Preprint of the authors:
Schuh, W.-D.; Brockamnn, J.M. (2018): The numerical treatment of covariance stationary processes in least squares collocation.
In: Freeden, (Ed.): *Handbuch der Geodäsie, Vol. 6 "Mathematical Geodesy", Springer Berlin Heidelberg*,
doi:10.1007/978-3-662-46900-2_95-1 (in print)

The numerical treatment of covariance stationary processes in least squares collocation

Wolf-Dieter Schuh, Jan Martin Brockmann
Institute of Geodesy and Geoinformation, University Bonn
53115 Bonn, Nussallee 17
schuh@geod.uni-bonn.de, brockmann@geod.uni-bonn.de

Abstract: Digital sensors provide long series of equispaced and often strongly correlated measurements. A rigorous treatment of this huge set of correlated measurements in a collocation approach is a big challenge. Standard procedures – applied in a thoughtless brute force approach – fail because these techniques are not suitable to handle such huge systems.

In this article two different strategies, denoted as covariance approach and filter approach, to handle such huge systems are contrasted. In the covariance approach various decorrelation strategies based on different Cholesky approaches to factorize the variance/covariance matrices are reviewed. The focus is on arbitrary distributed data sets with with a finite number of data. But also extensions to sparse systems resulting from finite covariance functions and on exploiting the Toeplitz structure which results in the case of equispaced systems are elaborated.

Apart from that filter approaches are discussed to perform a prewhitening strategy for the data and rearrange the whole model to work with this filtered data in a rigorous way. Here, the special focus is on autoregressive processes to model the correlations. Finite, causal, non-recursive filters are constructed as prewhitening filters for the data as well as the model. This approach is extreme efficient, but can only deal with infinite equispaced data sets.

In real data scenarios, finite sequences and data gaps must be handled as well. For the covariance approach this is straightforward but it is a serious problem for the filter approach. Therefore a combination of these approaches is constructed to select the best properties from each. Covariance matrices of equispaced data sets designed by recusively defined covariance seqeunces are represented by AR processes as well as by Cholesky factorized matrices. It is shown, that it is possible to switch between both strategies to get data gaps and the warm up phase for the filter approach under control.

Key Words:
Stochastic processes, collocation, decorrelation, filtering, equispaced data, finite data series, data gaps, Yule-Walker equations, Magic Square, recursive defined covariance sequences

Zusammenfassung: Digitale Sensoren liefern Zeitreihen von gleichabständigen und oft stark korrelierten Messungen. Eine strenge Auswertung, dieser zumeist umfangreichen Datensätze, in einem Kollokationsansatz stellt eine große Herausforderung dar. Standardverfahren sind nicht in der Lage solch große Systeme zu bewältigen.

In diesem Artikel werden zwei unterschiedliche Verfahren — der Kovarianzansatz und der Filteransatz — einander gegenübergestellt, um deren Potential für den Einsatz bei sehr großen Datenmengen zu diskutieren. Im Kovarianzansatz erfolgt die Modellierung der Korrelationen durch Kovarianzfunktionen. Die Dekorrelation der Messungen erfolgt über die Faktorisierung der Kovarianzmatrizen, wofür unterschiedliche Varianten der Cholesky Faktorisierung untersucht werden. Dünn besetzte Systeme hervorgerufen durch den Einsatz von finiten Kovarianzfunktionen, regelmäßige Toeplitz Systeme resultierend aus der regelmäßigen Abtastung, aber auch die Auswirkungen von lokalen Datenverlusten (Datenlöcher) werden untersucht und maßgeschneiderte Zerlegungsverfahren entwickelt.

Der Filteransatz bietet hingegen kaum Möglichkeiten flexibel auf Eigenheiten der Daten einzugehen. Für regelmäßige Abtastung von sehr langen Zeitreihen, ist die Dekorrelation über Filter extrem effizient. Durch die Modellierung der Korrelationen durch finite, autoregressive Prozesse (AR-Prozesse) werden kausale, nichtrekursive, finite Filter aufgebaut, die unendlich ausgedehnte Meßreihen effizient dekorrelieren können. Datenlöcher, aber auch die Initialisierung des Filterprozesses bewirken aber einen Datenverlust, der speziell bei großen Korrelationslängen durchaus dramatisch sein kann.

Messserien aus der Praxis bestehen aber aus endlich vielen Messungen, enthalten Unregelmäßigkeiten und Datenlöcher. Während der Kovarianzansatz ohne Probleme auf diese Daten anwendbar ist, sind beim effizienten Filteransatz spezielle Vorkehrungen (Näherungen) notwendig. Durch den Einsatz von rekursiv definierten Kovarianzsequenzen kann jedoch ein Übergang zwischen den beiden Ansätzen hergestellt werden. Die vollbesetzte Kovarianzmatrix kann mit Hilfe einer speziellen Variante der Cholesky Inversion — rekursives rückwärts Rändern — in eine dünnbesetzte Matrix zerlegt werden, die weitgehend eine Band-Toeplitz-Struktur aufweist, die dem kausalen, nichtrekursiven Filter im Filteransatz entspricht. Durch die Kombination der beiden Ansätze ist es somit möglich einen zeitvariablen, kausalen, nicht rekursiven Filter zu entwickeln, der endliche und leicht unregelmäßige Datensätze streng und effizient dekorrelieren kann.

Schlüsselwörter: Stochastische Prozesse, Kollokation, Dekorrelation, Filterung, gleichabständige Daten, finite Zeitreihen, Datenlöcher, Yule-Walker Gleichungen, Magisches Quadrat, rekursiv definierte Kovarianzsequenzen

Contents

1 Introduction

Digital sensors provide long series of equispaced measurements, which are often strongly correlated because of the high sampling rate and the characteristics of the measurement system. A rigorous treatment of this huge set of correlated measurements in an adjustment model or collocation approach is a big challenge. Simplified models or approximative solutions are often introduced to overcome the numerical effort e.g. short arc approaches (Mayer-Gürr, 2006; Schall et al., 2014), stochastic impulses (Beutler et al., 2006; Jäggi, 2007) or a periodic behavior of the correlations are assumed to get access to Fourier-techniques in the spectral domain (Bottoni und Barzaghi, 1993; Sansò und Tscherning, 2003). But for a rigorous treatment tailored solution strategies in connection with high performance computing are essential to overcome the huge numerical effort (Schuh, 1996b).

Therefore it is necessary to analyze the situation in detail. The measurement model in the collocation approach (Krarup, 1969; Moritz, 1973, 1980) consists of a deterministic part and a stochastic part which can further be divided into a signal part and a noise part. Whereas the signal part is characterized by a covariance stationary stochastic process, the noise part is typically a simple white noise process with uncorrelated components.

Because of the finite length of the data streams, the treatment of the correlations by covariance matrices seems to be preferable. Also data gaps can be easily accounted for in this representation. In addition, a sparse representation resulting from the use of finite covariance functions (Sansò und Schuh, 1987; Schuh, 1989; Gaspari und Cohn, 1999; Moreaux, 2008; Koch et al., 2010; Schuh et al., 2014; Schuh, 2016) opens the possibility to treat huge matrices. But, a large drawback is the deduction of the shape and the variability of the covariance function from the discrete data samples. Within the computation of the empirical covariance function and its approximation by an analytical, positive definite function it is not easy to consider and resolve secondary effects (Priestley, 2004, Sec. 5.3). Though, each covariance stationary process can be represented in the time/space domain as well as in the spectral domain. The four quantities, (a) the stochastic process itself, (b) the autocovariance sequence, (c) the spectral representation of the process, and (d) the power spectrum as spectral representation of the covariance function, form the corners of a square (*Magic Square*) and the relations between them are indicated by arrows. With respect to sensor measurements, equispaced, time-discrete, non-periodic stochastic processes are of special interest. The quantities and relations in the *Magic Square* can be found in Fig. 1 (Krasbutter et al., 2015).

Using the spectral representation of the covariance function, a large variety of different strategies can be applied for the estimation process. Depending on the data characteristics (bias, finite sequences) different strategies are discussed in Kay und Marple (1981) (see also Box und Jenkins (1970, Chap. 7), Priestley (2004, Sec. 5.3 and 5.4)). Furthermore, robust estimation techniques are extensively treated in Kleiner et al. (1979). For the robust estimation of the covariance function, the description of a stochastic process by an autoregressive model is of special interest. In that case, it suggests itself to use directly the coefficients of the modeled autoregressive process (AR-process)

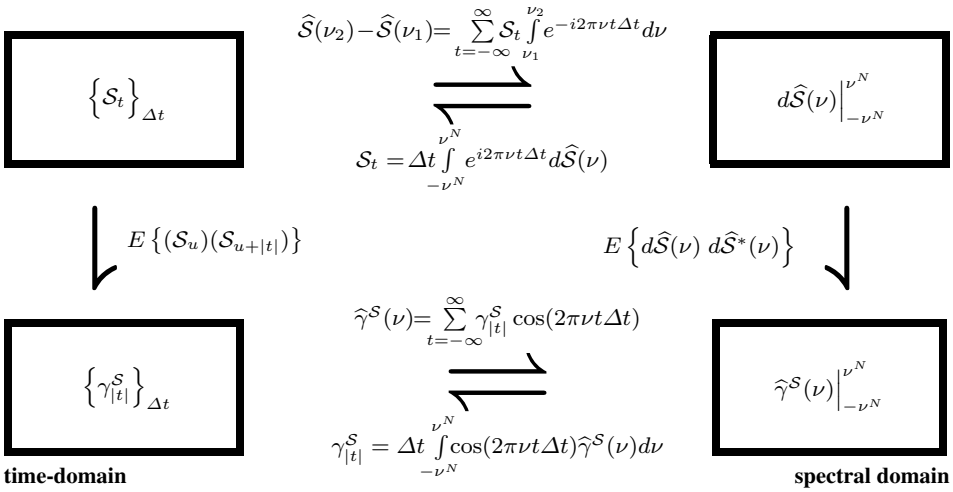


Fig. 1: *Magic Square of a covariance stationary, equispaced, infinite, real valued stochastic process. Upper left: Sequence of stochastic values with a sampling interval of Δt. Upper right: The spectral representation of the equispaced stochastic process forms a continuous, piecewise defined, complex valued, stochastic process $d\widehat{\mathcal{S}}(\nu)$, which is periodic with respect to the interval $[-\nu^N, \nu^N]$ (Krasbutter et al., 2015). ν^N represents the Nyquist frequency, which is fixed by $\nu^N = \frac{1}{2\Delta t}$. The transformation between time and spectral domain is performed by a stochastic Fourier transformThe back transformation is defined by a stochastic Fourier integral Lower left: real valued and even covariance sequence. Lower right: spectral density function $\widehat{\gamma}^{\mathcal{S}}(\nu)$, real valued and periodic in the interval $[-\nu^N, \nu^N]$. The discrete form of the Wiener-Chintschin theorem (Buttkus, 2000, Sect. 8.2.2) defines the transformation between the covariance sequence in the time-domain and the power spectrum. Because of the real valued and even covariance sequence, the transformation from the time domain to the spectral domain can be done by a discrete cosine transformation, while the inverse transformation is performed by a continuous cosine transformation.*

as a prewhitening filter to decorrelate the data, instead of the covariance approach. At a first glance this approach has a lot of benefits, as it performs efficient on huge equispaced data sets. But it has also two important drawbacks with respect to real finite measurement sequences: a filter warmup phase is necessary to handle the first data of the sequence and no data gaps are allowed (Schuh, 2003). In this paper we elaborate an approach, where the best properties of both strategies are combined.

The paper is organized as follows. In Sec. 2 the basics of the Least Squares Collocation (LSC) approach are summarized. Special focus in Sec. 3 is on the covariance approach and on strategies for the decorrelation of arbitrarily structured covariance matrices by Cholesky factorization. Different variations of the Cholesky factorization are used to design causal and anti-causal, recursive and non-recursive filters. In Sec. 4 the signal in the collocation approach is considered as an autoregressive process of a finite order p, shortly denoted as AR(p) process. The pros and cons of this ansatz are discussed in detail. In Sec. 5 the transition from the AR(p) process to the covariance approach is discussed. Recursive defined covariance functions for equispaced data are introduced. Their representations in the time and the spectral domain are elaborated and presented in compact form in the *Magic Square*. In Sec. 6 it is shown, that it is possible to switch between the covariance and the filter approach. In this way, data gaps and the warm up phases can be treated in the filter approach by a time variable AR processes. This findings are discussed in the resumé (Sec. 7). The Appendix outlines different strategies for the Cholesky factorization. Besides the common factorization in a lower and upper triangular matrix also the backward Cholesky factorization is discussed. Ccomputational strategies to compute the inverse of a system directly from the Cholesky factorized triangular matrices are elaborated. A flexible combination of all the possible factorization steps for a symmetric matrix as well as for their inverse is discussed and serves a basis for the integration of the covariance approach into the filter approach.

2 Least Squares Collocation

Following Moritz (1980, Sec. 14) (see also Schuh (2016, Sec. 3)) the measurement model for the observables $\boldsymbol{\mathcal{L}} \in \mathbb{R}^n$ in least squares collocation is given in the general form by

$$\boldsymbol{\mathcal{L}} = \boldsymbol{f}(\boldsymbol{\xi}) + \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}} \qquad \text{and} \qquad \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}\} + \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{N}}\}, \tag{1}$$

or in the linear form by

$$\boldsymbol{\mathcal{L}} = \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}} \qquad \text{and} \qquad \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}\} + \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{N}}\}, \tag{2}$$

where $\boldsymbol{f}(\boldsymbol{\xi})$ respectively $\boldsymbol{A}\boldsymbol{\xi}$ represents a systematic (deterministic) model part whereas the signal $\boldsymbol{\mathcal{S}}$ and the noise $\boldsymbol{\mathcal{N}}$ represents the stochastic part[1]. In contrast to the general covariance stationary stochastic process $\boldsymbol{\mathcal{S}}$, the noise process $\boldsymbol{\mathcal{N}}$ is characterized by independently and identical distributed random variables with expectation $\boldsymbol{E}\{\boldsymbol{\mathcal{N}}\} = \boldsymbol{0}_n$ and variance $\boldsymbol{\Sigma}\{\boldsymbol{\mathcal{N}}\} = \mathbb{1}_n\sigma^2$ and often shortly denoted as *white noise process*. The random variables $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{N}}$ are independent, $\boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{N}}\} = \mathbb{O}_n$ and therefore

$$\boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}}\} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}\} + \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{N}}\} := \boldsymbol{\Sigma} \tag{3}$$

holds for the stochastic model. The variance/covariance matrix $\boldsymbol{\Sigma}\{\boldsymbol{\mathcal{L}}\}$ will be often simply denoted as covariance matrix and marked by $\boldsymbol{\Sigma}$ in the following. The best linear unbiased estimator (BLUE) $\widetilde{\boldsymbol{\mathcal{X}}}$ for the true unknown functional parameters $\boldsymbol{\xi}$ is given by

$$\widetilde{\boldsymbol{\mathcal{X}}} = \left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{L}} \tag{4}$$

with the variance/covariance matrix

$$\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{X}}}\} = \left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}. \tag{5}$$

