

Summer 8-2-2012

Completely Recursive Least Squares and Its Applications

Xiaomeng Bian

University of New Orleans, xiaomeng.bian@gmail.com

Follow this and additional works at: <https://scholarworks.uno.edu/td>



Part of the [Power and Energy Commons](#), and the [Signal Processing Commons](#)

Recommended Citation

Bian, Xiaomeng, "Completely Recursive Least Squares and Its Applications" (2012). *University of New Orleans Theses and Dissertations*. 1518.

<https://scholarworks.uno.edu/td/1518>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

Completely Recursive Least Squares and Its Applications

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
In
Engineering and Applied Science
Electrical Engineering

by

Xiaomeng Bian

B. S. Zhejiang University, China, 2000

August, 2012

@Copyright 2012, Xiaomeng Bian

Dedication

This dissertation is dedicated to all I have ever learnt from. Sincerely.

Acknowledgement

First, I would like to give my sincere thanks to my families: my parents, uncles and aunties, brothers and sisters, and my in-heaven grandparents. Their selfless love and firm support encourage me to face all the challenges in my study-and-life time.

I would like to give my sincere thanks to Dr. George E. Ioup, Dr. Linxiong Li, Dr. Vesselin P. Jilkov, Dr. Huimin Chen and Dr. Zhide Fang for the helpful guidance, suggestions and comments, as well as the patience and time that you have provided. I would express my gratitude to Professor Jiaju Qiu, Professor Deqiang Gan and Professor Meiqin Liu from Zhejiang University, China. I do appreciate your kind instructions and encouragements in the past years. I would also say special thanks to Dr. Huimin Chen for your insightful discussions and comments on my papers and dissertation.

I would also like to express my gratitude to Dr. Ming Yang, Dr. Zhansheng Duan, Dr. Ryan R. Pitre, Mr. Arther G Butz, Mr. Yu Liu, Mr. Rastin Rastgoufard, Mr. Jiande Wu, Mr. Gang Liu, Dr. Bingjian Wang, Dr. Liping Yan, Miss Jing Hou and all the other colleagues and classmates I have met. I do enjoy the “attacking” atmosphere in our Information & System Lab. Your challenging questions do help me refine and improve my ideas.

Finally and foremost, I would like to thank my major advisor Dr. X. Rong Li for maturing me academically and for teaching me the art of critical thinking [R. R. Piter]. It is because of your strict and normative training and guidance that I can truly feel the beauty in the new realm. Thank you.

Table of Contents

List of Figures	viii
List of Tables	ix
Abstract	x
1. Introduction and Literature Review	001
1.1 Classification of LS, WLS and GLS	001
1.2 Review of Batch LS/WLS/GLS Solutions	002
1.2.1 Batch LS/WLS Methods	002
1.2.2 Batch GLS and LE-Constrained LS solutions	008
1.3 Recursive LS Approaches	009
1.3.1 Recursive Methods	009
1.3.2 RLS Initialization and Deficient-Rank Problems	011
1.4 Completely Recursive Least Squares (CRLS)	014
1.5 CRLS, LMMSE and KF.....	015
1.6 Power System State Estimation and Parameter Estimation	016
1.7 Our Work and Novelty	017
1.8 Outline.....	021
Nomenclature in Chapters 2-3	022
2. Completely Recursive Least Squares—Part I: Unified Recursive Solutions to GLS	024
2.1 Background	024
2.2 Problem Formulation	026
2.3 Theoretical Support	027
2.3.1 Preliminaries	027
2.3.2 Recursive solutions to unconstrained GLS	028
2.3.3 Recursive solutions to LE-constrained GLS	031
2.3.4 Recursive imposition of LE constraints	032
2.3.5 Recursive solution to coarse GLS	035
2.4 Unified Procedures and Algorithms	042
2.5 Performance analysis	045
2.6 Appendix	046
2.6.1 Appendix A: Proof of Proposition 1	046
2.6.2 Appendix B: Proof of Theorem 1	047
2.6.3 Appendix C: Proof of Fact 4	049
2.6.4 Appendix D: Proof of Theorem 2.....	050
2.6.5 Appendix E: Proof of Theorem 3	051
2.6.6 Appendix F: Proof of Proposition 2	053
2.6.7 Appendix G: Proof of Proposition 3.....	053
2.6.8 Appendix H: Proof of Theorem 5.....	054
3. Completely Recursive Least Squares—Part II: Recursive Initialization, Linear- Equality	

Imposition and Deficient-Rank Processing.....	055
3.1 Introduction	055
3.1.1 Background	055
3.1.2 Our Work	056
3.2 Problem Formulations	057
3.2.1 WLS and Its Batch Solutions	057
3.2.2 RLS Solutions	059
3.2.3 Our Goals	059
3.3 Derivations and Developments	060
3.3.1 Preliminaries	060
3.3.2 Development of Recursive Initialization	060
3.3.3 Deficient-Rank Processing.....	068
3.4 Overall Algorithm and Implementation	070
3.5 Performance Analysis.....	072
3.6 Appendices	074
3.6.1 Appendix A: Proof of Theorem 2	074
3.6.2 Appendix B: Proof of Theorem 5.....	076
3.6.3 Appendix C: Proof of Theorem 6.....	078
4. Linear Minimum Mean-Square Error Estimator and Unified Recursive GLS	079
4.1 Review of LMMSE Estimation and GLS	079
4.1.1 LMMSE Estimator	079
4.1.2 LMMSE Estimation Based on Linear Data Model	080
4.2 Verification of Generic Optimal Kalman Filter	082
Nomenclature in Chapters 5-6	087
5. Joint Estimation of State and Parameter with Synchrophasors—Part I: State Tracking	088
5.1 Introduction	088
5.1.1 Background and Motivation.....	088
5.1.2 Literature Review	089
5.1.3 Our Work.....	089
5.2 Formulation and State-Space Models	090
5.2.1 State Tracking with Uncertain Parameters	091
5.2.2 Prediction Model for State Tracking.....	091
5.2.3 Measurement Model of State Tracking.....	092
5.3 Adaptive Tracking of State.....	093
5.3.1 Basic Filter Conditioned on Correlation	093
5.3.2 Detection and Estimation of Abrupt Change in State	095
5.4 Procedure and Performance Analysis.....	096
5.4.1 Explanations on Prediction and Measurement Models	096
5.4.2 Implementation Issues.....	098
5.4.3 On Observability, Bad-Data Processing and Accuracy	099
5.5 Comparative Experiments.....	100

5.6 Conclusions	106
5.7 Appendices	106
5.7.1 Appendix A: Derivation of Filter on Correlated Errors and Noise	106
5.7.2 Appendix B: Derivation of Abrupt Change Estimation	109
6. Joint Estimation of State and Parameter with Synchrophasors—Part II: Para. Tracking	111
6.1 Introduction	111
6.1.1 Previous Work	111
6.1.2 Our Contributions.....	112
6.2 Parameter Estimation	113
6.3 Formulation of Parameter Tracking	114
6.3.1 Prediction Model.....	115
6.3.2 Measurement Model.....	115
6.4 Tracking of Line Parameters	116
6.4.1 Basic Filter	116
6.4.2 Adjustment of Transition Matrix	118
6.5 Error Evolution and Correlation Calculation	119
6.6 Procedures and Performance Analysis.....	122
6.6.1 Parameter Tracking and Overall Procedure	122
6.6.2 Accuracy and Complexity.....	122
6.6.3 A Practical Scheme	123
6.7 Simulations.....	124
6.7.1 Comparisons with Other Approaches	124
6.7.2 Practical Two-Step Implementation of the Approach.....	126
6.8 Conclusions	128
6.9 Appendix	128
7. Summary and Future Work	130
References	133
Vita	140

List of Figures

Fig. 5.1 State tracking procedure with abrupt-change detection.....	099
Fig. 5.2 Realistic 18-bus system	101
Fig. 5.3 Comparison on effect of parameter uncertainty.....	103
Fig. 5.4 Comparison on effect of abrupt-state-change detection	105
Fig. 5.5 Comparison with random-walk model	105
Fig. 5.6 Comparison with forecasting model	106
Fig. 6.1 Joint state and parameter estimation procedure	123
Fig. 6.2 12-bus 500kV system.....	125
Fig. 6.3 Performance comparison with EKF.....	126
Fig. 6.4 Practical scheme based on 12-bus system	127
Fig. 6.5 Practical scheme based on 12-bus system (zoomed in).....	128

List of Tables

TABLE 5.1 Brach Parameter (18-Bus System)	103
TABLE 5.2 RMSEs of First Experiment	104
TABLE 5.3 RMSEs of Secon Experiment	104
TABLE 6.1 Parameters (Realistic System)	125

Abstract

The matrix-inversion-lemma based recursive least squares (RLS) approach is of a recursive form and free of matrix inversion, and has excellent performance regarding computation and memory in solving the classic least-squares (LS) problem. It is important to generalize RLS for generalized LS (GLS) problem. It is also of value to develop an efficient initialization for any RLS algorithm.

In Chapter 2, we develop a unified RLS procedure to solve the unconstrained/linear-equality (LE) constrained GLS. We also show that the LE constraint is in essence a set of special error-free observations and further consider the GLS with implicit LE constraint in observations (ILE-constrained GLS).

Chapter 3 treats the RLS initialization-related issues, including rank check, a convenient method to compute the involved matrix inverse/pseudoinverse, and resolution of underdetermined systems. Based on auxiliary-observations, the RLS recursion can start from the first real observation and possible LE constraints are also imposed recursively. The rank of the system is checked implicitly. If the rank is deficient, a set of refined non-redundant observations is determined alternatively.

In Chapter 4, base on [Li07], we show that the linear minimum mean square error (LMMSE) estimator, as well as the optimal Kalman filter (KF) considering various correlations, can be calculated from solving an equivalent GLS using the unified RLS.

In Chapters 5 & 6, an approach of joint state-and-parameter estimation (JSPE) in power system monitored by synchrophasors is adopted, where the original nonlinear parameter problem is reformulated as two loosely-coupled linear subproblems: state tracking and parameter tracking. Chapter 5 deals with the state tracking which determines the voltages in JSPE, where dynamic behavior of voltages under possible abrupt changes is studied. Chapter 6 focuses on the subproblem of parameter tracking in JSPE, where a new prediction model for parameters with moving means is introduced. Adaptive filters are developed for the above two

subproblems, respectively, and both filters are based on the optimal KF accounting for various correlations. Simulations indicate that the proposed approach yields accurate parameter estimates and improves the accuracy of the state estimation, compared with existing methods.

Chapter 1: Introduction and Literature Review

1.1 Classification of Linear LS, WLS and GLS

The principle of *least squares* (LS), which was first invented independently by a few scientists and mathematicians such as C. F. Gauss, A. M. Legendre and R. Adrain [Stigler86][Li07], is a classic and standard approach to obtaining the optimal solution of an overdetermined system to minimize the sum of squared residuals.

The most popular and important interpretation of the LS approach is from the application in data fitting. That is, the best fitting in the LS sense is to approximate a set of parameters (estimands) such that the sum of squared residuals is minimized, where the residuals are differences between the measured observation values and the corresponding fitted ones [Wiki01]. In addition, the LS problem also has various names in different disciplines [Li07]. For instance, in some mathematical areas, LS may be treated as a special *minimum l_2 -norm* problem [Bjorck96]. In statistics, it is also formulated as a probabilistic problem widely used in *regression analysis* or *correlation analysis* [Freedman05] [Kleinbaum07]. In engineering, it is a powerful tool adopted in *parameter estimation*, *filtering*, *system identification*, and so on [Sorenson80] [Ba-Shalom01]. In particular, in the area of estimation, the LS formulation can be derived from the maximum-likelihood (ML) criterion if the observation errors are normally distributed. The LS estimator can also be treated as a moment estimator [Wiki01].

Roughly speaking, LS problems can be classified into linear and nonlinear cases, depending on whether the involved observation quantities are linear functions of the estimand or not. It is also well known that, with linearization techniques such as Gauss-Newton methods, a nonlinear LS problem may be converted to linearized iterative refinements. This dissertation focuses on *linear LS* solutions.

More generally, the object of the data fitting may be extended to minimize the sum of the weighted residual squares, which leads to the definition of least squares with weights. According to the features of their weighting matrices, linear LS problems with weights can be

largely categorized from the simple to the complex as LS, weighted LS (WLS), and generalized LS (GLS), where linear-equality (LE) constraints may be imposed. Note that in this dissertation the concept of WLS is limited to the case with a diagonal matrix while the GLS has a non-diagonal weighting matrix [Amemiya85] [Greene00] [Wiki02]. In addition, linear-inequality constraints may also be involved and can be treated as combinations of LE constraints [Lawson95].

Note that in some statistics books, “weighted least squares” may be used for LS problems with equal weights while those with distinguished weights are named generally- weighted least squares. As the equally-weighted LS and the conventional LS has the same solution plus mutually-proportional estimation-error covariances, we ignore their difference and follow the convention in [Lawson95] [Bjorck96]: WLS is for LS with distinguished (not all-equal) weights and GLS is for LS weighted by an arbitrary PD matrix.

In summary, we use the following LS/WLS categorizing list to present the LS solutions from simplest to most complex problem setup:

- Unconstrained LS → LE-constrained LS
- Unconstrained/LE-constrained WLS
- Unconstrained GLS → LE-constrained GLS
- Implicitly-LE-constrained (ILE-constrained) GLS

We focus on the development of the recursive unconstrained/LE-constrained/ILE-constrained LS/WLS/GLS solutions and the related initialization as well as deficient-rank processing. The study starts from the conventional RLS and its exact initializations, which are reviewed below.

1.2 Review of Batch LS/WLS/GLS Solutions

1.2.1 Batch LS/WLS Methods

As a classic minimization problem, the LS problem has been studied for more than two centuries. Many methods and algorithms have been developed and well surveyed in the past. Most commonly-recognized methods and algorithms are presented in well-known books, such

as [Lawson95] and [Bjorck96]. Among the exiting LS methods and topics, we will review those issues related to our research in detail. Roughly speaking, the (unconstrained) linear LS approach is to solve the following classic (unconstrained) linear LS problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J \quad (1.1a)$$

$$\text{with } J = \sum_{m=1}^M (z_m - H_m \mathbf{x})^T (z_m - H_m \mathbf{x}) = (\mathbf{z} - \mathbf{H}\mathbf{x})^T (\mathbf{z} - \mathbf{H}\mathbf{x}) \quad (1.1b)$$

where $\mathbf{x} = [\cdots x_n \cdots]^T$, $\mathbf{H} = [\cdots H_m^T \cdots]^T$, and $\mathbf{z} = [\cdots z_m \cdots]^T$. x_n is the n th to-be-determined quantity. H_m and z_m are the coefficient (row vector) and value of the m th observation, respectively, and M is the total observation number. Typically, $\hat{\mathbf{x}}$ can be obtained via normal-equation solution, QR-decomposition (of \mathbf{H}), Gauss elimination and so on. These methods are reviewed as follows.

The solutions can come from solving the following normal equation:

$$(\mathbf{H}^T \mathbf{H}) \hat{\mathbf{x}} = \mathbf{H}^T \mathbf{z} \quad (1.2a)$$

That is,

$$(\sum_{m=1}^M H_m^T H_m) \hat{\mathbf{x}} = \sum_{m=1}^M H_m^T z_m \quad (1.2b)$$

Clearly, if and only if (iff) $\text{rank}(\mathbf{H}) = N$, (1.2b) has a unique solution and the batch solution is

$$\begin{cases} \mathbf{C} = (\mathbf{H}^T \mathbf{H})^{-1} \\ \hat{\mathbf{x}} = \mathbf{C} \mathbf{H}^T \mathbf{z} \end{cases} \quad (1.2c)$$

where estimation-error-covariance-like (EEC-like) matrix \mathbf{C} is closely related to the covariance of estimation errors in engineering applications. Matrix triangularization and diagonalization techniques such as Cholesky decomposition (LL^T decomposition) can be used to decompose $\mathbf{H}^T \mathbf{H}$ and thus compute \mathbf{C} and $\hat{\mathbf{x}}$ efficiently [Martin&Wilkinson65] [Passino98], where the symmetric structure of $\mathbf{H}^T \mathbf{H}$ is advantageous. Several variants, as well as different ways to sequencing the CD operations, were discussed in [George81]. Reference [Golub96] also shows that LL^T decomposition and the QR decomposition method can lead to equal upper triangular matrices if \mathbf{H} has full column rank.

Actually, $\hat{\mathbf{x}}$ can also be determined directly from the nonsymmetric linear equation $\mathbf{H}\mathbf{x} = \mathbf{z}$. Gauss elimination with partial pivoting is used to solve $\mathbf{H}\mathbf{x} = \mathbf{z}$, and different Gauss-elimination based methods are surveyed in [Nobel73]. Among the existing approaches, the Peters-Wikinson method is a uniform one. It utilizes the LU factorization to reduce the original LS problem to a simplified one with a lower triangular coefficient. Because the solution is obtained from the triangularized coefficient and thus suffers less from rounding error, this method is numerically more stable than those using the normal equation directly.

More popularly, the QR decomposition can be employed to decompose the observation coefficient matrix into a product of an orthogonal square matrix \mathbf{Q} and an upper triangular matrix such that

$$\mathbf{H} = \mathbf{Q} \begin{bmatrix} \mathbf{R}^T & \mathbf{0}^T \end{bmatrix}^T \quad (1.3)$$

where \mathbf{R} is an upper triangular square matrix. Correspondingly, the observation-value vector \mathbf{z} is transformed into

$$\mathbf{z} = \mathbf{Q} \begin{bmatrix} \tilde{\mathbf{z}}_1^T & \tilde{\mathbf{z}}_2^T \end{bmatrix}^T \quad (1.4)$$

Accordingly, the objective function in (1.1a) becomes as simple as

$$J = (\tilde{\mathbf{z}}_1 - \mathbf{R}\mathbf{x})^T (\tilde{\mathbf{z}}_1 - \mathbf{R}\mathbf{x}) + \tilde{\mathbf{z}}_2^T \tilde{\mathbf{z}}_2 \quad (1.5)$$

Then $\hat{\mathbf{x}}$ can be obtained conveniently by solving the following linear equation

$$\mathbf{R}\hat{\mathbf{x}} = \tilde{\mathbf{z}}_1 \quad (1.6)$$

The solution can take advantage of the upper-triangular structure of \mathbf{R} and can be obtained by back substitution efficiently. Accordingly, the computational complexity of solving (1.6), which is denoted by the number of the involved floating-point operations (flops), is only as low as $O(N^2)$ (order of N^2). The computation accuracy is also high because the observation-coefficient-matrix decomposition suffers little from rounding errors [Bjorck96] [Golub96].

Subsequently, the major work is on constructing appropriate orthogonal matrix \mathbf{Q} . Many classic methods have been utilized, such as Householder reflection, Givens rotation, and classical or modified Gram-Schmidt [Horn85] [Press07] [Parlett00].

For instance, in the Householder transformation based method [Golub65a] [Householder58], the orthogonal matrix \mathbf{Q} is constructed as a product of a sequence of orthogonal matrix:

$$\mathbf{Q} = \mathbf{Q}_1 \cdots \mathbf{Q}_k \cdots \mathbf{Q}_N \quad (1.7)$$

where $\mathbf{Q}_k = \text{diag}(\mathbf{I}_{N-k}, \tilde{\mathbf{Q}}_k)$. $\tilde{\mathbf{Q}}_k$ is a Householder reflection matrix which is designed to satisfy

$$\tilde{\mathbf{Q}}_k \tilde{\mathbf{a}}_k = \|\tilde{\mathbf{a}}_k\|_2 [1 \ 0_2 \ \cdots \ 0_{N-k+1}]^T \quad (1.8)$$

and $\tilde{\mathbf{a}}_k$ is a subvector in the k th column of the following transformed intermediate observation-coefficient matrix:

$$\mathbf{Q}_{k-1} \cdots \mathbf{Q}_1 \mathbf{H} = \begin{bmatrix} \mathbf{R}_{k-1} & * \\ \mathbf{0} & [\tilde{\mathbf{a}}_k \ *] \end{bmatrix} \quad (1.9)$$

Consequently,

$$\tilde{\mathbf{Q}}_k = \mathbf{I} - 2\mathbf{b}_k \mathbf{b}_k^T / \|\mathbf{b}_k\|_2^2 \quad (1.10)$$

with $\mathbf{b}_k = \tilde{\mathbf{a}}_k - \|\tilde{\mathbf{a}}_k\|_2 [1 \ 0_2 \ \cdots \ 0_{N-k+1}]^T$.

In the Givens rotation based method [Givens58], the QRD matrix \mathbf{Q} is a product of a series of plane-rotation matrices which have a specific form as

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cos \theta & \mathbf{0} & \sin \theta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ 0 & -\sin \theta & 0 & \cos \theta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \quad (1.11)$$

Given a vector $\tilde{\mathbf{a}}$ which has the same size as $\tilde{\mathbf{Q}}$'s column and $\tilde{\mathbf{a}} = [\mathbf{a}_1^T \ \tilde{a}_1 \ \mathbf{a}_2^T \ \tilde{a}_2 \ \mathbf{a}_3^T]^T$, one has

$$\mathbf{a}' = \tilde{\mathbf{Q}} \tilde{\mathbf{a}} = [\mathbf{a}_1^T \ \tilde{a}_1 \cos \theta + \tilde{a}_2 \sin \theta \ \mathbf{a}_2^T \ -\tilde{a}_1 \sin \theta + \tilde{a}_2 \cos \theta \ \mathbf{a}_3^T]^T \quad (1.12)$$

It is clear that, if $\sin \theta = \tilde{a}_2 / \sqrt{\tilde{a}_1^2 + \tilde{a}_2^2}$ and $\cos \theta = \tilde{a}_1 / \sqrt{\tilde{a}_1^2 + \tilde{a}_2^2}$, then

$$\mathbf{a}' = [\mathbf{a}_1^T \ \sqrt{\tilde{a}_1^2 + \tilde{a}_2^2} \ \mathbf{a}_2^T \ 0 \ \mathbf{a}_3^T]^T \quad (1.13)$$

After a sequence of rotations as in (1.12)-(1.13) are applied to the observation coefficient matrix \mathbf{H} , the upper triangular form in (1.3) can be easily obtained. Particularly, fast Givens

rotation methods are also developed [Gentleman73] [Hammarling74] [Lawson79], where the multiplication number is reduced by adopting the scaled two-factor form of \mathbf{H} . Owing to the effect of scaling-factor update, the traditional fast rotations may suffer from underflow problems. Correspondingly, self-scaling fast rotations, which can monitor and rescale the size of the scaling factors, are developed in [Anda94] to overcome the underflow problems.

Both the Householder-transformation based and the Givens-rotation based approaches have good properties regarding computation and storage. For instance, the standard Householder factorization requires $N^2(M - \frac{1}{3}N)$ flops while the normal-equation method may use $\frac{1}{2}N^2(M + \frac{1}{3}N)$ flops. Hence, the Householder transformation method requires roughly the same computation as the \mathbf{LL}^T decomposition based normal-equation one for $M \approx N$ but has twice computation for $M \ll N$ [Bjorck96]. The standard Givens-rotation method takes more computation as $2N^2(M - \frac{1}{3}N)$ multiplications. However, the QR-decomposition methods have overwhelmingly higher accuracy than the normal-equation ones. They are numerically more stable since the solution does not involve $(\mathbf{H}^T \mathbf{H})$ but is determined from $\mathbf{H}\mathbf{x} = \mathbf{z}$ directly. The Givens rotations are easy to implement and also have convenient recursive forms (see next subsection).

As surveyed in [Lawson95] [Bjorck96], Gram-Schmidt orthogonalization is also employed to produce the orthogonal matrix \mathbf{Q} in the QR decomposition. The classic Gram-Schmidt method, which first appeared in [Schmidt1908], may lose orthogonality in some ill-conditions and is thus a theoretical tool rather than a good base for numerical algorithms. However, the modified Gram-Schmidt methods can reduce the risk of loss of orthogonality [Gulliksson95].

In addition, the singular value decomposition (SVD) can also be adopted to solve the LS problems. That is, for the LS problem in (1.1), given the SVD of \mathbf{H} as

$$\mathbf{H} = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^T \quad (1.14)$$

then the minimum-norm LS solution is

$$\hat{\mathbf{x}} = \mathbf{H}^+ \mathbf{z} = \mathbf{V} \begin{pmatrix} \mathbf{\Sigma}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U}^T \mathbf{z} \quad (1.15)$$

where superscript “+” stands for the Moore-Penrose pseudo inverse (MP inverse). $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_{N_H})$, σ_i is the square root of the i th eigenvalue of $\mathbf{H}^T \mathbf{H}$, and $\text{rank}(\mathbf{H}) = N_H$. The left and right singular vectors, which are stored in the orthogonal matrices \mathbf{U} and \mathbf{V} , are the corresponding eigenvectors of $\mathbf{H}^T \mathbf{H}$ and $\mathbf{H} \mathbf{H}^T$, respectively [Lawson95]. The first stable algorithm based on SVD was presented in [Golub65b], where \mathbf{H} is reduced to a bidiagonal matrix via Householder transformation of a Lanczos process such that the singular values and vectors refer to eigenvalues and eigenvectors of a special tridiagonal matrix [Bjorck96]. Later, adaptation and improvement are made to the QR algorithm [Golub68] [Golub70]. Newer Jacobi methods are also developed to improve the relative computation accuracy of the singular values in bidiagonal matrices [Kogbetliantz55] [Hestenes58]. Note that, in (1.14), $N_H \leq N$. If $N_H < N$, the SVD based solution leads to the minimum-norm LS solution, which is a powerful tool for the deficient-rank analysis discussed in Sec. 1.3.2.

WLS is generalized from LS, in which each observation is assigned a positive weight:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J \quad (1.16)$$

$$\text{with } J = \sum_{m=1}^M (z_m - H_m \mathbf{x})^T w_m (z_m - H_m \mathbf{x}) = (\mathbf{z} - \mathbf{H} \mathbf{x})^T \mathbf{W} (\mathbf{z} - \mathbf{H} \mathbf{x}) \quad (1.17a)$$

$$\text{Equivalently, } J = (\mathbf{z} - \mathbf{H} \mathbf{x})^T \mathbf{W} (\mathbf{z} - \mathbf{H} \mathbf{x}), \quad (1.17b)$$

$$\text{where } \mathbf{W} = \text{diag}(\dots, w_m, \dots) \quad (1.17c)$$

Theoretically speaking, since w_m can be easily decomposed into $w_m^{\frac{1}{2}} w_m^{\frac{1}{2}}$, problem (1.16) can be converted to an equivalent LS one without difficulty. As a result, all the batch LS methods, such as the normal-equation based, the QRD based and the SVD based algorithms, are applicable. The major issues arise from possible ill conditions deteriorated by significantly large weight ratios and special efforts are made to treat the extremely ill-conditioned situations [Vavasis94] [Hough94] [Bjorck80] [Duff94] [Gulliksson92] [Gulliksson95] [Anda94].

1.2.2 Batch GLS and LE-Constrained LS Solutions

Compared with WLS, GLS is a further generalized LS problem, where the weight \mathbf{W} can be an arbitrary positive definite (PD) matrix. Since a GLS problem can always be converted to an equivalent LS form by decomposing \mathbf{W} , those methods used for solving LS are all applicable to GLS. For instance, the normal equation of GLS is

$$(\mathbf{H}^T \mathbf{W} \mathbf{H}) \hat{\mathbf{x}} = \mathbf{H}^T \mathbf{W} \mathbf{z} \quad (1.18)$$

where $(\mathbf{H}^T \mathbf{W} \mathbf{H})$ remains symmetric. Then those normal-equation based methods used in LS, such as LL^T decomposition based ones, can still be utilized.

Methods based on the decomposition of \mathbf{W} are also widely adopted. That is,

$$\mathbf{W} = \tilde{\mathbf{W}} \tilde{\mathbf{W}}^T \quad (1.19)$$

Based on (1.19), generalized QRD and SVD methods are developed, where the subsequent decompositions of \mathbf{H} and $\tilde{\mathbf{W}}$ are performed separately to achieve better numerical accuracy [Paige90] [Van76]. References [Anderson91], [De92] and [Paige81] further discuss the implementation, application and extension of these generalized methods. In addition, how to decompose \mathbf{W} into the symmetric form (1.19) is also an issue. In principle, $\tilde{\mathbf{W}}$ can be the square root of \mathbf{W} , which is unique since \mathbf{W} is PD. The square root can be obtained by orthogonal decomposition, Jordan decomposition, Denman-Beavers iteration, the Babylonian method and so on [Higham86]. Particularly, $\tilde{\mathbf{W}}$ can also be a lower triangular matrix. Choleksy decomposition can be used. The computation can utilize $\tilde{\mathbf{W}}$'s triangular structure.

In addition, in practical applications, constraints may be imposed to the LS solutions. For example, in curve fitting, inequality constraints related to monotonicity, nonnegativity, and convexity and equality constraints related to continuity and smoothness may be involved [Zhu&Li07]. In the category of linear LS, linear-equality (LE) constrained problems are widely investigated, in which problem (1.1) is subject to a set of consistent LE constraints as

$$\mathbf{A} \mathbf{x} = \mathbf{B} \quad (1.20)$$

with $\mathbf{A} \in \mathbb{R}^{N_A \times N}$ and (without loss of generality) $\text{rank}(\mathbf{A}) = N_A$. One natural way to handle the LE-constrained LS is direct elimination, with which \mathbf{x} is reduced to a lower dimensional vector since the constraints imply that N_A components of \mathbf{x} are always linear combinations of the left $N - N_A$. Correspondingly, the original problem is then converted to an unconstrained LS problem with reduced dimensions equivalently [Bjorck67] [Lawson95]. The most popular practical methods to solve LE-constrained problems are based on the introduction of Lagrange multiplier [Chong07]. The weighting method is also widely adopted, where each LE constraint is treated as an observation with a “huge” weight [Anda96]. Although this method is very easy to implement, it may lead to poor numerical condition. The LE-constrained LS solution can also be obtained using the null-space method [Leringe70], based on which the close form of the solution can be:

$$\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{B} + [\mathbf{H}(\mathbf{I} - \mathbf{A}^+ \mathbf{A})]^+ (\mathbf{z} - \mathbf{H} \mathbf{A}^+ \mathbf{B}) \quad (1.21)$$

where \mathbf{A}^+ stands for the Moore Penrose generalized inverse of \mathbf{A} . If $\text{rank}([\mathbf{A}^T \ \mathbf{H}^T]^T) = N$, (1.21) is the unique solution; otherwise, (1.21) leads to the minimum-norm solution in rank-deficient problems [Wei92a] [Wei92b]. [Zhu&Li07] gives another null-space based form as in (24), which is a useful tool for the subsequent derivations in this dissertation. In addition, other techniques, such as the generalized SVD [Van85], are also introduced to solve and analyze the LE-constrained LS problem. In this dissertation, our purpose is to develop completely recursive LS which provides solutions (theoretically) identical to the batch ones. The above batch methods will provide a solid foundation for the subsequent development.

1.3 Recursive Approaches

1.3.1 Recursive Methods

The above batch methods, such as QRD, SVD and normal-equation methods, can be implemented recursively [Apolinario09]. That is, the current solution can be obtained by updating the previously-processed one (using old observation data) with new observations.

In those observation-coefficient-matrix decomposition based methods, recursive procedures mainly aim to construct the orthogonal matrix \mathbf{Q} recursively [Gill74]. For instance, in the Givens-rotation QRD methods, the rotations of the new observation coefficient can naturally take advantage of the existing upper triangular matrix, where the computation complexity is at $O(N^2)$ per cycle [Yang92]. The Gram-Schmidt decomposition can also be performed recursively in a stable way with $O(MN)$ per-cycle computational operations [Daniel76]. The recursion is still applicable to SVD, but the updating computation requires $O(MN^2)$ flops at each cycle [Bunch78], which is too much compared with recursive Givens rotation and Gram-Schmidt orthogonalization.

Particularly, the matrix-inversion (MI) lemma based RLS, which is a recursive normal-equation method, can obtain \mathbf{C} (and $\hat{\mathbf{x}}$) in another sequential way [Woodbury50] [Chen85]. We will further investigate the MI-lemma-based RLS and generalize it to solve GLS problems. The proposed recursive GLS (RGLS) techniques are also applicable to the QRD-RLS.

Concretely, initialized by

$$\mathbf{C}_{M_0} = (\sum_{m'=1}^{M_0} H_{m'}^T H_{m'}) = (\mathbf{H}_{M_0}^T \mathbf{H}_{M_0})^{-1} \quad (1.22)$$

the EEC-like matrix \mathbf{C} of the unconstrained LS problem (1.2c) can be computed exactly by the following recursion cycle:

$$\mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{C}_{m-1} H_m^T (H_m \mathbf{C}_{m-1} H_m^T + 1)^{-1} H_m \mathbf{C}_{m-1} \quad (1.23)$$

where M_0 is the number of initial observations and the recursion/data index $m > M_0$. $\hat{\mathbf{x}}_m$ can be calculated concurrently.

Furthermore, when a set of consistent LE constraints as in (1.20) is imposed, the unique solution exists iff $\text{rank}([\mathbf{A}^T \ \mathbf{H}^T]^T) = N$:

$$\begin{cases} \mathbf{C} = \mathbf{U}[\mathbf{U}^T (\sum_{m=1}^M H_m^T H_m) \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{B} + \mathbf{C} \mathbf{H}^T (\mathbf{z} - \mathbf{H} \mathbf{A}^+ \mathbf{B}) \end{cases} \quad (1.24)$$

where $\mathbf{H} = \mathbf{H}_M = [H_1^T \cdots H_M^T]^T$, $\mathbf{z} = \mathbf{z}_M = [z_1 \cdots z_M]^T$, \mathbf{U} satisfies $[\tilde{\mathbf{U}} \mathbf{U}][\tilde{\mathbf{U}} \mathbf{U}]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}) = \text{col}(\mathbf{A}^T)$ [Zhu&Li07]. Here “ $\text{col}(\mathbf{X})$ ” denotes the space spanned by all the columns of \mathbf{X} . Reference [Zhu&Li07] also shows that, the recursion formula in (1.23) is still applicable for the LE case once the LE constraint (1.20) has been imposed on the initialization appropriately. For instance, the recursion procedure for the EEC-like matrix should be initialized as

$$\mathbf{C}_{M_0} = \mathbf{U}[\mathbf{U}^T (\sum_{m'=1}^{M_0} H_{m'}^T H_{m'}) \mathbf{U}]^{-1} \mathbf{U}^T \quad (1.25)$$

where the iff condition $\text{rank}([\mathbf{A}^T \mathbf{H}_{M_0}^T]^T) = N$ is implicitly satisfied.

In addition, fast RLS methods are also developed [Ljung78] [Cioffi84], where the computation is reduced from some convenient properties of the data, such as the involved matrices' Toeplitz structure.

The RLS is particularly suitable for real-time applications since sequential algebraic operations at each cycle require low computation as well as fixed storage [Zhou02]. It has been widely applied to such areas as signal processing, control and communication [Mikeles07]. In adaptive-filtering applications, “RLS algorithms” are even referred in particular to RLS-based algorithms for problems with fading-memory weights [Haykin01]. As a normal-equation method, a major disadvantage of the RLS methods is that they have relatively poor numerical stability (for getting $\hat{\mathbf{x}}$), compared with direct observation-function coefficient factorization methods. Fortunately, recursive QRD methods, such as Givens rotations, can be combined to improve numerical stability [Proudlar88] [Cioffi90a] [Cioffi90b] [Li07].

1.3.2 RLS Initialization and Deficient-Rank Problems

In the MI-lemma based RLS, to guarantee that the LS solutions at each recursion cycle are identical to the corresponding batch ones, the procedure needs to start from an exact LS initialization, which leads to the RLS initialization problem. However, although RLS has been well studied and applied in the past decades, less attention has been paid to the RLS initialization. It is because RLS is mainly applied to low-dimensional and high-redundancy

problems in areas such as signal processing, where the calculations for the batch LS solutions based on a small number of initial observations may not be so costly in quite a few cases.

However, a simple initialization is still desired. In the original work, the following approximation is adopted [Albert65]:

$$\begin{cases} \mathbf{C}_0 = \alpha^{-1} \mathbf{I} \\ \hat{\mathbf{x}}_0 = \mathbf{0} \end{cases} \quad (1.26)$$

where α is a tiny positive number. With (1.26), the recursion (1.23) for the unconstrained RLS can start from the first piece of observation data. It is clear that the recursion initialized by (1.26) leads to the exact LS solution iff $\alpha \rightarrow 0$. Although it is hard to implement such an α^{-1} exactly, (1.26) is widely adopted where the effect of the approximate α may be trivial when observations keep accumulating. However, in some practical applications, the negative effect caused by $\alpha \neq 0$ may not be ignored and a too small α may deteriorate numerical conditions. In [Hubing91], an exact initialization scheme is studied, which makes full use of the special form of the initial observations in recursion-based adaptive filtering:

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} x_2 \\ x_1 \\ 0 \\ \vdots \end{bmatrix}, \quad \dots, \quad \mathbf{x}_N = \begin{bmatrix} x_N \\ x_{N-1} \\ \vdots \\ x_1 \end{bmatrix} \quad (1.27)$$

Furthermore, recursive QR-decomposition methods can also be used to transform the coefficient matrix of initial observations into the upper triangular form, based on which the initialization is recursive and easier to be determined [Haykin01]. Reference [Zhou02] introduces variants of the Greville formula (order recursion) to develop recursive initializations for RLS, where the recursion can also start from the first piece of data. We will study a simple recursive exact RLS initialization method using the RLS formulae. We will also apply the newly-developed method to solve high-dimensional and low-redundancy problems such as power system state estimation, where recursive initialization does play an important role in solving the LS state estimation problems. In addition, two accessorial issues should also be studied. One is whether and when the foregoing observations can support an exact and

unique RLS initialization, namely (parameter) observability analysis (or rank check) in engineering. In previous work, observability analysis usually requires extra numerical or topological analysis [Chan87] [Chan94] [Monticelli00]. Correspondingly, if all the observations can not uniquely determine a LS solution, then it leads to deficient-rank LS problems. There are two possible deficient-rank situations: (a) there are no sufficient observations, so that some to-be-determined quantities can not be uniquely determined (or the estimand is unobservable); (b) although the observations are sufficient in theory, the estimand is numerically unobservable due to ill conditions caused by round-off error, e.g., the involved matrix inverses exist in theory but can not be computed numerically. The first situation is theoretical rank deficiency while the second one is numerical rank deficiency [Stewart84]. In addition, the other ill-conditioned case is also treated as numerical rank deficiency, in which the observations are not sufficient in theory but still make the estimand observable due to the effect of round-off error. Topological analysis can detect theoretical deficiency of observations [Monticelli00] while numerical analysis can disclose the implementation details. The latter is well investigated in the past. For instance, SVD based method is used to determine the numerical rank of a matrix, where the singular values in Σ reveal the numerical condition of the LS problem [Manteuffel81]. Cholesky decomposition and QRD methods, such as Householder transformation, Givens rotation and modified Gram-Schmidt orthogonalizations, are also used to examine the numerical rank of the LS problem, where column pivoting is widely used [Golub65a]. Among these decomposition based methods, the SVD is the most reliable one to reveal the matrix rank in general [Bjorck96]. Once numerical rank deficiency is detected, caution has to be taken to fix or avoid this ill-conditioned situation [Dongarra79].

The second issue is how to handle problems with theoretically insufficient observations. Note that, at this stage, the concepts of solving LS problem and performing LS estimation are different. For a deficient-rank LS problem, there exist infinitely many feasible solutions to minimize the squared sum, among which the one having the minimum norm and a simple analytical form in batch is widely adopted [Lawson95] [Zhou02]. However, for LS estimation,

a deficient-rank situation means the estimand is not observable and no estimate exists. We treat the deficient-rank problem from the viewpoint of LS estimation and introduce a reduced-dimensional alternative estimate which is observable and has specific physical interpretations.

1.4 Completely Recursive Least Squares (CRLS)

The RLS is originally applied to solve unconstrained LS problems [Albert65]. Recently, [Zhu&Li07] found that the exact solution of LE-constrained LS can be obtained by the same recursion as for the unconstrained problem, provided that the RLS procedure is appropriately initialized. Owing to the excellent performance of the RLS regarding efficient real-time computation and low memory, it is worthwhile to apply the RLS techniques to perform the initialization of RLS recursively. It is also of value to generalize the conventional RLS as well as the corresponding recursive initialization method and develop an integrated solution for LE-constrained GLS problems. Consequently, there are two major issues worth studying. The first issue is how to generalize the RLS method and solve the WLS and GLS problems in similar recursive/sequential ways. It is clear that the square-root values of the diagonal elements of the diagonal weighting matrix in the WLS can be simply multiplied into the coefficient matrices. The WLS problem is thus converted to an LS one. In other words, a recursive WLS (RWLS) method can be developed from the RLS easily. Therefore, the major focus will be on the development of the recursive GLS (RGLS). The second issue is a recursive initialization applicable to all the RLS, RWLS and RGLS. Actually, in the RLS/RWLS and the newly-developed RGLS, the determination of exact solutions relies on appropriate initializations. We will study a simpler recursive initialization method for the RGLS as well as the RLS/RWLS, which is still based on the RLS formulae. In brief, with such a generalized RLS approach, the recursion can start from the first piece of data no matter what dependence the newly-arriving data have on the old data. The recursion can also be implemented for both unconstrained and explicitly/implicitly LE-constrained cases. In these senses, the newly-developed approach is named *Completely Recursive Least Squares* (CRLS).

1.5 CRLS, LMMSE and KF

In statistics and signal processing, the widely-used linear minimum mean-square error (LMMSE) estimator is a linear function, in fact, an affine function of the observation that achieves the smallest means-square error among all linear/affine estimators [Johnson04]. LMMSE estimator is the theoretical fundament of linear filtering: Kalman filtering, LMMSE filtering (for nonlinear problems), (steady-state) Wiener filtering, and so on [Li07].

As is known in [Li07] that, given a linear data model, an LMMSE estimator with complete/partial prior knowledge (i.e., prior mean, prior covariance and the cross- covariance between the random estimand and the measurement noise) can always be treated as the LMMSE estimator without prior information by unifying the prior mean as extra data. It is also true that a linear-data-model-based LMMSE estimator without prior knowledge may be a unification of Bayesian and classic linear estimation [Li07]. In other words, a linear-data-model-based LMMSE estimator for a random estimand may be mathematically identical to a linear WLS/GLS estimator using the PD joint covariance of the estimand and the measurement noise as the weight inverse. In addition, the GLS with a positive semi-definite (PSD) observation-error-covariance like (OEC-like) matrix has not been well posed, although the LMMSE estimator with PSD measurement-noise covariance has been studied. We will convert the LS with a PSD OEC-like matrix to an LE constrained GLS with implicit constraint. As a result, the linear-data-model-based LMMSE estimation with a PSD measurement-noise covariance, or more generally, the linear-data-model-based LMMSE estimator with a unified PSD joint covariance of the estimand and the measurement noise can also be obtained by solving the corresponding LS problems.

Furthermore, in the Kalman filter applied to linear systems with linear measurements, the prediction is from an LMMSE estimator using the data up to the most recent time while the update can be implemented from another LMMSE estimator using all the data up to the current time. Owing to the mathematical equivalence between the linear-data-model based LMMSE (without prior and with PD measurement error covariance) estimator and the GLS, the CRLS

can be applied to:

- 1) Verify the optimal Kalman filter accounting for the correlation between the prediction error and the measurement noise, which was first derived in the LMMSE sense [Li07].

- 2) Apply the CRLS to improve the sequential data-processing scheme for the optimal Kalman filter and to deal with various complicated situations caused by data correlation, PSD covariance, etc.

Furthermore, we will apply the correlation-accounting KF to develop a series of adaptive filtering techniques and to solve practical problems such as power system state estimation and parameter estimation.

1.6 Power System State Estimation and Parameter Estimation

Power system state estimation, introduced to power systems in 1960s [Schweppe69], is to estimate the involved bus voltages of a power system under steady-state conditions using real-time voltage/power/current data collected by the supervisor control and data acquisition (SCADA) system, where the steady-state system is usually modeled as a single-frequency, balanced and symmetric one and the measured quantities, such as voltage magnitude, active/reactive branch flow and active/reactive injection power, are all linear/nonlinear functions of the bus voltages (system state) [Monticelli00] [Meliopoulos01]. Most of the existing power system state estimation (SE) programs are formulated as static WLS problems with one-scan data [Monticelli00]. Dynamic state estimation (DSE) is not popularly applied due to practical limitations such as the complexity of measurement system and the inaccuracy of dynamic and measurement models. Parameter estimation (PE) is responsible for calibrating the suspicious measurement model parameters [Abur04], within which the bus voltages of interest and the unknown parameters are usually stacked as an augmented state [Zarco00]. Correspondingly, dynamic-estimation methods are preferred since they exploit data from multiple scans and take advantage of dynamic models [Leite87]. Unfortunately, similar obstacles as in the DSE for bus voltages are encountered and the estimation accuracy is not guaranteed. These dilemmas can be avoided in power systems metered by synchrophasors.

