

Doctoral Thesis

Probabilistic Data Reconciliation in Material Flow Analysis

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the TU Wien, Faculty of Civil Engineering

Dissertation

Probabilistischer Datenausgleich in der Materialflussanalyse

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaft
eingereicht an der Technischen Universität Wien, Fakultät für Bauingenieurwesen

von

Dipl.-Ing. **Oliver Cencic**

Matr.Nr.: 09140006

Petronell-C., Austria

Supervisor: Univ.-Prof. Dipl.-Ing. Dr.techn. **Helmut Rechberger**
Institute for Water Quality and Resource Management
Technische Universität Wien, Austria

Supervisor: Univ.-Doz. Dipl.-Ing. Dr. **Rudolf Frühwirth**
Institute of High Energy Physics (HEPHY)
Austrian Academy of Sciences, Austria

Assessor: Associate Prof. **Reinout Heijungs**
Department of Econometrics and Operations Research
Vrije Universiteit Amsterdam, The Netherlands

Assessor: Prof. **Julian M. Allwood**
Department of Engineering
University of Cambridge, United Kingdom

Vienna, February 2018

Acknowledgement

First of all, I would like to express my deepest gratitude to Dr. Rudolf Frühwirth for his patience and his unlimited support during the work on this theses. I learned so much about statistics and programming in MATLAB through all the discussion we had. It was motivating, inspiring, and a great honor to work with him. I also want to thank Prof. Helmut Rechberger for supervising my work and giving me the freedom to work on mathematical methodologies deviating from the actual field of research at the department. I would also like to express my appreciation to Prof. Reinout Heijungs and Prof. Julian Allwood for their evaluation of my thesis. Special thanks to my colleagues Nada Dzubur, Oliver Schwab, David Laner and Manuel Hahn for giving me the chance to internally discuss uncertainty matters after many years of solitary work on this topic. Thanks also to Alfred Kovacs for his marvelous work in programming the free software STAN. Finally, I want to thank my parents, my friends and my family for all their encouragement and support.

Abstract

Material Flow Analysis (MFA) is a tool that helps to model and quantify the flows and stocks of a system of interest. Due to unavoidable measurement or estimation errors, the observed values of flows and stocks are in conflict with known constraints such as the law of mass conservation. The basic idea of data reconciliation is to resolve these contradictions by statistically adjusting the collected data based on the assumption that their uncertainty is described by a probability density function (pdf).

Most solving techniques that have been developed over the last 60 years are based on a weighted least-squares minimization of the measurement adjustments subject to constraints involving observed variables, unknown variables and fixed quantities. The underlying main assumption of this approach is that of normally distributed (Gaussian) observation errors, with zero mean and known covariance matrix. In STAN, a freely available software that supports MFA and allows to consider data uncertainties, this approach has been implemented. Paper 1 of this cumulative doctoral thesis covers the mathematical foundation of the nonlinear data reconciliation algorithm incorporated in STAN and demonstrates its use on a hypothetical example from MFA.

In scientific models in general and in MFA models in particular, however, data is often not normally distributed. Thus, a different approach to data reconciliation, based on Bayesian reasoning, was developed within the scope of this thesis that can deal with arbitrary continuous probability distributions. Its main idea is to restrict the joint prior probability distribution of the observed variables with model constraints to get a joint posterior probability distribution. Because in general the posterior probability density function cannot be calculated analytically, it is shown that it has decisive advantages to sample from the posterior distribution by a Markov chain Monte Carlo (MCMC) method. From the resulting sample, the joint

pdf of observed and unobserved variables and its moments can be estimated, along with the marginal posterior densities, moments, quantiles, and other characteristics. Paper 2 covers the case of linear constraints while paper 3 deals with nonlinear constraints. In both papers, the method is illustrated by examples from MFA and chemical engineering.

Finally, the summary of this thesis contains two additional topics for the Bayesian approach, which haven't been covered by the papers 2 and 3: it is shown how to use copulas to implement correlated observations, and how to use M-estimators to get a reconciliation procedure that is robust against outlying observations and does not require any prior assumptions on the distribution of the outliers.

Kurzfassung

Die Materialflussanalyse (MFA) ist ein Werkzeug, das dabei hilft, die Flüsse und Lager eines zu untersuchenden Systems zu modellieren und zu quantifizieren. Auf Grund unvermeidlicher Mess- und Schätzfehler sind die erhobenen Daten im Widerspruch mit bekannten Zwangsbedingungen wie zum Beispiel dem Massenerhaltungsgesetz. Die grundlegende Idee des Datenausgleichs ist es, diese Widersprüche aufzulösen, indem die gesammelten Daten statistisch angepasst werden. Dabei wird angenommen, dass deren Unsicherheit durch eine Wahrscheinlichkeitsdichtefunktion beschrieben werden kann. Die meisten Lösungsverfahren, die in den letzten 60 Jahren entwickelt wurden, basieren auf einer Minimierung der gewichteten Quadrate der notwendigen Beobachtungsanpassungen (Methode der kleinsten Fehlerquadrate), bei der die zu erfüllenden Zwangsbedingungen beobachtete Variablen, unbekannte Variablen und fixe Größen enthalten können. Die zugrundeliegende Hauptannahme dieses Ansatzes ist, dass die Fehler der Beobachtungen normalverteilt sind, mit Mittelwert Null und bekannter Kovarianzmatrix. Dieser Ansatz wurde auch in STAN verwendet, einer frei erhältlichen Software für MFA, die die Berücksichtigung von Datenunsicherheiten unterstützt. Artikel 1 dieser kumulativen Dissertation behandelt die mathematischen Grundlagen des nichtlinearen Ausgleichsalgorithmus, der in STAN implementiert wurde und demonstriert seine Anwendung an einem hypothetischen Beispiel aus der MFA.

In wissenschaftlichen Modellen im allgemeinen, und in MFA-Modellen im speziellen, sind die verwendeten Daten jedoch oft nicht normalverteilt. Deshalb wurde im Rahmen dieser Doktorarbeit ein alternativer Zugang zum Datenausgleich entwickelt, der auf bayesschen Schlussfolgerungen basiert und mit beliebigen stetigen Wahrscheinlichkeitsverteilungen umgehen kann. Die Hauptidee dieses Ansatzes ist, die gemeinsame a-priori Wahrscheinlichkeitsverteilung der beobachteten Größen mit den Modellgleichungen einzuschränken, um die gemeinsame a-posteriori

Wahrscheinlichkeitsverteilung zu erhalten. Da im allgemeinen die a-posteriori Verteilung nicht analytisch berechnet werden kann, wird gezeigt, dass es erhebliche Vorteile bringt, die a-posteriori Verteilung mittels eines Markov-Ketten-Monte-Carlo-Verfahrens (MCMC) zu beproben. Aus der resultierende Stichprobe können die gemeinsame Wahrscheinlichkeitsverteilung, sowie die a-posteriori Randverteilungen, Momente, Quantile und andere Charakteristika der beobachteten und unbekannten Variablen berechnet werden. Artikel 2 deckt den Fall der linearen Randbedingungen ab, während sich Artikel 3 mit nicht linearen Zwangsbedingungen beschäftigt. In beiden Artikeln werden Beispiele aus der MFA und der chemischen Literatur verwendet, um die Anwendung der entwickelten Methode zu demonstrieren.

Zusätzlich enthält die Rahmenschrift dieser Doktorarbeit zwei Erweiterungen für den bayesschen Ansatz, die in den Artikeln 2 und 3 nicht behandelt wurden: (1) die Verwendung von Copulas für die Implementierung von korrelierten Beobachtungen und (2) die Verwendung von M-Schätzern, um eine Ausgleichsprozedur zu erhalten, die robust gegen Ausreißer ist und keine Annahmen über die Verteilung der Ausreißer benötigt.

Published articles and author's contribution

This thesis brings together the results of more than 10 years of research and builds upon three journal articles (see appendix):

Paper 1

Nonlinear data reconciliation in material flow analysis with software STAN

Oliver Cencic

Sustainable Environment Research, 2016, 26 (6)

DOI: 10.1016/j.serj.2016.06.00

Paper 2

A general framework for data reconciliation - Part I: Linear constraints

Oliver Cencic and Rudolf Frühwirth

Computers and Chemical Engineering, 2015, 75

DOI: 10.1016/j.compchemeng.2014.12.004

Paper 3

Data reconciliation of nonnormal observations with nonlinear constraints

Oliver Cencic and Rudolf Frühwirth

Journal of Applied Statistics, 2018, in press

DOI: 10.1080/02664763.2017.1421916

Paper 1 was completely written by myself. In the papers 2 and 3, I primarily contributed to the problem definition, development of the methodology, example preparation, interpretation of the results and implementation of the algorithms in MATLAB. The derivation of the mathematical/statistical proofs and the implementation of the examples in MATLAB were done by Rudolf Frühwirth.

Contents

1	Introduction	1
2	Methodology	3
2.1	Error Model	3
2.2	Weighted Least Squares Approach to DR	3
2.2.1	Linear Constraints	4
2.2.2	Nonlinear Constraints	8
2.2.3	Correlated Observations	10
2.2.4	Gross Error Detection	11
2.3	Bayesian Approach to DR	11
2.3.1	Linear Constraints	17
2.3.2	Nonlinear Constraints	18
2.3.3	Correlated Observations	24
2.3.4	Robust Reconciliation and Gross Error Detection	27
3	Conclusions and Outlook	35
	Bibliography	37
	Notation	41
	Abbreviations	47
	Appendix	49

1 Introduction

Collecting data is an important part of each modeling procedure. Due to the fact that information often originates from different sources, collected data is unavoidably of varying quality. If only the point estimators of observations are considered, known constraints such as the conservation laws of mass and energy are frequently violated. Considering also the uncertainties of these point estimators, data reconciliation (DR) can be applied to statistically adjust contradicting observations by using redundant information.

Since the first publication on DR in the context of process optimization (Kuehn and Davidson, 1961), a variety of techniques has been developed to deal with these problems. For a comprehensive review see Narasimhan and Jordache (2000); Romagnoli and Sanchez (2000); Bagajewicz (2010). Most of the proposed methods are based on a weighted least squares minimization of the measurement adjustments subject to constraints involving observed (measured or estimated) variables, unobserved variables and fixed quantities. This classical approach, based on the assumption that the observation errors are normally distributed with zero mean and known variance, has also been implemented in STAN, a freely available software that supports MFA and enables the consideration of uncertain data under nonlinear constraints (Cencic and Rechberger, 2008). The calculation algorithm of STAN allows to make use of redundant information to reconcile uncertain “conflicting” data (with DR) and subsequently to compute unknown variables including their uncertainties (with error propagation) (Cencic, 2016). For more detailed information about the software, visit the website www.stan2web.net.

In scientific models in general and in MFA models in particular, however, data is often known to be not normally distributed. If, for instance, a process model is correct (i.e. there are no model uncertainties), mass flows cannot take negative values, and mass fractions and transfer coefficients¹ are restricted to the unit in-

terval. Another example is provided by expert opinions that frequently have to be relied on in MFA due to missing data. These opinions are often modeled by uniform, triangular or trapezoidal distributions, depending on the expert's knowledge about the quantity under consideration. And finally, if a sufficient number of measurements of the quantity is available, either a parametric model can be fitted, or a nonparametric model such as the empirical distribution function or the kernel estimate can be used.

Therefore, an alternative approach to DR based on Bayesian reasoning was developed that is able to perform DR with arbitrarily distributed input data (Cencic and Frühwirth, 2015, 2018). The goal of this work is to deliver a methodology to be able to compare the results from the classical approach, using the assumption of normally distributed data, to those of the Bayesian approach, using arbitrary pdfs.

Note that in this thesis pdfs are used to express the uncertainties of variables. Thus, probabilistic DR is covered only. For a possibilistic approach to DR, where the uncertainties of variables are expressed with membership functions (fuzzy sets) instead, see Dubois et al. (2014); Dzubur et al. (2017).

¹A transfer coefficient describes the percentage of the input of a process that is transferred to a certain output flow.

2 Methodology

Remark: Because the notations used in the three papers of this doctoral thesis (Cencic, 2016; Cencic and Frühwirth, 2015, 2018) are slightly different (e.g. accents of variables), it was necessary to unify them in this summary in order to be able to give a consistent overview of the used methods and to show the connections between them.

2.1 Error Model

Observations are subject to observation errors that are of random or systematic nature. The respective error model can be written as

$$\tilde{\mathbf{x}} = \boldsymbol{\mu}_x + \boldsymbol{\epsilon} + \boldsymbol{\delta}, \quad (2.1)$$

where $\tilde{\mathbf{x}}$ is the vector of observations, $\boldsymbol{\mu}_x$ is the vector of true values of the observed variables \mathbf{x} , $\boldsymbol{\epsilon}$ is the vector of random errors (with expectation $E(\boldsymbol{\epsilon}) = \mathbf{0}$) and $\boldsymbol{\delta}$ is the vector of measurement biases.

In the following, it is assumed that $\boldsymbol{\delta} = \mathbf{0}$, i.e. the observations are free of systematic errors. How to deal with gross errors ($\boldsymbol{\delta} \neq \mathbf{0}$), see sections 2.2.4 and 2.3.4.

2.2 Weighted Least Squares Approach to DR

If the observation errors $\boldsymbol{\epsilon}$ are assumed to be normally distributed with zero mean and known joint covariance matrix,

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\tilde{\mathbf{x}}}), \quad (2.2)$$

the best estimates $\hat{\mathbf{x}}$ of the true but unknown values $\boldsymbol{\mu}_x$ of the observed variables \mathbf{x} can be found by minimizing the objective function

$$J(\mathbf{x}) = (\mathbf{x} - \tilde{\mathbf{x}})^T \mathbf{Q}_{\tilde{\mathbf{x}}}^{-1} (\mathbf{x} - \tilde{\mathbf{x}}) \quad (2.3)$$

with respect to \mathbf{x} , subject to the constraints

$$\mathbf{G}(\mathbf{y}; \mathbf{x}; \mathbf{z}) = \mathbf{0}. \quad (2.4)$$

\mathbf{x} is the vector of observed variables, \mathbf{y} is the vector of unobserved variables and \mathbf{z} is the vector of fixed (nonrandom) quantities.

2.2.1 Linear Constraints

If Eq. (2.4) is a set of linear constraints, the system of equations can be written as

$$\begin{aligned} \mathbf{G}(\mathbf{y}; \mathbf{x}; \mathbf{z}) &= \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{C}\mathbf{z} = \mathbf{0}, \\ &= \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{c} = \mathbf{0}, \end{aligned} \quad (2.5)$$

where \mathbf{A} , \mathbf{B} and \mathbf{C} are coefficient matrices, which, in the linear case, contain only fixed entries. \mathbf{c} is a vector of aggregated fixed quantities.

If by transformation of the linear equality constraints at least one equation can be found that contains no unknown but at least one observed variable, DR can be performed to improve the accuracy and precision of the observations.

To eliminate unobserved variables from the DR problem, Madron (1992) proposed to apply a Gauss-Jordan elimination to the coefficient matrix $(\mathbf{B}, \mathbf{A}, \mathbf{c})$ ¹ of the linear constraints to get its canonical form (= reduced row echelon form, RREF). After reordering the columns of the resulting matrix and the corresponding rows of

¹The comma (semicolon) denotes horizontal (vertical) concatenation of vectors and matrices (Matlab convention).

the variable vector, the initial system of constraints

$$\begin{pmatrix} \mathbf{B} & \mathbf{A} & \mathbf{c} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ 1 \end{pmatrix} = \mathbf{0} \quad (2.6)$$

can be written as

$$\begin{pmatrix} \begin{array}{ccc|ccc|c} \mathbf{I} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{E}_1 & \mathbf{E}_2 & \mathbf{e} \\ \mathbf{O} & \mathbf{I} & \mathbf{F}_0 & \mathbf{O} & \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{f}_u \\ \hline \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{I} & \mathbf{D}_1 & \mathbf{O} & \mathbf{d} \\ \hline \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{f}_z \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{O} \end{array} \end{pmatrix} \begin{pmatrix} \mathbf{y}_o \\ \mathbf{y}_{u_1} \\ \mathbf{y}_{u_2} \\ \mathbf{x}_{r_1} \\ \mathbf{x}_{r_2} \\ \mathbf{x}_n \\ 1 \end{pmatrix} = \mathbf{0}. \quad (2.7)$$

The structure of the resulting matrix in Eq. (2.7) simplifies the reconciliation procedure and provides useful information for variable classification.