The best linear unbiased predictor (BLUP) $\widetilde{\boldsymbol{\mathcal{S}}}$ for the signal component is given by

$$\widetilde{\boldsymbol{\mathcal{S}}} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{S}}+\boldsymbol{\mathcal{N}}\}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\Pi}^{\boldsymbol{\Sigma}^{-1}}_{S^\perp(\boldsymbol{A})}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{L}}\,. \tag{6}$$

where $\boldsymbol{\Pi}^{\boldsymbol{\Sigma}^{-1}}_{S^\perp(\boldsymbol{A})}$ denotes the projector into the orthogonal column space $S^\perp(\boldsymbol{A})$ of the matrix $\boldsymbol{A}$ with metric $\boldsymbol{\Sigma}^{-1}$, defined as

$$\boldsymbol{\Pi}^{\boldsymbol{\Sigma}^{-1}}_{S^\perp(\boldsymbol{A})} := \boldsymbol{\Sigma} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T. \tag{7}$$

Introducing the reduced observation vector

$$\widetilde{\Delta\boldsymbol{\mathcal{L}}} := \boldsymbol{\Pi}^{\boldsymbol{\Sigma}^{-1}}_{S^\perp(\boldsymbol{A})}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mathcal{L}} \tag{8}$$

we can rewrite (6) as

$$\widetilde{\boldsymbol{\mathcal{S}}} = \boldsymbol{\Sigma}\{\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{S}}+\boldsymbol{\mathcal{N}}\}\,\boldsymbol{\Sigma}^{-1}\widetilde{\Delta\boldsymbol{\mathcal{L}}}\,. \tag{9}$$

The variance of this estimator can be derived by variance propagation. Considering that

$$\boldsymbol{\Sigma}\{\widetilde{\Delta\boldsymbol{\mathcal{L}}}\} = \boldsymbol{\Pi}^{\boldsymbol{\Sigma}^{-1}}_{S^\perp(\boldsymbol{A})}\,, \tag{10}$$

[1] In this contribution random variables are denoted by calligraphic letters $\mathcal{L}$ and random vectors by $\boldsymbol{\mathcal{L}}$. Greek letters denote true values ξ and vectors of true values $\boldsymbol{\xi}$, whereas Latin letters represent fixed numbers or realizations a, vectors $\boldsymbol{a}$ and matrices $\boldsymbol{A}$ of numbers and realizations. $\mathbb{1}_n$ stands for the unity matrix with dimension $n \times n$ and $\mathbb{O}_n$ for a nullmatrix of dimension $n \times n$, $\mathbb{O}_{n_1 \times n_2}$ for a nullmatrix with n_1 rows and n_2 columns, whereas $\boldsymbol{0}_n$ denotes a null vector with n elements.

and taking into account the idempotence of the projector with respect to the metric $\boldsymbol{\Sigma}^{-1}$ Schuh (2016, Sec. 3.1), it follows that

$$\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{S}}}\} = \boldsymbol{\Sigma}\left\{\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{S}}+\boldsymbol{\mathcal{N}}\right\} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}\{\widetilde{\Delta\boldsymbol{\mathcal{L}}}\} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}\left\{\boldsymbol{\mathcal{S}}+\boldsymbol{\mathcal{N}}, \boldsymbol{\mathcal{S}}\right\} . \tag{11}$$

The variance of the estimation error $\boldsymbol{\mathcal{E}}_{\widetilde{\boldsymbol{\mathcal{S}}}} := \widetilde{\boldsymbol{\mathcal{S}}} - \boldsymbol{\mathcal{S}}$ can be computed by (cf. Moritz (1980, Sec. 14), Schuh (2016, Sec. 3))

$$\boldsymbol{\Sigma}\{\boldsymbol{\mathcal{E}}_{\widetilde{\boldsymbol{\mathcal{S}}}}\} = \boldsymbol{\Sigma}\left\{\boldsymbol{\mathcal{S}}\right\} - \boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{S}}}\} . \tag{12}$$

3 Covariance approach

In this paper we have a focus on the computation of the unknown parameters (4) and their covariance matrix (7). A strategy which is often used when covariances have to be considered is to decorrelate the observations in a pre-processing step. Therefore, the observation equations (2) are multiplied with a regular matrix $\boldsymbol{H} \in \mathbb{R}^{n\times n}$

$$\underbrace{\boldsymbol{H}\boldsymbol{\mathcal{L}}}_{:=\overline{\boldsymbol{\mathcal{L}}}} = \underbrace{\boldsymbol{H}\boldsymbol{A}}_{:=\overline{\boldsymbol{A}}}\boldsymbol{\xi} + \underbrace{\boldsymbol{H}\boldsymbol{\mathcal{S}}}_{:=\overline{\boldsymbol{\mathcal{S}}}} + \underbrace{\boldsymbol{H}\boldsymbol{\mathcal{N}}}_{:=\overline{\boldsymbol{\mathcal{N}}}} \tag{13}$$

with the objective, that the transformed observations $\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{H}\boldsymbol{\mathcal{L}}$ are uncorrelated and homogenized. That means for the propagated variance of $\overline{\boldsymbol{\mathcal{L}}}$

$$\boldsymbol{\Sigma}\left\{\overline{\boldsymbol{\mathcal{L}}}\right\} = \boldsymbol{\Sigma}\left\{\boldsymbol{H}\boldsymbol{\mathcal{L}}\right\} = \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^T \stackrel{!}{=} \mathbb{1}_n \tag{14}$$

must hold. This yields an alternate formulation for (2)

$$\overline{\boldsymbol{\mathcal{L}}} = \overline{\boldsymbol{A}}\boldsymbol{\xi} + \overline{\boldsymbol{\mathcal{S}}} + \overline{\boldsymbol{\mathcal{N}}} \qquad \text{with} \qquad \boldsymbol{\Sigma}\left\{\overline{\boldsymbol{\mathcal{L}}}\right\} = \boldsymbol{\Sigma}\left\{\overline{\boldsymbol{\mathcal{S}}} + \overline{\boldsymbol{\mathcal{N}}}\right\} = \mathbb{1}_n . \tag{15}$$

To find an adequate representation for $\boldsymbol{H}$ which satisfies (14), it is a straightforward step to factorize the variance/covariance matrix in two parts, where one part is the transposed of the other. This can be done e.g. by an eigenvalue decomposition $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^T$, where the columns of the matrix $\boldsymbol{U}$ represent the eigenvectors and the square roots of the corresponding eigenvalues are arranged in the diagonal of the diagonal matrix $\boldsymbol{\Lambda}^{1/2}$. A computational more efficient way is to factorize the covariance matrix $\boldsymbol{\Sigma}$ by the Cholesky (forward factorization) approach into a lower and upper triangular matrix, where the lower triangular matrix is the transposed of the upper triangular matrix $\boldsymbol{R}$,

$$\boldsymbol{\Sigma} = \boldsymbol{R}^T\boldsymbol{R} . \tag{16}$$

To satisfy condition (14),

$$\boldsymbol{H}\,\boldsymbol{R}^T\boldsymbol{R}\,\boldsymbol{H}^T \stackrel{!}{=} \mathbb{1}_n \tag{17}$$

must hold. This is clearly fulfilled for $\boldsymbol{H}^T = \boldsymbol{R}^{-1}$. Therefore the transformed quantities $\overline{\boldsymbol{\mathcal{L}}}$ and $\overline{\boldsymbol{A}}$ can be computed by

$$\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{H}\boldsymbol{\mathcal{L}} = \left(\boldsymbol{R}^T\right)^{-1}\boldsymbol{\mathcal{L}} \qquad \text{resp.} \qquad \overline{\boldsymbol{A}} = \boldsymbol{H}\boldsymbol{A} = \left(\boldsymbol{R}^T\right)^{-1}\boldsymbol{A} \tag{18}$$

which corresponds to the solutions of the systems of equations

$$\boldsymbol{R}^T\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{L}} \qquad \text{resp.} \qquad \boldsymbol{R}^T\overline{\boldsymbol{A}} = \boldsymbol{A} . \tag{19}$$

In an element-wise formulation, the solution can be determined by a forward substitution algorithm

$$\begin{bmatrix} r_{11} & & & \\ r_{12} & r_{22} & & \\ \vdots & \vdots & \ddots & \\ r_{13} & r_{23} & \dots & r_{nn} \end{bmatrix} \begin{bmatrix} \overline{\mathcal{L}}_1 \\ \overline{\mathcal{L}}_2 \\ \vdots \\ \overline{\mathcal{L}}_n \end{bmatrix} = \begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \\ \vdots \\ \mathcal{L}_n \end{bmatrix} \qquad \text{resp.} \qquad \overline{\mathcal{L}}_i = \frac{1}{r_{ii}}\Big(\mathcal{L}_i - \sum_{k=1}^{i-1} r_{ki}\overline{\mathcal{L}}_k\Big) \tag{20}$$

(cf. Appendix A.1). If we look to (20) as a filter equation, we identify this filter as a *linear, time variant, recursive, causal filter*.

This filter has some benefits because it can handle finite time series, is flexible with respect to data distribution (data gaps) and can handle sparse structures resulting from for instance finite covariance functions. But, this filter approach has also some cons, as this recursive procedures are time-consuming and its parallel usage is not straightforward. Time variant filters have a memory requiement of $\mathcal{O}\{n^2\}$ and have a computational complexity of $\mathcal{O}\{n^3\}$.

3.1 Variations of Cholesky decorrelation filters

The versatility of the Cholesky approach as documented in the Appendix can be used to construct different types of filters, also non-recursive (moving average) and anti-causal filters can be constructed. In a first approach an anti-causal, recursive filter can be constructed by the partition of the matrix $\boldsymbol{\Sigma}$ into an upper triangular matrix and its transposed,

$$\boldsymbol{\Sigma} = \overline{\boldsymbol{R}}\,\overline{\boldsymbol{R}}^T \tag{21}$$

Numerically this partitioning can be done by backward Cholesky factorization (83) as elaborated in Appendix A.4. The corresponding transformation matrix $\boldsymbol{H}$ for the decorrelation process $\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{H}\boldsymbol{\mathcal{L}}$ follows from the condition

$$\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^T = \boldsymbol{H}\overline{\boldsymbol{R}}\,\overline{\boldsymbol{R}}^T\boldsymbol{H}^T \stackrel{!}{=} \mathbb{1}_n \tag{22}$$

and yields $\boldsymbol{H} = \overline{\boldsymbol{R}}^{-1}$. To avoid the inversion of the matrix $\overline{\boldsymbol{R}}$, the decorrelation process in form of a matrix-vector multiplication

$$\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{H}\boldsymbol{\mathcal{L}} = \overline{\boldsymbol{R}}^{-1}\boldsymbol{\mathcal{L}} \tag{23}$$

is rearranged, such that the solution of

$$\overline{\boldsymbol{R}}\,\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{L}}. \tag{24}$$

has to be determined. For the solution, the upper triangular form of $\overline{\boldsymbol{R}}$ can be exploited starting form the last equation by a backward substitution. The decorrelation process for $\overline{\boldsymbol{\mathcal{L}}}$ resp. $\overline{\boldsymbol{A}}$ with respect to the factorization (21) can be seen as an *anti-causal, recursive filter*.

Both filters (20) and (24) discussed so far are recursive filters. However, non-recursive filters are preferable because of their better performance with respect to runtime and their straightforward implementation on parallel machines. In a first attempt it is assumed that the inverse $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix $\boldsymbol{\Sigma}$ is known. The inverse matrix $\boldsymbol{\Sigma}^{-1}$ is now factorized by the common Cholesky approach into a lower and upper triangular matrix. The name of this matrix can be chosen arbitrarily but according to the discussion in the Appendix, especially with respect to Fig. 10(c) we denote this matrix by $\overline{\boldsymbol{R}}^{-1}$. The Cholesky factorization of the inverse $\boldsymbol{\Sigma}^{-1}$ is given by

$$\boldsymbol{\Sigma}^{-1} = \big(\overline{\boldsymbol{R}}^{-1}\big)^T\overline{\boldsymbol{R}}^{-1}\,. \tag{25}$$

Considering the decorrelation condition (14) with respect to the inverse covariance matrix

$$\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H} = \boldsymbol{H}\big(\boldsymbol{\Sigma}^{-1}\big)^{-1}\boldsymbol{H}^T = \boldsymbol{H}\Big(\big(\overline{\boldsymbol{R}}^{-1}\big)^T\overline{\boldsymbol{R}}^{-1}\Big)^{-1}\boldsymbol{H}^T = \boldsymbol{H}\overline{\boldsymbol{R}}\,\overline{\boldsymbol{R}}^T\boldsymbol{H}^T \stackrel{!}{=} \mathbb{1}_n \tag{26}$$

it follows that the transformation matrix is $\boldsymbol{H} = \overline{\boldsymbol{R}}^{-1}$. It is not very efficient and numerically stable to firstly compute the inverse $\boldsymbol{\Sigma}^{-1}$ and then secondly perform the Cholesky factorization $\overline{\boldsymbol{R}}^{-1}$. In Fig. 10(c) it is shown, that the computation of $\overline{\boldsymbol{R}}^{-1}$ can be done starting from $\boldsymbol{\Sigma}$ by (79) organized by *recursive backward edging* strategy cf. (80), beginning with the last element. With the computed matrix $\overline{\boldsymbol{R}}^{-1}$ the decorrelation process can be performed by

$$\overline{\boldsymbol{\mathcal{L}}} = \overline{\boldsymbol{R}}^{-1}\boldsymbol{\mathcal{L}} \tag{27}$$

which then corresponds to an *anti-causal, non-recursive filter*.

To construct a *causal, non-recursive filter*, the inverse of the covariance matrix must be factorized by a *backward Cholesky factorization* (cf. Fig. 10(d)) into

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{R}^{-1}\left(\boldsymbol{R}^{-1}\right)^T . \tag{28}$$

The corresponding decorrelation matrix $\boldsymbol{H} = \left(\boldsymbol{R}^T\right)^{-1}$ results from the decorrelation condition (14) and can be computed by (87) by a *recursive forward edging* strategy cf. (80) starting with the first element. The decorrelation process is then given by

$$\overline{\boldsymbol{\mathcal{L}}} = \left(\boldsymbol{R}^T\right)^{-1}\boldsymbol{\mathcal{L}}. \tag{29}$$

Now, it corresponds to a *causal, non-recursive filter*, which is simple to apply from the numerical point of view and benefits from a sequential data availability.

3.2 Resumé about Cholesky decorrelation filters

In the last section various types of Cholesky factorization strategies are recapped. The focus of this consideration is not on the efficient computation of these factorizations, but on the achievable decorrelation results in terms of different recursive and non-recursive, causal and anti-causal filters.