The invention of synchrophasor, also known as synchronized phasor measurement unit (PMU), has led a revolution in SE since it yields linear measurement functions as well as accurate data within three to five cycles [Phadke93]. In spite of the involved instrumental channel errors [Sakis07] and the high cost, PMU has been tentatively used in centralized or distributed estimators [Phake86] [Zhao05] [Jiang07] and in bad-data detection [Chen06].

We aim at performing accurate parameter and state estimation in complex situations using synchrophasor data. An approach of joint state-and-parameter estimation, which is different from the state augmentation, is adopted, where the original nonlinear PE problem is reformulated as two loosely-coupled linear subproblems: state tracking and parameter tracking. First, as a result of the reformulation, the state tracking with possible abrupt voltage changes and correlated prediction-measurement errors is investigated, which can be applied to determine the voltages in a PE problem or to estimate the system state in a conventional DSE problem. Second, the parameter calibration of transmission network is also studied. For this high-dimension low-redundancy nonlinear parameter estimation, we propose a balanced method which adopts merits from both the extended Kalman filter (EKF) and the particle filter (PF). It follows the simple structure of EKF but further accounts for the uncertain effects such as involved bus voltages and high-order terms (of Taylor's expansion) in EKF as pseudo measurement errors correlated with prediction errors. Correspondingly, the recently-developed optimal filtering technique that can handle correlation is introduced. We also introduce random samples from the idea of PF to evolve the pseudo-error ensembles and to evaluate the statistics related to the pseudo errors, where the error-ensemble sampling does not rely on the measurement redundancy and is much easier to implement than PF. Based on this balanced method, the joint state-and-parameter estimation considering complicated behavior of voltages and parameters is discussed.

1.7 Our Work and Novelties

As reviewed in Sec. 1.3, the RLS is of a recursive form and free of matrix inversion, and thus is excellent regarding efficient real-time computation and low memory. It is worthwhile to

generalize the above RLS procedure and to solve the unconstrained/LE-constrained GLS problem in a similar recursive way. It is also of value to apply the RLS method for all the involved RLS initializations. Consequently, the work of this dissertation includes: 1) to extend the use of conventional to solve GLS problems in a similar recursive way; 2) to find efficient recursive methods to initialize the RLS as well as the newly-developed recursive GLS procedures; 3) to use the unified recursive GLS to improve the corresponding sequential-data-processing procedures of the optimal KF considering various correlations; 4) to exploit correlation-accounting KF based adaptive filtering approaches to perform power system state estimation with synchrophasor measurements and treat parameter calibration of the transmission network.

The generalization of the RLS for solving GLS problems is discussed in Chapter 2. Starting from the unconstrained/LE-constrained RLS, we will develop a recursive procedure to solve the unconstrained GLS, develop a similar recursive procedure applicable to the LE-constrained GLS, and show that the LE constraint is in essence a set of special observations free of observation errors and can be processed sequentially in any place in the data sequence. More generally, we will consider recursive ILE-constrained GLS. A unified recursive procedure is developed, which is applicable to ILE-constrained GLS as well as all the unconstrained/LE-constrained LS/WLS/GLS.

In Chapter 3, a recursive exact initialization applicable to all the RLS, RWLS and RGLS, is investigated. This chapter treats the RLS initialization-related issues, including rank check, a convenient method to compute the involved matrix inverse/pseudoinverse, and resolution of underdetermined systems. No extra non-RLS formula but an auxiliary-observation based procedure is utilized. The RLS recursion can start from the first real observation and possible LE constraints are also imposed recursively/sequentially. The rank of the system is checked implicitly. If the rank is full, the initialization and the subsequent RLS cycles can be integrated as a whole to yield exact LS estimates. If the rank is deficient, the procedure provides a mapping from the unobservable (original) estimand to a reduced-dimensional set of alternative

quantities which are linear combinations of the original quantities and uniquely determined. The consequent estimate is a set of refined non-redundant observations. The refinement is lossless in the WLS sense: if new observations are available later, it can take the role of the original data in the recalculation.

As shown in [Li07], the linear-data-model based linear minimum-mean-square-error (LMMSE) estimator without prior with PD measurement-error covariance is mathematically equivalent to the LS problem weighted by the measurement-error covariance. In Chapter 4, we show that the linear-data-model based linear minimum-mean-square-error (LMMSE) estimator can always be calculated from solving a unified ILE-constrained GLS. Consequently, the recursive GLS can be used to improve the sequential procedure of the optimal KF considering various correlations.

In Chapters 5 & 6, we aim at performing accurate parameter (and state) estimation in complex situations using synchrophasor data, based on the optimal KF accounting for the correlation between the measurement noise and the prediction error. An approach of joint state-and-parameter estimation, which is different from the state augmentation, is adopted, where the original nonlinear PE problem is reformulated as two loosely-coupled linear subproblems: state tracking and parameter tracking, respectively.

Chapter 5 focuses on the state tracking, which can be used to determine bus voltages in parameter estimation or to track the system state (dynamic state estimation). Dynamic behavior of bus voltages under possible abrupt changes is studied, using a novel and accurate prediction model. The measurement model is also improved. An adaptive filter based on optimal tracking with correlated prediction-measurement errors, including the module for abrupt-change detection and estimation, is developed. With the above settings, accurate solutions are obtained.

In Chapter 6, we study the parameter tracking and the techniques dealing with the coupling. A new prediction model for parameters with moving means is adopted. The uncertainty in the voltages is covered by pseudo measurement errors resulting in prediction-measurement-error

correlation. An error-ensemble-evolution method is proposed to evaluate the correlation. An adaptive filter based on the optimal filtering with the evaluated correlation is developed, where a sliding-window method is used to detect and adapt the moving tendency of parameters. Simulations indicate that the proposed approach yields accurate parameter estimates and improves the accuracy of the state estimation, compared with existing methods.

In brief, our contributions include:

- 1) Combining RLS formulae with a recursive decorrelation method and developing a recursive procedure for GLS;
- 2) Showing that an LE constraint is in essence a special observation free of error and can be processed using RLS formulae at any place in the observation sequence;
- 3) Developing a unified recursive GLS procedure which is also used for ILE-constrained GLS;
- 4) Designing a simple auxiliary observation based RLS initialization procedure which allows the recursion to start from the first piece of real observation data, where no extra technique but the RLS formulae is used;
- 5) Inventing a new method to handle a deficient-rank LS problem, with which a set of reduced-dimensional alternative estimates is provided for practical use;
- 6) When introducing the optimal KF accounting for prediction-measurement-error correlation to solve joint state and parameter estimation in power systems monitored by synchrophasors, we separate the original nonlinear problem as two coupled linear subproblems of state tracking and parameter tracking.
- 7) In state tracking, we propose a new pair of prediction model and measurement model; Develop a filtering method which can also detect the abrupt change; develop a new adaptive filtering algorithm based on optimal tracking with correlated prediction-measurement errors, including the module for the abrupt-change detection.
- 8) In parameter tracking, a new prediction model accounting for the effect of prior knowledge and moving parameter means is proposed; a new adaptive filter is developed, based on the optimal filtering with correlated prediction-measurement errors; a sliding-window method is

proposed to detect the moving tendency of parameters and adjust the transition matrix adaptively; a sample-based method, namely, error-ensemble evolution, is used to evaluate the correlation between pseudo measurement errors and prediction errors.

1.8 Outline

The rest of the dissertation is organized as follows. In Chapter 2, for the first part of the CRLS approach, we will investigate the unified recursive GLS. The second part of CRLS, which is on the exact recursive initialization and deficient-rank processing, is presented in Chapter 3. In Chapter 4, the CRLS is then applied to verify the optimal KF (based on the LMMSE criterion) which can handle the correlation between the prediction error and the measurement noise. In Chapters 5 and 6, the correlation-accounting KF is applied to develop a series of nonlinear filtering techniques to solve the problem of joint state and parameter estimation in power systems metered by synchrophasors, where the original joint problem is divided into two coupled subproblems as state tracking and parameter tracking.

Nomenclature in Chapters 2-3

The major notations used in Chapters 2 and 3 are listed below for quick reference.

i. Variables and Numbers

C	estimation-error-covariance like (EEC-like) matrix
H	(observation) coefficient (vector) in real-observation function
H	(observation) coefficient (matrix) and $\mathbf{H} = \begin{bmatrix} \vdots \\ H_m \\ \vdots \end{bmatrix}$
R	observation-error-covariance-like (OEC-like) matrix; weight inverse if PSD
x	estimand, full-/reduced-dimensional vector of to-be-determined quantities
z	observation
z	observation vector containing all z 's
M	total number of observations
N	dimension of the estimand, row number of x
T	total number of constraints

ii. Overlines

Let X be an original vector/matrix, then

\hat{X}	estimated X
\tilde{X}	augmentation from X
\tilde{X}	temporarily-used quantities
X'	after an LRC decorrelation
X''	related to the maximum-rank PD principal minor
X^D	used in an LRC decorrelation
X^*	related to observations excluding implicit constraints

X° related to implicit LE constraints

iii. Subscripts and Indices

au	of auxiliary observations
b	of the (selected) basis of the row space of $\tilde{\mathbf{H}}$
lec	linear-equality constrained
left	of the left (undeleted) auxiliary observations
sim	of a simple basis mapping \mathbf{x} to an \mathbf{x}_{sim} containing \mathbf{x}_{ud}
tran	after an equivalent linear transformation
uc	unconstrained
k	time index in Kalman filter
\tilde{m}	recursion index
m	observation (data) index
s	auxiliary-observation index
t	constraint index

In addition, notations not listed here are for temporary use only and are thus explained where they first appear.

Chapter 2: Completely Recursive Least Squares—Part I: Unified Recursive Solutions to Generalized Least Squares

2.1 Background

It is well known that, if and only if $\text{rank}(\mathbf{H}) = N$, the following LS problem has a unique solution:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J \quad (2.1)$$

$$\text{with } J = \sum_{m=1}^M (z_m - H_m \mathbf{x})(z_m - H_m \mathbf{x}) \quad (2.2)$$

where estimand \mathbf{x} is a vector containing all the N to-be-determined variables and $\hat{\mathbf{x}}$ the estimated \mathbf{x} . scalar z is observation and H observation coefficient (vector) in real-observation function. Observation-value vector \mathbf{z} contains all z 's and the rows of observation coefficient (matrix) \mathbf{H} contain the corresponding H 's. The total number of observation is M and m is observation (data) index.

The unique solution can be determined using a recursive procedure described by the following fact:

Fact 1 (Unconstrained RLS [Bjorck96]): In the unconstrained LS problem (2.1), if $M_0 < M$ and $\text{rank}(\mathbf{H}_{M_0}) = N$, the problem has the following recursive solution:

$$\begin{cases} \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \\ \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z_m - H_m \hat{\mathbf{x}}_{m-1}) \end{cases} \quad (2.3)$$

for $M_0 < m \leq M$ with

$$\begin{cases} \mathbf{S}_m \triangleq \mathbf{H}_m \mathbf{C}_{m-1} \mathbf{H}_m^T + 1 \\ \mathbf{K}_m \triangleq \mathbf{C}_{m-1} \mathbf{H}_m^T \mathbf{S}_m^{-1} \end{cases} \quad (2.4)$$

and

$$\begin{cases} \mathbf{C}_{M_0} \triangleq (\mathbf{H}_{M_0}^T \mathbf{H}_{M_0})^{-1} \\ \hat{\mathbf{x}}_{M_0} \triangleq \mathbf{C}_{M_0} \mathbf{H}_{M_0}^T \mathbf{z}_{M_0} \end{cases} \quad (2.5)$$

where M_0 is the number of initial observations uniquely determining the values in (2.5).

Furthermore, problem (2.1) may be subject to a set of consistent LE constraints as

$$\mathbf{Ax} = \mathbf{B} \quad (2.6)$$

with $\mathbf{A} \in \mathbb{R}^{N_A \times N}$ and (without loss of generality) $\text{rank}(\mathbf{A}) = N_A$. Correspondingly, the unique solution, which exists iff $\text{rank}([\mathbf{A}^T \ \mathbf{H}^T]^T) = N$, can be obtained recursively as follows:

Fact 2 (LE-constrained RLS [Zhu&Li07]): In the LE-constrained LS problem (2.1) subject to (2.6), if $M_0 < M$ and $\text{rank}(\tilde{\mathbf{H}}_{M_0}) = N$, the problem has the following recursive solution:

$$\begin{cases} \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \\ \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z_m - H_m \hat{\mathbf{x}}_{m-1}) \end{cases} \quad (2.7)$$

for $M_0 < m \leq M$ with

$$\begin{cases} \mathbf{S}_m \triangleq H_m \mathbf{C}_{m-1} H_m^T + 1 \\ \mathbf{K}_m \triangleq \mathbf{C}_{m-1} H_m^T \mathbf{S}_m^{-1} \end{cases} \quad (2.8)$$

and

$$\begin{cases} \mathbf{C}_{M_0} \triangleq \mathbf{U}[\mathbf{U}^T \mathbf{H}_{M_0}^T \mathbf{H}_{M_0} \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}}_{M_0} \triangleq \mathbf{A}^+ \mathbf{B} + \mathbf{C}_{M_0} \mathbf{H}_{M_0}^T [\mathbf{z}_{M_0} - \mathbf{H}_{M_0} \mathbf{A}^+ \mathbf{B}] \end{cases} \quad (2.9)$$

where $\tilde{\mathbf{H}}_{M_0} = [\mathbf{A}^T \ \mathbf{H}_{M_0}^T]^T$.

As shown in [Zhu&Li07], the above two recursive solutions have the same procedure except that the initializations are different. As reviewed in Chapter 1, the RLS is of a recursive form and free of matrix inversion, so it has efficient real-time computation and low memory storage. It is worthwhile to generalize the above RLS procedure and to solve the unconstrained/LE-constrained GLS problems in a similar recursive way. It is also of value to apply the RLS method for the involved initialization. Consequently, there are two major issues: (a) generalization of the RLS to solve GLS problems in similar recursive/sequential ways, which is discussed in this chapter; (b) recursive initialization applicable to the unified RLS, which is to be handled in Chapter 3.

2.2 Problem Formulations

For the convenience of description and practical applications, rather than the direct weighting matrix, an observation-error-covariance-like (OEC-like) matrix is preferred. In essence, an OEC-like matrix is the weight inverse if it is positive-definite (PD). In the reverse, a positive-semi-definite (PSD) OEC-like matrix means that the constraint data have not been explicitly distinguished from the observation data, which leads to the problem of LS with implicit LE constraint (ILE constrained LS).

The discussion begins with GLS with a PD OEC-like matrix. That is, given a PD matrix \mathbf{R} , the linear GLS weighted by \mathbf{R}^{-1} is to solve

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J \quad (2.10)$$

$$\text{with } J = (\mathbf{z} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{H}\mathbf{x}) \quad (2.11)$$

In general, \mathbf{R} can be a non-diagonal matrix. That is, for $1 < m \leq M$,

$$\mathbf{R}_m = \begin{bmatrix} \mathbf{R}_{m-1} & R_m \\ R_m^T & r_m \end{bmatrix} \quad (2.12)$$

First, when there is no constraint, the solution to problem (2.10) can be determined from the following normal equation:

$$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \hat{\mathbf{x}} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} \quad (2.13)$$

Second, when there is a set of (consistent) LE constraints

$$\mathbf{A}\mathbf{x} = \mathbf{B} \quad (2.14)$$

with $\mathbf{A} \in \mathbb{R}^{N_A \times N}$ and (without loss of generality) $\text{rank}(\mathbf{A}) = N_A$, the normal equation becomes

$$\begin{bmatrix} (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} \\ \mathbf{B} \end{bmatrix} \quad (2.15)$$

where $\boldsymbol{\lambda}$ is a Lagrange multiplier.

Subsequently, we aim to solving (2.13) and (2.15) in a recursive way similar to RLS, where a recursive decorrelation technique is adopted to perform a last-row-column decorrelation to

make \mathbf{R}_m diagonal prior to the application of the RLS formulae. Specifically, starting from the unconstrained/LE-constrained RLS, we will

- 1) Develop a recursive procedure to solve the unconstrained GLS as in (2.13);
- 2) Develop a similar recursive procedure applicable to the LE-constrained GLS as in (2.15);
- 3) Show that the LE constraint is in essence a set of special observations free of observation errors and can be processed sequentially in any place in the data sequence;
- 4) More generally, we will consider the GLS with implicit LE constraint (ILE-constrained GLS) in which \mathbf{R} is PSD. A unified recursive procedure is developed, which is applicable to ILE-constrained GLS as well as all the unconstrained/LE-constrained LS/WLS/GLS.

2.3 Theoretical Foundation and Results

The theoretical foundations and derivations for the recursive GLS are presented as follows.

2.3.1 Preliminaries

The following important identities and lemmas are crucial to the development of the recursive GLS.

Schur's Identity: If inverses of \mathbf{D} , $(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})$, \mathbf{G} and $(\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F})$ are legitimate, then

$$\begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{O} & \mathbf{P} \end{bmatrix} \quad (2.16)$$

where

$$\mathbf{M} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{E}\mathbf{P}\mathbf{F}\mathbf{D}^{-1} = (\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F})^{-1} \quad (2.17)$$

$$\mathbf{N} = -\mathbf{D}^{-1}\mathbf{E}\mathbf{P} = -\mathbf{M}\mathbf{E}\mathbf{G}^{-1} \quad (2.18)$$

$$\mathbf{O} = -\mathbf{P}\mathbf{F}\mathbf{D}^{-1} = -\mathbf{G}^{-1}\mathbf{F}\mathbf{M} \quad (2.19)$$

$$\mathbf{P} = (\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1} = \mathbf{G}^{-1} + \mathbf{G}^{-1}\mathbf{F}\mathbf{M}\mathbf{E}\mathbf{G}^{-1} \quad (2.20)$$

This well-known identity can be verified via blockwise elimination [Woodbury50]. Particularly, another useful identity, described as matrix inversion (MI) lemma, follows by substituting (2.20) into (2.18):

MI Lemma: If inverses of matrices \mathbf{D} , \mathbf{G} and $(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})$ exist, then

$$(\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F})^{-1} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{E}(\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E})^{-1}\mathbf{F}\mathbf{D}^{-1} \quad (2.21)$$

The MI lemma has the following important corollary which is the basis of the conventional RLS:

Corollary of MI Lemma: If inverses of matrices \mathbf{D} , \mathbf{G} and $(\mathbf{G} + \mathbf{E}^T\mathbf{D}^{-1}\mathbf{E})$ exist, then

$$(\mathbf{D} + \mathbf{E}^T\mathbf{G}^{-1}\mathbf{E})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{E}^T(\mathbf{E}\mathbf{D}^{-1}\mathbf{E}^T + \mathbf{G})^{-1}\mathbf{E}\mathbf{D}^{-1} \quad (2.22)$$

Actually, this corollary contains such a bidirectional causal relation as: Given \mathbf{D}^{-1} and \mathbf{G}^{-1} , the existence of $(\mathbf{G} - \mathbf{E}^T\mathbf{D}^{-1}\mathbf{E})^{-1}$ is equivalent to the existence of $(\mathbf{D} + \mathbf{E}^T\mathbf{G}^{-1}\mathbf{E})^{-1}$, which is an important basis for Facts 1 and 2 regarding the conventional unconstrained and LE-constrained RLS solutions, respectively (see the Introduction).

2.3.2 Recursive solutions to unconstrained GLS

First, the recursive solution to the unconstrained GLS is investigated, which is identical to the following batch one:

Fact 3 (Batch solution to unconstrained GLS): Iff $\text{rank}(\mathbf{H}) = N$, the normal equation (2.13) has a unique solution:

$$\begin{cases} \mathbf{C} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \\ \hat{\mathbf{x}} = \mathbf{C} \mathbf{H}^T \mathbf{W} \mathbf{z} \end{cases} \quad (2.23)$$

In general, iff $\text{rank}(\mathbf{H}_m) = N$, the GLS problem with data up to m has a unique solution:

$$\begin{cases} \mathbf{C}_m = (\mathbf{H}_m^T \mathbf{W}_m \mathbf{H}_m)^{-1} \\ \hat{\mathbf{x}}_m = \mathbf{C}_m \mathbf{H}_m^T \mathbf{W}_m \mathbf{z}_m \end{cases} \quad (2.24)$$

The batch solutions, as in (2.23) and (2.24), can be computed in the same recursive way as the conventional RLS as long as the following decorrelation is applied.

Definition 1 (Last-row-column (LRC) decorrelation): Given PD \mathbf{R}_m (2.12), the following pair of nonsingular matrices \mathbf{Q}_m and \mathbf{Q}_m^T can diagonalize the last row and column of \mathbf{R}_m :

$$\begin{aligned}\mathbf{R}'_m &= \mathbf{Q}_m^T \mathbf{R}_m \mathbf{Q}_m \\ &= \text{diag}\{\mathbf{R}_{m-1}, r_m - D_m^T R_m\}\end{aligned}\quad (2.25)$$

$$\text{with } \mathbf{Q}_m = \begin{bmatrix} \mathbf{I}_{(m-1) \times (m-1)} & -D_m \\ \mathbf{0}^T & 1 \end{bmatrix} \text{ and } D_m = \mathbf{R}_{m-1}^{-1} R_m \quad (2.26)$$

Equation (2.25) can be easily verified. In addition, using a series of successive LRC decorrelations, \mathbf{R}_m can be transformed into a diagonal matrix, as in Proposition 1.

Proposition 1 (Inverse decomposition of a PD matrix): Given PD \mathbf{R}_m (2.12), the following equality holds for $1 < m \leq M$:

$$\mathbf{Q}_{1:m}^T \mathbf{R}_m \mathbf{Q}_{1:m} = \text{diag}\{r_1, r'_2, \dots, r'_m\} \quad (2.27)$$

with

$$\mathbf{Q}_{1:m} = \begin{bmatrix} 1 & -D_2 & \cdots & -D_m \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (2.28a)$$

$$r'_m = r_m - D_m^T R_m \quad (2.28b)$$

$$\text{and } D_m = \mathbf{R}_{m-1}^{-1} R_m \quad (2.28c)$$

A proof of this proposition is given in Appendix A. Note that, although \mathbf{R}_m is transformed into a diagonal matrix, the transformation in (2.27) is conceptually different from the conventional matrix diagonalization where the pairwise transformation matrices are mutual inverses. In fact, the one in (2.27) is in essence an inverse process of the symmetric indefinite factorization which is an alternative form of the Cholesky decomposition (factorization) [Watkins91] [Ogita12]. In previous work as in [Petkovic09] [Karlsson06], this inverse transformation is usually treated as a two-stage procedure of decomposition and inversion [Ogita10]. Proposition 1 provides a method to obtain the inverse decomposition directly. Actually, as shown in (2.25), the transformation in (2.27) is not only a mathematical inversion

process of the Cholesky decomposition. It has specific interpretations of error-correlation deduction in practical applications where \mathbf{R} is set as observation error covariance (see Chapter 4). In particular, it is clear that $\mathbf{Q}_{1:m}$ in (2.28a) can be computed recursively as

$$\mathbf{Q}_{1:m} = \begin{bmatrix} \mathbf{Q}_{1:(m-1)} & -D_m \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{1:(m-1)} & -\mathbf{R}_{m-1}^{-1} R_m \\ 0 & 1 \end{bmatrix} \quad (2.28d)$$

which is similar to the recursive (inverse) Cholesky decomposition [Gustavson00] [Bjarne01] [Bjarne02]. In fact, the recursive computation in (2.28d) can be further simplified. Applying this recursive transformation as well as the RLS formulae to the unconstrained GLS, the recursive solution can thus be developed as follows.

Theorem 1 (Recursive GLS): If $\text{rank}(\mathbf{H}_{M_0}) = N$ and $M_0 < M$, the unconstrained GLS problem (2.10) has the following recursive solution:

$$\begin{cases} \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z'_m - H'_m \hat{\mathbf{x}}_{m-1}) \\ \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{cases} \quad (2.29)$$

for $M_0 < m \leq M$ with

$$\begin{cases} \mathbf{S}_m \triangleq H'_m \mathbf{C}_{m-1} H_m'^T + r'_m \\ \mathbf{K}_m \triangleq \mathbf{C}_{m-1} H_m'^T \mathbf{S}_m^{-1} \end{cases} \quad (2.30)$$

$$\begin{cases} H'_m \square H_m - D_m'^T \mathbf{H}_{m-1}' \\ r'_m \square r_m - D_m'^T R'_m \\ z'_m \square z_m - D_m'^T \mathbf{z}_{m-1}' \end{cases} \quad (2.31a)$$

$$\begin{cases} R'_m = \mathbf{Q}_{1:(m-1)}^T R_m \\ D'_m \square \mathbf{D}_{1:m-1} R'_m \end{cases} \quad (2.31b)$$

$$\mathbf{Q}_{1:m} = \begin{bmatrix} \mathbf{Q}_{1:(m-1)} & -\mathbf{Q}_{1:(m-1)} D'_m \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (2.32a)$$

$$\begin{cases} \mathbf{H}'_m = [\mathbf{H}_{m-1}'^T H_m'^T]^T \\ \mathbf{D}_m = \text{diag}(\mathbf{D}_{1:m-1}, r_m'^{-1}) \\ \mathbf{z}'_m = [\mathbf{z}_{m-1}'^T z_m'^T]^T \end{cases} \quad (2.32b)$$

and

$$\begin{cases} \mathbf{C}_{M_0} \triangleq (\mathbf{H}_{M_0}^T \mathbf{R}_{M_0}^{-1} \mathbf{H}_{M_0})^{-1} \\ \hat{\mathbf{x}}_{M_0} \triangleq \mathbf{C}_{M_0} \mathbf{H}_{M_0}^T \mathbf{R}_{M_0}^{-1} \mathbf{z}_{M_0} \end{cases} \quad (2.33a)$$

$$\begin{cases} \mathbf{D}_{M_0} = \text{diag}\{r_1^{-1}, (r_2 - R_2^T r_1^{-1} R_2)^{-1}, \dots, \\ \dots, (r_{M_0} - R_{M_0}^T r_{M_0-1}^{-1} R_{M_0})^{-1}\} \\ \mathbf{H}'_{M_0} = \mathbf{Q}_{1:M_0}^T \mathbf{H}_{M_0} \\ \mathbf{z}'_{M_0} = \mathbf{Q}_{1:M_0}^T \mathbf{z}_{M_0} \end{cases} \quad (2.33b)$$

$$\mathbf{Q}_{1:M_0} = \begin{bmatrix} 1 & -r_1^{-1} R_2 & \cdots & -\mathbf{R}_{M_0-1}^{-1} R_{M_0} \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (2.33c)$$

A proof of this theorem is given in Appendix B. Compared with the conventional RLS in Fact 1, the recursive solution to the unconstrained GLS has an additional LRC- decorrelation process described by (2.32a) and (2.32b). In reverse, if \mathbf{R}_m is diagonal for all m 's, which means $R_m = \mathbf{0}$, then D'_m is always equal to zero. Correspondingly, (2.31a)-(2.32b) can be omitted and the recursive procedure is degenerated into Fact 1. In other words, the unconstrained RLS is a special case of the solution to the unconstrained GLS.

2.3.3 Recursive solutions to LE-constrained GLS

The recursive procedure to the unconstrained GLS in Theorem 1 can be generalized to solve the LE-constrained GLS. The corresponding solution is identical to the batch one described in the following fact:

Fact 4 (Batch solution to LE-constrained GLS): Let $\tilde{\mathbf{H}} \square [\mathbf{A}^T \ \mathbf{H}^T]^T$, iff $\text{rank}(\tilde{\mathbf{H}}) = N$, the normal equation (2.15) has a unique solution:

$$\begin{cases} \mathbf{C} = \mathbf{U}[\mathbf{U}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{B} + \mathbf{C} \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{z} - \mathbf{H} \mathbf{A}^+ \mathbf{B}] \end{cases} \quad (2.34)$$

where superscript “+” stands for Moore-Penrose pseudo inverse (MP inverse). \mathbf{U} satisfies $[\tilde{\mathbf{U}} \ \mathbf{U}][\tilde{\mathbf{U}} \ \mathbf{U}]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}) = \text{col}(\mathbf{A}^T)$ [Zhu&Li07]. “ $\text{col}(\mathbf{X})$ ” denotes the space spanned by

all the columns of \mathbf{X} . A simple proof for Fact 4 is given in Appendix C. Similarly, iff $\text{rank}(\tilde{\mathbf{H}}_m) = N$, the LE-constrained GLS (2.10) subject to (2.14) with data up to m has a unique solution

$$\begin{cases} \mathbf{C}_m = \mathbf{U}[\mathbf{U}^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{B} + \mathbf{C}_m \mathbf{H}_m^T \mathbf{R}_m^{-1} [\mathbf{z}_m - \mathbf{H}_m \mathbf{A}^+ \mathbf{B}] \end{cases} \quad (2.35)$$

These batch solutions (2.34) and (2.35) can be computed recursively, based on the following theorem.

Theorem 2 (LE-constrained recursive GLS): If $\text{rank}([\mathbf{A}^T \mathbf{H}_{M_0}^T]^T) = N$ and $M_0 < M$, the LE-constrained GLS problem defined by (2.10) subject to (2.14), has the following recursive solution:

$$\begin{cases} \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z'_m - H'_m \hat{\mathbf{x}}_{m-1}) \\ \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{cases} \quad (2.36)$$

for $M_0 < m \leq M$ with

$$\begin{cases} \mathbf{C}_{M_0} \triangleq \mathbf{U}[\mathbf{U}^T \mathbf{H}_{M_0}^T \mathbf{R}_{M_0}^{-1} \mathbf{H}_{M_0} \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}}_{M_0} = \mathbf{A}^+ \mathbf{B} + \mathbf{C}_{M_0} \mathbf{H}_{M_0}^T (\mathbf{z}_{M_0} - \mathbf{H}_{M_0} \mathbf{A}^+ \mathbf{B}) \end{cases} \quad (2.37)$$

where the involved quantities $\mathbf{S}_m, \mathbf{K}_m, D_m, H'_m, r'_m, z'_m, \mathbf{Q}_{1:m}$ and so on can be determined by (2.30)-(2.32b). A proof is given in Appendix D. Clearly, the recursive solution to the LE-constrained GLS problem has the same procedure as the unconstrained one after the constraint is properly imposed on the initialization. It is also a method generalized from the conventional LE-constrained RLS via applying the LRC decorrelation. In fact, as shown in the next subsection, the LE constraint can be treated as a special and simple “observation” and the LE-constrained solution can thus be obtained as simply as the unconstrained one.

2.3.4 Recursive imposition of LE constraints

Furthermore, the constraint imposition in the LE-constrained GLS can be implemented in a more flexible and simpler way, based on the following theorem.

Theorem 3 (Recursive LE constraint imposition): Given $\text{rank}([\mathbf{A}^T \mathbf{H}^T]^T) = N$, the LE-constrained GLS solution \mathbf{C}_{lec} (and $\hat{\mathbf{x}}_{\text{lec}}$) in (2.34) has the following properties:

1) If $\text{rank}(\mathbf{H}) = N$ (unconstrained solution \mathbf{C}_{uc} thus exists), then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{uc}} - \mathbf{C}_{\text{uc}} \mathbf{A}^T (\mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{C}_{\text{uc}} \\ \hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{x}}_{\text{uc}} + \mathbf{C}_{\text{uc}} \mathbf{A}^T (\mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T)^{-1} (\mathbf{B} - \mathbf{A} \hat{\mathbf{x}}_{\text{uc}}) \end{cases} \quad (2.38)$$

2) If $[\mathbf{A}^T \mathbf{H}^T]^T$ is an $N \times N$ matrix, then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{lec}, N \times N} \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{H} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} \end{bmatrix} [\mathbf{A}^T \mathbf{H}^T]^{-1} \right) \\ \hat{\mathbf{x}}_{\text{lec}} = ([\mathbf{A}^T \mathbf{H}^T]^T)^{-1} [\mathbf{B}^T \mathbf{z}^T]^T \end{cases} \quad (2.39a)$$

3) Given $\mathbf{H}_1 = [\mathbf{H} \mathbf{A}_1]$, $\mathbf{W}_1 = \text{diag}(\mathbf{W}, \mathbf{W}_{A1})$ and \mathbf{W}_{A1} is PD, if the rows of \mathbf{A}_1 are all from \mathbf{A} , then

$$\begin{cases} \mathbf{C}_1 = \mathbf{C}_{\text{lec}} \\ \hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_{\text{lec}} \end{cases} \quad (2.39b)$$

where $\hat{\mathbf{x}}_1$ (and \mathbf{C}_1) is the LE-constrained solution which are subject to (2.6) and based on observations data having coefficient \mathbf{H}_1 , weight \mathbf{W}_1 and value $\mathbf{z}_1 = [\mathbf{z}^T \mathbf{z}_{A1}^T]^T$. The added-in auxiliary observation data (See the general definition in Chapter 4) has the same coefficient as the constraint related to \mathbf{A}_1 but the weight \mathbf{W}_{A1} is a given PD matrix. The corresponding \mathbf{z}_{A1} may be unequal to \mathbf{B} .

4) If $\mathbf{A} = [\mathbf{A}_1^T \mathbf{A}_2^T]^T$ and $\text{rank}([\mathbf{H}^T \mathbf{A}_1^T]^T) = N$, then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{lec}, 2} \\ \hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{x}}_{\text{lec}, 2} \end{cases} \quad (2.40a)$$

with

$$\begin{cases} \mathbf{S}_{\text{lec}, 2} \square \mathbf{A}_2 \mathbf{C}_{\text{lec}, 1} \mathbf{A}_2^T, \mathbf{K}_2 \square \mathbf{C}_{\text{lec}, 1} \mathbf{A}_2^T \mathbf{S}_{\text{lec}, 2}^{-1} \\ \mathbf{C}_{\text{lec}, 2} \square \mathbf{C}_{\text{lec}, 1} - \mathbf{K}_2 \mathbf{S}_{\text{lec}, 2} \mathbf{K}_2^T \\ \hat{\mathbf{x}}_{\text{lec}, 2} = \hat{\mathbf{x}}_{\text{lec}, 1} + \mathbf{K}_2 (\mathbf{B}_2 - \mathbf{A}_2 \hat{\mathbf{x}}_{\text{lec}, 1}) \end{cases} \quad (2.40b)$$

and

$$\begin{cases} \mathbf{C}_{\text{lec},1} = \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_1)^{-1} \mathbf{U}_1^T \\ \hat{\mathbf{x}}_{\text{lec},1} = \mathbf{A}_1^+ \mathbf{B}_1 + \mathbf{C}_{\text{lec},1} \mathbf{H}^T (\mathbf{z} - \mathbf{H} \mathbf{A}_1^+ \mathbf{B}_1) \end{cases} \quad (2.40c)$$

where $[\tilde{\mathbf{U}}_1 \ \mathbf{U}_1][\tilde{\mathbf{U}}_1 \ \mathbf{U}_1]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$.

A proof of Theorem 3 is in Appendix E. In this theorem, Statement 1) provides a formula to compute the LE-constrained solution from the corresponding unconstrained solution (if the latter exists). Statement 2) presents the specific solution which is exactly determined by the observation and constraint. Statement 3) shows that those observations which have the same coefficient as the constraint have no effect on the solution. Statement 4) indicates that multiple LE constraints can be imposed sequentially.

In addition, \mathbf{C}_{uc} in Statement 1) and $\mathbf{C}_{\text{lec},1}$ in Statement 4) can both be viewed as special initializations for the recursive GLS, within which no constraint or only a part of constraints, accompanied by a part of observations, are processed. In Statement 4), the constraints related to \mathbf{A}_2 can be inserted into any place in the remaining-observation sequence. It is because: the observations prior to the constraint \mathbf{A}_2 plus the existing initialization form a new augmented initialization with partial constraints; according to statement 4), after being processed, the \mathbf{A}_2 constraints and the previous observations & constraints compose another new initialization. Correspondingly, the fact that the overall procedure leads to the exact GLS solution can be shown by Theorem 2.

In summary, combining Theorems 2 and 3, it can be concluded that the LE constraints can be wholly or partially imposed onto the initialization, and they can also be processed posterior to the initialization if the corresponding unconstrained solution exists. After the initialization, the remaining constraints can be processed sequentially at any place in the to-be-processed observation sequence. In other words, an LE constraint plays a role of a special observation free of error: $r_A = 0$, and the crossing terms between the constraint and other

observations/constraints $R_A = \mathbf{0}$. When the constraint is processed in the recursive GLS,

$$\mathbf{S}_{\text{lec}} = \mathbf{A}\mathbf{C}_{-1}\mathbf{A}^T = \mathbf{A}'\mathbf{C}_{-1}\mathbf{A}'^T = \lim_{r \rightarrow 0} \mathbf{A}\mathbf{C}_{-1}\mathbf{A}^T + r\mathbf{I} \quad (2.41)$$

where \mathbf{C}_{-1} is the unconstrained/LE-constrained covariance-like matrix prior to the current constraint imposition.

2.3.5 Recursive solution to GLS with Implicit LE constraint

According to the above observation-constraint unification, the unified observations in the LE-constrained GLS problem defined by (2.10) subject to (2.14) has $\tilde{\mathbf{H}} = [\mathbf{H}^T \mathbf{A}^T]^T$ and $\tilde{\mathbf{z}} = [\mathbf{z}^T \mathbf{B}^T]^T$, and the corresponding unified weight is

$$\tilde{\mathbf{W}} = \text{diag}\{\mathbf{R}^{-1}, [+ \infty \cdots + \infty]\} \quad (2.42a)$$

One major challenge using $\tilde{\mathbf{W}}$ is that “ $+\infty$ ” may not be exactly expressed in practical applications. Alternatively, if we start from the OEC-like matrix

$$\tilde{\mathbf{R}} = \text{diag}\{\mathbf{R}, \mathbf{0}\} \quad (2.42b)$$

then the unified form of the LE-constrained GLS, as well as the solution, can take advantage of the recursive-constraint-imposition formula (2.41) and can thus avoid the possible numerical problem caused by infinite weights.

More generally, starting from a PSD OEC-like matrix, the unified form of the LE-constrained GLS can be extended to resolve some application problems where the LE constraint is embedded in the data sequence implicitly. Correspondingly, an equivalent explicit GLS formulation and the (recursive) solutions are exploited. This combination of GLS formulation and solution is in essence a GLS problem with a set of implicit LE constraints.

The recursive solution to the GLS with implicit LE constraint (ILE-constrained GLS) is investigated, which is based on the following decorrelation techniques.

Proposition 2 (LRC decorrelation of a special PSD matrix): Consider the following special matrix

$$\mathbf{R}_{a+2} = \begin{bmatrix} \mathbf{R}_{a+1} & R_{a+2} \\ R_{a+2}^T & r_{a+2} \end{bmatrix} \quad (\text{Note: not as generic as } \mathbf{R}_{m=a+1}) \quad (2.43a)$$

with $\mathbf{R}_{a+1} = \begin{bmatrix} \mathbf{R}_a & \mathbf{R}_{a,a+1} \\ \mathbf{R}_{a,a+1}^T & \mathbf{R}_{a+1,a+1} \end{bmatrix}$ and $R_{a+2} = \begin{bmatrix} R_{a+2}''^T & R_{a+1,a+2}^T \end{bmatrix}^T$, where r_{a+2} is a scalar, $\mathbf{R}_{a+1,a+1}$ is a scalar/square matrix, and vector R_{a+2}'' occupies the same rows as \mathbf{R}_a . If \mathbf{R}_{a+1} is PSD, \mathbf{R}_a is PD, and $\text{rank}(\mathbf{R}_{a+1}) = \text{rank}(\mathbf{R}_a)$, then

$$\tilde{\mathbf{Q}}_{a+2}^T \mathbf{R}_{a+2} \tilde{\mathbf{Q}}_{a+2} = \text{diag}(\mathbf{R}_{a+1}, r_{a+2} - R_{a+2}''^T \mathbf{R}_a^{-1} R_{a+2}'') \quad (2.43b)$$

$$\text{with } \tilde{\mathbf{Q}}_{a+2} = \begin{bmatrix} \mathbf{I}_{a+1} & -\mathbf{R}_a^{-1} R_{a+2}'' \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

This proposition discloses that the last row and column of the PSD matrix \mathbf{R}_{a+2} defined in (2.43a) can be diagonalized using \mathbf{R}_a , the leading maximum-rank PD principal minor of \mathbf{R}_{a+1} , and the corresponding cross subvector R_{a+2}'' only. A proof is given in Appendix F.

Furthermore, from [Meyer00], for an arbitrary PSD \mathbf{R}_m as in (2.12), there exists a pair of permutation matrices $\tilde{\mathbf{Q}}_m^T$ and $\tilde{\mathbf{Q}}_m$ with which the orders of the rows and columns of \mathbf{R}_{m-1} can be symmetrically adjusted such that

$$\tilde{\mathbf{Q}}_m^T \mathbf{R}_m \tilde{\mathbf{Q}}_m = \begin{bmatrix} \tilde{\mathbf{R}}_{m-1} & \tilde{R}_m \\ \tilde{R}_m^T & r_m \end{bmatrix} = \tilde{\mathbf{R}}_m \quad \text{with} \quad \tilde{\mathbf{R}}_{m-1} = \begin{bmatrix} \mathbf{R}_{m-1}'' & * \\ * & * \end{bmatrix},$$

and $\text{rank}(\mathbf{R}_{m-1}'') = \text{rank}(\mathbf{R}_{m-1})$, where $\tilde{R}_m = \begin{bmatrix} R_m'' \\ * \end{bmatrix}$ and R_m'' comes from the same rows (of \mathbf{R}_{m-1}) as

\mathbf{R}_{m-1}'' . In other words, \mathbf{R}_{m-1}'' is a maximum-rank PD principal minor of \mathbf{R}_{m-1} and is promoted forwards as in $\tilde{\mathbf{R}}_{m-1}$ (with accordance to Proposition 2). All the not-shown terms are marked by “*”. Applying Proposition 2 to $\tilde{\mathbf{R}}_m$, we have

$$\tilde{\mathbf{Q}}_m'^T \tilde{\mathbf{Q}}_m^T \mathbf{R}_m \tilde{\mathbf{Q}}_m \tilde{\mathbf{Q}}_m' = \text{diag}(\tilde{\mathbf{R}}_{m-1}, r_m') \quad (2.44)$$

$$\text{with } \tilde{\mathbf{Q}}'_m = \begin{bmatrix} \mathbf{I}_{m-1} & -\tilde{D}_m \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad \tilde{D}_m = \begin{bmatrix} -\mathbf{R}_{m-1}''^{-1} R_m'' \\ \mathbf{0} \end{bmatrix} \quad (2.45a)$$

$$\text{and } r'_m = r_m - R_m''^T \mathbf{R}_{m-1}''^{-1} R_m'' \quad (2.45b)$$

Note that the permutation-matrix pair of $\tilde{\mathbf{Q}}_m^T$ and $\tilde{\mathbf{Q}}_m$ only adjusts the orders of \mathbf{R}_m 's rows and columns and thus turns \mathbf{R}_m and \mathbf{R}_{m-1} into $\tilde{\mathbf{R}}_m$ and $\tilde{\mathbf{R}}_{m-1}$, respectively. The other pair of $\tilde{\mathbf{Q}}_m'^T$ and $\tilde{\mathbf{Q}}'_m$ has no effect on the orders of $\tilde{\mathbf{R}}_m$'s rows and columns. Therefore, the reverse pair of $\tilde{\mathbf{Q}}_m$ and $\tilde{\mathbf{Q}}_m'^T$ can recover the original orders of the adjusted rows and columns of \mathbf{R}_m . Obviously, the adjustment also turns $\tilde{\mathbf{R}}_{m-1}$ into \mathbf{R}_{m-1} . That is,

$$\tilde{\mathbf{Q}}_m \tilde{\mathbf{Q}}_m'^T \tilde{\mathbf{Q}}_m^T \mathbf{R}_m \tilde{\mathbf{Q}}_m \tilde{\mathbf{Q}}_m' \tilde{\mathbf{Q}}_m^T = \text{diag}(\mathbf{R}_{m-1}, r'_m) \quad (2.46a)$$

Namely,

$$\tilde{\mathbf{Q}}_m^T \mathbf{R}_m \tilde{\mathbf{Q}}_m = \text{diag}(\mathbf{R}_{m-1}, r'_m) \quad (2.46b)$$

$$\text{with } \tilde{\mathbf{Q}}_m = \tilde{\mathbf{Q}}_m \tilde{\mathbf{Q}}_m' \tilde{\mathbf{Q}}_m^T = \begin{bmatrix} \mathbf{I}_{m-1} & -\tilde{D}_m \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (2.47a)$$

Here, the m th row of $\tilde{\mathbf{Q}}_m$ has always a form as $[\mathbf{0}^T \ 1]$, so it can be easily verified that

$$\begin{bmatrix} -\tilde{D}_m \\ 1 \end{bmatrix} = \tilde{\mathbf{Q}}_m \begin{bmatrix} -\mathbf{R}_{m-1}''^{-1} R_m'' \\ \mathbf{0} \\ 1 \end{bmatrix} \quad (2.47b)$$

In other words, \tilde{D}_m is adjusted from \tilde{D}_m in such a way: keeping the orders of $-\mathbf{R}_{m-1}''^{-1} R_m''$'s entries and inserting these zeros in the “0” of \tilde{D}_m to those rows which are not occupied by \mathbf{R}_{m-1}'' .

