisting of the basis functions evaluated at the satellite’s positions \( r_k \) with \( k = 1, ..., N \),
\[
V = \begin{pmatrix}
\Phi_1(r_1) & \Phi_2(r_1) & \ldots & \Phi_I(r_1) \\
\Phi_1(r_2) & \Phi_2(r_2) & \ldots & \Phi_I(r_2) \\
\vdots & \vdots & \ddots & \vdots \\
\Phi_1(r_N) & \Phi_2(r_N) & \ldots & \Phi_I(r_N)
\end{pmatrix}_{N \times I}.
\]  
(4.23)

### 4.2.2 Gravity

Under consideration of Eq. (3.5), gravity at positions along the arcs of the satellite’s orbit can be calculated according to
\[
g(r) = \nabla V(r) = \sum_{i=1}^{I} a_i \nabla \Phi_i(r).
\]  
(4.24)

Again, this can be expressed by the corresponding matrix-vector product
\[
\begin{pmatrix}
g(r_1) \\
\vdots \\
g(r_N)
\end{pmatrix} = Gx,
\]  
(4.25)
4.2. The Gravity Field and its Functionals in Terms of Splines

where the elements of the matrix $\tilde{G}$ contain the gradients of the basis functions,

$$
\tilde{G} = \begin{pmatrix}
\nabla \Phi_1(r_1) & \nabla \Phi_2(r_1) & \ldots & \nabla \Phi_I(r_1) \\
\nabla \Phi_1(r_2) & \nabla \Phi_2(r_2) & \ldots & \nabla \Phi_I(r_2) \\
\ldots & \ldots & \ldots & \ldots \\
\nabla \Phi_1(r_N) & \nabla \Phi_2(r_N) & \ldots & \nabla \Phi_I(r_N)
\end{pmatrix}_{3N \times I}.
$$

(4.26)

Usually, the gravity vector is obtained by differentiating the gravitational potential with respect to the coordinates of a local north-oriented frame and a subsequent rotation into a cartesian Earth-fixed reference frame. Here a different method is applied, as the derivatives of the basis functions are directly calculated with respect to the cartesian coordinates of an Earth-fixed frame by applying the chain rule,

$$
\nabla \Phi = \begin{pmatrix}
\frac{\partial \Phi}{\partial x} \\
\frac{\partial \Phi}{\partial y} \\
\frac{\partial \Phi}{\partial z}
\end{pmatrix} = \begin{pmatrix}
\frac{\partial \Phi}{\partial r} \cdot \frac{\partial r}{\partial x} + \frac{\partial \Phi}{\partial t} \cdot \frac{\partial t}{\partial x} \\
\frac{\partial \Phi}{\partial r} \cdot \frac{\partial r}{\partial y} + \frac{\partial \Phi}{\partial t} \cdot \frac{\partial t}{\partial y} \\
\frac{\partial \Phi}{\partial r} \cdot \frac{\partial r}{\partial z} + \frac{\partial \Phi}{\partial t} \cdot \frac{\partial t}{\partial z}
\end{pmatrix},
$$

(4.27)

with the partial derivatives of the basis functions with respect to $r$ and $t$,

$$
\frac{\partial \Phi}{\partial r} = \sum_{n=2}^{\infty} \frac{-(n+1)}{R} k_n \left(\frac{R}{r}\right)^{n+2} P_n(t),
$$

(4.28)

$$
\frac{\partial \Phi}{\partial t} = \sum_{n=2}^{\infty} k_n \left(\frac{R}{r}\right)^{n+2} \frac{dP_n(t)}{dt},
$$

(4.29)

and with the partial derivatives of $r$ and $t$ with respect to the cartesian coordinates,

$$
\frac{\partial r}{\partial \alpha} = \frac{\alpha}{r} \quad \text{and} \quad \frac{\partial t}{\partial \alpha} = \frac{\alpha t}{r^2} \quad \text{with} \quad \alpha = x, y, z.
$$

(4.30)

Here $\alpha$ represents the coordinates of $r$ and $\alpha_i$ the coordinates of $r_i$. The positions used for the calculation of the matrix in Eq. (4.26) have to be given in an Earth-fixed co-rotating coordinate system, thus gravity in Eq. (4.25) is given in this Earth-fixed frame as well. In contrast to this, the functionals of the satellite movement that will serve as functional models in the gravity field determination process, as will be described in Section 4.3, refer to the inertial reference frame. Therefore, they require gravity to be formulated in this frame. The corresponding transformation can be expressed as

$$
G := RG,
$$

(4.31)

with $G$ denoting the matrix related to the inertial reference frame. Using the rotation matrices of the Earth rotation, as given in the IERS Conventions (McCarthy and Petit 2004) of the International Earth Rotation and Reference System Service (IERS), the relationship between the International Celestial Reference Frame (ICRF) as a realization of the quasi-inertial system and the International Terrestrial Reference Frame (ITRF) as realization of an Earth fixed reference system can be established. The rotations from the ITRF to the ICRF for each observation time can be given by the matrices $R(t_k)$ arranged in the blockdiagonal matrix

$$
R = \begin{pmatrix}
R(t_1) \\
\vdots \\
R(t_N)
\end{pmatrix}.
$$

(4.32)

Gravity in the ICRF can be obtained by inserting $G$ instead of $\tilde{G}$ in Eq. (4.25).

4.2.3 Gravity Gradient

In case of satellite gravity gradiometry, the gravity gradient components serve as observations. Therefore, the functional model for this type of observations is required as well. The gravity gradient is calculated as second derivatives of the gravitational potential, expressed in terms of spline functions according to

$$
\nabla^2 V(r) = \sum_{i=1}^{I} a_i \nabla^2 \Phi_i(r).
$$

(4.33)
The corresponding matrix-vector product can be formulated as follows,

$$
\left( \nabla \nabla V(r_1) \right) = T x,
$$

(4.34)

with the block matrix $T$ containing the second derivatives of the basis functions,

$$
T = \left( \begin{array}{ccc}
\nabla \nabla \Phi_1(r_1) & \nabla \nabla \Phi_2(r_1) & \cdots & \nabla \nabla \Phi_I(r_1) \\
\nabla \nabla \Phi_1(r_2) & \nabla \nabla \Phi_2(r_2) & \cdots & \nabla \nabla \Phi_I(r_2) \\
\vdots & \vdots & \ddots & \vdots \\
\nabla \nabla \Phi_1(r_N) & \nabla \nabla \Phi_2(r_N) & \cdots & \nabla \nabla \Phi_I(r_N)
\end{array} \right).
$$

(4.35)

Here $\nabla \nabla \Phi$ denotes the Hessian matrix of the second derivatives of a basis function,

$$
\nabla \nabla \Phi = \left\{ \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} \right\}_{\alpha \beta}
$$

with $\alpha, \beta = x, y, z$.

(4.36)

Depending on the specific application, these second derivatives can be arranged either in a $3 \times 3$ tensor or in a column vector, as will be shown below. Again the chain rule is applied in order to directly calculate the derivatives in the cartesian Earth-fixed frame,

$$
\Phi_{\alpha \beta} = \Phi_{\alpha \alpha} = \frac{\partial \Phi}{\partial r} \cdot \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} + \frac{\partial \Phi}{\partial t} \cdot \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} + \frac{\partial^2 \Phi}{\partial r^2} \cdot \frac{\partial r}{\partial \alpha} \cdot \frac{\partial \beta}{\partial \beta}
$$

\[+ \frac{\partial^2 \Phi}{\partial r \partial t} \cdot \frac{\partial r}{\partial \alpha} \cdot \frac{\partial \beta}{\partial t} + \frac{\partial^2 \Phi}{\partial \beta \partial t} \cdot \frac{\partial \beta}{\partial \alpha} \cdot \frac{\partial \beta}{\partial t} \]

(4.37)

with the second derivatives of the basis functions with respect to $r$ and $t$,

$$
\frac{\partial^2 \Phi}{\partial r^2} = \sum_{n=2}^{\infty} \frac{(n+2)(n+1)}{R^2} k_n \left( \frac{R}{r} \right)^{n+3} P_n(t),
$$

(4.38)

$$
\frac{\partial^2 \Phi}{\partial r \partial t} = \sum_{n=2}^{\infty} \frac{(n+1)}{R} k_n \left( \frac{R}{r} \right)^{n+2} \frac{dP_n(t)}{dt},
$$

(4.39)

$$
\frac{\partial^2 \Phi}{\partial t^2} = \sum_{n=2}^{\infty} k_n \left( \frac{R}{r} \right)^{n+2} \frac{d^2 P_n(t)}{dt^2}.
$$

(4.40)

The second derivatives of $r$ and $t$ with respect to the cartesian coordinates can be formulated as follows,

$$
\frac{\partial^2}{\partial \alpha \partial \beta} = \frac{1}{r} \delta_{\alpha \beta} - \frac{\alpha \beta}{r^3} \quad \text{and} \quad \frac{\partial^2}{\partial \alpha \partial \beta} = -\frac{t}{r^2} \delta_{\alpha \beta} - \frac{\alpha \beta + \alpha \beta t}{r^3} + \frac{3 \alpha \beta t}{r^2},
$$

(4.41)

with $\alpha, \beta = x, y, z$.

Again $\alpha$ and $\beta$ are the coordinates of $r$ and $\alpha_i$ and $\beta_i$ the coordinates of $r_i$. The observations of the second derivatives are given in the reference frame of the gradiometer, the so-called ‘gradiometer reference frame’ (GRF), oriented along the three axes of the gradiometer instrument with origin in the nominal intersection of the three one-axis gradiometers. In contrast to that, the derivatives in the formulas given above are formulated in an Earth-fixed reference frame (ITRF) with origin located in the geocenter, $z$-axis directed to the pole, $x$-axis fixed in the equatorial plane in the direction of the Greenwich meridian, and $y$-axis as completion to a right-handed system. Therefore, the observation equation has to be transformed from the ITRF into the GRF. When the second derivatives for each observation point $r_k$ are arranged according to

$$
T(r_k) = \nabla \nabla \Phi = \begin{pmatrix}
\Phi_{xx} & \Phi_{xy} & \Phi_{xz} \\
\Phi_{yx} & \Phi_{yy} & \Phi_{yz} \\
\Phi_{zx} & \Phi_{zy} & \Phi_{zz}
\end{pmatrix},
$$

(4.42)
this transformation can be performed by using the transformation formula for tensors,
\[ T(r_k) = D(r_k)\overline{T}(r_k)D(r_k)^T. \] (4.43)

\( T \) is the tensor of second derivatives in the Earth-fixed frame and \( \overline{T} \) the tensor referring to the gradiometer reference frame. The matrix \( D \) is the \( 3 \times 3 \) rotation matrix transforming the former frame into the latter. Further information about the transformation of gravity gradients from one coordinate system into another can be found in Koop (1993). For practical calculations, it can often be reasonable to combine the independent tensor components in a column vector according to
\[ \mathbf{t}(r_k) = (\Phi_{xx}, \Phi_{xy}, \Phi_{xz}, \Phi_{yy}, \Phi_{yz}, \Phi_{zz})^T. \] (4.44)

Then the transformation described by Eq. (4.43) can be reformulated as a single matrix-vector operation, as, for example, described by Ditmar et al. (2003),
\[ \mathbf{t}(r_k) = R(r_k)\mathbf{t}(r_k), \] (4.45)

with \( \mathbf{t} \) denoting the rotated values. The matrix \( R(r_k) \) relates the vector \( \mathbf{t}(r_k) \) to the vector \( \mathbf{t}(r_k) \); it emerges from carrying out the two rotations in Eq. (4.43). When the transformation matrices for the different observation points are combined into one blockdiagonal matrix according to
\[ R = \begin{pmatrix} R(r_1) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & R(r_N) \end{pmatrix}, \] (4.46)

then the transformation can be applied to the matrix \( \overline{T} \) in Eq. (4.35) as follows,
\[ \mathbf{T} = RT. \] (4.47)

The gravity gradient related to the GRF can be obtained by inserting \( \mathbf{T} \) instead of \( \overline{T} \) in Eq. (4.34).