Decomposition	Factorization	Decorrelation matrix	Decorrelation of $\boldsymbol{\mathcal{L}}$
$\boldsymbol{\Sigma} = \boldsymbol{R}^T\boldsymbol{R}$	forward Cholesky reduction compute $\boldsymbol{R}$ from $\boldsymbol{\Sigma}$ (63)	$\boldsymbol{H} = \left(\boldsymbol{R}^T\right)^{-1}$	$\overline{\boldsymbol{\mathcal{L}}} = \left(\boldsymbol{R}^T\right)^{-1}\boldsymbol{\mathcal{L}} \Longleftrightarrow \boldsymbol{R}^T\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{L}}$ $\begin{bmatrix} * & & \\ * & \boldsymbol{R}^T & \\ * & * & * \end{bmatrix}\begin{bmatrix} \\ \overline{\boldsymbol{\mathcal{L}}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{\mathcal{L}} \\ \\ \end{bmatrix}$ causal, recursive filter
$\boldsymbol{\Sigma} = \overline{\boldsymbol{R}}\,\overline{\boldsymbol{R}}^T$	backward Cholesky reduction compute $\overline{\boldsymbol{R}}$ from $\boldsymbol{\Sigma}$ (82)	$\boldsymbol{H} = \overline{\boldsymbol{R}}^{-1}$	$\overline{\boldsymbol{\mathcal{L}}} = \overline{\boldsymbol{R}}^{-1}\boldsymbol{\mathcal{L}} \Longleftrightarrow \overline{\boldsymbol{R}}\,\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{\mathcal{L}}$ $\begin{bmatrix} * & * & * \\ & \overline{\boldsymbol{R}} & * \\ & & * \end{bmatrix}\begin{bmatrix} \\ \overline{\boldsymbol{\mathcal{L}}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{\mathcal{L}} \\ \\ \end{bmatrix}$ anti-causal, recursive filter
$\boldsymbol{\Sigma} = \left((\overline{\boldsymbol{R}}^{-1})^T\overline{\boldsymbol{R}}^{-1}\right)^{-1}$	recursive backward edging Compute $\overline{\boldsymbol{R}}^{-1}$ from $\boldsymbol{\Sigma}$ (79)	$\boldsymbol{H} = \overline{\boldsymbol{R}}^{-1}$	$\overline{\boldsymbol{\mathcal{L}}} = \overline{\boldsymbol{R}}^{-1}\boldsymbol{\mathcal{L}}$ $\begin{bmatrix} \\ \overline{\boldsymbol{\mathcal{L}}} \\ \\ \end{bmatrix} = \begin{bmatrix} * & * & * \\ & \overline{\boldsymbol{R}}^{-1} & * \\ & & * \end{bmatrix}\begin{bmatrix} \\ \boldsymbol{\mathcal{L}} \\ \\ \end{bmatrix}$ anti-causal, non-recursive filter
$\boldsymbol{\Sigma} = \left(\boldsymbol{R}^{-1}(\boldsymbol{R}^{-1})^T\right)^{-1}$	recursive forward edging Compute $\boldsymbol{R}^{-1}$ from $\boldsymbol{\Sigma}$ (87)	$\boldsymbol{H} = \left(\boldsymbol{R}^T\right)^{-1}$	$\overline{\boldsymbol{\mathcal{L}}} = \left(\boldsymbol{R}^T\right)^{-1}\boldsymbol{\mathcal{L}}$ $\begin{bmatrix} \\ \overline{\boldsymbol{\mathcal{L}}} \\ \\ \end{bmatrix} = \begin{bmatrix} * & & \\ * & \left(\boldsymbol{R}^T\right)^{-1} & \\ * & * & * \end{bmatrix}\begin{bmatrix} \\ \boldsymbol{\mathcal{L}} \\ \\ \end{bmatrix}$ causal, non-recursive filter

Fig. 2: *Decorrelation strategies by various types of Cholesky factorizations.*

Fig. 2 illustrates the different possibilities to factorize the covariance matrix by the Cholesky approach. Summarizing the pros and cons it can be stated that the covariance approach in connection with the different Cholesky factorization techniques is extremely flexible, can handle arbitrarily distributed data sets, but is computational very demanding. Huge data sets can only be handled, when finite covariance functions are applied and the sparsity of the matrix can be utilized and preserved in the different factorizations. Especially for equispaced data the filter approach, which is introduced in the upcoming section, seems preferably.

4 Least Squares Collocation - Filter Approach

In contrast to the general collocation approach in the last section, we will now have a look on equispaced data sets. For that purpose, we define the observation equations in a row-wise sequential formulation as

$$\mathcal{L}_t = \boldsymbol{A}_t\boldsymbol{\xi} + \mathcal{S}_t + \mathcal{N}_t \qquad t \in \mathbb{Z}, \tag{30}$$

where the random variable $\mathcal{L}_t$ describes the measurement process in time $T = T_0 + t\Delta t$. $\boldsymbol{A}_t$ denotes the t^{th}-row of the design matrix $\boldsymbol{A}$ and $\boldsymbol{\xi}$ denotes the unknown true parameters. The random variables $\mathcal{S}_t$ and $\mathcal{N}_t$ are associated with stochastic processes $\{\mathcal{S}_t\}_{\Delta t}$ and $\{\mathcal{N}_t\}_{\Delta t}$, which can be represented by an auto-regressive parameter model in the parameters $\alpha_i, i = 1, \ldots, p$ (AR(p) process)

$$\mathcal{S}_t = \alpha_1\left(\mathcal{S}_{t-1} + \mathcal{N}_{t-1}\right) + \alpha_2\left(\mathcal{S}_{t-2} + \mathcal{N}_{t-2}\right) + \ldots + \alpha_p\left(\mathcal{S}_{t-p} + \mathcal{N}_{t-p}\right) + \mathcal{E}_t. \tag{31}$$

$\mathcal{E}_t$ and $\mathcal{N}_t$ are member of two infinite white noise processes $\{\mathcal{E}_t\}_{\Delta t}$ and $\{\mathcal{N}_t\}_{\Delta t}$ characterized by independent and identically distributed random variables with expectation $E\{\mathcal{E}_t\} = 0$, $E\{\mathcal{N}_t\} = 0$, and variances $\sigma^2_{\mathcal{E}}$ resp. $\sigma^2_{\mathcal{N}}$,

$$\{\mathcal{E}_t\}_{\Delta t} \sim WN\left(\boldsymbol{0}, \mathbb{1}_{\infty}\sigma^2_{\mathcal{E}}\right); \quad \{\mathcal{N}_t\}_{\Delta t} \sim WN\left(\boldsymbol{0}, \mathbb{1}_{\infty}\sigma^2_{\mathcal{N}}\right); \quad \Sigma\{\mathcal{E}_t, \mathcal{N}_k\} = 0; \qquad t, k \in \mathbb{Z}\,. \tag{32}$$

The goal is to find the best linear unbiased estimator $\widetilde{\boldsymbol{\mathcal{X}}}$ cf. (4) for the true unknown parameters $\boldsymbol{\xi}$ with minimal variance $\boldsymbol{\Sigma}\{\widetilde{\boldsymbol{\mathcal{X}}}\}$. We use a whitening strategy and reorder the AR(p)-process (31) with respect to the white noise component $\mathcal{E}_t$

$$\mathcal{S}_t - \alpha_1(\mathcal{S}_{t-1} + \mathcal{N}_{t-1}) - \alpha_2(\mathcal{S}_{t-2} + \mathcal{N}_{t-2}) - \ldots - \alpha_p(\mathcal{S}_{t-p} + \mathcal{N}_{t-p}) = \mathcal{E}_t \tag{33}$$

and expand both sides with $\mathcal{N}_t$ and re-write the equation for $t, t-1, t-2, \ldots$ in matrix-vector form

$$\underbrace{\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & & \\ -\alpha_p & \ldots & -\alpha_2 & -\alpha_1 & 1 & & \\ & -\alpha_p & \ldots & -\alpha_2 & -\alpha_1 & 1 & \\ & & -\alpha_p & \ldots & -\alpha_2 & -\alpha_1 & 1 \\ & & & \ddots & & \ddots & \ddots & \ddots \end{bmatrix}}_{\boldsymbol{H}} \underbrace{\begin{bmatrix} \vdots \\ \mathcal{S}_{t-2} + \mathcal{N}_{t-2} \\ \mathcal{S}_{t-1} + \mathcal{N}_{t-1} \\ \mathcal{S}_t + \mathcal{N}_t \\ \vdots \end{bmatrix}}_{\boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}}} = \underbrace{\begin{bmatrix} \vdots \\ \mathcal{E}_{t-2} + \mathcal{N}_{t-2} \\ \mathcal{E}_{t-1} + \mathcal{N}_{t-1} \\ \mathcal{E}_t + \mathcal{N}_t \\ \vdots \end{bmatrix}}_{\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{N}}} \tag{34}$$

or

$$\boldsymbol{H}\big(\boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}}\big) = \boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{N}} \tag{35}$$

which can be interpreted as a linear filter (cf. Fig. 3).

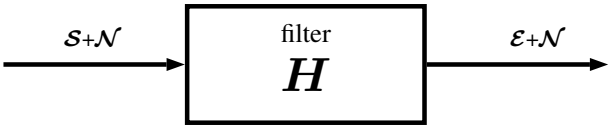


Fig. 3: *Pre-whitening filter*

This pre-whitening filter is characterized by the following attributes, it is linear, time invariant, non-recursive and causal. It can be applied to the observation equations (2) by multiplication from the left side with the filter matrix $\boldsymbol{H}$. Afterwards, they read

$$\overline{\boldsymbol{\mathcal{L}}} = \overline{\boldsymbol{A}}\boldsymbol{\xi} + \boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{N}} \qquad \text{with} \quad \boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{N}} \sim WN\big(\mathbf{0}, \mathbb{1}_\infty(\sigma^2_{\boldsymbol{\mathcal{E}}} + \sigma^2_{\boldsymbol{\mathcal{N}}})\big)\,, \tag{36}$$

where $\overline{\boldsymbol{\mathcal{L}}} = \boldsymbol{H}\boldsymbol{\mathcal{L}}$ and $\overline{\boldsymbol{A}} = \boldsymbol{H}\boldsymbol{A}$ denote the filtered measurements and the column-wise filtered design matrix. The colored noise $\boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{N}}$ is transformed to white noise $\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{N}}$ by applying the filter matrix as shown in (35). The parameters $\alpha_1, \ldots \alpha_p$ of the AR(p)-process can be derived in an iterative procedure from the residuals (cf. Box und Jenkins (1970); Kay und Marple (1981); Priestley (2004); Krasbutter et al. (2014)). The pre-whitening filter approach has a lot of pros like the flexibility in filter design and the possibility to cascade filters, the causality of the filter allows a sequential adjustment approach with a computational complexity of $\mathcal{O}\{n\}$, which allows for an efficient computation in cache-conform algorithms and can be easily tailored for parallel computations with a scalability $\mathcal{O}\{\text{cores}\}$. Despite the pros, also some cons can be identified because the filter can only deal with infinite, equispaced, uninterrupted (gapless) and outlier-free data sets. These needs are not realistic, when considering real measurement time series. These results in the open questions:

- How to initialize the filter process? Often, filter warmup phases are proposed. This warm up automatically results in a loss of data (Schuh, 2003), which should be avoided especially for short time series.
- How to deal with data gaps and outliers? Should data gaps be filled with interpolated or approximated data? For instance (Siemes, 2008, Sec. 4.3) proposed a sophisticated approach, where besides the deterministic part also the stochastic part is estimated for the gap filling values.

All these points can be well treated by the covariance approach (cf. Sec. 3) and need no special attention. But, as mentioned above the covariance approach has the disadvantage of the loss of computational efficiency. Therefore, it is reasonable to propose a strategy for the combination of these two approaches.

5 Magic Square — Transition from an AR(p)-process to a covariance function

At the end of the last section the pro and cons of the covariance approach as well as of the filter approach are stated. In this section we seek for a combination of both approaches, which benefits from the pros of both strategies. Therefore it is necessary to find the connection between the covariance sequence and the AR(p) coefficients.

The autocovariance sequence $\left\{\gamma^{\mathcal{Z}}_{|t|}\right\}_{\Delta t}$ of a general time discrete covariance stationary process $\left\{\mathcal{Z}_t\right\}_{\Delta t}$ is defined via

$$\gamma^{\mathcal{Z}}_{|k|} := E\left\{\left(\mathcal{Z}_t - E\left\{\mathcal{Z}_t\right\}\right)\left(\mathcal{Z}_{t-k} - E\left\{\mathcal{Z}_{t-k}\right\}\right)\right\}; \qquad k \in \mathbb{Z}\,. \tag{37}$$

Introducing the AR(p)-process $\left\{\mathcal{S}_t\right\}_{\Delta t}$ given by (31) connected with the assumption of white noise processes $\mathcal{E}_t$ and $\mathcal{N}_t$ it can be stated that $E\left\{\mathcal{S}_t\right\}=0$. Therefore the autocovariance sequence (37) coincides with the autocorrelation sequence, i.e.

$$\gamma^{\mathcal{S}}_{|k|} = E\left\{\mathcal{S}_t\mathcal{S}_{t-k}\right\}; \qquad k \in \mathbb{Z}\,. \tag{38}$$

For simplicity, but without restriction on the universality, only the signal part $\mathcal{S}_t$ and not the combined signal plus noise part $\mathcal{S}_t+\mathcal{N}_t$ is considered in the following. The extension for the noise part is straightforward. The autocovariance $\gamma^{\mathcal{S}}_{|k|}$ for the interval k with respect to the AR(p) process (31) is obtained from

$$\gamma^{\mathcal{S}}_{|k|} = \alpha_1 E\left\{\mathcal{S}_{t-1}\mathcal{S}_{t-k}\right\} + \alpha_2 E\left\{\mathcal{S}_{t-2}\mathcal{S}_{t-k}\right\} + \ldots + \alpha_p E\left\{\mathcal{S}_{t-p}\mathcal{S}_{t-k}\right\} + E\left\{\mathcal{E}_t\mathcal{S}_{t-k}\right\}; \quad k \in \mathbb{Z}\,. \tag{39}$$

For the evaluation of this equation for different k the following relations have to be considered. It can be stated that preceding signals $\mathcal{S}_{t-k}$, $k>0$ and the noise $\mathcal{E}_t$ at stage t are independent and therefore $E\left\{\mathcal{S}_{t-k}\mathcal{E}_t\right\} = 0$. Furthermore the signal $\mathcal{S}_t$ at stage t includes the noise $\mathcal{E}_t$ which yields to $E\left\{\mathcal{E}_t\mathcal{S}_t\right\}=\sigma^2_{\mathcal{E}}$. Considering covariance stationarity, the covariances depend only on the time lag k and not on the stage t, consequently

$$E\left\{\mathcal{S}_{t-1}\mathcal{S}_{t-k}\right\} = E\left\{\mathcal{S}_t\mathcal{S}_{t-(k-1)}\right\} = \gamma^{\mathcal{S}}_{|k-1|}\,. \tag{40}$$

Evaluating (39) for different values k yields

$$\begin{array}{lllllll}
k=0: & \gamma^{\mathcal{S}}_0 & = \alpha_1\gamma^{\mathcal{S}}_1 & + \alpha_2\gamma^{\mathcal{S}}_2 & + \ldots + \alpha_p\gamma^{\mathcal{S}}_p & + \sigma^2_{\mathcal{E}} \\
k=1: & \gamma^{\mathcal{S}}_1 & = \alpha_1\gamma^{\mathcal{S}}_0 & + \alpha_2\gamma^{\mathcal{S}}_1 & + \ldots + \alpha_p\gamma^{\mathcal{S}}_{p-1} & \\
k=2: & \gamma^{\mathcal{S}}_2 & = \alpha_1\gamma^{\mathcal{S}}_1 & + \alpha_2\gamma^{\mathcal{S}}_0 & + \ldots + \alpha_p\gamma^{\mathcal{S}}_{p-2} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \\
k=p: & \gamma^{\mathcal{S}}_p & = \alpha_1\gamma^{\mathcal{S}}_{p-1} & + \alpha_2\gamma^{\mathcal{S}}_{p-2} & + \ldots + \alpha_p\gamma^{\mathcal{S}}_0 & \\
\hline
k=p+1: & \gamma^{\mathcal{S}}_{p+1} & = \alpha_1\gamma^{\mathcal{S}}_p & + \alpha_2\gamma^{\mathcal{S}}_{p-1} & + \ldots + \alpha_p\gamma^{\mathcal{S}}_1 & \\
k=p+2: & \gamma^{\mathcal{S}}_{p+2} & = \alpha_1\gamma^{\mathcal{S}}_{p+1} & + \alpha_2\gamma^{\mathcal{S}}_p & + \ldots + \alpha_p\gamma^{\mathcal{S}}_2 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & .
\end{array} \tag{41}$$

This equations are known as *Yule-Walker equations* (Box und Jenkins, 1970, Sec. 3.2) and provide the connection between the AR(p) coefficients $\alpha_j, j=1,\ldots,p$ and the covariance sequence $\left\{\gamma^{\mathcal{S}}_{|t|}\right\}_{\Delta t}, t\in\mathbb{Z}$. The Yule-Walker equations can be divided into two parts. In the upper part only the covariances $\gamma^{\mathcal{S}}_0$ to $\gamma^{\mathcal{S}}_p$ are involved, whereas in the lower part each additional equation adds a new autocovariance factor $\gamma^{\mathcal{S}}_{p+i}$ to the system. It should be noted that the covariance sequence is recursively defined and is therefore denoted as *recursive defined covariance sequence*. The sequence of covariances uniquely defined by the first $p+1$ covariances, $\gamma^{\mathcal{S}}_0,\ldots,\gamma^{\mathcal{S}}_p$.