Actually, these zero rows have no effect on the LRC decorrelation. That is,

$$r'_m = r_m - \tilde{D}_m^T R_m = r_m - R_m''^T \mathbf{R}_{m-1}''^{-1} R_m'' \quad (2.48a)$$

Similarly,

$$\begin{cases} H'_m \square H_m - \tilde{D}_m^T \mathbf{H}_{m-1} = H_m - R_m''^T \mathbf{R}_{m-1}''^{-1} \mathbf{H}_{m-1}'' \\ z'_m \square z_m - \tilde{D}_m^T \mathbf{z}_{m-1} = z_m - R_m''^T \mathbf{R}_{m-1}''^{-1} \mathbf{z}_{m-1}'' \end{cases} \quad (2.48b)$$

where \mathbf{H}_{m-1}'' occupies the rows of \mathbf{H}_{m-1} which have the same (row) numbers as those of \mathbf{R}_{m-1}'' in \mathbf{R}_{m-1} , so is \mathbf{z}_{m-1}'' . In brief, the last row and column of a PSD matrix \mathbf{R}_m can always be diagonalized using a pair of $\check{\mathbf{Q}}_m^T$ and $\check{\mathbf{Q}}_m$. \check{D}_m in $\check{\mathbf{Q}}_m$, as well as \tilde{D}_m , depends only on \mathbf{R}_{m-1}'' (a maximum-rank PD principal minor of \mathbf{R}_{m-1}) and the corresponding cross subvector R_m'' . To determine \mathbf{R}_{m-1}'' , the following proposition can be utilized.

Proposition 3 (Determination of maximum-rank PD principal minor): Consider the following PSD (or PD) matrix

$$\mathbf{R}_{a+1} = \begin{bmatrix} \mathbf{R}_a & R_{a+1} \\ R_{a+1}^T & r_{a+1} \end{bmatrix}$$

Suppose that \mathbf{R}_a'' is a maximum-rank PD minor of \mathbf{R}_a . Correspondingly, R_{a+1} 's subvector

R_{a+1}'' occupies the same rows as \mathbf{R}_a'' does. If $r_{a+1}' (\square r_{a+1} - R_{a+1}''^T \mathbf{R}_a''^{-1} R_{a+1}'') \neq 0$, then

$\mathbf{R}_{a+1}'' = \begin{bmatrix} \mathbf{R}_a'' & R_{a+1}'' \\ R_{a+1}''^T & r_{a+1}'' \end{bmatrix}$ is a maximum-rank PD minor of \mathbf{R}_{a+1} ; otherwise, if $r_{a+1}' = 0$, then \mathbf{R}_a'' is a

maximum-rank PD principal minor of \mathbf{R}_{a+1} .

A proof of Proposition 3 is given in Appendix G. Using Proposition 3, a maximum-rank PD principal minor of each \mathbf{R}_m can be determined recursively. The implementation of Proposition 3 is presented in Procedure 1 in Sec. 2.4. Based on Propositions 2 and 3, we define the generalized LRC decorrelation as follows.

Definition 2 (Generalized LRC decorrelation): Given a PSD \mathbf{R}_m as in (2.12), the pair of $\check{\mathbf{Q}}_m^T$ and $\check{\mathbf{Q}}_m$ defined by (2.47a) can perform a decorrelation over the last row and column (as in (2.46b)).

Clearly, after applying a set of successive generalized LRC decorrelations to \mathbf{R}_m , we can draw a conclusion similar to Proposition 1 as follows.

Proposition 4 (Inverse decomposition of a PSD matrix): Given PSD \mathbf{R}_m as in (2.12)

with $\mathbf{R}_1 = r_1 \neq 0$, the following equalities hold for $1 < m \leq M$,

$$\tilde{\mathbf{Q}}_{1:m}^T \mathbf{R}_m \tilde{\mathbf{Q}}_{1:m} = \text{diag}\{r_1, r'_2, \dots, r'_m\} \quad (2.49)$$

$$\text{with } \tilde{\mathbf{Q}}_{1:m} = \begin{bmatrix} 1 & -\tilde{D}_2 & \dots & -\tilde{D}_m \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 \end{bmatrix} \quad (2.50a)$$

where r'_m and \tilde{D}_m are defined as in (2.45b) and (2.47a), respectively.

It is clear that $\tilde{\mathbf{Q}}_{1:m}$ can be obtained recursively as

$$\tilde{\mathbf{Q}}_{1:m} = \begin{bmatrix} \tilde{\mathbf{Q}}_{1:(m-1)} & -\tilde{D}_m \\ 0 & 1 \end{bmatrix} \quad (2.50b)$$

which is similar to (2.28d). In addition, it can also be verified that

$$\begin{cases} \mathbf{z}'_m \square \tilde{\mathbf{Q}}_{1:m}^T \mathbf{z}_m = [z_1^T \ z_2'^T \ \dots \ z_m'^T]^T \\ \mathbf{H}'_m \square \tilde{\mathbf{Q}}_{1:m}^T \mathbf{H}_m = [H_1^T \ H_2'^T \ \dots \ H_m'^T]^T \end{cases} \quad (2.51)$$

where z'_m and H'_m are defined in (2.48b).

Particularly, if $r_1 = 0$, the decorrelation starts from the first m' with $r_{m'} \neq 0$.

Based on the above inverse decomposition, the ILE-constrained GLS, which in essence is GLS with implicit LE constraints, can be defined as follows.

Definition 3 (ILE-constrained GLS): For a GLS problem with data \mathbf{z} , coefficient \mathbf{H} and PSD OEC-like matrix \mathbf{R} (as in (2.12)), there exists $\tilde{\mathbf{Q}}_{1:M}$ defined as in (2.50a) such that

$$\begin{cases} \mathbf{R}' = \tilde{\mathbf{Q}}_{1:M}^T \mathbf{R} \tilde{\mathbf{Q}}_{1:M} = \text{diag}\{r_1, r'_2, \dots, r'_M\} = \text{diag}\{\dots r'_m \dots\} \\ \mathbf{z}' = \tilde{\mathbf{Q}}_{1:M}^T \mathbf{z} = [\dots z_m'^T \dots]^T \\ \mathbf{H}' = \tilde{\mathbf{Q}}_{1:M}^T \mathbf{H} = [\dots H_m'^T \dots]^T \end{cases} \quad (2.52a)$$

Let

$$\mathbf{R}^* = \text{diag}\{r_1, r_2^*, \dots, r_{M^*}^*\} \quad (2.52b)$$

where sequence $r_1^*, r_2^*, \dots, r_{M^*}^*$ contains all the nonzero diagonal elements of \mathbf{R}' in the original order. \mathbf{z}^* and \mathbf{H}^* store the elements of \mathbf{z}' and \mathbf{H}' from the same rows, respectively. In reverse, \mathbf{z}° and \mathbf{H}° store the remaining columns of \mathbf{z}' and \mathbf{H}' corresponding to the zero diagonal elements of \mathbf{R}' , respectively, where the possible components having all-zero coefficients have been ruled out. The solution to the ILE-constrained GLS is then equivalent to solving the following LE-constrained problem:

$$\begin{cases} \hat{\mathbf{x}} = \min_{\mathbf{x}} (\mathbf{z}^* - \mathbf{H}^* \mathbf{x})^T \mathbf{R}^{*-1} (\mathbf{z}^* - \mathbf{H}^* \mathbf{x}) \\ s. t. \quad \mathbf{H}^\circ \mathbf{x} = \mathbf{z}^\circ \end{cases} \quad (2.53)$$

It is clear that the problem (2.53) can be solved recursively. Note that, in Sec. 2.3.4, it has been shown that (multiple) LE constraints can be processed sequentially in an arbitrary order and each piece can be inserted in any place in the observation sequence. Consequently, the constraint data (denoted by \mathbf{H}° and \mathbf{z}°) and the observation data (by \mathbf{H}^* , \mathbf{z}^* and \mathbf{R}^*) can be arranged in the original order as in \mathbf{H}' , \mathbf{z}' , and \mathbf{R}' . Combined with the recursive inverse decomposition in Proposition 4, the recursive solution to the ILE-constrained GLS is developed as follows.

Theorem 4 (Recursive ILE-constrained GLS): If $\text{rank}(\mathbf{H}_{M_0}) = N$, $M_0 < M$, and the LE-constraint coefficient \mathbf{H}° is of full row rank, the ILE-constrained GLS problem described by Definition 3 has the following recursive solution:

$$\begin{cases} \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (\mathbf{z}'_m - \mathbf{H}'_m \hat{\mathbf{x}}_{m-1}) \\ \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{cases} \quad (2.54)$$

for $M_0 < m \leq M$ with

$$\begin{cases} \mathbf{S}_m \triangleq \mathbf{H}'_m \mathbf{C}_{m-1} \mathbf{H}'_m^T + \mathbf{r}'_m \\ \mathbf{K}_m \triangleq \mathbf{C}_{m-1} \mathbf{H}'_m^T \mathbf{S}_m^{-1} \end{cases} \quad (2.55)$$

$$\begin{cases} \mathbf{r}'_m = \mathbf{r}_m - \tilde{\mathbf{D}}_m^T \mathbf{R}_m = \mathbf{r}_m - \mathbf{R}_m^{*T} \mathbf{R}_{m-1}^{*-1} \mathbf{R}_m^* \\ \mathbf{H}'_m \square \mathbf{H}_m - \tilde{\mathbf{D}}_m^T \mathbf{H}_{m-1} = \mathbf{H}_m - \mathbf{R}_m^{*T} \mathbf{R}_{m-1}^{*-1} \mathbf{H}_{m-1}^* \\ \mathbf{z}'_m \square \mathbf{z}_m - \tilde{\mathbf{D}}_m^T \mathbf{z}_{m-1} = \mathbf{z}_m - \mathbf{R}_m^{*T} \mathbf{R}_{m-1}^{*-1} \mathbf{z}_{m-1}^* \end{cases} \quad (2.56)$$

and

$$\begin{cases} \mathbf{C}_{M_0} = \mathbf{U}_{M_0} [\mathbf{U}_{M_0}^T \mathbf{H}_{M_0}^{*T} \mathbf{R}_{M_0}^{*-1} \mathbf{H}_{M_0}^* \mathbf{U}_{M_0}]^{-1} \mathbf{U}_{M_0}^T \\ \hat{\mathbf{x}}_{M_0} = \mathbf{H}_{M_0}^{\circ+} \mathbf{z}_{M_0}^{\circ} + \mathbf{C}_{M_0} \mathbf{H}_{M_0}^{*T} \mathbf{R}_{M_0}^{*-1} (\mathbf{z}_{M_0}^* - \mathbf{H}_{M_0}^* \mathbf{H}_{M_0}^{\circ+} \mathbf{z}_{M_0}^{\circ}) \end{cases} \quad (2.57)$$

where $\mathbf{z}_{M_0}^*$, $\mathbf{H}_{M_0}^*$, $\mathbf{R}_{M_0}^*$, $\mathbf{z}_{M_0}^{\circ}$ and $\mathbf{H}_{M_0}^{\circ}$ are the components of \mathbf{z}^* , \mathbf{H}^* , \mathbf{R}^* , \mathbf{z}° and \mathbf{H}° in (2.53) which are related to the first M_0 data pieces (for an initialization). Actually, the term $\mathbf{R}_m^{nT} \mathbf{R}_{m-1}^{n-1}$ in (2.56) can also be obtained recursively using reduced computation, which is presented in detail by Procedure 2 in Sec 2.4 (see (2.69) and (2.70)).

In this theorem, two conditions need to be satisfied, which require \mathbf{H}° to be of full row rank and \mathbf{H}_{M_0} to be of full column rank, respectively. The first one requires that no redundant constraint exist in the LE constraint set. Note that, different from Theorem 2, constraint redundancy checking is needed in Theorem 4 because the constraint set is implicitly contained in data and incorrectly recorded data may cause redundant constraints. Regarding the second one, it can be easily shown that

$$\mathbf{H}_{M_0} = \tilde{\mathbf{Q}}_{M_0}^T \tilde{\mathbf{Q}}_{1:M_0}^T [\mathbf{H}_{M_0}^{*T} \quad \mathbf{H}_{M_0}^{\circ T}]^T \quad (2.58)$$

So $\text{rank}(\mathbf{H}_{M_0}) = \text{rank}([\mathbf{H}_{M_0}^{*T} \quad \mathbf{H}_{M_0}^{\circ T}]^T)$. According to Theorem 2, iff these two conditions are satisfied, an exact initialization, as well as the unique ILE-constrained GLS solution, is guaranteed. Subsequently, efficient tools for checking the two conditions are also developed: The follow-up Theorem 5 can check the constraint redundancy in an easy way; Chapter 3 will further handle issues related to the exact initialization, such as: i) how to check the rank of the coefficient matrix efficiently, ii) how to perform an exact initialization recursively, and iii) how to deal with the deficient-rank situation from a simple practical viewpoint.

Theorem 5 (Detection of redundant LE constraints): In the LE-constrained GLS defined by (2.10) and subject to $\tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{B}}$ with $\tilde{\mathbf{A}} = [\mathbf{A}_1^T \quad \mathbf{A}_2^T]^T$, given $\text{rank}([\mathbf{H}^T \quad \mathbf{A}_1^T]^T) = N$ and $\text{rank}(\mathbf{A}_1) = N_{A_1}$, then $\text{rank}(\tilde{\mathbf{A}}) = N_{\tilde{\mathbf{A}}}$ iff

$$\mathbf{S}_2 \neq 0 \quad (2.59)$$

where $\mathbf{S}_2 \sqsubseteq \mathbf{A}_2 \mathbf{C}_1 \mathbf{A}_2^T$, $\mathbf{C}_1 = \mathbf{U}_1 [\mathbf{U}_1^T (\mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1) \mathbf{U}_1]^{-1} \mathbf{U}_1^T$, and \mathbf{U}_1 satisfies $[\tilde{\mathbf{U}}_1 \ \mathbf{U}_1][\tilde{\mathbf{U}}_1 \ \mathbf{U}_1]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$.

A proof of Theorem 5 is given in Appendix H. This theorem indicates that no extra work but simply calculating \mathbf{S}_m can tell whether a piece of new LE constraint is redundant with respect to the processed ones or not. Once a redundant constraint is detected, a further judgment on constraint consistency can be made by comparing $\mathbf{A}_2 \hat{\mathbf{x}}_1$ and \mathbf{B}_2 . That is, if $\mathbf{A}_2 \hat{\mathbf{x}}_1 = \mathbf{B}_2$, then the constraint $\mathbf{A}_2 \mathbf{x} = \mathbf{B}_2$ has already been previously “imposed” and can thus be ignored. Otherwise, if $\mathbf{A}_2 \hat{\mathbf{x}}_1 \neq \mathbf{B}_2$, then the constraint is inconsistent with one or more previously-imposed constraint. Note that, in practical applications, the constraint inconsistency is mostly caused by incorrectly-recorded data. Therefore, data correctness needs to be rechecked if an inconsistent LE constraint is found. Finally, it can be verified that the solution of the LE-constrained GLS is identical to that of the corresponding ILE-constrained GLS with a PSD OEC-like matrix. Therefore, the latter is a unified and generalized recursive GLS.

2.4 Unified Procedures and Algorithms

During the course of decorrelating the PSD OEC-like matrix, the LRC decorrelation employs only a maximum-rank PD principal minor of the to-be-decorrelated matrix as well as the corresponding cross subvector. Therefore, the determination of a maximum-rank PD principal minor is crucial. This work can be handled by the following recursive procedure efficiently.

Procedure 1: Determination of maximum-rank PD principal minor

1) $m = 1, m^* = 1$,

$$\mathbf{R}_{m^*}^D = \mathbf{R}_m'' = \mathbf{R}_1 = r_1 \quad (r_1 \neq 0)$$

$$\mathbf{R}_{\mathbf{O}_m} = 1$$

2) $m := m + 1$:

Using the row numbers in $R_{o_{m-1}}$, pick R_m'' out of R_m ;

$$r_m' = r_m - R_m''^T R_{m-1}'' R_m''$$

2.1) If $r_m' = 0$, then

$$R_m'' = R_{m-1}''$$

2.2) If $r_m' \neq 0$, then $m^* := m^* + 1$:

$$R_{m^*}^D = R_m'' \quad (2.60a)$$

$$R_{m^*}^D = R_m'' = \begin{bmatrix} R_{m-1}'' & R_m'' \\ R_m''^T & r_m \end{bmatrix} \quad (2.60b)$$

$$R_{o_m} = [R_{o_m} \ m]$$

3) Repeat 2) till $m = M$.

Here, m^* is used to index different recorded maximum-rank PD principal minors. This indexing method can bring convenience to the follow-up procedure design.

Note that Procedure 1 is mainly for the use of illustrating the determination of R_m'' . In fact, in the recursive ILE-constrained GLS, it is $R_m''^{-1}$ (or $(R_{m^*}^D)^{-1}$ identically) that is directly involved in the computation. According to Proposition 1, there exists

$$Q_{1:(m^*)}^D = \begin{bmatrix} 1 & -(R_1^D)^{-1} R_2^D & \cdots & -(R_{m^*-1}^D)^{-1} R_{m^*}^D \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (2.61)$$

such that

$$(R_{m^*}^D)^{-1} = Q_{1:(m^*)}^D \cdot \text{diag}\{r_1^{-1}, (r_2^D)^{-1}, \dots, (r_{m^*}^D)^{-1}\} \cdot Q_{1:(m^*)}^{D^T} \quad (2.62)$$

where it can be easily verified that, for $1 < i \leq m^*$,

$$r_i^D = r_i^* \quad (2.63)$$

Therefore, (2.62) shows that the calculation of $(\mathbf{R}_{m^*}^D)^{-1}$ can take advantage of the previously-processed decorrelations, just as in Theorem 1. In addition, $\mathbf{Q}_{1:(m^*)}^D$ can also be obtained recursively from $\mathbf{Q}_{1:(m^*-1)}^D$ and $(\mathbf{R}_{m^*-1}^D)^{-1} R_{m^*-1}^D$. Both of these facts are adopted in the following recursive procedure to solve the ILE-constrained GLS comprehensively:

Procedure 2: Recursive ILE-constrained GLS

1) Initialization: $m = M_0$, $m^* = M_0^*$

$$\begin{cases} \mathbf{C}_m = \mathbf{C}_{M_0} \\ \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{M_0} \end{cases} \quad (\text{Same as in (2.56)}) \quad (2.64)$$

$$\mathbf{Q}_{1:M_0^*}^D = \begin{bmatrix} 1 & -r_1^{-1} R_2^D & \cdots & -(\mathbf{R}_{M_0^*-1}^D)^{-1} R_{M_0^*}^D \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix} \quad (2.65)$$

$$\mathbf{R}_{m^*}^D = \mathbf{R}_m'' = \mathbf{R}_{M_0}'' \quad (2.66)$$

$$\begin{cases} \mathbf{D}\mathbf{i}_{m^*}^* = \text{diag}\{r_1^{-1}, r_2^{*-1}, \dots, r_{M_0^*}^{*-1}\} \\ \mathbf{z}_{m^*}^* = \mathbf{Q}_{1:M_0^*}^{D^T} \mathbf{z}_{M_0}'' = \mathbf{z}_{M_0}^* \\ \mathbf{H}_{m^*}^* = \mathbf{Q}_{1:M_0^*}^{D^T} \mathbf{H}_{M_0}'' = \mathbf{H}_{M_0}^* \end{cases} \quad (2.67)$$

$$\mathbf{R_o}_m = \mathbf{R_o}_{M_0}; \mathbf{R_c}_m = \mathbf{R_c}_{M_0} \quad (2.68)$$

where $\mathbf{R_o}_{M_0}$ stores the row (and column) numbers of \mathbf{R}_{M_0}'' in \mathbf{R}_{M_0} and $\mathbf{R_c}_{M_0}$ stores all the other row numbers in \mathbf{R}_{M_0} . M_0^* is dimension of \mathbf{R}_{M_0}'' . Redundant LE constraints have been fixed.

2) Recursion: $m := m + 1$

2.1) Direct data processing:

Using the row numbers in $\mathbf{R_o}_{m-1}$, pick R_m'' out of R_m ;

$$\begin{cases} R_m' = \mathbf{Q}_{1:m^*}^{D^T} R_m'' \\ D_m' \sqcap \mathbf{D}\mathbf{i}_{m^*}^* R_m' \end{cases} \quad (2.69)$$

$$\begin{cases} r'_m = r_m - R_m'^T D'_m \\ H'_m = H_m - D_m'^T \mathbf{H}_m^* \\ z'_m = z_m - D_m'^T \mathbf{z}_m^* \end{cases} \quad (2.70)$$

$$\begin{cases} \mathbf{S}_m \triangleq H'_m \mathbf{C}_{m-1} H_m'^T + r'_m \\ \mathbf{K}_m \triangleq \mathbf{C}_{m-1} H_m'^T \mathbf{S}_m^{-1} \end{cases} \quad (2.71)$$

2.1.1) If $\mathbf{S}_m \neq 0$,

$$\begin{cases} \hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z'_m - H'_m \hat{\mathbf{x}}_{m-1}) \\ \mathbf{C}_m = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{cases} \quad (2.72)$$

2.1.2) If $\mathbf{S}_m = 0$ (redundant LE constraint),

Fix the constraint set recorded by $\mathbf{R}_{\mathbf{c}_m}$;

2.2) Reserve-data processing:

2.2.1) If $r'_m = 0$ (constraint),

$$\mathbf{R}_{\mathbf{o}_m} = \mathbf{R}_{\mathbf{o}_{m-1}}; \quad \mathbf{R}_{\mathbf{c}_m} = [\mathbf{R}_{\mathbf{c}_{m-1}} \quad m] \quad (2.73)$$

2.2.2) If $r'_m \neq 0$ (observation),

$$\mathbf{R}_{\mathbf{o}_m} = [\mathbf{R}_{\mathbf{o}_{m-1}} \quad m]; \quad \mathbf{R}_{\mathbf{c}_m} = \mathbf{R}_{\mathbf{c}_{m-1}} \quad (2.74)$$

$$m^* := m^* + 1 \quad (2.75)$$

$$\mathbf{Q}_{1:m^*}^{\mathbf{D}} = \begin{bmatrix} \mathbf{Q}_{1:(m^*-1)}^{\mathbf{D}} & -\mathbf{Q}_{1:(m^*-1)}^{\mathbf{D}} D'_m \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (2.76)$$

$$\begin{cases} \mathbf{D}_{m^*}^{\mathbf{i}} = \text{diag}\{\mathbf{D}_{m^*-1}^{\mathbf{i}}, r_m'^{-1}\} \\ \mathbf{H}_{m^*}^* = [\mathbf{H}_{m^*-1}^{*T} \quad H_m'^T]^T \\ \mathbf{z}_{m^*}^* = [\mathbf{z}_{m^*-1}^{*T} \quad z_m'^T]^T \end{cases} \quad (2.77)$$

3) Go to 2) till $m = M$.

Particularly, in the Initialization described by step 1), the determination of the initial values

\mathbf{C}_{M_0} and $\hat{\mathbf{x}}_{M_0}$ in (2.64) can be completed by the work in Chapter 3.

2.5 Performance Analysis

As a generalized approach, the recursive GLS inherits the recursion-oriented advantages of the conventional RLS [Zhu&Li07]. It remains a recursive form. The overall computation is free of direct matrix-inverse operation although matrix inverses are involved in both the EEC-like matrices and the weight calculations. Particularly, it is also revealed that the LE constraint can be recursively imposed as a special and even simplified observation set without resorting to MP-inverse operation. Therefore, the recursive GLS has much more reduced computational complexity than traditional LE-constrained GLS methods and is thus suitable for real-time applications. That is, the number of algebraic operations, as well as that of the required memory locations, is of $O(m^2)$ (order of m^2) at the m th recursion cycle while the corresponding batch solutions can cost as high as $O(M^2N)$. Note that, because of the increment of nonzero elements contained in each \mathbf{R}_m (i.e., R_m) amounts to m , the operations in each cycle may not remain fixed but increase with m . It is because the terms $R'_m = \mathbf{Q}_{1:m}^{D^T} R''_m$ in (2.69) and $H'_m = H_m - D_m'^T \mathbf{H}_m^*$ introduce increasingly more computation during the course of data accumulation. In particular, the major computation in the step of reserve-data processing as in 2.2) can be handled in two parallel threads since the formulae (2.70)-(2.72) and the formulae (2.74)-(2.77) do not rely on each other.

2.6 Appendix

2.6.1 Appendix A

Proof of Proposition 1:

According to (2.25), for $1 < m \leq M$, we have

$$\mathbf{Q}_m^T \mathbf{R}_m \mathbf{Q}_m = \mathbf{R}'_m$$

So $\mathbf{R}_m = (\mathbf{Q}_m^T)^{-1} \mathbf{R}'_m \mathbf{Q}_m^{-1}$

$$= (\mathbf{Q}_m^T)^{-1} \cdot \text{diag}\{\mathbf{R}_{m-1}, r'_m\} \cdot \mathbf{Q}_m^{-1}$$

$$= (\mathbf{Q}_m^T)^{-1} \begin{bmatrix} (\mathbf{Q}_{m-1}^T)^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \cdot \text{diag}\{\mathbf{R}'_{m-1}, r'_m\} \cdot \begin{bmatrix} \mathbf{Q}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{Q}_m^{-1} \quad (2.78)$$

Further diagonalizing \mathbf{R}'_{m-1} in the similar way, we can get that

$$\mathbf{R}_m = (\mathbf{Q}_{1:m}^T)^{-1} \cdot \text{diag}\{r_1, r'_2, \dots, r'_m\} \cdot \mathbf{Q}_{1:m}^{-1} \quad (2.79)$$

with

$$\mathbf{Q}_{1:m}^{-1} \square \begin{bmatrix} \mathbf{Q}_2^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \dots \begin{bmatrix} \mathbf{Q}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{Q}_m^{-1}$$

Equivalently,

$$\begin{aligned} \mathbf{Q}_{1:m} \square \mathbf{Q}_m \begin{bmatrix} \mathbf{Q}_{m-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \dots \begin{bmatrix} \mathbf{Q}_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \\ = \begin{bmatrix} 1 & -D_2 & \dots & -D_m \\ 0 & 1 & & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 \end{bmatrix} \end{aligned} \quad (2.80)$$

From (2.79),

$$\mathbf{Q}_{1:m}^T \mathbf{R}_m \mathbf{Q}_{1:m} = \text{diag}\{r_1, r'_2, \dots, r'_m\}. \text{ Done.}$$

2.6.2 Appendix B

Proof of Theorem 1:

According to (2.78),

$$\mathbf{R}_m^{-1} = \mathbf{Q}_m \cdot \text{diag}\{\mathbf{R}_{m-1}^{-1}, r_m'^{-1}\} \cdot \mathbf{Q}_m^T$$

Then it can be verified that

$$\begin{aligned} & \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \\ &= [\mathbf{H}_{m-1}^T \ H_m^T] \mathbf{Q}_m \cdot \text{diag}\{\mathbf{R}_{m-1}^{-1}, r_m'^{-1}\} \cdot \mathbf{Q}_m^T [\mathbf{H}_{m-1}^T \ H_m^T]^T \\ &= \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} + H_m'^T r_m'^{-1} H_m' \end{aligned} \quad (2.81)$$

$$\text{with } \begin{cases} H'_m = H_m - R_m^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} \\ r'_m = r_m - R_m^T \mathbf{R}_{m-1}^{-1} R_m \\ z'_m = z_m - R_m^T \mathbf{R}_{m-1}^{-1} \mathbf{z}_{m-1} \end{cases} \quad (2.82)$$

Furthermore, the fact that $\text{rank}(\mathbf{H}_{M_0}) = N$ ensures the existence of \mathbf{C}_{m-1} (for $M_0 < m \leq M$).

Therefore, similar to the conventional RLS in Fact 1 [Zhu&Li07], we can show from (2.22) that

$$\begin{aligned} \mathbf{C}_m &= (\mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m)^{-1} \\ &= (\mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} + H_m'^T r_m'^{-1} H_m')^{-1} \\ &= \mathbf{C}_{m-1} - \mathbf{C}_{m-1} H_m'^T (H_m' \mathbf{C}_m H_m'^T + r_m')^{-1} H_m' \mathbf{C}_m \\ &= \mathbf{C}_{m-1} - \mathbf{K}_m H_m' \mathbf{C}_{m-1} = \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{aligned} \quad (2.83)$$

and

$$\begin{aligned} \hat{\mathbf{x}}_m &= \mathbf{C}_m \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{z}_m \\ &= \mathbf{C}_m [\mathbf{H}_{m-1}^T \ H_m'^T] \mathbf{Q}_m \cdot \text{diag}\{\mathbf{R}_{m-1}^{-1}, r_m'^{-1}\} \cdot \mathbf{Q}_m^T [\mathbf{z}_{m-1}^T \ z_m^T]^T \\ &= \mathbf{C}_m [\mathbf{H}_{m-1}^T \ H_m'^T] \cdot \text{diag}\{\mathbf{R}_{m-1}^{-1}, r_m'^{-1}\} \cdot [\mathbf{z}_{m-1}^T \ z_m'^T]^T \\ &= \mathbf{C}_m \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{z}_{m-1} + \mathbf{C}_m H_m'^T r_m'^{-1} z_m' \end{aligned}$$

According to (2.83),

$$\begin{aligned} \mathbf{C}_m \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{z}_{m-1} &= (\mathbf{I}_{m-1} - \mathbf{K}_m H_m') \mathbf{C}_{m-1} \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{z}_{m-1} \\ &= (\mathbf{I}_{m-1} - \mathbf{K}_m H_m') \hat{\mathbf{x}}_{m-1} \\ \mathbf{C}_m H_m'^T r_m'^{-1} z_m' &= \mathbf{C}_{m-1} H_m'^T [\mathbf{I}_{m-1} - (H_m' \mathbf{C}_{m-1} H_m'^T + r_m')^{-1} H_m' \mathbf{C}_{m-1} H_m'^T] r_m'^{-1} z_m' \\ &= \mathbf{C}_{m-1} H_m'^T (H_m' \mathbf{C}_{m-1} H_m'^T + r_m')^{-1} [H_m' \mathbf{C}_{m-1} H_m'^T + r_m' - H_m' \mathbf{C}_{m-1} H_m'^T] r_m'^{-1} z_m' \\ &= \mathbf{K}_m z_m' \end{aligned}$$

As a result, for $M_0 < m \leq M$,

$$\hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z'_m - H'_m \hat{\mathbf{x}}_{m-1})$$

Eventually,

$$\begin{cases} \hat{\mathbf{x}}_M = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} \\ \quad = \hat{\mathbf{x}}_{M-1} + \mathbf{K}_M (z'_M - H'_M \hat{\mathbf{x}}_{M-1}) \\ \mathbf{C}_M = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = \mathbf{C}_{M-1} - \mathbf{K}_M \mathbf{S}_M \mathbf{K}_M^T \end{cases} \quad (2.84)$$

which is identical to the batch solution (2.23).

Furthermore, according to Proposition 1, for $1 < m \leq M$,

$$\begin{aligned} D_m &= \mathbf{R}_{m-1}^{-1} R_m \\ &= \mathbf{Q}_{1:(m-1)} \mathbf{D}_{m-1} \mathbf{Q}_{1:(m-1)}^T R_m \\ &= \mathbf{Q}_{1:(m-1)} \mathbf{D}_{m-1} R'_m = \mathbf{Q}_{1:(m-1)} D'_m \end{aligned} \quad (2.85)$$

So in (2.83),

$$H'_m = H_m - R_m^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} = H_m - D_m'^T \mathbf{Q}_{1:(m-1)}^T \mathbf{H}_{m-1} = H_m - D_m'^T \mathbf{H}'_{m-1}$$

$$r'_m = r_m - D_m'^T R'_m, \text{ and } z'_m = z_m - D_m'^T \mathbf{z}'_{m-1}$$

Correspondingly,

$$\mathbf{Q}_{1:m} = \begin{bmatrix} \mathbf{Q}_{1:(m-1)} & -D_m \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{1:(m-1)} & -\mathbf{Q}_{1:(m-1)} D'_m \\ \mathbf{0}^T & 1 \end{bmatrix}$$

and the formulae at $m = M_0$ come from the definitions.

2.6.3 Appendix C

Proof of Fact 4:

Since \mathbf{R} is PD, there exists $\mathbf{R}^{-\frac{1}{2}}$ such that $\mathbf{R}^{-\frac{1}{2}} = (\mathbf{R}^{-\frac{1}{2}})^T$ and $\mathbf{R}^{-\frac{1}{2}} \mathbf{R}^{-\frac{1}{2}} = \mathbf{R}^{-1}$. Then (2.15) can be rewritten as

$$\begin{bmatrix} (\mathbf{H}'^T \mathbf{H}') & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{H}'^T \mathbf{z}' \\ B \end{bmatrix} \quad (2.86)$$

with $\mathbf{H}' = \mathbf{R}^{-\frac{1}{2}} \mathbf{H}$ and $\mathbf{z}' = \mathbf{R}^{-\frac{1}{2}} \mathbf{z}$. Applying (1.11) (according to [Zhu&Li07]), iff

$\text{rank}([\mathbf{A}^T \ \mathbf{H}'^T]^T) = N$, (2.86) as well as (2.15) has a unique solution as

$$\begin{cases} \mathbf{C} = \mathbf{U}[\mathbf{U}^T (\sum_{m=1}^M \mathbf{H}_m'^T \mathbf{H}_m') \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{B} + \mathbf{C} \mathbf{H}'^T (\mathbf{z}' - \mathbf{H}' \mathbf{A}^+ \mathbf{B}) \end{cases} \quad (2.87)$$

Here, (2.87) is identical to (2.34). In addition,

$$[\mathbf{A}^T \ \mathbf{H}'^T]^T = \text{diag}(\mathbf{I}, \mathbf{R}^{-\frac{1}{2}}) [\mathbf{A}^T \ \mathbf{H}^T]^T.$$

Therefore,

$$\text{rank}([\mathbf{A}^T \ \mathbf{H}'^T]^T) = \text{rank}([\mathbf{A}^T \ \mathbf{H}^T]^T) \quad (2.88)$$

Fact 4 is thus proven.

2.6.4 Appendix D

Proof of Theorem 2:

Since $\text{rank}([\mathbf{A}^T \ \mathbf{H}_{M_0}'^T]^T) = N$, $\mathbf{T}_m (\square (\mathbf{U}^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{U})^{-1})$ exists for $M_0 < m \leq M$. Similar to (2.83), we have

$$\mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m = \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} + \mathbf{H}_m'^T \mathbf{r}_m'^{-1} \mathbf{H}_m'$$

So

$$\begin{aligned} \mathbf{C}_m &= \mathbf{U}[\mathbf{U}^T \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m \mathbf{U}]^{-1} \mathbf{U}^T \\ &= \mathbf{U}[\mathbf{U}^T \mathbf{H}_{m-1}^T \mathbf{R}_{m-1}^{-1} \mathbf{H}_{m-1} \mathbf{U} + \mathbf{U}^T \mathbf{H}_m'^T \mathbf{r}_m'^{-1} \mathbf{H}_m' \mathbf{U}]^{-1} \mathbf{U}^T \end{aligned} \quad (2.89)$$

Applying (2.22) to the matrix inverse in (2.89), we get

$$\mathbf{C}_m = \mathbf{U}[\mathbf{T}_{m-1} - \mathbf{T}_{m-1} \mathbf{U}^T \mathbf{H}_m'^T (\mathbf{H}_m' \mathbf{U} \mathbf{T}_{m-1} \mathbf{U}^T \mathbf{H}_m'^T + \mathbf{r}_m')^{-1} \mathbf{U} \mathbf{T}_{m-1}] \mathbf{U}^T$$

Note that $\mathbf{U} \mathbf{T}_{m-1} \mathbf{U}^T = \mathbf{C}_{m-1}$. Therefore,

$$\begin{aligned} \mathbf{C}_m &= \mathbf{C}_{m-1} - \mathbf{C}_{m-1} \mathbf{H}_m'^T (\mathbf{H}_m' \mathbf{C}_{m-1} \mathbf{H}_m'^T + \mathbf{r}_m')^{-1} \mathbf{H}_m' \mathbf{C}_{m-1} \\ &= \mathbf{C}_{m-1} - \mathbf{K}_m \mathbf{S}_m \mathbf{K}_m^T \end{aligned} \quad (2.90)$$

According to (2.35),

$$\hat{\mathbf{x}}_m = \mathbf{A}^+ \mathbf{B} + \mathbf{C}_m \mathbf{H}_m^T \mathbf{R}_m^{-1} [\mathbf{z}_m - \mathbf{H}_m \mathbf{A}^+ \mathbf{B}]$$

Similar to the proof of Theorem 1, by substituting \mathbf{C}_m with (2.90), we can show that

$$\hat{\mathbf{x}}_m = \hat{\mathbf{x}}_{m-1} + \mathbf{K}_m (z'_m - \mathbf{H}'_m \hat{\mathbf{x}}_{m-1}) \quad (2.91)$$

2.6.5 Appendix E

Proof of Theorem 3:

Statement 1):

This well-known result can be verified from the normal equation (2.15) explicitly. It discloses that, if $\text{rank}(\mathbf{H}) = N$, \mathbf{C}_{lec} can be computed from $\mathbf{C}_{\text{uc}} (= (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1})$ conveniently via “deducting” a constraint-related term.

Statement 2):

Since $[\mathbf{A}^T \ \mathbf{H}^T]^T$ is a full-rank square matrix, $\hat{\mathbf{x}}_{\text{lec}}$ is an exactly-determined GLS solution:

$$\hat{\mathbf{x}}_{\text{lec}} = ([\mathbf{A}^T \ \mathbf{H}^T]^T)^{-1} [\mathbf{B}^T \ \mathbf{z}^T]^T \square [\mathbf{T}_1 \ \mathbf{T}_2] [\mathbf{B}^T \ \mathbf{z}^T]^T = \mathbf{T}_1 \mathbf{B} + \mathbf{T}_2 \mathbf{z} \quad (2.93)$$

According to (2.34), we also have

$$\hat{\mathbf{x}}_{\text{lec}} = (\mathbf{A}^+ - \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A}^+) \mathbf{B} + \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z} \quad (2.94)$$

These two $\hat{\mathbf{x}}_{\text{lec}}$ ’s are identical for arbitrary \mathbf{z} and \mathbf{B} , so

$$\mathbf{T}_1 = \mathbf{A}^+ - \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A}^+, \quad \mathbf{T}_2 = \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{R}^{-1}$$

In $\mathbf{C}_{\text{lec}, N \times N}$ defined in (2.39a), replace $([\mathbf{A}^T \ \mathbf{H}^T]^T)^{-1}$ with $[\mathbf{T}_1 \ \mathbf{T}_2]$, then

$$\mathbf{C}_{\text{lec}, N \times N} = \mathbf{T}_2 \mathbf{R} \mathbf{T}_2^T = \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{C}_{\text{lec}} \quad (2.95)$$

Substituting $\mathbf{U}[\mathbf{U}^T (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \mathbf{U}]^{-1} \mathbf{U}^T$ for \mathbf{C}_{lec} , we get

$$\mathbf{C}_{\text{lec}, N \times N} = \mathbf{U}[\mathbf{U}^T (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \mathbf{U}]^{-1} \mathbf{U}^T = \mathbf{C}_{\text{lec}}.$$

Similarly, $\hat{\mathbf{x}}_{\text{lec}, N \times N} = \hat{\mathbf{x}}_{\text{lec}}.$

Statement 3):

We have

$$\mathbf{A}_1 \mathbf{C}_{\text{lec}} = \mathbf{A}_1 \mathbf{U} [\mathbf{U}^T (\mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1) \mathbf{U}]^{-1} \mathbf{U}^T \quad (2.96)$$

where $[\tilde{\mathbf{U}} \mathbf{U}] [\tilde{\mathbf{U}} \mathbf{U}]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}) = \text{col}(\mathbf{A}^T)$.

Since the rows of \mathbf{A}_1 are all from \mathbf{A} ,

$$\mathbf{A}_1 \mathbf{U} = \mathbf{0}, \quad \mathbf{A}_1 \mathbf{C}_{\text{lec}} = \mathbf{0}. \quad (2.97)$$

Compute \mathbf{C}_1 from \mathbf{C}_{lec} :

$$\mathbf{C}_1 = \mathbf{C}_{\text{lec}} - \mathbf{C}_{\text{lec}} \mathbf{A}_1^T (\mathbf{A}_1 \mathbf{C}_{\text{lec}} \mathbf{A}_1^T + \mathbf{W}_{A1}^{-1})^{-1} \mathbf{A}_1 \mathbf{C}_{\text{lec}} \quad (2.98)$$

Taking (2.97) in, we get

$$\mathbf{C}_1 = \mathbf{C}_{\text{lec}} \quad (2.99)$$

Similarly, $\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_{\text{lec}}$. End.

Statement 4):

Let $\mathbf{H}' = \mathbf{R}^{-\frac{1}{2}} \mathbf{H}$ and $\mathbf{z}' = \mathbf{R}^{-\frac{1}{2}} \mathbf{z}$. Then the original LE-constrained GLS is equivalently transformed to the following LE-constrained LS:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J \quad (2.100a)$$

$$\text{with } J = (\mathbf{z}' - \mathbf{H}' \mathbf{x})^T (\mathbf{z}' - \mathbf{H}' \mathbf{x}) \quad (2.100b)$$

subject to

$$\mathbf{A} \mathbf{x} = \mathbf{B} \quad (2.100c)$$

Correspondingly, we also have

$$\begin{cases} \mathbf{C}_{\text{lec},1} = \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}'^T \mathbf{R}^{-1} \mathbf{H}' \mathbf{U}_1)^{-1} \mathbf{U}_1^T = \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}'^T \mathbf{H}' \mathbf{U}_1)^{-1} \mathbf{U}_1^T \\ \hat{\mathbf{x}}_{\text{lec},1} = \mathbf{A}_1^+ \mathbf{B}_1 + \mathbf{C}_{\text{lec},1} \mathbf{H}'^T (\mathbf{z}' - \mathbf{H}' \mathbf{A}_1^+ \mathbf{B}_1) = \mathbf{A}_1^+ \mathbf{B}_1 + \mathbf{C}_{\text{lec},1} \mathbf{H}'^T (\mathbf{z}' - \mathbf{H}' \mathbf{A}_1^+ \mathbf{B}_1) \end{cases} \quad (2.101)$$

Then the Statement 2) is equivalent to that

$$\begin{cases} \mathbf{C}'_{\text{lec}} = \mathbf{C}_{\text{lec},2} \\ \hat{\mathbf{x}}'_{\text{lec}} = \hat{\mathbf{x}}_{\text{lec},2} \end{cases} \quad (2.102)$$

A proof of this LS problem will be presented in Chapter 3.