Given that all elements of Eq. (2.7) exist

- \mathbf{y}_o are “observable” unknown variables that can be computed from the constraints.
- \mathbf{y}_{u_1} and \mathbf{y}_{u_2} are “unobservable” unknown variables that cannot be computed from the constraints.
- \mathbf{x}_{r_1} and \mathbf{x}_{r_2} are “redundant” observed variables. Each of these variables could be computed from the rest of the redundant observations if its value was missing. Thus, it would become an observable unknown variable.
- \mathbf{x}_n are “nonredundant” observed variables. None of these variables could be computed from the rest of the observations if its value was missing. Thus, it would become an unobservable unknown variable.

- \mathbf{F}_0 is a matrix where each row contains at least one nonzero element. Zero columns in \mathbf{F}_0 indicate that the corresponding unobserved variables are not included in the constraints.
- \mathbf{D}_1 is a matrix where each column contains at least one nonzero element.
- \mathbf{E}_1 , \mathbf{E}_2 , \mathbf{F}_1 and \mathbf{F}_2 are matrices with arbitrary content. Zero columns in matrix $(\mathbf{E}_2; \mathbf{F}_2)$ indicate that the corresponding redundant observed variables are not included in the constraints.
- \mathbf{e} , \mathbf{f}_u and \mathbf{d} are column vectors with arbitrary content.
- f_z is a scalar that is either 0 or 1.

Zero rows at the bottom of the matrix exist if

- the given constraints included dependent equations that have been eliminated during the Gauss-Jordan elimination procedure,
- there are given/transformed constraints containing constant noncontradicting input data only.

If the constant input data violate given/transformed constraints, $f_z = 1$. In this case, the respective contradictions have to be resolved before being able to solve the system of equations, resulting in a zero row with $f_z = 0$.

In all cases, zero rows can be ignored because they have no influence on the solution of the equation system.

The remaining equations can then be written as

$$\mathbf{y}_o + \mathbf{E}_1 \mathbf{x}_{r_2} + \mathbf{E}_2 \mathbf{x}_n + \mathbf{e} = \mathbf{0}, \quad (2.8)$$

$$\mathbf{y}_{u_1} + \mathbf{F}_0 \mathbf{y}_{u_2} + \mathbf{F}_1 \mathbf{x}_{r_2} + \mathbf{F}_2 \mathbf{x}_n + \mathbf{f}_u = \mathbf{0}, \quad (2.9)$$

$$\mathbf{x}_{r_1} + \mathbf{D}_1 \mathbf{x}_{r_2} + \mathbf{d} = \mathbf{0}. \quad (2.10)$$

Eq. (2.10) is a set of equations, containing observed variables and fixed quantities only, which is normally not satisfied by the observations. However, it can be used to adjust the observations by DR. Note that Eq. (2.10) is free of nonredundant observed variables \mathbf{x}_n . That is the reason why \mathbf{x}_n is not adjusted during DR

provided that the observations \mathbf{x}_r and \mathbf{x}_n are not correlated.

Eq. (2.9) represents a set of equations that cannot be solved because each involved equation contains at least two unobservable variables, one from \mathbf{y}_{u_1} and least one from \mathbf{y}_{u_2} .

Eq. (2.8) is a set of equations that can be used to compute the observable variables \mathbf{y}_o .

Note that \mathbf{y}_o is a function of \mathbf{x}_{r_2} and \mathbf{x}_n , and \mathbf{x}_{r_1} is a function of \mathbf{x}_{r_2} only. Thus, in section 2.3, $(\mathbf{x}_{r_2}; \mathbf{x}_n)$ is denoted as the vector of free observed variables \mathbf{w} of the equation system, and \mathbf{x}_{r_1} as the vector of dependent observed variables \mathbf{u} .

For the sake of simplicity, we assume in the following that $\mathbf{y} = (\mathbf{y}_o; \mathbf{y}_{u_1}; \mathbf{y}_{u_2})$ and $\mathbf{x} = (\mathbf{x}_{r_1}; \mathbf{x}_{r_2}; \mathbf{x}_n) = (\mathbf{u}; \mathbf{w})$, i.e. the entries of \mathbf{y} and \mathbf{x} are already in the right order to reach the structure of the matrix in Eq. (2.7) without having to reorder any columns after the Gauss-Jordan elimination. Additionally, the classification of the observed variables will be ignored even if it could be exploited to accelerate the computation.

By removing the unobservable variables \mathbf{y}_{u_1} and \mathbf{y}_{u_2} from the variable vector, and deleting the corresponding rows and columns of the coefficient matrix in Eq. (2.7) (by deleting columns 2 and 3, and row 2), the constraints can be rewritten as

$$\left(\begin{array}{c|c|c} \mathbf{I} & \mathbf{E} & \mathbf{e} \\ \hline \mathbf{O} & \mathbf{D} & \mathbf{d} \end{array} \right) \begin{pmatrix} \mathbf{y}_o \\ \mathbf{x} \\ 1 \end{pmatrix} = \mathbf{0}. \quad (2.11)$$

with $\mathbf{E} = (\mathbf{O}, \mathbf{E}_1, \mathbf{E}_2)$ and $\mathbf{D} = (\mathbf{I}, \mathbf{D}_1, \mathbf{O})$.

Eq. (2.10) then becomes

$$\mathbf{D}\mathbf{x} + \mathbf{d} = \mathbf{0}. \quad (2.12)$$

The result of minimizing the objective function (Eq. (2.3)) subject to the now reduced set of constraints (Eq. (2.12)) can be found by using the classical method

of Lagrange multipliers (solved first in Kuehn and Davidson (1961)):

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} - \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}^T (\mathbf{D} \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}^T)^{-1} (\mathbf{D} \tilde{\mathbf{x}} + \mathbf{d}) \quad (2.13)$$

The best estimates of the observable unknown variables $\hat{\mathbf{y}}_o$ can then be calculated from Eq. (2.8):

$$\hat{\mathbf{y}}_o + \mathbf{E} \hat{\mathbf{x}} + \mathbf{e} = \mathbf{0} \quad (2.14)$$

The variance-covariance matrices $\mathbf{Q}_{\hat{\mathbf{x}}}$ of the reconciled observations $\hat{\mathbf{x}}$, and $\mathbf{Q}_{\hat{\mathbf{y}}_o}$ of the best estimates $\hat{\mathbf{y}}_o$ of the observable unknown variables can be computed by error propagation from Eqs. (2.13) and (2.14) leading to

$$\mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{I} - \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}^T (\mathbf{D} \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}^T)^{-1} \mathbf{D}) \mathbf{Q}_{\tilde{\mathbf{x}}}, \quad (2.15)$$

$$\mathbf{Q}_{\hat{\mathbf{y}}_o} = \mathbf{E} \mathbf{Q}_{\hat{\mathbf{x}}} \mathbf{E}^T. \quad (2.16)$$

The complete variance-covariance matrix of all estimated variables can be written as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{\hat{\mathbf{y}}_o} & -\mathbf{E} \mathbf{Q}_{\hat{\mathbf{x}}} \\ -\mathbf{Q}_{\hat{\mathbf{x}}} \mathbf{E}^T & \mathbf{Q}_{\hat{\mathbf{x}}} \end{pmatrix}. \quad (2.17)$$

Fully worked examples can be found in Brunner and Rechberger (2016, section 2.3).

2.2.2 Nonlinear Constraints

Nonlinear DR problems that contain only equality constraints can be solved using iterative techniques based on successive linearization and analytical solution of the linear reconciliation problem (Narasimhan and Jordache, 2000).

If Eq. (2.4) is a set of nonlinear constraints, a linear approximation can be obtained from a first order Taylor series expansion:

$$\mathbf{G}(\mathbf{y}; \mathbf{x}; \mathbf{z}) \approx \mathbf{J}_y(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z})(\mathbf{y} - \hat{\mathbf{y}}) + \mathbf{J}_x(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z})(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{G}(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z}) = \mathbf{0} \quad (2.18)$$

This can be written as

$$\begin{pmatrix} \mathbf{B} & \mathbf{A} & \mathbf{c} \end{pmatrix} \begin{pmatrix} \mathbf{y} - \hat{\mathbf{y}} \\ \mathbf{x} - \hat{\mathbf{x}} \\ 1 \end{pmatrix} = \mathbf{0}, \quad (2.19)$$

where \mathbf{A} and \mathbf{B} are the Jacobi matrices $\mathbf{J}_x = \partial \mathbf{G} / \partial \mathbf{x}$ and $\mathbf{J}_y = \partial \mathbf{G} / \partial \mathbf{y}$, respectively, and \mathbf{c} is the vector of the residuals of the equality constraints \mathbf{G} , all evaluated at the expansion point $(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z})$.

The only differences to the linear case (Eq. (2.6)) are:

- the variable vector contains the differences to the expansion point instead of the variable values themselves,
- the entries of \mathbf{A} , \mathbf{B} and \mathbf{c} may also contain functions of variables (evaluated at the expansion point), in contrast to only constant entries in the case of linear constraints.

Because of the latter, the solution must be found iteratively.

Applying the same procedure as described in section 2.2.1, the reduced constraints can be written as

$$(\mathbf{y}_o - \hat{\mathbf{y}}_o) + \mathbf{E}(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{e} = \mathbf{0}, \quad (2.20)$$

$$\mathbf{D}(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{d} = \mathbf{0}. \quad (2.21)$$

The solution of minimizing the objective function (Eq. (2.3)) subject to the reduced set of constraints (Eq. (2.21)) can again be found by using the classical method of Lagrange multipliers:

$$\hat{\mathbf{x}}_{i+1} = \tilde{\mathbf{x}} - \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}_i^T (\mathbf{D}_i \mathbf{Q}_{\tilde{\mathbf{x}}} \mathbf{D}_i^T)^{-1} (\mathbf{D}_i (\tilde{\mathbf{x}} - \hat{\mathbf{x}}_i) + \mathbf{d}_i) \quad (2.22)$$

The best estimates of the observable unknown variables \mathbf{y}_o can be computed from

Eq. (2.20):

$$\hat{\mathbf{y}}_{o,i+1} = \hat{\mathbf{y}}_{o,i} - \mathbf{E}_i(\hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i) - \mathbf{e}_i \quad (2.23)$$

In the first iteration, the observations $\tilde{\mathbf{x}}$ are taken as initial estimates $\hat{\mathbf{x}}_1$. If there are also unobserved variables, an educated guess of the initial estimates $\hat{\mathbf{y}}_{o,1}$ has to be provided by the user. Alternatively, e.g. the constraint consensus method (Chinneck, 2004) can be employed to find proper starting values for unobserved variables.

If the new estimates $\hat{\mathbf{x}}_{i+1}$ and $\hat{\mathbf{y}}_{o,i+1}$ are significantly different from the previous estimates $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{y}}_{o,i}$, respectively, another iteration is performed by re-expanding the nonlinear constraints at the updated expansion point $(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z})$ (see Eq. (2.18)). Note that the new $\hat{\mathbf{y}}$ also contains the initial estimates of the unobservable unknown variables $\hat{\mathbf{y}}_u$. As soon as convergence is reached, the procedure is stopped and the complete variance-covariance matrix is computed from Eqs. (2.15), (2.16) and (2.17), as in the linear case.

A fully worked example can be found in Cencic (2016).

2.2.3 Correlated Observations

In the case of normally distributed measurement errors, correlations between the observations $\tilde{\mathbf{x}}$ can be easily introduced by modifying their joint covariance matrix $\mathbf{Q}_{\tilde{\mathbf{x}}}$.

If the correlation matrix \mathbf{R} is given, the corresponding covariance matrix $\mathbf{Q}_{\tilde{\mathbf{x}}}$ can be computed from

$$\mathbf{Q}_{\tilde{\mathbf{x}}} = \text{diag}(\boldsymbol{\sigma}_{\tilde{\mathbf{x}}}) \mathbf{R} \text{diag}(\boldsymbol{\sigma}_{\tilde{\mathbf{x}}}), \quad (2.24)$$

where $\text{diag}(\boldsymbol{\sigma}_{\tilde{\mathbf{x}}})$ is a diagonal matrix constructed from the vector of standard deviations of the observations.

2.2.4 Gross Error Detection

Beyond their statistical uncertainty, the observations may also be corrupted by gross errors δ such as biased observations or faulty readings. If these gross errors are not detected and eliminated or at least down-weighted, the reconciled values will be biased.

If the observations follow normal distributions, there are various test statistics with known distribution under the null hypothesis of no gross errors. These can be used for detecting or identifying corrupted observations. For instance, the mere presence of gross errors can be detected by a test on the global chi-square statistic (Almasy and Sztano, 1975; Madron et al., 1977). In Tamhane and Mah (1985), two tests were discussed that identify the contaminated observation(s) so that they can be removed from the reconciliation process, the nodal test and the measurement test (see also Madron (1992)). Instead of identifying and removing observations with gross errors, the approach taken by robust methods is to give them smaller weight or larger variance during reconciliation. In Johnston and Kramer (1995), a maximum likelihood rectification technique was proposed that is closely related to robust regression. The use of M-estimators in general and of redescending M-estimators in particular has been discussed extensively in the literature, see e.g. Arora and Biegler (2001); Özyurt and Pike (2004); Hu and Shao (2006); Llanos et al. (2015). Finally, the methods proposed in Alhaj-Dibo et al. (2008); Yuan et al. (2015) describe simultaneous reconciliation and gross error detection based on prior information about the distribution of the gross errors.

For comprehensive reviews on gross error detection techniques with illustrative examples, see Narasimhan and Jordache (2000); Romagnoli and Sanchez (2000); Bagajewicz (2010).

2.3 Bayesian Approach to DR

In the context of MFA, in most cases, the precise but unknown true value of a quantity of interest is to be estimated (e.g. the mass of residual solid waste produced in Austria in the year 2016). The respective uncertainty of this estimator is of

epistemic nature (in contrast to aleatory variability) because it could be reduced if more information were available. Often, the assumption of normally distributed observation errors is not justified (e.g. mass flows, for instance, cannot take negative values, and mass fractions and transfer coefficients are restricted to the unit interval). The more information about the quantity of interest is available, the better the shape of the pdf of the estimator can be modeled. If a sufficient number of observations of the quantity is available, either a parametric model can be fitted, or a nonparametric model such as the empirical distribution function or the kernel estimate can be used.

If no observation is available, expert opinions are often used instead to restrict the possible location of the true value of a quantity of interest. These opinions are frequently modeled by uniform, triangular or trapezoidal distributions, depending on the expert's knowledge about the quantity under consideration.

As the objective function in Eq. (2.3) uses only the first two moments of the distributions, in all of the above mentioned cases, it is not possible to take into account the complete information contained in the joint pdf of the observations. Consequently, the reconciliation problem cannot be fully solved by minimizing an objective function of this form. Only in the case of linear constraints, the constrained least-squares estimator is unbiased and a linear function of the observations, and therefore gives the correct mean and covariance matrix of the reconciled values. Their distribution, however, is not known, and it is not possible to compute quantiles or higher moments such as the skewness.

This problem was solved in Cencic and Frühwirth (2015) by a Bayesian method that gives the joint (posterior) distribution of the reconciled variables under linear constraints for arbitrary continuous (prior) distributions of the observations. In Cencic and Frühwirth (2018), the method was extended to nonlinear constraints.

The main idea of this approach is to restrict the joint prior probability distribution of the observed variables with model constraints to get their joint posterior probability distribution. Thus, the posterior distribution is the prior distribution conditional on the constraints, and not on observed data (which are already part of the prior distribution).