Rewriting the Yule-Walker equations (41) to matrix-vector notation helps for a better understanding,

$$\begin{bmatrix} \gamma_0^{\mathcal{S}} \\ \gamma_1^{\mathcal{S}} \\ \gamma_2^{\mathcal{S}} \\ \vdots \\ \gamma_p^{\mathcal{S}} \\ \hline \gamma_{p+1}^{\mathcal{S}} \\ \gamma_{p+2}^{\mathcal{S}} \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma_1^{\mathcal{S}} & \gamma_2^{\mathcal{S}} & \cdots & \gamma_p^{\mathcal{S}} \\ \gamma_0^{\mathcal{S}} & \gamma_1^{\mathcal{S}} & \cdots & \gamma_{p-1}^{\mathcal{S}} \\ \gamma_1^{\mathcal{S}} & \gamma_0^{\mathcal{S}} & \cdots & \gamma_{p-2}^{\mathcal{S}} \\ \vdots & \vdots & & \vdots \\ \gamma_{p-1}^{\mathcal{S}} & \gamma_{p-2}^{\mathcal{S}} & \cdots & \gamma_0^{\mathcal{S}} \\ \hline \gamma_p^{\mathcal{S}} & \gamma_{p-1}^{\mathcal{S}} & \cdots & \gamma_1^{\mathcal{S}} \\ \gamma_{p+1}^{\mathcal{S}} & \gamma_p^{\mathcal{S}} & \cdots & \gamma_2^{\mathcal{S}} \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ \vdots \end{bmatrix} \sigma_{\mathcal{E}}^2 \,. \tag{42}$$

Reordering the first $p+1$ equations (upper part) with respect to the autocovariance sequence $\gamma_0^{\mathcal{S}}, \gamma_1^{\mathcal{S}}, \ldots \gamma_p^{\mathcal{S}}$, it follows

$$\begin{array}{rrrrrrrrrrrrr} (-1)\gamma_0^{\mathcal{S}} + & & \alpha_1\gamma_1^{\mathcal{S}} + & & \alpha_2\gamma_2^{\mathcal{S}} + \ldots + & \alpha_{p-1}\gamma_{p-1}^{\mathcal{S}} + & \alpha_p\gamma_p^{\mathcal{S}} & = & -\sigma_{\mathcal{E}}^2 \\ \alpha_1\gamma_0^{\mathcal{S}} + & (-1+\alpha_2)\gamma_1^{\mathcal{S}} + & & & \alpha_3\gamma_2^{\mathcal{S}} + \ldots + & \alpha_p\gamma_{p-1}^{\mathcal{S}} + & 0\gamma_p^{\mathcal{S}} & = & 0 \\ \alpha_2\gamma_0^{\mathcal{S}} + & (\alpha_1+\alpha_3)\gamma_1^{\mathcal{S}} + & (-1+\alpha_4)\gamma_2^{\mathcal{S}} + & & \ldots + & 0\gamma_{p-1}^{\mathcal{S}} + & 0\gamma_p^{\mathcal{S}} & = & 0 \\ \vdots & & \vdots & & & \vdots & \vdots & & \\ \alpha_p\gamma_0^{\mathcal{S}} + & \alpha_{p-1}\gamma_1^{\mathcal{S}} + & \alpha_{p-2}\gamma_2^{\mathcal{S}} + & & \ldots + & \alpha_1\gamma_{p-1}^{\mathcal{S}} + & (-1)\gamma_p^{\mathcal{S}} & = & 0 \end{array} \,. \tag{43}$$

This can be expressed in matrix-vector notation by

$$\left(\begin{bmatrix} -1 & & & & \\ \alpha_1 & -1 & & & \\ \alpha_2 & \alpha_1 & -1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \alpha_{p-1} & \alpha_{p-2} & \alpha_{p-3} & \ldots & -1 \\ \alpha_p & \alpha_{p-1} & \alpha_{p-2} & \ldots & \alpha_1 \quad -1 \end{bmatrix} + \begin{bmatrix} 0 & \alpha_1 & \ldots & \alpha_{p-2} & \alpha_{p-1} & \alpha_p \\ 0 & \alpha_2 & \ldots & \alpha_{p-1} & \alpha_p & \\ 0 & \alpha_3 & \ldots & \alpha_p & & \\ \vdots & \vdots & \iddots & & & \\ 0 & \alpha_p & & & & \\ 0 & & & & & \end{bmatrix} \right) \begin{bmatrix} \gamma_0^{\mathcal{S}} \\ \gamma_1^{\mathcal{S}} \\ \gamma_2^{\mathcal{S}} \\ \vdots \\ \gamma_{p-1}^{\mathcal{S}} \\ \gamma_p^{\mathcal{S}} \end{bmatrix} = \begin{bmatrix} -\sigma_{\mathcal{E}}^2 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \,. \tag{44}$$

Together with the equations of the lower part of (41)

$$\begin{array}{rrrrrrr} \gamma_{p+1}^{\mathcal{S}} = & \alpha_1\gamma_p^{\mathcal{S}} & + \alpha_2\gamma_{p-1}^{\mathcal{S}} & + \ldots + & \alpha_p\gamma_1^{\mathcal{S}} \\ \gamma_{p+2}^{\mathcal{S}} = & \alpha_1\gamma_{p+1}^{\mathcal{S}} & + \alpha_2\gamma_p^{\mathcal{S}} & + \ldots + & \alpha_p\gamma_2^{\mathcal{S}} \\ \vdots & \vdots & \vdots & & \vdots \end{array} \,, \tag{45}$$

these equations form the *reorganized Yule-Walker equations* (Schuh et al., 2014; Schuh, 2016). The covariance sequence $\left\{\gamma_{|t|}^{\mathcal{S}}\right\}_{\Delta t}, t \in \mathbb{Z}$ can be deduced via the solution of (44) directly from the coefficients $\alpha_j, j = 1, \ldots, p$ of the AR(p)-process and the recursion formulas (45). And vice versa, the coefficients of the AR(p)-process $\alpha_j, j = 1, \ldots, p$, can be derived from the solution of the equations upper part of the equation system (42) from the covariance sequence $\gamma_1^{\mathcal{S}}, \ldots, \gamma_p^{\mathcal{S}}$. The error variance $\sigma_{\mathcal{E}}^2$ results from the first row of (42). The Yule-Walker equations and its reorganized representation allow for direct transformation between the autocovariance sequence and the coefficients of an AR(p)-process. It should be noted, that the stochastic process defined by an AR(p)-process (31) yields to a special type of covariance sequence, which is denoted as a *recursive defined covariance sequence*.

To find the representation of the AR(p)-process in the spectral domain, we reformulate the AR(p)-process as a moving average process. With the reordered infinite sequence of deterministic coefficients

$$\left\{\overline{\alpha}_t\right\}_{\Delta t} := \left\{ \ldots 0 \; 1 - \alpha_1 \; -\alpha_2 \ldots - \alpha_p \, 0 \ldots \right\} \tag{46}$$

a convolution of the sequence with the sequence of the stochastic signal $\left\{\mathcal{S}_t\right\}_{\Delta t}$ results in a white noise sequence $\left\{\mathcal{E}_t\right\}_{\Delta t}$,

$$\left\{\overline{\alpha}_t\right\}_{\Delta t} * \left\{\mathcal{S}_t\right\}_{\Delta t} = \left\{\mathcal{E}_t\right\}_{\Delta t}. \tag{47}$$

The representation in the spectral domain is given by

$$H(\nu)\, d\widehat{\mathcal{S}}(\nu) = d\widehat{\mathcal{E}}(\nu) \tag{48}$$

where $H(\nu)$ denotes the transfer function of $\left\{\overline{\alpha}_t\right\}_{\Delta t}$, which is defined by the Fourier transform of the non-periodic infinite sequence $\left\{\overline{\alpha}_t\right\}_{\Delta t}$

$$H(\nu) = \mathcal{F}\left\{\left\{\overline{\alpha}_t\right\}_{\Delta t}\right\} = 1 - \sum_{j=1}^{p} \alpha_j e^{-i2\pi\nu j\Delta t}\,, \tag{49}$$

which is defined in agreement with the definition in Fig. 1. The spectral representation of the signal immediately follows from (48)

$$d\widehat{\mathcal{S}}(\nu) = \frac{d\widehat{\mathcal{E}}(\nu)}{1 - \sum\limits_{j=1}^{p} \alpha_j e^{-i2\pi\nu j\Delta t}}. \tag{50}$$

As the autocovariance sequence in the time domain is represented by the expectation of the correlation of the stochastic process with itself, the correlation in the time domain can be represented by an element-wise multiplication with the conjugate quantities in the spectral domain. Whereas the denominator is deterministic the stochastic part in the numerator is transformed by the expectation operator to $\sigma^2_{\mathcal{E}}$ (Schuh, 2016, Fig. 5).

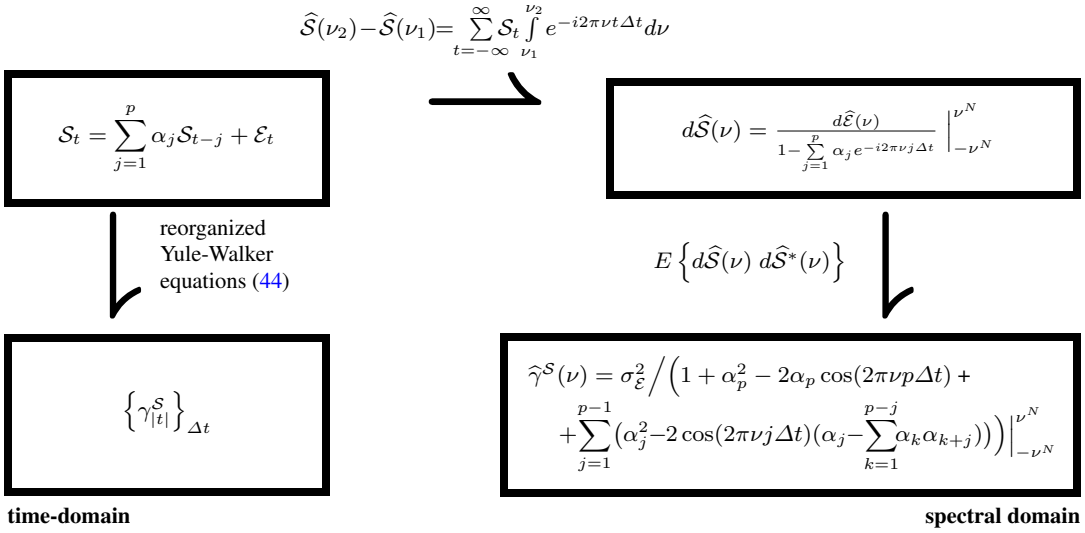


Fig. 4: *Magic Square of a autoregressive process of order p*

With the help of the *Magic Square* representation, these relations can be documented in the time domain as well as in the spectral domain (Schuh et al., 2014; Schuh, 2016). We can find closed analytical relations between all the quantities, if we assume that the AR(p) coefficients $\alpha_j, j = 1, \ldots, p$ are given (cf. Fig 4). Especially the power spectrum can be expressed in a closed form, which is not possible if we use the standard Wiener-Chintschin Theorem for the transformation of the covariance sequence into the power spectrum (cf. Fig. 1).

6 Combined Approach

In the last sections two different approaches to deal with correlated time series in the least squares collocation approach were compared. In the covariance approach, the correlations are expressed by the covariance function and collected in the variance/covariance matrix. The Cholesky approach is used for the decorrelation of the measure-

ments. In contrast to the covariance approach, the filter approach uses an AR(p)-process to describe the correlations in the time series. Applying a causal, non-recursive filter to the measurements (35) as well as to the columns of the design matrix, an alternative problem (36) with uncorrelated measurements is identified.

In the last section it is shown that the Yule-Walker equations provide a direct transition between the covariance sequence and the coefficients of the AR(p)-process. Because of the required equispaced data distribution for the AR(p)-process, the continuous covariance function becomes a recursively defined covariance sequence. Considering covariance stationarity, this covariance matrix of the filter approach for equispaced data has a Toeplitz structure. To be more precise, the covariance matrix is an infinite Toeplitz structured matrix.

In the case of equispaced data, the covariance approach yields to a Toeplitz structured covariance matrix as well. But in contrast to the infinite matrix of the filter approach, the covariance matrix in the covariance approach is restricted to a finite dimension. Let us denote the number of measurements by n, then the variance/covariance matrix is of dimension $n\times n$ and can be seen as a resected diagonal block of the infinite Toeplitz matrix. In contrast to an infinite system, the inverse of a finite Toeplitz matrix is not Toeplitz structured again. Fast $\mathcal{O}(n\log n)$ solutions strategies based on Fourier transformations are only possible for infinite or periodic systems (Bottoni und Barzaghi, 1993), but also for band Toeplitz systems (Schuh, 1996b, App. A.7). For finite Toeplitz systems, Levinsons-Durbin algorithm Levinson (1947); Durbin (1960); Trench (1964) provides a very efficient solution with respect to the number of operations $\mathcal{O}(n^2)$ and required memory $\mathcal{O}(n)$.

In the last section, the duality between the filter and covariance approach was already elaborated. Now, we would like to answer the question, whether it is possible to use this duality and the given relations for an efficient and flexible decorrelation. The filter approach results in a causal, finite, non-recursive filter. At the end of Sec. 3.1, a causal, non-recursive filter was constructed by the Cholesky approach as well (by computing $\left(\boldsymbol{R}^T\right)^{-1}$ by the *recursive forward edging* strategy (84)). We use this analogy, to identify a direct connection between these two strategies.