### 4.3 Functional Model

In the following section, the functional models providing the relationship between the unknown gravity field parameters and the different types of observations provided by the satellite missions CHAMP, GRACE, and GOCE will be derived. These observation principles are the concept of precise orbit determination (POD) derived from satellite-to-satellite tracking in the high-low mode, satellite-to-satellite tracking in the low-low mode, and satellite gravity gradiometry. The functional models can be applied to different gravity field representations, e.g. spherical harmonic expansions or radial basis functions. The representation of the different gravity field functionals in terms of radial basis functions, as described in Section 4.2, leads to the specific observation equations for the regional gravity field recovery. To exploit all the advantages of a regional gravity field recovery, only the data being associated with the respective regional area has to be used. Therefore, the observation equations are established for satellite data over the selected regional recovery area, while the coverage with satellite data should be slightly larger than the recovery region itself to prevent the solution from geographical truncation effects. This aspect has to be considered with care especially for the long wavelength gravity field features. Thus the orbit information and additional satellite data have to be cut out over the regional area. Consequently, the employed observation models necessarily have to be based on short arcs of the satellite’s orbit. The use of short arcs presents further advantages. Unmodeled disturbing forces acting on the satellite do not accumulate and, therefore, only have a limited impact on the solution. A further benefit of short arcs lies in the fact that one can deal with data gaps more easily, as a new arc can be started after each data gap. In this way, the observations along an arc can be regarded to have equal distances in time without gaps. Therefore, there is no necessity to consider respective exceptions during the data processing.
4.3.1 Precise Orbit Determination

All three satellite missions CHAMP, GRACE, and GOCE carry a receiver for the Global Positioning System (GPS) in order to acquire satellite-to-satellite tracking data in the high-low configuration. From this data, the satellites’ orbits can be calculated by means of precise orbit determination (POD), see, for example, Svehla and Rothacher (2001). In case of CHAMP, this POD information provides the primary type of observations, but it establishes the basis for the functional model of the mission GRACE as well and is also used in the GOCE processing. Different approaches to process POD data have been proposed, among them are the acceleration approach, as developed by Austen and Reubelt (2000), the acceleration approach using double differences, as introduced by Ditmar and van Eck van der Sluijs (2004), and the energy integral approach as a very popular tool in satellite geodesy. In the context of determining a gravity field solution from CHAMP data, it was, for example, used by Gerlach et al. (2003). An overview of the different methods is given in Ilk et al. (2005). In the following, only the approach actually used for the calculation procedure in this thesis will be described. It is based on an integral equation approach, first proposed by Schneider (1968) in the context of orbit determination and then applied to the gravity field determination by Reigber (1969). The following procedure was applied successfully to the calculation of a global CHAMP gravity field model in the space domain by Mayer-Gürr et al. (2005). More details of the method can be found in Mayer-Gürr (2006).

The relation between the satellite’s movement along the orbit and the forces f acting on the satellite is described by Newton’s equation of motion, here formulated with respect to a unity mass element,

\[ \ddot{r}(t) = f(t, r, \dot{r}), \]  

(4.48)

with \( f(t) \) denoting the satellite’s acceleration for a given time \( t \) and \( r \) and \( \dot{r} \) describing the satellite’s position and velocity, respectively. Typically, this differential equation is non-linear, as the force function does not depend linearly on the satellite’s positions. Integrating the differential equation twice results in an integral equation for the satellite’s orbit. With the boundary values

\[ r_A := r(t_A), \quad r_B := r(t_B), \quad t_A < t_B, \]  

(4.49)

for the start and end position of the arc of the satellite’s orbit. This integral equation can be formulated as the solution of a boundary value problem according to

\[ r(\tau) = (1 - \tau)r_A + \tau r_B - T^2 \int_{\tau'}=0^{1} K(\tau, \tau') f(\tau', r, \dot{r}) d\tau', \]  

(4.50)

with the normalized time

\[ \tau = \frac{t - t_A}{T} \quad \text{with} \quad T = t_B - t_A \]  

(4.51)

and the integral kernel

\[ K(\tau, \tau') = \begin{cases} \tau (1 - \tau'), & \tau \leq \tau', \\ \tau' (1 - \tau), & \tau' \leq \tau. \end{cases} \]  

(4.52)

Eq. (4.50) is a Fredholm type integral equation of the second kind. The unknowns in this equation are the boundary values \( r_A \) and \( r_B \) and the specific force function \( f(\tau', r, \dot{r}) \) along the orbit. In the following, the positions \( r \) in the force function will be substituted by the positions \( r \), measured by GPS. In case of POD measurements, they can be assumed as sufficiently accurate, as has been investigated in Mayer-Gürr (2006). Therefore, the dependence of the force function on the unknown satellite orbit can be omitted. When the same integral equation is used to derive observation equations for the highly accurate satellite-to-satellite tracking measurements, this dependence has to be considered, as will be described in Section 4.3.1.1. The satellite’s velocity \( \dot{r} \) only influences the force function due to the surface forces acting on the satellite. Those surface forces can directly be determined by the onboard accelerometer, therefore the dependence of the
force function on the satellite’s velocity can be neglected as well. This leads to the POD observations to the
simplification \( f(\tau', \mathbf{r}, \mathbf{r}) = f(\tau') \).

As observations serve the satellite positions \( \mathbf{r}(\tau_k) \) which are determined at \( N \) discrete, equidistant normed
points in time \( \tau_k \) calculated by

\[
\tau_k = \frac{k - 1}{N - 1} \quad \text{for } k = 1, \ldots, N.
\]

(4.53)

As equation (4.50) is evaluated for each observed position \( \mathbf{r}(\tau_k) \) along the arc of the satellite’s orbit, one
obtains a system of equations which can be formulated in matrix notation as follows,

\[
\mathbf{l} = \mathbf{Bb} + \mathbf{h}.
\]

(4.54)

The vector \( \mathbf{l} \) represents the observed positions of the satellite,

\[
\mathbf{l} = \begin{pmatrix} \mathbf{r}(\tau_1) \\ \vdots \\ \mathbf{r}(\tau_N) \end{pmatrix}.
\]

(4.55)

The vector \( \mathbf{b} \) contains the boundary values and the matrix \( \mathbf{B} \) the corresponding normalized observation
times,

\[
\mathbf{B} = \begin{pmatrix} (1 - \tau_1) & \tau_1 \\ (1 - \tau_2) & \tau_2 \\ \vdots & \vdots \\ (1 - \tau_N) & \tau_N \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{r}_A \\ \mathbf{r}_B \end{pmatrix}.
\]

(4.56)

The vector \( \mathbf{h} \) stands for the integral evaluated at the \( N \) observation epochs,

\[
\mathbf{h} = \begin{pmatrix} h(\tau_1) \\ \vdots \\ h(\tau_N) \end{pmatrix} \quad \text{with} \quad h(\tau_k) = -T^2 \int_0^1 K(\tau_k, \tau') f(\tau') d\tau'.
\]

(4.57)

In order to establish a relation between the observations and the unknown parameters \( \mathbf{x} \), the integral is
expressed as a linearized functional of these parameters,

\[
\mathbf{h} = \mathbf{h}_0 + \mathbf{A} \mathbf{x} \quad \text{with} \quad (\mathbf{A})_{ki} = \frac{\partial h(\tau_k)}{\partial x_i}.
\]

(4.58)

This results in the following linear system of observation equations,

\[
\mathbf{l} - \mathbf{h}_0 = \mathbf{Bb} + \mathbf{Ax},
\]

(4.59)

with the vector of unknown parameters according to Eq. (4.21). The elements of matrix \( \mathbf{A} \) can be derived
by applying the chain rule, i.e. by partially differentiating the integral with respect to the force function and then
differentiating \( f \) with respect to the unknown parameters,

\[
(\mathbf{A})_{ki} = \frac{\partial h(\tau_k)}{\partial x_i} = \frac{\partial h(\tau_k)}{\partial f(\tau')} \frac{\partial f(\tau')}{\partial x_i}.
\]

(4.60)

It reads expressed in matrix notation

\[
\mathbf{A} = \mathbf{KG} \quad \text{with} \quad (\mathbf{K})_{kj} = \frac{\partial h(\tau_k)}{\partial f_j} \quad \text{and} \quad (\mathbf{G})_{ji} = \frac{\partial f_j}{\partial x_i}.
\]

(4.61)
The setup of matrix $G$ has already been described in Section 4.2.2; it is given according to Eq. (4.26) under consideration of Eq. (4.31). The specific forces along the satellite’s orbit can then be expressed in terms of the unknown gravity field functionals according to

$$
\begin{pmatrix}
  f(\tau_1) \\
  \vdots \\
  f(\tau_N)
\end{pmatrix} = Gx + a,
$$

(4.62)

with the vector $a$ representing the reference gravity field, the tide models, and the disturbing surface forces as determined by the onboard accelerometer. The observation vector is reduced by these forces which implies that the accelerometer measurements and force models are assumed to be error-free. The matrix $K$ represents the numerical integration of the integral (4.57). By means of quadrature methods, a linear relationship between $h$ and $f$ can be established,

$$
\begin{pmatrix}
  h(\tau_1) \\
  \vdots \\
  h(\tau_N)
\end{pmatrix} = K \begin{pmatrix}
  f(\tau_1) \\
  \vdots \\
  f(\tau_N)
\end{pmatrix}.
$$

(4.63)

For more details concerning the calculation of $K$, refer to Mayer-Gürr (2006). Using this matrix $K$, the approximate values $h_0$ for the integral can be calculated from the reference accelerations according to

$$
h_0 = Ka.
$$

(4.64)

This leads to the following Gauss-Markoff model,

$$
1 - Ka = KGx + Bb + \epsilon,
$$

(4.65)

where $b$ and $x$ are the vectors of unknown parameters, and $\epsilon$ denotes the noise.

### 4.3.1.1 Refinement of the Functional Model

The functional model described so far can be used in the processing of POD measurements. The observation equations can be modified and subsequently applied to establish the linearized model for intersatellite K-band measurements as in case of the GRACE mission as well. The superior accuracy of the K-band observations results in high demands on the accuracy of the functional model. The intersatellite ranges can be observed with an accuracy of a few $\mu$m, thus the errors occurring from the approximations made in the model itself have to be significantly smaller than these measurement errors. Those approximation errors can originate from the linearization procedure, from the numerical integration, and from values introduced into the functional model. The first one can be considered by good approximate values or by an iteration of the estimation procedure. The numerical integration is non-critical as well, as shown by Mayer-Gürr (2006). But the accuracy of the satellite’s positions that are introduced into the functional model poses a more serious problem. In the integral (4.57), the force function $f$ is evaluated at the specific positions of the satellite at specific observation epochs. This results in the force function to depend not only on time, but on the position as well: $f(r(\tau))$. In case of the POD observations, the inaccuracies of the GPS positions were neglected, but this is not possible in case of K-band measurements. The positions can be determined with an accuracy of a few cm, which is not sufficient compared to the high precision measurements provided by the K-band instrument. Therefore, Mayer-Gürr (2006) proposed the use of refined observation equations. This refinement procedure is shortly described in the following.