Example 1: Let us assume that there are three observed variables x_1 , x_2 and x_3 with the prior density $f(x_1, x_2, x_3)$ defined on \mathbb{R}^3 . The constraint equation $x_3 = 0.2 x_1 x_2$ defines a surface in \mathbb{R}^3 . If the prior density is restricted to points in this surface and normalized to 1, the joint posterior density of x_1, x_2, x_3 is obtained. By computing the marginal distributions of the posterior we get the posterior densities of x_1 , x_2 and x_3 , respectively.

Figure 2.1 shows an instance of this problem, with x_1, x_2, x_3 independent, $f_1(x_1) = \gamma(x_1; 2, 2)$, $f_2(x_2) = \gamma(x_2; 3, 1.5)$ and $f_3(x_3) = \gamma(x_3; 6, 1.7)$, where $\gamma(x; a, b)$ is the density of the gamma distribution with parameters a and b :

$$\gamma(x; a, b) = \frac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)}$$

The values of $f(x_1, x_2, x_3)$ are shown color-coded. □

Notes: (1) The Bayesian approach also allows to consider inequality constraints by introducing slack variables. For instance, the inequality constraint $x_1 \leq f(x_2, x_3)$ can be transformed into the equality constraint $x_1 + x_S = f(x_2, x_3)$ with a slack variable $x_S \geq 0$, which is assumed to have a proper or an improper prior enforcing positivity. (2) If multiple observations (priors) of the same quantity are available, additional equality constraints have to be added stating that the posteriors of the observations have to be identical. (3) While in the classical weighted least squares approach only the point estimators of true values are reconciled, in the Bayesian approach the complete pdfs are taken into account.

As shown in section 2.2.1, the n_x observed variables \mathbf{x} can be split into n_w free variables \mathbf{w} and n_u dependent variables \mathbf{u} . The n_{y_o} observable variables \mathbf{y}_o and the dependent observed variables \mathbf{u} can then be expressed as functions of the free variables \mathbf{w} :

$$\mathbf{y}_o = \mathbf{k}(\mathbf{w}) \tag{2.25}$$

$$\mathbf{u} = \mathbf{h}(\mathbf{w}) \tag{2.26}$$

In Cencic and Frühwirth (2018) it was shown that the posterior density of the free

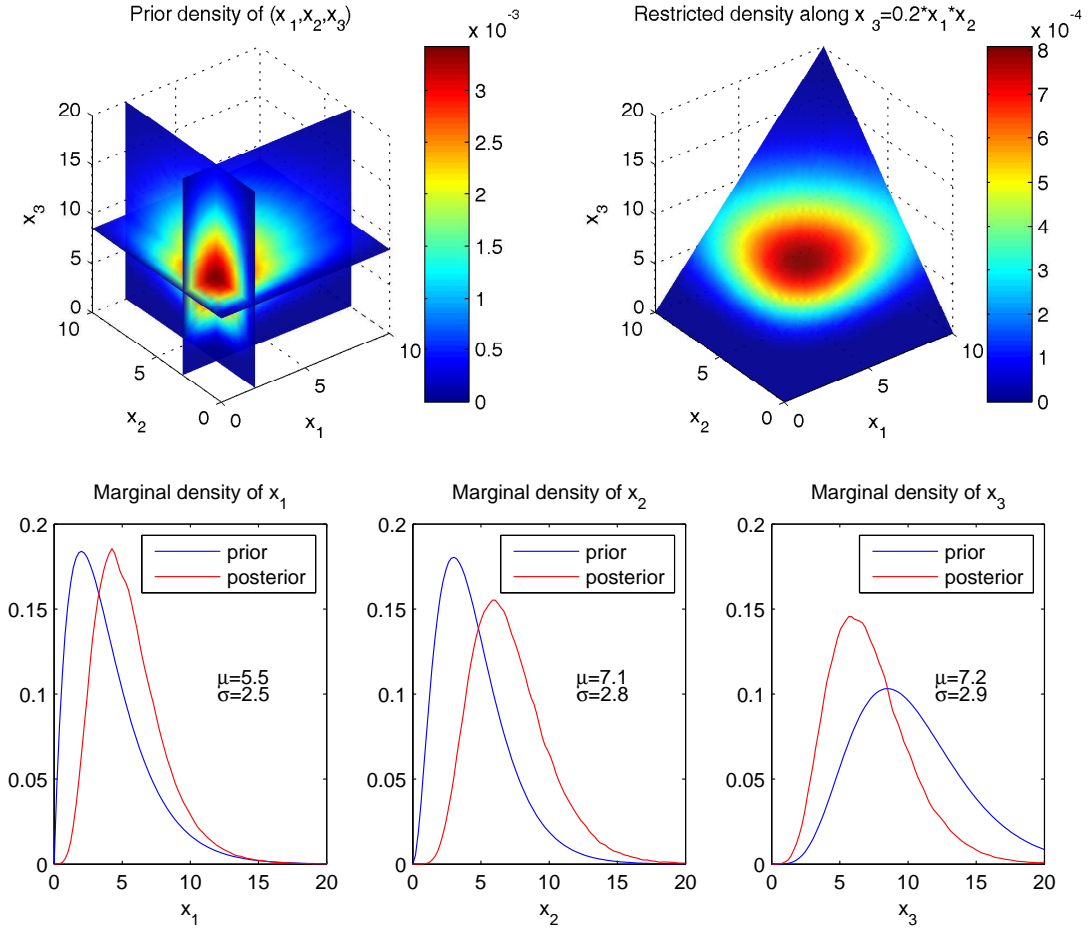


Figure 2.1: Visualization of the densities in Example 1. Top left: prior density $f(x_1, x_2, x_3)$. Top right: prior density $f(x_1, x_2, x_3)$ on the surface $x_3 = 0.2x_1x_2$. Bottom: observed (prior, blue) and reconciled (posterior, red) densities of x_1, x_2, x_3 .

variables \mathbf{w} can be written as

$$\pi(\mathbf{w}) = \frac{f(\mathbf{h}(\mathbf{w}), \mathbf{w}) V(\mathbf{w})}{\int_{\mathbf{W}} f(\mathbf{h}(\mathbf{w}), \mathbf{w}) V(\mathbf{w}) d\mathbf{w}}, \quad (2.27)$$

where \mathbf{W} is the domain of the free variables \mathbf{w} , $f(\mathbf{u}, \mathbf{w})$ is the joint prior density of the observed variables, and $V(\mathbf{w}) d\mathbf{w} = dS(\mathbf{w})$ is a differential element of the constraint manifold S defined by Eq. (2.26). S is of dimension n_w and can thus be a curve, a surface, a volume, or a hypervolume.

$V(\mathbf{w})$ can be computed from

$$V(\mathbf{w}) = \sqrt{|\mathbf{I} + \mathbf{H}^T \mathbf{H}|}, \quad (2.28)$$

where \mathbf{H} is the Jacobian of the function \mathbf{h} ,

$$\mathbf{H}(\mathbf{w}) = \frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}}, \quad (2.29)$$

and $\mathbf{I} + \mathbf{H}^T \mathbf{H}$ is the metric tensor of the induced metric in \mathbf{S} (O'Neill, 1983).

If the observed variables are independent, their joint prior density factorizes into the marginal densities $f_u(\mathbf{u})$ and $f_w(\mathbf{w})$, and the posterior density of \mathbf{w} is equal to

$$\pi(\mathbf{w}) = \frac{f_u(\mathbf{h}(\mathbf{w})) f_w(\mathbf{w}) V(\mathbf{w})}{\int_{\mathbf{W}} f_u(\mathbf{h}(\mathbf{w})) f_w(\mathbf{w}) V(\mathbf{w}) d\mathbf{w}}. \quad (2.30)$$

The reason why it is not possible to state the posterior distribution $\pi(\mathbf{x})$ of all observed variables explicitly is demonstrated by the following example: Let us assume there are two measured variables x_1 and x_2 , with given joint prior distribution $f(x_1, x_2)$, which have to obey the constraint $x_2 = h(x_1)$. The constraint can be visualized as a 1-dimensional cut of the 2-dimensional density $f(x_1, x_2)$. Taking the cut as a new 1-dimensional coordinate system, there exists a posterior density along this cut. But, because the cut has no area, there is no posterior density defined with respect to the original 2-dimensional coordinate system, which is denoted as singular pdf. However, the 1-dimensional posterior density along the cut can be transformed into a corresponding pdf with respect to a different 1-dimensional reference system (the one of the so called free variable), which in our case could be the x_1 -axis or the x_2 -axis of the originally given 2-dimensional reference system. Note that in this example it is completely arbitrary which variable to choose as the free variable, provided the transformation is done correctly by taking into account the metric structure, i.e., the arc length of the curve as a function of x_1 or x_2 .

The explicit calculation of the posterior density in Eqs. (2.27) and (2.30) can be avoided by generating a random sample from the posterior distribution by means

of Markov chain Monte Carlo (MCMC) methods (Robert and Casella, 2004; Liu, 2004; Brooks et al., 2011). This has two advantages: 1. the normalization constant (= denominator of Eqs. (2.27) and (2.30)) is irrelevant; 2. the corresponding sample values of the dependent variables \mathbf{y}_o and \mathbf{u} can be computed by Eqs. (2.25) and (2.26). It is therefore straightforward to estimate posterior marginals, expectations, variances and covariances of *all* variables from the full sample.

The first MCMC sampling algorithm was presented in Metropolis et al. (1953), which was generalized later in Hastings (1970). Its goal is to get a large representative sample from a posterior distribution $\pi(\mathbf{w})$ that cannot be sampled directly or whose normalizing constant is not known. The requirements for this sampler are

- a function proportional to the posterior (in our case the numerator of Eqs. (2.27) and (2.30)), which can be evaluated at any point \mathbf{w} ,
- a proposal density $p(\mathbf{w}_{target}|\mathbf{w}_{source}) = p(\mathbf{w}_{target} - \mathbf{w}_{source})$ to suggest where to go next, which can also be evaluated at any point \mathbf{w} , and from which random vectors can be generated.

The proposed position $\dot{\mathbf{w}}$ ($= \mathbf{w}_{target}$), which depends on the current position \mathbf{w}_i ($= \mathbf{w}_{source}$) in the chain, is accepted with a certain probability $\alpha(\mathbf{w}_i, \dot{\mathbf{w}})$, i.e. if a uniform random number drawn from the unit interval is smaller than or equal to

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{\pi(\dot{\mathbf{w}}) p(\mathbf{w}_i|\dot{\mathbf{w}})}{\pi(\mathbf{w}_i) p(\dot{\mathbf{w}}|\mathbf{w}_i)} \right). \quad (2.31)$$

Otherwise, the current position \mathbf{w}_i is appended to the sample instead.

In Cencic and Frühwirth (2015) it was argued that the sampler best suited to the context of DR is the independence sampler (Chib and Greenberg, 1995; Liu, 1996; Brooks et al., 2011), in which the proposal values $\dot{\mathbf{w}}$ are drawn from a proposal density $p(\mathbf{w})$ independent of the current position. The acceptance probability of the sampler is given by

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{\pi(\dot{\mathbf{w}}) p(\mathbf{w}_i)}{\pi(\mathbf{w}_i) p(\dot{\mathbf{w}})} \right). \quad (2.32)$$

In the case of independent observations, this is equivalent to

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{f_u(\mathbf{h}(\dot{\mathbf{w}})) f_w(\dot{\mathbf{w}}) V(\dot{\mathbf{w}}) p(\mathbf{w}_i)}{f_u(\mathbf{h}(\mathbf{w}_i)) f_w(\mathbf{w}_i) V(\mathbf{w}_i) p(\dot{\mathbf{w}})} \right). \quad (2.33)$$

Note that the normalizing constant of $\pi(\mathbf{w})$ cancels in Eqs. (2.31), (2.32) and (2.33), so there is no need to compute it. If the proposal density is chosen as $p(\mathbf{w}) = f_w(\mathbf{w})$, Eq. (2.33) reduces to

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{f_u(\mathbf{h}(\dot{\mathbf{w}})) V(\dot{\mathbf{w}})}{f_u(\mathbf{h}(\mathbf{w}_i)) V(\mathbf{w}_i)} \right). \quad (2.34)$$

In the general case of correlated observations, the acceptance probability has to be computed according to Eq. (2.32), with a suitable proposal density $p(\mathbf{w})$.

2.3.1 Linear Constraints

Using \mathbf{w} and \mathbf{u} instead of \mathbf{x} , Eq. (2.7) can be rewritten as

$$\mathbf{G}(\mathbf{y}; \mathbf{u}; \mathbf{w}) = \begin{pmatrix} \mathbf{I} & \mathbf{O} & \mathbf{E} & \mathbf{e} \\ \mathbf{O} & \mathbf{I} & \mathbf{D} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{y}_o \\ \mathbf{u} \\ \mathbf{w} \\ 1 \end{pmatrix} = \mathbf{0}, \quad (2.35)$$

with $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2)$, $\mathbf{D} = (\mathbf{D}_1, \mathbf{O})$, $\mathbf{u} = \mathbf{x}_{r_1}$ and $\mathbf{w} = (\mathbf{x}_{r_2}; \mathbf{x}_n)$.

Due to the Gauß-Jordan elimination, the observable unknown variables \mathbf{y}_o , which are linear functions of \mathbf{w} only, are eliminated from the DR problem, simplifying the constraints to

$$\mathbf{u} = \mathbf{h}(\mathbf{w}) = -\mathbf{D}\mathbf{w} - \mathbf{d}. \quad (2.36)$$

The observable unknown variables can be computed from

$$\mathbf{y}_o = \mathbf{k}(\mathbf{w}) = -\mathbf{E}\mathbf{w} - \mathbf{e}. \quad (2.37)$$

Thus, for any given \mathbf{w} the corresponding \mathbf{u} and \mathbf{y}_o can be computed from Eqs. (2.36) and (2.37), which is prerequisite for the Bayesian approach.

In the case of linear constraints, $V(\mathbf{w})$ is a constant and cancels in the posterior densities defined in Eqs. (2.27) and (2.30).

Fully worked examples can be found in Cencic and Frühwirth (2015).

2.3.2 Nonlinear Constraints

If the explicit computation of the dependent variables \mathbf{y}_o and \mathbf{u} as functions of the chosen free variables \mathbf{w} (see Eqs. (2.25) and (2.26)) is not feasible, the solution has to be computed by numerical methods. The algorithms that can be employed to this purpose fall into two categories. Algorithms in the first category use gradient information, algorithms in the second category do not, i.e. are gradient-free. A typical example of the first category is the Newton-Raphson algorithm.

Adapting Eq. (2.35) for nonlinear constraints yields

$$\mathbf{G}(\mathbf{y}; \mathbf{u}; \mathbf{w}) \approx \begin{pmatrix} \mathbf{I} & \mathbf{O} & \mathbf{E} & \mathbf{e} \\ \mathbf{O} & \mathbf{I} & \mathbf{D} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{y}_o - \hat{\mathbf{y}}_o \\ \mathbf{u} - \hat{\mathbf{u}} \\ \mathbf{w} - \hat{\mathbf{w}} \\ 1 \end{pmatrix} = \mathbf{0}. \quad (2.38)$$

For given \mathbf{w} , the corresponding \mathbf{y}_o and \mathbf{u} are to be computed. With initial educated guesses $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{u}}$, and the choice $\hat{\mathbf{w}} = \mathbf{w}$, Eq. (2.38) reduces to

$$(\mathbf{y}_o - \hat{\mathbf{y}}_o) + \mathbf{e} = \mathbf{0}, \quad (2.39)$$

$$(\mathbf{u} - \hat{\mathbf{u}}) + \mathbf{d} = \mathbf{0}, \quad (2.40)$$

leading to the update equations

$$\hat{\mathbf{y}}_{o,i+1} = \hat{\mathbf{y}}_{o,i} - \mathbf{e}_i, \quad (2.41)$$

$$\hat{\mathbf{u}}_{i+1} = \hat{\mathbf{u}}_i - \mathbf{d}_i. \quad (2.42)$$

If $\hat{\mathbf{y}}_{o,i+1}$ and $\hat{\mathbf{u}}_{i+1}$ are significantly different from $\hat{\mathbf{y}}_{o,i}$ and $\hat{\mathbf{u}}_i$, respectively, another iteration is performed with re-expanding the nonlinear constraints at the updated expansion point $(\hat{\mathbf{y}}; \hat{\mathbf{x}}; \mathbf{z}) = (\hat{\mathbf{y}}; \hat{\mathbf{u}}; \hat{\mathbf{w}}; \mathbf{z})$ as described in section 2.2.2. Note that the new $\hat{\mathbf{y}}$ also contains the initial estimates of the unobservable unknown variables $\hat{\mathbf{y}}_u$. Convergence is guaranteed only if the initial $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{u}}_1$ are sufficiently close to the final solution.