2.6.6 Appendix F

Proof of Proposition 2:

$$\text{Let } \check{\mathbf{Q}}_{a+1} = \begin{bmatrix} \mathbf{I}_a & -\mathbf{R}_a^{-1}\mathbf{R}_{a,a+1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \text{ and}$$

$$\check{\mathbf{Q}} = \begin{bmatrix} \mathbf{I}_a & -\mathbf{R}_a^{-1}\mathbf{R}_{a,a+1} & -\mathbf{R}_a^{-1}R_{a+2}'' \\ \mathbf{0}^T & \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0} & 1 \end{bmatrix}.$$

Clearly, $\check{\mathbf{Q}}_{a+1}$, $\check{\mathbf{Q}}_{a+2}$ and $\check{\mathbf{Q}}$ all have full rank. We have

$$\check{\mathbf{Q}}_{a+1}^T \mathbf{R}_{a+1} \check{\mathbf{Q}}_{a+1} = \text{diag}(\mathbf{R}_a, \mathbf{R}_{a+1,a+1} - \mathbf{R}_{a,a+1}^T \mathbf{R}_a^{-1} \mathbf{R}_{a,a+1}) \quad (2.103)$$

Since $\text{rank}(\mathbf{R}_{a+1}) = \text{rank}(\mathbf{R}_a)$,

$$\mathbf{R}'_{a+1,a+1} = \mathbf{R}_{a+1,a+1} - \mathbf{R}_{a,a+1}^T \mathbf{R}_a^{-1} \mathbf{R}_{a,a+1} = \mathbf{0} \quad (2.104)$$

Furthermore,

$$\check{\mathbf{Q}}^T \mathbf{R}_{a+2} \check{\mathbf{Q}} = \begin{bmatrix} \mathbf{R}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{R}'_{a+1,a+1} & R'_{a+1,a+2} \\ \mathbf{0}^T & R_{a+1,a+2}'^T & r_{a+2} - R_{a+2}''^T \mathbf{R}_a^{-1} R_{a+2}'' \end{bmatrix} \quad (2.105)$$

with $R'_{a+1,a+2} = R_{a+1,a+2} - \mathbf{R}_{a,a+1}^T \mathbf{R}_a^{-1} R_{a+2}''$ and $\mathbf{R}'_{a+1,a+1} = \mathbf{0}$ (from (2.104)). If $R'_{a+1,a+2} \neq \mathbf{0}$,

$$\text{rank} \left(\begin{bmatrix} \mathbf{R}'_{a+1,a+1} & R'_{a+1,a+2} \\ R_{a+1,a+2}'^T & r_{a+2} - R_{a+2}''^T \mathbf{R}_a^{-1} R_{a+2}'' \end{bmatrix} \right) \geq 2$$

which implies that $\text{rank}(\mathbf{R}_{m+1}) \geq \text{rank}(\mathbf{R}_m) + 2$. It is a contradiction. Therefore,

$$\begin{cases} R'_{a+1,a+2} = R_{a+1,a+2} - \mathbf{R}_{a,a+1}^T \mathbf{R}_a^{-1} R_{a+2}'' = \mathbf{0} \\ R_{a+1,a+2}^T - R_{a+2}''^T \mathbf{R}_a^{-1} \mathbf{R}_{a,a+1} = \mathbf{0}^T \end{cases} \quad (2.106)$$

Applying (2.106) to $\check{\mathbf{Q}}_{a+2}^T \mathbf{R}_{a+2} \check{\mathbf{Q}}_{a+2}$, (2.43b) is thus proven.

2.6.7 Appendix G

Proof of Proposition 3:

Since $r_i \neq 0$, we have

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ R_{a+1}''^T \mathbf{R}_a''^{-1} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_a'' & \mathbf{0} \\ \mathbf{0}^T & r'_{a+1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ R_{a+1}''^T \mathbf{R}_a''^{-1} & 1 \end{bmatrix}^T = \begin{bmatrix} \mathbf{R}_a'' & R_{a+1}'' \\ R_{a+1}''^T & r \end{bmatrix}$$

It is clear that $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ R_{a+1}''^T \mathbf{R}_a''^{-1} & 1 \end{bmatrix}$ is a full-rank square matrix.

Therefore, if $r'_{a+1} \neq 0$, then $\begin{bmatrix} \mathbf{R}_a'' & \mathbf{0} \\ \mathbf{0}^T & r'_{a+1} \end{bmatrix}$, as well as $\begin{bmatrix} \mathbf{R}_a'' & R_{a+1}'' \\ R_{a+1}''^T & r \end{bmatrix}$, is PD. Thus,

$\mathbf{R}_{a+1}'' = \begin{bmatrix} \mathbf{R}_a'' & R_{a+1}'' \\ R_{a+1}''^T & r \end{bmatrix}$ is a maximum-rank PD principal minor of \mathbf{R}_{a+1} .

If $r'_{a+1} = 0$, then \mathbf{R}_{a+1} can be diagonalized as $\text{diag}(\mathbf{R}_a, 0)$. Thus, $\text{rank}(\mathbf{R}_{a+1}) = \text{rank}(\mathbf{R}_a)$.

\mathbf{R}_a'' is still a maximum-rank PD principal minor of \mathbf{R}_{a+1} .

2.6.8 Appendix H

Proof of Theorem 5:

$[\tilde{\mathbf{U}}_1 \mathbf{U}_1][\tilde{\mathbf{U}}_1 \mathbf{U}_1]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$ imply that, for an arbitrary row vector A_a ,

$$A_a \mathbf{U}_1 = \mathbf{0}^T \Leftrightarrow A_a^T \in \text{col}(\mathbf{A}_1^T). \quad (2.107)$$

First, if $\tilde{\mathbf{A}}$ has full row rank, then $A_2^T \notin \text{col}(\mathbf{A}_1^T)$. According to (2.107), we have $A_2 \mathbf{U}_1 \neq \mathbf{0}^T$.

As a result,

$$\mathbf{S}_2 = A_2 \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_1)^{-1} \mathbf{U}_1^T A_2^T \neq 0.$$

Second, if $\tilde{\mathbf{A}}$ does not have full row rank but \mathbf{A}_1 does, then $A_2^T \in \text{col}(\mathbf{A}_1^T)$. So $A_2 \mathbf{U}_1 = \mathbf{0}^T$ and

$$\mathbf{S}_2 = A_2 \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_1)^{-1} \mathbf{U}_1^T A_2^T = 0$$

Hence, the theorem is proven.

Chapter 3: Completely Recursive Least Squares—Part II:

Recursive Initialization and Deficient-Rank Processing

3.1 Introduction

3.1.1 Background

As reviewed in Chapter 1, a great many methods have been developed to solve the fundamental and classic problem of linear LS, among which the matrix-inversion-lemma (MI lemma) based recursive least-squares (RLS) is a milestone. The RLS provides LS solutions in a sequential way and is thus excellent in computation and storage [Albert65]. It is very suitable to real-time LS applications and has been widely applied into such areas as signal processing, control and communication [Goodwin77] [Ljung87] [Chen85] [Haykin01] [Passion98] [Mikles07].

The RLS was originally applied to solve unconstrained LS problems [Albert65], in which the equivalency between the recursive and the corresponding batch LS solutions is guaranteed once the MI-lemma based recursion starts from an exact LS solution on the full-rank initial data. Recently, reference [Zhou&Li07] discovers that the exact solutions of the linear-equality (LE) constrained LS problem can be obtained by the same (unconstrained) recursion cycle, provided that the RLS procedure is appropriately initialized. We aim to make the unconstrained RLS “complete” in two aspects: in Chapter 2 (Part I), a unified recursive procedure for all LS, WLS and GLS problems with or without LE constraints is developed, where an accompanying recursive decorrelation procedure is utilized; in this chapter, we continue discussing appropriate RLS initialization techniques. As reviewed in Sec. 1.4.2, less attention has been paid to the RLS initialization in the low-dimensional and high-redundancy environment. However, for high-dimensional and low-redundancy applications, such as power system state estimation, it is necessary to develop an efficient exact RLS initialization. In the existing methods reviewed in Sec. 1.4.2, approximate ones usually are not robust because a

better approximation can lead to a poorer numerical condition [Albert65]; it is also not always realistic to start the recursion with forgoing data having special observation functions for simple exact initializations [Haykin01] [Hubing91] [Albert65]. Reference [Zhou02] introduces variants of the Greville formula (order recursion) to develop recursive initializations for RLS. We will study a simple recursive initialization method for RLS, where only the RLS formulae are used. In addition, two accessory issues should also be studied. One is whether and when the foregoing observations can support an exact and unique RLS initialization, namely (parameter) observability analysis (or rank check) in engineering. A good practical observability analysis tool is expected to distinguish theoretical rank deficiency and numerical rank deficiency (see Sec. 1.4.2). The second issue is how to deal with problems with insufficient observations. Although the minimum-norm solution, which is the one (among infinitely many feasible estimates) has the minimum norm and also has a simple analytical form in batch, is widely adopted in theory [Lawson95] [Zhou02], other choices that may bring more convenience in practice should also be considered.

Exploiting an efficient and exact initialization for RLS, including the implicit rank check and deficient-rank processing, is worthwhile and expected to make the RLS more applicable to large-dimensional applications where exact initializations are costly.

3.1.2 Our Work

This chapter treats the initialization-related issues in the RLS problems, including rank check, a convenient method to compute the involved matrix inverse/pseudoinverse, and resolution of underdetermined systems. No extra non-RLS formula but an auxiliary-observation based procedure is utilized.

The main ideas include: a) introducing a set of simple auxiliary observations to construct a simple fake RLS initialization; b) developing a series of simple tools to detect and remove these auxiliary observations as soon as possible (ASAP). More concretely, with our method,

- 1) The RWLS recursion can start from the first real observation and possible LE constraints are also imposed recursively;

2) The rank of the system is checked implicitly. If the rank is full, the initialization and the subsequent RLS cycles can be integrated as a whole to yield exact LS solutions;

3) If the rank is deficient, the procedure provides a mapping from the unobservable (original) estimand to a reduced-dimensional set of alternative variables which are linear combinations of the original variables and uniquely determined. The consequent estimate is a set of refined non-redundant observations. The refinement is lossless in the WLS sense: if new observations are available later, it can take the role of the original data in the recalculation.

4) Furthermore, other simple mapping matrices, which lead to alternative-variable sets reserving all the uniquely-determined original variables, can also be easily constructed from the output of the procedure.

With the above settings, our recursive initialization is efficient and the RLS is thus more applicable to large-scale applications where exact initializations are costly.

In addition, this set of techniques is also applicable to the recursive GLS (RLS) discussed in Chapter 2 because the proposed initialization is free of the diagonalization process in the GLS.

3.2 Problem Formulation

The purpose is to present a series of initialization techniques to complement the RLS and solve linear WLS problems with or without LE constraint in a completely recursive manner. The deficient-rank case in which the estimand can not be uniquely determined by the observation-and-constraint set is also considered.

3.2.1 WLS and Its Batch Solutions

In a WLS problem

$$\mathbf{W} = \text{diag}(\mathbf{w}_{1:M}) = \text{diag}([w_1 \cdots w_M]) \quad (3.1a)$$

Correspondingly,

$$\begin{cases} \mathbf{z} = \mathbf{z}_M = \mathbf{z}_{1:M} = [z_1 \cdots z_M]^T \\ \mathbf{H} = \mathbf{H}_M = \mathbf{H}_{1:M} = [H_1^T \cdots H_M^T]^T \end{cases} \quad (3.1b)$$

The objective is to solve

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} J \quad (3.2)$$

$$\text{with } J = \mathbf{H}^T \mathbf{W} \mathbf{H} = \sum_{m=1}^M (z_m - H_m \mathbf{x})^T w_m (z_m - H_m \mathbf{x}) \quad (3.3)$$

where estimand \mathbf{x} is a vector containing all the N to-be-determined variables and $\hat{\mathbf{x}}$ the estimated \mathbf{x} . H is observation coefficient (vector) in real-observation function, and z observation value. Correspondingly, observation coefficient (matrix) \mathbf{H} contains all H 's and observation-value vector \mathbf{z} contains all z 's. The total number of observation is M and m is observation (data) index.

First, when there is no constraint, the normal equation is

$$(\sum_{m=1}^M H_m^T w_m H_m) \hat{\mathbf{x}} = \sum_{m=1}^M H_m^T w_m z_m \quad (3.4)$$

which has a unique solution if and only if $\text{rank}(\mathbf{H}) = N$:

$$\begin{cases} \mathbf{C}_{\text{uc}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} = (\sum_{m=1}^M H_m^T w_m H_m)^{-1} \\ \hat{\mathbf{x}}_{\text{uc}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{z} = \mathbf{C}_{\text{uc}} \sum_{m=1}^M H_m^T w_m z_m \end{cases} \quad (3.5)$$

Second, when there is a set of (consistent) LE constraints

$$\mathbf{A} \mathbf{x} = \mathbf{B} \quad (3.6)$$

with constraint coefficient $\mathbf{A} \in \mathbb{R}^{T \times N}$, constant $\mathbf{B} \in \mathbb{R}^{T \times 1}$, and (without loss of generality) $\text{rank}(\mathbf{A}) = T$, the normal equation is

$$\begin{bmatrix} (\sum_{m=1}^M H_m^T w_m H_m) & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \sum_{m=1}^M H_m^T w_m z_m \\ \mathbf{B} \end{bmatrix} \quad (3.7)$$

which has a unique solution if and only if $\text{rank}(\tilde{\mathbf{H}}) = N$:

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{U}[\mathbf{U}^T \mathbf{H}^T \mathbf{W} \mathbf{H} \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}}_{\text{lec}} = \mathbf{A}^+ \mathbf{B} + \mathbf{C}_{\text{lec}} \mathbf{H}^T \mathbf{W} [\mathbf{z} - \mathbf{H} \mathbf{A}^+ \mathbf{B}] \end{cases} \quad (3.8)$$

where $\tilde{\mathbf{H}} \in [\mathbf{A}^T \mathbf{H}^T]^T$. $\boldsymbol{\lambda}$ is Lagrange's multiplier. Superscript “+” stands for the Moore-Penrose (MP) inverse. \mathbf{U} satisfies $[\tilde{\mathbf{U}} \mathbf{U}][\tilde{\mathbf{U}} \mathbf{U}]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}) = \text{col}(\mathbf{A}^T)$, where “ $\text{col}(\mathbf{X})$ ” denotes the space spanned by all the columns of \mathbf{X} .

3.2.2 RLS Solutions

The RLS computes the unconstrained WLS solution recursively. That is, the solution in (3.5) can be calculated by repeating the following recursion cycle (and increasing \tilde{m}) sequentially:

$$\begin{cases} \mathbf{S}_{\text{uc},\tilde{m}} \triangleq \mathbf{H}_{\tilde{m}+M_0} \mathbf{C}_{\text{uc},\tilde{m}-1} \mathbf{H}_{\tilde{m}+M_0}^T + \mathbf{W}_{\tilde{m}+M_0}^{-1} \\ \mathbf{K}_{\text{uc},\tilde{m}} \triangleq \mathbf{C}_{\text{uc},\tilde{m}-1} \mathbf{H}_{\tilde{m}+M_0}^T \mathbf{S}_{\text{uc},\tilde{m}}^{-1} \\ \hat{\mathbf{x}}_{\text{uc},\tilde{m}} = \hat{\mathbf{x}}_{\text{uc},\tilde{m}-1} + \mathbf{K}_{\text{uc},\tilde{m}} (z_{\tilde{m}+M_0} - \mathbf{H}_{\tilde{m}+M_0} \hat{\mathbf{x}}_{\text{uc},\tilde{m}-1}) \\ \mathbf{C}_{\text{uc},\tilde{m}} = \mathbf{C}_{\text{uc},\tilde{m}-1} - \mathbf{K}_{\text{uc},\tilde{m}} \mathbf{S}_{\text{uc},\tilde{m}} \mathbf{K}_{\text{uc},\tilde{m}}^T \end{cases} \quad (3.9)$$

where $\text{rank}(\mathbf{H}_{1:M_0}) = N$ and the initial values are

$$\begin{cases} \mathbf{C}_{\text{uc},0} = (\sum_{m=1}^{M_0} \mathbf{H}_m^T \mathbf{W}_m \mathbf{H}_m)^{-1} \\ \hat{\mathbf{x}}_{\text{uc},0} = \mathbf{C}_{\text{uc},0} \sum_{m=1}^{M_0} \mathbf{H}_m^T \mathbf{W}_m \mathbf{z}_m \end{cases} \quad (3.10)$$

which are identical to the batch solution with the first M_0 data.

Furthermore, reference [Zhu&Li07] shows that the recursion (3.9) is also exact for (3.8), provided that the constraints are imposed initially:

$$\begin{cases} \mathbf{C}_{\text{lec},0} = \mathbf{U}[\mathbf{U}^T (\sum_{m=1}^{M_0} \mathbf{H}_m^T \mathbf{W}_m \mathbf{H}_m) \mathbf{U}]^{-1} \mathbf{U}^T \\ \hat{\mathbf{x}}_{\text{lec},0} = \mathbf{A}^+ \mathbf{B} + \mathbf{C}_{\text{lec},0} (\sum_{m=1}^{M_0} \mathbf{H}_m^T \mathbf{W}_m) [\mathbf{z} - \mathbf{H} \mathbf{A}^+ \mathbf{B}] \end{cases} \quad (3.11)$$

where $\text{rank}([\mathbf{A}^T \ \mathbf{H}_{1:M_0}^T]^T) = N$.

3.2.3 Our Goals

It is not always easy to find the initial values $\mathbf{C}_{\text{uc},0}$ and $\hat{\mathbf{x}}_{\text{uc},0}$ from the forgoing data, and much less for $\mathbf{C}_{\text{lec},0}$ and $\hat{\mathbf{x}}_{\text{lec},0}$ since $\mathbf{C}_{\text{lec},0}$ is positive semi-definite (PSD) and thus the MP inverse is involved. The proposed initialization is to overcome these difficulties. It performs initializations, imposes LE constraints, and determines the minimal initial data set, in a simple and totally recursive way. In addition, the rank of the system is checked implicitly. If the rank is full, an exact initialization is obtained. If the rank is deficient, a practical deficient-rank

processing method is proposed to provide a set of refined non-redundant observations.

3.3 Derivations and Developments

The theoretical foundations and derivations for the recursive initialization are still based on those for the RLS, which include the following preliminaries.

3.3.1 Preliminaries

As reviewed in Sec. 2.3.1, the Schur's Identity leads to the matrix-inversion (MI) lemma. The MI lemma has a well-known corollary which is the basis of RLS:

Corollary of MI Lemma: If inverses of matrices \mathbf{D} , \mathbf{G} and $(\mathbf{G} - \mathbf{E}^T \mathbf{D}^{-1} \mathbf{E})$ exist, then

$$(\mathbf{D} + \mathbf{E}^T \mathbf{G}^{-1} \mathbf{E})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{E}^T (\mathbf{E} \mathbf{D}^{-1} \mathbf{E}^T + \mathbf{G})^{-1} \mathbf{E} \mathbf{D}^{-1} \quad (3.18)$$

Particularly, it contains a bidirectional causal relation:

$$\text{Given } \mathbf{D}^{-1} \text{ and } \mathbf{G}^{-1}, (\mathbf{G} - \mathbf{E}^T \mathbf{D}^{-1} \mathbf{E})^{-1} \Leftrightarrow (\mathbf{D} + \mathbf{E}^T \mathbf{G}^{-1} \mathbf{E})^{-1}$$

which is an important base to the subsequent development.

3.3.2 Development of Recursive Initialization

In our recursive initialization method, we will first introduce a set of simple auxiliary observations to construct a fake initialization which allows the RLS recursion to start from the piece of observation data. Then during the course of real observation data processing, we delete these auxiliary observations ASAP using the corresponding canceling observations. Therefore, to present the recursive initialization, three types of unconventional observations are introduced first:

Simple auxiliary observation (SAO) – a make-up simple observation (direct measurement of estimand component) utilized for simplicity. Specifically, N different SAO's are adopted in this chapter, among which the s th one has the following observation coefficient, data, and (simple) weight:

$$H_{\text{au},s} (= I_s \square [0_1 \cdots 0_1 1_s 0 \cdots 0_N]), \quad z_{\text{au},s}, \quad w_{\text{au},s} = p_s^{-1} \quad (3.19)$$

where s is SAO index. $p_s > 0$ and the values of $z_{\text{au},s}$ and p_s can be assigned by the user for

convenience.

Canceling observation (CO) – a fake observation utilized to counteract the influence of a specific auxiliary or real observation (RO). Specifically, suppose that the coefficient, value and weight of the to-be-removed simple observation are \ddot{H} , \ddot{z} and $\ddot{w} (> 0)$. Then the corresponding simple canceling observation (SCO) are \ddot{H} , \ddot{z} and $-\ddot{w}$, respectively.

Pseudo observation (PO) – an “observation” that corresponds to a LE constraint. Specifically, the i th constraint is a PO having b_i , A_i and $+\infty$ as data, observation coefficient and weight, respectively.

Basic Tools

An observation may be removed from the processed observation set using the RLS-based method conveniently [Albert65]. We introduce the concept of CO and further consider whether an observation is removable in LE-constrained/unconstrained problems, which leads to the following tool.

Theorem 1 (Observation removal): In the LE-constrained WLS problem (3.2) with (3.6), given $\mathbf{C}_1 = (\mathbf{I} - \mathbf{U}(\mathbf{U}^T \mathbf{H}^T \mathbf{W} \mathbf{H} \mathbf{U})^{-1} \mathbf{U}^T)$ with $\mathbf{H}^T \mathbf{W} \mathbf{H} = \mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1 + \mathbf{H}_2^T \mathbf{W}_2 \mathbf{H}_2$, $\mathbf{C}_2 (\mathbf{I} - \mathbf{U}(\mathbf{U}^T \mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1 \mathbf{U})^{-1} \mathbf{U}^T)$ exists if and only if $(\mathbf{H}_2 \mathbf{C}_1 \mathbf{H}_2^T - \mathbf{W}_2^{-1})^{-1}$ exists. If \mathbf{C}_2 exists, then

$$\mathbf{C}_2 = \mathbf{C}_1 - \mathbf{C}_1 \mathbf{H}_2^T (\mathbf{H}_2 \mathbf{C}_1 \mathbf{H}_2^T - \mathbf{W}_2^{-1})^{-1} \mathbf{H}_2 \mathbf{C}_1 \quad (3.20)$$

Proof:

In Theorem 1, \mathbf{C}_1 exists if and only if $\text{rank}(\tilde{\mathbf{H}}) = N$ (see the derivation for (3.8) in [Zhu&Li07]). The MI Lemma is applied to expand \mathbf{C}_2 and the if-and-only-if condition requires that $\text{rank}(\tilde{\mathbf{H}})$ remain N after $\mathbf{H}_2^T \mathbf{W}_2 \mathbf{H}_2$ is removed.

Theorem 1 has the following corollary:

Corollary 1.1 (Simple observation removal): In the LE-constrained WLS problem (3.2)

with (3.6), given $\mathbf{C}_1 (\square \mathbf{U}(\mathbf{U}^T \mathbf{H}^T \mathbf{W} \mathbf{H} \mathbf{U})^{-1} \mathbf{U}^T)$ with $\mathbf{H}^T \mathbf{W} \mathbf{H} = \mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1 + H_2^T w_2 H_2$, $\mathbf{C}_2 (\square \mathbf{U}(\mathbf{U}^T \mathbf{H}_1^T \mathbf{W}_1 \mathbf{H}_1 \mathbf{U})^{-1} \mathbf{U}^T)$ exists if and only if $H_2 \mathbf{C}_1 H_2^T \neq w_2^{-1}$ holds. If \mathbf{C}_2 exists, then

$$\mathbf{C}_2 = \mathbf{C}_1 - \mathbf{C}_1 H_2^T (H_2 \mathbf{C}_1 H_2^T - w_2^{-1})^{-1} H_2 \mathbf{C}_1 \quad (3.21)$$

Clearly, this corollary is a simplified case of Theorem 1 and is to remove a single observation. It indicates that removing a simple observation is equivalent to adding in the corresponding SCO with the “negative weight”. It is clear that the theorem and corollary is also applicable to the unconstrained WLS problem. This is the key to the unification of the unconstrained and LE-constrained RLS.

The Initialization Procedure

Our recursive initialization utilizes the above basic tools to introduce and remove auxiliary observations. Consequently, an initializing procedure, which is the major focus to be discussed, is developed to perform a recursive and exact initialization for the LE-constrained linear LS. The development takes advantage of the following properties of the LE-constrained WLS solutions:

Theorem 2 (Properties of LE-constrained WLS): Given $\text{rank}([\mathbf{A}^T \mathbf{H}^T]^T) = N$, \mathbf{C}_{lec} (and $\hat{\mathbf{x}}_{\text{lec}}$) in (3.8) has the following properties:

1) If $\text{rank}(\mathbf{H}) = N$ (\mathbf{C}_{uc} thus exists), then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{uc}} - \mathbf{C}_{\text{uc}} \mathbf{A}^T (\mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{C}_{\text{uc}} \\ \hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{x}}_{\text{uc}} + \mathbf{C}_{\text{uc}} \mathbf{A}^T (\mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T)^{-1} (\mathbf{B} - \mathbf{A} \hat{\mathbf{x}}_{\text{uc}}) \end{cases} \quad (3.22)$$

2) If $[\mathbf{A}^T \mathbf{H}^T]^T$ is an $N \times N$ matrix, then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{lec}, N \times N} (\square \begin{bmatrix} \mathbf{A} \\ \mathbf{H} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} \end{bmatrix} [\mathbf{A}^T \mathbf{H}^T]^T)^{-1} \\ \hat{\mathbf{x}}_{\text{lec}} = ([\mathbf{A}^T \mathbf{H}^T]^T)^{-1} [\mathbf{B}^T \mathbf{z}^T]^T \end{cases} \quad (3.23a)$$

3) Given $\mathbf{H}_1 = [\mathbf{H} \mathbf{A}_1]$, $\mathbf{W}_1 = \text{diag}(\mathbf{W}, \mathbf{W}_{A1})$, and \mathbf{W}_{A1} is PD, if the rows of \mathbf{A}_1 are all from \mathbf{A} ,

then

$$\begin{cases} \mathbf{C}_1 = \mathbf{C}_{\text{lec}} \\ \hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_{\text{lec}} \end{cases} \quad (3.23b)$$

where $\hat{\mathbf{x}}_1$ (and \mathbf{C}_1) is the LE-constrained solution subject to (3.6) and using real observation data having $\mathbf{H}_1, \mathbf{W}_1$ and $\mathbf{z}_1 = [\mathbf{z}^T \mathbf{z}_{A1}^T]^T$ as the coefficient, weight and value, respectively.

4) If $\mathbf{A} = [\mathbf{A}_1^T \mathbf{A}_2^T]^T$ and $\text{rank}([\mathbf{H}^T \mathbf{A}_1^T]^T) = N$, then

$$\begin{cases} \mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{lec},2} \\ \hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{x}}_{\text{lec},2} \end{cases} \quad (3.24a)$$

with

$$\begin{cases} \mathbf{S}_{\text{lec},2} \square \mathbf{A}_2 \mathbf{C}_{\text{lec},1} \mathbf{A}_2^T, \mathbf{K}_2 \square \mathbf{C}_{\text{lec},1} \mathbf{A}_2^T \mathbf{S}_{\text{lec},2}^{-1} \\ \mathbf{C}_{\text{lec},2} \square \mathbf{C}_{\text{lec},1} - \mathbf{K}_2 \mathbf{S}_{\text{lec},2} \mathbf{K}_2^T \\ \hat{\mathbf{x}}_{\text{lec},2} = \hat{\mathbf{x}}_{\text{lec},1} + \mathbf{K}_2 (\mathbf{B}_2 - \mathbf{A}_2 \hat{\mathbf{x}}_{\text{lec},1}) \end{cases} \quad (3.24b)$$

and

$$\begin{cases} \mathbf{C}_{\text{lec},1} = \mathbf{U}_1 (\mathbf{U}_1^T \mathbf{H}^T \mathbf{W} \mathbf{H} \mathbf{U}_1)^{-1} \mathbf{U}_1^T \\ \hat{\mathbf{x}}_{\text{lec},1} = \mathbf{A}_1^+ \mathbf{B}_1 + \mathbf{C}_{\text{lec},1} \mathbf{H}^T \mathbf{W} (\mathbf{z} - \mathbf{H} \mathbf{A}_1^+ \mathbf{B}_1) \end{cases} \quad (3.24c)$$

where $[\tilde{\mathbf{U}}_1 \mathbf{U}_1][\tilde{\mathbf{U}}_1 \mathbf{U}_1]^T = \mathbf{I}$ and $\text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$.

A proof of Theorem 2 is given in Appendix A. In fact, this theorem is a corollary of Theorem 3 in Part I (Chapter 2), where Statement 1) provides a way to compute \mathbf{C}_{lec} from \mathbf{C}_{uc} , and Statement 2) presents a specific \mathbf{C}_{lec} . Statement 3) shows that, if observation coefficient \mathbf{H}_1 and constraint coefficient \mathbf{A} have common rows \mathbf{A}_1 , then the observations related to \mathbf{A}_1 have no effect on the eventual solution. Statement 4) indicates that multiple constraints can be imposed recursively (sequentially). Just as Chapter 2 declares, these properties imply that the LE constraints can be treated as special observations (POs) with $+\infty$ as weights (or equivalently the observation errors are zero). Each constraint can be processed sequentially in any place in the overall observation-and-constraint data sequence. Correspondingly,

$$\mathbf{S}_{\text{lec}} = \lim_{w \rightarrow +\infty} \mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T + w^{-1} \mathbf{I} = \mathbf{A} \mathbf{C}_{\text{uc}} \mathbf{A}^T \quad (3.25)$$

In this chapter, for the convenience of description, it is assumed that the LE constraint is explicitly given and is imposed in the initialization. In fact, the consequent initialization techniques can be applied to solve problems with implicit LE constraints.

Based on the above theorems and corollaries, an LE-constrained Initialization (LECI) procedure is designed to initialize the RLS with LE constraint, which is also applicable to unconstrained problems after omitting the Constraint and SCO recursion in its step 2) bellows.

LECI Procedure

1) Auxiliary initialization: $\tilde{m} = 0, t = 0, m = 0$

$$\begin{cases} \tilde{\mathbf{C}}_{\tilde{m}=0} = (\sum_{s=1}^N H_{\text{au},s}^T w_{\text{au},s} H_{\text{au},s})^{-1} = \text{diag}([p_1 \ \dots \ p_N]) \\ \hat{\mathbf{y}}_{\tilde{m}=0} = [z_{\text{au},1} \ \dots \ z_{\text{au},N}]^T \\ \text{Rec_au} = [0 \ \dots \ 0]_{1 \times N}, \tilde{\mathbf{H}}_{\text{b},\tilde{m}=0} = \mathbf{I} \end{cases} \quad (3.26)$$

2) Constraint and SCO (C&S) recursion: $t := t + 1$

2.1) Constraint imposing: $\tilde{m} := \tilde{m} + 1$

$$\begin{cases} \tilde{\mathbf{S}}_{\tilde{m}} \sqsubseteq \mathbf{A}_t \tilde{\mathbf{C}}_{\tilde{m}-1} \mathbf{A}_t^T \\ \tilde{\mathbf{K}}_{\tilde{m}} \sqsubseteq \tilde{\mathbf{C}}_{\tilde{m}-1} \mathbf{A}_t^T \tilde{\mathbf{S}}_{\tilde{m}}^{-1} \\ \hat{\mathbf{y}}_{\tilde{m}} = \hat{\mathbf{y}}_{\tilde{m}-1} + \tilde{\mathbf{K}}_{\tilde{m}} (C_t - \mathbf{A}_t \hat{\mathbf{y}}_{\tilde{m}-1}) \\ \tilde{\mathbf{C}}_{\tilde{m}} = \tilde{\mathbf{C}}_{\tilde{m}-1} - \tilde{\mathbf{K}}_{\tilde{m}} \tilde{\mathbf{S}}_{\tilde{m}} \tilde{\mathbf{K}}_{\tilde{m}}^T \\ \tilde{\mathbf{H}}_{\text{b},\tilde{m}} = \tilde{\mathbf{H}}_{\text{b},\tilde{m}-1} \end{cases} \quad (3.27)$$

2.2) Conditional SCO processing:

For s from 1 to N , if $\text{Rec_au}(s) = 1$ and the (current) $\tilde{\mathbf{C}}_{\tilde{m}}$'s s th diagonal entry

$$\tilde{c}_{s,s} \neq p_s : \tilde{m} := \tilde{m} + 1$$

$$\begin{cases} \tilde{\mathbf{S}}_{\tilde{m}} \sqsubseteq \tilde{c}_{s,s} - p_s \\ \tilde{\mathbf{K}}_{\tilde{m}} \sqsubseteq \tilde{\mathbf{C}}_{\tilde{m}-1} H_{\text{au},s}^T \tilde{\mathbf{S}}_{\tilde{m}}^{-1} \\ \hat{\mathbf{y}}_{\tilde{m}} = \hat{\mathbf{y}}_{\tilde{m}-1} + \tilde{\mathbf{K}}_{\tilde{m}} (z_{\text{au},s} - H_{\text{au},s} \hat{\mathbf{y}}_{\tilde{m}-1}) \\ \tilde{\mathbf{C}}_{\tilde{m}} = \tilde{\mathbf{C}}_{\tilde{m}-1} - \tilde{\mathbf{K}}_{\tilde{m}} \tilde{\mathbf{S}}_{\tilde{m}} \tilde{\mathbf{K}}_{\tilde{m}}^T \end{cases} \quad (3.28a)$$

$$\begin{cases} \text{Rec_au}_s = 1, \tilde{\mathbf{H}}_{b,\tilde{m}} = [\tilde{\mathbf{H}}_{b,\tilde{m}-1}^T A_t^T]^T \\ \text{remove } H_{\text{au},s} \text{ out of } \tilde{\mathbf{H}}_{b,\tilde{m}} \end{cases} \quad (3.28b)$$

The C&S recursion is completed at $t = T$.

3) RO and SCO (R&S) recursion: $m := m + 1$

3.1) Real-observation (RO) processing: $\tilde{m} := \tilde{m} + 1$

$$\begin{cases} \tilde{\mathbf{S}}_{\tilde{m}} \triangleq H_m \tilde{\mathbf{C}}_{\tilde{m}-1} H_m^T + w_m^{-1} \\ \tilde{\mathbf{K}}_{\tilde{m}} \triangleq \tilde{\mathbf{C}}_{\tilde{m}-1} H_m^T \tilde{\mathbf{S}}_{\tilde{m}}^{-1} \\ \hat{\mathbf{y}}_{\tilde{m}} = \hat{\mathbf{y}}_{\tilde{m}-1} + \tilde{\mathbf{K}}_{\tilde{m}} (\mathbf{z}_m - H_m \hat{\mathbf{y}}_{\tilde{m}-1}) \\ \tilde{\mathbf{C}}_{\tilde{m}} = \tilde{\mathbf{C}}_{\tilde{m}-1} - \tilde{\mathbf{K}}_{\tilde{m}} \tilde{\mathbf{S}}_{\tilde{m}} \tilde{\mathbf{K}}_{\tilde{m}}^T \\ \tilde{\mathbf{H}}_{b,\tilde{m}} = \tilde{\mathbf{H}}_{b,\tilde{m}-1} \end{cases} \quad (3.29)$$

3.2) Conditional SCO processing:

Same as in 2.2)

The R&S recursion is stopped after $\text{Rec_au} = [1 \cdots 1]$ or $m = M$.

4) Output Rec_au , \tilde{M} ($\square m$), $\tilde{\mathbf{H}}_b$ ($\square \tilde{\mathbf{H}}_{b,\tilde{m}}$), $\hat{\mathbf{y}}$ ($\triangleq \hat{\mathbf{y}}_{\tilde{m}}$) and $\tilde{\mathbf{C}}$ ($\triangleq \tilde{\mathbf{C}}_{\tilde{m}}$).

Here, t is constraint index and \tilde{m} recursion index.

Analysis (and proof):

In the LE-constrained WLS, a full column rank $[\mathbf{A}^T \tilde{\mathbf{H}}^T]^T \Leftrightarrow$ existence of $\tilde{\mathbf{C}}_{\text{lec}}$. Therefore, the existence of $\tilde{\mathbf{C}}_{\tilde{m}}$ is equivalent to the full rank of $\tilde{\mathbf{H}}_{\tilde{m}} (\square [\mathbf{H}_{\text{au},\text{left}}^T \mathbf{A}_{\text{lt}}^T \mathbf{H}_{\text{lt},m}^T]^T)$.

In the procedure, step 1) introduces N SAO's to build an auxiliary initialization with $\tilde{\mathbf{H}}_0 = \mathbf{I}$, which starts the RLS recursion from the first constraint or RO. In the constraint-processing, T constraints are imposed sequentially according to the statement 2) of Theorem 2. After each constraint processing in step 2.1), step 2.2) checks whether one (at most one, shown later) of the remaining SAO's is removable, based on Corollary 1.1. If yes, the corresponding SCO is added. Similarly, in the subsequent data-processing recursion, step 3.1) processes a new RO while step 3.2) checks (and deletes) a removable SAO. Note that the checking-and-deleting mechanism adopted in both C&S and R&S recursions removes the N auxiliary rows in $\tilde{\mathbf{H}}_0$ as

soon as possible (ASAP), provided that $\text{rank}(\tilde{\mathbf{H}}_{\tilde{m}})$ remains N for all \tilde{m} . Therefore, before Rec_au flips to all-ones, a recursion cycle to process a new constraint or RO is always executable since the MI-lemma conditions for updating $\tilde{\mathbf{C}}_{\tilde{m}}$ is always satisfied. In particular, the “for” loops in 2.2) and 3.2) keep checking all the remaining SAO’s for removable one in each cycle. The found SAO is deleted at once. As a result, “ASAP” is guaranteed. In this design, $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}}$, which is to store a set of base rows (from $\tilde{\mathbf{H}}_{\tilde{m}}$) spanning the N -dimensional space, can be constructed as follows:

i) $\tilde{\mathbf{H}}_{\mathbf{b},0} = \mathbf{I}$ with rows from the N SAO’s; ii) When a new constraint or RO is processed, $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}} = \tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}-1}$ since the rows in $\tilde{\mathbf{H}}_{\tilde{m}-1}$ are all passed to $\tilde{\mathbf{H}}_{\tilde{m}}$; iii) Once a SAO is removed after a constraint or RO processing, the new row A_t (or H_m) plus the $N-1$ rows (excluding the newly-removed SAO row) in $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}-1}$ composes a new basis $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}}$.

In iii), the newly-constructed $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}}$ must be a basis. Otherwise, $\text{rank}(\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}}) = N-1$ since $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}}$ inherits $N-1$ base rows from $\tilde{\mathbf{H}}_{\mathbf{b},\tilde{m}-1}$, which means that A_t (or H_m) is a linear combination of the $N-1$ base rows. Accordingly, $\text{rank}(\tilde{\mathbf{H}}_{\tilde{m}})$ should remain N if A_t (or H_m) is removed from $\tilde{\mathbf{H}}_{\tilde{m}}$. However, after the removal, the new $\tilde{\mathbf{H}}_{\tilde{m}}$ is equal to a fake $\tilde{\mathbf{H}}_{\tilde{m}-1}$ formed by deleting the SAO row which is undeletable at $\tilde{m}-1$. This is a contradiction.

The above analysis also implies that no more than one SAO can be removed within each C&S or R&S recursion cycle. Thus the “for” loop in steps 2.2) and 3.2) can find at most one removable SAO.

Furthermore, when the overall procedure is terminated, the constraint or RO rows stored in $\tilde{\mathbf{H}}_{\mathbf{b}}$ must compose a basis of $\text{row}([\mathbf{A}^T \mathbf{H}_{1:M}^T]^T)$ since the SAO rows left in the eventual $\tilde{\mathbf{H}}_{\tilde{m}}$ (and so $\tilde{\mathbf{H}}_{\mathbf{b}}$) are not removable. In particular, if the N SAO rows are all removed, then

$\tilde{\mathbf{H}}_b$ contains N constraint or RO rows and the corresponding $[\mathbf{A}^T \mathbf{H}_{1:m}^T]^T$ has full column rank. End of Analysis.

Clearly, the above initialization procedure does not require extra work to check the rank of $\tilde{\mathbf{H}}$ in advance. It gets $\text{rank}(\tilde{\mathbf{H}})$ as a byproduct. In brief, the outputs of the procedure are summarized by the following theorems.

Theorem 3 (Rank check by LECI): In the LECI for the LE-constrained WLS defined by (3.2) with (3.6), if the output $\tilde{\mathbf{H}}_b$ contains $N_1 (\leq N)$ constraint or RO rows, then $\text{rank}(\tilde{\mathbf{H}}) = N_1$ and the $N_1 (\leq N)$ rows compose a basis of $\text{row}(\tilde{\mathbf{H}})$.

In the case that $\text{rank}(\tilde{\mathbf{H}}) = N$, we have

Theorem 4 (Full-rank initialization by LECI): In the LECI for the LE-constrained WLS defined by (3.2) with (3.6), if $\text{rank}(\tilde{\mathbf{H}}) = N$, then \tilde{M} and \tilde{C} are identical to the minimal M_0 such that $\text{rank}([\mathbf{A}^T \mathbf{H}_{1:M_0}^T]^T) = N$ and the corresponding initial estimation error covariance like matrix $\mathbf{C}_{\text{lec},0} (\square \mathbf{U}(\mathbf{U}^T \mathbf{H}_{1:M_0}^T \mathbf{W}_{1:M_0} \mathbf{H}_{1:M_0} \mathbf{U})^{-1} \mathbf{U}^T)$ exists, respectively.

In the deficient-rank case (i.e., $\text{rank}(\tilde{\mathbf{H}}) < N$), a reduced-dimensional alternative estimand, which comes from the basis of $\text{row}(\tilde{\mathbf{H}})$ and has a dimension of $N_1 (< N)$, can be determined using \mathbf{H}_b and \tilde{C} conveniently. The corresponding estimate is for a set of refined constraints and RO's. Once new observations are available later, the estimate plus the corresponding estimation-error covariance like matrix can take the role of the original data in the recalculation (according to the two-step LS [Haupt96]). In this sense, the refinement is lossless in WLS processing (see Theorem 5). Actually, this reducing-dimension mapping can provide good practical resolutions to the deficient problem which is discussed next. Rather than the conventional minimum-norm solution which has great value in theoretical development, the proposed reducing-dimension mapping may be directly applied to many practical problems.

3.3.3 Deficient-Rank Processing

To study the deficient-rank WLS, the problem after a linear equivalent transformation is first investigated as follows.

Fact 1: In the LE-constrained WLS defined by (3.2) with (3.6), if \mathbf{T} is a nonsingular square matrix and $\mathbf{x}_{\text{tran}} = \mathbf{T}\mathbf{x}$, then

$$\begin{cases} \hat{\mathbf{x}}_{\text{tran}} = \mathbf{T}\hat{\mathbf{x}} \\ \mathbf{C}_{\text{tran}} = \mathbf{T}\mathbf{C}_{\text{lec}}\mathbf{T}^T \end{cases} \quad (3.30)$$

Here, if \mathbf{x}_{tran} is treated as the new estimand, then $\mathbf{H}_{\text{tran}} = \mathbf{H}\mathbf{T}^{-1}$, $\mathbf{A}_{\text{tran}} = \mathbf{A}\mathbf{T}^{-1}$, $\mathbf{U}_{\text{tran}} = \mathbf{T}\mathbf{U}$ and $\mathbf{W}_{\text{tran}} = \mathbf{W}$ in the new formulation. Using (3.8), (3.30) is verified.

Base on Fact 1, the following theorem can be derived to estimate the reduced- dimensional alternative estimand for a set of refined and lossless constraints and RO's.