To find the similarities between these two approaches the Yule-Walker equation (42) is rewritten

$$\begin{bmatrix} \gamma_0^{\mathcal{S}} & \gamma_1^{\mathcal{S}} & \cdots & \gamma_p^{\mathcal{S}} & \gamma_{p+1}^{\mathcal{S}} & \cdots \\ \gamma_1^{\mathcal{S}} & \gamma_0^{\mathcal{S}} & \cdots & \gamma_{p-1}^{\mathcal{S}} & \gamma_p^{\mathcal{S}} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ \gamma_p^{\mathcal{S}} & \gamma_{p-1}^{\mathcal{S}} & \cdots & \gamma_0^{\mathcal{S}} & \gamma_1^{\mathcal{S}} & \cdots \\ \gamma_{p+1}^{\mathcal{S}} & \gamma_p^{\mathcal{S}} & \cdots & \gamma_1^{\mathcal{S}} & \gamma_0^{\mathcal{S}} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha_1 \\ \vdots \\ -\alpha_p \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} \sigma_{\mathcal{E}}^2 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix} \quad . \tag{51}$$

With respect to the covariance approach the covariance matrix $\boldsymbol{\Sigma}$ is factorized into $\left(\boldsymbol{R}^{-1}(\boldsymbol{R}^{-1})^T\right)^{-1}$ (cf. Fig 2), where $\boldsymbol{R}^{-1}$ can be computed by *recursive forward edging* (84). Using the matrix $\boldsymbol{\Sigma}$ instead of $\boldsymbol{N}^{-1}$ and additionally $\boldsymbol{R}^{-1}$ instead of $\bar{\boldsymbol{R}}$ compared to Fig. 10(d) for the backward Cholesky approach, the following equation results

$$\begin{bmatrix} \gamma_0^{\mathcal{S}} & \gamma_1^{\mathcal{S}} & \cdots & \gamma_j^{\mathcal{S}} \\ \gamma_1^{\mathcal{S}} & \gamma_0^{\mathcal{S}} & \cdots & \gamma_{j-1}^{\mathcal{S}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_j^{\mathcal{S}} & \gamma_{j-1}^{\mathcal{S}} & \cdots & \gamma_0^{\mathcal{S}} \end{bmatrix} \begin{bmatrix} r_{jj}^{(-1)} r_{jj}^{(-1)} \\ r_{j-1,j}^{(-1)} r_{jj}^{(-1)} \\ \vdots \\ r_{1j}^{(-1)} r_{jj}^{(-1)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{for} \qquad j = 0, \ldots, n \ . \tag{52}$$

This equation corresponds to the problem to find the first row of the inverse of matrix $\boldsymbol{\Sigma}$(1:j,1:j). This can be done very efficient for large Toeplitz systems by the Levinson-Durbin algorithm (Levinson, 1947; Durbin, 1960), but for smaller systems also a forward edging process corresponding to (80) is suitable. Comparing (52) and (51) we see that for $j \geq p$ the two systems corresponds to each other and the solution of (52) can be found by an element wise comparison with the solution vector in (51)

$$
\begin{aligned}
r_{jj}^{(-1)} &= \frac{1}{\sigma_\mathcal{E}}; \\
r_{j-k,j}^{(-1)} &= -\frac{\alpha_k}{\sigma_\mathcal{E}}; && k = 1,\dots,p \\
r_{k,j}^{(-1)} &= 0; && k = p+1,\dots \quad .
\end{aligned} \tag{53}
$$

It can be seen that the Cholesky factors for the rows $j \geq p$ can be immediately deduced from the coefficients of the AR(p)-process only by a linear scaling with respect to the variance $\sigma_\mathcal{E}$. The same values, shifted by one column, are obtained row-by-row. Thus, the Cholesky factorization of the Toeplitz structured covariance matrix corresponding to the AR(p)-process is a lower band matrix with $p+1$ elements per row and has again Toeplitz structure for the rows $j \geq p$. For the rows $j < p$ the coefficients can be determined by (52) in a rigorous way, because of the finite shape of the covariance matrix. On the left side of Fig. 5 the results for an AR(6) process are depicted. The blue elements mark the Toeplitz structure with the scaled AR(6) coefficients deduced from (53) or strictly speaking from the recursion (second) part of the Yule-Walker equations (42), whereas the yellow and red elements represent the warm-up phase determined by (52).

This matrix can be seen as causal, non-recursive filter with variable filter coefficients for the warm up phase. They are due to the truncation of the infinite sequence to a finite sequence at the beginning of the sequence. Because of the causality, no additional disturbances at the end of the sequence arises. It has to be noted, that the corresponding variance/covariance matrix is a dense matrix, but the factorized form derived by the backward Cholesky approach is a band matrix due to the recursively defined covariance sequence. In the common case the relation is vice versa. If the system is sparse (e.g. because of a finite covariance function), the Cholesky factorized system is also sparse but its inverse is dense. Here, we start with the dense system and compute the Cholesky factorization of the sparse inverse system which covariance function is deduced from a finite AR(p) process.

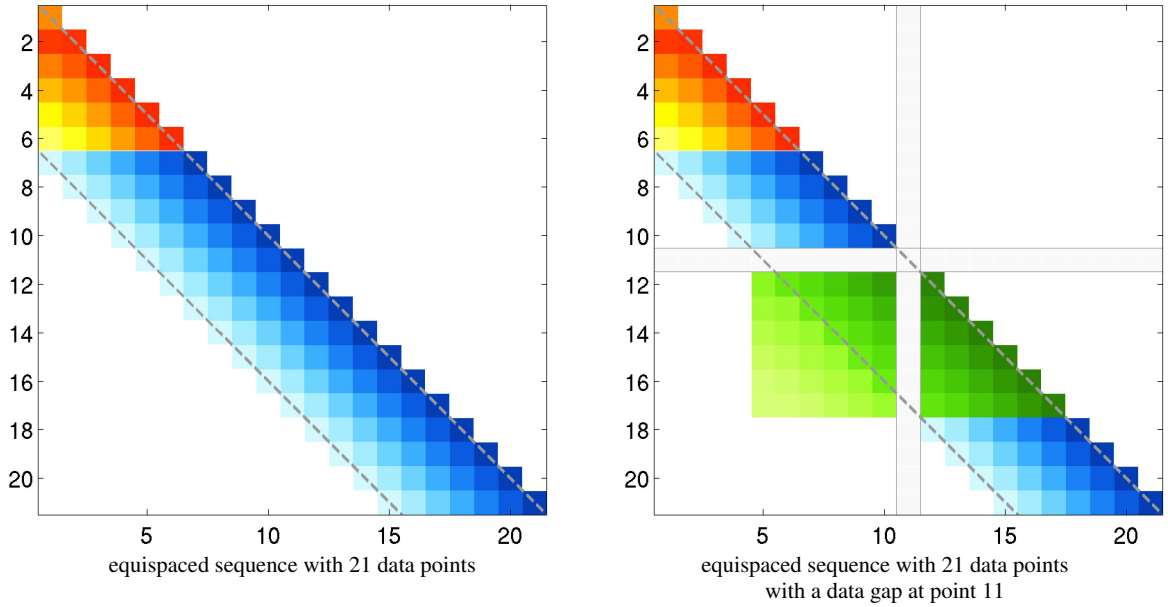


Fig. 5: *Causal, non-recursive filter for a AR(6) process deduced by the backward Cholesky approach by recursive forward edging. The blue elements correspond after a scaling (53) exactly to the AR(p) coefficients. The yellow and red part at the beginning represents the warm-up part. On the left side a finite sequence with 21 equispaced data points is shown, whereas on the right side the data point at position 11 is missing. The green elements mark coefficients which are affected by this data gap.*

Real measurement series are not only characterized by a finite number of measurements but also affected by data gaps. In many practical filter applications an interpolated value is used to preserve the equispaced structure (Siemes, 2008, Sec. 4.3). In contrast to this, the covariance approach has no problems with these data gaps. Only

the corresponding columns and rows from the variance/covariance matrix has to be eliminated to account for the missing data. Using equispaced data the elimination of the rows and columns destroys – or better disturbs – the Toeplitz structure. This is of special interest for the rows $k \geq p$ where the Toeplitz structure is also reflected by the causal filter. A detailed look to the recursion part of the Yule-Walker equations (42) shows, that starting with the data gap the next p recursion equations are disturbed, but afterwards, the Toeplitz structure restores and the AR(p) coefficients fulfill the recursion equations. This means that only p rows are influenced by a data gap. The disturbed coefficients can be computed by the Cholesky approach, starting with a stack of p standard rows, the data gap and p additional rows. This computation can then be inserted into the equispaced filter (matrix). In Fig. 5 on the right side the impact of a data gap at point 11 illustrates this behavior.

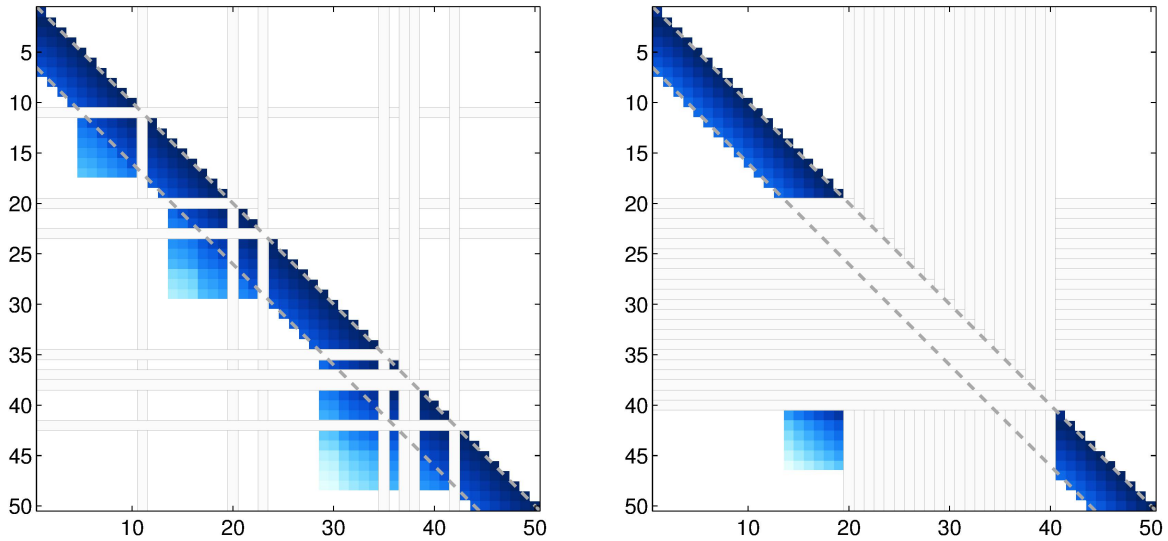


Fig. 6: *Causal, non-recursive filter for a AR(6) process deduced by the backward Cholesky approach by recursive forward edging. On the left side a finite sequence with 50 equispaced data points with missing data at the points 11, 20, 23, 35, 37, 38, and 42 are shown, where as on the right side a continuous data gap from point 20 to point 40 is simulated.*

Fig. 6 shows the situation for several data gaps. Different situations can occur. Isolated gaps as discussed before (here at point 11), but also combined data gaps, where the space between two gaps is smaller than the process order p (here at point 20 and 23) or four gaps (35, 35, 38, and 42) interact with each other (Fig. 6, left). On the right side a continuous data gap between 20 to 40 is shown. All these situations can be fixed by a local Cholesky approach, where corresponding divergent rows can be locally computed in strict manner by the Cholesky approach.

7 Resumé

Huge variance/covariance matrices are the bottlenecks for least squares collocation. In this article different approaches are discussed to overcome this problem. The covariance approach is on the one hand extremely flexible with respect to data distribution, can rigorously handle data sets with a finite number of data and can treat data gaps without any problems. But the numerical effort – especially for huge data sets – is enormous. Efficient numerical methods, high performance computing but also finite covariance functions in connection with sparse solvers are necessary to handle such huge systems. For the solution of the symmetric covariance systems the Cholesky algorithm is a very common and efficient tool. Different decorrelation strategies based on the Cholesky factorization are discussed and a large variety to design causal and anti-causal, recursive and non-recursive filters are shown.

Further strategies are discussed to handle Toeplitz systems for equispaced data sets. But the filter approach, where the correlations in an equispaced time series are represented by an autoregressive process, seems to be unbeatable for this equispaced data distribution. The resulting discrete, causal, non-recursive filters can be efficiently applied. But for real data applications and real measurement series the large drawback of the filter approach is, that this filter approach works only on infinite equispaced data series. Also data gap are not acceptable. This paper shows a strategy to combine both approaches, the combination selects the best properties from both. Considering the Yule-Walker equations a recursively defined covariance sequence can be defined which allows to switch between the covariance approach and the filter approach. Thus an efficient algorithm is deduced which can handle in a rigorous and efficient way finite sequences of equispaced data sets which can include data gaps.

Appendix:

A Cholesky approach - revisited

The Cholesky approach was first published posthumously by Benoit (1924) in *Bulletin Géodésique* and based on an unpublished manuscript from Cholesky (1910). The solution of symmetric linear equation systems is a frequently asked problem in numerical mathematics and data analysis. The efficient and stable solution of the normal equations as well as the factorization and inversion of covariance matrices is an inherent task in least squares adjustment and least squares collocation. Thus, many geodesists use the Cholesky approach and developed a lot of strategies to adopt this method for their special applications (Poder und Tscherning, 1973; Meissl, 1982). Sparse solvers (Snay, 1976; Schuh, 1981) as well as partial inversion strategies (Hanson, 1978; Auzinger und Schuh, 1998) are developed, studies on the numerical stability and the influence of rounding errors in huge systems are performed (Meissl, 1980; Ernst und Schuh, 2012), but also strategies to generate correlated signals for Monte Carlo approaches (Alkhatib und Schuh, 2007) are developed. Usually Cholesky factorization is connected to symmetric positive definite systems, but in Fig. 7 also other possible applications for infinite, rectangular, asymmetric and ill-posed systems are mentioned. These systems are out of scope of this article, but in many situations the Cholesky approach is extremely efficient and numerically stable applicable.

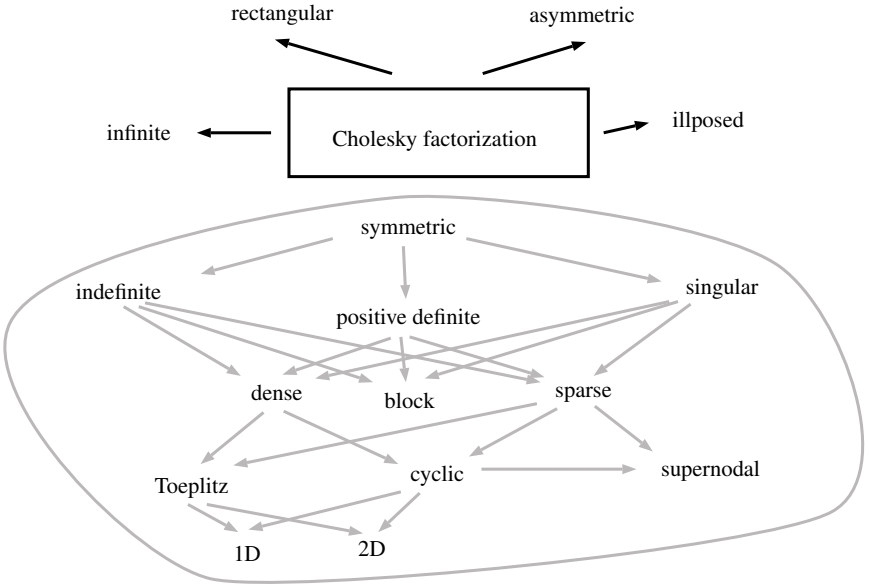


Fig. 7: *Application of the Cholesky approach on different types of systems. This figure is focused on the application on symmetric systems and shows the variety of different strategies only for this special type.*

A.1 Cholesky solution

The Cholesky approach was originally developed for the solution of a symmetric linear equation system by the decomposition into a lower and upper triangular matrix. In contrast to the LU decomposition the diagonal elements of both matrices are chosen equal to eliminate the degree of freedom. For symmetric systems the hermitian matrix $\boldsymbol{N} \in \mathbb{C}^{n\times n}$ is decomposed into triangular matrices, where one is the conjugate transposed of the other. In this work we choose the notation

$$\boldsymbol{N} = \boldsymbol{R}^H \boldsymbol{R}\,, \tag{54}$$

where $\boldsymbol{R}$ denotes a upper (right) triangular matrix and $\boldsymbol{R}^H$ its conjugate transposed (hermitian) form. If $\boldsymbol{N} \in \mathbb{R}^{n\times n}$ is a positive definite, real valued, symmetric matrix the decomposition is real valued, and thus

$$\boldsymbol{N} = \boldsymbol{R}^T \boldsymbol{R}\,, \tag{55}$$

with $\boldsymbol{R} \in \mathbb{R}^{n\times n}$. This approach denoted a *Cholesky forward factorization* is often used for least squares problems, especially for Gauss-Markov models and adjustment models with condition equations. The resulting positive definite normal equations can be decomposed very efficiently (roughly twice as efficient as the LU decomposition) and numerically stable (Meissl, 1980; Ernst und Schuh, 2012) by the Cholesky approach.