Note that, in this context, the finally found $\hat{\mathbf{y}}$, $\hat{\mathbf{u}}$ and $\hat{\mathbf{w}}$ are not estimated parameters of distributions, as in chapter 2.2.1 and 2.2.2, but a set of numbers complying with the constraints. Thus, for any given \mathbf{w} the corresponding \mathbf{u} and \mathbf{y} can be computed by this iterative procedure, which is prerequisite for the Bayesian approach.

If the Newton-Raphson iteration (or any other gradient-based method) fails to converge, a gradient-free approach can be applied. For example, the objective function

$$J(\mathbf{y}; \mathbf{u}) = \|\mathbf{G}(\mathbf{y}; \mathbf{u}; \mathbf{w})\|_2 \quad (2.43)$$

can be minimized for given \mathbf{w} with respect to \mathbf{y} and \mathbf{u} by the simplex algorithm (Nelder and Mead, 1965).

Gradient-based methods need an initial expansion point, gradient-less methods need a starting point. For a dependent observed variable the natural choice is the mode of the prior distribution of the variable. If there are unobserved variables, an educated guess of the starting point should be sufficient to find the correct solution by the simplex algorithm. Alternative methods such as the constraint consensus method (Chinneck, 2004) can also be employed to find starting values for unobserved variables. The simplex algorithm is less sensitive to the starting point than gradient-based methods, as it is possible to leave a local minimum by restarting the search with a sufficiently large initial simplex.

The application of the independence sampler additionally requires the computation of $\mathbf{H} = \partial \mathbf{u} / \partial \mathbf{w}$ to derive $V(\mathbf{w})$, which is part of the posterior density

(see Eqs. (2.27) and (2.30)). It follows from Eq. (2.38) that

$$\mathbf{H} = -\mathbf{D}. \quad (2.44)$$

Example 2: The example in Figure 2.2 shows a splitting process with one input and two output flows. All variables, three mass flows and two transfer coefficients, are measured ($n = n_x = 5$). The constraints are the mass balance of the process and two transfer coefficient equations ($n_c = 3$):

$$x_1 = x_2 + x_3$$

$$x_2 = x_1 t_2$$

$$x_3 = x_1 t_3$$

The observations are assumed to be independent with the following prior distributions:

$$x_1 \sim \text{Trap}(80, 90, 110, 120)$$

$$x_2 \sim \text{Tria}(40, 60, 80)$$

$$x_3 \sim \text{Unif}(30, 50)$$

$$t_2 \sim \text{Unif}(0.5, 0.7)$$

$$t_3 \sim \text{Unif}(0.3, 0.5)$$

$\text{Unif}(a, b)$ denotes the uniform distribution in the interval $[a, b]$, $\text{Tria}(a, b, c)$ denotes the triangular distribution in the interval $[a, c]$ with maximum at b , and $\text{Trap}(a, b, c, d)$ denotes the trapezoidal distribution in the interval $[a, d]$ that is uni-

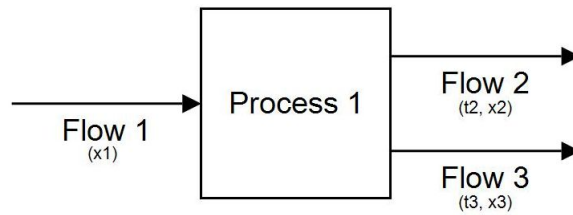


Figure 2.2: Flowsheet of Example 2.3 with three flows and one processes.

form in $[b, c]$.

For this simple example, $\mathbf{u} = \mathbf{h}(\mathbf{w})$ can be written in closed form. If e.g. $\mathbf{w} = (x_3; t_3)$ is selected as the $n_w = n - n_c = 5 - 3 = 2$ free variables, $\mathbf{u} = (x_1; x_2; t_2) = \mathbf{h}(\mathbf{w})$ becomes

$$\mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \\ t_2 \end{pmatrix} = \begin{pmatrix} x_3/t_3 \\ x_3/t_3 - x_3 \\ 1 - t_3 \end{pmatrix}.$$

Note that not all choices of \mathbf{w} lead to a feasible solution for the dependent variables: e.g. $\mathbf{w} = (t_2; t_3)$ leads to $\mathbf{u} = (x_1; x_2; x_3) = (0; 0; 0)$.

$V(\mathbf{w})$ can be derived from Eq. (2.28):

$$\mathbf{H}(\mathbf{w}) = \frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}} = \begin{pmatrix} \frac{\partial x_1}{\partial x_3} & \frac{\partial x_1}{\partial t_3} \\ \frac{\partial x_2}{\partial x_3} & \frac{\partial x_2}{\partial t_3} \\ \frac{\partial t_2}{\partial x_3} & \frac{\partial t_2}{\partial t_3} \end{pmatrix} = \begin{pmatrix} \frac{1}{t_3} & -\frac{x_3}{t_3^2} \\ (\frac{1}{t_3} - 1) & -\frac{x_3}{t_3^2} \\ 0 & -1 \end{pmatrix}$$

$$V(\mathbf{w}) = \sqrt{|\mathbf{I} + \mathbf{H}^T \mathbf{H}|} = \sqrt{(3x_3^2 + 4t_3^4 - 4t_3^3 + 4t_3^2)/t_3^4}$$

Because the analytical computation of $V(\mathbf{w})$ gets laborious pretty fast even for small models, the numerical solution is often to be preferred:

The Taylor series expansion of the original constraints leads to

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 0 & -1 & 0 & \hat{x}_1 - \hat{x}_2 - \hat{x}_3 \\ -\hat{t}_2 & 1 & -\hat{x}_1 & 0 & 0 & \hat{x}_2 - \hat{x}_1 \hat{t}_2 \\ -\hat{t}_3 & 0 & 0 & 1 & -\hat{x}_1 & \hat{x}_3 - \hat{x}_1 \hat{t}_3 \end{array} \right) \begin{pmatrix} x_1 - \hat{x}_1 \\ x_2 - \hat{x}_2 \\ t_2 - \hat{t}_2 \\ x_3 - \hat{x}_3 \\ t_3 - \hat{t}_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

After the Gauss-Jordan elimination, these constraints can be written as

$$\left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & -\frac{1}{\hat{t}_3} & \frac{\hat{x}_1}{\hat{t}_3} & \frac{\hat{x}_3 - \hat{t}_3 \hat{x}_1}{\hat{t}_3} \\ 0 & 1 & 0 & \frac{\hat{t}_3 - 1}{\hat{t}_3} & \frac{\hat{x}_1}{\hat{t}_3} & \frac{\hat{t}_3 \hat{x}_2 - \hat{x}_3 + \hat{t}_3 \hat{x}_3}{\hat{t}_3} \\ 0 & 0 & 1 & \frac{\hat{t}_2 + \hat{t}_3 - 1}{\hat{t}_3 \hat{x}_1} & -\frac{\hat{t}_2 - 1}{\hat{t}_3} & \frac{\hat{x}_3(\hat{t}_2 + \hat{t}_3 - 1)}{\hat{t}_3 \hat{x}_1} \end{array} \right) \begin{pmatrix} x_1 - \hat{x}_1 \\ x_2 - \hat{x}_2 \\ t_2 - \hat{t}_2 \\ x_3 - \hat{x}_3 \\ t_3 - \hat{t}_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{I} & \mathbf{D} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{u} - \hat{\mathbf{u}} \\ \mathbf{w} - \hat{\mathbf{w}} \\ 1 \end{pmatrix} = \mathbf{0}.$$

Note that normally the Gauss-Jordan elimination is performed numerically. For comparison reasons, here it was done analytically.

Using the Newton-Raphson algorithm, the corresponding \mathbf{u} for given \mathbf{w} are computed iteratively via update equation Eq. (2.42). After convergence, the constraints can be written as

$$\left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & -\frac{1}{\hat{t}_3} & \frac{\hat{x}_3}{\hat{t}_3^2} & 0 \\ 0 & 1 & 0 & \frac{\hat{t}_3 - 1}{\hat{t}_3} & \frac{\hat{x}_3}{\hat{t}_3^2} & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array} \right) \begin{pmatrix} x_1 - \hat{x}_1 \\ x_2 - \hat{x}_2 \\ t_2 - \hat{t}_2 \\ x_3 - \hat{x}_3 \\ t_3 - \hat{t}_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Comparing the result with the analytical derivation of $\mathbf{H}(\mathbf{w})$, it can be seen that $\mathbf{H}(\mathbf{w}) = -\mathbf{D}$, as stated in Eq. (2.44).

By using the independence sampler, a sample $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_L)$ of size L is drawn from the posterior distribution of the free variables \mathbf{w} . The corresponding sample $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_L)$ of the dependent variables \mathbf{u} is computed by using

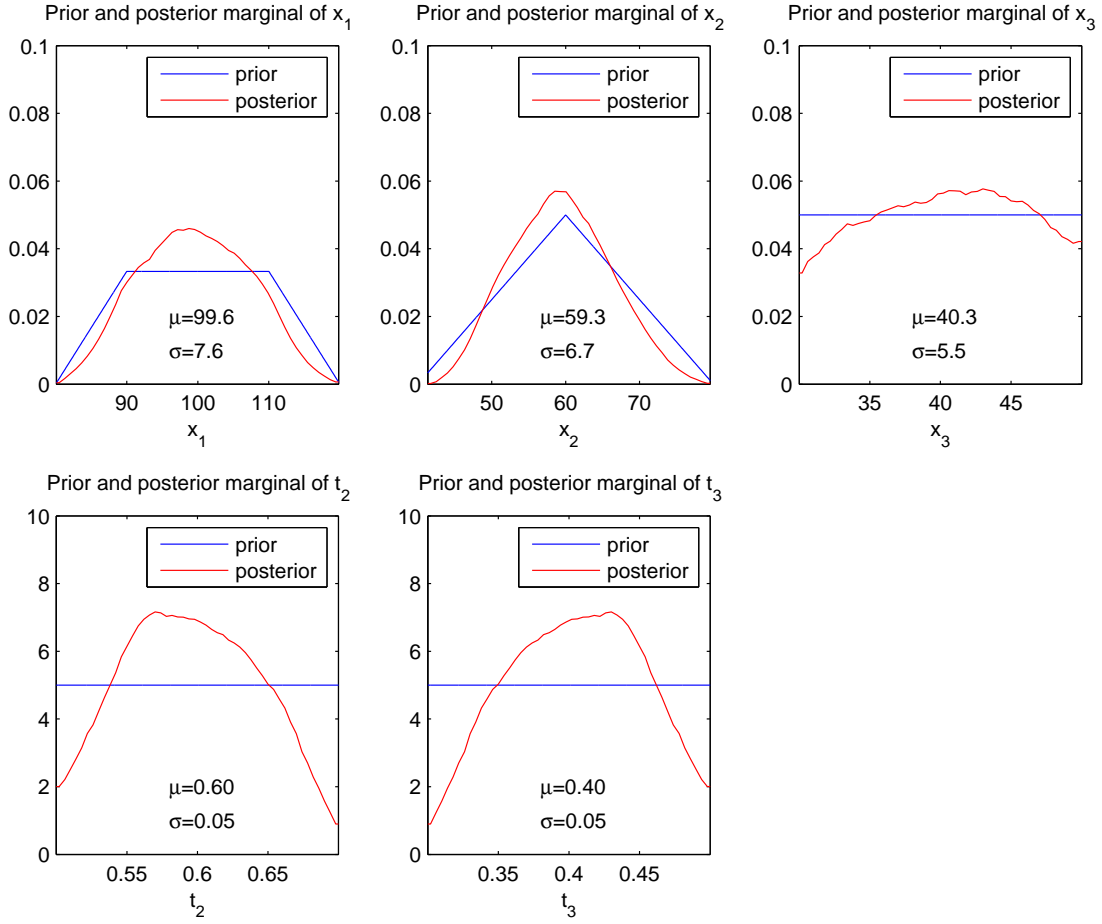


Figure 2.3: Priors and smoothed marginal posteriors of x_1, x_2, x_3, t_2, t_3 in Example 2

Eq. (2.26). Figure 2.3 shows the smoothed²marginal posterior densities with their means and standard deviations. The latter are estimated from the posterior sample and are given with a precision that is commensurate with their standard error. \square

Other worked examples can be found in Cencic and Frühwirth (2018).

²We have used the function `smooth` in the MATLAB Curve Fitting Toolbox.

2.3.3 Correlated Observations

In the case of normal observations, correlations between the observed variables can be introduced by modifying their joint covariance matrix. In the nonnormal case, the standard way of imposing correlations on observations with given marginal distributions is via a copula.

A function $C : [0, 1]^d \rightarrow [0, 1]$ is called a d -dimensional copula if it is the joint cumulative distribution function (cdf) of a d -dimensional random vector on the unit hypercube $[0, 1]^d$ with uniform marginals (Nelsen, 2006). In the following, the Gaussian copula $C_G(\boldsymbol{\xi}; \mathbf{R})$ with $\boldsymbol{\xi} \in [0, 1]^d$ and correlation matrix \mathbf{R} is used. Its density c_G is given by

$$c_G(\boldsymbol{\xi}; \mathbf{R}) = \frac{1}{\sqrt{\det \mathbf{R}}} \exp \left[-\frac{1}{2} \Phi^{-1}(\boldsymbol{\xi})^T (\mathbf{R}^{-1} - \mathbf{I}) \Phi^{-1}(\boldsymbol{\xi}) \right], \quad \boldsymbol{\xi} \in [0, 1]^d, \quad (2.45)$$

where Φ^{-1} is the inverse distribution function of the d -dimensional standard normal distribution.

In the general form of the posterior density (Eq. (2.27)), $f(\mathbf{u}; \mathbf{w})$ is the joint density of all observed variables, which in the presence of correlations no longer factorizes into the marginal densities of \mathbf{u} and \mathbf{w} (cf. Eq. (2.30)).

Let $f_k(x_k)$ denote the prior marginal density of the (observed) variable x_k and $F_k(x_k)$ its cdf, where $k = 1, \dots, n_x$. \mathbf{R} is the assumed or estimated correlation matrix of \mathbf{x} . The following lemma shows how to compute the joint density function of \mathbf{x} and its correlation matrix.

Lemma. *Assume that $\boldsymbol{\xi}$ is distributed according to the Gaussian copula $C_G(\boldsymbol{\xi}; \mathbf{R})$ and $x_k = F_k^{-1}(\xi_k)$, $k = 1, \dots, n_x$. Then:*

(a) *The joint density of \mathbf{x} is equal to*

$$g(\mathbf{x}) = c_G(F_1(x_1), \dots, F_{n_x}(x_{n_x}); \mathbf{R}) \prod_{k=1}^{n_x} f_k(x_k) \quad (2.46)$$

and the marginal density of x_k is equal to $f_k(x_k)$.