Theorem 5 (Reduced-dimensional WLS): In the LECI procedure for the LE-constrained WLS defined by (3.2) with (3.6), if $\text{rank}(\check{\mathbf{H}}) < N$ and the rows of $\check{\mathbf{H}}_{\text{arb}}$ compose the basis of $\text{row}(\check{\mathbf{H}})$, then $\mathbf{x}_{\text{arb}} (\square \check{\mathbf{H}}_{\text{arb}}\mathbf{y})$ is uniquely determined, and

$$\begin{cases} \mathbf{C}_{\text{arb}} = \check{\mathbf{H}}_{\text{arb}} \tilde{\mathbf{C}} \check{\mathbf{H}}_{\text{arb}}^T \\ \hat{\mathbf{x}}_{\text{arb}} = \check{\mathbf{H}}_{\text{arb}} \hat{\mathbf{y}} \end{cases}, \quad (3.31)$$

Clearly, the LECI procedure has provided $\check{\mathbf{H}}_{\text{b}}$ which stores a basis of $\text{row}(\check{\mathbf{H}})$. Consequently, the alternative estimand \mathbf{x}_{b} which is a set of refined (and constrained) RO's can be easily estimated by Theorem 5. Note that, from (3.52), we also have

$$\tilde{\mathbf{C}}_{\text{left}} = [\mathbf{0} \ \mathbf{I}_{\text{left}}] \tilde{\mathbf{C}} [\mathbf{0} \ \mathbf{I}_{\text{left}}]^T \quad (3.32)$$

Based on (3.31) and (3.32), it can be easily verified that

$$\begin{bmatrix} \mathbf{C}_{\text{arb}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{\text{left}} \end{bmatrix} = \begin{bmatrix} \check{\mathbf{H}}_{\text{arb}} \\ \mathbf{0} \ \mathbf{I}_{\text{left}} \end{bmatrix} \tilde{\mathbf{C}} \begin{bmatrix} \check{\mathbf{H}}_{\text{arb}} \\ \mathbf{0} \ \mathbf{I}_{\text{left}} \end{bmatrix}^T \quad (3.33)$$

where $\begin{bmatrix} \tilde{\mathbf{H}}_{\text{arb}} \\ \mathbf{0} \mathbf{I}_{\text{left}} \end{bmatrix}$, as well as $\begin{bmatrix} \tilde{\mathbf{H}}_{\text{b}} \\ \mathbf{0} \mathbf{I}_{\text{left}} \end{bmatrix}$, has full rank. From (3.33), it can be stated that, given $\tilde{\mathbf{H}}_{\text{arb}}, \mathbf{C}_{\text{arb}}, [\mathbf{0} \mathbf{I}_{\text{left}}]$ and the corresponding $\tilde{\mathbf{C}}_{\text{left}}$, $\tilde{\mathbf{C}}$ can be recovered, based on which the LECI can go on whenever new data arrive. Here, the coefficient of the remaining SAO by the LECI procedure, which is assumed to be $[\mathbf{0} \mathbf{I}_{\text{left}}]$ (for simplicity without loss of generality), is determined by $\tilde{\mathbf{H}}_{\text{b}}$. \mathbf{C}_{left} is also from the output of the LECI. In fact, neither of these conditions is mandatory. First, $\tilde{\mathbf{C}}_{\text{left}}$ can be an arbitrary positive diagonal matrix since it is always identical to $\mathbf{W}_{\text{left}}^{-1}$ which is a set of originally-given arbitrary positive numbers (see (3.53)). Furthermore, we can construct a different SAO set which has the minimum number of elements to augment $\tilde{\mathbf{H}}_{\text{arb}}$ till full rank. The only difference is that the diagonal elements of $\tilde{\mathbf{C}}_{\text{left}}$ will occupy new diagonal positions of (new) $\tilde{\mathbf{C}}$. Therefore, it can be concluded that the refined data according to (3.54) is lossless in the WLS sense.

Furthermore, due to the rank deficiency of $\tilde{\mathbf{H}}$, some components of \mathbf{x} may be uniquely determined while some others can not. Denote all the uniquely-determined components as \mathbf{x}_{ud} , and then the next question is: How to find \mathbf{x}_{ud} and further construct a basis of $\text{row}(\tilde{\mathbf{H}})$, stored in $\tilde{\mathbf{H}}_{\text{sim}}$, to make $\mathbf{x}_{\text{sim}} (\square \tilde{\mathbf{H}}_{\text{sim}} \mathbf{y})$ contain \mathbf{x}_{ud} .

Based on the LECI, the following theorem provides a way to finding \mathbf{x}_{ud} :

Theorem 6 (Simple basis determination): In the LECI procedure for the LE-constrained WLS defined by (3.2) with (3.6), if $\text{rank}(\tilde{\mathbf{H}}) < N$ and the undeletable-SAO-row index-number set is S_{left} , then $I_i(\square [0_1 \cdots 0_{i-1} \ 1_i \ 0_{i+1} \cdots 0_N]) \notin \text{row}(\tilde{\mathbf{H}})$ if and only if $\tilde{\mathbf{C}}$'s entry $c_{i,j}$ is not zero for at least one $j (\in S_{\text{left}})$.

A proof of Theorem 6 is given in Appendix C. This theorem shows that $C_{i,\text{left}} = \mathbf{0} \Leftrightarrow I_i \in \text{row}(\tilde{\mathbf{H}}) \Leftrightarrow \mathbf{x}_i \in \mathbf{x}_{\text{ud}}$. Suppose that \mathbf{x}_{ud} contains N_{ud} entries. Then the corresponding N_{ud} different I_i 's should be in $\tilde{\mathbf{H}}_{\text{sim}}$. Accordingly, the other $N_1 - N_{\text{ud}}$ rows of $\tilde{\mathbf{H}}_{\text{sim}}$ can be found from $\tilde{\mathbf{H}}_{\text{b}}$ as follows:

a) Delete (from $\tilde{\mathbf{H}}_{\text{b}}$) the N_{ud} columns storing the coefficients of \mathbf{x}_{ud} ; b) Delete (from the modified $\tilde{\mathbf{H}}_{\text{b}}$) the $N - N_1$ columns storing the coefficients of \mathbf{x}_{left} ; c) Find $N_1 - N_{\text{ud}}$ linearly-independent rows in the “shrunk” $\tilde{\mathbf{H}}_{\text{b}}$; d) Augment the linearly-independent rows: the coefficients of \mathbf{x}_{ud} are set zero while those of \mathbf{x}_{left} are assigned the original values (in the original $\tilde{\mathbf{H}}_{\text{b}}$). Note that the rank of the shrunk $\tilde{\mathbf{H}}_{\text{b}}$ is $N_1 - N_{\text{ud}}$ after the deleting operations in a) and b). Thus the operation to find a basis of $\text{row}(\tilde{\mathbf{H}}_{\text{b}})$ is always valid in c). Obviously, with this method, finding $\tilde{\mathbf{H}}_{\text{sim}}$ is equivalent to finding the basis of the shrunk $\tilde{\mathbf{H}}_{\text{b}}$. In general, the latter has a smaller dimension. Existing rank-check methods [Chan87], including the one proposed in Theorem 3, can be employed.

3.4 Overall Algorithm and Implementation

The overall CRLS algorithm, which includes the initialization, LE-constraint imposition and rank check in the LECI procedure as well as the conventional RLS cycles, can obtain the exact recursive solutions (up to current data) ASAP. It comprises the following steps:

1) Auxiliary initialization:

Same as step 1) in the LECI procedure;

2) C&S recursion:

Same as step 2) in the LECI procedure;

3) Initial R&S recursion:

Same as step 3) in the LECI procedure, except that the recursion is completed

if $\text{Rec_au} = [1 \cdots 1]$ or $m = M$. If $\text{Rec_au} = [1 \cdots 1]$ and $m < M$, turn to 4);
if $\text{Rec_au} = [1 \cdots 1]$ and $m = M$, turn to 5); if $\text{Rec_au} \neq [1 \cdots 1]$, turn to 6).

4) Conventional RO recursion: $m := m + 1, \tilde{m} := \tilde{m} + 1$

$$\begin{cases} \tilde{\mathbf{S}}_{\tilde{m}} \triangleq \mathbf{H}_m \tilde{\mathbf{C}}_{\tilde{m}-1} \mathbf{H}_m^T + \mathbf{w}_m^{-1}, \tilde{\mathbf{K}}_{\tilde{m}} \triangleq \tilde{\mathbf{C}}_{\tilde{m}-1} \mathbf{H}_m^T \tilde{\mathbf{S}}_{\tilde{m}}^{-1} \\ \hat{\mathbf{y}}_{\tilde{m}} = \hat{\mathbf{y}}_{\tilde{m}-1} + \tilde{\mathbf{K}}_{\tilde{m}} (z_m - \mathbf{H}_m \hat{\mathbf{y}}_{\tilde{m}-1}) \\ \tilde{\mathbf{C}}_{\tilde{m}} = \tilde{\mathbf{C}}_{\tilde{m}-1} - \tilde{\mathbf{K}}_{\tilde{m}} \tilde{\mathbf{S}}_{\tilde{m}} \tilde{\mathbf{K}}_{\tilde{m}}^T \end{cases} \quad (3.35)$$

This recursion is completed at $m = M$.

5) Termination. Output $\hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{y}}_{\tilde{m}}$ and $\mathbf{C}_{\text{lec}} = \tilde{\mathbf{C}}_{\tilde{m}}$.

6) Reduced-dimensional processing:

$$\begin{cases} \hat{\mathbf{x}}_{\text{b}} = \tilde{\mathbf{H}}_{\text{b},\tilde{m}} \hat{\mathbf{y}}_{\tilde{m}} \\ \mathbf{C}_{\text{b}} = \tilde{\mathbf{H}}_{\text{b},\tilde{m}} \tilde{\mathbf{C}}_{\tilde{m}} \tilde{\mathbf{H}}_{\text{b},\tilde{m}}^T \end{cases} \quad (3.36)$$

7) Termination. Output $\hat{\mathbf{x}}_{\text{b}}, \mathbf{C}_{\text{b}}$ and $\tilde{\mathbf{H}}_{\text{b},\tilde{m}}$.

In particular, if the numerical condition of the observation set is not poor, $\tilde{\mathbf{C}}_0$ in step 1) can be simply set to the identity matrix \mathbf{I} . Otherwise, simple values, which may be close to the diagonal numbers in \mathbf{C}_{lec} , can be assigned to each p_s individually, which will guarantee that the introduction of SAO's does not degrade the numerical condition of the RO set. In step 3), if the eventual $\text{Rec_au} = [1 \cdots 1]$, then the minimal M_0 making $[\mathbf{A}^T \mathbf{H}_{1:M_0}^T]^T$ full-column-rank is thus determined and $M_0 = m$. Otherwise, if $\text{Rec_au} \neq [1 \cdots 1]$ after M RO's are all processed, then the RO set is insufficient to determine \mathbf{x} uniquely. Correspondingly, the reduced-dimensional estimand $\tilde{\mathbf{H}}_{\text{b}} \mathbf{x}$ can be estimated alternatively in 6). As Theorem 5 states, $\tilde{\mathbf{H}}_{\text{b}}$ is not the only mapping matrix. In fact, all the bases of $\text{row}(\tilde{\mathbf{H}})$ are candidates. In particular, Theorem 7 provides an easy way to find the uniquely- determined components (of \mathbf{x}) stored as \mathbf{x}_{ud} . Consequently, \mathbf{x}_{sim} , which contains \mathbf{x}_{ud} and is identical to \mathbf{x}_{b} through an equivalent linear transform, can be used as the new reduced- dimensional alternative estimand.

Note that step 6) is not necessarily implemented recursively. In essence, it is a batch problem and a recursive method could be employed as long as the method has good performance in the application.

3.5 Performance Analysis

The RLS has well-known advantages in time and storage. Its recursion does not involve any matrix inversion operation and thus has a lower computational complexity. In general, without considering the initialization, the computation of the RLS is $O(N^2)$ (order of N^2) per cycle while those of batch LS algorithms are usually $O(N^3)$ (or with a lower order but a much larger leading coefficient [Stoer02] [Golub96] [Horn85]). In some specific applications, the computation of the fast RLS algorithms can even be reduced to $O(N)$. Furthermore, both the number of algebraic operations and the amount of required memory locations at each cycle are fixed. Particularly, the sequential data processing can make use of the time over the data-accumulation period flexibly. Thus the RLS is particularly suitable to real-time low-dimensional applications where high-redundancy data are involved.

In the CRLS, via introducing SAO's, the initialization of the RLS is performed by the simple auxiliary initialization plus the minimum C&S and R&S cycles. Clearly, the CRLS inherits and enhances good properties of the RLS since the recursion starts from the first piece of (real) data and the LE constraints are also imposed recursively. To our knowledge, the CRLS initialization is the simplest among the recursive methods to initializing the RLS exactly. On the other hand, owing to the recursion, the CRLS initialization has better performance than those by batch methods. For instance, Cholesky decomposition (CD) has been commonly used as an efficient tool to compute the inverse of a symmetric PD matrix. Compared with a batch initialization using the CD-based matrix inversion, the superiority of the CRLS (initialization) arises from the following aspects:

- 1) The proposed method can distribute the processing time over the data-accumulation period while the CD-based batch one usually has to wait till all data are available.

2) In general, the computation of processing a RO (or a LE constraint) consists of $\frac{3}{2}N^2$ multiplications and $\frac{3}{2}N^2$ additions per cycle, and that of a SCO requires $\frac{1}{2}N^2$ multiplications and $\frac{1}{2}N^2$ additions per cycle only. Therefore, without considering the effect of accumulating data, the CRLS still has a low computational complexity relative to the CD-based one (see [Golub96]).

3) With the CRLS, the initialization of LE-constrained RLS solution, which (in the batch form) usually involves MP inverses, is made as simple as for the unconstrained ones now.

4) In sparse applications, the CRLS can benefit more from the sparsity because its recursion can make full use of the sparse structure of the observation coefficients.

5) The observability analysis in the CRLS requires no extra computation while the batch one usually needs additional pre-processing to check the rank of \mathbf{H} (or $[\mathbf{A}^T \mathbf{H}^T]^T$). The result by the CRLS is numerically consistent with the existence of \mathbf{C} in calculation.

Obviously, the CRLS is still applicable to the traditional RLS problems, such as adaptive filtering [Haykin01], which are usually of a low dimension but have high-redundancy data. Then the efficiency can be enhanced to some extent since the computation can benefit from a generically simpler initialization.

For high-dimensional WLS applications, partly limited by the costly batch initialization, the RLS has not been widely adopted. Particularly, in high-dimensional but low-redundancy cases, the initialization procedure can take over most of the processing period and the RLS thus helps little. Fortunately, the CRLS now provides a simple and recursive initialization for the RLS and starts the recursion from the first RO. Therefore, the CRLS approach, when applied to real-time high-dimensional applications can shorten the data-processing period significantly.

In addition, the concept of auxiliary observations, with which the initialization of the CRLS has been simplified significantly, can also be extended and utilized to improve the detection and identification of outliers and also to ease a possible wrong-data correction (see [Albert65] for details).

3.6 Appendices

3.6.1 Appendix A

Proof of Theorem 2:

Statement 1):

This well-know existing result can be derived from the normal equation (3.7) explicitly.

Statement 2):

This statement is a corollary of Statement 2) in Theorem 3 of Part I (Chapter 3) since LS is a special WLS with all-one weights.

Statement 3):

This statement is a corollary of Statement 3) in Theorem 3 of Part I (Chapter 3).

Statement 4):

First, if $\text{rank}(\mathbf{H}) = N$, then \mathbf{C}_{lec} is given by (3.22), where

$$\mathbf{A}\mathbf{C}_{\text{uc}}\mathbf{A}^T = \begin{bmatrix} \mathbf{A}_1\mathbf{C}_{\text{uc}}\mathbf{A}_1^T & \mathbf{A}_1\mathbf{C}_{\text{uc}}\mathbf{A}_2^T \\ \mathbf{A}_2\mathbf{C}_{\text{uc}}\mathbf{A}_1^T & \mathbf{A}_2\mathbf{C}_{\text{uc}}\mathbf{A}_2^T \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{E} \\ \mathbf{F} & \mathbf{G} \end{bmatrix}$$

It is clear that \mathbf{D} and \mathbf{G} are PD since \mathbf{C}_{uc} is PD and \mathbf{A} , \mathbf{A}_1 and \mathbf{A}_2 are of full row rank. We also have

$$\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E} = \mathbf{A}_2[\mathbf{C}_{\text{uc}} - \mathbf{C}_{\text{uc}}\mathbf{A}_1^T(\mathbf{A}_1\mathbf{C}_{\text{uc}}\mathbf{A}_1^T)^{-1}\mathbf{A}_1\mathbf{C}_{\text{uc}}]\mathbf{A}_2^T$$

According to (3.8),

$$\mathbf{C}_{\text{lec},1} = \mathbf{U}_1(\mathbf{U}_1^T\mathbf{H}^T\mathbf{W}\mathbf{H}\mathbf{U}_1)^{-1}\mathbf{U}_1^T = \mathbf{C}_{\text{uc}} - \mathbf{C}_{\text{uc}}\mathbf{A}_1^T\mathbf{S}_{\text{lec},1}^{-1}\mathbf{A}_1\mathbf{C}_{\text{uc}} \quad (3.37)$$

So

$$\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E} = \mathbf{A}_2\mathbf{U}_1(\mathbf{U}_1^T\mathbf{H}^T\mathbf{W}\mathbf{H}\mathbf{U}_1)^{-1}\mathbf{U}_1^T\mathbf{A}_2^T \quad (3.38)$$

According to [Meyer00],

$$\text{rank}(\mathbf{A}_2\mathbf{U}_1) = \text{rank}(\mathbf{A}_2) - \dim(\text{Null}(\mathbf{U}_1^T) \cap \text{col}(\mathbf{A}_2^T))$$

Here, $[\tilde{\mathbf{U}}_1 \ \mathbf{U}_1][\tilde{\mathbf{U}}_1 \ \mathbf{U}_1]^T = \mathbf{I}$, $\text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$, and \mathbf{A} , \mathbf{A}_1 and \mathbf{A}_2 are all of full row rank.

Thus,

$$\text{Null}(\mathbf{U}_1^T) = \text{col}(\tilde{\mathbf{U}}_1) = \text{col}(\mathbf{A}_1^T)$$

Clearly, $\text{col}(\mathbf{A}_1^T) \cap \text{col}(\mathbf{A}_2^T) = \mathbf{0}^T$. Therefore,

$$\text{rank}(\mathbf{A}_2 \mathbf{U}_1) = \text{rank}(\mathbf{A}_2) \quad (3.39)$$

As a result, $\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E}$ is PD. Similarly, we can also show that $\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F}$ is also PD.

Since \mathbf{D}, \mathbf{G} , $\mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E}$ and $\mathbf{D} - \mathbf{E}\mathbf{G}^{-1}\mathbf{F}$ are all PD, Schur's identity is applicable. That is,

$$(\mathbf{A}\mathbf{C}_{uc}\mathbf{A}^T)^{-1} = \begin{bmatrix} \mathbf{A}_1\mathbf{C}_{uc}\mathbf{A}_1^T & \mathbf{A}_1\mathbf{C}_{uc}\mathbf{A}_2^T \\ \mathbf{A}_2\mathbf{C}_{uc}\mathbf{A}_1^T & \mathbf{A}_2\mathbf{C}_{uc}\mathbf{A}_2^T \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}$$

with

$$\mathbf{T}_{11} = \mathbf{S}_{lec,1}^{\circ -1} - \mathbf{S}_{lec,1}^{\circ -1} \mathbf{A}_1 \mathbf{C}_{uc} \mathbf{A}_2^T \mathbf{S}_{lec,2}^{\circ -1} \mathbf{A}_2 \mathbf{C}_{uc} \mathbf{A}_1^T \mathbf{S}_{lec,1}^{\circ -1}$$

$$\mathbf{T}_{22} = \mathbf{S}_{lec,2}^{\circ -1}$$

$$\mathbf{T}_{12} = -\mathbf{S}_{lec,1}^{\circ -1} \mathbf{A}_1 \mathbf{C}_{uc} \mathbf{A}_2^T \mathbf{S}_{lec,2}^{\circ -1}$$

$$\mathbf{T}_{21} = -\mathbf{S}_{lec,2}^{\circ -1} \mathbf{A}_2 \mathbf{C}_{uc} \mathbf{A}_1^T \mathbf{S}_{lec,1}^{\circ -1}$$

where

$$\mathbf{S}_{lec,1}^{\circ} \square \mathbf{A}_1 \mathbf{C}_{uc} \mathbf{A}_1^T \quad (\text{PD, the inverse thus exists})$$

$$\mathbf{S}_{lec,2}^{\circ} \square \mathbf{A}_2 (\mathbf{C}_{uc} - \mathbf{C}_{uc} \mathbf{A}_1^T \mathbf{S}_{lec,1}^{\circ -1} \mathbf{A}_1 \mathbf{C}_{uc}) \mathbf{A}_2^T = \mathbf{G} - \mathbf{F}\mathbf{D}^{-1}\mathbf{E}$$

Substituting $\mathbf{T}_{11}, \mathbf{T}_{12}, \mathbf{T}_{21}$ and \mathbf{T}_{22} into \mathbf{C}_{lec} in (3.22) and substituting $\mathbf{C}_{lec,1}$ in (3.37) into $\mathbf{C}_{lec,2}$ in

(3.24), we can verify that

$$\mathbf{C}_{lec} = \mathbf{C}_{lec,2} \quad (3.40)$$

Second, if $\text{rank}(\mathbf{H}) < N$, then introduce a set of auxiliary observations which have $\mathbf{A}_1, \mathbf{I}_1$, and \mathbf{z}_1 as the coefficient, weight and value, respectively. Clearly, $\tilde{\mathbf{H}} = [\mathbf{H}^T \mathbf{A}_1^T]^T$ has full column rank:

$$\begin{cases} \tilde{\mathbf{C}}_{\text{lec},1} = \mathbf{U}_1[\mathbf{U}_1^T(\mathbf{H}^T\mathbf{W}\mathbf{H} + \mathbf{A}_1^T\mathbf{A}_1)\mathbf{U}_1]^{-1}\mathbf{U}_1^T \\ \tilde{\mathbf{C}}_{\text{lec}} = \mathbf{U}[\mathbf{U}^T(\mathbf{H}^T\mathbf{W}\mathbf{H} + \mathbf{A}_1^T\mathbf{A}_1)\mathbf{U}]^{-1}\mathbf{U}^T \end{cases} \quad (3.41)$$

According to (3.40), $\tilde{\mathbf{C}}_{\text{lec}}$ can also be calculated recursively as:

$$\tilde{\mathbf{C}}_{\text{lec},1} \rightarrow \tilde{\mathbf{C}}_{\text{lec},2} = \tilde{\mathbf{C}}_{\text{lec}} \quad (3.42)$$

Furthermore, since \mathbf{A}_1 is a part of the constraint coefficient, according to Statement 3, the effect of the auxiliary observations “disappear” automatically. That is,

$$\begin{cases} \tilde{\mathbf{C}}_{\text{lec},1} = \mathbf{C}_{\text{lec},1} \\ \tilde{\mathbf{C}}_{\text{lec}} = \mathbf{C}_{\text{lec}} \end{cases} \quad (3.43)$$

Therefore, the following recursive procedure is exact:

$$\tilde{\mathbf{C}}_{\text{lec},1} \rightarrow \mathbf{C}_{\text{lec},2} = \mathbf{C}_{\text{lec}} \quad (3.44)$$

Combining (3.40) and (3.44) yields $\mathbf{C}_{\text{lec}} = \mathbf{C}_{\text{lec},2}$. Similarly, $\hat{\mathbf{x}}_{\text{lec}} = \hat{\mathbf{x}}_{\text{lec},2}$. This completes the proof.

3.6.2 Appendix B

Proof of Theorem 5:

First, suppose that $\tilde{\mathbf{H}}_{\text{arb}} = \tilde{\mathbf{H}}_{\text{b}}$ and $\mathbf{x}_{\text{b}} (\square \tilde{\mathbf{H}}_{\text{b}}\mathbf{y})$ is to be determined. Without loss of generality, assuming that the undeleted SAO rows in $\tilde{\mathbf{H}}_{\text{b}}$ is $[\mathbf{0} \ \mathbf{I}_{\text{left}}]$, then

$$\tilde{\mathbf{H}}_{\text{b}} = \begin{bmatrix} \tilde{\mathbf{H}}_{\text{b}} \\ \mathbf{H}_{\text{left}} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}}_{\text{b}} \\ \mathbf{0} \ \mathbf{I}_{\text{left}} \end{bmatrix}, \quad \mathbf{y}_{\text{b}} = \tilde{\mathbf{H}}_{\text{b}}\mathbf{y} = \begin{bmatrix} \mathbf{x}_{\text{b}} \\ \mathbf{x}_{\text{left}} \end{bmatrix} \quad (3.45)$$

Reformulate the WLS problem using \mathbf{y}_{b} as the estimand:

$$\begin{cases} \hat{\mathbf{y}}_{\text{b}} = \min_{\mathbf{y}_{\text{b}}} J = (\mathbf{z} - \mathbf{H}_{\text{yb}}\mathbf{y}_{\text{b}})^T \mathbf{W}(\mathbf{z} - \mathbf{H}_{\text{yb}}\mathbf{y}_{\text{b}}) \\ \quad + (\mathbf{z}_{\text{left}} - \mathbf{H}_{\text{left}}\mathbf{y}_{\text{b}})^T \mathbf{W}_{\text{left}}(\mathbf{z}_{\text{left}} - \mathbf{H}_{\text{left}}\mathbf{y}_{\text{b}}) \\ \text{s. t. } \mathbf{A}_{\text{yb}}\mathbf{y}_{\text{b}} = \mathbf{B} \end{cases} \quad (3.46)$$

with $\mathbf{H}_{\text{yb}}\mathbf{y}_{\text{b}} = \mathbf{H}\mathbf{y}$, $\mathbf{A}_{\text{yb}}\mathbf{y}_{\text{b}} = \mathbf{A}\mathbf{y}$, and $\mathbf{H}_{\text{left}}\mathbf{y}_{\text{b}} = \mathbf{x}_{\text{left}}$. The rows of \mathbf{H} and \mathbf{A} are linear combinations

of the rows of $\check{\mathbf{H}}_b$, so $\mathbf{H}_{yb} = [\mathbf{H}_{yb1} \ \mathbf{0}]$, $\mathbf{A}_{yb} = [\mathbf{A}_{yb1} \ \mathbf{0}]$, $\mathbf{H}_{\text{left}} = [\mathbf{0} \ \mathbf{I}_{\text{left}}]$ (3.47)

and $\text{rank}([\mathbf{A}_{yb1}^T \ \mathbf{H}_{yb1}^T]^T) = \text{rank}(\check{\mathbf{H}})$ since $[\mathbf{A}_{yb}^T \ \mathbf{H}_{yb}^T]^T \check{\mathbf{H}}_b = \check{\mathbf{H}}$. Then (3.46) is equivalent to the following two independent subproblems that are over and exactly determined, respectively,

$$\begin{cases} \hat{\mathbf{x}}_b = \min_{\mathbf{x}_b} J_b = (\mathbf{z} - \mathbf{H}_{yb1} \mathbf{x}_b)^T \mathbf{W} (\mathbf{z} - \mathbf{H}_{yb1} \mathbf{x}_b) \\ \text{s. t. } \mathbf{A}_{yb1} \mathbf{x}_b = B \end{cases} \quad (3.48)$$

and $\hat{\mathbf{x}}_{\text{left}} = \min_{\mathbf{x}_{\text{left}}} J_{\text{left}} = (\mathbf{z}_{\text{left}} - \mathbf{x}_{\text{left}})^T \mathbf{W}_{\text{left}} (\mathbf{z}_{\text{left}} - \mathbf{x}_{\text{left}})$ (3.49)

Apply the RLS to solve (3.48) and (3.46) concurrently. In particular, select $\check{\mathbf{H}}_b$ and $\check{\mathbf{H}}_b$ for the initializations, respectively:

$$\mathbf{C}_{b,0} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\text{Hb}}^{-1} \end{bmatrix}, \quad \tilde{\mathbf{C}}_{yb,0} = \begin{bmatrix} \mathbf{C}_{b,0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\text{left}}^{-1} \end{bmatrix} \quad (3.50)$$

which are based on (3.23). Using (3.9) and (3.47), it is verified that $\mathbf{C}_{yb,\tilde{m}} = \begin{bmatrix} \mathbf{C}_{b,\tilde{m}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\text{left}}^{-1} \end{bmatrix}$ for

all \tilde{m} and eventually

$$\tilde{\mathbf{C}}_{yb} = \begin{bmatrix} \mathbf{C}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\text{left}}^{-1} \end{bmatrix} \quad (3.51)$$

According to Fact 1,

$$\tilde{\mathbf{C}}_{yb} = \check{\mathbf{H}}_b \tilde{\mathbf{C}} \check{\mathbf{H}}_b^T = \begin{bmatrix} \check{\mathbf{H}}_b \tilde{\mathbf{C}} \check{\mathbf{H}}_b^T & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{\text{left}} \end{bmatrix} \quad (3.52)$$

Thus,
$$\begin{cases} \mathbf{C}_b = \check{\mathbf{H}}_b \tilde{\mathbf{C}} \check{\mathbf{H}}_b^T \\ \tilde{\mathbf{C}}_{\text{left}} = \mathbf{W}_{\text{left}}^{-1} \end{cases} \quad (3.53)$$

Second, solve $\mathbf{x}_{\text{arb}} = \check{\mathbf{H}}_{\text{arb}} \mathbf{y}$. Since the rows in both $\check{\mathbf{H}}_{\text{arb}}$ and $\check{\mathbf{H}}_a$ compose basis of $\text{row}(\check{\mathbf{H}})$, there exists a nonsingular square matrix \mathbf{T} such that $\check{\mathbf{H}}_{\text{arb}} = \mathbf{T} \check{\mathbf{H}}_b$. According to Fact 1 and (3.53),

$$\mathbf{C}_{\text{arb}} = \mathbf{T} \mathbf{C}_b \mathbf{T}^T = \check{\mathbf{H}}_{\text{arb}} \tilde{\mathbf{C}} \check{\mathbf{H}}_{\text{arb}}^T \quad (3.54)$$

$\hat{\mathbf{x}}_{\text{arb}}$ can be verified similarly.

3.6.3 Appendix C

Proof of Theorem 6:

Without loss of generality, assume that $\tilde{\mathbf{H}}_{\text{b}}$ has the same simple form as in (3.45). Now recalculate $\tilde{\mathbf{C}}$ from \mathbf{C}_{yb} given by (3.51). Since $N_1 = \text{rank}(\tilde{\mathbf{H}}) < N$, there exists a I_i such that $I_i \in \text{row}(\tilde{\mathbf{H}}_{\text{b}})$ but $I_i \notin \text{row}(\tilde{\mathbf{H}})$. Let

$$I_i = [\boldsymbol{\alpha}_i \ \boldsymbol{\beta}_i] \tilde{\mathbf{H}}_{\text{b}} = \boldsymbol{\alpha}_i \tilde{\mathbf{H}}_{\text{b}} + \boldsymbol{\beta}_i \mathbf{H}_{\text{left}}$$

Then $y_i = [\boldsymbol{\alpha}_i \ \boldsymbol{\beta}_i] \mathbf{y}_{\text{b}}$, $\mathbf{y}_{\text{left}} = \mathbf{x}_{\text{left}} = [\mathbf{0} \ \mathbf{I}_{\text{left}}] \mathbf{y}_{\text{b}}$.

According to (3.30), $\tilde{\mathbf{C}}$'s cross entries between y_i and \mathbf{y}_{left} form a vector as:

$$C_{i,\text{left}} = [\boldsymbol{\alpha}_i \ \boldsymbol{\beta}_i] \tilde{\mathbf{C}}_{\text{b}} [\mathbf{0} \ \mathbf{I}_{\text{left}}]^T = \boldsymbol{\beta}_i \mathbf{W}_{\text{left}}^{-1}$$

Clearly, $C_{i,\text{left}} \neq \mathbf{0} \Leftrightarrow \boldsymbol{\beta}_i \neq \mathbf{0} \Leftrightarrow I_i \notin \text{row}(\tilde{\mathbf{H}})$.

Chapter 4: Linear Minimum Mean-Square Error Estimator and Unified Recursive GLS

4.1 Overview of LMMSE Estimation and LS

4.1.1 LMMSE Estimation

Following [Li07] closely, the widely-used linear minimum mean-square error (LMMSE) estimator can achieve the smallest means-square error among all linear/affine estimators [Johnson04]. That is, the LMMSE estimator of a random estimand \mathbf{x} is

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \arg \min_{\hat{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mathbf{b}} \text{MSE}(\hat{\mathbf{x}}) \quad (4.1a)$$

which is also equivalent to

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \arg \min_{\hat{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mathbf{b}} \text{mse}(\hat{\mathbf{x}}) = \arg \min_{\hat{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mathbf{b}} \|\mathbf{x} - \hat{\mathbf{x}}\| \quad (4.1b)$$

where \mathbf{z} is observation data, and \mathbf{A} and \mathbf{b} are parameters to be determined. The orthogonality principle for a random estimand can be applied to solve problem (4.1), where the solution $\hat{\mathbf{x}}$ should be an unbiased estimate first [Li07]. As a result,

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \bar{\mathbf{x}} + \mathbf{A}(\mathbf{z} - \bar{\mathbf{z}}) \quad (4.2a)$$

and \mathbf{A} is determined by the following normal equation:

$$\mathbf{C}_{\mathbf{xz}} - \mathbf{A}\mathbf{C}_{\mathbf{z}} = \mathbf{0} \quad (4.2b)$$

where $\bar{\mathbf{x}}$ and $\mathbf{C}_{\mathbf{x}}$ are mean and covariance of \mathbf{x} , $\bar{\mathbf{z}}$ and $\mathbf{C}_{\mathbf{z}}$ are mean and covariance of the observation \mathbf{z} , and $\mathbf{C}_{\mathbf{xz}}$ is the crosscovariance between \mathbf{x} and \mathbf{z} . Consequently, in general,

$$\mathbf{A} = \mathbf{C}_{\mathbf{xz}} \mathbf{C}_{\mathbf{z}}^+ \quad (4.3)$$

In most cases, $\mathbf{C}_{\mathbf{z}}$ is PD. Then the LMMSE estimator becomes

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \bar{\mathbf{x}} + \mathbf{C}_{\mathbf{xz}} \mathbf{C}_{\mathbf{z}}^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \quad (4.4)$$

with which the mean-square error (MSE) matrix is

$$\text{MSE}(\hat{\mathbf{x}}) = \mathbf{C}_x - \mathbf{C}_{xz} \mathbf{C}_z^{-1} \mathbf{C}_{xz}^T \quad (4.5)$$

Actually, if all the involved mean-and-covariance information is given, then (4.2) is the formula of the LMMSE estimator **with complete prior** (the mean and covariance of \mathbf{x}). However, it is also possible that no prior information but only data is available in some practical applications. This LMMSE estimator **with no prior** will be discussed as follows, based on the linear data model.

4.1.2 LMMSE Estimation Based on Linear Data Model

Furthermore, consider that the data model is linear:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v} \quad (4.6)$$

where the mean and covariance of observation error \mathbf{v} are $\bar{\mathbf{v}}$ and \mathbf{C}_v and the crosscovariance between \mathbf{x} and \mathbf{v} is \mathbf{C}_{xv} . Based on this linear model, the mean, covariance and crosscovariance related to data can be derived as

$$\begin{cases} \bar{\mathbf{z}} = \mathbf{H}\bar{\mathbf{x}} + \bar{\mathbf{v}} \\ \mathbf{C}_z = \mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_v + \mathbf{H}\mathbf{C}_{xv} + (\mathbf{H}\mathbf{C}_{xv})^T \\ \mathbf{C}_{xz} = \mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv} \end{cases} \quad (4.7)$$

Correspondingly, the LMMSE estimator can be obtained by substituting $\bar{\mathbf{z}}$, \mathbf{C}_z and \mathbf{C}_{xz} in (4.2a) and (4.2b) with those in (4.7), which is also known as the *best linear unbiased estimation* (BLUE) for an random estimand:

$$\begin{cases} \hat{\mathbf{x}}_{\text{LMMSE}} = \bar{\mathbf{x}} + (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})[\mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_v + \mathbf{H}\mathbf{C}_{xv} + (\mathbf{H}\mathbf{C}_{xv})^T]^{-1}(\mathbf{z} - \mathbf{H}\bar{\mathbf{x}} - \bar{\mathbf{v}}) \\ \text{MSE}(\hat{\mathbf{x}}) = \mathbf{C}_x - (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})[\mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_v + \mathbf{H}\mathbf{C}_{xv} + (\mathbf{H}\mathbf{C}_{xv})^T]^{-1} \\ \quad \cdot (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})^T \end{cases} \quad (4.8)$$

In the that \mathbf{C}_z is PD:

$$\begin{cases} \hat{\mathbf{x}}_{\text{LMMSE}} = \bar{\mathbf{x}} + (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})[\mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_v + \mathbf{H}\mathbf{C}_{xv} + (\mathbf{H}\mathbf{C}_{xv})^T]^{-1}(\mathbf{z} - \mathbf{H}\bar{\mathbf{x}} - \bar{\mathbf{v}}) \\ \text{MSE}(\hat{\mathbf{x}}) = \mathbf{C}_x - (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})[\mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_v + \mathbf{H}\mathbf{C}_{xv} + (\mathbf{H}\mathbf{C}_{xv})^T]^{-1} \\ \quad \cdot (\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_{xv})^T \end{cases} \quad (4.9)$$

Here, (4.8) is exactly the formula of the BLUE estimator with complete prior of \mathbf{x} , where the

mean and covariance of \mathbf{x} , as well as the crosscovariance \mathbf{C}_{xv} , are assumed to be given. When $\bar{\mathbf{x}}$, \mathbf{C}_x , and \mathbf{C}_{xv} are all unknown, the existence of the LMMSE estimator can be checked by the following fact [Li03]:

Fact 4.1: If the prior of an estimand \mathbf{x} is unknown, an LMMSE estimator based on linear data model $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$ with known $E(\mathbf{v}) = \bar{\mathbf{v}}$ exists if and only if \mathbf{H} has full column rank; if exists, the estimator is given by:

$$\hat{\mathbf{x}} = \mathbf{K}(\mathbf{z} - \bar{\mathbf{v}}) \quad (4.10)$$

where \mathbf{K} is the solution of the following constrained minimization problem:

$$\begin{cases} \mathbf{K} = \arg \min_{\mathbf{K}} \text{MSE}(\hat{\mathbf{x}}) = \tilde{\mathbf{K}} \mathbf{C}_v \tilde{\mathbf{K}}^T \\ \text{s. t. } \mathbf{K} \mathbf{H} = \mathbf{I} \end{cases} \quad (4.11)$$

If the observation-error covariance \mathbf{C}_v is PD, then

$$\mathbf{K} = (\mathbf{H}^T \mathbf{C}_v^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_v^{-1} \quad (4.12)$$

Correspondingly,

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{C}_v^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_v^{-1} (\mathbf{z} - \bar{\mathbf{v}}) \quad (4.13)$$

$$\text{and } \text{MSE}(\hat{\mathbf{x}}_{\text{LMMSE}}) = (\mathbf{H}^T \mathbf{C}_v^{-1} \mathbf{H})^{-1} \quad (4.14)$$

On the other hand, if \mathbf{C}_v is PSD, then

$$\mathbf{K} = \mathbf{H}^+ \{ \mathbf{I} - \mathbf{C}_v [(\mathbf{I} - \mathbf{H} \mathbf{H}^+) \mathbf{C}_v (\mathbf{I} - \mathbf{H} \mathbf{H}^+)]^+ \} \quad (4.15)$$

In fact, as discussed in [Li07], a linear-data-model based LMMSE estimator with complete or partial prior can be converted into another one with no prior by treating the prior mean as data using the following formula:

$$\tilde{\mathbf{z}} = \begin{bmatrix} \bar{\mathbf{x}}' \\ \mathbf{z} \end{bmatrix}, \quad \tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H} \end{bmatrix}, \quad \mathbf{E}(\tilde{\mathbf{v}}) = \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{v}} \end{bmatrix}, \quad \tilde{\mathbf{C}}_v = \begin{bmatrix} \mathbf{C}_{x'} & -\mathbf{C}_{x'v} \\ -\mathbf{C}_{x'v} & \mathbf{C}_v \end{bmatrix} \quad (4.16)$$

where \mathbf{x}' contains complete/partial components of \mathbf{x} . $\bar{\mathbf{x}}'$, $\mathbf{C}_{x'}$ and $\mathbf{C}_{x'v}$ are prior mean, covariance and crosscovariance (between \mathbf{x}' and \mathbf{v}) of \mathbf{x}' , respectively. \mathbf{I} and $\mathbf{0}$ are of the appropriate

dimensions. Therefore, it can be concluded that, under a linear data model, the LMMSE estimator with no prior is a unification of LMMSE estimators with complete/partial/no prior knowledge.

In addition, it is clear that, under the condition that the observation-error covariance \mathbf{C}_z is PD, the LMMSE estimator with no prior in (4.12)-(4.14), is mathematically identical to a GLS solution which is based on the same linear data model as in (4.6) except that the estimand \mathbf{x} is nonrandom. Furthermore, it can also be verified that, if \mathbf{C}_z is PSD, the LMMSE estimator with no prior, which is described by (4.15), is mathematically equivalent to the GLS with implicit LE constraint (ILE constrained GLS) defined in Chapter 2. Actually, [Li07] has presented a quasi-recursive form the LMMSE estimator. It can be shown that, based on the linear data model in (4.6), the quasi-recursive form of the LMMSE estimator mathematically coincides with the LRC-decorrelation in the recursive ILE-constrained GLS. As a result, the solution to a linear-data-model LMMSE estimation without prior is also mathematically identical to a unified GLS solution. Therefore, the linear-data-model-based LMMSE estimation without prior with a PD/PSD measurement noise covariance, or more generally, the linear-data-model-based LMMSE estimator with a unified PD/PSD joint covariance of the estimand and the measurement noise can be obtained by solving the corresponding LS problems studied in Chapter 2. The calculation can thus take advantage of the recursive procedure of the unified GLS.

4.2 Verification of Optimal Kalman Filter with Various Correlations

It is well known that LMMSE estimation is the theoretical basis for linear filtering, such as Kalman filter, LMMSE filter (for nonlinear problems) and (steady-state) Wiener filter. For instance, in the Kalman filter for linear systems with linear measurements, the prediction is from an LMMSE estimator using the data up to the most recent time while the update can be implemented from another LMMSE estimator using all the data up to the current time. Owing to the mathematical equivalence between the linear-data-model based LMMSE and the GLS,

the LMMSE estimators, as well as the Kalman filter, can make full use of the improvement of the LS computation. The detailed discussion is presented as follows.

According to [Li07], consider the following stochastic linear discrete-time system,

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{\Gamma}_{k-1}\mathbf{w}_{k-1} \quad (4.17)$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k + \mathbf{E}_k\mathbf{u}_k \quad (4.18)$$

where k is time index. \mathbf{x} and \mathbf{u} are system state and (possible) input, respectively. \mathbf{w} and \mathbf{v} are process noise and measurement noise, respectively. \mathbf{F} , \mathbf{G} and \mathbf{E} are transition matrix, input gain matrix, and input-output matrix, respectively. \mathbf{H} is output (measurement coefficient) matrix and $\mathbf{\Gamma}$ is process-noise coefficient matrix. Here, the mean-and- covariance assumptions are crucial:

$$\begin{cases} E[\mathbf{w}_k] = \bar{\mathbf{w}}_k \\ \text{cov}(\mathbf{w}_i, \mathbf{w}_j) = \mathbf{Q}_i \delta_{i-j} \end{cases} \quad (4.19)$$

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (4.20)$$

$$\begin{cases} \text{cov}(\mathbf{x}_0, \mathbf{w}_k) = \mathbf{0} \\ \text{cov}(\mathbf{x}_0, \mathbf{v}_k) = \mathbf{0} \end{cases} \quad (4.21)$$

$$\begin{cases} E[\mathbf{v}_k] = \bar{\mathbf{v}}_k \\ \text{cov}(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{R}_i \delta_{i-j} \end{cases} \quad (4.22)$$

and

$$\text{cov}(\mathbf{v}_i, \mathbf{w}_j) = \mathbf{0} \quad (4.23)$$

where \mathbf{Q}_i and \mathbf{R}_i are PD matrices. (4.19) means that process noises at different instants are uncorrelated. Similarly, (4.22) requires that the measurement noises at different instants be uncorrelated. (4.22) and (4.23) indicate that the crosscovariance, such as the one between the initial state \mathbf{x}_0 and process noise \mathbf{w}_k , between \mathbf{x}_0 and measurement noise \mathbf{v}_k , and between \mathbf{v}_i and \mathbf{w}_j , are all zero. When the above assumptions are satisfied, the standard Kalman filter,

of which the predictor-and-corrector form is described by the following formulae, can estimate the state \mathbf{x}_k optimally in the sense of LMMSE:

Prediction:

$$\begin{cases} \hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{G}_{k-1} \bar{\mathbf{u}}_{k-1} + \mathbf{\Gamma}_{k-1} \bar{\mathbf{w}}_{k-1} \\ \hat{\mathbf{z}}_{k|k-1} = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} + \bar{\mathbf{v}}_k + \mathbf{E}_k \bar{\mathbf{u}}_k \\ \mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{\Gamma}_{k-1} \mathbf{Q}_{k-1} \mathbf{\Gamma}_{k-1}^T \end{cases} \quad (4.24)$$

Update:

$$\begin{cases} \mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \\ \mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1} \end{cases} \quad (4.25)$$

$$\begin{cases} \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \\ \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \end{cases} \quad (4.26)$$

In particular, under the given assumptions, the prediction error and the measurement noise are also uncorrelated:

$$\text{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}, \mathbf{v}_k^T) = \mathbf{0} \quad (4.27)$$

However, it is likely that (4.27) does not hold due to some accidental reasons such as $\text{cov}(\mathbf{w}_k, \mathbf{v}_k^T) \neq \mathbf{0}$. Denote that

$$\text{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}, \mathbf{v}_k^T) = \mathbf{M}_k \neq \mathbf{0} \quad (4.28)$$

Then [Li07] shows that Kalman filter still work – both the prediction and the update are optimal in the LMMSE sense.