The solution $\boldsymbol{x} \in \mathbb{R}^n$ of the linear symmetric system

$$\boldsymbol{N}\boldsymbol{x} = \boldsymbol{n} \tag{56}$$

with the symmetric matrix $\boldsymbol{N} \in \mathbb{R}^{n\times n}$ and the right hand side $\boldsymbol{n} \in \mathbb{R}^n$ can be efficiently determined after the factorization

$$\boldsymbol{R}^T \underbrace{\boldsymbol{R}\boldsymbol{x}}_{:=\,\boldsymbol{z}} = \boldsymbol{n} \tag{57}$$

in a two step approach by solving the lower triangular system

$$\boldsymbol{R}^T \boldsymbol{z} = \boldsymbol{n} \tag{58}$$

for the auxiliary unknowns $\boldsymbol{z} \in \mathbb{R}^n$ in a *forward substitution* step and then solving the upper triangular system

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{z} \tag{59}$$

for the unknowns $\boldsymbol{x}$ in a *backward substitution* step . The formulas for each single coefficient of the forward and backward substitution can be derived immediately from (58)

$$z_i = \left(n_i - \sum_{k=1}^{i-1} r_{ki} z_k \right) / r_{ii}\,, \qquad i = 1, \ldots, n, \tag{60}$$

and (59)

$$x_i = \left(z_i - \sum_{k=i+1}^{n} r_{ik} x_k \right) / r_{ii}\,, \qquad i = n, \ldots, 1 \quad . \tag{61}$$

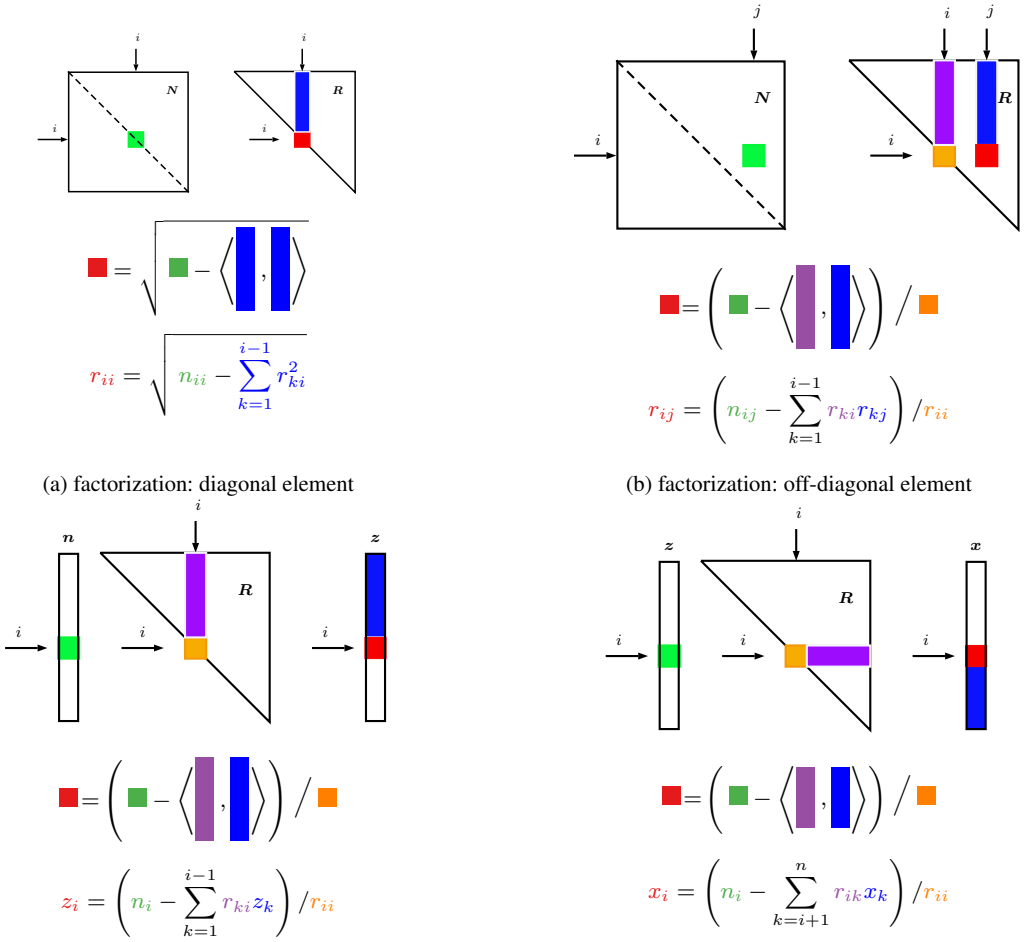
The coefficients of the triangular matrix $\boldsymbol{R}$ can be derived from the identity (55). Written in expanded form

$$\begin{bmatrix} n_{11} & n_{12} & \ldots & n_{1n} \\ n_{12} & n_{22} & \ldots & n_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ n_{13} & n_{23} & \ldots & n_{nn} \end{bmatrix} = \begin{bmatrix} r_{11} & & & \\ r_{12} & r_{22} & & \\ \vdots & \vdots & \ddots & \\ r_{13} & r_{23} & \ldots & r_{nn} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ & r_{22} & \ldots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}, \tag{62}$$

it can be seen immediately that n^2 – respectively $\frac{1}{2}n(n+1)$ linear independent – identities results from this equation. They can be used to derive the Cholesky coefficients by

$$\left.\begin{aligned} r_{ii} &= \sqrt{n_{ii} - \sum_{k=1}^{i-1} r_{ki}^2} \\ r_{ij} &= \left(n_{ij} - \sum_{k=1}^{i-1} r_{ki} r_{kj}\right) / r_{ii}\,, \qquad j = i{+}1, \ldots, n \end{aligned}\right\} \quad i{=}1, \ldots, n\,. \tag{63}$$

It should be mentioned, that the three required loops with respect to i, j and k can be organized in accordance with the used storage scheme of the matrices. These numerical considerations are not in the focus of this article (see for instance Schuh (1996b, p. 138ff)).



(a) factorization: diagonal element

(b) factorization: off-diagonal element

(c) substitution: forward

(d) substitution: backward

Fig. 8: *Graphical interpretation of the factorization steps* (63)*, the forward substitution step* (60) *and the backward substitution step* (61)*.*

Fig. 8 summarizes the four steps which are required to determine a solution of a linear symmetric system with the Cholesky approach. It should be mentioned, that the factorization step for off-diagonal elements exactly corresponds to the forward substitution step. The first three steps use a column access only, whereas for the backward substitution a row and a column access is proposed.

These figures provide a feeling, what happens during the factorization of sparse input matrices. If an element n_{ij} is initially zero, it is unchanged during factorization, if the scalar product of the column vectors r_{ki} and r_{kj} vanishes. This is the case especially for envelope systems, where the column above a zero element is zero as well. Therefore the envelope structure of a sparse symmetric system is preserved during Cholesky factorization (George und Liu, 1981; Meissl, 1982). Special numbering schemes are developed (Snay, 1976; Schuh, 1981, 1996b; Boxhammer und Schuh, 2006) to determine the order of the parameters in the matrix, to optimize the structure to preserve sparsity also during the factorization step. The design of covariance matrices by finite covariance functions (Sansò und Schuh, 1987; Gaspari und Cohn, 1999; Koch et al., 2010; Schuh, 2016) yields to sparse systems, which can efficiently solved by the Cholesky approach.

Fig. 8 can also be used to get an impression of the case of an indefinite system. In the factorization step of the diagonal element, the value under the square-root becomes negative. Therefore it is necessary to introduce the imaginary unit i. In case this happens, the divisor in all off-diagonal factorization steps is imaginary, therefore the entire row consists of imaginary values. The main operation during the whole factorization and substitution steps are scalar products between two columns. If a row contains imaginary values in both column elements, the result is not complex anymore, but real valued. The product of these two elements is not added during the computation of the scalar product, but must be subtracted. Therefore, indefinite systems can be solved easily by the Cholesky approach by introducing an index vector to indicate which rows involve imaginary numbers. As the normal equations of the commonly used Gauss-Markov model, or the covariance matrices are positive definite systems, this extension of the Cholesky approach is rarely used in geodesy. But the normal equations of the Gauss-Helmert model and also the Gauss-Markov model with restrictions are indefinite systems and the above mentioned extension is crucial to take benefit from the stable and efficient Cholesky approach also for indefinite systems.

A.2 Cholesky inversion

In this section we are looking for the inverse of the matrix $\boldsymbol{N}$, where we assume, that the Cholesky factorized matrices $\boldsymbol{R}$ are already computed. A common used strategy is to invert the triangular matrix $\boldsymbol{R}$ and apply

$$\boldsymbol{N}^{-1} = \left(\boldsymbol{R}^T\boldsymbol{R}\right)^{-1} = \boldsymbol{R}^{-1}\left(\boldsymbol{R}^T\right)^{-1}. \tag{64}$$

Here we want to elaborate a direct strategy to switch from $\boldsymbol{R}$ to the inverse matrix $\boldsymbol{N}^{-1}$.

The positive definite matrix $\boldsymbol{N}$ and its inverse $\boldsymbol{N}^{-1}$ can be written in block notation as

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ \boldsymbol{N}_{12}^T & \boldsymbol{N}_{22} \end{bmatrix}, \qquad \boldsymbol{N}^{-1} = \begin{bmatrix} \boldsymbol{N}_{11}^{(-1)} & \boldsymbol{N}_{12}^{(-1)} \\ (\boldsymbol{N}_{12}^{(-1)})^T & \boldsymbol{N}_{22}^{(-1)} \end{bmatrix}, \tag{65}$$

where $\boldsymbol{N}_{11}^{(-1)}$ denotes the upper left block in the inverse $\boldsymbol{N}^{-1}$. It has to be carefully distinguished from the inverse of the upper left block of the matrix $\boldsymbol{N}$, which is denoted by $\boldsymbol{N}_{11}^{-1}$. The partitioning of the matrix into blocks is performed in a way, that $\boldsymbol{N}_{11} \in \mathbb{R}^{n_1 \times n_1}$ and $\boldsymbol{N}_{22} \in \mathbb{R}^{n_2 \times n_2}$ are quadratic. The identity

$$\boldsymbol{N}^{-1}\boldsymbol{N} = \mathbb{1}_n \qquad \begin{bmatrix} \boldsymbol{N}_{11}^{(-1)} & \boldsymbol{N}_{12}^{(-1)} \\ (\boldsymbol{N}_{12}^{(-1)})^T & \boldsymbol{N}_{22}^{(-1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ \boldsymbol{N}_{12}^T & \boldsymbol{N}_{22} \end{bmatrix} = \begin{bmatrix} \mathbb{1}_{n_1} & \mathbb{O}_{n_1 \times n_2} \\ \mathbb{O}_{n_2 \times n_1} & \mathbb{1}_{n_2} \end{bmatrix}, \tag{66}$$

can be used to find a representation of the blocks of the inverse matrix. They read

$$\boldsymbol{N}_{22}^{(-1)} = \left(\boldsymbol{N}_{22} - \boldsymbol{N}_{12}^T\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\right)^{-1}, \tag{67}$$

$$\boldsymbol{N}_{12}^{(-1)} = -\boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\boldsymbol{N}_{22}^{(-1)}, \tag{68}$$

$$\boldsymbol{N}_{11}^{(-1)} = \boldsymbol{N}_{11}^{-1} + \boldsymbol{N}_{11}^{-1}\boldsymbol{N}_{12}\boldsymbol{N}_{22}^{(-1)}\boldsymbol{N}_{12}^T\boldsymbol{N}_{11}^{-1}. \tag{69}$$

Performing a Cholesky factorization of the block partitioned matrix $\boldsymbol{N}$

$$\begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ (\boldsymbol{N}_{12})^T & \boldsymbol{N}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{11}^T & \\ (\boldsymbol{R}_{12})^T & \boldsymbol{R}_{22}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ & \boldsymbol{R}_{22} \end{bmatrix} \tag{70}$$

the relation between the matrices $\boldsymbol{N}_{ij}$ and the matrices $\boldsymbol{R}_{ij}$ are given by

$$\boldsymbol{N}_{11} = \boldsymbol{R}_{11}^T \boldsymbol{R}_{11} \qquad \boldsymbol{N}_{12} = \boldsymbol{R}_{11}^T \boldsymbol{R}_{12} \qquad \boldsymbol{N}_{22} = \boldsymbol{R}_{12}^T \boldsymbol{R}_{12} + \boldsymbol{R}_{22}^T \boldsymbol{R}_{22} \; . \tag{71}$$

Substituting these quantities in (67) to (69) the well-known relations between the blocks of the inverse matrix and the Cholesky factorized block matrices can be derived as

$$\boldsymbol{N}_{22}^{(-1)} = \left(\boldsymbol{R}_{22}^T \boldsymbol{R}_{22}\right)^{-1}, \tag{72}$$

$$\boldsymbol{N}_{12}^{(-1)} = -\boldsymbol{R}_{11}^{-1} \boldsymbol{R}_{12} \boldsymbol{N}_{22}^{(-1)}, \tag{73}$$

$$\boldsymbol{N}_{11}^{(-1)} = \left(\boldsymbol{R}_{11}^T \boldsymbol{R}_{11}\right)^{-1} - \boldsymbol{N}_{12}^{(-1)} \boldsymbol{R}_{12}^T (\boldsymbol{R}_{11}^{-1})^T \; . \tag{74}$$

To avoid the inversion of large systems the computations are performed by a backward substitution row-by-row starting with the element $n_{nn}^{(-1)}$

$$\left.\begin{aligned} n_{nn}^{(-1)} &= \tfrac{1}{r_{nn}^2} \\ n_{ij}^{(-1)} &= -\tfrac{1}{r_{ii}} \sum_{k=i+1}^{n} r_{ik} n_{kj}^{(-1)} \; , \qquad j = i{+}1, \ldots, n \\ n_{ii}^{(-1)} &= \tfrac{1}{r_{ii}^2} - \tfrac{1}{r_{ii}} \sum_{k=i+1}^{n} r_{ik} n_{ik}^{(-1)} \end{aligned}\right\} \quad i{=}n, \ldots, 1 \; . \tag{75}$$

Fig. 9 gives a graphical interpretation of (75) . It shows the necessary operations to invert the row i. At this step it is assumed, that for the rows $i+1, \ldots, n$ the inverse elements are already computed and for the rows $1, \ldots i$ the Cholesky factorized form is available. By this *backward substitution* process the whole inverse can be computed.

The computation of the inverse matrix, with the Cholesky factorization available, can be performed in the order of $\mathcal{O}(\frac{1}{3}n^3)$ operations, which is twice expensive as the Cholesky factorization step itself, where only $\mathcal{O}(\frac{1}{6}n^3)$ operations are necessary. All in all the inversion (Cholesky and recursive backward edging) process is in the order of $\mathcal{O}(\frac{1}{2}n^3)$. It is well-known, that in general the inversion destroys sparsity, such that the inverse is typically a dense matrix.