(b) *The joint correlation matrix of \mathbf{x} is equal to \mathbf{R} in first-order Taylor approximation.*

Proof. Assertion (a) follows from the relations $\xi_k = F_k(\mathbf{x}_k)$ and the transformation theorem for densities:

$$\begin{aligned} g(\mathbf{x}) &= c_G(\xi, \dots, \xi_{n_x}; \mathbf{R}) \left| \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} \right| \\ &= c_G(F_1(x_1), \dots, F_{n_x}(x_{n_x}); \mathbf{R}) \prod_{k=1}^{n_x} f_k(x_k) \end{aligned}$$

Assertion (b) can be proved by noting that the linear error propagation from $\boldsymbol{\xi}$ to \mathbf{x} has the following form:

$$\text{Cov}(\mathbf{x}) = \mathbf{J} \text{Cov}(\boldsymbol{\xi}) \mathbf{J}^T, \quad \text{with} \quad \mathbf{J} = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\xi}} \quad (2.47)$$

\mathbf{J} is diagonal, which reduces the error propagation to a rescaling of the variables. As correlation matrices are invariant under such a transformation, the correlation matrix of \mathbf{x} is equal to \mathbf{R} in first-order approximation. \square

The simplest proposal density $p(\mathbf{w})$ is the product of the marginal densities of the free variables:

$$p(\mathbf{w}) = \prod_{k=1}^{n_w} f_{w,k}(w_k) \quad (2.48)$$

The sampling algorithm is summarized in the box Algorithm 1.

Example 3: The modified sampler is illustrated using the linear model in Cencic and Frühwirth (2015, Example 4.3) which has nine flows and four processes. Variables x_3 and x_8 are unobserved. The other variables are modeled by lognormal priors. The effect of setting $\rho_{24} = 0.7$, $\rho_{56} = -0.4$, $\rho_{57} = -0.3$, $\rho_{67} = -0.6$ is shown in Figure 2.4. The posterior densities of x_5 and x_6 are significantly affected by the correlations, whereas the other variables hardly change. \square

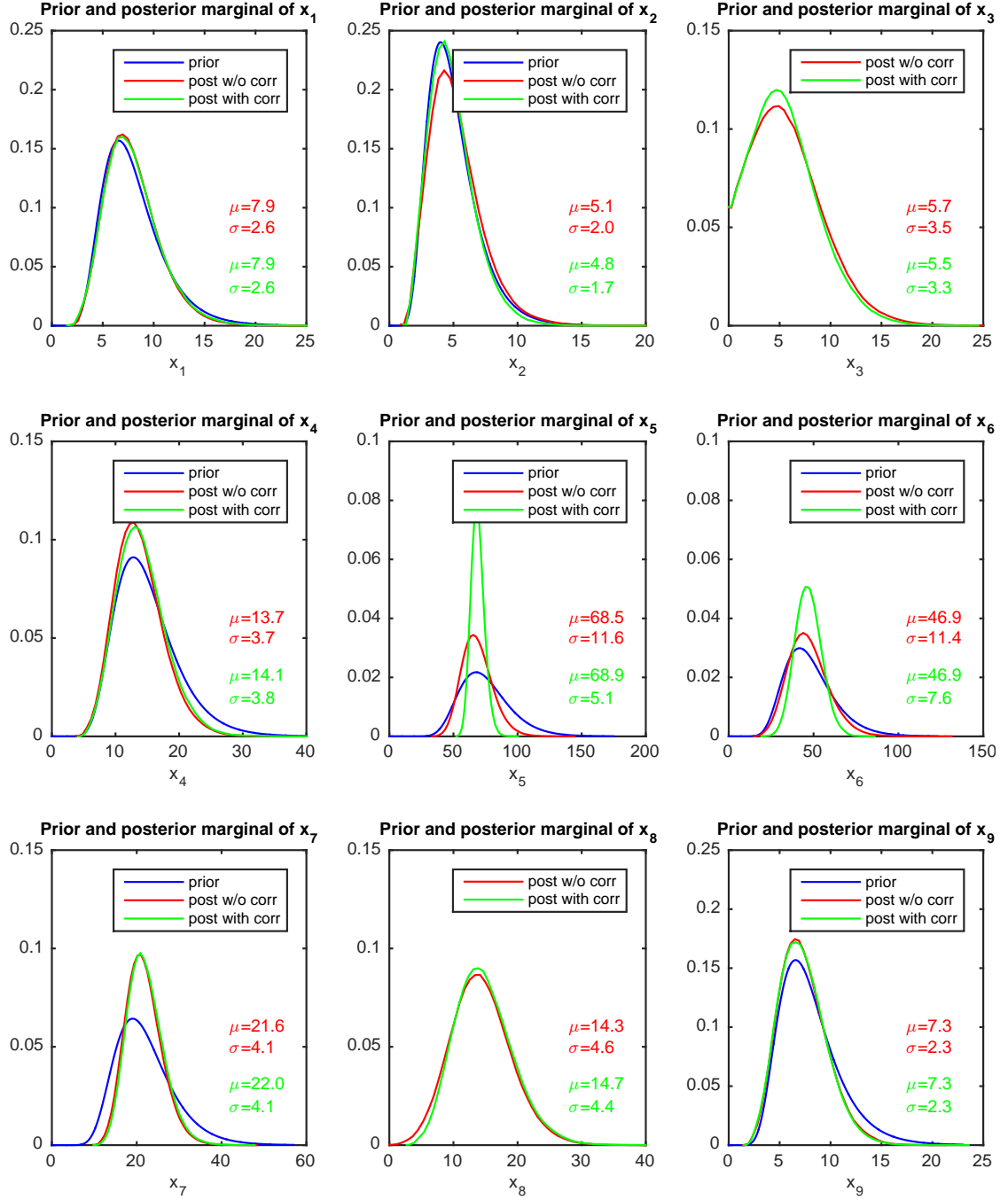


Figure 2.4: Priors and smoothed posterior marginals without and with correlations of all variables in Example 3. The improper priors of x_3 and x_8 cannot be shown.

Algorithm IS-CO

IS-CO1. Set $i = 1$ and choose the sample size L .

IS-CO2. Generate the starting value \mathbf{w}_1 by drawing independent random numbers from the prior marginal distribution of the free variables and compute $\mathbf{u}_1 = \mathbf{h}(\mathbf{w}_1)$, $\mathbf{x}_1 = (\mathbf{u}_1; \mathbf{w}_1)$. If $g(\mathbf{x}_1) = 0$, repeat the procedure until $g(\mathbf{x}_1) > 0$.

IS-CO3. Generate a proposal value $\dot{\mathbf{w}}$ by drawing independent random numbers from the prior marginal distribution of the free variables and compute $\dot{\mathbf{u}} = \mathbf{h}(\dot{\mathbf{w}})$, $\dot{\mathbf{x}} = (\dot{\mathbf{u}}; \dot{\mathbf{w}})$.

IS-CO4. Compute the acceptance probability α :

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{g(\dot{\mathbf{x}}) V(\dot{\mathbf{w}}) p(\mathbf{w}_i)}{g(\mathbf{x}_i) V(\mathbf{w}_i) p(\dot{\mathbf{w}})} \right), \quad (2.49)$$

with g as in Eq. (2.46).

IS-CO5. Draw a uniform random number $\xi \in [0, 1]$.

IS-CO6. If $\xi \leq \alpha$, accept the proposal and set $\mathbf{x}_{i+1} = \dot{\mathbf{x}}$; otherwise set $\mathbf{x}_{i+1} = \mathbf{x}_i$.

IS-CO7. Increase i by 1. If $i < L$, go to IS-CO3, otherwise stop sampling.

Algorithm 1: Independence sampler for correlated observations

2.3.4 Robust Reconciliation and Gross Error Detection

All the methods for gross error detection mentioned in chapter 2.2.4 have in common that they rely explicitly or implicitly on the assumption of normally distributed observations. In the case of nonnormal distributions, the distributions of residuals or chi-square like statistics can in general no longer be computed explicitly, so that tests based on these statistics are not feasible. However, it turns out that M-estimators can be generalized to nonnormal and even asymmetric distributions, leading to a reconciliation procedure that is robust against outlying observations and does not require any prior assumptions on the distribution of the outliers.

In the case of normal observations, an M-estimator of location can be implemented as an iterated reweighted least-squares estimator (Hampel et al., 1986; Huber, 2004). The reweighting process is tantamount to rescaling or dilating the distribution of an outlying observation around its mean, which is also the mode and the

median in the normal case. The dilation can be immediately generalized to non-normal symmetric distributions, for instance Student's t -distribution, the Laplace distribution or the hyperbolic secant distribution, all of which fall off to zero less rapidly than the normal distribution. In the case of unimodal, but asymmetric distributions, however, the mode of the density is to be preferred over the mean as the parameter of location (Dalenius, 1965; Porzio et al., 2015) and therefore as the center of the dilation.³ This results in the following prescription for the dilation of an unimodal distribution with density $f(x)$ and mode m by a factor s :

$$f_s(x) = \frac{1}{s} f(m + (x - m)/s) \quad (2.50)$$

The mode is invariant under the dilation, and the standard deviation is multiplied by s . If the observation has to be positive, the dilated prior density can be truncated and renormalized (see Figure 2.5). If the prior density is improper, as in the case of a variable with only a positivity constraint, rescaling has no effect and is skipped. The user is free to implement a different dilation algorithm or to refrain from dilation altogether if suggested by the problem, especially if the distribution to be dilated is not unimodal or confined to a finite interval. A typical example is the Beta distribution which is confined to the interval $[0, 1]$ and may have modes both at 0 and at 1, in which case it is difficult to come up with a reasonable prescription.

The scaling factor s should be larger than 1 for outlying observations, and equal to or close to 1 for regular observations. In order to compute s , a measure of distance between the prior and the posterior marginals of the observed variables is introduced. In analogy to the normal case, the distance equals the difference of the mean values, divided by the standard deviation of the prior:

$$r_i = \frac{E_{\text{prior}}[x_i] - E_{\text{posterior}}[x_i]}{\sigma_{\text{prior}}[x_i]} \quad (2.51)$$

The weight w_i (a number between 0 and 1) of observation i is calculated from the distance r_i . In the normal case, applying a weight w to an observation means multiplying the prior standard deviation by $1/\sqrt{w}$. The equivalent for a nonnormal

³If the density attains its maximal value in an entire interval, the mode is defined as the center of the modal interval.

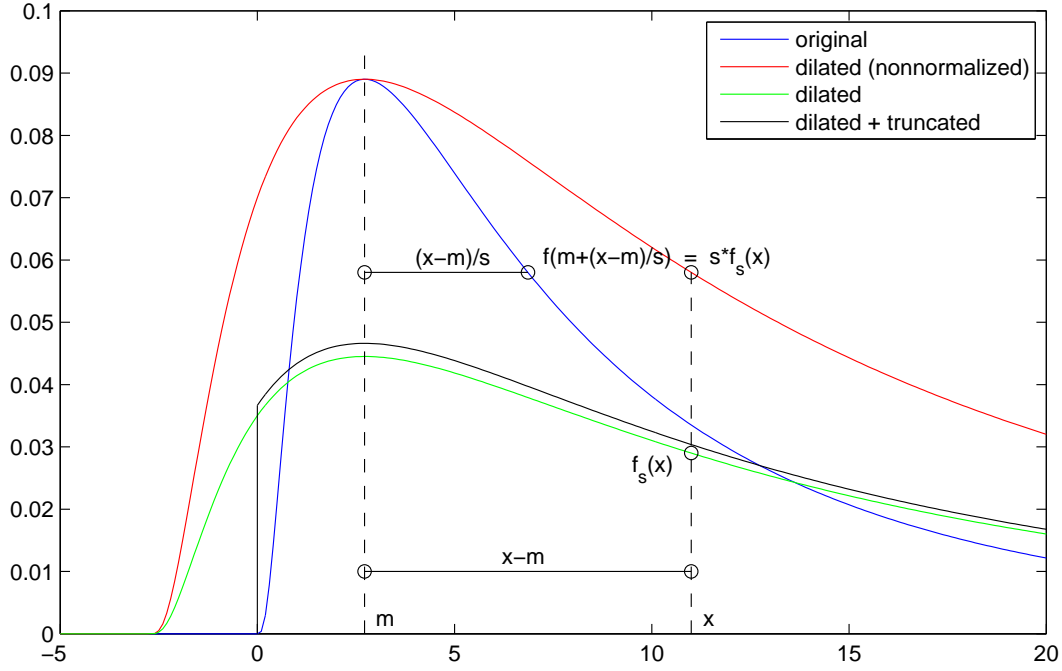


Figure 2.5: Visualization of the dilation procedure: The original prior (blue line) is dilated around the mode by factor s . The resulting dilated prior (red line) needs to be normalized by dividing its density by s . The normalized dilated prior (green line) is truncated at zero and renormalized (black line).

observation is to dilate its prior density by the factor $s = 1/\sqrt{w}$ around the mode.

For the problem considered here, one of the redescending M-estimators proposed in Frühwirth and Waltenberger (2008) has been used, because redescending M-estimators are particularly insensitive to gross errors or extreme outliers (Shevlyakov et al., 2008). The weight function has been slightly modified to ensure that $w(0) = 1$:

$$w(r_i) = \frac{\varphi(r_i)}{\varphi(r_i) + \varphi(c)} \frac{\varphi(0) + \varphi(c)}{\varphi(0)} \quad (2.52)$$

In principle, the function $\varphi(r)$ can be any symmetric standardized density function with infinite support. If φ is the standard normal density, the weight function decays to 0 very quickly with rising $|r|$ (see Figure 2.6, dotted line). To avoid this, φ has — somewhat arbitrarily — been chosen as the density of a Student- t distribution with 4 degrees of freedom. The cut value c can be interpreted as

a (fuzzy) boundary which discriminates between “good” and “bad”, or “inlying” and “outlying” observations. The weight function $w(|r|)$ with $c = 2.5$ is shown by the full line in Figure 2.6. If necessary, the number of degrees of freedom can be adapted to the problem at hand, or a different family of densities can be chosen.

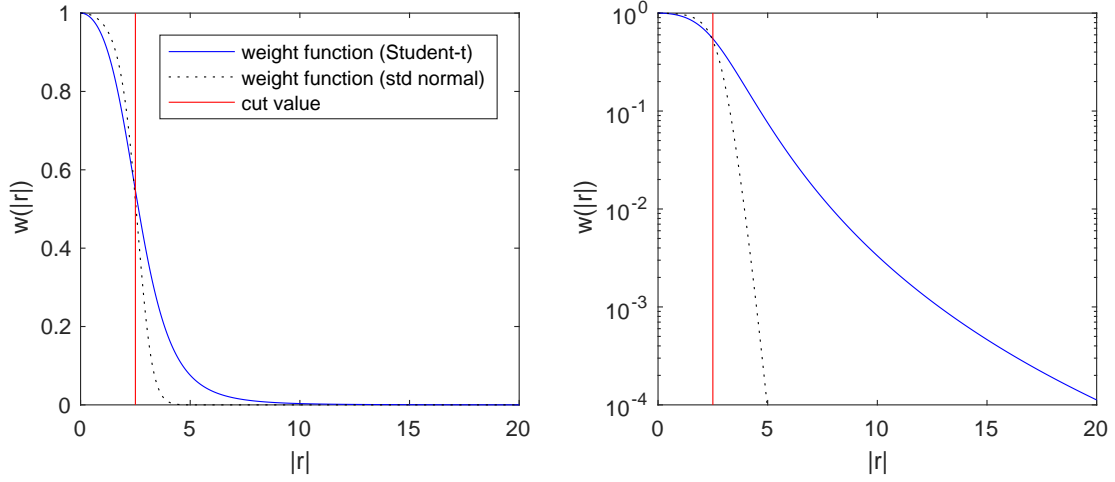


Figure 2.6: The weight function of Eq. (2.52) (full line) on a linear scale (left) and a logarithmic scale (right). The weight function based on the standard normal density (dotted line) is shown for comparison.

The sampler can be run as usual, but has to be iterated. After each run of the sampler, the distances r_i and the corresponding weights w_i are recomputed, the prior distributions are dilated accordingly, and the sampler is run again, until convergence of the weights. Note that the distances (Eq. (2.51)) have to be computed using the spread of the original, undilated prior.

Example 4: The robustified sampler is demonstrated on the linear model in Figure 2.7. The nonnormal priors of the five flows are as follows:

$$x_1 \sim \text{Trap}(160, 180, 190, 210)$$

$$x_2 \sim \text{Trap}(85, 105, 115, 135)$$

$$x_3 \sim \text{Tria}(145, 170, 195)$$

$$x_4 \sim \text{Unif}(70, 90)$$

$$x_5 \sim \text{Unif}(45, 65)$$

where $\text{Trap}(a, b, c, d)$ is the trapezoidal distribution with support $[a, d]$ and modal interval $[b, c]$. $\text{Tria}(a, b, c)$ is the triangular distribution with support $[a, c]$ and mode b . $\text{Unif}(a, b)$ is the uniform distribution with support $[a, b]$.