The evaluation of the integral at noisy positions $r_\epsilon$ results in positions $\hat{r}$,

$$
\hat{r}(\tau) = (1 - \tau)r_A + \tau r_B + T^2 \int_0^1 K(\tau, \tau')f(r_\epsilon(\tau')) d\tau'.
$$

(4.66)
The difference compared to the true positions \( \mathbf{r} \) can be calculated according to
\[
\mathbf{r}(\tau) - \hat{\mathbf{r}}(\tau) = T^2 \int_0^1 K(\tau, \tau') [f(\mathbf{r}(\tau')) - f(\mathbf{r}_e(\tau'))] \, d\tau'.
\] (4.67)

The above equation can be given in a simplified version,
\[
\mathbf{r} - \hat{\mathbf{r}} = K [f(\mathbf{r}) - f(\mathbf{r}_e)],
\] (4.68)
with the linear integral operator
\[
K = T^2 \int_0^1 K(\tau, \tau') \, d\tau'.
\]

Linearizing the force function,
\[
f(\mathbf{r}) = f(\mathbf{r}_e) + \nabla f \big|_{\mathbf{r}_e} \cdot (\mathbf{r} - \mathbf{r}_e) + \ldots,
\] (4.69)
yields
\[
\mathbf{r} - \hat{\mathbf{r}} = K \nabla f \cdot (\mathbf{r} - \mathbf{r}_e).
\] (4.70)

The insertion of Eq. (4.70) instead of \( \hat{\mathbf{r}} \) into Eq. (4.66) and some rearrangements lead to
\[
[I - K \nabla f(\mathbf{r}_e)] (\mathbf{r} - \mathbf{r}_e) = K f(\mathbf{r}_e) + \mathbf{b} - \mathbf{r}_e,
\] (4.71)
with \( I \) denoting the unity operator and \( \mathbf{b}(\tau) = (1 - \tau) \mathbf{r}_A + \tau \mathbf{r}_B \) being the functional of the line of sight connection. As long as the inverse of the operator \([I - K \nabla f(\mathbf{r}_e)]\) exists, the noise-free positions can be calculated during the linearization process of the force function,
\[
\Delta \mathbf{r} = \mathbf{r} - \mathbf{r}_e = [I - K \nabla f(\mathbf{r}_e)]^{-1} [K f(\mathbf{r}_e) + \mathbf{b} - \mathbf{r}_e].
\] (4.72)

This equation can be discretized analogously to the procedure used with the POD observations,
\[
\Delta \mathbf{r} = (I - \mathbf{K} \mathbf{T})^{-1} (\mathbf{K} f + \mathbf{B} \mathbf{b} - \mathbf{r}_e),
\] (4.73)
with \( \mathbf{K} \) being the matrix of the numerical quadrature and \( \mathbf{B} \) being the matrix containing the functional of the boundary values as in (4.56). \( \mathbf{T} \) is the matrix of the gravity gradients according to
\[
\mathbf{T} = \begin{pmatrix}
\nabla f(\tau_1) & 0 \\
\vdots & \ddots \\
0 & \nabla f(\tau_N)
\end{pmatrix}.
\] (4.74)

Eq. (4.73) can now be used to calculate error-free positions,
\[
\mathbf{r} = \mathbf{r}_e + \Delta \mathbf{r}.
\] (4.75)

### 4.3.2 Low-Low Satellite-to-Satellite Tracking

For the processing of satellite-to-satellite tracking data in the low-low configuration, as provided by the K-band measurement instrument onboard the GRACE mission, different approaches have been proposed as well. Among them are the employment of the energy integral, as proposed by 

\[ Jekeli \ (1999), \] the acceleration approach, as, for example, used by 

\[ Sharifi \ and \ Keller \ (2005), \] and the Hammerstein-Schneider method, see, for example, 

\[ Ilk \ (1984). \] The approach that is used in the calculations presented here is based on the integral equation described for the POD observations in Eq. (4.50). This method is explained in more detail in 

\[ Mayer-Gürr \ et al. \ (2006); \] in the following it will shortly be reviewed.

If precise intersatellite functionals as line-of-sight ranges or range-rate measurements are available as in case of the GRACE mission, the mathematical model can be derived by projecting the relative vector between the two satellites onto the line-of-sight connection,
\[
\rho(\tau) = \mathbf{e}_{12}(\tau) \cdot (\mathbf{r}_2(\tau) - \mathbf{r}_1(\tau)),
\] (4.76)
with $\rho$ denoting the range between the two satellites. The vectors $r_2(\tau)$ and $r_1(\tau)$ describe the positions of the two GRACE satellites, and $e_{12}$ is the unit vector in line-of-sight direction,

$$e_{12}(\tau) = \frac{\mathbf{r}_{12}(\tau)}{\|\mathbf{r}_{12}(\tau)\|} \quad \text{with} \quad \mathbf{r}_{12}(\tau) = \mathbf{r}_2(\tau) - \mathbf{r}_1(\tau). \quad (4.77)$$

Analogous formulae can be derived for range-rate and range-acceleration measurements, as described in Mayer-Gürr (2006). Eq. (4.76) is a non-linear functional of the unknown parameters $x$. In order to use it in a Gauss-Markoff model, it has to be linearized,

$$\rho = \rho_0 + \frac{\partial \rho}{\partial x} \bigg|_{x_0} \cdot \Delta x + ... \quad (4.78)$$

The relation between the range measurements and the unknown parameters is not stated explicitly in Eq. (4.76), but via the satellites’ positions along the orbits. Thus the partial derivatives of the range measurements are derived by applying the chain rule and first differentiating with respect to the positions of the two satellites. Then the positions are differentiated with respect to the force function $f$ and finally the force function with respect to unknown parameters,

$$\frac{\partial \rho}{\partial x} = \frac{\partial \rho}{\partial \mathbf{r}_1} \cdot \frac{\partial \mathbf{r}_1}{\partial f} \cdot \frac{\partial f}{\partial x} + \frac{\partial \rho}{\partial \mathbf{r}_2} \cdot \frac{\partial \mathbf{r}_2}{\partial f} \cdot \frac{\partial f}{\partial x}. \quad (4.79)$$

The first term of each product is the differentiation of Eq. (4.76),

$$\frac{\partial \rho}{\partial \mathbf{r}_1} = -e_{12}, \quad \frac{\partial \rho}{\partial \mathbf{r}_2} = e_{12}. \quad (4.80)$$

The second part of the partial derivatives, the linearized relationship between the positions and the unknown parameters $\partial \mathbf{r}_i/\partial x = \partial \mathbf{r}_i/\partial f \cdot \partial f/\partial x$, corresponds to the observation equations for the POD case, as given by Eq. (4.65). The observation equations for the range measurements are then obtained from the ones for the POD observations by multiplication with the derivatives in Eq. (4.80). This equals a projection onto the line of sight connection. It has to be pointed out that, in case of K-band observations, the refined satellite’s positions given by Eq. (4.75) have to be introduced into the model, as the accuracy of the original GPS positions is not sufficient. The non-linear relation (4.76) requires a reference gravity field model to be introduced as approximation for the unknown parameters. Furthermore, the range observations have to be reduced by the influence of the other modelled forces, such as surface and tidal forces. The reduced observations are introduced as pseudo-observations on the left side of Eq. (4.76). For these pseudo-observations a covariance matrix can be determined, including the noise model of the satellite ranging system and of the accelerometer observations, see Mayer-Gürr (2006).

### 4.3.3 Satellite Gravity Gradiometry (GOCE)

The gradiometer instrument, as in case of the GOCE mission, determines the gravity gradient consisting of the second derivatives of the gravitational potential. The observation equation for these types of measurements is given by

$$\nabla \nabla V(\mathbf{r}) = \sum_{i=1}^{l} a_k \nabla \nabla \Phi_i(\mathbf{r}). \quad (4.81)$$

The design matrix $A$ corresponds to the matrix $T$ already described by Eq. (4.35),

$$A = T. \quad (4.82)$$

In this way, the matrix $A$ contains the observation equations that have been rotated into the gradiometer reference frame. It should be pointed out that in practical observations not all of the nine tensor components will be measured with equal accuracy. On the contrary, the three components on the main diagonal of the gravity gradient tensor will be determined with superior accuracy. Therefore, these three components will serve as primary observations.
5. Solving the System of Observation Equations

The following chapter deals with the solution of the observation equations that are set up in Chapter 4.3. The solution is performed in the sense of a least squares adjustment, as introduced in Section 4.1. Particular interest is dedicated to the ill-posedness of the downward continuation process that is inevitable when dealing with data in satellite altitude and aiming at the determination of the gravity field on the Earth's surface, especially in case of space-localizing basis functions. Therefore, in the first part of this chapter, the concept and the implications of ill-posed problems are reviewed. To overcome the issues imposed by the downward continuation, the problem can be stabilized by the introduction of prior information. This corresponds to a regularization according to Tikhonov. Thus in the second part of this chapter, the regularization process is described for the applied case of a Tikhonov regularization. Subsequently, the implications of this kind of regularization are specified for the case of a parameterization by the space localizing basis functions used within this thesis. The regularization parameter is determined by a variance component estimation procedure, as described afterwards. Finally, the concept of a regionally adapted regularization will be introduced, as this plays a major role for the benefit of the regional gravity field recovery. In this sense, the regularization procedure presents an important module for the regional approach presented in this thesis.

5.1 Inverse Problems, Ill-posed Problems

When dealing with geodetic problems, especially in satellite geodesy, the concepts of inverse and ill-posed problems is of particular importance. The issue has been discussed widely within the geodetic community, see, for example, Moritz (1966), Schwarz (1971), Rummel et al. (1979), or Ilk (1984). A short introduction will be given in this section. An overview concerning this topic from the mathematical point of view can be found in Louis (1989) or Hansen (1997).

Let $X$ and $Y$ be normed vector spaces, and let $A : X \rightarrow Y$ be the physical model describing the linear mapping between the two spaces according to

$$ Ax = y, \quad (5.1) $$

with $x \in X$ and $y \in Y$. The task of calculating the output $y$ from a given set of $x$ is described as direct problem, whereas the calculation of the model parameters indirectly from a given set of observations is specified as inverse problem. An inverse problem is called 'well posed' according to Hadamard (1923), as long as $A$ is bijective, and the inverse operator $A^{-1}$ is continuous. $A$ being bijective means that the equation is solvable for all $y \in Y$ (surjective), and that it has a unique solution (injective). Requiring the inverse operator to be continuous implicates that the solution $x$ depends continuously on the data (with the concept of continuity introduced in Appendix A.2). This implies that small changes in $x$ result in small changes in $y$. In this case, the problem can be considered as being stable. In the context of a linear operator, the requirement of continuity is equivalent to assuming that the operator is bounded. As soon as any one of these requirements (existence, uniqueness, and stability) is violated, the problem is said to be an 'improperly posed' or 'ill-posed' problem. In geodesy one frequently has to deal with inverse problems, as the model parameters to be determined are usually not subject to direct observations. Inverse problems are frequently ill-posed, as very often at least one of the conditions mentioned above is not satisfied. An example of the non-uniqueness of the solution is the problem of determining the mass distribution in the interior of the Earth from the exterior gravitational potential, as infinitely many mass distributions can result in the same exterior gravity field. The reason for ill-posedness that will be further investigated in this thesis is the lack of stability. In case of satellite geodesy, the missing stability can have several causes, the most important of them being listed below:
5. Solving the System of Observation Equations

The calculation of gravity field functionals on the Earth’s surface from measurements in satellite altitude causes an ill-posed problem, as during the downward continuation the signal is amplified especially in the high frequency part of the spectrum. This matter will be described in more detail in Section 5.1.1.

Irregular data distribution or data gaps (such as the polar gap problem in case of the GOCE mission) lead to instabilities.

A measurement instrument might not be able to recover the whole spectral domain. (For example, the GOCE gradiometer is not sensitive with respect to long wavelength features of the gravity field.)

5.1.1 Ill-posedness of the Downward Continuation Process

In this thesis, the focus in the context of ill-posed problems is on the downward continuation process in case of gravity field recovery from satellite data. Therefore, this problem shall be dealt with in more detail. The instability of this process can be observed by analyzing the compact upward continuation operator, as defined by Eq. (3.70).

\[ U(x, x_r) = \sum_{n=0}^{\infty} \frac{1}{R} \left( \frac{R}{r} \right)^{n+1} Y_{nm} \left( \frac{x_r}{|x_r|} \right) Y_{nm}(x). \] (5.2)

This operator can be decomposed into its singular values, as described in Section A.2.0.1, with the orthogonal singular functions \( Y_{nm} \) and the singular values \( \sigma_{nm} = \frac{1}{R} \left( \frac{R}{r} \right)^{n+1} \). (5.3)

Details concerning these spectral relationships can be found in Meissl (1971) or Rummel and van Gelderen (1995). The decay of the singular values towards zero with increasing frequency \( n \) becomes evident in Eq. (5.3). Thus high frequencies are associated with small singular values, which indicates the smoothing property of the kernel. This is a very important characteristic of compact operators, as defined by Eq. (A.37). A consequence of this characteristic is the fact that the inverse of a compact linear operator cannot be bounded. The inverse of the upward continuation operator is the downward continuation operator with its singular values approaching infinity with increasing \( n \). This illustrates the amplification of small frequencies during the downward continuation process. As boundedness and continuity are equivalent concepts when dealing with linear operators, the downward continuation operator cannot be continuous. Thus it violates Hadamard’s continuity assumption for well-posed problems.

5.1.2 Singular Value Decomposition of the Design Matrix

The instability of a problem given by Eq. (4.6) can be identified by an examination of the spectral behavior of the design matrix \( A \). Therefore, the theory of spectral decomposition (of finite dimensional matrices) will shortly be reviewed in the following, as described, for example, by Hansen (1997). For the case of infinite dimensional operators, the singular value decomposition is described in Appendix A.2.0.1. The concepts introduced there for infinite problems can help to understand the finite dimensional problem, as the design matrix \( A \) represents a discretization of the infinite dimensional operator.