Theorem 4.1 (Optimal Kalman filter considering prediction-measurement correlation):

In the linear system described by (4.17)-(4.18), given that the requirements in (4.19)-(4.23) are all satisfied up to $k = \tilde{k} - 1$ and (4.19) and (4.20) are still valid at $k = \tilde{k}$, even if at least one of the requirements in (4.21) and (4.22) is not satisfied, which leads to $\text{cov}(\mathbf{x}_{\tilde{k}} - \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}-1}, \mathbf{v}_{\tilde{k}}^T) = \mathbf{M}_{\tilde{k}} \neq \mathbf{0}$, then the state $\mathbf{x}_{\tilde{k}}$ can be still estimated optimally in the LMMSE sense:

Prediction:

$$\begin{cases} \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}-1} = \mathbf{F}_{\tilde{k}-1} \hat{\mathbf{x}}_{\tilde{k}-1|\tilde{k}-1} + \mathbf{G}_{\tilde{k}-1} \bar{\mathbf{u}}_{\tilde{k}-1} + \mathbf{\Gamma}_{\tilde{k}-1} \bar{\mathbf{w}}_{\tilde{k}-1} \\ \hat{\mathbf{z}}_{\tilde{k}|\tilde{k}-1} = \mathbf{H}_{\tilde{k}} \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}-1} + \bar{\mathbf{v}}_{\tilde{k}} + \mathbf{E}_{\tilde{k}} \bar{\mathbf{u}}_{\tilde{k}} \\ \mathbf{P}_{\tilde{k}|\tilde{k}-1} = \mathbf{F}_{\tilde{k}-1} \mathbf{P}_{\tilde{k}-1|\tilde{k}-1} \mathbf{F}_{\tilde{k}-1}^T + \mathbf{\Gamma}_{\tilde{k}-1} \mathbf{Q}_{\tilde{k}-1} \mathbf{\Gamma}_{\tilde{k}-1}^T \end{cases} \quad (4.29)$$

Update:

$$\begin{cases} \mathbf{S}_{\tilde{k}} = \mathbf{H}_{\tilde{k}} \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}^T + \mathbf{R}_{\tilde{k}} + \mathbf{H}_{\tilde{k}} \mathbf{M}_{\tilde{k}} + (\mathbf{H}_{\tilde{k}} \mathbf{M}_{\tilde{k}})^T \\ \mathbf{K}_{\tilde{k}} = (\mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}^T + \mathbf{M}_{\tilde{k}}) \mathbf{S}_{\tilde{k}}^{-1} \end{cases} \quad (4.30)$$

$$\begin{cases} \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}} = \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}-1} + \mathbf{K}_{\tilde{k}} (\mathbf{z}_{\tilde{k}} - \hat{\mathbf{z}}_{\tilde{k}|\tilde{k}-1}) \\ \mathbf{P}_{\tilde{k}|\tilde{k}} = \mathbf{P}_{\tilde{k}|\tilde{k}-1} - \mathbf{K}_{\tilde{k}} \mathbf{S}_{\tilde{k}} \mathbf{K}_{\tilde{k}}^T \end{cases} \quad (4.31)$$

In fact, Theorem 4.1 has been shown in [Li07] by treating the update as an LMMSE estimator based on the prediction and the current data. We want to verify this theorem using the recursive GLS formulae since the LMMSE estimation and GLS (considering ILE constraint) are mathematically equivalent. Subsequently, the sequential computation of the optimal KF considering correlation can take advantage of the recursive GLS.

Verification (of the optimal KF considering prediction-measurement correlation):

Since the requirements in (4.19)-(4.23) are all satisfied up to $k = \tilde{k} - 1$ and those in (4.19) and (4.20) are still valid at $k = \tilde{k}$, the prediction should be the same as the standard KF, which is given by (4.29).

The update also comes from an LMMSE estimator based on the current data and the predicted estimate which is the prior. According to the unification of LMMSE estimators in (4.16):

$$\tilde{\mathbf{z}}_{\tilde{k}} = \begin{bmatrix} \bar{\mathbf{x}}'_{\tilde{k}|\tilde{k}-1} \\ \mathbf{z}_{\tilde{k}} \end{bmatrix}, \quad \tilde{\mathbf{H}}_{\tilde{k}} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_{\tilde{k}} \end{bmatrix}, \quad \mathbf{E}(\tilde{\mathbf{v}}_{\tilde{k}}) = \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{v}}_{\tilde{k}} \end{bmatrix}, \quad \tilde{\mathbf{R}}_{\tilde{k}} = \begin{bmatrix} \mathbf{P}_{\tilde{k}|\tilde{k}-1} & -\mathbf{M} \\ -\mathbf{M}^T & \mathbf{R}_{\tilde{k}} \end{bmatrix} \quad (4.32)$$

Applying Theorem 2.1, we have

$$\begin{cases} \mathbf{S}_{\tilde{k}} = \mathbf{H}_{\tilde{k}}' \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}'^T + \mathbf{R}_{\tilde{k}}' \\ \mathbf{K}_{\tilde{k}} = \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}'^T \mathbf{S}_{\tilde{k}}^{-1} \end{cases} \quad (4.33)$$

with

$$\begin{cases} \mathbf{H}'_{\tilde{k}} = \mathbf{H}_{\tilde{k}} + \mathbf{M}^T \mathbf{P}_{\tilde{k}|\tilde{k}-1}^{-1} \mathbf{I} \\ \mathbf{R}'_{\tilde{k}} = \mathbf{R}_{\tilde{k}} + \mathbf{M}^T \mathbf{P}_{\tilde{k}|\tilde{k}-1}^{-1} \mathbf{M} \\ \mathbf{z}_{\tilde{k}} = \mathbf{z}_{\tilde{k}} + \mathbf{M}^T \mathbf{P}_{\tilde{k}|\tilde{k}-1}^{-1} \hat{\mathbf{x}}_{\tilde{k}|\tilde{k}-1} \end{cases} \quad (4.34)$$

which comes from (2.82). Accordingly,

$$\begin{cases} \mathbf{S}_{\tilde{k}} = \mathbf{H}'_{\tilde{k}} \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}'_{\tilde{k}}^T + \mathbf{R}'_{\tilde{k}} = \mathbf{H}_{\tilde{k}} \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}^T + \mathbf{M}^T \mathbf{H}_{\tilde{k}}^T + \mathbf{H}_{\tilde{k}} \mathbf{M} + \mathbf{R}_{\tilde{k}} \\ \mathbf{K}_{\tilde{k}} = \mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}'^T \mathbf{S}_{\tilde{k}}^{-1} = (\mathbf{P}_{\tilde{k}|\tilde{k}-1} \mathbf{H}_{\tilde{k}}^T + \mathbf{M}_{\tilde{k}}) \mathbf{S}_{\tilde{k}}^{-1} \end{cases} \quad (4.35)$$

which is identical to (4.30). The subsequent (4.31) can be shown similarly.

Correspondingly, we can apply the CRLS to implement a sequential data processing scheme for the optimal KF considering correlation. In addition, we can also use the recursive GLS to deal with various complicated situations caused by data correlation, PSD covariance, and so on.

Actually, the formulae of the optimal KF considering correlation can be adopted in nonlinear filtering applications where the predicted state and the current measurement error are widely correlated due to linearization. For instance, we will apply the optimal KF considering correlation to develop a series of adaptive filtering techniques and solve practical problems such as power system state estimation and parameter estimation.

Nomenclature in Chapters 5-6

Subscripts in Chapters 5-6 are listed below for quick reference:

c	related to the center of \mathbf{x} or \mathbf{p}
k	time index
$k k-1$	time index for prediction from instant $k-1$ to k
$k k$	time index for update at instant k
mn	to-be-calibrated line connecting nodes m and n
$mn0$	m side (to the ground) of the line connecting nodes m and n
p	related to the unknown parameter vector \mathbf{p}
r and i	real and imaginary components
st	arbitrary line connecting nodes s and t
$st0$	s side (to the ground) of the line connecting nodes s and t
x	related to voltage state \mathbf{x}

Multiple subscripts are separated by a comma.

In addition, some variables and symbols, which have been adopted in previous Chapters 2-4, may be used again but have specific or different interpretations in Chapter 5-6. These specific interpretations are used by default unless the variables adopted in Chapters 2-4 are really involved in Chapters 5-6.

Chapter 5: Joint Estimation of State and Parameter with Synchrophasors—Part I: State Tracking

5.1 Introduction

5.1.1 Background and Motivation

Most power system state estimation (SE) programs are formulated as static weighted-least-squares (WLS) problems with one-scan data [Monticelli00]. Dynamic state estimation (DSE) is not popularly applied due to practical limitations such as the complexity of the measurement system and the inaccuracy of dynamic and measurement models. In fact, parameter estimation (PE) is responsible for calibrating the suspicious measurement model parameters [Abur04], within which the bus voltages of interest and the unknown parameters are usually stacked as an augmented state [Zarco00]. Correspondingly, dynamic-estimation methods are preferred since they exploit data from multiple scans and take advantage of dynamic models [Leite87]. Unfortunately, similar obstacles as in the DSE for bus voltages are encountered and the estimation accuracy is not guaranteed. These dilemmas can be avoided in power systems metered by synchrophasors.

The invention of synchrophasor, also known as synchronized phasor measurement unit (PMU), has led a revolution in SE since it yields linear measurement functions as well as accurate data within three to five cycles [Phadke93] [Phadke02]. In spite of the involved instrumental channel errors [Sakis07] and the high cost, PMU has been tentatively used in centralized or distributed estimators [Phadke86] [Zhao05] [Jiang07] and in bad-data detection [Chen06].

In Chapters 5 & 6, we aim at performing accurate parameter (and state) estimation in complex situations using synchrophasor data. An approach of joint state-and-parameter estimation, which is different from the state augmentation, is adopted, where the original nonlinear PE problem is reformulated as two loosely-coupled linear subproblems: state

tracking and parameter tracking, respectively. This chapter is on the state tracking with possible abrupt voltage changes and correlated prediction-measurement errors, which can be applied to determine the voltages in a PE problem or to estimate the system state in a conventional DSE problem.

5.1.2 Literature Review

DSE appeared soon after the static SE was introduced in the 1970s. Reference [Debs70] employs the extended Kalman filter (EKF) and uses the random-walk based prediction model

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_k \quad (5.1)$$

to predict the state \mathbf{x} from time $k-1$ to k , where the state offset \mathbf{w} may be nonwhite. (5.1) is later adopted in parameter estimation with \mathbf{w} simplified as zero-mean white noise [Zarco00]. It has several limitations (see Sec. 5.4.1). Furthermore, an improved model incorporating the dynamic nature of the system via forecasting is suggested as

$$\mathbf{x}_k = \mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{G}_{k-1} + \mathbf{w}_k \quad (5.2)$$

where \mathbf{w} is now zero-mean white noise. Coefficients \mathbf{F} and \mathbf{G} can be determined by the linear exponential smoothing method [Leite83]. With this model, EKF is also applicable [Leite87]. In addition, [Shih02] improves the robustness of the filtering algorithm using an exponential weight function. In [Mandal95] [Sinha99], \mathbf{G} is interpreted as known control actions, and [Sinha99] applies artificial neural network (ANN) to estimate \mathbf{G} . However, (5.2) can not handle a possible abrupt change between instants $k-1$ and k [Lin03], since \mathbf{F} and \mathbf{G} only depend on the past estimates. References [Lin03] and [Huang04] introduce techniques of fuzzy control to relieve this difficulty. Nevertheless, (5.1) and (5.2) are not very applicable (see Sec. 5.4.1). We prefer to find an efficient method from a different point of view.

5.1.3 Our Work

The purpose is to track the bus voltages using multiple-scan synchrophasor data. Contributions of our work are summarized as follows:

- 1) The system dynamic behavior is reanalyzed, which results in an improved prediction model.

- 2) A prediction model is proposed, which accounts for the effects of minor shifts and abrupt changes comprehensively.
- 3) The measurement model is improved via introducing a concept of pseudo measurement error for the uncertainty in line parameters.
- 4) A method for detecting the abrupt change is proposed. It makes use of an inherent property of the filter: estimation-error covariance is independent of the state.
- 5) An adaptive filtering algorithm based on optimal tracking with correlated prediction-measurement errors, including the module for the abrupt-change detection, is developed.

The solutions in the above setting yield accurate results and provide a reliable support for the parameter tracking presented in chapter 6. The work has been published in [Bian11a].

5.2 Formulation and State-Space Models

5.2.1 State Tracking with Uncertain Parameters

Multiscan synchrophasor data are used to estimate voltages of interest. Some steady-state parameters of transmission lines may also need to be calibrated. A generic measurement model (at time k) can be written as

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{p}_k) + \mathbf{v}_k \quad (5.3)$$

where

\mathbf{z} vector including all one-scan measurements

\mathbf{x} state containing node voltages of interest, $\mathbf{x} = [x_{r,1} \ x_{r,2} \ x_{i,2} \ \cdots \ x_{r,N} \ x_{i,N}]^T$

\mathbf{p} vector containing uncertain line admittance, e.g., uncertain admittance of a single line

(by the π -type circuit) is $[g_{mn} \ g_{mn0} \ b_{mn} \ b_{mn0}]^T$

g_{mn} serial conductance in the π -type circuit for a line connecting nodes m and n

g_{mn0} half shunt conductance in the π -type circuit for a line connecting nodes m and n

b_{mn} serial susceptance of the π -type circuit for a line connecting nodes m and n

b_{mn0} half shunt susceptance of the π -type circuit for a line connecting nodes m and n

\mathbf{v} vector of total measurement noise

Assuming that the state and the measured data are in (or have been preprocessed into) the rectangular form (see Sec. 5.4.1), the measurement function $\mathbf{h}(\mathbf{x}, \mathbf{p})$ contains four types of scalar functions (represented by h) as follows.

Bus-voltage complex synchrophasors:

$$\begin{cases} h_{r,s} = x_{r,s} \\ h_{i,s} = x_{i,s} \end{cases} \quad (5.4)$$

Branch-current complex synchrophasors:

$$\begin{cases} h_{r,st} = (x_{r,s} - x_{r,t})g_{st} - (x_{i,s} - x_{i,t})b_{st} + x_{r,s}g_{st0} - x_{i,s}b_{st0} \\ h_{i,st} = (x_{i,s} - x_{i,t})g_{st} + (x_{r,s} - x_{r,t})b_{st} + x_{i,s}g_{st0} + x_{r,s}b_{st0} \end{cases} \quad (5.5)$$

$\mathbf{h}(\mathbf{x}, \mathbf{p})$ is nonlinear if \mathbf{x} is augmented with \mathbf{p} . Our approach treats this nonlinear problem as two loosely-coupled linear subproblems, namely, state tracking and parameter tracking. With this approach, \mathbf{p} is roughly estimated in advance for the use of the state tracking. In other words, it is uncertain rather than unknown. To our knowledge, state tracking with uncertain line parameters is first discussed in this paper. In addition, the proposed approach can be simplified and applied to the conventional cases where \mathbf{p} is exactly known.

5.2.2 Prediction Model for State Tracking

We introduce a new prediction model

$$\mathbf{x}_k = \mathbf{x}_{c,k-1} + \mathbf{u}_k + \mathbf{w}_{x,k} \quad (5.6a)$$

$$\mathbf{x}_{c,k} = \mathbf{x}_{c,k-1} + \mathbf{u}_k \quad (5.6b)$$

Substituting (5.6b) into (5.6a), the model is equivalent to

$$\begin{cases} \mathbf{x}_k = \mathbf{x}_{c,k} + \mathbf{w}_{x,k} \\ \mathbf{x}_{c,k} = \mathbf{x}_{c,k-1} + \mathbf{u}_k \end{cases} \quad (5.7)$$

where

\mathbf{x}_c mean (center) of \mathbf{x}

\mathbf{w}_x vector of zero-mean white process noise

\mathbf{u} abrupt change to be detected

Here, (5.6a)-(5.6b) describe the standard dynamics of the system while (5.7) clearly reveals two facts: (a) \mathbf{x}_k varies around $\mathbf{x}_{c,k}$ over time, and (b) $\mathbf{x}_{c,k}$ may change abruptly. Detailed explanation of this model is given in Sec. 5.4.1.

5.2.3 Measurement Model of State Tracking

The standard measurement model with known admittance is

$$\mathbf{z}_{x,k} = \mathbf{H}_x \mathbf{x}_k + \mathbf{v}_{x,k} \quad (5.8)$$

where

\mathbf{H}_x coefficient matrix derived from (5.4)-(5.5) by treating all conductance and susceptance as constants

\mathbf{z}_x vector of measurements in state tracking

\mathbf{v}_x vector of measurement noise in state tracking

Since \mathbf{H}_x may contain uncertain admittance parameters, equation (5.8) can be expressed as

$$\mathbf{z}_{x,k} = \mathbf{H}_{x,k}^\circ \mathbf{x}_k + \mathbf{v}_{x,k}^\circ \quad (5.9)$$

where

$$\mathbf{v}_{x,k}^\circ = \Delta \mathbf{H}_{x,k} (\hat{\mathbf{x}}_{k|k-1} + \Delta \mathbf{x}_{k|k-1}) + \mathbf{v}_{x,k} \quad (5.10)$$

$$\Delta \mathbf{H}_{x,k} = \mathbf{H}_x - \mathbf{H}_{x,k}^\circ, \quad \Delta \mathbf{x}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$$

$\hat{\mathbf{x}}_{k|k-1}$ predicted \mathbf{x}

\mathbf{H}_x° (time-dependent) coefficient matrix obtained from \mathbf{H}_x by replacing \mathbf{p} with an estimate $\hat{\mathbf{p}}$.

Note that the original noises \mathbf{w}_x and \mathbf{v}_x are zero-mean, white Gaussian, and mutually independent.

5.3 Adaptive Tracking of State

5.3.1 Basic Filter Considering Correlation

The mean of the (pseudo) measurement noise $\mathbf{v}_{x,k}^\circ$ in (5.10) is not necessarily zero. $\mathbf{v}_{x,k}^\circ$ and the prediction error $\Delta\mathbf{x}_{k|k-1}$ may be correlated. Thus the following quantities are considered:

$$\bar{\mathbf{v}}_{x,k}^\circ = E[\mathbf{v}_{x,k}^\circ], \quad \mathbf{R}_{x,k}^\circ = \text{cov}[\mathbf{v}_{x,k}^\circ] \quad (5.11)$$

$$\tilde{\mathbf{A}}_k = \begin{bmatrix} \mathbf{A}_k \\ \mathbf{A}_{c,k} \end{bmatrix} = \begin{bmatrix} \text{cov}(\Delta\mathbf{x}_{k|k-1}, \mathbf{v}_{x,k}^\circ) \\ \text{cov}(\Delta\mathbf{x}_{c,k|k-1}, \mathbf{v}_{x,k}^\circ) \end{bmatrix} \quad (5.12)$$

The optimal Kalman filter considering the correlation between errors in the prediction and the measurements, which is recently developed in [Li07], is used to develop the state-tracking procedure (see Appendix A). The following is the algorithm in the predictor-corrector form, which is convenient for implementation and coding:

1) Initialization: $k = 0$

$$\hat{\mathbf{x}}_{0|0} = \hat{\mathbf{x}}_{c,0|0}; \quad \mathbf{P}_{0|0} = \mathbf{P}_{c,0|0} \quad (5.13)$$

2) Recursion: $k := k + 1$

a) Prediction:

Assume $\hat{\mathbf{u}}_k = \mathbf{0}$

$$\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_{c,k-1|k-1}; \quad \hat{\mathbf{x}}_{c,k|k-1} = \hat{\mathbf{x}}_{c,k-1|k-1} \quad (5.14)$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}_{c,k-1|k-1} + \mathbf{Q}_{x,k}; \quad \mathbf{P}_{c,k|k-1} = \mathbf{P}_{c,k-1|k-1} \quad (5.15)$$

$$\hat{\mathbf{z}}_{x,k|k-1} = \mathbf{H}_{x,k}^\circ \hat{\mathbf{x}}_{k|k-1} + \bar{\mathbf{v}}_{x,k}^\circ \quad (5.16)$$

b) Update:

$$\mathbf{S}_k = \mathbf{H}_{x,k}^\circ \mathbf{P}_{k|k-1} (\mathbf{H}_{x,k}^\circ)^T + \mathbf{R}_{x,k}^\circ + \mathbf{H}_{x,k}^\circ \mathbf{A}_k + (\mathbf{H}_{x,k}^\circ \mathbf{A}_k)^T \quad (5.17)$$

$$\begin{cases} \mathbf{K}_k = [\mathbf{P}_{k|k-1}(\mathbf{H}_{x,k}^\circ)^T + \mathbf{A}_k] \mathbf{S}_k^{-1} \\ \mathbf{K}_{c,k} = [\mathbf{P}_{c,k|k-1}(\mathbf{H}_{x,k}^\circ)^T + \mathbf{A}_{c,k}] \mathbf{S}_k^{-1} \end{cases} \quad (5.18)$$

As a result,

$$\Delta \mathbf{z}_{x,k} = \mathbf{z}_{x,k} - \hat{\mathbf{z}}_{x,k|k-1} \quad (5.19)$$

$$\begin{cases} \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \Delta \mathbf{z}_{x,k} \\ \hat{\mathbf{x}}_{c,k|k} = \hat{\mathbf{x}}_{c,k|k-1} + \mathbf{K}_{c,k} \Delta \mathbf{z}_{x,k} \end{cases} \quad (5.20)$$

$$\begin{cases} \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \\ \mathbf{P}_{c,k|k} = \mathbf{P}_{c,k|k-1} - \mathbf{K}_{c,k} \mathbf{S}_k \mathbf{K}_{c,k}^T \end{cases} \quad (5.21)$$

c) Adaptation:

Reset $\hat{\mathbf{x}}_{c,k|k-1} = \hat{\mathbf{x}}_{c,k-1|k-1} + \hat{\mathbf{u}}_k$ if $\hat{\mathbf{u}}_k$ is detected and estimated; (see Sec. 5.3.2 for details)

where

\mathbf{P} covariance of estimation errors in $\hat{\mathbf{x}}$

\mathbf{P}_c covariance of estimation errors in $\hat{\mathbf{x}}_c$

\mathbf{S} covariance of measurement prediction error

\mathbf{K} filter gain

\mathbf{Q}_x covariance of process noise \mathbf{w}_x , assumed known

Subscript $k|k-1$ is time index for prediction from instant $k-1$ to k and $k|k$ is time index for update at instant k .

In particular, $\bar{\mathbf{v}}_{x,k}^\circ$, $\mathbf{R}_{x,k}^\circ$ and $\bar{\mathbf{A}}_k$ are assumed known in the above procedure. An error-ensemble-evolution method is used to evaluate these quantities. First, this method generates ensembles of random quantities such as $\Delta \mathbf{x}_{0|0}$, $\Delta \mathbf{x}_{c,0|0}$, $\mathbf{w}_{x,k}$ and $\mathbf{v}_{x,k}$ using given mean and covariance information. Consequently, the ensembles of predicted and updated estimation errors evolve as follows:

$$\begin{cases} \Delta \mathbf{x}_{c,klk-1}^{(j)} = \Delta \mathbf{x}_{c,k-1lk-1}^{(j)} \\ \Delta \mathbf{x}_{klk-1}^{(j)} = \Delta \mathbf{x}_{c,klk-1}^{(j)} + \mathbf{w}_{x,k}^{(j)} \end{cases} \quad (5.22)$$

$$\begin{cases} \Delta \mathbf{x}_{c,klk}^{(j)} = \Delta \mathbf{x}_{c,klk-1}^{(j)} - \mathbf{K}_{c,k} \left(\mathbf{H}_{x,k}^{\circ} \Delta \mathbf{x}_{klk-1}^{(j)} + \mathbf{v}_{x,k}^{\circ(j)} - \bar{\mathbf{v}}_{x,k}^{\circ} \right) \\ \Delta \mathbf{x}_{klk}^{(j)} = \Delta \mathbf{x}_{klk-1}^{(j)} - \mathbf{K}_k \left(\mathbf{H}_{x,k}^{\circ} \Delta \mathbf{x}_{klk-1}^{(j)} + \mathbf{v}_{x,k}^{\circ j} - \bar{\mathbf{v}}_{x,k}^{\circ} \right) \end{cases} \quad (5.23)$$

where superscript j is a sample index in ensemble. Second, using (5.10), the ensemble of $\mathbf{v}_{x,k}^{\circ}$ can be obtained from the (derived or given) parameter-error ensemble. Finally, sample mean and sample covariance are used for $\bar{\mathbf{v}}_{x,k}^{\circ}$, $\mathbf{R}_{x,k}^{\circ}$ and $\check{\mathbf{A}}_k$, respectively. Details of this method and the derivation for (5.22)-(5.23) can be found in Chapter 6.

5.3.2 Detection and Estimation of Abrupt Change in State

Two hypotheses are considered for abrupt change detection, and the decided one is used in the above filter. That is,

$$H_0: \mathbf{u}_k = 0 \quad \text{vs.} \quad H_1: \mathbf{u}_k \neq 0$$

As shown in Appendix B, the possible abrupt change and the different innovations under the two hypotheses form an important linear “measurement equation” as

$$\Delta \mathbf{z}_{x,k}^{-} = \mathbf{H}_{x,k}^{\circ} \mathbf{u}_k + \Delta \mathbf{z}_{x,k} \quad (5.24)$$

where $\Delta \mathbf{z}_x^{-}$ is vector of innovations under H_0 , and $\Delta \mathbf{z}_x$ is under H_1 . Consequently, \mathbf{u}_k is estimated as

$$\hat{\mathbf{u}}_k = [(\mathbf{H}_{x,k}^{\circ})^T \mathbf{S}_k^{-1} \mathbf{H}_{x,k}^{\circ}]^{-1} (\mathbf{H}_{x,k}^{\circ})^T \mathbf{S}_k^{-1} \Delta \mathbf{z}_x^{-} \quad (5.25)$$

Note that \mathbf{S}_k holds the same value under both hypotheses. This important fact comes from the property of the Kalman filter: Estimation error covariance is independent of the state.

Under H_0 , $\hat{\mathbf{u}}_k$ is approximately Gaussian:

$$\hat{\mathbf{u}}_k \sim N(0, \mathbf{L}), \text{ with } \mathbf{L} = [(\mathbf{H}_{x,k}^{\circ})^T \mathbf{S}_k^{-1} \mathbf{H}_{x,k}^{\circ}]^{-1} \quad (5.26)$$

Accordingly, a significance test can be performed as

$$T_u = \hat{\mathbf{u}}_k^T \mathbf{L}^{-1} \hat{\mathbf{u}}_k \geq c \quad (5.27)$$

where the test statistic T_u is chi-square distributed under H_0 , the threshold c is determined by

$$P\{T_u \geq c\} = \alpha \quad (5.28)$$

and α is the desired probability of false alarm.

If (5.27) is not satisfied, H_1 is rejected and $\hat{\mathbf{u}}_k$ is declared to be zero. Otherwise, H_0 is rejected and $\hat{\mathbf{u}}_k$ is estimated by (5.25).

In addition, calculated \mathbf{S}_k may be affected by $\hat{\mathbf{u}}_k$ through $\tilde{\mathbf{A}}_k$ and $\mathbf{R}_{x,k}^\circ$. Iteration can be applied if necessary.

5.4 Procedure and Performance Analysis

5.4.1 Explanations on Prediction and Measurement Models

In essence, the prediction model (5.6) (without \mathbf{u}_k) is an improved random-walk model. In the model of (5.1), the state is

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_{x,k} = \cdots = \mathbf{x}_0 + \sum_{j=1}^k \mathbf{w}_{x,j} \quad (5.29)$$

where the covariance of cumulative noise $\sum_{j=1}^k \mathbf{w}_{x,j}$ increases with k . This is inconsistent with the fact that the normal state does not deviate far from the nominal center due to the effect of the control system. The new model assumes that \mathbf{x}_k varies around a deterministic center $\mathbf{x}_{c,k}$ by $\mathbf{w}_{x,k}$. \mathbf{x}_k does not deviate far from \mathbf{x}_{k-1} since the noise $\mathbf{w}_{x,k} - \mathbf{w}_{x,k-1}$ still has constant covariance. Clearly, this model is better over a long time horizon.

The forecasting model in (5.2) introduces the coefficients \mathbf{F}_k and \mathbf{G}_k into the random-walk model. However, in power systems, especially in those monitored by synchrophasors, measurement redundancy is not high. It is not easy to accurately determine excessive additional model parameters since they indirectly reduce the measurement redundancy. As a result, a “smooth” but inaccurate trajectory may be obtained with this model. Furthermore, the

forecasting model has a risk of instability. In fact, \mathbf{F} and \mathbf{G} totally rely on the past data and they may mismatch the current true model. In the filtering procedures, the filter also assigns increasing weights to the prediction. Thus the past data are over weighted. When the filter update does not mitigate a serious mismatch error using new data with incorrectly lower weights, \mathbf{F} and \mathbf{G} are adapted worse. Consequently, the prediction at the next instant deviates further away from the true state, especially when there is an abrupt change. In extreme cases, the filter will diverge. In contrast, the new model in (5.6) accounts for the effects of the control system, the small state shifting and the abrupt state change, which can describe the behavior of the realistic system accurately. It also assigns relatively higher weights to new data. Moreover, it is much more concise and easier to utilize in the filtering procedure.

Practical implementations can also take advantage of the conciseness of the new model. For instance, \mathbf{Q}_x , the covariance of \mathbf{w}_x , can be evaluated easily with (5.6): The estimated state based on static WLS estimation methods can be used as samples; the sample covariance minus the covariance of (static) estimation error is roughly equal to \mathbf{Q}_x . On the contrary, with (5.2) and even with (5.1), \mathbf{Q}_x is not easy to evaluate. Furthermore, \mathbf{u}_k , which is for an abrupt change that exists in reality but was ignored in previous work, also plays a beneficial role. Once it is detected correctly, it does compensate the effect of the change. However, even if it responds to a false alarm, which ought to be avoided in the design, the misoperations may also improve estimation accuracy, especially when the process noise is much larger than measurement errors. In addition, the estimated \mathbf{u}_k can be used to process bad data (see Sec. 5.4.2).

To obtain a linear measurement model in (5.8) and simplify the filtering procedure, both the bus voltage variables and the measured voltage and current phasors are assumed in the rectangular form. The theoretical synchrophasor measurements are inherently in the rectangular form [Phadke08]. The IEEE standard C37.118-2005 suggests that synchrophasors provide data in both rectangular and angular forms [IEEEStd05]. However, some types of

synchrophasors may be manufactured to provide angular-formed data only. Then the coordinate transformation for the data and accompanying amplified measurement errors, which is investigated in [Bi08], can be employed. Eventually, for the practical use, the estimated state can almost be converted into the angular form without loss of accuracy.

5.4.2 Implementation Issues

The proposed approach performs DSE for power systems in generic situations where uncertain line parameters are involved. The state tracking procedure is depicted in Fig. 5.1. The two modules in the dashed box, regarding covariance calculation and parameter tracking, will be discussed in Chapter 6.

In fact, this approach can be simplified to solve conventional DSE problems, where all the line parameters are known and no correlation exists between the prediction and the measurement errors. That is, $\mathbf{H}_{x,k}^\circ = \mathbf{H}_x$, $\mathbf{R}_{x,k}^\circ = \mathbf{R}_x$, $\bar{\mathbf{v}}_{x,k}^\circ$ and $\check{\mathbf{A}}_k = \mathbf{0}$. Correspondingly, in Fig. 5.1, the modules in the dashed box are not needed any more.

Particularly, the Kalman filter, including the recently-introduced considering-correlation one, has another important property which can benefit practical issues related to large system scale, observability analysis, bad-data processing and so on. That is, the updated state $\hat{\mathbf{x}}_{k|k}$ is mathematically equivalent to a WLS estimate which takes both $\mathbf{z}_{x,k}$ and the predicted state $\hat{\mathbf{x}}_{k|k-1}$ as data. Correspondingly, $\mathbf{v}_{x,k}^\circ$ and $\Delta \mathbf{x}_{k|k-1}$ are taken as measurement errors. In this sense, when the proposed approach is applied to large-scale systems, existing WLS-based distributed computation techniques can be applied. For instance, the multiarea state-estimation technique in [Conejo07] and [Ebrahimian00] can be employed to complete the update in the filter. Moreover, the system now is linear. The distributed estimation can be handled much more easily and quickly since iteration may be avoided. Furthermore, the sparsity of $\mathbf{H}_{x,k}$ can even improve the efficiency more significantly because the measured data now are processed in small blocks.

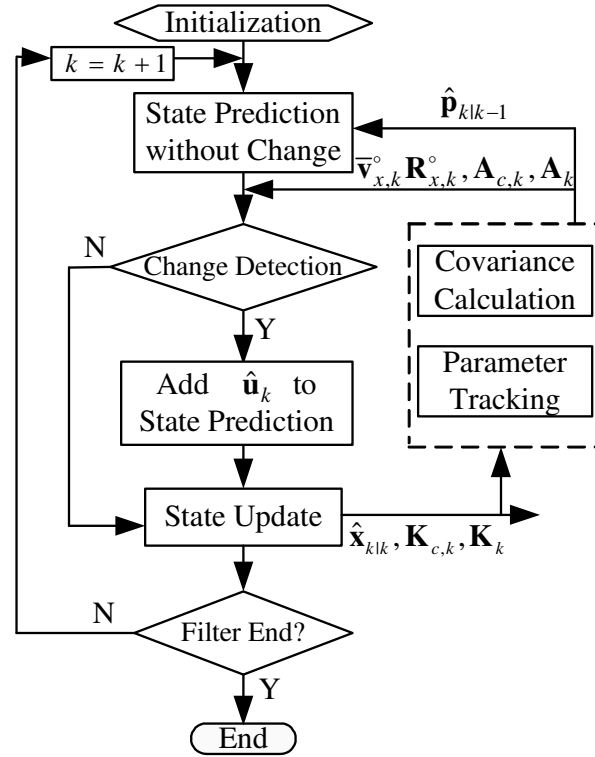


Fig. 5.1 State tracking procedure with abrupt-change detection

For systems with hybrid synchrophasor and (conventional) SCADA data or with SCADA data only, the proposed dynamic model is still applicable. To preserve the linearity of measurements, the SCADA data can be preprocessed by traditional state estimators and the estimated voltages can be treated as refined linear measurements [Zhou06]. The proposed approach and models are then applied with synchrophasor data plus the refined voltages. For SCADA-data only systems, the iterative EKF [Bar-Shalom01], which usually performs better than EKF, can be adopted for filtering. With the model in (5.6), the computation at each instant is almost the same as that of the static WLS estimation.

5.4.3 On Observability, Bad-Data Processing and Accuracy

In the static SE, although the measurement system in design always guarantees the observability, sometimes missing data may make the system unobservable. With

synchrophasors, observability analysis becomes easy since the bus-voltage and branch-current phasors are metered directly [Gou01]. The mathematical equivalence between the filter's update and the WLS estimation implies that the requirement for observability in our approach is similar to but simpler than that in the static SE. In fact, with the proposed approach, observability analysis is almost dispensable. When missing data occur occasionally, a predicted state still makes the system state observable at the current time instant.

Traditional bad-data processing techniques [Monticelli00] [Abur04] are still applicable to preprocess the data if the one-scan redundancy is adequate. However, the proposed approach itself can reduce the effect of bad data through the estimated change $\hat{\mathbf{u}}_k$. For example, if a bad-data situation occurs at time k , it results in a detected $\hat{\mathbf{u}}_k$. At time $k+1$, it is very likely that another $\hat{\mathbf{u}}_{k+1}$, which balances $\hat{\mathbf{u}}_k$ out, is detected. If found, the state at time k be recalculated.

The solutions at each stage are obtained based on the best linear unbiased estimation (BLUE) criterion [Li03]: The state estimates are optimal in the sense of linear minimum-mean-square-error if there is no abrupt change; if abrupt changes occur and are detected, the overall solution is not necessarily optimal but the linear WLS estimates of the changes do improve the state-tracking accuracy.

5.5 Comparative Experiments

Simulations are performed on a realistic 18-bus system, where comparative experiments regarding estimation accuracy and tracking convergence in three scenarios, such as (a) an abrupt state change occurs and the line parameters are all known; (b) no abrupt change occurs but one branch contains uncertain admittance; and (c) both abrupt state changes and uncertain parameters are involved, are carried out.

In the following examples, the involved parameter estimates do not change. Cases with varying parameter estimates will be tested in Chapter 6 where parameter-tracking techniques are discussed. The true, measured and estimated (voltage and current) phasors are all in the rectangular form and the output (estimates) may be converted into the angular form without

loss of accuracy. The true states are simulated by adding zero-mean Gaussian noise to the base-case network data. The synchronous measurements of bus voltages and branch currents are generated similarly, where the errors outside the 3σ region are discarded to exclude the effect of bad data.

The realistic system depicted in Fig. 5.2, which is from a 220kV network supplying power for a major city in East China, is employed. The line parameters are listed in Table 5.1.

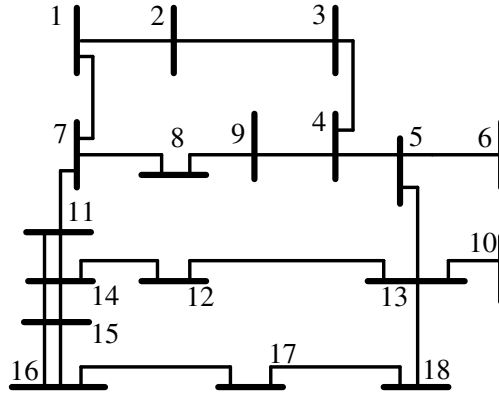


Fig. 5.2 Realistic 18-bus system

In the experiments, bus 1 is designated as the reference bus. The standard deviations of the real and imaginary voltage process noises are 0.005 and 0.01, respectively. All the bus voltages and branch currents are metered by synchrophasors (SMU-type PMU) and the covariance matrices of the errors in voltage and current measurements are

$$\mathbf{R}_v = \begin{bmatrix} 4 & 0.4 \\ 0.4 & 1 \end{bmatrix} \times 10^{-6} \text{ and } \mathbf{R}_c = \begin{bmatrix} 25 & 2.25 \\ 2.25 & 2.25 \end{bmatrix} \times 10^{-6}$$

which are set to be greater than manufacturer's specifications.

The classic measure, root mean square errors (RMSE), which evaluates the average performance (or accuracy) of the state tracking, is adopted, where N_{run} is the total number of Monte-Carlo runs:

TABLE 5.1
BRANCH PARAMETERS (18-BUS SYSTEM)

Br.	1-2	2-3	3-4	1-7	7-8	8-9	9-4	4-5	5-6	7-11
g_{mn}	23.708	9.084	8.129	8.049	9.552	15.742	16.444	11.978	13.835	12.842
b_{mn}	-106.124	-39.765	-35.744	-35.407	-54.456	-69.619	-72.827	-52.742	-62.466	-57.213
b_{mn0}	0.0121	0.0316	0.0362	0.0376	0.0036	0.0182	0.0180	0.0249	0.0205	0.0230
Br.	5-13	11-14	14-12	12-13	13-10	14-15	15-16	13-18	16-17	17-18
g_{mn}	10.413	7.892	13.893	18.076	27.760	9.157	4.046	6.009	7.380	8.006
b_{mn}	-45.449	-38.169	-60.418	-81.728	-122.667	-45.284	-19.586	-26.422	-32.371	-34.958
b_{mn0}	0.0288	0.1202	0.0218	0.0250	0.0117	0.1020	0.2427	0.0491	0.0413	0.0380

g_{mn} , b_{mn} and b_{mn0} are serial conductance, serial susceptance and shunt susceptance, respectively. Shunt conductance is always zero.

$$\text{RMSE} = \left[\frac{1}{N_{run}} \sum_{j=1}^{N_{run}} (\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)^T (\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k) \right]^{\frac{1}{2}} \quad (5.30)$$

First, branches 4-9 and 12-13 are assumed suspicious. The erroneous admittances, arranged in the order of serial conductance, serial susceptance and half shunt susceptance, are $[15.0 \ -75.0 \ 0.015]^T$ and $[18.0 \ -82.0 \ 0.024]^T$, respectively. The corresponding standard deviations for the uncertainty are 1.0, 1.2 and 0.002 for branch 4-9 and 1.0, 1.0 and 0.002 for branch 13-12.

The proposed approach considering erroneous parameters' uncertainty, the one ignoring the uncertainty, and the ideal one using the correct parameters are compared. Fig. 5.3 gives 3 sets of voltage RMSEs with 200 Monte Carlo runs by different approaches. The results by the proposed approach with a sample size of 200 are also presented in Table 5.2. It is seen that the trajectory by the proposed approach is a little poorer than the ideal one due to the uncertainty of parameters. It is also clear that ignoring the uncertainty in erroneous parameters leads to inaccurate estimates.

Second, abrupt state changes are introduced into the new experiment while branches 4-9 and 12-13 remain uncertain. The other system and measurement configurations are the same as those in the previous case. An abrupt change occurs at $k = 11$, which is centered at bus 4 and affects several adjacent buses. The increment complex-voltage pairs at buses 9, 4, 5, 3 and 8

are:

$$\begin{aligned} &-0.02500 - j0.01745, \quad 0.04000 - j0.03490, \quad 0.02000 - j0.01396, \\ &-0.01200 - j0.00872, \quad -0.01500 - j0.00872 \end{aligned}$$

The change-detection threshold is $\alpha = 0.01$. A simulation with 200 runs is conducted. The RMSE results by the proposed one are in Table 5.3. Fig. 5.4 also shows two sets of voltage RMSEs using the proposed state tracking and the one without detection and estimation of change, respectively.

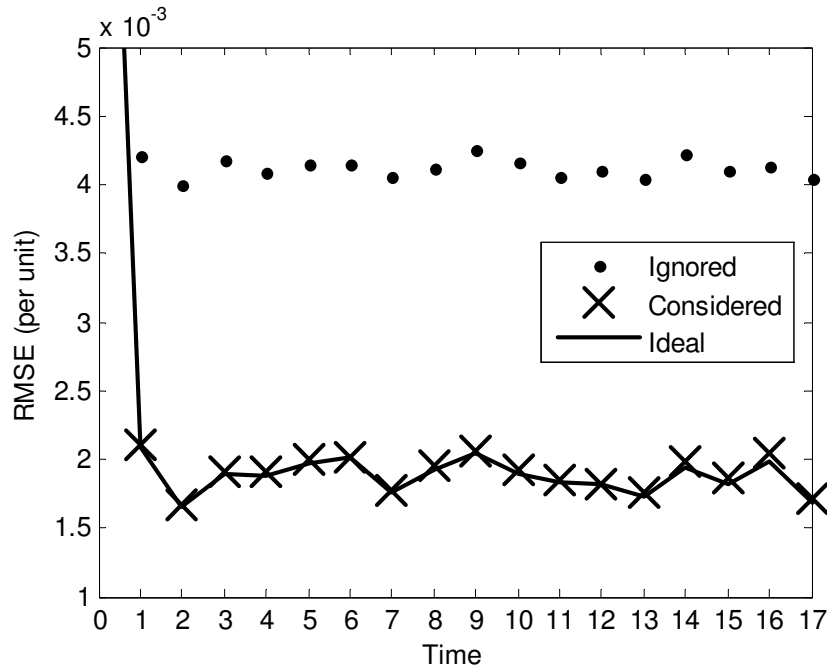


Fig. 5.3 Comparison on effect of parameter uncertainty

At $k = 11$, the values in the ‘Accounted’ trajectory are much smaller than those in the ‘Unaccounted’ one. In addition, at instants 5-10, the ‘Accounted’ values, which should coincide with the ‘Unaccounted’, are still smaller. It implies that the change-detection-and-estimation process has responded to false alarms, as discussed in Sec. 5.4.1. These ‘unexpected’ operations help improve the estimation accuracy. Thus it is shown that the

TABLE 5.2

RMSES OF FIRST EXPERIMENT (BY PROPOSED APPROACH)

Time	0	1	2	3	4	5	6	7	8
Value	0.00942	0.00210	0.00166	0.00191	0.00191	0.00200	0.00201	0.00177	0.00195
Time	9	10	11	12	13	14	15	16	17
Value	0.00205	0.00191	0.00184	0.00182	0.00176	0.00198	0.00186	0.00204	0.00170

TABLE 5.3

RMSES OF SECOND EXPERIMENT (BY PROPOSED APPROACH)

Time	0	1	2	3	4	5	6	7	8
Value	0.00910	0.00249	0.00246	0.00224	0.00246	0.00231	0.00213	0.00248	0.00220
Time	9	10	11	12	13	14	15	16	17
Value	0.00228	0.00212	0.00206	0.00207	0.00223	0.00223	0.00214	0.00217	0.00201

method of change detection and estimation is necessary and effective. In particular, in this most complicated case, the average processing time for each prediction-update cycle (one instant) is less than 0.05 second when the approach is implemented via MATLAB on a computer with a single-core 3.0GHz CPU.