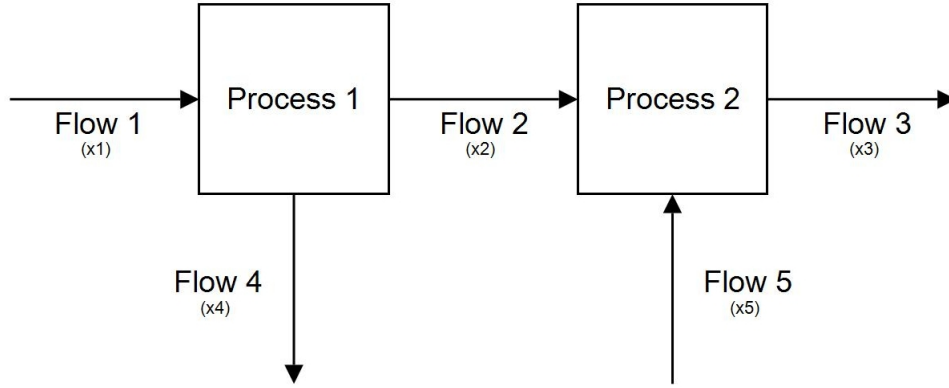


Figure 2.7: Flowsheet of the simple model used to demonstrate robust reconciliation and gross error detection.

The robustified sampler converges after three iterations, and the final weights are between 0.99 and 1. The acceptance rate⁴ is 58%. Figure 2.8 shows the prior and the reconciled (posterior) densities of the flows. If the prior distribution of x_2 is shifted by 50, the acceptance rate in the first iteration is only 2.5%, and the posterior distributions are strongly biased by the gross error. The robustified sampler estimator converges after 22 iterations, with a final acceptance rate of 63%. The weight of x_2 is 0.12, the other weights are between 0.98 and 1. Figure 2.9 shows the prior and the reconciled (posterior) densities of the flows. Note that the prior density of x_2 is dilated by a factor of about 3, whereas the other priors are hardly changed. The posteriors are very similar to the posteriors in Figure 2.8.

If the prior distribution of x_2 is shifted by 100, the acceptance rate is 0, and no weights can be computed. If this is the case, the priors have to be approximated by normal densities, and the final weights of the standard M-estimator are then used as the initial weights of the robustified sampler. The results are shown in Figure 2.10. The prior of x_2 is dilated by a factor of about 15. Note that with a shift of 100, x_2 has a very small weight, so that the posteriors look more similar to the posteriors without shift than with shift 50.

⁴Number of accepted proposed values in Markov chain in relation to sample size L .

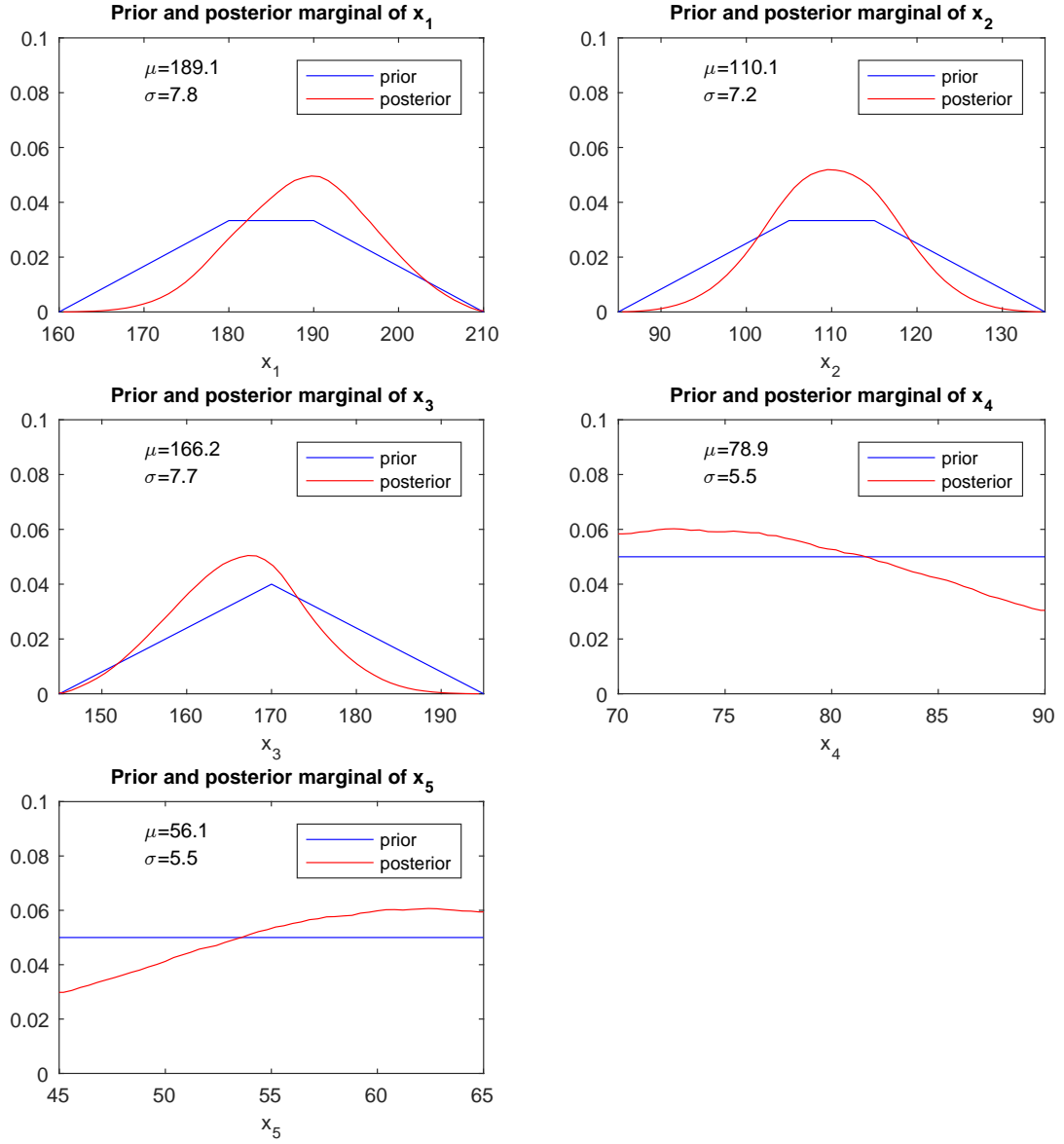


Figure 2.8: Priors and smoothed marginal posteriors of x_1, \dots, x_5 .

The computation of the initial weights by the M-estimator with approximating normal densities can also be used if the acceptance rate is not strictly 0, but small enough to raise suspicion of a gross error. If the prior of x_2 is shifted by 50, this procedure gives virtually identical results, but the robustified sampler converges much faster, after only three iterations. \square

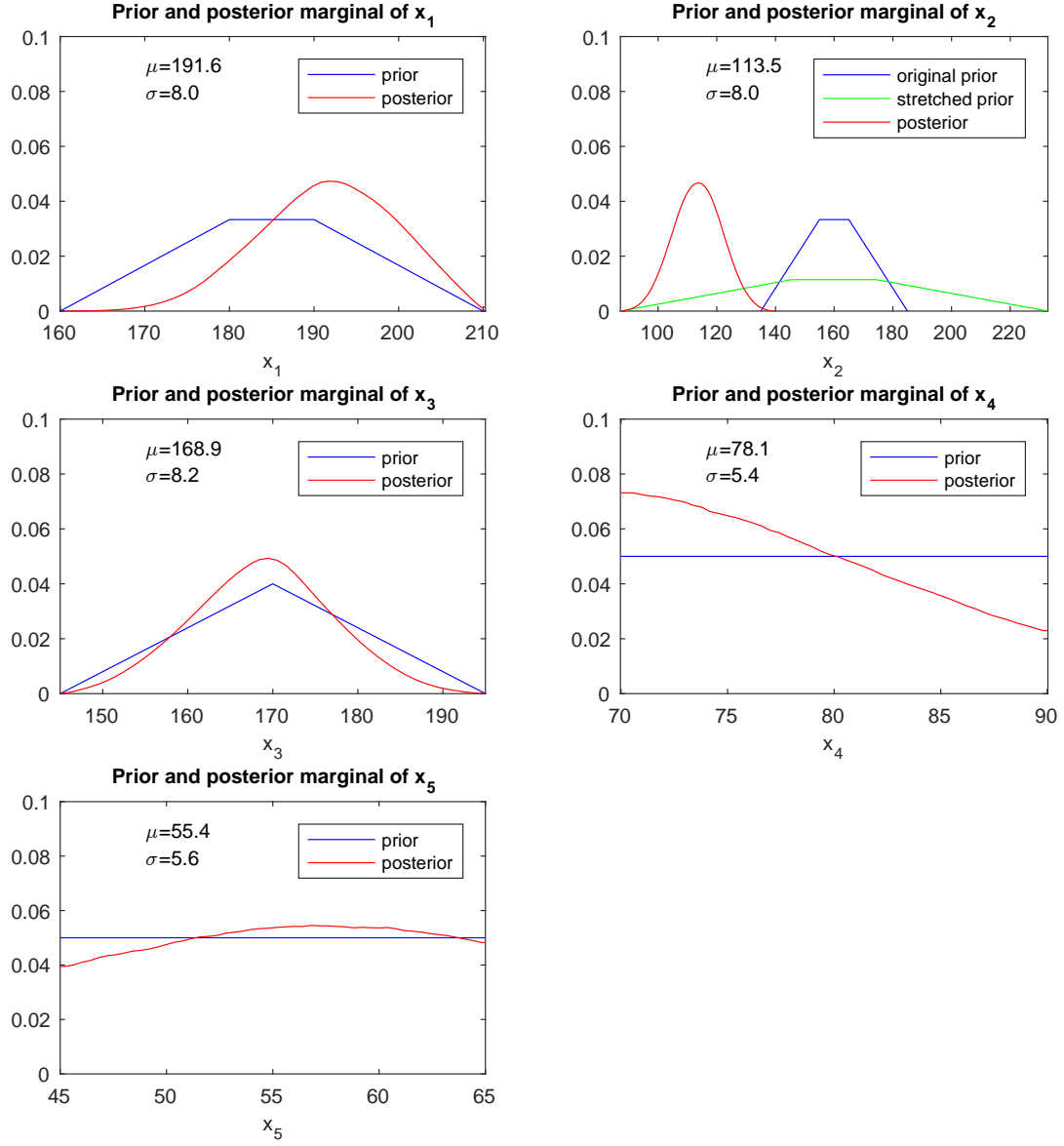


Figure 2.9: Priors and smoothed marginal posteriors of x_1, \dots, x_5 . The prior of x_2 is shifted by 50. The dilated prior of x_2 is also shown.

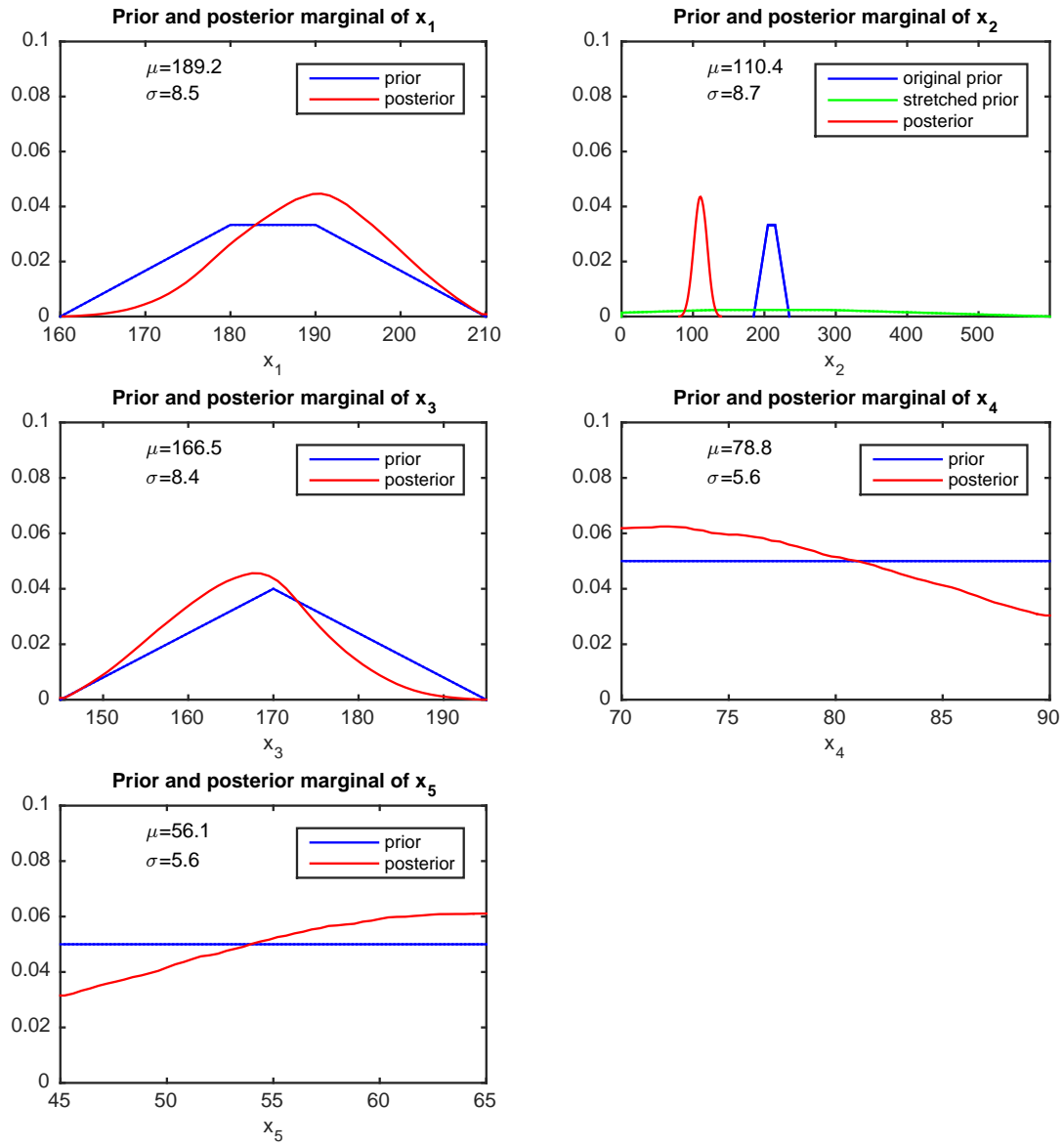


Figure 2.10: Priors and smoothed marginal posteriors of x_1, \dots, x_5 . The prior of x_2 is shifted by 100. The dilated prior of x_2 is also shown.

3 Conclusions and Outlook

The classical weighted least squares approach to DR, which assumes normally distributed observation errors, poses the problem that only in case of linear constraints the estimation errors of the adjusted observations are normally distributed again. In case of nonlinear constraints, however, the normally distributed results gained from linearization can differ substantially from the precise solution, especially if the involved uncertainties are large. Additionally, in scientific models, the assumption of normally distributed input data is often not justified.

Therefore, in Cencic and Frühwirth (2018), the Bayesian approach to DR for linear constraints, presented in Cencic and Frühwirth (2015), was further developed and extended to nonlinear constraints.

The advantages of the Bayesian approach are: First, arbitrary continuous pdfs can be used to describe the uncertainty of the observations. Second, even nonparametric estimators of the pdf are allowed, provided that it is possible to draw a random sample from them. Third, not only means, variances and covariances of observed and unobserved variables can be computed a posteriori, but also various other characteristics of the marginal posteriors, such as the mode, skewness, quantiles, and HPD intervals.