The rectangular matrix \( A \in \mathbb{R}_{n \times u} \) with rank \( n \geq u \) can be decomposed into

\[ A = U \Sigma V^T = \sum_{i=1}^{u} u_i \sigma_i v_i^T, \quad \Sigma = \begin{pmatrix} \Sigma_u \\ 0 \end{pmatrix}, \] (5.4)

with the orthogonal matrices \( U = (u_1, \ldots, u_n) \in \mathbb{R}_{n \times n} \) and \( V = (v_1, \ldots, v_u) \in \mathbb{R}_{u \times u} \). The matrix \( \Sigma \in \mathbb{R}_{n \times u} \) contains the diagonal matrix \( \Sigma_u = \text{diag}(\sigma_1, \ldots, \sigma_u) \) with the singular values \( \sigma_i \) that can be arranged in descending order according to

\[ \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_u > 0. \] (5.5)
The vectors \( \mathbf{v}_i \) and \( \mathbf{u}_i \) are the corresponding left and right singular vectors, respectively, satisfying the following relations,

\[
\begin{align*}
\mathbf{A} \mathbf{v}_i &= \sigma_i \mathbf{u}_i, \\
\mathbf{A}^T \mathbf{u}_i &= \sigma_i \mathbf{v}_i.
\end{align*}
\] (5.6)

The singular values of the design matrix can be related to the eigenvalues of the normal equation matrix, as the \( \mathbf{v}_i \) are the eigenvectors of \( \mathbf{A}^T \mathbf{A} \) and the \( \mathbf{u}_i \) the eigenvectors of \( \mathbf{A} \mathbf{A}^T \) with the same eigenvalues \( \lambda_i \) for both matrices. The singular values are then given by \( \sigma_i = \sqrt{\lambda_i} \). As described in Section 5.1.1, singular vectors corresponding to small singular values can be associated with high frequencies. Using this characteristic, the following relations,

\[
\begin{align*}
\mathbf{Ax} &= \sum_{i=1}^{n} \sigma_i (x^T \mathbf{v}_i) \mathbf{u}_i.
\end{align*}
\] (5.7)

From the above Eq. (5.7), it becomes obvious that by the mapping these high frequencies of \( \mathbf{x} \) are damped more strongly than the lower frequencies due to the multiplication with smaller singular values \( \sigma_i \). The opposite is the case as to the solution of the inverse problem,

\[
\hat{\mathbf{x}} = \sum_{i=1}^{n} \frac{1}{\sigma_i} (y^T \mathbf{u}_i) \mathbf{v}_i.
\] (5.8)

Eq. (5.8) can be considered as spectral decomposition of \( \hat{\mathbf{x}} \) with the coefficients \( \frac{1}{\sigma_i} (y^T \mathbf{u}_i) \) indicating the spectral properties of \( \hat{\mathbf{x}} \). The amplification of the smaller frequencies becomes evident. As formulated by Eq. (A.40), the observations must not be arbitrarily rough, as the coefficients \( (y^T \mathbf{u}_i) \) have to decay faster than the singular values \( \sigma_i \), which imposes a smoothness condition on the observations. In case of a finite dimensional operator, (5.8) always converges. But the vector \( \hat{\mathbf{x}} \) is the discretization of a continuous function which can only be continuous if, on the average, the frequency band decays with increasing frequency. Therefore, the requirement holds for discrete problems as well (Discrete Picard condition, for more details see, e.g., HANSEN 1997). Uncorrelated noise does not decline with higher frequencies, on the contrary, white noise is of equal magnitude for all frequencies, which poses a contradiction to the Picard condition.

The Picard condition is equivalent to demanding the observations \( \mathbf{y} \) to be in the range \( \mathcal{R}(\mathbf{A}) = \text{span}(\mathbf{u}_1, ..., \mathbf{u}_n) \) of the matrix \( \mathbf{A} \). But again, due to measurement noise present in the data, this cannot necessarily be assumed. The noise does not generally belong to \( \mathcal{R}(\mathbf{A}) \). As a solution to Eq (5.1) only exists for \( \mathbf{y} \in \mathcal{R}(\mathbf{A}) \), the system of equations cannot be solved in the rigorous way, but only the distance between \( \mathbf{y} \) and \( \mathbf{Ax} \) can be minimized. This implies that \( \hat{\mathbf{x}} \) in Eq. (5.8) is the following solution in the least squares sense,

\[
\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{y},
\] (5.9)

with \( \mathbf{A}^+ \) denoting the generalized inverse or Moore-Penrose inverse, see, for example, GROETSCHE (1977) or BJÖRCK (1996),

\[
\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \sum_{i=1}^{n} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T.
\] (5.10)

In case of compact operators in a Hilbert space, described in Appendix A.2, it is stated that a compact operator in the infinite dimensional case can never have a continuous and therefore bounded inverse. In case of a finite dimensional operator, if the inverse exists, it is also bounded and thus continuous. Therefore, strictly speaking, Hadamard’s continuity principle is not violated. But, nevertheless, the solution becomes numerically unstable, and the amplified errors in the observations contaminate the solution.

A linear system of the form (5.1) is regarded as ill-posed if the singular values of \( \mathbf{A} \) descend very fast. Crucial is the ratio of the largest singular value compared to the smallest singular value that is referred to as condition number of the matrix,

\[
\text{cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_n}.
\] (5.11)
5. Solving the System of Observation Equations

In case of an ill-posed problem, it tends to take very large numbers. The matrix is then said to be ill conditioned. From Eq. (5.6) it can be assumed that singular values equal to zero span the null space $N(A)$ of the matrix $A$, as defined by Eq. (A.22). This results in the solution of Eq. (5.1) not to be unique any more. In case of singular values very close to zero, the system might be uniquely solvable from a strictly mathematical point of view, but will become severely unstable.

If the design matrix is ill conditioned, the data by itself is insufficient to obtain a stable solution. Referring to least squares estimation, this implies that the minimization of the square sum of the residuals,

$$J(x) = \frac{1}{\sigma^2} (Ax - y)^T P_c (Ax - y) := \|Ax - y\|^2_{C(c)},$$  \hspace{1cm} (5.12)

by itself does not lead to a stable solution. Therefore, additional prior information about the solution has to be introduced. These can be a-priori values for the unknown parameters, e.g. derived from existing models, and smoothness assumptions in terms of an a-priori given covariance matrix of the unknown parameters. The stabilization procedure is known as regularization and will be further described in the next section.

5.2 Regularization

In literature a variety of regularization methods has been proposed; an overview can be found, e.g., in Louis (1989), Bouman (1998), or Kusche (2002). Among them are the Tikhonov regularization (Tikhonov 1963), the truncated singular value decomposition (see, for example, Hansen 1987), and regularization techniques taking advantage of the regularizing character of iterative solution strategies (Landweber 1951) such as conjugate gradients. In this thesis, only the Tikhonov regularization has been applied, therefore it will be described in more detail in the following.

5.2.1 Tikhonov Regularization

This regularization method was independently developed by Tikhonov (1963) and Phillips (1962) and is, therefore, also referred to as Tikhonov-Philips regularization. The procedure is based on the minimization of the functional (5.12) under consideration of a ‘penalty term’ $\|s\|^2_K$ that describes the norm of the signal $s$ given as linear functional $Lx$ of the unknown parameters $x$,

$$s = Lx.$$  \hspace{1cm} (5.13)

$\|s\|^2_K$ stands for a smoothing norm, for example given by the inner product of a reproducing kernel Hilbert space with kernel $K$ according to (3.21),

$$\|s\|^2_K = \langle s, s \rangle_K.$$  \hspace{1cm} (5.14)

The requirement of this norm to be finite,

$$\|s\|^2_K \leq c < \infty,$$  \hspace{1cm} (5.15)

does not only introduce a bound on the norm of the solution, but does also imply a smoothness condition defined by the kernel $K$ in equivalence with the smoothness assumption imposed by the kernel of an RKHS in Section 3.2. The resulting functional to minimize (5.12) and (5.14) simultaneously can be formulated as follows,

$$J_\alpha(x) = \|Ax - y\|^2_{C(c)} + \alpha \|s\|^2_K,$$  \hspace{1cm} (5.16)

with $\alpha$ being the regularization or smoothing parameter. The minimization of Eq. (5.16) constitutes a compromise between minimizing the residual norm and keeping the penalty term $\|s\|^2_K$ small. Due to the ill-posedness of the original model, small errors in the data lead to large deviations in the solution. In order to avoid this, the norm of the solution has to be limited.
5.2. Regularization in the Finite Dimensional Model

As smoothing kernel (in the following in the discrete sense defined by Eq. (3.21)) can serve an a-priori known covariance matrix of the signal,

\[ K = C(s) = \sigma^2_s P_s^{-1}, \]

(5.17)

with \( \sigma_s \) being the variance factor of the signal. The scalar product in an RKHS defined by a (finite-dimensional) matrix can be formulated according to (3.21) as the \( L_2 \) inner product containing the inverse of the kernel matrix following

\[ \|s\|_{C(s)}^2 = \langle s, s \rangle_{C(s)} = \left\langle s, \frac{1}{\sigma^2_s} P_s s \right\rangle. \]

(5.18)

Under consideration of the prior information

\[ E(s) = 0 \quad \text{and} \quad C(s) = \sigma^2_s P_s^{-1}, \]

(5.19)

an equivalent formulation to (5.16) of the Tikhonov regularization based on the Bayes-approach is given by

\[ J_\alpha(x) = \left\| \begin{pmatrix} A \\ L \end{pmatrix} x - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2_{C}, \]

(5.20)

with the inverse of the extended covariance matrix

\[ \tilde{C}^{-1} = \begin{pmatrix} \frac{1}{\sigma^2_c} P_c & 0 \\ 0 & \frac{1}{\sigma^2_s} P_s \end{pmatrix} = \frac{1}{\sigma^2_c} \begin{pmatrix} P_c & 0 \\ 0 & \sigma^2_s P_s \end{pmatrix}. \]

(5.21)

The expectation value \( E(s) = 0 \) can be reasoned by the fact that (especially in a regional refinement approach) the estimated signal always represents a residual field to a (global) reference field. The minimum \( x_\alpha \) of (5.16) and (5.20) is defined by the unique solution of the regularized normal equations,

\[ x_\alpha = \left( A^T L^T \begin{pmatrix} \sigma^2_c P_c & 0 \\ 0 & \sigma^2_s P_s \end{pmatrix} \begin{pmatrix} A \\ L \end{pmatrix} \right)^{-1} A^T L^T \begin{pmatrix} \sigma^2_c P_c & 0 \\ 0 & \sigma^2_s P_s \end{pmatrix} \begin{pmatrix} y \\ 0 \end{pmatrix} \]

(5.22)

The regularization parameter \( \alpha \) in Eq. (5.16) is interpreted here as signal-to-noise ratio,

\[ \frac{1}{\alpha} = \frac{\sigma^2_c}{\sigma^2_s}. \]

(5.23)

This enables the regularized solution to be expressed according to

\[ x_\alpha = (A^T P_c A + \alpha L^T P_s L)^{-1} A^T P_c y. \]

(5.24)

In this context, \( x_\alpha \) presents an unbiased least squares estimate of the unknown parameters under consideration of the prior information given in (5.19), when the Tikhonov regularization is interpreted as introduction of prior information in the Bayesian sense (Koch 1990). It shall be mentioned, however, that the question whether the regularized solution can be regarded as unbiased is a controversially discussed matter, refer to, e.g., Xu and Rummel (1994), Xu et al. (2006), and Koch and Kusche (2007).