Third, the proposed approach accompanied by the newly-introduced dynamic model in (5.6) is compared with a traditional approach using the random-walk model in (5.1) and an approach using the forecasting model in (5.2). In the forecasting model, the coefficients \mathbf{F} and \mathbf{G} are determined by the classic linear exponential smoothing method [Leite83].

The existing approaches do not consider the abrupt state changes and the possible uncertain parameters, on which the proposed approach has already won out. To make a ‘fair’ comparison, only the case without abrupt changes and uncertain parameters is tested. All the other system and measurement configurations are the same as those in the first test of Sec. 5.5.5. The RMSE trajectories by the proposed approach and the random-walk model based approach are presented in Fig. 5.5. The plots indicate that the random-walk model leads to results inferior to those by the proposed one.

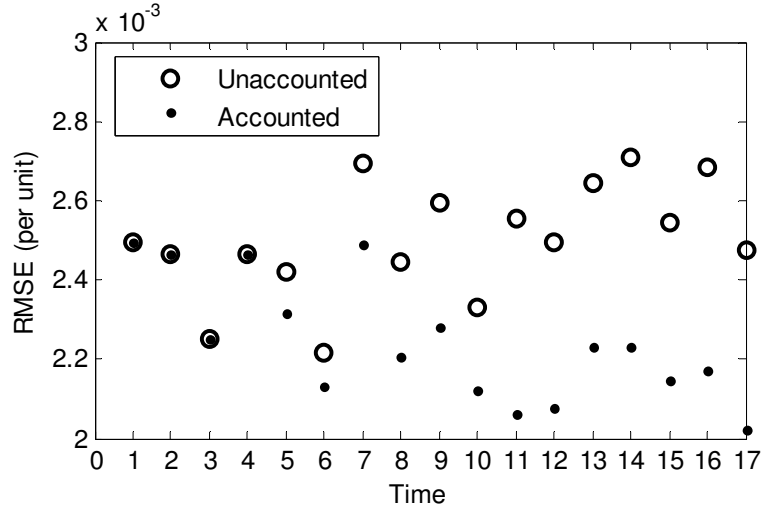


Fig. 5.4 Comparison on effect of abrupt-state-change detection

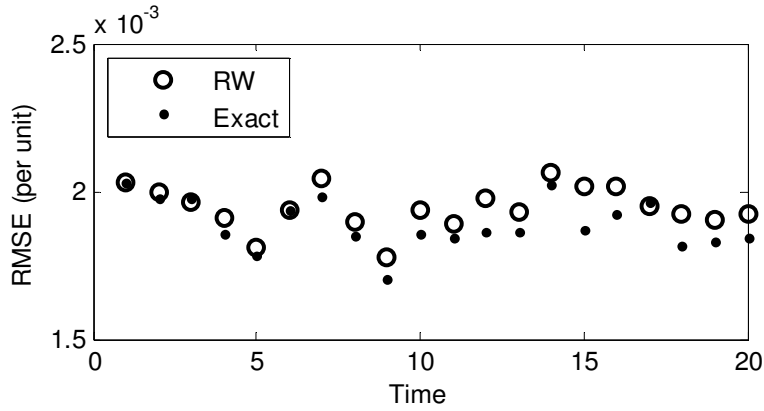


Fig. 5.5 Comparison with random-walk model

Fig. 5.6 plots the RMSE trajectories to compare the proposed approach with the forecasting model based approach. The setting for the forecasting model is the same as that in [Leite83]. Clearly, the results are consistent with the analysis in Sec. 5.4.1 and the proposed approach has better accuracy. In field applications, the PMU accuracy and the number of measurements may not be as high as in the simulation setting, and it is also possible that some measured data are not valid at some instants. In these realistic situations, the improved performance of the

proposed approach over the compared ones in Figs. 5.4, 5.5 and 5.6 will be more significant since the estimation accuracy now relies more on the prediction model.

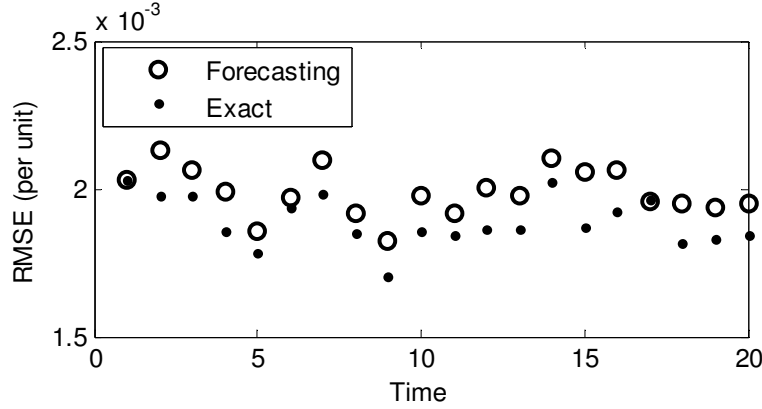


Fig. 5.6 Comparison with forecasting model

5.6 Conclusions

This chapter presents a novel approach to state tracking with correlated prediction-measurement errors and possible abrupt state change. Prediction models have been analyzed, leading to a new model. An adaptive filtering procedure, accompanied by detection and estimation of abrupt changes, has been developed, which also considers the correlation between the prediction and the pseudo measurement errors. Simulations illustrate the performance of the approach under multiple conditions. High estimation accuracy is achieved with this approach. In addition, the approach can be simplified and applied to conventional DSE problems with synchrophasor data. It can also be introduced to other state estimation problems with conventional SCADA data.

5.7 Appendices

5.7.1 Appendix A: Derivation of Filter on Correlated Errors and Noise

The linear system tracked by the standard Kalman filter (KF) usually assumes whiteness and non-correlation. Reference [Li07] presents an optimal KF dealing with various correlations. The following derivation, which follows [Li07] and adopts the predictor-corrector form of KF [Bar-Shalom01] that is easy for implementation and coding, assumes that the prediction and

the measurement errors are correlated. For the system described by

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{c,k} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{F}_{k-1} \\ \mathbf{0} & \mathbf{F}_{k-1} \end{bmatrix} \tilde{\mathbf{x}}_{k-1} + \begin{bmatrix} \mathbf{M}_k \\ \mathbf{M}_k \end{bmatrix} + \begin{bmatrix} \mathbf{w}_k \\ \mathbf{0} \end{bmatrix} \quad (5.31)$$

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k = [\mathbf{H}_k \quad \mathbf{0}] \tilde{\mathbf{x}}_k + \mathbf{v}_k \quad (5.32)$$

with \mathbf{M} for a general input and \mathbf{w} for zero-mean white noise, one predictor-corrector cycle (at one time instant) of the filter includes prediction and update as follows.

1) Prediction:

The prediction is exactly the same as that in the standard KF. That is,

$$\hat{\tilde{\mathbf{x}}}_{k|k-1} = \begin{bmatrix} \mathbf{0} & \mathbf{F}_{k-1} \\ \mathbf{0} & \mathbf{F}_{k-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{k-1|k-1} \\ \hat{\mathbf{x}}_{c,k-1|k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{M}_k \\ \mathbf{M}_k \end{bmatrix}$$

or $\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_{c,k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{c,k-1|k-1} + \mathbf{M}_k \quad (5.33)$

The corresponding mean-square error (MSE) matrix is

$$\begin{aligned} \tilde{\mathbf{P}}_{k|k-1} &= \text{cov}(\tilde{\mathbf{x}}_k - \hat{\tilde{\mathbf{x}}}_{k|k-1}) = \tilde{\mathbf{F}}_{k-1} \tilde{\mathbf{P}}_{k-1|k-1} \tilde{\mathbf{F}}_{k-1}^T + \text{diag}(\mathbf{Q}_k, \mathbf{0}) \\ &= \begin{bmatrix} \mathbf{F}_{k-1} \mathbf{P}_{c,k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_k & \mathbf{F}_{k-1} \mathbf{P}_{c,k-1|k-1} \mathbf{F}_{k-1}^T \\ \mathbf{F}_{k-1} \mathbf{P}_{c,k-1|k-1} \mathbf{F}_{k-1}^T & \mathbf{F}_{k-1} \mathbf{P}_{c,k-1|k-1} \mathbf{F}_{k-1}^T \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{c,k|k-1} \\ \mathbf{P}_{c,k|k-1} & \mathbf{P}_{c,k|k-1} \end{bmatrix} \end{aligned}$$

where $\text{cov}(\cdot)$ is for covariance.

In other words,

$$\begin{cases} \mathbf{P}_{k|k-1} = \mathbf{P}_{c,k|k-1} + \mathbf{Q}_k \\ \mathbf{P}_{c,k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{c,k-1|k-1} \mathbf{F}_{k-1}^T \end{cases} \quad (5.34)$$

2) Update:

The update includes the following steps similar to those in the standard KF.

The predicted measurement is

$$\hat{\mathbf{z}}_{k|k-1} = \tilde{\mathbf{H}}_k \hat{\tilde{\mathbf{x}}}_{k|k-1} + \bar{\mathbf{v}}_k = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} + \bar{\mathbf{v}}_k \quad (5.35)$$

where $\bar{\mathbf{v}}$ is the mean of \mathbf{v} . By definition, the innovation is

$$\Delta \mathbf{z}_k = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} \quad (5.36)$$

The covariance of the innovation (assuming measurement-prediction-error correlation exists) is

$$\begin{aligned}\mathbf{S}_k &= \text{cov}(\Delta \mathbf{z}_k) = \tilde{\mathbf{H}}_k \tilde{\mathbf{P}}_{k|k-1} \tilde{\mathbf{H}}_k^T + \mathbf{R}_k + \tilde{\mathbf{H}}_k \tilde{\mathbf{A}}_k + (\tilde{\mathbf{H}}_k \tilde{\mathbf{A}}_k)^T \\ &= \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k + \mathbf{H}_k \mathbf{A}_k + (\mathbf{H}_k \mathbf{A}_k)^T\end{aligned}\quad (5.37)$$

where $\tilde{\mathbf{H}}_k \tilde{\mathbf{A}}_k + (\tilde{\mathbf{H}}_k \tilde{\mathbf{A}}_k)^T$ is introduced by the newly-counted correlation. The filter gain is also modified with $\tilde{\mathbf{A}}_k$. That is,

$$\begin{aligned}\tilde{\mathbf{K}}_k &= (\tilde{\mathbf{P}}_{k|k-1} \tilde{\mathbf{H}}_k^T + \tilde{\mathbf{A}}_k) \mathbf{S}_k^{-1} \\ &= \left(\begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{c,k|k-1} \\ \mathbf{P}_{c,k|k-1} & \mathbf{P}_{c,k|k-1} \end{bmatrix} \begin{bmatrix} \mathbf{H}_k^T \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{A}_k \\ \mathbf{A}_{c,k} \end{bmatrix} \right) \mathbf{S}_k^{-1} \\ &= \begin{bmatrix} \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{A}_k \\ \mathbf{P}_{c,k|k-1} \mathbf{H}_k^T + \mathbf{A}_{c,k} \end{bmatrix} \mathbf{S}_k^{-1} = \begin{bmatrix} \mathbf{K}_k \\ \mathbf{K}_{c,k} \end{bmatrix}\end{aligned}$$

In other words,

$$\begin{cases} \mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{A}_k \\ \mathbf{K}_{c,k} = \mathbf{P}_{c,k|k-1} \mathbf{H}_k^T + \mathbf{A}_{c,k} \end{cases}\quad (5.38)$$

Then the updated estimate and its MSE matrix are

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \begin{bmatrix} \hat{\mathbf{x}}_{k|k}^T & \hat{\mathbf{x}}_{c,k|k}^T \end{bmatrix}^T = \hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{K}}_k \Delta \mathbf{z}_k \\ &= \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \hat{\mathbf{x}}_{c,k|k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_k \\ \mathbf{K}_{c,k} \end{bmatrix} \Delta \mathbf{z}_k\end{aligned}\quad (5.39)$$

$$\begin{aligned}\tilde{\mathbf{P}}_{k|k} &= \text{cov}(\tilde{\mathbf{x}}_k - \hat{\mathbf{x}}_{k|k}) = \tilde{\mathbf{P}}_{k|k-1} - \tilde{\mathbf{K}}_k \mathbf{S}_k \tilde{\mathbf{K}}_k^T \\ &= \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{c,k|k-1} \\ \mathbf{P}_{c,k|k-1} & \mathbf{P}_{c,k|k-1} \end{bmatrix} - \begin{bmatrix} \mathbf{K}_k \\ \mathbf{K}_{c,k} \end{bmatrix} \mathbf{S}_k \begin{bmatrix} \mathbf{K}_k \\ \mathbf{K}_{c,k} \end{bmatrix}^T = \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{k|k} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{P}_{c,k|k} \end{bmatrix}\end{aligned}\quad (5.40)$$

with $\mathbf{D}_{11} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T$,

$$\mathbf{D}_{21} = \mathbf{P}_{c,k|k-1} - \mathbf{K}_{c,k} \mathbf{S}_k \mathbf{K}_k^T,$$

$$\mathbf{D}_{12} = \mathbf{P}_{c,k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_{c,k}^T,$$

and $\mathbf{D}_{22} = \mathbf{P}_{c,k|k-1} - \mathbf{K}_{c,k} \mathbf{S}_k \mathbf{K}_{c,k}^T$. In summary,

$$\begin{cases} \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \Delta \mathbf{z}_k \\ \hat{\mathbf{x}}_{c,k|k} = \hat{\mathbf{x}}_{c,k|k-1} + \mathbf{K}_{c,k} \Delta \mathbf{z}_k \\ \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \\ \mathbf{P}_{c,k|k} = \mathbf{P}_{c,k|k-1} - \mathbf{K}_{c,k} \mathbf{S}_k \mathbf{K}_{c,k}^T \end{cases} \quad (5.41)$$

5.7.2 Appendix B: Derivation of Abrupt Change Estimation

This derivation, inspired by the input estimation approach in [Bar-Shalom01], is based on the model in (5.6). Hypotheses H_0 and H_1 are considered as follows.

1) $H_0: \mathbf{u}_k = 0$

According to (5.14), the prediction is

$$\hat{\mathbf{x}}_{k|k-1}^- = \hat{\mathbf{x}}_{c,k-1|k-1}^- \quad (5.42)$$

Then the corresponding innovation is

$$\Delta \mathbf{z}_{x,k}^- = \mathbf{z}_{x,k} - \mathbf{H}_{x,k}^\circ \hat{\mathbf{x}}_{k|k-1}^- = \mathbf{z}_{x,k} - \mathbf{H}_{x,k}^\circ \hat{\mathbf{x}}_{c,k-1|k-1}^- \quad (5.43)$$

2) $H_1: \mathbf{u}_k \neq 0$,

From (5.14), the prediction is

$$\hat{\mathbf{x}}_{k|k-1} = \hat{\mathbf{x}}_{c,k-1|k-1} + \mathbf{u}_k \quad (5.44)$$

The innovation under H_1 is

$$\begin{aligned} \Delta \mathbf{z}_{x,k} &= \mathbf{z}_{x,k} - \mathbf{H}_{x,k}^\circ \hat{\mathbf{x}}_{k|k-1} \\ &= \mathbf{z}_{x,k} - \mathbf{H}_{x,k}^\circ \hat{\mathbf{x}}_{c,k-1|k-1} - \mathbf{H}_{x,k}^\circ \mathbf{u}_k \end{aligned} \quad (5.45)$$

Comparing (5.43) with (5.45), and also considering the initial condition $\hat{\mathbf{x}}_{c,k-1|k-1}^- = \hat{\mathbf{x}}_{c,k-1|k-1}$, we have

$$\Delta \mathbf{z}_{x,k}^- = \mathbf{H}_{x,k}^\circ \mathbf{u}_k + \Delta \mathbf{z}_{x,k} \quad (5.46)$$

which implies that $\Delta \mathbf{z}_{x,k}^-$ and $\Delta \mathbf{z}_{x,k}$ form a measurement-and-noise pair with respect to \mathbf{u}_k .

In the filter in Sec. 5.3, the covariance \mathbf{S}_k does not depend on the concrete state values in the

filter. Thus $\Delta \mathbf{z}_{x,k}^-$ and $\Delta \mathbf{z}_{x,k}$ have the same covariance. Accordingly, the WLS estimate of \mathbf{u}_k is

$$\hat{\mathbf{u}}_k = [(\mathbf{H}_{x,k}^\circ)^T \mathbf{S}_k^{-1} \mathbf{H}_{x,k}^\circ]^{-1} \mathbf{H}_{x,k}^\circ \mathbf{S}_k^{-1} \Delta \mathbf{z}_{x,k}^- \quad (5.47)$$

Chapter 6: Joint Estimation of State and Parameter with Synchronphasors—Part II: Parameter Tracking

6.1 Introduction

6.1.1 Previous Work

In more and more modern power systems, synchronphasors are deployed to meter bus voltages directly, and branch-current phasors are also collected to maintain measurement redundancy [Phadke02] [Martin08]. Consequently, accurate model parameters are required by state estimators. Unfortunately, incorrect parameters exist and their percentage can be up to 30% in some utilities' database [Kusic04]. Thus parameter estimation is crucial for power system state estimation and other steady-state applications.

As surveyed by [Zarco00], parameter-estimation methods using residual sensitivity analysis or augmenting the parameters into the state vector, with data from a single scan, have been proposed [Liu92] [Liu95] [Alsac98]. These static-estimation methods are usually limited because the measurement redundancy is inadequate in practice. Dynamic-estimation methods using multi-scan data are generally preferred [Debs74]. They have better performance, but require accurate dynamic models and relatively more computation. Typically, the dynamic state vector is augmented [Debs74] [Slutsker96] and the extended Kalman filter (EKF) is applied since the measurement functions are nonlinear. However, EKF-based approaches do not perform well if the problems are highly nonlinear or some components of the augmented state set are sensitive to measurement noise [Li04]. In addition, although several artificial intelligent techniques have been introduced to improve prediction models in state estimation [Sinha99] [Lin03] [Huang04], less attention has been paid to the dynamic behavior of the involved voltages and parameters in parameter estimation. An inappropriate model may lead to serious model mismatch errors, especially for filtering over a long period.

In essence, the EKF-based approaches (in power system state estimation) are industrial applications of nonlinear filtering techniques which are largely classified into two categories: point estimation and density estimation [Debs74]. In addition to EKF, other point estimation methods, such as unscented filter [Julier04] and Gaussian filter [Ito00], employ different approximation techniques to track nonlinear systems. Their performance is limited if the specific system is not well approximated. Density-estimation methods, such as Fokker-Planck equation based methods [Ahmed98] and particle filtering (PF) [Djuric03], try to provide exact or approximate posterior probability density functions. However, they require sufficient one-scan measurements and the computational requirements are demanding, especially in high-dimensional cases.

For high-dimension low-redundancy nonlinear parameter estimation, we propose a balanced method which adopts merits from both EKF and PF. It follows the simple structure of EKF but further accounts for the uncertain effects such as involved bus voltages and high-order terms (of Taylor's expansion) in EKF as the pseudo measurement errors correlated with the prediction errors. Correspondingly, a recently-developed optimal filtering technique that can handle correlation [Li07] is introduced (see Chapter 5). We also introduce random samples from the idea of PF to evolve the pseudo-error ensembles and to evaluate the statistics related to the pseudo errors, where the error-ensemble sampling does not rely on the measurement redundancy and is much easier to implement than PF. Based on this balanced method, the joint state-and-parameter estimation considering complicated behaviors of voltages and parameters is discussed.

6.1.2 Our Contributions

The problem of parameter estimation with synchrophasor data is formulated as two loosely-coupled subproblems: state tracking (presented in Chapter 5) and parameter tracking. This chapter focuses on the parameter tracking, where only line parameters are estimated and the involved bus voltages are now treated as uncertain “measurement- function coefficients”. The work includes:

1) A new prediction model accounting for the effect of prior knowledge and moving parameter means is proposed. The measurement model is also improved by introducing a concept of pseudo measurement error to cover the uncertainty in the estimated voltages, high-order terms and measurement errors.

2) An adaptive filter is developed, based on the optimal filtering with correlated prediction-measurement errors.

3) A sliding-window method is proposed to detect the moving tendency of parameters and adjust the transition matrix adaptively. It is embedded in the adaptive filter.

4) A sample-based method, namely, error-ensemble evolution, is used to evaluate the correlation between pseudo measurement errors and prediction errors.

5) In addition, a refinery-and-adaptation method is suggested for the use of realistic systems where the data-acquisition rate is high. Specifically, with the above setting, parameters from different lines can be accurately calibrated in parallel.

In particular, the work of this chapter has been published in [Bian11b].

6.2 Parameter Estimation

The goal is to calibrate the steady-state parameters in the π -type equivalent circuits (Pi-Model) of transmission lines using multi-scan synchrophasor data. The bus voltages at the nodes connected by these lines, which are included in the measurement function, also need to be determined. To obtain accurate estimates of the two voltages, a few adjacent buses usually are involved. Accordingly, the overall measurement equation can be written as

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{p}_k) + \mathbf{v}_k \quad (6.1)$$

where

\mathbf{z} vector of measurements for \mathbf{p} and \mathbf{x}

\mathbf{p} vector of admittance over to-be-calibrated lines, e.g., admittance of a single line

(by the π -type circuit) is $[g_{mn} \ g_{mn0} \ b_{mn} \ b_{mn0}]^T$

g_{mn}	serial conductance in the π -type circuit for a line connecting nodes m and n
g_{mn0}	half shunt conductance in the π -type circuit for a line connecting nodes m and n
b_{mn}	serial susceptance of the π -type circuit for a line connecting nodes m and n
b_{mn0}	half shunt susceptance of the π -type circuit for a line connecting nodes m and n
\mathbf{x}	bus voltages
$\mathbf{h}(\cdot)$	function (set) relating \mathbf{p} and \mathbf{x} with measurements
\mathbf{v}	vector of measurement noise

In conventional parameter estimation, \mathbf{p} and \mathbf{x} are solved as an augmented state, which leads to a typical nonlinear estimation problem and estimation accuracy is not guaranteed. We formulate this problem as two loosely-coupled subproblems: state tracking and parameter tracking, incorporated by an error-ensemble-evolution method dealing with the coupling. In one (state or parameter) tracking procedure, the other estimand (\mathbf{p} or \mathbf{x}) is taken as uncertain “coefficients” in measurement function related to the current one. Consequently, the dynamic behaviors of the state and the parameters are studied separately since they do act differently in reality and the effect of the uncertain coefficients is represented by pseudo measurement errors. Specific filters are also developed. Specifically, in parameter tracking, the measurement set includes the involved branch-current data only and each current phasor pair is exclusively incident to one branch. This implies that parameters from different lines can be calibrated separately.

The state tracking has been presented in Chapter 5. This chapter discusses the overall approach and focuses mainly on the adaptive parameter tracking with uncertain voltages and the method to evaluate the involved correlation.

6.3 Formulation of Parameter Tracking

As discussed in Sec. 6.2, admittances from different transmission lines can be processed in parallel. Thus, a single transmission line is considered in the subsequent discussion.

6.3.1 Prediction Model

The dynamic behavior of line parameters is modeled as

$$\begin{cases} \mathbf{p}_k = \mathbf{p}_{c,k} + \mathbf{w}_{p,k} \\ \mathbf{p}_{c,k} = f(\mathbf{p}_{c,k-1}, k) \end{cases} \quad (6.2)$$

where

\mathbf{p}_c mean (center) of \mathbf{p}

\mathbf{w}_p vector of zero-mean process noise in \mathbf{p} , caused by small disturbances (e.g., wind or rain)

$f(\cdot)$ function for deterministic moving tendency of \mathbf{p}_c

This prediction model emphasizes two facts: (a) rather than a random walk, the true values vary randomly around a deterministic “center”; (b) the center is not necessarily constant. It may move over time since the transmission line can shrink or expand from heat, moisture, etc. Specifically, an adaptive model is utilized to approximate this moving tendency:

$$\mathbf{p}_{c,k} = \mathbf{F}_{p,k-1} \mathbf{p}_{c,k-1}, \quad (6.3)$$

where transition matrix \mathbf{F}_p is adjusted by the sliding-window method (see Sec. 6.4.2).

6.3.2 Measurement Model

The voltages have been estimated using the scheme presented in Chapter 5. Thus only the synchronized branch currents through the to-be-calibrated line are considered in the measurement set. All the data are in the rectangular form. For those synchrophasors providing data in the angular-form only, the coordinate transformation used in [Bi08] can be employed to preprocess the data. The measurement model is

$$\mathbf{z}_{p,k} = \mathbf{H}_{p,k} \mathbf{p}_k + \mathbf{v}_{p,k}, \quad (6.4)$$

where

$$\mathbf{H}_{p,k} = \begin{bmatrix} x_{r,m} - x_{r,n} & x_{i,n} - x_{i,m} & x_{r,m} & -x_{i,m} \\ x_{i,m} - x_{i,n} & x_{r,m} - x_{r,n} & x_{i,m} & x_{r,m} \\ x_{r,n} - x_{r,m} & x_{i,m} - x_{i,n} & x_{r,n} & -x_{i,n} \\ x_{i,n} - x_{i,m} & x_{r,n} - x_{r,m} & x_{i,n} & x_{r,n} \end{bmatrix}_k \quad (6.5)$$

\mathbf{z}_p vector of branch-current measurements

\mathbf{v}_p vector of branch-current measurement noise

x entry of \mathbf{x}

Note that (6.5) is for two pairs of complex branch currents in reverse directions. For more redundant current observations, (6.5) will be extended. Furthermore, since not \mathbf{x}_k but its estimate $\hat{\mathbf{x}}_k$ is available, (6.4) is refined as

$$\mathbf{z}_{p,k} = \mathbf{H}_{p,k}^\circ \mathbf{p}_k + \mathbf{v}_{p,k}^\circ, \quad (6.6)$$

where $\mathbf{H}_{p,k}^\circ$ is a coefficient matrix obtained from $\mathbf{H}_{p,k}$ by replacing \mathbf{x}_k with $\hat{\mathbf{x}}_k$. The pseudo measurement noise is

$$\mathbf{v}_{p,k}^\circ = \Delta \mathbf{H}_{p,k} (\hat{\mathbf{p}}_{k|k-1} + \Delta \mathbf{p}_{k|k-1}) + \mathbf{v}_{p,k}, \quad (6.7)$$

where $\hat{\mathbf{p}}_{k|k-1}$ is predicted estimate of \mathbf{p} . $\Delta \mathbf{p}_{k|k-1} = \mathbf{p}_k - \hat{\mathbf{p}}_{k|k-1}$, $\Delta \mathbf{H}_{p,k} = \mathbf{H}_{p,k} - \mathbf{H}_{p,k}^\circ$, and $\Delta \mathbf{H}_{p,k}$ depends on $\Delta \mathbf{x}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$.

The updated estimate $\hat{\mathbf{x}}_{k|k}$ is used as $\hat{\mathbf{x}}_k$ in this chapter. The process and (original) measurement noises are assumed zero-mean, white Gaussian, and mutually independent.

6.4 Tracking of Line Parameters

6.4.1 Basic Filter

The adaptive filter for the parameter tracking is arranged in the predictor-corrector form and includes recursive steps as below (also see Fig. 6.1).

1) Initialization: $k = 0$

$$\hat{\mathbf{p}}_{0|0} = \hat{\mathbf{p}}_{c,0|0}; \quad \tilde{\mathbf{P}}_{0|0} = \tilde{\mathbf{P}}_{c,0|0} \quad (6.8)$$

2) Recursion: $k := k + 1$

Determine $\mathbf{F}_{p,k-1}$ (see Sec. 6.4.2)

Prediction:

$$\hat{\mathbf{p}}_{k|k-1} = \hat{\mathbf{p}}_{c,k|k-1}; \quad \hat{\mathbf{p}}_{c,k|k-1} = \mathbf{F}_{p,k-1} \hat{\mathbf{p}}_{c,k-1|k-1} \quad (6.9)$$

$$\begin{cases} \tilde{\mathbf{P}}_{k|k-1} = \tilde{\mathbf{P}}_{c,k|k-1} + \mathbf{Q}_{p,k} \\ \tilde{\mathbf{P}}_{c,k|k-1} = \mathbf{F}_{p,k-1} \tilde{\mathbf{P}}_{c,k-1|k-1} \mathbf{F}_{p,k-1}^T \end{cases} \quad (6.10)$$

$$\hat{\mathbf{z}}_{p,k|k-1} = \mathbf{H}_{p,k}^\circ \hat{\mathbf{p}}_{k|k-1} + \bar{\mathbf{v}}_{p,k}^\circ \quad (6.11)$$

Update:

$$\tilde{\mathbf{S}}_k = \mathbf{H}_{p,k}^\circ \tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_{p,k}^{\circ T} + \mathbf{R}_{p,k}^\circ + \mathbf{H}_{p,k}^\circ \tilde{\mathbf{A}}_k + \tilde{\mathbf{A}}_k^T \mathbf{H}_{p,k}^{\circ T} \quad (6.12)$$

$$\begin{cases} \tilde{\mathbf{K}}_k = (\tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_{p,k}^{\circ T} + \tilde{\mathbf{A}}_k) \tilde{\mathbf{S}}_k^{-1} \\ \tilde{\mathbf{K}}_{c,k} = (\tilde{\mathbf{P}}_{c,k|k-1} \mathbf{H}_{p,k}^{\circ T} + \tilde{\mathbf{A}}_{c,k}) \tilde{\mathbf{S}}_k^{-1} \end{cases} \quad (6.13)$$

$$\Delta \mathbf{z}_{p,k} = \mathbf{z}_{p,k} - \hat{\mathbf{z}}_{p,k|k-1} \quad (6.14)$$

Finally,

$$\begin{cases} \hat{\mathbf{p}}_{k|k} = \hat{\mathbf{p}}_{k|k-1} + \tilde{\mathbf{K}}_k \Delta \mathbf{z}_{p,k} \\ \hat{\mathbf{p}}_{c,k|k} = \hat{\mathbf{p}}_{c,k|k-1} + \tilde{\mathbf{K}}_{c,k} \Delta \mathbf{z}_{p,k} \end{cases} \quad (6.15)$$

$$\begin{cases} \tilde{\mathbf{P}}_{k|k} = \tilde{\mathbf{P}}_{k|k-1} - \tilde{\mathbf{K}}_k \tilde{\mathbf{S}}_k \tilde{\mathbf{K}}_k^T \\ \tilde{\mathbf{P}}_{c,k|k} = \tilde{\mathbf{P}}_{c,k|k-1} - \tilde{\mathbf{K}}_{c,k} \tilde{\mathbf{S}}_k \tilde{\mathbf{K}}_{c,k}^T \end{cases} \quad (6.16)$$

Here,

$\tilde{\mathbf{P}}$ covariance of estimation error in $\hat{\mathbf{p}}$

$\tilde{\mathbf{P}}_c$ covariance of estimation error in $\hat{\mathbf{p}}_c$

The following quantities are assumed known:

\mathbf{Q}_p covariance of parameter process noise \mathbf{w}_p

\mathbf{R}_p covariance of original measurement noise \mathbf{v}_p

In addition, the mean and covariance of $\mathbf{v}_{p,k}^\circ$ ($\bar{\mathbf{v}}_p^\circ$ and \mathbf{R}_p°), and the correlation between the prediction and the pseudo measurement errors which includes

$$\tilde{\mathbf{A}}_k = \text{cov}(\Delta \mathbf{p}_{k|k-1}, \mathbf{v}_{p,k}^\circ), \quad (6.17)$$

$$\tilde{\mathbf{A}}_{c,k} = \text{cov}(\Delta \mathbf{p}_{c,k|k-1}, \mathbf{v}_{p,k}^\circ), \quad (6.18)$$

should be preprocessed. It is discussed in detail in Sec. 6.5.

6.4.2 Adjustment of Transition Matrix

Detection of the moving tendency in the parameter center is based on the following fact: As opposite to that the mismatched prediction model drives the predicted values farther and farther away, the newly-arrived measurements tend to pull the updated values back toward the truth. Consequently, a sliding-window method based on evidence from N_{det} successive instants is proposed, which contains two hypotheses as

H_0 : The center keeps increasing in the current window;

H_1 : The center keeps decreasing in the current window.

In the case that neither H_0 nor H_1 is correct, the center is assumed to remain constant in the current window.

\mathbf{F}_p is usually assumed diagonal and all the diagonal elements can be adjusted in a similar way. Let \hat{p} be a representative element in $\hat{\mathbf{p}}$ and f_p be the corresponding diagonal entry of \mathbf{F}_p , respectively. Three decision rules are designed as follows:

- 1) Decide H_0 if $\hat{p}_{k|k} / \hat{p}_{k|k-1} > 1$ occurs at least $N_0 + \frac{1}{2} N_{\text{det}}$ times within a window of N_{sw} instants.
- 2) Decide H_1 if $\hat{p}_{k|k} / \hat{p}_{k|k-1} < 1$ occurs at least $N_1 + \frac{1}{2} N_{\text{det}}$ times within the current window.
- 3) Slide the window forward by one step if neither H_0 nor H_1 is decided.

To obtain a smooth parameter-tracking trajectory, we propose a scheme to tune f_p after either H_0 or H_1 is declared:

- a) N_0 and N_1 can be set equal, and a small value such as 1, 2 and 3 can be used;
- b) N_{det} is preset according to field experience (by observing the tracking trajectory);
- c) Past data can be used to reduce the false-alarm probability and eliminate possible delay.

That is, an adjustment window may have $N_{\text{adj}} (> N_{\text{det}})$ instants. If H_0 (or H_1) still holds within the adjustment window, continue tuning f_p ; otherwise, end and move the detection window.

d) In the adjustment window, f_p is turned up (or down) conservatively. For instance, if $\hat{p}_{klk} / \hat{p}_{klk-1} > 1$ occurs $N_{\text{temp}} (\geq N_0 + \frac{1}{2} N_{\text{det}})$ times, f_p should be turned up until N_{temp} is reduced by 1. This stop criterion is milder than those in H_0 and H_1 . In addition, the calculated prediction covariance at the first instant of the window is multiplied by $\beta (> 1.0)$, which assigns more weight to the data than to the prediction since the prediction model is detected to be inaccurate.

In fact, the introduction of $N_0, N_1, N_{\text{det}}, N_{\text{adj}}$, and β makes the adjustment flexible to handle. The well-fitted trajectory is expected to be smooth. Otherwise, when the tuned f_p vibrates frequently, N_0 and N_1 have to be turned up to reduce the false-alarm probability; when f_p does not vibrate frequently but has abrupt increments, N_0 and N_1 have to be turned down to reduce the missing probability.

6.5 Error Evolution and Correlation Calculation

This section deals with the coupling between the state tracking and the parameter tracking. Specifically, it calculates $\bar{\mathbf{v}}_p^\circ, \mathbf{R}_p^\circ, \tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}_c$. It also evaluates $\bar{\mathbf{v}}_x^\circ, \mathbf{A}, \mathbf{A}_c$ and \mathbf{R}_x° for the state tracking in Chapter 5. We will not calculate the above quantities analytically. An error-ensemble-evolution method is used to evaluate them numerically. The main ideas

include: (i) evolve the ensemble of estimation errors for the coupled state and parameter via error evolution; (ii) combine two sets of error samples to get the ensemble of pseudo errors; and (iii) evaluate the relevant quantities.

Details of the error evolution are given in the Appendix. The correlation calculation consists of the following steps:

First, perform initialization: $k = 0$

- a) Obtain $\Delta \mathbf{x}_{c,00}^{(j)}$ by sampling from $N(0, \mathbf{P}_{c,00})$
- b) Obtain $\Delta \mathbf{p}_{c,00}^{(j)}$ by sampling from $N(0, \tilde{\mathbf{P}}_{c,00})$

Second, begin recursion:

- 1) Set $k := k + 1$
- 2) Obtain $\mathbf{v}_{x,k}^{(j)}$ when $\hat{\mathbf{p}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k-1}$ are ready:
 - a) Obtain $\mathbf{w}_{p,k}^{(j)}$ by sampling from $N(0, \mathbf{Q}_{p,k})$
 - b) Evolve $\begin{cases} \Delta \mathbf{p}_{c,k|k-1}^{(j)} = \mathbf{F}_{p,k-1} \Delta \mathbf{p}_{c,k-1|k-1}^{(j)} \\ \Delta \mathbf{p}_{k|k-1}^{(j)} = \Delta \mathbf{p}_{c,k|k-1}^{(j)} + \mathbf{w}_{p,k}^{(j)} \end{cases}$
 - c) Obtain $\mathbf{w}_{x,k}^{(j)}$ by sampling from $N(0, \mathbf{Q}_{x,k})$
 - d) Evolve $\begin{cases} \Delta \mathbf{x}_{c,k|k-1}^{(j)} = \Delta \mathbf{x}_{c,k-1|k-1}^{(j)} \\ \Delta \mathbf{x}_{k|k-1}^{(j)} = \Delta \mathbf{x}_{c,k|k-1}^{(j)} + \mathbf{w}_{x,k}^{(j)} \end{cases}$
 - e) Obtain $\mathbf{v}_{x,k}^{(j)}$ by sampling from $N(0, \mathbf{R}_{x,k})$
 - f) Obtain $\mathbf{v}_{x,k}^{(j)} = \Delta \mathbf{H}_{x,k}^{(j)} (\hat{\mathbf{x}}_{k|k-1} + \Delta \mathbf{x}_{k|k-1}^{(j)}) + \mathbf{v}_{x,k}^{(j)}$, where $\Delta \mathbf{H}_{x,k}^{(j)}$ only depends on $\Delta \mathbf{p}_{k|k-1}^{(j)}$
- 3) Calculate

$$\bar{\mathbf{v}}_{x,k}^{\circ} = \frac{1}{N_{sample}} \sum_j \mathbf{v}_{x,k}^{(j)} \quad (6.19)$$

$$\mathbf{R}_{x,k}^{\circ} = \frac{1}{N_{sample}} \sum_j (\mathbf{v}_{x,k}^{(j)} - \bar{\mathbf{v}}_{x,k}^{\circ})(\mathbf{v}_{x,k}^{(j)} - \bar{\mathbf{v}}_{x,k}^{\circ})^T \quad (6.20)$$

$$\mathbf{A}_k = \frac{1}{N_{sample}} \sum_j \Delta \mathbf{x}_{k|k-1}^{(j)} (\mathbf{v}_{x,k}^{\circ(j)} - \bar{\mathbf{v}}_{x,k}^{\circ})^T \quad (6.21)$$

$$\mathbf{A}_{c,k} = \frac{1}{N_{sample}} \sum_j \Delta \mathbf{x}_{c,k|k-1}^{(j)} (\mathbf{v}_{x,k}^{\circ(j)} - \bar{\mathbf{v}}_{x,k}^{\circ})^T \quad (6.22)$$

4) Obtain $\mathbf{v}_{p,k}^{\circ(j)}$ when $\hat{\mathbf{p}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$ are ready:

a) Update $\Delta \mathbf{x}_{k|k}^{(j)}$ via error evolution

$$\begin{aligned} \Delta \mathbf{z}_k^{(j)} &= \mathbf{H}_{x,k}^{\circ} \Delta \mathbf{x}_{k|k-1}^{(j)} + \mathbf{v}_{x,k}^{\circ(j)} - \bar{\mathbf{v}}_{x,k}^{\circ} \\ \begin{cases} \Delta \mathbf{x}_{c,k|k}^{(j)} = \Delta \mathbf{x}_{c,k|k-1}^{(j)} - \mathbf{K}_{c,k} \Delta \mathbf{z}_k^{(j)} \\ \Delta \mathbf{x}_{k|k}^{(j)} = \Delta \mathbf{x}_{k|k-1}^{(j)} - \mathbf{K}_k \Delta \mathbf{z}_k^{(j)} \end{cases} \end{aligned}$$

b) Obtain $\mathbf{v}_{p,k}^{\circ(j)} = \Delta \mathbf{H}_{p,k}^{\circ(j)} (\hat{\mathbf{p}}_{k|k-1} + \Delta \mathbf{p}_{k|k-1}^{(j)}) + \mathbf{v}_{p,k}^{(j)}$, where $\Delta \mathbf{H}_{p,k}^{\circ(j)}$ depends only on $\Delta \mathbf{x}_{k|k}^{(j)}$, and $\mathbf{v}_{p,k}^{(j)}$ is a part of $\mathbf{v}_{x,k}^{(j)}$

5) Calculate

$$\bar{\mathbf{v}}_{p,k}^{\circ} = \frac{1}{N_{sample}} \sum_j \mathbf{v}_{p,k}^{\circ(j)}, \quad (6.23)$$

$$\mathbf{R}_{p,k}^{\circ} = \frac{1}{N_{sample}} \sum_j (\mathbf{v}_{p,k}^{\circ(j)} - \bar{\mathbf{v}}_{p,k}^{\circ})(\mathbf{v}_{p,k}^{\circ(j)} - \bar{\mathbf{v}}_{p,k}^{\circ})^T \quad (6.24)$$

$$\tilde{\mathbf{A}}_k = \frac{1}{N_{sample}} \sum_j \Delta \mathbf{p}_{k|k-1}^{(j)} (\mathbf{v}_{p,k}^{\circ(j)} - \bar{\mathbf{v}}_{p,k}^{\circ})^T \quad (6.25)$$

$$\tilde{\mathbf{A}}_{c,k} = \frac{1}{N_{sample}} \sum_j \Delta \mathbf{p}_{c,k|k-1}^{(j)} (\mathbf{v}_{p,k}^{\circ(j)} - \bar{\mathbf{v}}_{p,k}^{\circ})^T \quad (6.26)$$

6) Update $\Delta \mathbf{p}_{k|k}^{(j)}$

$$\Delta \mathbf{z}_k^{(j)} = \mathbf{H}_{p,k}^{\circ} \Delta \mathbf{p}_{k|k-1}^{(j)} + \mathbf{v}_{p,k}^{\circ(j)} - \bar{\mathbf{v}}_{p,k}^{\circ} \quad (6.27)$$

$$\begin{cases} \Delta \mathbf{p}_{c,k|k}^{(j)} = \Delta \mathbf{p}_{c,k|k-1}^{(j)} - \tilde{\mathbf{K}}_{c,k} \Delta \mathbf{z}_k^{(j)} \\ \Delta \mathbf{p}_{k|k}^{(j)} = \Delta \mathbf{p}_{k|k-1}^{(j)} - \tilde{\mathbf{K}}_k \Delta \mathbf{z}_k^{(j)} \end{cases} \quad (6.28)$$

Here the recursion ends if k reaches the maximum number. j is the sample index, N_{sample} the total sample size, and $N(\cdot)$ Gaussian probability density function.

6.6 Procedures and Performance Analysis

6.6.1 Parameter Tracking and Overall Procedure

We have presented three major components of the joint state and parameter estimation. From the viewpoint of parameter tracking, the estimated voltages are uncertain “measurement coefficients” and the uncertainty is represented by pseudo measurement errors. A similar strategy holds when discussed from the viewpoint of state tracking. The error ensemble keeps evolving errors for the evaluation of the statistics related to pseudo errors in both tracking procedures.

Furthermore, the dynamic behaviors of the voltages and the parameters are different. The former may jump abruptly while the latter may increase (or decrease) slowly but persistently. Correspondingly, appropriate detection and adaptation techniques have been developed.

The overview for the joint state-and-parameter estimation is depicted in Fig. 6.1. The proposed method consists of parameter-tracking, state-tracking (in Chapter 5) and correlation-calculation. The parameter tracking procedure is located on the left side.