The main idea of the method is to restrict the joint prior probability distribution of the observed variables with model constraints to get a joint posterior probability distribution. The derived joint posterior of the free observed variables is sampled by a MCMC method using the independence sampler, and the dependent variables (observed and unobserved) are computed from this sample.

By construction, all individual elements of the Markov chain satisfy the constraints, but the sample mean in the nonlinear case in general does not. If a representative value of the posterior distribution satisfying the constraints is required, the element

of the Markov chain with the smallest distance from the sample mean can be selected. Possible distance measures are, among others, the L_1 , the L_2 and the L_∞ distance.

It was shown that, for nonlinear constraints, it is essential to consider the metric of the constraint manifold. The posterior density derived in Cencic and Frühwirth (2018) contains the term $V(\mathbf{w})$ (Eq. (2.28)), which in the linear case is constant and cancels. If $V(\mathbf{w})$ is neglected in the nonlinear case, the posterior is no longer invariant under the choice of the free variables.

For the Bayesian approach, in chapter 2.3.3 it was shown how to incorporate correlated observations, and in chapter 2.3.4 how to perform robust gross error detection with the help on redescending M-estimators. Note that both topics are not covered in the three papers of this thesis.

In subsequent work, the Bayesian method will be applied to more extensive real life examples in order to compare the results to alternative approaches such as classical weighted least squares, fuzzy sets (possibilistic approach to DR), and another Bayesian approach that was published recently in Lupton and Allwood (2018).

Bibliography

- Alhaj-Dibo, M., Maquin, D., and Rago, J. Data reconciliation: A robust approach using a contaminated distribution. *Control Engineering Practice*, 16:159–170, 2008.
- Almasy, G. and Sztano, T. Checking and Correction of Measurements on the Basis of Linear System Model. *Problems of Control and Information Theory*, 4:57–69, 1975.
- Arora, N. and Biegler, L. Redescending estimators for data reconciliation and parameter estimation. *Computers and Chemical Engineering*, 25:1585–1599, 2001.
- Bagajewicz, M. *Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations*. McGraw-Hill, New York, 2010.
- Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall, Boca Raton, 2011.
- Brunner, P. and Rechberger, H. *Handbook of Material Flow Analysis: For Environmental, Resource, and Waste Engineers, Second Edition*. CRC Press, Boca Raton, Florida, 2016.
- Cencic, O. Nonlinear data reconciliation in material flow analysis with software STAN. *Sustainable Environment Reserach*, 26:291–298, 2016.
- Cencic, O. and Frühwirth, R. A general framework for data reconciliation—Part I: Linear constraints. *Computers and Chemical Engineering*, 75:196–208, 2015.
- Cencic, O. and Frühwirth, R. Data reconciliation of nonnormal observations with nonlinear constraints. *Journal of Applied Statistics*, 2018.
- Cencic, O. and Rechberger, H. Material flow analysis with software STAN. *Journal of Environmental Engineering and Management*, 18(1):3–7, 2008.

- Chib, S. and Greenberg, E. Understanding the Metropolis–Hastings Algorithm. *The American Statistician*, 49(4):327–335, 1995.
- Chinneck, J. The Constraint Consensus Method for Finding Approximately Feasible Points in Nonlinear Programs. *INFORMS Journal on Computing*, 16(3): 255–265, 2004.
- Dalenius, T. The Mode—A Neglected Statistical Parameter. *Journal of the Royal Statistical Society A*, 128(1):110–117, 1965.
- Dubois, D., Fargier, H., Ababou, M., and Guyonnet, D. A fuzzy constraint-based approach to data reconciliation in material flow analysis. *International Journal of General Systems*, 43(8):787–809, 2014.
- Dzubur, N., Sunanta, O., and Laner, D. A fuzzy set-based approach to data reconciliation in material flow modeling. *Applied Mathematical Modelling*, 43: 464–480, 2017.
- Frühwirth, R. and Waltenberger, W. Redescending M-estimators and Deterministic Annealing. *Austrian Journal of Statistics*, 37(3&4):301–317, 2008.
- Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York, 1986.
- Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–107, 1970.
- Hu, M. and Shao, H. Theory Analysis of Nonlinear Data Reconciliation and Application to a Coking Plant. *Industrial & Engineering Chemistry Research*, 45: 8973–8984, 2006.
- Huber, P. *Robust Statistics: Theory and Methods*. John Wiley & Sons, New York, 2004.
- Johnston, L. and Kramer, M. Maximum Likelihood Data Rectification: Steady-State Systems. *AIChE Journal*, 24(11):2415–2426, 1995.
- Kuehn, D. and Davidson, H. Computer control - II. Mathematics of control. *Chemical Engineering Progress*, 57(6), 1961.

- Liu, J. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6:113–119, 1996.
- Liu, J. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2004.
- Llanos, C., Sánchez, M., and Maronna, R. Robust Estimators for Data Reconciliation. *Industrial & Engineering Chemistry Research*, 54:5096–5105, 2015.
- Lupton, R. C. and Allwood, J. M. Incremental Material Flow Analysis with Bayesian Inference. *Journal of Industrial Ecology*, in press, 2018.
- Madron, F. *Process Plant Performance*. Ellis Horwood, New York, 1992.
- Madron, F., Veverka, V., and Venecek, V. Statistical Analysis of Material Balance of a Chemical Reactor. *AIChE Journal*, 57(6):482–486, 1977.
- Metropolis, N. et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21:1087–1091, 1953.
- Narasimhan, S. and Jordache, C. *Data Reconciliation and Gross Error Detection. An Intelligent Use of Process Data*. Gulf Publishing Company, Houston, 2000.
- Nelder, J. and Mead, R. A Simplex Method for Function Minimization. *Computer Journal*, 7:308–313, 1965.
- Nelsen, R. *An Introduction to Copulas*. Springer, New York, 2006.
- O’Neill, B. *Semi-Riemannian Geometry*. Academic Press, San Diego, 1983.
- Özyurt, D. and Pike, R. Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. *Computers and Chemical Engineering*, 28:381–402, 2004.
- Porzio, G., Ragozini, G., Liebscher, S., and Kirschstein, T. Minimum volume peeling: A robust nonparametric estimator of the multivariate mode. *Computational Statistics & Data Analysis*, 93:456–468, 2015.
- Robert, C. and Casella, C. *Monte Carlo Statistical Methods*. Springer, New York, 2004.

- Romagnoli, J. and Sanchez, M. *Data Processing and Reconciliation for Chemical Process Operations*. Academic Press, San Diego, 2000.
- Shevlyakov, G., Morgenthaler, S., and Shurygin, A. Redescending M-estimators. *Journal of Statistical Planning and Inference*, 138(10):2906 – 2917, 2008.
- Tamhane, A. and Mah, R. Data Reconciliation and Gross Error Detection in Chemical Process Networks. *Technometrics*, 27(4):409–422, 1985.
- Yuan, Y., Khatibisepehr, S., Huang, B., and Li, Z. Bayesian method for simultaneous gross error detection and data reconciliation. *AIChE Journal*, 61(10): 3232–3248, 2015.

Notation

a	parameter of a distribution
\mathbf{A}	coefficient matrix of observed variables
b	parameter of a distribution
\mathbf{B}	coefficient matrix of unobserved variables
c	parameter of a distribution
\mathbf{c}	vector of aggregated constant quantities
C	copula
\mathbf{C}	coefficient matrix of constant quantities
c_G	probability density function of the Gaussian copula
C_G	cumulative distribution function of the Gaussian copula
d	parameter of a distribution
\mathbf{d}	vector, which is a submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{D}	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
dS	differential element of the constraint manifold S
\mathbf{D}_1	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{e}	vector, which is a submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
E	expectation
\mathbf{E}	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{E}_1	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{E}_2	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
f	distribution function
F	cumulative distribution function
f_s	dilated distribution function

f_u	joint prior density of dependent observed variables
\mathbf{f}_u	vector, which is a submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
f_w	joint prior density of free observed variables
f_z	scalar that is either 0 or 1
\mathbf{F}_0	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{F}_1	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
\mathbf{F}_2	submatrix of $\text{RREF}(\mathbf{B}, \mathbf{A}, \mathbf{c})$
g	joint prior density of correlated observed variables
\mathbf{G}	vector of equality constraint equations
\mathbf{h}	vector of functions of free observed variables to compute dependent observed variables
\mathbf{H}	partial derivatives of function \mathbf{h} with respect to \mathbf{w}
i	counter of iterations
\mathbf{I}	identity matrix
J	objective function to be minimized
\mathbf{J}	partial derivatives of \mathbf{x} with respect to $\boldsymbol{\xi}$
\mathbf{J}_x	partial derivatives of equality constraints \mathbf{G} with respect to observed variables \mathbf{x}
\mathbf{J}_y	partial derivatives of equality constraints \mathbf{G} with respect to unknown variables \mathbf{y}
k	counter of observed variables
\mathbf{k}	vector of functions of free observed variables to compute observable unknown variables
L	sample size
L_1	Manhattan distance
L_2	Euclidean distance
L_∞	Chebyshev distance
m	mode
\mathbf{m}	vector containing one element of the Markov chain

n	number of variables
\mathcal{N}	normal distribution
n_c	number of independent equality constraint equations
n_u	number of dependent observed variables
n_w	number of free observed variables
n_x	number of observed variables
n_{y_o}	number of observable unknown variables
\mathbf{O}	null matrix
p	proposal density for MCMC algorithm
\mathbf{Q}	variance-covariance matrix of estimated observable unknown and reconciled observed variables
$\mathbf{Q}_{\hat{x}}$	variance-covariance matrix of reconciled observed variables
$\mathbf{Q}_{\tilde{x}}$	variance-covariance matrix of observations of observed variables
$\mathbf{Q}_{\hat{y}_o}$	variance-covariance matrix of estimated observable unknown variables
r	measure of distance between prior and posterior distribution
\mathbf{R}	correlation matrix
t	transfer coefficient
s	dilation factor
S	constraint manifold; domain of the constraint manifold
\mathbf{u}	vector of dependent observed variables
V	square root of the determinant of the metric tensor
w	weight, a number in the interval $[0, 1]$
\mathbf{w}	vector of free observed variables
W	domain of the free variables \mathbf{w}
x	observed variable
\mathbf{x}	vector of observed variables
\mathbf{x}_n	vector of nonredundant observed variables
\mathbf{x}_{r_1}	vector of dependent redundant observed variables

\mathbf{x}_{r_2}	vector of free redundant observed variables
\mathbf{y}	vector of unknown variables
\mathbf{y}_o	vector of observable unknown variables; subset of \mathbf{y}
\mathbf{y}_{u_1}	vector of 'dependent' unobservable unknown variables; subset of \mathbf{y}
\mathbf{y}_{u_2}	vector of 'free' unobservable unknown variables; subset of \mathbf{y}
\mathbf{z}	vector of constant quantities
$\mathbf{0}$	null vector

Greek symbols:

α	probability of acceptance
γ	gamma distribution
Γ	complete gamma function
δ	vector of measurement biases of observed variables
ϵ	vector of random errors of observed variables
μ	mean value
$\boldsymbol{\mu}_x$	vector of true values of observed variables
σ	standard deviation
$\boldsymbol{\sigma}$	vector of standard deviations
φ	symmetric standardized density function with infinite support
Φ^{-1}	inverse distribution function of the multi-dimensional standard normal distribution
ξ	uniform random numbers in the interval $[0,1]$
$\boldsymbol{\xi}$	vector of uniform random numbers in the interval $[0,1]$
π	joint posterior distribution
ρ	correlation coefficient

Superscripts:

- \sim measured/observed values
- $\hat{}$ estimated values
- \cdot proposed values

Summary of unified notation:

Thesis	Paper 1	Paper 2	Paper 3
\mathbf{d}	\mathbf{A}_{rz}		
\mathbf{D}	\mathbf{A}_{rx}		
\mathbf{e}	\mathbf{A}_{cz}^*		
\mathbf{E}	\mathbf{A}_{cx}^*		
f_u		f_d	
f_w		f_f	
\mathbf{G}	\mathbf{f}		
J	$F(\mathbf{x})$		$F(\mathbf{x})$
n		N	
n_c		K	
n_x	n	I	
n_{y_o}	o	J	n_y
$\mathbf{Q}_{\hat{x}}$	\mathbf{Q}_x		
$\mathbf{Q}_{\tilde{x}}$	\mathbf{Q}		Σ
$\mathbf{Q}_{\hat{y}_o}$	\mathbf{Q}_{y^*}		
\mathbf{u}		\mathbf{v}_d	
\mathbf{w}		\mathbf{v}_f	
$\dot{\mathbf{w}}$		$\hat{\mathbf{w}}$	
\mathbf{y}_o	\mathbf{y}^*	\mathbf{y}	\mathbf{y}

Abbreviations

cdf	cumulative distribution function
Cov	covariance
DR	data reconciliation
e.g.	for example
i.e.	that is; in other words
IS-CO	independence sampler - correlated observations
MCMC	Markov chain Monte Carlo
MFA	material flow analysis
pdf	probability density function
RREF	reduced row echelon form
STAN	software for subSTance flow ANalysis
Trap	trapezoidal distribution function
Tria	triangular distribution function
Unif	uniform distribution function

Appendix

Paper 1:

Nonlinear data reconciliation in material flow analysis with software STAN

Oliver Cencic

Sustainable Environment Research 2016, 26 (6)

DOI: 10.1016/j.serj.2016.06.00



Contents lists available at ScienceDirect

Sustainable Environment Research

journal homepage: www.journals.elsevier.com/sustainable-environment-research/

Technical note

Nonlinear data reconciliation in material flow analysis with software STAN



Oliver Cencic

Institute for Water Quality, Resource and Waste Management, Vienna University of Technology, Vienna A-1040, Austria

ARTICLE INFO

Article history:
Received 18 January 2016
Received in revised form
26 March 2016
Accepted 19 April 2016
Available online 16 June 2016

Keywords:
Material flow analysis
STAN
Nonlinear data reconciliation
Error propagation

ABSTRACT

STAN is a freely available software that supports Material/Substance Flow Analysis (MFA/SFA) under the consideration of data uncertainties. It is capable of performing nonlinear data reconciliation based on the conventional weighted least-squares minimization approach, and error propagation. This paper summarizes the mathematical foundation of the calculation algorithm implemented in STAN and demonstrates its use on a hypothetical example from MFA.

© 2016 Chinese Institute of Environmental Engineering, Taiwan. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Material flow analysis (MFA) is a systematic assessment of the flows and stocks of materials within a system defined in space and time [1]. Due to the fact that direct measurements are scarce for macro-scale MFA (e.g., regions, countries), additional data are often taken from other sources of varying quality such as official statistics, reports or expert estimates [2]. Because all these sources are subject to uncertainties, practitioners are frequently confronted with data that are in conflict with model constraints. These contradictions can be resolved by applying data reconciliation, a statistical method that helps to find the most likely values of measured quantities. While most of the model constraints are linear (e.g., mass flow balances of individual processes), in some cases also nonlinear equations (e.g., concentration or transfer coefficient equations) are involved leading to nonlinear data reconciliation problems.

A variety of techniques has been developed in the last 50 years to deal with these problems. Most of them are based on a weighted least squares minimization of the measurement adjustments subject to constraints involving reconciled (measured), unknown (unmeasured) and fixed variables [3–5]. This approach is also

implemented in STAN (Fig. 1), a freely available software that supports MFA/SFA (Substance Flow Analysis) and enables the consideration of data uncertainties [6]. The calculation algorithm of STAN allows to make use of redundant information to reconcile uncertain “conflicting” data (with data reconciliation) and subsequently to compute unknown variables including their uncertainties (with error propagation). For more detailed information about the software, see appendix A or visit the website www.stan2web.net.

In this paper, the mathematical foundation of the calculation algorithm implemented in STAN is explained and its application demonstrated on a hypothetical example from MFA. A detailed description of the notation used in this paper can be found in appendix B.

2. Example

As example, a simple model with seven mass flows and three processes (Fig. 2) is investigated where the mass flows are represented by the variables m_1 to m_7 . Additionally, a (nonconstant) transfer coefficient tc_{34} is given defining how much of flow 3 is transferred into flow 4.