The choice of the regularization parameter \( \alpha \) is a crucial task, as it resembles the trade-off between the fitting of the solution to the given data set and the norm and smoothness of the regularized solution. The larger the regularization parameter is chosen, the stronger is the dampening and the smoother the solution. The parameter is to be chosen in accordance with the given signal. On the one hand, the solution has to be regularized as strongly as necessary to obtain a reasonable solution. On the other hand, the signal should not be dampened too much in order not to lose any information contained in the data. Several procedures to choose this parameter have been proposed, e.g. the L-curve criterion (Hansen 1992), the generalized cross validation (going back to Wahra 1977, see also Hansen 1987), minimizing the total mean square error (Xu 1992), or the discrepancy principle (Morozov 1966). In this thesis, the regularization parameter is interpreted as signal-to-noise ratio and is determined by variance component estimation as proposed by Koch and Kusche (2001). Details of this method are given in Section 5.2.4.
5.2.3 Regularization and Splines

As described in Chapter 5.2, the regularization procedure aims at smoothing the solution of an ill-posed problem. In order to do so, certain smoothing conditions are imposed on the solution by adding a penalty term to the minimization functional according to (5.16). This penalty term $\|s\|_K^2$ restricts the norm of a linear functional of the unknown parameters, in this case the spline coefficients $a_i$ being arranged in the vector $x = (a_0, ..., a_I)^T$. The linear functional $Lx$ of the unknown parameters can then be interpreted as

$$Lx = s(x) = \sum_{i=1}^{I} a_i \Phi_i(x, x_i). \tag{5.25}$$

Considering the a-priori information given in (5.19) and taking (5.17) into account, the functional (5.16) can be reformulated according to

$$J_\alpha(x) = (Ax - y)^T P_s (Ax - y) + \alpha \langle s, s \rangle_C. \tag{5.26}$$

The smoothing condition is induced by using the a-priori covariance matrix of the signal $C = C(s)$, in case of the gravitational potential given by (3.58). With the signal expressed in terms of radial basis functions according to (3.74), the scalar product is defined by

$$\langle s, s \rangle_C = \left( \sum_i a_i \Phi_i, \sum_k a_k \Phi_k \right)_C = \sum_{i,k} a_i a_k \langle \Phi_i, \Phi_k \rangle_C. \tag{5.27}$$

If the scalar products of the basis functions $\langle \Phi_i, \Phi_k \rangle_C$ are combined in the regularization matrix $R$ with the elements

$$R_{ik} = \langle \Phi_i, \Phi_k \rangle_C, \tag{5.28}$$

then (5.26) becomes

$$J_\alpha(x) = (Ax - y)^T P_s (Ax - y) + \alpha \cdot x^T Rx. \tag{5.29}$$

The regularized solution for the minimization of the functional $J_\alpha(x)$ is then given by

$$x_\alpha(x) = (A^T P_s A + \alpha R)^{-1} A^T P_s y. \tag{5.30}$$

Next, the nature of the regularization matrix $R$ shall be investigated. Therefore, the elements of the regularization matrix in (5.28) will be calculated. In the limit, the inner products represent integrals over the sphere. If the covariance function (3.58) is chosen as reproducing kernel, and the basis functions are chosen according to (3.73), the elements of the regularization matrix result in

$$R_{i,k} = \langle \Phi_i, \Phi_k \rangle_C = \int_{\Omega} \frac{1}{k^2} \left[ \sum_{n=2}^{\infty} \sum_{m=-n}^{n} k_n Y_{nm}(x) Y_{nm}(x_i) \right] \left[ \sum_{n=2}^{\infty} \sum_{m=-n}^{n} k_n Y_{nm}(x) Y_{nm}(x_k) \right] \tag{5.31}$$

$$= \sum_{n=2}^{\infty} \sum_{m=-n}^{n} Y_{nm}(x_i) Y_{nm}(x_k) = \delta(x_i, x_k).$$

$\delta(x_i, x_k)$ denotes the Dirac functional, which implies that it has the value zero for two different functions, but is infinite in the case $i = k$ with the integral over $\delta$ being one. Thus in a certain sense, the basis functions diagonalize the regularization matrix, as any two functions located at two different points are decorrelated with respect to the inner product defined by the kernel. This is very much appreciated, as it allows the separation of the regularization matrix, as will be described in Section 5.2.5. But the problem is the fact that each $\Phi_i$ itself does not possess finite energy concerning the scalar product defined by $C$. This means that the norm with respect to $C$ is not finite,

$$\|\Phi_i\|_C = \infty. \tag{5.32}$$
which results in the elements on the main diagonal of $R$ to become infinite. The implication of Eq. (5.32) is the fact that the basis functions $\Phi_i$ do not belong to the RKHS defined by the covariance function $C$. They are not smooth enough to satisfy the requirement imposed by $C$ as reproducing kernel, as has already been stated in Eq. (3.90).

If one is interested in adapting the basis functions in such a way to make them part of the Hilbert space defined by the reproducing kernel $C$, the basis functions would have to become smoother. This could be achieved by introducing a damping factor of $(1/\sqrt{1+\epsilon})^n$ ($\epsilon$ being a small positive number) into the series expansion,

$$\tilde{\Phi}_i(x, x_i) = \sum_{n=2}^{\infty} \sqrt{2n+1} \cdot k_n \cdot \left( \frac{1}{\sqrt{1+\epsilon}} \right)^n P(x \cdot x_i),$$

leading to the following elements of the regularization matrix,

$$R_{i,k} = \langle \tilde{\Phi}_i, \tilde{\Phi}_k \rangle_C = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \left( \frac{1}{1+\epsilon} \right)^n Y_{nm}(x_i)Y_{nm}(x_k).$$

Eq. (5.34) reveals that, in contrast to the original basis functions, the modified spline kernels lack the orthogonality regarding the respective inner product.

The above considerations have been made for non-bandlimited spline functions. In practical calculations, however, the respective basis functions always have to be bandlimited with an upper degree $N$ leading to the functions defined in Section 3.4.3. From this it follows that the series expansion used in the calculation of the scalar product in (5.31) is truncated at degree $N$ as well. When dealing with a truncated series, of course, the problem of infinite energy does not exist any more. Hence, in case of bandlimited spline functions, the energy on the main diagonal of the regularization matrix becomes finite, even without introduction of a damping factor. However, the inner product (5.31) can only be orthogonal if the summation is performed up to $N = \infty$. Thus the bandlimited spline functions lose their strict orthogonality in the sense of the inner product with respect to $C$. The elements $R_{i,k}$ of the regularization matrix reflect the correlations between two different basis functions located at two different nodal points $x_i$ and $x_k$.

An ideal case would provide a set of basis functions being decorrelated by the scalar product induced by the covariance function as reproducing kernel, but at the same time having finite energy with respect to the norm defined by the same kernel. The simultaneous satisfaction of both requirements would result in the regularization matrix becoming the unit matrix (or at least differing from the unit matrix only by a constant factor which could be absorbed in the regularization parameter). However, the fulfillment of both conditions is not easily available (especially for bandlimited basis functions) due to the issues of convergence of the scalar product and the loss of orthogonality discussed above.

Nevertheless, the regularization matrix is to be approximated by the unit matrix. This provides the advantage of separating the regularization matrix and calculating different matrices for different areas, as will be discussed in more detail in Section 5.2.5. The approximation

$$R = I$$

leads to the functional $J_\alpha(x)$ of Eq. (5.29) to be modified according to

$$J_\alpha(x) = (Ax - y)^T P_\epsilon (Ax - y) + \alpha \cdot x^T x,$$

resulting in the regularized solution

$$x_\alpha(x) = (A^T P_\epsilon A + \alpha I)^{-1} A^T P_\epsilon y.$$

It is understood that the use of $I$ instead of $R$ is only an approximation, therefore it has to be investigated in how far this approximation has any effect on the solution. This will be attended to in the following considerations and in a simulation scenario described in Section 7.2.
To illustrate the effect of the non-orthogonality of the bandlimited spline kernels and to show the effect of the modified basis functions (5.33), examples of regularization matrices are plotted in Fig. 5.1. The spline kernels in this example have been developed up to $N = 14$ and have been arranged on a matching triangle vertex grid (see Section 3.5.1.6) of level 4. The left hand side of the figure shows the regularization matrix resulting from the (bandlimited) inner products for the unmodified basis functions given by (5.31). The matrix is normalized to provide the value one for the elements on the main diagonal. The figure specifies that the elements apart from the main diagonal are significantly smaller than one, the highest correlation between two different spline kernels being about 0.12, as listed in Tab. 5.1. Even though this represents only one example of a specific basis function and grid combination, the situation does not change much when the regularization matrix is calculated for higher resolution spline kernels and their corresponding nodal point arrangements. The highest correlation factor is always around 10% of the value on the main diagonal. On the right hand side of Fig. 5.1, the same regularization matrix is plotted for modified spline kernels and their inner products given by (5.34), here presented with a modification factor of $(1/\sqrt{1+\epsilon})^n = 0.9$. This again represents only an exemplary value, the more $\epsilon \to 0$, the more the regularization matrix resembles the original one. While the overall sum of the absolute values of the elements apart from the main diagonal decreases with the modification factor, the highest correlation and the standard deviation of the off-diagonal elements increase. The impact of the modification factor on the basis functions and the resulting regularization matrices will be investigated in a simulation scenario in Section 7.2. There the impact of the unit matrix to be used as regularization matrix will be investigated, too.

<table>
<thead>
<tr>
<th></th>
<th>highest correlation</th>
<th>standard deviation</th>
<th>sum of absolute values</th>
</tr>
</thead>
<tbody>
<tr>
<td>original kernel</td>
<td>0.128337</td>
<td>6.08808</td>
<td>1163.07</td>
</tr>
<tr>
<td>modified kernel</td>
<td>0.285343</td>
<td>9.01617</td>
<td>921.36</td>
</tr>
</tbody>
</table>

Table 5.1: Difference between the regularization matrix and the unit matrix, triangle vertex level 4, $N = 14$

Mathematically, the impact of the use of the unit matrix as regularization matrix can be exploited by investigating the difference $D$ between regularized normal equations with fully occupied regularization matrix compared to the normal equations with the use of the unit matrix according to

$$D = (A^T P A + \alpha I)^{-1} - (A^T P A + \alpha R)^{-1}.$$

(5.38)
With the simplifications $B = (A^T P_x A + \alpha I)$ and $B' = (A^T P_x A + \alpha R)$, this reads
\[ D = B^{-1} - B'^{-1}. \] (5.39)

Factoring out the matrices $B'^{-1}$ and $B^{-1}$ results in
\[ D = B^{-1}(I - BB'^{-1}) = B^{-1}(B' - B)B'^{-1}. \] (5.40)

Resubstituting the original matrix expressions results in the following difference,
\[ D = (A^T P_x A + \alpha I)^{-1}(R - I)(A^T P_x A + \alpha R)^{-1}. \] (5.41)

This implies that the difference between the regularized normal equations can directly be expressed by the difference $(R - I)$ between the matrices themselves.

It shall be pointed out that a regularization matrix equivalent to Eq. (5.34) would be obtained in case of unmodified basis functions if the modified covariance function $\bar{C}$, proposed in Eq. (3.69), was applied
\[ \langle \bar{\Phi}_i, \bar{\Phi}_k \rangle_C = \langle \Phi_i, \Phi_k \rangle_{\bar{C}}. \] (5.42)

In this case, the basis functions (even without damping factor) would directly belong to the RKHS defined by $\bar{C}$. Even though the regularization matrix equals the one given in Eq. (5.34), the basis functions used in the modeling differ in both cases. The impact of this modified covariance matrix and the resulting regularization matrix is also investigated in the simulation scenario presented in Section 7.2.

### 5.2.4 Variance Component Estimation

If different kinds of observations are to be combined, the determination of the correct relative weighting of the observations is essential to receive a correct result. In the same context, the choice of the regularization parameter can be treated if the regularization is interpreted as prior information in the Bayesian sense. For both tasks the method of variance component estimation (VCE), as described by Koch and Kusche (2001), can be applied.

The solution can be estimated from a system of (combined) normal equations $N$ which is accumulated as a weighted sum of the normal equation systems $N_k$ of the individual observation groups. This can be formulated according to
\[ N \hat{x} = n \quad \text{with} \quad N = \sum_k \frac{1}{\sigma_k^2} N_k \quad \text{and} \quad n = \sum_k \frac{1}{\sigma_k^2} n_k. \] (5.43)

The weighting factors are the reciprocal variances of the normal equations,
\[ \sigma_k^2 = \frac{\Omega_k}{r_k}, \] (5.44)

with
\[ \Omega_k = e_k^T P_k e_k = (A_k \hat{x} - l_k)^T P_k (A_k \hat{x} - l_k), \] (5.45)

being the square sum of the residuals of the $k$th group of observations and
\[ r_k = n_k - \frac{1}{\sigma_k^2} \text{tr} (N_k N_k^{-1}) \] (5.46)

their partial redundancy with $n_k$ denoting the number of observations in the $k$th group. The partial redundancies sum up to the overall redundancy $\sum_k r_k = n - u$. The solution $\hat{x}$ and the variances $\sigma_k^2$ are unknown a-priori, consequently an iterative procedure is inevitable, as illustrated in the flow chart of Fig. 5.2.
As in this thesis the method of VCE is primarily utilized for the task of the determination of the regularization parameter, this special case of the Tikhonov regularization ought to be stated explicitly as well. Here the described approximation of the regularization matrix by the unit matrix is assumed, which leads to the following system,

\[ N = \frac{1}{\sigma^2} A^T P_A A + \frac{1}{\sigma^2} I \quad \text{and} \quad n = \frac{1}{\sigma^2} A^T P_A l, \]  

with \( \sigma \) being the standard deviation of the observations and \( \sigma_s \) being the standard deviation of the signal. Both variance components can be treated as above described.