In contrast to many other engineering and research areas, where only the solution of the system is requested, in statistics, adjustment theory and geodesy the inverse of the system is of special interest, because the uncertainties and the covariances are deduced from the inverse. The knowledge of this information is crucial in most applications. Sometimes, we are only interested in special elements of the inverse, e.g. the main diagonal elements or elements in a defined bandwidth or off-diagonal elements between special nodes which are characterized by their neighborhood or by points directly connected by measurements. These are typically the entries, which are already non zero in the sparse normal equations. Error propagation within geodetic networks (Schuh, 1981), *Helmert-Blocking*(Wolf, 1978) or *nested dissection* (George, 1973) partitioning strategies to solve large finite element mesh systems and *kite-structures* (Schuh, 1996a,b) to perform a spherical harmonic analysis with irregularly fully populated low degree and block-diagonal structured high degree fields are typical applications.

In connection with the new adjustment of the North American horizontal datum (Bossler, 1976) a special algorithm to compute the inverse elements for an envelope structured sparse matrices was introduced and denoted as *partial inverse* . It is shown by Hanson (1978), that all the inverse elements within the envelope can be computed in a strict manner without any information about the inverse elements outside of the envelope. This is due to the fact, that the non-computed elements within each column of the inverse corresponds exactly to the zero-elements within the Cholesky reduced row. Therefore, the scalar product in (75) (cf. also the computation of $\boldsymbol{N}_{i,i+1:n}^{(-1)}$ in Fig. 9) for all elements within the envelope is not influenced by inverse elements, that are positioned outside the envelope. This argument holds also for special structures coming from a nested dissection partitioning or for kite-structures (Auzinger und Schuh, 1998).

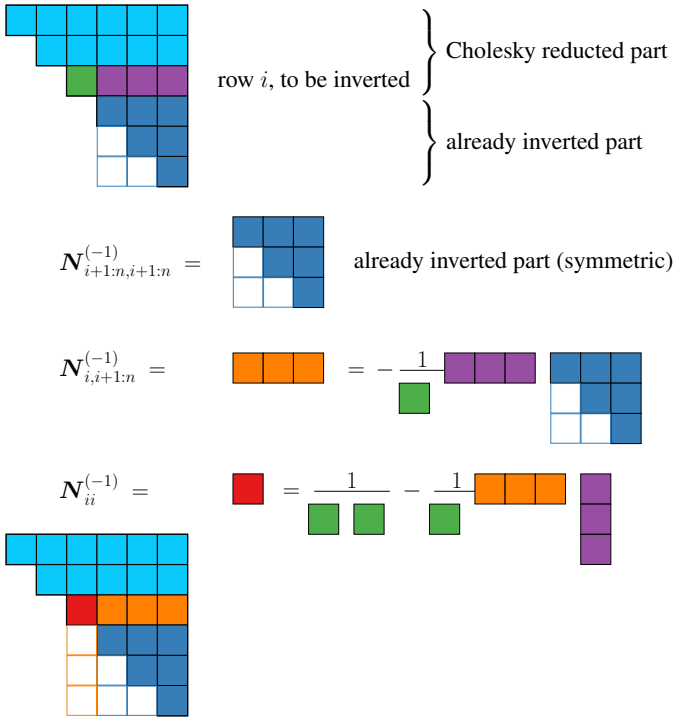


Fig. 9: *Graphical interpretation of the inversion step (75) starting from a Cholesky factorized matrix. By a backward process the inverse matrix is computed row-by-row. In contrast to the upper triangular Cholesky matrix $\boldsymbol{R}$ the inverse matrix $\boldsymbol{N}^{-1}$ is a symmetric matrix. The skeleton-like squares marks this symmetric part, which is typically not stored.*

A.3 Cholesky factorization of a matrix starting from its inverse

In many statistical applications the inverse of the matrix (covariance matrix) plays a central role. For some special applications, it can be useful to compute the Cholesky factors of the initial matrix $\boldsymbol{N}$ directly from the inverse matrix $\boldsymbol{N}^{-1}$ (covariance matrix). To obtain an algorithm for that purpose, it is necessary to reorder the relations between the Cholesky coefficients and the elements of the inverse given in (75) with respect to the Cholesky coefficients. For the last diagonal element from the inverse $n^{(-1)}_{nn}$, we get immediately the last Cholesky coefficient by

$$r_{nn} = 1/\sqrt{n^{(-1)}_{nn}}\,. \tag{76}$$

Now in a step-wise approach the Cholesky coefficients can be computed row-by-row by rearranging (75) with respect to the Cholesky coefficients. This yields to

$$n^{(-1)}_{ij} r_{ii} + n^{(-1)}_{i+1,j} r_{i,i+1} + \ldots + n^{(-1)}_{n,j} r_{in} = 0\,, \qquad j = i+1, \ldots, n \tag{77}$$

and respectively

$$n^{(-1)}_{ii} r_{ii} + n^{(-1)}_{i+1,i} r_{i,i+1} + \ldots + n^{(-1)}_{n,i} r_{in} = \frac{1}{r_{ii}}\,. \tag{78}$$

After multiplying all equations by r_{ii} and a rearranging, we end up with a system of linear equations

$$\begin{bmatrix} r_{ii}r_{ii} \\ r_{ii}r_{i,i+1} \\ \vdots \\ r_{ii}r_{in} \end{bmatrix} = \begin{bmatrix} n_{ii}^{(-1)} & n_{i,i+1}^{(-1)} & \dots & n_{in}^{(-1)} \\ n_{i,i+1}^{(-1)} & n_{i+1,i+1}^{(-1)} & \dots & n_{i+1,n}^{(-1)} \\ \vdots & \vdots & \ddots & \vdots \\ n_{in}^{(-1)} & n_{i-1,n}^{(-1)} & \dots & n_{nn}^{(-1)} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{for} \qquad i = n-1, \dots, 1 \ . \tag{79}$$

The Cholesky coefficients of row i multiplied by r_{ii} are therefore equivalent to the first row of the corresponding sub-matrix of the inverse. The square root of the first element provides r_{ii} and all the other elements can be computed by dividing the elements by r_{ii}. On a first glance this factorization process seems to be very inefficient, because in each step the computation of the inverse of the sub-block is necessary. But this successive inversion can be done by *recursive backward edging*

$$\left[\begin{array}{c|c} n_{11} & \boldsymbol{n}_{12} \\ \hline \boldsymbol{n}_{12}^T & \boldsymbol{N}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \left(n_{11} - \boldsymbol{n}_{12}\boldsymbol{N}_{22}^{-1}\boldsymbol{n}_{12}^T\right)^{-1} & -n_{11}^{(-1)}\boldsymbol{n}_{12}\boldsymbol{N}_{22}^{-1} \\ \hline -\boldsymbol{N}_{22}^{-1}\boldsymbol{n}_{12}^T n_{11}^{(-1)} & \boldsymbol{N}_{22}^{-1} + \boldsymbol{N}_{22}^{-1}\boldsymbol{n}_{12} n_{11}^{(-1)} \boldsymbol{n}_{12}^T \boldsymbol{N}_{22}^{-1} \end{array}\right] = \left[\begin{array}{c|c} n_{11}^{(-1)} & \boldsymbol{n}_{12}^{(-1)} \\ \hline (\boldsymbol{n}_{12}^T)^{(-1)} & \boldsymbol{N}_{22}^{(-1)} \end{array}\right]^{-1} \tag{80}$$

assuming that the inverse $\boldsymbol{N}_{22}^{-1}$ respectively $\boldsymbol{N}_{i+1:n,i+1:n}^{-1}$ are already computed and the inverse of the by one row and column extended matrix should be determined $\boldsymbol{N}_{i:n,i:n}^{-1}$ (Schuh, 2003, p.13f). The required operations of the overall procedure is in the order of $\mathcal{O}(\frac{1}{2}n^3)$. In applications from signal processing, often deal with covariance matrices resulting from equispaced data. These matrices are Toeplitz structured and the Levinson-Durbin recursion allows for a very efficient solution with $\mathcal{O}(2n^2)$ operations (Levinson (1947); Durbin (1960); Golub und van Loan (1983, p.128-129); Schuh (2003, p.291ff)).

A.4 Backward Cholesky factorization

In the standard Cholesky factorization procedure a symmetric matrix $\boldsymbol{N}$ is split into a lower triangular matrix and an upper triangular matrix, where one is the transposed of the other, $\boldsymbol{N} = \boldsymbol{R}^T\boldsymbol{R}$, where $\boldsymbol{R}$ denotes the upper (right) triangular matrix. In the *backward Cholesky factorization* the symmetric matrix is factorized into an upper triangular matrix and a lower triangular matrix. This yields to

$$\boldsymbol{N} = \overline{\boldsymbol{R}}\,\overline{\boldsymbol{R}}^T, \tag{81}$$

or in extended notation to

$$\begin{bmatrix} n_{11} & n_{12} & \dots & n_{1n} \\ n_{12} & n_{22} & \dots & n_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ n_{13} & n_{23} & \dots & n_{nn} \end{bmatrix} = \begin{bmatrix} \overline{r}_{11} & \overline{r}_{12} & \dots & \overline{r}_{1n} \\ & \overline{r}_{22} & \dots & \overline{r}_{2n} \\ & & \ddots & \vdots \\ & & & \overline{r}_{nn} \end{bmatrix} \begin{bmatrix} \overline{r}_{11} & & & \\ \overline{r}_{12} & \overline{r}_{22} & & \\ \vdots & \vdots & \ddots & \\ \overline{r}_{13} & \overline{r}_{23} & \dots & \overline{r}_{nn} \end{bmatrix} . \tag{82}$$

This factorization can be seen as the standard Cholesky approach for a top-down and left-right flipped symmetric system. The factorization starts at the last diagonal element n_{nn} and processes row by row to the top. The notation with the bar on the top should express, that in general the coefficients r_{ij} differ from $\overline{r}_{ij}$. From the identity (82), the recursion formulas for the computation of the coefficients $\overline{\boldsymbol{R}}$ can be immediately deduced

$$\left.\begin{aligned} \overline{r}_{ii} &= \sqrt{n_{ii} - \sum_{k=i+1}^{n} \overline{r}_{ik}^2} \\ \overline{r}_{ij} &= \left(n_{ij} - \sum_{k=j+1}^{n} \overline{r}_{jk}\overline{r}_{ik}\right) / \overline{r}_{jj} \ , \qquad j = n, \dots, i+1 \end{aligned}\right\} \quad i = n, \dots, 1 \ . \tag{83}$$

In correspondence to Sec. A.2, the recursion formulas for the coefficients of the inverse of the symmetric matrix can be computed by *forward substitution*,

$$\left.\begin{aligned} n_{11}^{(-1)} &= \tfrac{1}{\bar{r}_{11}^2} \\ n_{ij}^{(-1)} &= -\tfrac{1}{\bar{r}_{jj}} \sum_{k=1}^{j-1} \bar{r}_{kj} n_{ik}^{(-1)} \,, \qquad i = 1, \ldots, j-1 \\ n_{jj}^{(-1)} &= \tfrac{1}{\bar{r}_{jj}^2} - \tfrac{1}{\bar{r}_{jj}} \sum_{k=1}^{j-1} \bar{r}_{kj} n_{kj}^{(-1)} \end{aligned}\right\} \quad j = 1, \ldots, n \tag{84}$$

and can be reorganized with respect to the computation of the backward Cholesky factors $\bar{r}_{ij}$ from the inverse $\boldsymbol{N}^{-1}$

$$n_{i1}^{(-1)}\bar{r}_{1j} + n_{i2}^{(-1)}\bar{r}_{2j} + \ldots + n_{ij}^{(-1)}\bar{r}_{jj} = 0 \,, \qquad i = 1, \ldots, j-1 \tag{85}$$

$$n_{jj}^{(-1)}\bar{r}_{jj} + n_{1j}^{(-1)}\bar{r}_{1j} + \ldots + n_{j-1,j}^{(-1)}\bar{r}_{j-1,j} = \frac{1}{\bar{r}_{jj}} \qquad . \tag{86}$$

The computation of the backward Cholesky factors of the column j can be organized as the solution of the linear system of equations

$$\begin{bmatrix} \bar{r}_{1j}\bar{r}_{jj} \\ \bar{r}_{2j}\bar{r}_{jj} \\ \vdots \\ \bar{r}_{jj}\bar{r}_{jj} \end{bmatrix} = \begin{bmatrix} n_{11}^{(-1)} & n_{12}^{(-1)} & \ldots & n_{1j}^{(-1)} \\ n_{12}^{(-1)} & n_{22}^{(-1)} & \ldots & n_{2j}^{(-1)} \\ \vdots & \vdots & \ddots & \vdots \\ n_{1j}^{(-1)} & n_{2j}^{(-1)} & \ldots & n_{jj}^{(-1)} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \qquad \text{for} \qquad j = 2, \ldots, n \;\; , \tag{87}$$

as a *recursive forward edging* process. If the inverse matrix is Toeplitz structured the Levinson-Durbin algorithm (Levinson, 1947; Durbin, 1960; Trench, 1964) can be applied.

A.5 Resumé on Cholesky factorization

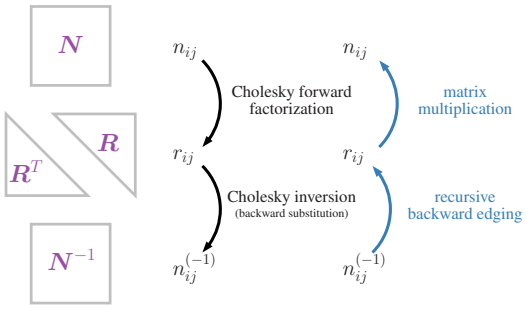
In the last sections two different approaches of the Cholesky factorization are discussed. The standard or *forward Cholesky factorization* (63), where the symmetric matrix $\boldsymbol{N}$ is split into a lower triangular matrix and an upper one

$$\boldsymbol{N} = \boldsymbol{R}^T \boldsymbol{R} \tag{88}$$

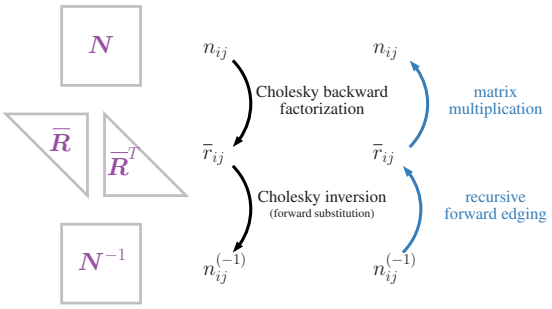
and the *backward Cholesky factorization* (83), where the symmetric matrix $\boldsymbol{N}$ is factorized into

$$\boldsymbol{N} = \overline{\boldsymbol{R}}\, \overline{\boldsymbol{R}}^T \,. \tag{89}$$

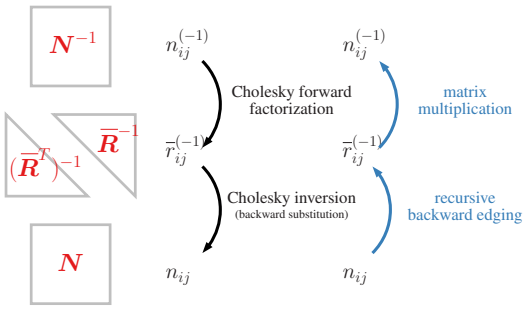
It is important to note that in general $\boldsymbol{R}$ is different from $\overline{\boldsymbol{R}}$. In a further step, the inverse matrix $\boldsymbol{N}^{-1}$ can be computed directly from the Cholesky factorized triangular matrices $\boldsymbol{R}$ or $\overline{\boldsymbol{R}}$ by *backward substitution* (75) or *forward substitution* (84). It is also presented, how the Cholesky factorized matrices $\boldsymbol{R}$ and $\overline{\boldsymbol{R}}$ of the initial matrix $\boldsymbol{N}$ can be derived from the given inverse matrix $\boldsymbol{N}^{-1}$. The matrix $\boldsymbol{R}$ can be computed by *recursive backward edging* (79) and the matrix $\overline{\boldsymbol{R}}$ by *recursive forward edging* (87). The circle can be closed by a simple matrix multiplication (88) or (89) to come back to the initial matrix $\boldsymbol{N}$ from the triangular matrices $\boldsymbol{R}$ and $\overline{\boldsymbol{R}}$. Fig. 10(a) and (b) summarizes this possible paths.