6.6.2 Accuracy and Complexity

The filtering is based on the newly-developed optimal Kalman filter considering prediction-measurement-error correlation. When the parameter center moves persistently, the adaptation mechanism works out (not necessarily optimal yet) accurate results.

In practice, several detailed modules can still be simplified to save computation. For instance, when estimation errors of parameters are large, the related branch-current measurements contribute little to the state tracking. If the one-scan measurement redundancy allows, the currents can be discarded temporarily to simplify the ensemble evolution.

Based on the fact that the calibration of the serial and shunt admittance of a transmission line only involves the incident branch currents and the bus voltages at two ends, parameters of

different branches can be tracked in parallel. In addition, the above fact also implies that the estimated voltages from other estimators such as WLS estimators in EMS can be applied similarly. Correspondingly, the injection powers and branch flows are now measured data. The error-ensemble evolution can be simplified except that estimation accuracy may be poor using traditional meters.

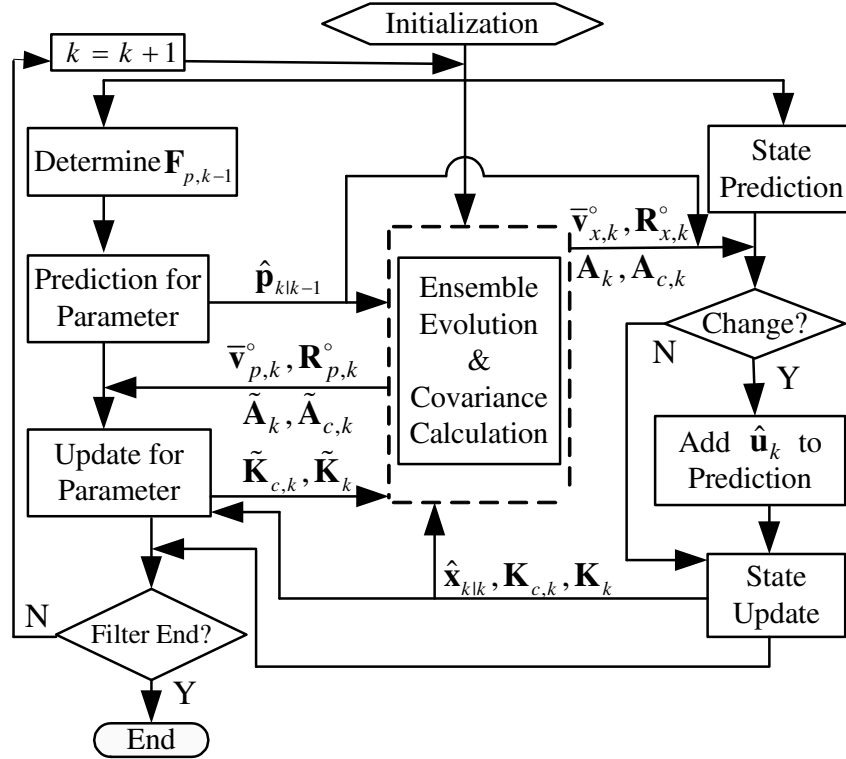


Fig. 6.1 Joint state and parameter estimation procedure

6.6.3 A Practical Scheme

For a realistic measurement system with synchrophasors, the data-acquisition rate is as high as multiple times per second. Within a short period (e.g., 5 or 10 seconds), the moving effect of parameter means as well as the process noise is negligible. As a result, an efficient and practical two-step refinery-and-adaptation scheme is suggested as follows.

First, the prediction model described by (6.2)-(6.3) can be simplified as

$$\mathbf{p}_t = \mathbf{p}_{c,t} = \mathbf{p}_{c,t-1} \quad (6.29)$$

The (simplified) basic filter and error-ensemble evolution can be utilized to roughly estimate \mathbf{p} using data within each period, where the filter is reinitialized at each first instant to release the dependence on the previous period. Second, $\hat{\mathbf{p}}_{tlt}$ and $\tilde{\mathbf{P}}_{tlt}$ at the last instant of each period can be treated as the refined data and put into the adaptive filter (including the adaptation mechanism) with the model

$$\begin{cases} \mathbf{p}_k = \mathbf{p}_{c,k} + \mathbf{w}_{p,k} \\ \mathbf{p}_{c,k} = \mathbf{F}_{p,k-1} \mathbf{p}_{c,k-1} \end{cases} \quad (6.30)$$

At this step, $\hat{\mathbf{p}}_{tlt}$ are direct “measurement” of \mathbf{p}_k and no correlation calculation is needed.

6.7 Simulations

The purposes are to show the performance of the parameter tracking, the effectiveness of the correlation calculation, and the performance of the overall approach. Tests were carried out on a realistic system for comparison. In the tests, the system states and to-be-determined parameters at different times were simulated by adding zero-mean Gaussian noise to the base-case network data. The synchronously-measured data were also generated in this way.

6.7.1 Comparisons with Other Approaches

The proposed approach is compared with an improved EKF approach and an ideal approach to be explained next.

In the standard EKF approach, the predicted values of the augmented state (voltages and parameters) are used in the Jacobian matrix during the linearization. In this example, it is improved by a stepwise strategy which estimates the voltages and the parameters one by one. Then the updated voltages are available to the Jacobian matrix in the parameter tracking.

In the ideal approach, the system state is assumed to be known “perfectly”, which is not feasible in practice but provides a reference for comparison.

The experiment is on a realistic system depicted in Fig. 6.2. This system is a 550kV bulk network supplying power for a major city in China. The line parameters are listed in Table 6.1. Bus 1 is the reference bus. The standard deviations of the real and imaginary voltage process

noises are 0.003 and 0.006, respectively. All the bus voltages and branch currents are metered by synchrophasors (SMU-type PMU) and the covariance matrices of errors in voltage and current measurements are

which are set to be greater than manufacturer's specifications.

Fig. 6.2 12-bus 500kV system

Quantity	Values					
Brn	1-2	2-3	3-4	4-5	5-6	6-7
g_{mn}	2.5820	3.3731	3.8111	2.6971	1.6613	1.4351
b_{mn}	-50.8697	-59.4294	-64.0039	-52.3099	-31.6074	-27.5095
b_{mn0}	0.0214	0.0156	0.0142	0.0190	0.0308	0.0359
Brn	6-8	6-9	9-10	9-11	11-12	12-1
g_{mn}	2.6627	5.3806	4.6541	3.2204	1.2825	1.8301
b_{mn}	-52.6787	-89.7985	-83.6753	-63.0862	-23.2173	-35.4743
b_{mn0}	0.0188	0.0109	0.0119	0.0165	0.0422	0.0297

$$\text{RMSE} = \left[\frac{1}{N_{\text{run}}} \sum_{j=1}^{N_{\text{run}}} (\hat{\mathbf{p}}_{k|k} - \mathbf{p}_k)^T (\hat{\mathbf{p}}_{k|k} - \mathbf{p}_k) \right]^{1/2} \quad (6.31)$$

is adopted, where N_{run} is the total number of Monte-Carlo runs.

The resultant trajectories averaged over 200 runs are plotted in Fig. 6.3. It shows that the proposed approach is much better than the improved EKF approach. The RMSE values from the proposed approach converge faster. The difference between the proposed and the ideal trajectories, caused by the uncertainty in the state estimates, is not large.

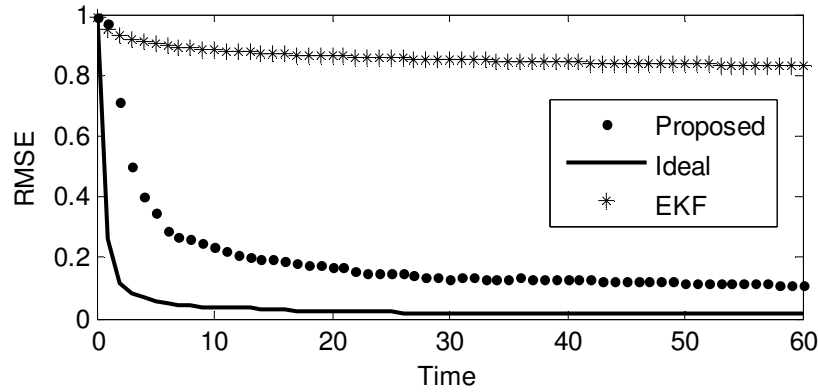


Fig. 6.3 Performance comparison with EKF

6.7.2 Practical Two-Step Implementation of the Approach

This part contains a particular test on the practical refinery-and-adaptation scheme discussed in Sec. 6.6.3. The experiment is implemented on the realistic system in Fig. 6.2, where the centers of line parameters move persistently. All the other system and measurement configurations are the same as those in Sec. 6.7.1. The data-acquisition rate is 5 times per second, and the observation interval covers 25 minutes. That is, 7500 pieces of data are recorded. Specifically, the line parameter in branch 6-9 moves at rates of 0.99999, 0.99997 and 0.99998 over intervals 1500-3200, 3201-4600 and 4601-6000, respectively.

First, the parameters are refined every 20 instants by the simplified tracking procedure using the model (6.29) and the error ensemble evolution method. Second, the simplified adaptive

tracking procedure with the model (6.30) is applied with the 375 pieces of refined data. The lengths of windows are $N_{\text{det}} = 20$ and $N_{\text{adj}} = 30$, respectively. The thresholds are $N_0 = N_1 = 2$. The amplification rate is $\beta = 1.02$. In contrast, another tracking procedure ignoring the possible center moving is also applied.

Fig. 6.4 plots the comparison results by the two approaches. In practice, the ratio of estimation errors to the true values may also be interested. As a result, the RMSE measure defined in (6.31) is modified: each component of $\hat{\mathbf{p}}_{klk} - \mathbf{p}_k$ is divided by the corresponding true value and the eventual result is averaged over the components. Fig. 6.5 is a zoomed-in version of Fig. 6.4. Clearly, the ‘Proposed’ trajectory can reach as low as 0.2% and is significantly better than the ‘Ignored’ one.

In addition, the processing time is less than 0.01 second per instant on a computer with a single-core 3.0GHz CPU.

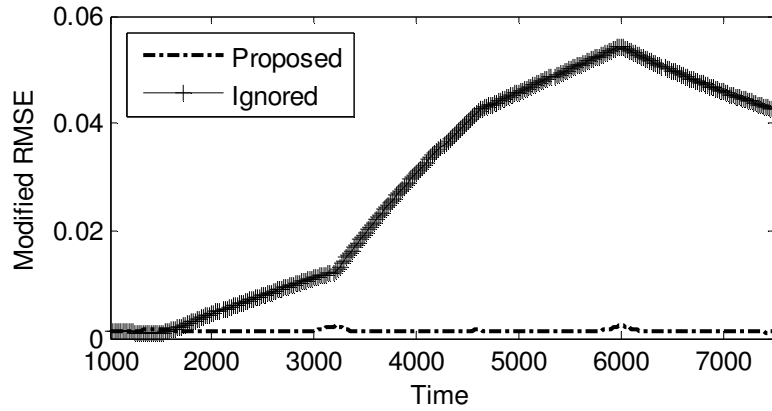


Fig. 6.4 Practical scheme based on 12-bus system

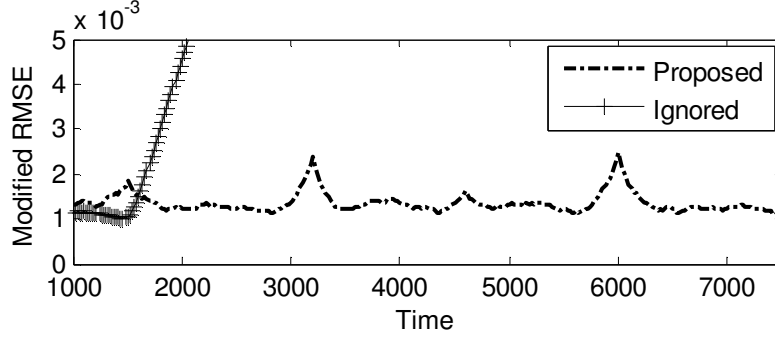


Fig. 6.5 Practical scheme based on 12-bus system (zoomed in)

6.8 Conclusions

Chapters 5-6 present an approach to jointly and dynamically estimate the transmission-line parameters and the bus voltages of interest. It contains two loosely-coupled procedures, namely state tracking and parameter tracking. An error-ensemble-evolution method is responsible for dealing with the coupling. Accurate models are studied separately. An adaptive filtering procedure has been developed to estimate the voltage state, accompanied by the detection and estimation of abrupt changes. Another adaptive filter including the adjustment of transition matrix has also been developed for the parameter tracking. Both filters are based on the generic (optimal) Kalman filter conditioned on prediction-measurement-error correlation. Simulations illustrate the performance of the whole approach under normal operation conditions and under the condition that abrupt state changes occur. The necessity of the error-ensemble evolution, and the accuracy and superiority of the proposed approaches have also been verified. The overall approach, as well as the two procedures, can be applied to other state and parameter estimation problems with traditional SCADA data.

6.9 Appendix:

Error Evolution Formulae

From the dynamic model for the parameter in (6.3), we get

$$\mathbf{p}_k = \mathbf{p}_{c,k} + \mathbf{w}_{p,k}; \quad \mathbf{p}_{c,k} = \mathbf{F}_{p,k-1} \mathbf{p}_{c,k-1} \quad (6.32)$$

According to (6.9), the prediction at time k is

$$\hat{\mathbf{p}}_{k|k-1} = \hat{\mathbf{p}}_{c,k|k-1} = \mathbf{F}_{p,k-1} \hat{\mathbf{p}}_{c,k-1|k-1}. \quad (6.33)$$

Then the prediction errors are

$$\begin{aligned} \Delta \mathbf{p}_{c,k|k-1} &= \mathbf{p}_{c,k} - \hat{\mathbf{p}}_{c,k|k-1} = \mathbf{F}_{p,k-1} (\mathbf{p}_{c,k-1} - \hat{\mathbf{p}}_{c,k-1|k-1}) \\ &= \mathbf{F}_{p,k-1} \Delta \mathbf{p}_{c,k-1|k-1}, \end{aligned} \quad (6.34)$$

$$\begin{aligned} \Delta \mathbf{p}_{k|k-1} &= \mathbf{p}_k - \hat{\mathbf{p}}_{k|k-1} \\ &= \mathbf{p}_{c,k} + \mathbf{w}_{p,k} - \hat{\mathbf{p}}_{c,k|k-1} = \Delta \mathbf{p}_{c,k|k-1} + \mathbf{w}_{p,k}. \end{aligned} \quad (6.35)$$

From (6.15) and (6.16), the updated parameters at time k is

$$\begin{cases} \hat{\mathbf{p}}_{k|k} = \hat{\mathbf{p}}_{k|k-1} + \tilde{\mathbf{K}}_k \Delta \mathbf{z}_{p,k} \\ \hat{\mathbf{p}}_{c,k|k} = \hat{\mathbf{p}}_{c,k|k-1} + \tilde{\mathbf{K}}_{c,k} \Delta \mathbf{z}_{p,k} \end{cases}. \quad (6.36)$$

Then the corresponding estimation errors are

$$\begin{aligned} \Delta \mathbf{p}_{k|k} &= \mathbf{p}_k - \hat{\mathbf{p}}_{k|k} = \mathbf{p}_k - \hat{\mathbf{p}}_{k|k-1} - \tilde{\mathbf{K}}_k \Delta \mathbf{z}_{p,k} = \mathbf{p}_k - \hat{\mathbf{p}}_{k|k-1} - \tilde{\mathbf{K}}_k (\mathbf{z}_{p,k} - \mathbf{H}_p^\circ \hat{\mathbf{p}}_{k|k-1} - \bar{\mathbf{v}}_{p,k}^\circ) \\ &= \Delta \mathbf{p}_{k|k-1} - \tilde{\mathbf{K}}_k (\mathbf{H}_p^\circ \mathbf{p}_k + \mathbf{v}_{p,k}^\circ - \mathbf{H}_p^\circ \hat{\mathbf{p}}_{k|k-1} - \bar{\mathbf{v}}_{p,k}^\circ) = \Delta \mathbf{p}_{k|k-1} - \tilde{\mathbf{K}}_k (\mathbf{H}_p^\circ \Delta \mathbf{p}_{k|k-1} + \mathbf{v}_{p,k}^\circ - \bar{\mathbf{v}}_{p,k}^\circ) \\ \Delta \mathbf{p}_{c,k|k} &= \Delta \mathbf{p}_{c,k|k-1} - \tilde{\mathbf{K}}_{c,k} (\mathbf{H}_p^\circ \Delta \hat{\mathbf{p}}_{k|k-1} + \mathbf{v}_{p,k}^\circ - \bar{\mathbf{v}}_{p,k}^\circ) \end{aligned}$$

In summary,

$$\begin{cases} \Delta \mathbf{p}_{c,k|k-1} = \mathbf{F}_{p,k-1} \Delta \mathbf{p}_{c,k-1|k-1} \\ \Delta \mathbf{p}_{k|k-1} = \Delta \mathbf{p}_{c,k|k-1} + \mathbf{w}_{p,k} \\ \Delta \mathbf{p}_{c,k|k} = \Delta \mathbf{p}_{c,k|k-1} - \tilde{\mathbf{K}}_{c,k} (\mathbf{H}_p^\circ \Delta \mathbf{p}_{k|k-1} + \mathbf{v}_{p,k}^\circ - \bar{\mathbf{v}}_{p,k}^\circ) \\ \Delta \mathbf{p}_{k|k} = \Delta \mathbf{p}_{k|k-1} - \tilde{\mathbf{K}}_k (\mathbf{H}_p^\circ \Delta \mathbf{p}_{k|k-1} + \mathbf{v}_{p,k}^\circ - \bar{\mathbf{v}}_{p,k}^\circ) \end{cases} \quad (6.37)$$

Similarly, for the bus-voltage state in Chapter 5, we can get

$$\begin{cases} \Delta \mathbf{x}_{c,k|k-1} = \Delta \mathbf{x}_{c,k-1|k-1} \\ \Delta \mathbf{x}_{k|k-1} = \Delta \mathbf{x}_{c,k|k-1} + \mathbf{w}_{x,k} \\ \Delta \mathbf{x}_{c,k|k} = \Delta \mathbf{x}_{c,k|k-1} - \mathbf{K}_{c,k} (\mathbf{H}_x^\circ \Delta \mathbf{x}_{k|k-1} + \mathbf{v}_{x,k}^\circ - \bar{\mathbf{v}}_{x,k}^\circ) \\ \Delta \mathbf{x}_{k|k} = \Delta \mathbf{x}_{k|k-1} - \mathbf{K}_k (\mathbf{H}_x^\circ \Delta \mathbf{x}_{k|k-1} + \mathbf{v}_{x,k}^\circ - \bar{\mathbf{v}}_{x,k}^\circ) \end{cases} \quad (6.38)$$

Chapter 7: Summary and Future Work

The least-squares (LS) principle, including the weighted least-squares (WLS), is widely introduced to various scientific and technological fields. A great many methods have been developed to solve the fundamental and classic LS problem, among which the matrix-inversion-lemma based recursive least-squares (RLS) is a milestone. The RLS is of recursive form and free of matrix inversion, and thus has excellent performance regarding the efficient real-time computation and low memory storage. We generalize the RLS procedure and to solve the unconstrained/ LE-constrained generalized LS (GLS) problem in a similar recursive way. We also apply the RLS method for all the involved initializations. The newly-developed methods are integrated as completely-recursive LS (CRLS).

Correspondingly, in Chapter 2, the generalization of the RLS for solving GLS problems is discussed. Concretely, starting from the unconstrained/LE-constrained RLS, we develop recursive procedures applicable to the unconstrained/LE-constrained GLS, and show that the LE constraint is in essence a set of special observations free of observation errors and can be processed sequentially in any place in the data sequence. More generally, we also treat the ILE-constrained GLS. A unified recursive procedure is developed, which is applicable to ILE-constrained GLS as well as all the unconstrained/LE-constrained LS/WLS/GLS.

In Chapter 3, a recursive exact initialization applicable to all the RLS, RWLS and RGLS, is investigated. This chapter treats the RLS initialization-related issues, including rank check, a convenient method to compute the involved matrix inverse/pseudoinverse, and resolution of underdetermined systems. No extra non-RLS formula but an auxiliary-observation based procedure is utilized. The RLS recursion can start from the first real observation and possible LE constraints are also imposed recursively; the rank of the system is checked implicitly. If the rank is full, the initialization and the subsequent RLS cycles can be integrated as a whole to yield exact LS solutions. If the rank is deficient, the procedure provides a mapping from the unobservable (original) estimand to a reduced-dimensional set of alternative variables which

are linear combinations of the original variables and uniquely determined. The consequent estimate is a set of refined non-redundant observations. The refinement is lossless in the WLS sense: if new observations are available later, it can take the role of the original data in the recalculation.

In summary, the CRLS approach has the following good properties: The proposed method can distribute the processing time (including the initialization) over the data-accumulation period; The CRLS has a low computational complexity; With the CRLS, the initialization of LE-constrained RLS solution, which (in the batch form) usually involves MP inverses, is made as simple as for the unconstrained ones now; In sparse applications, the CRLS can benefit more from the sparsity because its recursion can make full use of the sparse structure of the observation coefficients; The observability analysis in the CRLS requires no extra computation. The result by the CRLS is numerically consistent with the existence of \mathbf{C} in calculation.

In Chapter 4, we demonstrate the mathematical equivalence between the linear-data-model based linear minimum-mean-square-error (LMMSE) estimator and the ILE-constrained GLS. We also suggest to use the recursive ILE-constrained GLS to improve the sequential procedure of the optimal KF considering prediction-measurement-error correlation.

In Chapters 5 & 6, we perform accurate parameter (and state) estimation in complex situations using synchrophasor data, based on the optimal KF considering the correlation between the measurement noise and the prediction error. An approach of joint state-and-parameter estimation, which is different from the state augmentation, is adopted, where the original nonlinear PE problem is reformulated as two loosely-coupled linear subproblems: state tracking and parameter tracking, respectively.

An error-ensemble-evolution method is responsible for dealing with the coupling between the state tracking and the parameter tracking. Accurate models are studied separately. An adaptive filtering procedure has been developed to estimate the voltage state, accompanied by the detection and estimation of abrupt changes. Another adaptive filter including the

adjustment of transition matrix has also been developed for the parameter tracking. Both filters are based on the optimal KF conditioned on prediction-measurement-error correlation. Simulations illustrate the performance of the whole approach under normal operation conditions and under the condition that abrupt state changes occur. The necessity of the error-ensemble evolution, and the accuracy and superiority of the proposed approaches have also been verified. The overall approach, as well as the two procedures, can be applied to other state and parameter estimation problems with traditional SCADA data.

As declared in this dissertation, we have great interest in applying the newly-developed CRLS approach to solve practical applications. For instance, the joint-state-and-parameter estimation in power system based on synchrophasors is an application of the optimal KF considering prediction-measurement-error correlation, where the filter can be verified and the corresponding sequential procedure can be improved by the CRLS. In the future, we also aim to utilizing the proposed recursive RLS initialization technique to solve high-dimensional and low-redundancy practical problems. For instance, the application to power system state estimation with synchrophasors is quite attractive.

References

- [001] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*, Marcel Dekker, NY, 2004
- [002] N. U. Ahmed, *Linear and Nonlinear Filtering for Scientists and Engineers*, World Scientific, Singapore, 1998
- [003] A. Albert and R. W. Sittler, "A method for computing least squares estimators that keep up with the data," *SIAM Journal on Control and Optimization*, vol. 3, no. 3, pp. 384-417, 1965
- [004] O. Alsac, N. Vempati, B. Stott, and A. Monticelli, "Generalized state estimation," *IEEE Trans. Power Syst.*, vol. 13, no.3, pp. 1069-1075, Aug. 1998
- [005] T. Amemiya, *Advanced Econometrics*, Cambridge, Harvard University Press, MA, 1985
- [006] A. A. Anda and H. Park, "Fast plane rotations with dynamic scaling", *SIAM J. Matrix Anal. Appl.*, vol.15, pp. 162-174, 1994
- [007] A. A. Anda and H. Park, "Self-scaling fast rotations for stiff least squares problems", *Linear Algebra Appl.* , vol. 234, pp. 137-162, 1996
- [008] E. Anderson, Z. Bai, and J. Dongarra, "Generalized QR factorization and its applications," *Tech. Report CS-91-131*, Computer Science Department, University of Tennessee, Knoxville, TN, 1991
- [009] B. S. Andersen, J. A. Gunnels, F. G. Gustavson, J. Wasniewski, "A recursive formulation of the inversion of symmetric positive definite matrices in packed storage data format," in *Proc. 2002 (6th) Intern. Conf. Applied Parallel Computing Advanced Scientific Computing*, PARA 2002, pp. 287-296, 2002
- [010] B. S. Andersen, J. Wasniewski, F. G. Gustavson, "A recursive formulation of Cholesky factorization of a matrix in packed storage," *ACM Trans. Math. Softw.* vol. 27, no. 2, pp. 214-244, 2001
- [011] J. A. Apolinario, *QRD-RLS Adaptive Filtering*, 1st Edition, Springer, NY, 2009
- [012] J. L. Barlow and S. L. Handy, "The direct solution of weighted and equality constrained least squares problems," *SIAM J. Sci. Statist. Comput.*, vol. 9, pp. 704-716, 1988
- [013] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, John Wiley and Sons, NY, 2001
- [014] T. S. Bi, X. H. Qin, and Q. X. Yang, "A novel hybrid state estimator for including synchronized phasor measurements," *Electric Power Systems Research*, vol. 78, no. 8, pp. 1343-1352, Aug. 2008
- [015] X. Bian, X. R. Li, H. Chen and D. Gan, "Joint estimation of state and parameter with synchrophasors – part I: state tracking," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1196-1208, 2011a
- [016] X. Bian, X. R. Li, H. Chen and D. Gan, "Joint estimation of state and parameter with synchrophasors – part II: parameter tracking," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1209-1220, 2011b
- [017] A. Bjorck and G. H. Golub, "Iterative refinement of linear least squares solution by Householder transformation," *BIT*, vol. 7, pp. 322-337, 1967
- [018] A Bjorck and I. S. Duff, "A direct method for the solution of sparse linear least squares problems," *Linear Algebra Appl.*, vol. 34, pp. 43-67, 1980
- [019] A. Bjorck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996
- [020] O. Bretscher, *Linear Algebra With Applications, 3rd Edition*. Prentice Hall, Upper Saddle River, NJ, 1995

- [021] J. R. Bunch and C. P. Nielsen, "Updating the singular value decomposition," *Numer. Math.*, vol. 31, pp.111-129, 1978
- [022] T. Chan, "Rank revealing QR-factorizations," *Linear Algebra Applications*, vol. 88/89, pp. 67-82, 1987
- [023] T. Chan, "Low-rank revealing QR-factorizations," *Numerical Linear Algebra Applications*, vol. 1, pp. 33-44, 1994
- [024] H. F. Chen, *Recursive estimation and control for stochastic systems*, John Wiley and Sons, NY, 1985
- [025] J. Chen and A. Abur, "Placement of PMUs to enable bad data detection in state estimation," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1608-1615, May 2006
- [026] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization (3rd edition)*, John Wiley & Sons, NY, 2008
- [027] J. M. Cioffi and T. Kailath, "Fast, Recursive-least-squares Transversal Filters for Adaptive filtering," *IEEE Trans. Acoustic, Speech Signal Processing*, vol. ASSP-32, pp. 304-337, 1984
- [028] J. M. Cioffi, "The fast adaptive ROTOR's RLS algorithm," *IEEE Trans. Acoustic, Speech Signal Processing*, vol. ASSP-38, pp. 631-653, 1990
- [029] J. M. Cioffi, "The fast Householder RLS Adaptive Filter," in *Proc. ICASSP*, Albuquerque, NM. Apr. 1990
- [030] A. J. Conejo, S. de la Torre, and M. Canas, "An optimization approach to multiarea state estimation", *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 213-221, Feb 2007
- [031] A. S. Debs and R. E. Larson, "A dynamic estimator for tracking the state of a power system," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-89, no. 7, pp. 1670-1678, Oct. 1970
- [032] A. S. Debs, "Estimation of steady-state power system model parameters," *IEEE Trans. Power Apparatus and Syst.*, vol. PAS 93, pp. 1260-1268, 1974
- [033] J. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart, "Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization," *Math. Comp.*, vol. 30, pp. 772-795, 1976
- [034] B. De Moor and P. Van Dooren, "Generalizations of the singular value and QR decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 13, pp.993-1014, 1992
- [035] P. M. Djuric, J. H. Kotecha, and J. Zhang, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19-38, Sep. 2003
- [036] J. Dongarra, J. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users' Guide*, SIAM, Philadelphia, 1979
- [037] I. S. Duff, "The solution of augmented systems," in *Numerical Analysis 1993: Proceedings of the 15th Dundee Conference*, June 1993; D. F. Griffiths and G. A. Watson, Pitman Research Notes in Mathematics, Longman Scientific and Technical, Harlow, UK, 1994
- [038] R. Ebrahimian and R. Baldick, "State estimation distributed processing", *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp. 1240-1246, Nov. 2000
- [039] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press, 2005
- [040] W. M. Gentleman, "Least squares computations by Givens transformations without squares roots," *J. Inst. Maths. Applic.*, vol.12, pp. 329-336, 1973
- [041] J. A. George and J. W.-H. Liu, *Computer Solution of Large Scale Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1981
- [042] P. E. Gill, G. H. Golub, W. Murry, and M. Saunders, "Methods for modifying matrix factorizations," *Math. Comp.*, vol. 28, pp. 505-535, 1974
- [043] W. Givens, "Computation of plane unitary rotations transforming a general matrix to triangular form," *SIAM J. Appl. Math.*, vol. 6, pp. 26-50, 1958

- [044] G. H. Golub, "Numerical methods for solving least squares problems," *Numer. Math.*, vol. 7, pp. 206-216, 1965
- [045] G. H. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *SIAM J. Numer. Anal. Ser. B*, vol. 2, pp. 205-224, 1965
- [046] G. H. Golub, "Least squares, singular values and matrix approximations," *Aplikace Matematiky*, vol. 13, pp. 44-51, 1968
- [047] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solution," *Numer. Math.*, vol. 14, pp. 403-420, 1970
- [048] M. Gulliksson and P. -A. Wedin, "Modifying the QR decomposition to constrained and weighted linear least squares," *SIAM J. Matrix. Anal. Appl.* vol. 13, no. 4, pp. 1298-1313, 1992
- [049] M. Gulliksson, "On modified Gram-Schmidt for weighted and constrained linear least squares," *BIT*, vol. 35, pp. 458-473, 1995
- [050] G. Golub, C. Van Loan, *Matrix Computations*, John Hopkins, 1996
- [051] B. Gou and A. Abur, "An improved measurement placement algorithm for network observability", *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 819-824, Nov. 2001
- [052] W. H. Greene. *Econometric Analysis*, 4th Edition, Macmillan, NY, 2000
- [053] F. G. Gustavson and I. Jonsson, "Minimal-storage high-performance Cholesky factorization via recursion and blocking," *IBM J. Res. Develop.*, vol. 44, no. 6, pp. 823-850, 2000
- [054] S. Hammarling, "A note on modifications to the Givens plane rotation," *J. Inst. Maths. Applic.* vol. 13, pp. 215-218, 1974
- [055] G. T. Haupt, N. J. Kasdin, G. M. Keiser and B. W. Parkinson, "Optimal recursive iterative algorithm for discrete nonlinear least-squares estimation", *AIAA Journal of Guidance, Control, and Dynamics*, vol. 19, no. 3, pp.643-649, 1996
- [056] S. Haykin, *Adaptive Filter Theory*, 4th Edition, Prentice-Hall, Englewood Cliffs, NJ, 2001
- [057] M. R. Hestenes, "Inversion of matrices by biorthogonalization and related results," *J. Soc. Indust. Appl. Math.*, vol. 6, pp. 51-90, 1958
- [058] N. J. Higham, "Newton's method for the Matrix Square Root," *Mathematics of Computation*, vol. 46, 1986
- [059] R. A. Horn, C. R. Johnson, *Matrix analysis*, Cambridge University Press, NY, 1985
- [060] P. D. Hough and S. A. Vavasis, "Complete orthogonal decomposition for weighted least squares," *Tech. Report CTC93-TR203*, Advanced Computing Research Institute, Cornell University, Ithaca, NY, 1994
- [061] A. S. Householder, "Unitary triangularization of a nonsymmetric matrix," *J. Assoc. Comput. Mach.*, vol. 5, pp. 339-342, 1958
- [062] S.-J. Huang and J.-M. Lin, "Enhancement of anomalous data mining in power system predicting-aided state estimation," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 610-619, Feb 2004
- [063] N. Hubing and S. Alexander, "Statistical analysis of initialization methods for RLS adaptive filters," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1793-1804, 1991
- [064] *IEEE Standard for Synchrophasors for Power Systems*, IEEE Std. C37. 118-2005
- [065] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Automatic Control*, vol. 45, no. 3, pp. 910-927, May 2000
- [066] W. Jiang, V. Vittal, and G. T. Heydt, "A distributed state estimator utilizing synchronized phasor measurments," *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 563-570, May 2007

- [067] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no.3, pp. 401-422, 2004
- [068] D. Johnson, "Minimum mean squared error estimators", *Connexions*, Nov. 2004
- [069] L. Karlsson, *Computing Explicit Matrix Inverses by recursion*, Master's Thesis in Computing Science, Department of Computer Science, Umea University, Sweden, Feb. 2006
- [070] D. G. Kleinbaum, L. L. Kupper, A. Nizam, and K. E. Muller, *Applied Regression Analysis and Multivariable Methods, 4th Edition*, Duxbury Press, Belmont, CA, 2007
- [071] E. G. Kogbetliantz, "Solution of linear equations by diagonalization of coefficients matrix," *Quart. Appl. Math.*, vol. 13, pp. 123-132, 1955
- [072] G. L. Kusic and D. L. Garrison, "Measurement of transmission line parameters from SCADA data," in *Power Systems Conf. and Exposition 2004, IEEE/PES*, vol.1, pp. 440-445
- [073] C. L. Lawson, R. J. Hanson, D. R. Kincaid, and F. T. Krogh, "Basic linear algebra subprograms for Fortran usage," *ACM Trans. Math. Software*, vol. 5, pp. 308-323, 1979
- [074] C. L. Lawson, R. J. Hanson, *Solving Least Squares Problems*, SIAM, Philadelphia, PA, 1995
- [075] A. M. Leite da Silva, M. B. Co Coutto Filho, and J. M. C. Cantera, "An efficient dynamic state estimation algorithm including bad data processing," *IEEE Trans. Power Syst.*, vol. 2, no. 4, pp. 1050-1058, Nov. 1987
- [076] A. M. Leite da Silva, M. B. Co Coutto Filho, and J. F. de Queiroz, "State forecasting in electric power systems," *IEEE Proceedings*, vol. 130, no. 5, pp. 237-244, Sep. 1983
- [077] O. Leringe and P.-A. Wedin, "A comparison between different methods to compute a vector x which minimizes $\|Ax-b\|_2$ when $Gx=h$," *Tech. Report*, Department of Computer Science, Lund University, Lund, Sweden, 1970
- [078] L. Ljung, M. Morf, and D. Falconer, "Fast calculation of Gain Matrices for recursive estimation schemes," *Int. J. Control*, vol. 27, pp. 1-19, 1978
- [079] X. R. Li, *Fundamentals of Applied Estimation and Filtering*, Course notes, University of New Orleans, 2007, Ch. 15
- [080] X. R. Li, Y. Zhu, J. Wang, and C. Han, "Optimal linear estimation fusion—part I: unified fusion rules," *IEEE Trans. Information Theory*, vol. 49, no. 9, pp. 2192-2208, Sep. 2003
- [081] X. R. Li and V. P. Jilkov, "A survey of maneuvering target tracking: approximation techniques for nonlinear filtering," in *Proc. 2004 SPIE Conf. Signal and Data Processing of Small Targets*, vol. 5428, pp. 537-550, 2004
- [082] J.-M. Lin, S.-J. Huang, and K.-R. Shih, "Application of sliding surface-enhanced fuzzy control for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 570-577, May 2003
- [083] J.-M. Lin, S.-J. Huang, and K.-R. Shih, "Application of sliding surface-enhanced fuzzy control for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 570-577, May 2003
- [084] W.-H. E. Liu, F. F. Wu, and S.-M. Lun, "Estimation of parameter errors from measurement residuals in state estimation," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 81-89, Feb. 1992
- [085] W.-H. E. Liu and S.-L. Lim, "Parameter error identification and estimation in power system state estimation," *IEEE Trans. Power Syst.*, vol. 10, no.1, pp. 200-209, Feb. 1995
- [086] J. K. Mandal, A. K. Sinha, and L. Roy, "Incorporating nonlinearities of measurement function in power system dynamic state estimation," *IEE Proc. Gener. Transm. Distrib.*, vol. 142, no. 3, pp. 289-296, May 1995

- [087] T. A. Manteuffel, "An interval analysis approach to rank determination in linear least squares problems," *SIAM J. Sci. Statist. Comp.*, vol. 2, pp. 335-348, 1981
- [088] R. S. Martin and J. H. Wilkinson, "Handbook series linear algebra: symmetric decomposition of positive definite band matrices," *Numerical Mathematci*, vol. 7, no. 5, pp.355-361, 1965
- [089] K. E. Martin, D. Hamai, M. G. Adamiak, et al, "Exploring the IEEE standard C37.118-2005 synchrophasors for power systems," *IEEE Trans. Power Delivery*, vol. 23, no. 4, pp. 1805-1811, Oct. 2008
- [090] A. Meliopoulos, B. Fardanesh, S. Zelingher, "Power system state estimation: modeling error effects and impact on system operation", HICSS '01 Proceedings of the 34th annual Hawaii international conference on system sciences, vol. 2, pp. 2012, 2001; www.hicss.hawaii.edu/hicss_34/pdfs/cssar06.pdf
- [091] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, PA, 2000
- [092] J. Mikles and M. Fikar, *Process modeling, identification, and control II: identification and optimal control*, Springer-Verlag Berlin Heidelberg, NY, 2007
- [093] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no.2, Feb. 2000
- [094] B. Nobel, "Methods for computing the Moore-Penrose generalized inverse and related matters," in *Generalized Inverses and Applications, Proceedings of an Advanced Seminar*, the University of Wisconsin-Madison, 1973
- [095] T. Ogita, "Accurate matrix factorization: Inverse LU and inverse QR factorizations," *SIAM J. Matrix Anal. Appl.*, vol.31 (5): pp. 2477-2497, 2010
- [096] T. Ogita and S. Oishi, "Accurate and robust inverse Cholesky factorization," *Nonlinear Theory and Its Applications, IEICE*, vol.3, no.1, pp.103-111, 2012
- [097] C. C. Paige and M. A. Saunders, "Toward a generalized singular value decomposition," *SIAM J. Numer. Anal.* vol.18, pp. 398-405, 1981
- [098] C. C. Paige, "Some aspects of generalized QR factorizations," in *Reliable Numerical Computation*, Clarendon Press, Oxford, UK, pp. 71-91, 1990
- [099] B. N. Parlett, "The QR algorithm," *Computing Sci. Eng.*, no. 2, pp. 38-42, 2000
- [100] K. M. Passino and S. Yurkovich, *Fuzzy control*, Addison-Wesley, Menlo Park, California, 1998
- [101] A. G. Phadke, "Synchronized phasor measurements in power system," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10-15, Apr. 1993
- [102] M. D. Petkovic, P. S. Stanimirovic, "Generalized inversion is not harder than matrix multiplication," *Journal of Computational and Applied Mathematics*, v.230 (1), 270-282, 2009
- [103] A. G. Phadke, "Synchronized phasor measurements ~ a historical overview," in *Transmission and Distribution Conf. and Exhibition 2002: Asia Pacific, IEEE/PES*, vol. 1, pp. 476-479
- [104] A. G. Phadke, J. S. Thorp, and K. J. Karimi, "State estimation with phasor measurements," *IEEE Trans. Power Syst.*, vol. 1, no. 1, pp. 233-240, Feb. 1986
- [105] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*, Springer Science + Business Media, NY, 2008
- [106] A. G. Phadke, "Synchronized phasor measurements ~ a historical overview," in *Transmission and Distribution Conf. and Exhibition 2002: Asia Pacific, IEEE/PES*, vol. 1, pp. 476-479, 2002
- [107] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes: The art of scientific computing (3rd edition)*, Cambridge University Press, NY, 2007
- [108] I. K. Proudler, J. G. McWhirter, and T. J. Shepard, "Fast QRD-based Algorithms for least-squares linear

- prediction,” in *IMA Conf. Proc.*, Warwick, England, 1988
- [109] A. P. Sakis Meliopoulos, G. J. Cokkinides, and F. Galvan, “Advances in the supercalibrator concept-practical implementations,” in *Proc. 2007 40th Hawaii Intern. Conf. Systems Sciences*, pp. 01-10, 2007
- [110] E. Schmidt, “Über die Auflösung linearer Gleichungen mit unendlich vielen Unbekannten,” *Rend. Circ. Mat. Palermo. Ser. I*, vol. 25, pp. 53-77, 1908
- [111] K.-R. Shih and S.-J. Huang, “Application of a robust algorithm for dynamic state estimation of a power system,” *IEEE Trans. Power Syst.*, vol. 17, no. 1, pp. 141-147, Feb 2002
- [112] A. K. Sinha and J. K. Mondal, “Dynamic state estimator using ANN based bus load prediction,” *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1219-1225, Nov. 1999
- [113] I. W. Slutsker, S. Mokhtari, and K. A. Clements, “Real time recursive parameter estimation in energy management system,” *IEEE Trans. Power Syst.*, vol. 11, no.3, pp. 1393-1399, Aug. 1996
- [114] H. W. Sorenson, *Parameter Estimation*, Marcel Dekker, NY, 1980
- [115] G. W. Stewart, “Rank degeneracy,” *SIAM J. Sci. Statist. Comput.*, vol.5, pp. 403-413, 1984
- [116] S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*, Cambridge, Belknap Press of Harvard University Press, MA, 1986
- [117] C. F. Van Loan, “Generalizing the singular value decomposition,” *SIAM J. Numer. Anal.*, vol. 13, pp. 76-83, 1976
- [118] C. F. Van Loan, “On the method of weighting for equality constrained least squares,” *SIAM J. Numer. Anal.*, vol. 22, pp. 851-864, 1985
- [119] S. A. Vavasis, “Stable numerical algorithms for equilibrium systems,” *SIAM J. Matrix Anal. Appl.*, vol. 15, pp. 1108-1131, 1994
- [120] D. Watkins, *Fundamentals of Matrix Computations*, Wiley, NY, 1991
- [121] M. Wei, “Algebraic properties of rank-deficient equality constrained and weighted least squares problem,” *Linear Algebra Appl.*, 161, pp. 27-43, 1992
- [122] M. Wei, “Perturbation theory for the rank-deficient equality constrained and weighted least squares problem,” *SIAM J. Numer. Anal.*, vol. 29, pp. 1462-1481, 1992
- [123] Wiki01: http://en.wikipedia.org/wiki/Least_squares
- [124] Wiki02: http://en.wikipedia.org/wiki/Generalized_least_squares
- [125] M. Woodbury, “Inverting Modified Matrices”, *Memorandum Rept. 42*, Statistical Research Group, Princeton University, NJ, 1950
- [126] B. Yang and J. F. Bohme, “Rotation-based RLS algorithms: Unified derivations, numerical properties, and parallel implementations,” *IEEE Trans. Signal Processing*, vol. 40, pp. 1151-1167, May 1992
- [127] P. Zarco and A. G. Exposito, “Power system parameter estimation: a survey,” *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 216-222, Feb. 2000
- [128] L. Zhao and A. Abur, “Multiarea state estimation using synchronized phasor measurements,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611-617, May 2005
- [129] J. Zhou, Y. Zhu, X. R. Li and Z. You, “Variants of the Greville formula with applications to exact recursive least squares,” *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 1, pp. 150-164, 2002
- [130] M. Zhou, V. A. Centeno, J. S. Thorp and A. G. Phadke, “An alternative for including phasor measurements in state estimators”, *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930-1937, Nov. 2006

- [131] Y. Zhu and X. R. Li, "Recursive least squares with linear constraints," *Communication in Information and Systems*, vol. 7, no.3, pp. 287-312, 2007

Vita

The author was born in Yangzhou, Jiangsu, China. He obtained his Bachelor's degree of engineering from Zhejiang University, China in 2000. He joined the University of New Orleans graduate program to pursue a PhD in Engineering and Applied Science, Electrical Engineering, and became a member of Professor X. Rong Li's research group in 2006.