### 5.2.5 Regionally Adapted Regularization

The process of variance component estimation delivers the optimal regularization parameter under consideration of the given signal-to-noise ratio. In case of a regional gravity field determination, this results in one regularization parameter tailored optimally to the respective recovery region. This is an improvement in comparison to a global gravity field parameterization which allows only one regularization factor for the complete Earth, resulting in an overall mean damping of the gravity field features. But even within smaller geographical areas, the gravity field features may vary significantly. Therefore, it seems reasonable to further adapt the regularization procedure. The proposed approach does not take into account only one regularization matrix with one associated regularization parameter per region, but allows several matrices with respective parameters,

\[ N = \frac{1}{\sigma^2} A^T P_A A + \frac{1}{\sigma^2} R_1 + \ldots + \frac{1}{\sigma^2} R_n. \]  

The original regularization matrix is, therefore, split up into individual regularization matrices \( R_i \), each belonging to a regional regularization area \( i \). To separate the original regularization matrix, the approximation by a unit matrix (5.35) made above is very convenient. \( R \) can only be separated that easily if the basis
functions are assumed to be orthogonal with respect to the inner product given by \((5.31)\). In this case, each individual regularization matrix \(R_i\) is a diagonal matrix that features a ‘one’ for each regional spline parameter located inside the corresponding region and a ‘zero’ for parameters belonging to basis functions outside the regularization group,

\[
R_i(j, j) = \begin{cases} 
1 & \text{for } j \text{ inside } i \\
0 & \text{for } j \text{ outside } i
\end{cases}
\]  

(5.49)

The original identity matrix, as applied in the Tikhonov regularization process, has thus been divided into single diagonal matrices according to the membership of the respective unknown parameters to the different regularization groups,

\[
R_1 + \ldots + R_n = I.
\]  

(5.50)

The possibility of adapting the regularization procedure in this particular way is a unique feature of a field parameterization by space localizing basis functions, as each unknown parameter is related to a particular geographical location. This is an inevitable premise when the elements of the regularization matrix are supposed to be assigned to a certain region. The separation of a geographical region into different regularization areas is exemplarily illustrated in Fig. 5.3. Here the two regularization areas are the continental and the oceanic regions, resulting in the following normal equation matrix,

\[
N = \frac{1}{\sigma_t^2} A^T PA + \frac{1}{\sigma_{\text{Land}}^2} R_{\text{Land}} + \frac{1}{\sigma_{\text{Ocean}}^2} R_{\text{Ocean}}.
\]  

(5.51)

The separation into land and ocean areas can propose a reasonable choice in certain regions, where the gravity field information on the oceans is significantly less rough compared to the continent areas. In the presence of deep sea trenches or rough ocean bottom topography, however, this might not always be a valid assumption. In this case, different choices for the regularization areas are inevitable.
5.3 Relationship Between Spline Approximation and Collocation

In the following, the relationships shall be investigated that exist between the (regularized) least squares solution parameterized by the space localizing basis functions defined in Section 3.4.2 and the least squares collocation approach. The considerations will be performed without going into detail about the concept of least squares collocation; information about this method can be found in the literature, examples are Moritz (1962), Krarup (1969), and Moritz (1978).

The following formula specifies the case of least squares collocation with noise. It describes the prediction of a signal $s$ on the basis of a set of observations $I$,

$$s = C_{s,l} \left( C_{l,l} + \alpha P^{-1}_e \right)^{-1} l,$$

with the autocovariance matrix $C_{l,l}$ of the signal part of the observations, the cross-covariance matrix $C_{s,l}$ between the signal, and the observations and the covariance matrix $P$ of the observation noise. The factor $\alpha$ can be regarded as signal-to-noise ratio,

$$\frac{1}{\alpha} = \frac{\sigma_s}{\sigma_e}.$$

The solution obtained by least squares collocation is optimal in the sense of minimal variance on the basis of the given observations (Moritz 1980). In the following, it will be assumed that the observations as well as the functionals to be predicted at different positions are given as gravitational potential. The same prediction as given in Eq. (5.52) can be performed by using the spline representation,

$$s = Bx.$$

The vector $x$ contains the estimated spline parameters $a_i$ given in Eq. (3.74), and $B$ represents the prediction matrix according to Eq. (3.76) with the elements

$$(B)_{ki} = \sum_{n=2}^{N} \sigma_n \cdot P_n (\cos(x_k \cdot y_i)) = \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(y_k)Y_{nm}(x_i).$$

Here the $y_k$ are the positions to which the signal is predicted, while $x_i$ denote the nodal points of the spline kernels. By substituting the solution of the least squares adjustment for the estimated parameters in Eq. (5.54), the following relationship is obtained,

$$s = B \left( A^T P A + \alpha I \right)^{-1} A^T P l.$$

Rearranging Eq. (5.56) by making use of matrix identities as given in, for example, Koch (1997) yields

$$s = BA^T \left( AA^T + \alpha P^{-1}_e \right)^{-1} l.$$

Comparing Eq. (5.52) and Eq. (5.57) reveals that the two formulas exhibit certain similarities. They would be identical if the following relationships would hold,

$$BA^T = C_{s,l} \quad \text{and} \quad AA^T = C_{l,l}.$$

Carrying out the matrix multiplications results for the individual matrix elements in

$$BA^T = \sum_{j=1}^{I} \left( \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(y_k)Y_{nm}(x_i) \right) \left( \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(y_j)Y_{nm}(x_i) \right) \approx \int Y_{nm}(x_i)Y_{nm}(x_j) d\Omega = \delta_{nm} \delta_{ij},$$

$$\approx \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n^2}{\sqrt{2n+1}} Y_{nm}(y_k)Y_{nm}(y_j) = C_{s,l(kj)}.$$
and

\[
\mathbf{A A^T}_{(hj)} = \sum_{i=1}^{l} \left( \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(y_h) Y_{nm}(x_i) \right) \left( \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(y_j) Y_{nm}(x_i) \right) \\
= \sum_{n=2}^{N} \sum_{m=-n}^{n} \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n \sigma_m}{\sqrt{2n+1} \sqrt{2n+1}} Y_{nm}(y_h) Y_{nm}(y_j) \sum_{i=1}^{l} Y_{nm}(x_i) Y_{nm}(x_i) \\
\approx \sum_{n=2}^{N} \sum_{m=-n}^{n} \frac{\sigma_n^2}{2n+1} Y_{nm}(y_h) Y_{nm}(y_j) = C_{l,l(hj)},
\]

with \(y_h\) and \(y_j\) denoting the observation points. A comparison with Eq. (3.58) allows the assumption that the matrix products \(\mathbf{B A^T}\) and \(\mathbf{A A^T}\) in Eq. (5.57) show certain resemblances to the covariance matrices \(C_{s,l}\) and \(C_{l,l}\). This is a consequence of the shape coefficients of the radial basis functions \(k_n\) in Eq. (3.73) being chosen according to Eq. (3.77) as the square root of the coefficients of the covariance function. This can be regarded as evidence that the choice of the shape coefficients seems to be a reasonable one. Nevertheless, Eqs. (5.59) and (5.60) only represent approximations. The matrix products would only become exactly the covariance matrices if the sum over the basis functions \(\sum_{i=1}^{l} Y_{nm}(x_i) Y_{nm}(x_i)\) would turn into the integral over the sphere \(\int_{\Omega} Y_{nm}(x_i) Y_{nm}(x_i) d\Omega\). This would require an infinitely dense distribution of spline kernels over the surface of the sphere. Since this is not the case in practical calculations, no one-to-one relationship between the spline approximation as treated in this thesis and the least squares collocation can be derived from the above considerations. However, they confirm the usefulness of the basis functions. Further investigations of the discussed relationships are necessary.
6. From Regional to Global Gravity Fields

For a wide variety of applications, the calculation of regional gravity field solutions meets exactly the given requirements. Nevertheless, for some other applications, it seems to be useful to derive a global gravity field model by spherical harmonics without losing the details of a regional zoom-in. In the geodetic context, a set of spherical harmonics is often required, as it can easily be distributed and generally be handled without further instructions. Furthermore, it might be reasonable to deliver a gravity field parameterized by spherical harmonics for purposes of comparison. Most of the existing gravity field models are provided in terms of spherical harmonic coefficients, and, for instance, comparisons of accuracy are frequently performed on the basis of error degree variances. Thus if the gravity field model is to be published and acknowledged by the geodetic user community, the allocation of a spherical harmonic expansion appears to be desirable. In the first section of this chapter, it is derived how the spherical harmonic coefficients can directly be calculated from the coefficients of a global model parameterized by radial basis functions. If individual regional gravity field solutions are available with global coverage, then the determination of the spherical harmonic coefficients can be performed by means of quadrature methods, as described in the second section of this chapter.

6.1 Conversion from a Global Spline Representation to Spherical Harmonics

In principle, a gravity field representation based on space localizing basis functions cannot only be used for regional models, but for the representation of global gravity fields as well. The spline kernels are then located at a global grid, and a global set of observations is needed. Such a global spline representation can be transformed into a series of spherical harmonics without loss of information. This can be understood from the following considerations. Starting point is a representation of the gravitational potential in terms of splines,

\[
V(r) = \sum_{i=1}^{I} a_i \Phi_i(r) = \sum_{i=1}^{I} a_i \left[ \frac{GM}{R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sqrt{2n+1} k_n P_n \left( \frac{r_i}{R} \right) \right] \notag
\]

\[
= \sum_{i=1}^{I} a_i \left[ \frac{GM}{R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} k_n \sum_{m=-n}^{n} Y_{nm} \left( \frac{r_i}{R} \right) Y_{nm} \left( \frac{r}{R} \right) \right]. \tag{6.1}
\]

Eq. (6.1) can be reordered to yield

\[
V(r) = \frac{GM}{R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=-n}^{n} \left( \sum_{i=1}^{I} a_i k_n Y_{nm} \left( \frac{r_i}{R} \right) \right) Y_{nm} \left( \frac{r}{R} \right). \tag{6.2}
\]

Comparing this with the expansion of the gravitational potential in terms of spherical harmonics according to Eq. (3.42),

\[
V(r) = \frac{GM}{R} \sum_{n=2}^{\infty} \left( \frac{R}{r} \right)^{n+1} \sum_{m=-n}^{n} c_{nm} Y_{nm} \left( \frac{r}{R} \right), \tag{6.3}
\]

the spherical harmonic coefficients can be computed directly from the spline coefficients according to

\[
c_{nm} = \sum_{i=1}^{I} a_i k_n Y_{nm} \left( \frac{r_i}{R} \right), \tag{6.4}
\]
6.2 Patching of Individual Regional Solutions

If regional solutions have been calculated independently, the determination of a spherical harmonic expansion cannot simply be performed by applying Eq. (6.4). In order to avoid truncation effects, the satellite data has to be taken in an area slightly larger than the evaluation area itself. The radial basis functions are not perfectly space localizing, therefore those located in the boundary area have to be taken into account as well, as they have influence on the gravity field values in the inner evaluation area. The resulting global solution, however, is only to be composed of the inner areas without the boundaries. Therefore, the patching of the regional solutions has to be based on perfectly localizing values, and the strategy described by Eq. (6.4) does not work for individually calculated solutions. Thus a different strategy has to be used. The concept that has been applied in this thesis is the merging of the regional solutions to a global one by means of quadrature methods, as introduced in Eicker et al. (2004). From the regional solutions, the gravity functionals, parameterized by the spline representation in the specific regions, can be calculated at the nodes of a global grid. This can be performed, in principle with arbitrary resolution, and the continuous modeling by spherical splines allows the functionals to be evaluated at arbitrary grid points. The calculation of the spherical harmonic coefficients occurs in a second step by applying tailored quadrature formulas, as introduced in Section 6.2.1.