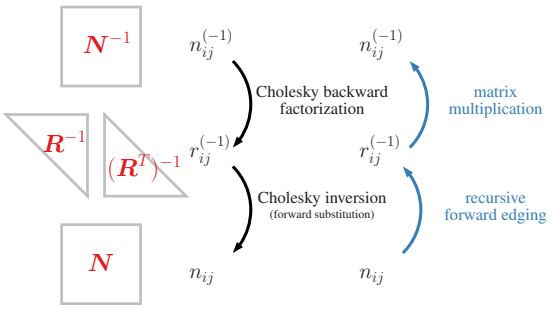
(a) (forward) Cholesky factorization of $\boldsymbol{N}$



(b) backward Cholesky factorization of $\boldsymbol{N}$



(c) Cholesky approach of $\boldsymbol{N}^{-1}$



(d) backward Cholesky approach of $\boldsymbol{N}^{-1}$

Fig. 10: *Review the Cholesky factorization steps applied to $\boldsymbol{N}$ and $\boldsymbol{N}^{-1}$.*

A further interesting perspective is to start with the inverse matrix $\boldsymbol{N}^{-1}$ instead of the matrix $\boldsymbol{N}$ and set up the same sequence of Cholesky factorization steps as done before. Formally the inversion of the matrix $\boldsymbol{N}$ with respect to the *forward Cholesky factorization* (88) can be written

$$\boldsymbol{N}^{-1} = \left(\boldsymbol{R}^T\boldsymbol{R}\right)^{-1} = \boldsymbol{R}^{-1}\left(\boldsymbol{R}^{-1}\right)^T \tag{90}$$

and

$$\boldsymbol{N}^{-1} = \left(\overline{\boldsymbol{R}}\overline{\boldsymbol{R}}^T\right)^{-1} = \left(\overline{\boldsymbol{R}}^{-1}\right)^T\overline{\boldsymbol{R}}^{-1} \tag{91}$$

with respect to *backward Cholesky factorization* (89). It is now crucial that in contrast to the common factorized matrices their inverse forms $\boldsymbol{R}^{-1}$ and $\overline{\boldsymbol{R}}^{-1}$ occur. Changing the view, also the inverse matrix $\boldsymbol{N}^{-1}$ can be factorized into a lower and upper triangular matrix by the *forward Cholesky factorization* (55) we see, that we get immediately $\overline{\boldsymbol{R}}^{-1}$ or by the *backward Cholesky factorization* (83) $\boldsymbol{R}^{-1}$ respectively. This means, starting from $\boldsymbol{N}^{-1}$ the inverses of the triangular matrices can be directly computed. In the next steps the Cholesky inversion by *backward substitution* (75) or *forward substitution* (84) can be computed to get the inverse of the inverse which is for sure the matrix $\boldsymbol{N}$ again. But also the other way is now possible, i.e. the computation of the matrix $\overline{\boldsymbol{R}}^{-1}$ directly from the matrix $\boldsymbol{N}$ by *recursive backward edging* (79) and by *recursive forward edging* (87) the matrix $\boldsymbol{R}^{-1}$. In Fig. 10 all these factorization steps are summarized for direct factorization of the matrix $\boldsymbol{N}$ as well as for $\boldsymbol{N}^{-1}$.

The Cholesky approach allows a lot of variations to factorize a symmetric matrix and opens a way to directly compute $\boldsymbol{R}$ and $\overline{\boldsymbol{R}}$ or $\boldsymbol{R}^{-1}$ or $\overline{\boldsymbol{R}}^{-1}$. It should be mentioned that some of the factorization steps can be extreme efficiently computed for Toeplitz structures resulting from equispaced data distribution.

References

Alkhatib, H., W.-D. Schuh (2007): Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scaled least squares problems. *Journal of Geodesy*, 70:53–66. doi:10.1007/s00190-006-0034-z.

Auzinger, T., W.-D. Schuh (1998): High-degree spherical harmonic analysis combining gridded and random distributed data sets. *Physics and Chemistry of the Earth*, 23(1):19–23. doi:10.1016/S0079-1946(97)00236-X. BibTeX, Pdf.

Benoit (1924): Note Sur Une Méthode de Résolution des équations Normales Provenant de L'Application de la Méthode des Moindres Carrés a un Système D'équations Linéaires en Nombre Inférieur a Celui des Inconnues. — Application de la Méthode a la Résolution D'un Système Defini D'éQuations LinéAires. *Bulletin géodésique*, 2(1):67–77. ISSN 0007-4632. doi:10.1007/BF03031308.

Beutler, G., A. Jäggi, U. Hugentobler, L. Mervart (2006): Efficient satellite orbit modelling using pseudo-stochastic parameter. *J. Geodesy*, 80:353–372. doi:10.1007/s00190-006-0072-6.

Bossler, J. D. (1976): The new adjustment of the North American horizontal datum. *Eos, Transactions American Geophysical Union*, 57(8):557–562. ISSN 2324-9250. doi:10.1029/EO057i008p00557.

Bottoni, G., R. Barzaghi (1993): Fast collocation. *Bulletin Géodésique*, 67:119–126.

Box, G., G. Jenkins (1970): *Time Series Analysis forcasting and control*. Holden-Day.

Boxhammer, C., W.-D. Schuh (2006): GOCE gravity field modeling: computational aspects - free kite numbering scheme. Rummel, R., C. Reigber, M. Rothacher, G. Boedecker, U. Schreiber, J. Flury, (Hrsg.), *Observation of the Earth System from Space*, S. 209–224. Springer, Berlin - Heidelberg. doi:10.1007/3-540-29522-4_15. BibTeX, Pdf.

Buttkus, B. (2000): *Spectral Analysis and Filter Theory in Applied Geophysics*. Springer–Verlag, Berlin, Heidelberg.

Cholesky, A.-L. (1910): Sur la résolution numérique des systèmes d'équations linéaires. *(Société des amis de la Bibliothèque et de l'Histoire de l'École polytechnique, reprint: Bulletin de la Sabix [En ligne], 39, p 81–95 | 2005)*. URL *http://sabix.revues.org/529*.

Durbin, J. (1960): The fitting of time series models. *Rev. inst. Int. Stat.*, 28:233–243.

Ernst, A., W.-D. Schuh (2012): The effect of reordering strategies on rounding errors in large, sparse equation systems. Sneeuw, N., P. Novák, M. Crespi, F. Sansò, (Hrsg.), *VII. Hotine-Marussi-Symposium, IAG Symposia*, Band 137 *Lecture Notes in Earth Sciences*, S. 99–104. Springer, Berlin - Heidelberg. doi:10.1007/978-3-642-22078-4_15. BibTeX, Pdf.

Gaspari, G., S. Cohn (1999): Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

George, A. (1973): Nested dissection of a regular finite element mesh. *SIAM J. Numer. Anal.*, 10:345–363. doi:10.1137/0710032.

George, A., J. W.-H. Liu (1981): *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall.

Golub, G., C. van Loan (1983): *Matrix Computations*. North Oxford Academic, Oxford.

Hanson, R. (1978): A posteriori error propagation. *Proceedings of the "2nd International Symposium on Problems Related to the Redefinition of North American Geodetic Networks"*, S. 427–445. Arlington, Virginia (April 24-28, 1978).

Jäggi, A. (2007): Pseudo-stochastic orbit modelling of Low Earth Satellites using the Global Positioning System. *Geodätisch-geophysikalische Arbeiten in der Schweiz*, Band 73. Schweizerische Geodätische Kommission. URL *http://boris.unibe.ch/id/eprint/25278*.

Kay, S., S. Marple (1981): Spectrum Analysis - A Modern Perspective. *Proceedings of the IEEE*, 69(11):1380–1419. ISSN 0018-9219.

Kleiner, B., R. Martin, D. Thomson (1979): Robust Estimation Of Power Spectra. *Journal of the Royal Statistical Society Series B-Methodological*, 41(3):313–351. ISSN 0035-9246.

Koch, K., H. Kuhlmann, W.-D. Schuh (2010): Approximating covariance matrices estimated in multivariate models by estimated auto- and cross-covariances. *J. Geodesy*, 84(6):383–397. doi:10.1007/s00190-010-0375-5. BibTeX.

Krarup, T. (1969): A contribution to the mathematical foundation of physical geodesy. *Geodætisk Institut, Meddelelse n. 44, København*.

Krasbutter, I., J. M. Brockmann, B. Kargoll, W.-D. Schuh (2014): Adjustment of digital filters for decorrelation of GOCE SGG data. Flechtner, F., N. Sneeuw, W.-D. Schuh, (Hrsg.), *Observation of the System Earth from Space - CHAMP, GRACE, GOCE and future missions.*, Band 20 *Advanced Technologies in Earth Sciences, GEOTECHNOLOGIEN Science Report*, S. 109–114. Springer. doi:10.1007/978-3-642-32135-1_14. BibTeX, Pdf.

Krasbutter, I., B. Kargoll, W.-D. Schuh (2015): Magic Square of Real Spectral and Time Series Analysis with an Application to Moving Average Processes. Kutterer, H., F. Seitz, H. Alkhatib, M. Schmidt, (Hrsg.), *The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11), IAG Symposia*, Band 140 *International Association of Geodesy Symposia*, S. 9–14. Springer. ISBN 978-3-319-10827-8. doi:10.1007/978-3-319-10828-5_2. BibTeX, Pdf.

Levinson, N. (1947): The Wiener RMS (Root Mean Square) error criterion in filter design and prediction. *Journal of Mathematics & Physics*, No.25:261–278.

Mayer-Gürr, T. (2006): *Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE*. Dissertation, Promotion an der Landwirtschaftlichen Fakultät der Universität Bonn, Schriftenreihe des Instituts für Geodäsie und Geoinformation der Rheinischen Friedrich-Wilhelms-Universität, Folge 9. URL *http://nbn-resolving.de/urn:nbn:de:hbz:5N-09047*.

Meissl, P. (1980): A priori prediction of roundoff error accumulation in the solution of a super-large geodetic normal equation system. NOAA / National Ocean Survey's National Geodetic Survey (NGS), Rockville, Md. URL *http://trove.nla.gov.au/work/19473491?selectedversion=NBD2223841*.

Meissl, P. (1982): Least squares adjustment: A modern approach. *Mitteilungen der Geodätischen Institute der TU Graz*, Band 43. Geodätischen Institute der TU Graz, Graz. URL *ftp://skylab.itg.uni-bonn.de/schuh/Separata_Meissl/meissl_82b.pdf*.

Moreaux, G. (2008): Compactly supported radial covariance functions. *Journal of Geodesy*, 82(7):431–443. ISSN 0949-7714. doi:10.1007/s00190-007-0195-4.

Moritz, H. (1973): *Least-squares collocation*. Deutsche Geodätische Kommission, München. Reihe A, 75.

Moritz, H. (1980): *Advanced Physical Geodesy*. Wichmann, Karlsruhe.
Poder, K., C. Tscherning (1973): *Cholesky's method on a computer*. Internal Report No. 8, The Danish Geodetic Institute.
Priestley, M. (2004): *Spectral Analysis and Time Series*. Elsevier Academic Press, Amsterdam.
Sansò, F., W.-D. Schuh (1987): Finite covariance functions. *Bulletin Géodésique*, 61(4):331–347. doi:10.1007/BF02520559. BibTeX.
Sansò, F., C. Tscherning (2003): Fast spherical collocation: theory and examples. *J. Geodesy*, 77:101–112.
Schall, J., A. Eicker, J. Kusche (2014): The ITG-Goce02 gravity field model from GOCE orbit and gradiometer data based on the short arc approach. *Journal of Geodesy*, 88(4):403–409. ISSN 0949-7714. doi:10.1007/s00190-014-0691-2.
Schuh, W.-D. (1981): *Programmierung rationeller Algorithmen zur Umordnung, Auflösung und Inversion der Normalgleichungen geodätischer Netze*. Diplomarbeit, Technische Universität Graz.
Schuh, W.-D. (1989): Kollokation - zu rechenaufwendig? *ZAMM, Z. angew. Math. Mech. 69*, 4:T73–T75. URL *http://onlinelibrary.wiley.com/doi/10.1002/zamm.19890690403/pdf*.
Schuh, W.-D. (1996a): Least squares adjustment of high degree spherical harmonics. Jacobsen, B. e., (Hrsg.), *Inverse Methods - Interdisciplinary Elements of Methodology, Computation and Application*, Lecture Notes in Earth Sciences 63, S. 276–283. Springer, Heidelberg. doi:10.1007/BFb0011786. BibTeX, PDF.
Schuh, W.-D. (1996b): *Tailored numerical solution strategies for the global determination of the earth's gravity field*, Band 81 *Mitteilungen der Geodätischen Institute*. Technische Universität Graz (TUG), Graz. URL *ftp://skylab.itg.uni-bonn.de/schuh/Separata/schuh_96.pdf*. BibTeX, PDF.
Schuh, W.-D. (2003): *Numerische Verfahren zur geodätischen Optimierung*. Skriptum. Theoretische Geodäsie, Universität Bonn.
Schuh, W.-D. (2016): Signalverarbeitung in der Physikalischen Geodäsie. Freeden, W., R. Rummel, (Hrsg.), *Handbuch der Geodäsie*, Band Erdmessung und Satellitengeodäsie *Springer Reference Naturwissenschaften*, S. 73–121. Springer Berlin Heidelberg. ISBN 978-3-662-47099-2. doi:10.1007/978-3-662-47100-5_15. BibTeX, PDF.
Schuh, W.-D., I. Krasbutter, B. Kargoll (2014): Korrelierte Messung - was nun? Neuner, H., (Hrsg.), *Zeitabhängige Messgrößen - Ihre Daten haben (Mehr-)Wert*, Band 74 *DVW-Schriftenreihe*, S. 85 – 101. Wißner, Augsburg. BibTeX, PDF.
Siemes, C. (2008): *Digital Filtering Algorithms for Decorrelation within Large Least Squares Problems.* Dissertation, Landwirtschaftliche Fakultät der Universität Bonn, Bonn. URL *http://nbn-resolving.de/urn:nbn:de:hbz:5N-13749*. BibTeX.
Snay, R. (1976): Reducing the profile of sparse symmetric matrices. *Bulletin Géodésique*, 50:341–352.
Trench, W. (1964): An algorithm for the inversion of finite Toeplitz matrices. *SIAM J. Soc. Indust. Appl. Math.*, 12:515–522.
Wolf, H. (1978): The Helmert Block Method - Its Origin and Development. *Second International Symposium on Problems Related to the Redefinition of the North American Geodetic networks*, S. 319–326. US Department of Commerce, Washington.

Index