6.2.1 Quadrature Methods

Quadrature formulas in general, as described for example by Bronstein and Semendjajew (1995), are a widely-used tool for the numerical evaluation of definite integrals. In the context of this thesis, quadrature procedures are applied to calculate spherical harmonic coefficients from discrete gravity field values provided on a spherical grid. This results in the numerical solution of the integral (3.39),

$$ c_{nm} = \frac{1}{4\pi} \int_{\Omega} f(x) Y_{nm}(x) d\Omega. $$

(6.5)

The integral is to be approximated by a weighted sum of function values $f(x_i)$ at given points $x_i$ resulting in

$$ c_{nm} = \frac{1}{4\pi} \sum_{i=1}^{l} w_i f(x_i) Y_{nm}(x_i), $$

(6.6)

where the weights $w_i$ can be regarded as the surface element associated with the grid point. The weights sum up to the surface area of the unit sphere,

$$ \sum_{i=1}^{l} w_i = 4\pi. $$

(6.7)

Eq. (6.6) only provides the correct result if the orthonormality of the spherical harmonics is preserved when converting the continuous integral to a discrete (weighted) sum, i.e. if the following relationship holds for the discrete case,

$$ \langle Y_{nm}(x), Y_{n'm'}(x) \rangle_W = \frac{1}{4\pi} \sum_{i=1}^{l} Y_{nm}(x_i) Y_{n'm'}(x_i) w_i = \delta_{nm} \delta_{n'm'}. $$

(6.8)

This is not generally true for arbitrary point distributions and corresponding weights $w_i$. The problem of discrete orthogonality can be addressed by investigating the corresponding weighted least squares adjustment procedures of estimating spherical harmonic coefficients from gridded data, as, for example, described by Sneeuw (1994). If the basis functions $Y_{nm}$ are combined in the design matrix $A$, and the weights are combined in the diagonal matrix $W$, the least squares estimate of the coefficients $\hat{c}$ is given by

$$ \hat{c} = (A^T WA)^{-1} A^T W f. $$

(6.9)
In order to compare this result with a given quadrature formula, Eq. (6.6) can be formulated in matrix notation as well,
\[ c = A^T W f. \] (6.10)

Obviously, the two results are identical in case of a normal equation matrix \[ N = A^T W A \] being the identity matrix,
\[ N = A^T W A = I. \] (6.11)

Every element of \( N \) represents the weighted inner product of the spherical harmonics according to Eq. (6.8). Therefore, the values of \( N \) apart from the main diagonal represent the amount of non-orthogonality of the corresponding discrete spherical harmonics. How much the normal equation matrix differs from a diagonal matrix depends on the grid point density on the one hand and on the grid pattern and the corresponding weights on the other hand. Therefore, it can be stated that a quadrature formula is the better suited for calculating spherical harmonic coefficients, the more accurately the orthonormality relations for spherical harmonics (3.32) are preserved.

In the following, several of these quadrature formulas will be introduced, each being characterized by the distribution of their quadrature nodes on the one hand and by the corresponding quadrature weights on the other hand. Firstly, two exact quadrature rules are investigated, namely the Gauss-Legendre quadrature and the Driscoll-Healy quadrature. Subsequently, the quadrature method is also tried out with the remaining point distributions described in Section 3.5.1.

### 6.2.1.1 Driscoll-Healy Quadrature

It was found out by Neumann (1838) that exact numerical quadrature up to a spherical harmonic degree \( N = L - 1 \) is possible by using \( 2L \) circles of latitude. They can be chosen arbitrarily, but have to be distinct. For example, an equidistant distribution of parallels is possible. This quadrature method is also quoted as 'Neumann’s first method', see, for example, Sneeuw (1994). Driscoll and Healy (1994) have developed a closed representation for the weights of such an equi-angular spaced point distribution. These weights are applied in the following investigations, therefore the corresponding quadrature method is identified as Driscoll-Healy quadrature. The weights for the numerical integration along parallels are the analytical solution of
\[ \sum_{j=0}^{2L-1} \tilde{w}_i P_k(\cos \frac{j\pi}{2L}) = 2\delta_{k0}. \] (6.12)

This leads to the following weights, their derivation can be found in Driscoll and Healy (1994) with additional explanations provided in Mohlenkamp (1997),
\[ \tilde{w}_i = 4 \frac{\sin(\vartheta_i)}{2L} \sum_{l=0}^{L-1} \frac{\sin [(2l+1) \vartheta_i]}{2l+1}. \] (6.13)

To obtain the weights to be applied in Eq. (6.6), the \( \tilde{w}_i \) have to be multiplied with \( \Delta\lambda \),
\[ w_i = \tilde{w}_i \cdot \Delta\lambda. \] (6.14)

### 6.2.1.2 Gauss-Legendre Quadrature

This method can be found in the work of Neumann (1838) as well, for this reason it is also referred to as 'Neumann’s second method' (see, for instance, Sneeuw (1994) for a historical review on the different quadrature methods). Gaussian quadrature methods are widely applied in numerical analysis, here can be
6.2. Patching of Individual Regional Solutions

referred, e.g., to Lanczos (1956), and are used in geodetic applications as well, see, for example, Payne (1971) and Colombo (1981). In contrast to the Driscoll-Healy quadrature described above, the Gauss-Legendre quadrature allows the recovery of a spherical harmonic expansion of degree $N = L - 1$ from only $L$ circles of latitude. On the other hand, the parallels cannot be chosen arbitrarily, but have to be located along the zeros of the Legendre polynomial of degree $L$. Therefore, the quadrature nodes of the Gauss-Legendre quadrature method coincide with the grid points of the Gaussian grid, as described in Section 3.5.1.2. They feature equi-angular spacing along the circles of latitude, whereas along the meridians the nodal points are located at the $L$ zeros of the Legendre polynomial of degree $L$. The orthogonality relations can best be identified when considering the specification of the spherical harmonic basis functions, as given by Eq. (3.47).

It features a separation of the $Y_{nm}$ into a trigonometric function depending only on the longitude $\lambda$ and the associated Legendre functions depending only on the co-latitude $\vartheta$. This allows to separately investigate the behavior of the basis functions in longitudinal and in latitudinal direction. The determination of the spherical harmonic coefficients can be split into a two-step procedure,

\[
\begin{align*}
    a_m(\vartheta) & = \frac{1}{1 + \delta_{m0} \pi} \int_{0}^{2\pi} f(x) \left\{ \cos(m\lambda) \right\} d\lambda, \\
    b_m(\vartheta) & = \frac{1}{4} \int_{0}^{\pi} \left\{ a_m(\vartheta) \right\} P_{nm}(\cos \vartheta) \sin \vartheta d\vartheta.
\end{align*}
\]

In the following, the analysis of the discretization of the integrals (6.15) and (6.16) will be described separately.

**Longitudinal Direction** The equi-angular spacing in longitudinal direction features $2L$ points along each circle of latitude. On such a regular configuration, the trigonometric functions obey discrete orthogonality relations,

\[
\begin{align*}
    \sum_{i=0}^{2L-1} \cos m\lambda_i \cos \tilde{m}\lambda_i & = (1 + \delta_{m0} + \delta_{mL}) L \delta_{mm}, \\
    \sum_{i=0}^{2L-1} \sin m\lambda_i \sin \tilde{m}\lambda_i & = (1 + \delta_{m0} + \delta_{mL}) L \delta_{mm}, \\
    \sum_{i=0}^{2L-1} \cos m\lambda_i \sin \tilde{m}\lambda_i & = 0.
\end{align*}
\]

Thus the conversion from the integral in Eq. (6.15) to a discrete sum delivers the exact result. The separate calculation in longitudinal direction is also beneficial from a computational point of view, as fast Fourier techniques (FFT) can be applied to the evaluation of Eq. (6.15).

**Latitudinal Direction** More crucial is the quadrature along the meridians. Here it can be made use of one-dimensional Gaussian quadrature procedures, as, for example, described in Bronstein and Semendjajew (1995), applied to function values evaluated at points defined by $t_i = \cos \theta_i$. By the Gaussian quadrature rule a polynomial of degree $2L - 1$ can exactly be integrated by the evaluation of $L$ function values,

\[
\int_{a}^{b} f(t) dt = \sum_{i=1}^{L} f(t_i) w_i.
\]

If the evaluation nodes $t_i$ are chosen as the zeros of orthogonal polynomials. In case of the Gauss-Legendre quadrature, the orthogonal polynomials are the Legendre polynomials $P_n$, \(^1\) and the integration interval equals $[a, b] = [-1, 1]$. The weights $w_i$ can be derived using the characteristics of orthogonal polynomials.

\(^1\)Equivalent Gaussian quadrature methods exist for Laguerre, Hermite, and Chebychev polynomials as well. They each require an additional, specific weighting function $q(t)$ to be introduced into the integral in (6.18). In case of the Gauss-Legendre quadrature, this weighting function equals a constant with the value of one and is, therefore, omitted in the above considerations.
as explained in more detail in Appendix B. For the one-dimensional quadrature along the meridians, the weights at a given co-latitude are \( \theta_i \) depend on the number of parallels \( L \) and are given by

\[
\tilde{w}_i(L) = \frac{2}{(1 - t_i^2)(P_L'(\cos(\theta_i)))^2},
\]

(c.f. Stroud and Secrest (1966). Again, the multiplication with \( \Delta \lambda \) yields the final weights to be used in Eq. (6.6),

\[
w_i = \tilde{w}_i \cdot \Delta \lambda.
\]

The relationship between the recoverable polynomial of degree \( 2L - 1 \) and the at maximum possible spherical harmonic degree \( N = L - 1 \) to be determined from the given point distribution will be investigated in the following.

### 6.2.1.3 Maximum Degree Determinable by Exact Methods

The spherical harmonic analysis with the goal of calculating spherical harmonic coefficients from function values is formulated by Eq. (6.5). Considering Eq. (3.38), it becomes obvious that this requires the integration of products of spherical harmonics \( Y_{nm} \), as illustrated by

\[
c_{nm} = \frac{1}{4\pi} \int f(x) Y_{nm}(x)d\Omega = \frac{1}{4\pi} \int \left( \sum_{n=0}^{N} \sum_{m=-n}^{n} c_{nm}Y_{nm}(x) \right) Y_{nm}(x)d\Omega.
\]

Thus the key point of the discretization process is to deliver a quadrature method valid for products of two spherical harmonics. Again, interpreting the integration in latitudinal and longitudinal direction separately leads to the following integrals,

\[
\int_{0}^{2\pi} \int_{0}^{\pi} P_{nm}(\cos \vartheta) \left\{ \frac{\cos(m\lambda)}{\sin(m\lambda)} \right\} P_{\bar{m}m}(\cos \vartheta) \left\{ \frac{\cos(\bar{m}\lambda)}{\sin(\bar{m}\lambda)} \right\} \sin \vartheta d\vartheta.
\]

For \( m = \bar{m} \) this results in evaluating the integral

\[
\int_{0}^{\pi} P_{nm}(\cos \vartheta)P_{\bar{m}m}(\cos \vartheta) \sin \vartheta d\vartheta.
\]

The product of the two associated Legendre functions \( P_{nm}(\cos \vartheta) \) and \( P_{\bar{m}m}(\cos \vartheta) \) is again a polynomial of degree at most \( 2N \) when \( N \) is the maximum degree to be determined. Therefore, the applied quadrature rule along the meridians needs to have an accuracy level (i.e. the at maximum determinable polynomial degree) of at least \( 2N \). If quadrature nodes are located arbitrarily, \( l \) nodes can provide an accuracy level of \( l - 1 \) (Bronstein and Semendjajew 1995). Therefore, at least \( 2N + 1 \) circles of latitude are required to achieve the accuracy level of \( 2N \) and thus to exactly evaluate the integral (6.23). This illustrates that, in case of the Driscoll-Healy quadrature, the \( 2L \) points along the meridians allow the determination of a maximum spherical harmonic degree \( N = L - 1 \). If the quadrature nodes are chosen at the zeros of the Legendre polynomial of degree \( L \), as in case of the Gauss-Legendre quadrature, the accuracy level of the corresponding quadrature rule can be enhanced. In this case \( l \) nodes allow an accuracy level of \( 2l + 1 \), or, in other words, the \( L \) meridians allow the recovery of polynomials up to degree \( 2L + 1 \). From this it results that the at maximum recoverable spherical harmonic coefficients are again of degree \( N = L - 1 \). This can be shown, as the product of the Legendre functions has then a maximum degree of \( 2 \cdot (L - 1) = 2L - 2 \) which is smaller than the possible polynomial degree of \( N = L - 1 \). In contrast to the Driscoll-Healy quadrature, however, this resolution can be achieved with only about half the number of circles of latitude. As both grids feature the same number of \( 2L \) grid points in longitudinal direction, the Gauss-Legendre quadrature requires only half the number of overall evaluation points. This is the reason why the Gauss-Legendre quadrature will be applied to the calculation of spherical harmonic expansions from regional gravity field solutions. The drawback of the irregular spacing of the quadrature nodes is not relevant, as the continuous representation of the regional solutions by radial basis functions allows the evaluation of the functionals at arbitrary points.
6.2.1.4 Quadrature in Case of Arbitrary Point Distributions

In principle, the quadrature formula (6.6) can be applied to any kind of point distribution. For reasons of completeness and in order to point out the differences to the exact quadrature procedures described above, the performance of the rest of the grids, described in Section 3.5.1, have been investigated as well. Thereby special emphasis is placed on the question of discrete orthogonality of the spherical harmonic basis functions, as this defines the accuracy of the quadrature method. In Fig. 6.1 the different normal equation matrices, as given by Eq. (6.11), are displayed for the calculation of spherical harmonic coefficients up to degree \( n = 10 \). The elements of the normal equation matrices contain the products of the spherical harmonic basis functions and, therefore, illustrate the discrete orthogonality relations.

**Geographical Grid**  The quadrature weights are chosen as the surface area associated with each grid point,

\[
 w_i = \int_{\lambda_i - \Delta \lambda}^{\lambda_i + \Delta \lambda} \int_{\vartheta_i - \Delta \vartheta}^{\vartheta_i + \Delta \vartheta} 2 \cdot \Delta \lambda \sin(\Delta \vartheta) \sin(\vartheta_i). \tag{6.24}
\]

The discrete orthogonality relations existing in case of the geographical grid can be investigated as to the longitudinal and latitudinal direction separately, as described above. Along the circles of latitude, the 2\( L \) equi-angular spaced points again preserve the discrete orthogonality of the trigonometric functions, as given by Eq. (6.17). Therefore, the \( C_{nm} \) and \( S_{nm} \) are orthogonal for different orders \( m \), and, within the same order, the sine and cosine terms are orthogonal as well. Along meridians, the orthogonality among the associated Legendre functions is destroyed by the transition from the continuous to the discrete case. But the (anti-)symmetry characteristics of the associated Legendre functions,

\[
P_{nm}(- \cos \vartheta) = (-1)^{n-m} P_{nm}(\cos \vartheta), \tag{6.25}
\]

can be exploited. They ensure within the same order the independence of the coefficients of even degrees and the coefficients of odd degrees. The resulting normal equation matrix is displayed in the upper left part of Fig. 6.1. The coefficients are ordered by order, within each order by degree, with \( c_{nm} \) and \( s_{nm} \) alternating. The orthogonality between different orders, between sine and cosine, and between even and odd degrees becomes evident. The blocks along the main diagonal show the dependencies within the same order, the lack of further blocks indicates the orthogonality among different orders. Reordering the matrix by combining the sine and cosine terms and within each of those the even and odd degrees, leads to the well-known block-diagonal structure. The orthogonality relations and the resulting normal equation matrices for different numbering schemes in case of the geographical grid can be found in Schuh (1996).

**Reuter Grid**  Basically, the Reuter grid shows the same regularities and symmetries as described for the geographical grid. Along each circle of latitude, the spacing between grid points is again equi-angular, and the co-latitudes are located symmetrically with respect to the equator. Therefore, in principle, the same discrete orthogonality are valid, as is the case for the geographical grid. This becomes evident in the upper right part of Fig. 6.1 by the blocks along the main diagonal exhibiting equal structure. Their magnitude is larger due to the declining point density towards the poles. The normal equation matrix of the Reuter grid does, however, possess additional correlation blocks with non-zero elements apart from those described above. They are caused by the circles of latitude near the poles for which the number of points \( \gamma_i \) calculated by Eq. (3.103) becomes very small. The resulting under-sampling of the trigonometric functions yields correlations between distinctive orders \( m \) and \( \bar{m} \). As to the quadrature weights, the surface elements calculated from the Voronoi cells, as described in Section 3.5.2.1, can be applied. But due to the approximately uniform distribution, the use of these weights influences the quadrature results significantly less than, for example, in case of the geographical grid.
Triangle Vertex and Triangle Center Grid  For the two different triangle grids, the normal equation matrices look very much alike. Therefore, only the one for the grid triangle vertex is displayed in the middle left part of Fig. 6.1. But the conclusions are valid for the triangle center grid as well. Even though the grids are not explicitly constructed based on equal-angular spacing between grid points, quite similar structures can be observed compared to the Reuter grid. The original icosahedron features a somehow regular and symmetrical structure. And obviously, certain symmetries and regular distributions enabling discrete orthogonality are preserved during the densification. Again, the surface elements calculated from the Voronoi cells can be used as quadrature weights.

Quasi Random Grid  In contrast to the other point distributions described in this chapter, the quasi random grid does not feature any discrete orthogonalities among different spherical harmonic basis functions, as illustrated by an almost fully occupied normal equation matrix in the middle right part of Fig. 6.1. The grid points are neither located regularly along parallels, nor are they arranged symmetrically to the equator. The only exception is the order \( m = 0 \), as here the orthogonality between odd and even degrees can again be observed. This can be explained by the fact that when the dependency on the longitude \( \lambda \) is omitted (as is the case for the spherical harmonics of order zero), the quasi random sequence positions the grid points symmetrically to the equator. Again, the surface elements calculated from the Voronoi cells can be used as quadrature weights.

6.2.1.5 Practical Calculations and Aliasing

Weight Matrix  The quadrature weights, as determined in case of the Gauss-Legendre and the Driscoll-Healy quadrature, are chosen according to the requirement of realizing discrete orthogonality of the spherical harmonic basis functions. The weights can be combined in the weight matrix \( W \), which results in the normal equation matrix (6.11) becoming the identity matrix. This is only the case, however, for this particular choice of \( W \). Introducing an additional weight matrix, e.g. accounting for the stochasticity of the data, will destroy the orthogonality. So far, the patching process has been performed without error propagation from the regional to the global solutions. Therefore, no additional weight matrix is introduced. If the error propagation will be included in the process, this aspect will have to be taken into account.

Aliasing  An aliasing effect occurs if a signal is undersampled, i.e. if the number of sampling points is not sufficient to completely reconstruct the original signal. The result is that the signal at higher frequencies become indistinguishable from that at certain lower frequencies (they are said to become 'aliases' of each other); thus they distort or create signal at lower frequencies. The aliasing effect is well-known from signal analysis, see, for example, SMITH (1997). In order to investigate the corresponding effect in the case of spherical harmonics on the sphere, it proves to be reasonable to again investigate the behavior in longitudinal and latitudinal direction separately. In the following, the different effects are described, and each of them is illustrated by an example in Fig. 6.2. In the context of calculating spherical harmonic coefficients from gravity field functionals sampled at discrete data points, the aliasing problem has to be kept in mind. The grid (and thus the sampling) used for the (exact) quadrature methods described above has to be chosen sufficiently dense to avoid the effects that will be described in more detail below.

As pointed out in Section 6.2.1.3, the point setting of a Gauss-grid with \( L \) parallels allows the exact quadrature in latitudinal direction (i.e. along the meridians) up to a spherical harmonic degree of \( N = L - 1 \) using the Gauss-Legendre quadrature method. The required inner products of two associated Legendre functions of degree \( n \) and \( \bar{n} \), respectively, are accurate in the discrete case as long as \( n + \bar{n} < 2L \). As soon as the original signal contains frequencies higher than \( N \), an aliasing error occurs. If \( n = L + k \), then the inner products for \( \bar{n} \geq L - k \) cannot be calculated correctly. The upper part of Fig. 6.2 illustrates this effect. Exemplarily, it shows the errors in spherical harmonic coefficients calculated from an original signal expanded up to degree \( N = 80 \). The quadrature was performed on a grid with \( L = 60 \) circles of latitude, which would have allowed the exact determination of coefficients up to degree \( N = 59 \). As \( n \) in this scenario, can reach a degree up to \( n = L + 20 \), the coefficients of degree higher than \( \bar{n} = L - 20 = 40 \) are affected due to the
Figure 6.1: Normal equation matrices for the calculation of spherical harmonic coefficients up to degree $N = 10$ from gridded data.
undersampling. In contrast to Fourier analysis, there is no one-to-one correspondence between a distinctive high frequency component being mapped onto a single lower frequency (as will be described below in case of the longitudinal direction). On the contrary, the undersampled function appears as a linear combination of the lower frequency functions up to degree \( \bar{n} \geq L - k \). Only in case of \( n + \bar{n} \) odd, the odd symmetry characteristics of the Legendre functions provides for the inner product to become correctly zero.

The quadrature in longitudinal direction corresponds to the sampling of trigonometric functions; therefore aliasing phenomena are equal to those arising in the case of Fourier analysis. In this context, the so-called sampling theorem declares that in order to reconstruct a signal exactly, the sampling frequency \( f_s \) has to be more than twice the highest frequency \( f_{\text{max}} \) present in the original signals,

\[
f_s > 2f_{\text{max}}.
\] (6.26)

The frequency representing half of the sampling rate is denoted as Nyquist frequency, and all frequencies smaller than the Nyquist frequency can be reconstructed from the present sampling. As soon as higher frequencies are present, the aliasing effects occur. In case of the Gauss grid, there are \( 2L \) sampling points along each circle of latitude, thus the Nyquist frequency equals a spherical harmonic order of \( m = L \), and orders up to \( m = L - 1 \) can be reconstructed correctly. Concerning the aliasing phenomena, it can be differentiated between reflective aliasing and periodic aliasing. The term 'reflective aliasing' specifies the fact that a frequency of order \( m = L + k \) corrupts the coefficients of order \( L - k \), as can be understood by considering

\[
\cos \left( (L + k) \frac{2\pi}{2L} \right) = \cos \left( (2L + (k - L)) \frac{2\pi}{2L} \right) \\
= \cos(j \frac{2\pi}{2L}) \cos \left( (k - L) \frac{2\pi}{2L} \right) - \sin(j \frac{2\pi}{2L}) \sin \left( (k - L) \frac{2\pi}{2L} \right) = \cos \left( (L - k) \frac{2\pi}{2L} \right) \tag{6.27}
\]

for \( j = 0 \ldots 2L \). Equivalent considerations can be made concerning the sine functions resulting in the relation

\[
\sin \left( (L + k) \frac{2\pi}{2L} \right) = - \sin \left( (L - k) \frac{2\pi}{2L} \right). \tag{6.28}
\]

This is illustrated in the middle part of Fig. 6.2 for a point setting of \( L = 30 \) parallels enabling the reconstruction up to a maximum order of \( m = 29 \). The original signal that was to be reconstructed only consisted of the coefficients \( c_{80,38} \) and \( s_{80,38} \), all the other coefficients being set to zero. The results confirm that order \( m = L + 8 = 38 \) is mapped onto order \( \bar{m} = L - 8 = 22 \). The fact that the single coefficients influence coefficients of different degrees again demonstrates the aliasing effect in latitudinal direction, as described above. Apart from the reflective aliasing, the periodic character of the trigonometric functions leads to periodic aliasing as well. This can be understood from

\[
\exp \left( i(2L + k) \frac{2\pi}{2L} \right) = \exp(i2\pi) \exp \left( ik \frac{2\pi}{2L} \right) = \exp \left( ik \frac{2\pi}{2L} \right). \tag{6.29}
\]

This shows that the functions \( \exp(i(2L + k)\lambda) \) are indistinguishable from the function \( \exp(i\lambda) \), thus coefficients of order \( m = 2L + k \) cause errors in the coefficients of order \( m = k \). This effect is displayed in the lower part of Fig. 6.2, again for a setting with \( L = 30 \). This time all coefficients apart from \( c_{80,68} \) and \( s_{80,68} \) were set equal to zero. The periodic aliasing can be observed by the order \( m = 68 = 2L + k = 60 + 8k \) causing corruption of the order \( \bar{m} = 8 \). At the same time order \( \bar{m} = 2L - k = 52 \) is corrupted as well, as the reflective aliasing described above also occurs for multiples of \( L \).
Figure 6.2: Aliasing effects caused by the undersampling of spherical harmonic functions. Upper part: aliasing in latitudinal direction, middle part: reflective aliasing in longitudinal direction, lower part: periodic aliasing in longitudinal direction.

Aliasing of the $P_{n,m}, L = 60, N = 80$

Reflective aliasing caused by $c_{90,38}$ and $s_{90,38}, L = 30$

Periodic aliasing caused by $c_{90,68}$ and $s_{90,68}, L = 30$