The constraints of this model are the mass balances of the three processes (linear equations f_1 to f_3) and the additional relation between flow 3 and flow 4 defined by the transfer coefficient tc_{34} (nonlinear equation f_4):

E-mail address: oliver.cencic@tuwien.ac.at.

Peer review under responsibility of Chinese Institute of Environmental Engineering.

<http://dx.doi.org/10.1016/j.serj.2016.06.002>

2468–2039/© 2016 Chinese Institute of Environmental Engineering, Taiwan. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

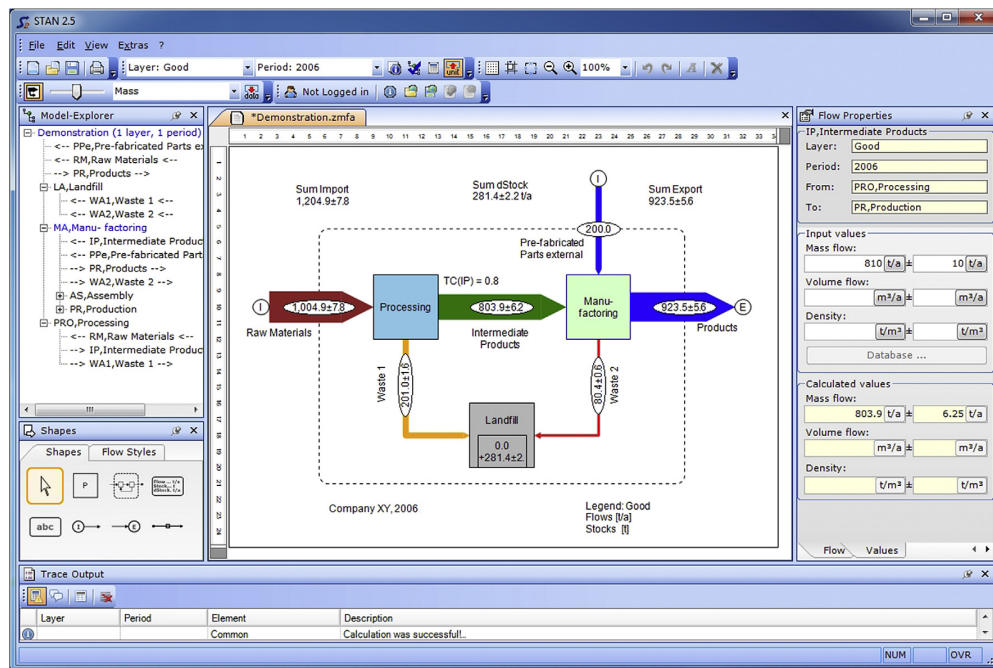


Fig. 1. The user interface of software STAN.

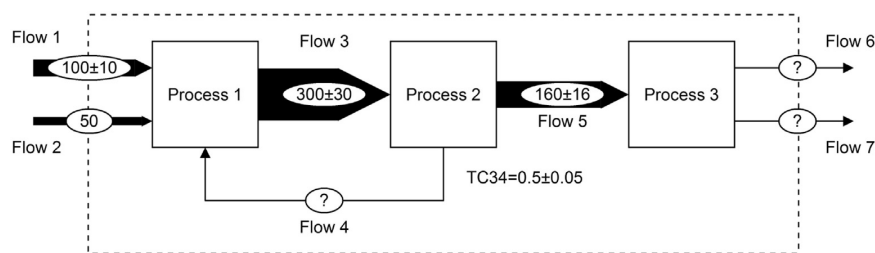


Fig. 2. Flowsheet example with seven flows and three processes.

$$f_1 = m_1 + m_2 + m_4 - m_3 = 0,$$

$$f_2 = m_3 - m_4 - m_5 = 0,$$

$$f_3 = m_5 - m_6 - m_7 = 0,$$

$$f_4 = m_4 - m_3 \cdot tc_{34} = 0.$$

Even though the nonlinearity in this example is marginal (only equation f_4 is nonlinear), it is sufficient to demonstrate the calculation procedure and the necessary preprocessing of the equation system in the nonlinear case.

The measured variables m_1 , m_3 , m_5 and tc_{34} are assumed to be normally distributed specified by their mean values (measurements) and standard errors, while variable m_2 is assigned a constant value. The variables m_4 , m_6 and m_7 are unknown. The respective values of the variables are listed in Table 1 and displayed in Fig. 2.

Trying to compute the unmeasured values without considering the uncertainties of the measurements, the following problems are encountered:

Table 1

List of available input data.

Variable name	Measurement of mass flow	Standard error of measurement
m_1	100	10
m_2	50	0
m_3	300	30
m_4	?	?
m_5	160	16
m_6	?	?
m_7	?	?
tc_{34}	0.5	0.05

Firstly, there are multiple ways to compute m_4 with different results. Calculated from the balance equation of process 1 (f_1), $m_4 = 150$, from the balance equation of process 2 (f_2), $m_4 = 140$, or from the transfer coefficient equation (f_4), $m_4 = 150$. Because one of the values is contradicting the others, it has to be checked if the contradiction can be resolved by adjusting (reconciling) the measured (uncertain) data, or if there are really conflicting constant values involved in the problem.

Secondly, there is not enough information given to compute m_6 and m_7 . That does not look like a major issue but could prevent the automatic computation of other unknown variables when using linear algebra.

In the following the mathematical foundation of the nonlinear data reconciliation algorithm will be derived step by step and each step immediately applied to the presented example.

3. Mathematical foundation

3.1. Theory (Part 1)

The general data reconciliation problem can be formulated as a weighted least-squares optimization problem by minimizing the objective function (Eq. (1)) subject to equality constraints (Eq. (2)).

$$F(\mathbf{x}) = (\tilde{\mathbf{x}} - \mathbf{x})^T \mathbf{Q}^{-1} (\tilde{\mathbf{x}} - \mathbf{x}), \quad (1)$$

$$\mathbf{f}(\mathbf{y}, \mathbf{x}, \mathbf{z}) = \mathbf{0}. \quad (2)$$

$\tilde{\mathbf{x}}$ is the vector of measurements of random variables with the true but unknown mean values μ_x . These measurements $\tilde{\mathbf{x}}$ are subject to measurement errors \mathbf{e} (Eq. (3)), which are assumed to be normally distributed with zero mean and known variance-covariance matrix \mathbf{Q} (Eq. (4)).

$$\tilde{\mathbf{x}} = \mu_x + \mathbf{e}, \quad (3)$$

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}). \quad (4)$$

\mathbf{x} is the vector of reconciled (adjusted) measurements, which are the best estimates of μ_x computed by data reconciliation (Eq. (5)). \mathbf{x} has to fulfill the model constraints.

$$\mathbf{x} = \hat{\mu}_x. \quad (5)$$

\mathbf{y} is the vector of estimates of unknown (unmeasured) random variables, and \mathbf{z} is a vector of constant values.

Nonlinear data reconciliation problems which contain only equality constraints can be solved by using iterative techniques based on successive linearization and analytical solution of the linear reconciliation problem [4]. In STAN even linear constraints will be treated as if they were nonlinear. In these cases, the solution will be found after two iterations.

A linear approximation of the nonlinear constraints can be obtained from a first order Taylor series expansion of Eq. (2) at the expansion point $\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z}$:

$$\mathbf{f}(\mathbf{y}, \mathbf{x}, \mathbf{z}) \approx \mathbf{J}_y(\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z})(\mathbf{y} - \hat{\mathbf{y}}) + \mathbf{J}_x(\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z})(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{f}(\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z}) = \mathbf{0}, \quad (6)$$

or

$$\mathbf{f}(\mathbf{y}, \mathbf{x}, \mathbf{z}) \approx \begin{pmatrix} \hat{\mathbf{J}}_y & \hat{\mathbf{J}}_x & \hat{\mathbf{f}} \end{pmatrix} \begin{pmatrix} \mathbf{y} - \hat{\mathbf{y}} \\ \mathbf{x} - \hat{\mathbf{x}} \\ 1 \end{pmatrix} = \mathbf{0}. \quad (7)$$

As the initial estimates $\hat{\mathbf{x}}$ of the reconciled measurements \mathbf{x} the measured values $\tilde{\mathbf{x}}$ are used. The initial estimates $\hat{\mathbf{y}}$ of the unknown values \mathbf{y} have to be provided by the user. The Jacobi matrices $\mathbf{J}_y, \mathbf{J}_x$ (derivations of \mathbf{f} with respect to the unknown and measured variables, respectively) and the vector of equality constraints \mathbf{f} are

evaluated with respect to $\hat{\mathbf{y}}, \hat{\mathbf{x}}$ and \mathbf{z} leading to $\hat{\mathbf{J}}_y, \hat{\mathbf{J}}_x$ and $\hat{\mathbf{f}}$, where the latter contains the constraints residuals.

Linearizing the nonlinear constraints and assuming the input data to be normally distributed ensures the results of data reconciliation to be also normally distributed. The limitations of this approach are discussed in Section 4.

3.2. Example (Part 1)

Grouping the variables into unknown, measured and fixed variables, $\mathbf{y} = (m_4, m_6, m_7)^T$, $\mathbf{x} = (m_1, m_3, m_5, tc_{34})^T$ and $\mathbf{z} = (m_2) = (50)$. As initial estimates $\hat{\mathbf{x}}$ of the reconciled measurements \mathbf{x} , the measurements $\tilde{\mathbf{x}}$ themselves are taken:

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} = (\tilde{m}_1, \tilde{m}_3, \tilde{m}_5, \tilde{tc}_{34})^T = (100, 300, 160, 0.5)^T.$$

In this example, the standard errors of the individual measurements are assumed to be 10% of the measured values. Because, in general, the measurement errors are assumed to be independent, i.e., there are no covariances, the variance-covariance matrix \mathbf{Q} has nonzero entries only in the diagonal, representing the variance of the measurement errors. Therefore, the variance-covariance matrix is

$$\mathbf{Q} = \begin{pmatrix} 10^2 & 0 & 0 & 0 \\ 0 & 30^2 & 0 & 0 \\ 0 & 0 & 16^2 & 0 \\ 0 & 0 & 0 & 0.05^2 \end{pmatrix}.$$

The choice of the covariance matrix \mathbf{Q} influences the results of data reconciliation considerably. Thus, the measurement or estimation error has to be determined as precisely as possible.

The initial estimates $\hat{\mathbf{y}}$ of the unknown values \mathbf{y} are computed from f_1 and f_3 with

$$\hat{m}_4 = \tilde{m}_3 - \tilde{m}_1 - \tilde{m}_2,$$

$$\hat{m}_6 = \tilde{m}_7 = \frac{\tilde{m}_5}{2},$$

leading to

$$\hat{\mathbf{y}} = (\hat{m}_4, \hat{m}_6, \hat{m}_7)^T = (150, 80, 80)^T.$$

The coefficient matrix $(\mathbf{J}_y \ \mathbf{J}_x \ \mathbf{f})$ is evaluated with respect to $\hat{\mathbf{y}}, \hat{\mathbf{x}}$ and \mathbf{z} with

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} = \begin{pmatrix} m_1 + m_2 + m_4 - m_3 \\ m_3 - m_4 - m_5 \\ m_5 - m_6 - m_7 \\ m_4 - m_3 \cdot tc_{34} \end{pmatrix},$$

$$\mathbf{J}_y = \frac{\partial \mathbf{f}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial f_1}{\partial m_4} & \frac{\partial f_1}{\partial m_6} & \frac{\partial f_1}{\partial m_7} \\ \frac{\partial f_2}{\partial m_4} & \frac{\partial f_2}{\partial m_6} & \frac{\partial f_2}{\partial m_7} \\ \frac{\partial f_3}{\partial m_4} & \frac{\partial f_3}{\partial m_6} & \frac{\partial f_3}{\partial m_7} \\ \frac{\partial f_4}{\partial m_4} & \frac{\partial f_4}{\partial m_6} & \frac{\partial f_4}{\partial m_7} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & -1 \\ 1 & 0 & 0 \end{pmatrix},$$

$$J_x = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial m_1} & \frac{\partial f_1}{\partial m_3} & \frac{\partial f_1}{\partial m_5} & \frac{\partial f_1}{\partial tc_{34}} \\ \frac{\partial f_2}{\partial m_1} & \frac{\partial f_2}{\partial m_3} & \frac{\partial f_2}{\partial m_5} & \frac{\partial f_2}{\partial tc_{34}} \\ \frac{\partial f_3}{\partial m_1} & \frac{\partial f_3}{\partial m_3} & \frac{\partial f_3}{\partial m_5} & \frac{\partial f_3}{\partial tc_{34}} \\ \frac{\partial f_4}{\partial m_1} & \frac{\partial f_4}{\partial m_3} & \frac{\partial f_4}{\partial m_5} & \frac{\partial f_4}{\partial tc_{34}} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -tc_{34} & 0 & -m_3 \end{pmatrix},$$

leading to

$$(\hat{J}_y \hat{J}_x \hat{f}) = \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & -1 & 0 & -10 \\ 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -0.5 & 0 & -300 & 0 \end{array} \right).$$

3.3. Theory (Part 2)

If by transformation of a nonlinear set of equations at least one equation can be found that contains no unknown and at least one measured variable, data reconciliation can be performed to improve the accuracy of the measurements. Madron [7] proposed to apply the Gauss-Jordan elimination to the coefficient matrix of a linear or linearized set of equations in order to decouple the unknown variables from the data reconciliation process. The structure of the resulting matrix, known as the reduced row echelon form (rref) or canonical form, can also be used to classify the involved variables, detect contradictions in constant input data, and eliminate dependent equations from the constraints.

A matrix is in rref when it satisfies the following conditions [8]:

- All zero rows (if there are any) are at the bottom of the matrix.
- The leading entry of each nonzero row after the first occurs to the right of the leading entry of the previous row.
- The leading entry in any nonzero row is 1.
- All entries in the column above and below a leading 1 are zero.

Fig. 3 shows a numerical example of a matrix in rref.

To transform a matrix into rref, the following elementary row operations can be applied to the matrix [9]:

- Interchange two rows.
- Multiply any row by a nonzero element.
- Add a multiple of one row to another.

A	y			x			z
c	1	0	X	0	0	X	X
	0	1	X	0	0	X	X
r	0	0	0	1	0	X	X
	0	0	0	0	1	X	X
t	0	0	0	0	0	0	0
	0	0	0	0	0	0	0

Fig. 3. Example of a matrix in reduced row echelon form split into submatrices. X entries can be any number.

If the Gauss-Jordan elimination is applied to the coefficient matrix $(\hat{J}_y \hat{J}_x \hat{f})$, the resulting matrix, which is in rref, can be split into the following submatrices:

$$\mathbf{A} = \text{rref}(\hat{J}_y \hat{J}_x \hat{f}) = \begin{pmatrix} \mathbf{A}_{cy} & \mathbf{A}_{cx} & \mathbf{A}_{cz} \\ \mathbf{0} & \mathbf{A}_{rx} & \mathbf{A}_{rz} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{tz} \end{pmatrix} = (\mathbf{A}_y \mathbf{A}_x \mathbf{A}_z) = \begin{pmatrix} \mathbf{A}_c \\ \mathbf{A}_r \\ \mathbf{A}_t \end{pmatrix}. \quad (8)$$

The corresponding transformed linearized set of equations \mathbf{g} can then be expressed as

$$\mathbf{g}(\mathbf{y}, \mathbf{x}, \mathbf{z}) \approx \begin{pmatrix} \mathbf{A}_{cy} & \mathbf{A}_{cx} & \mathbf{A}_{cz} \\ \mathbf{0} & \mathbf{A}_{rx} & \mathbf{A}_{rz} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{tz} \end{pmatrix} \begin{pmatrix} \mathbf{y} - \hat{\mathbf{y}} \\ \mathbf{x} - \hat{\mathbf{x}} \\ 1 \end{pmatrix} = \mathbf{0}. \quad (9)$$

Fig. 3 shows a numerical example of how to split matrix \mathbf{A} into its submatrices.

The columns of matrix \mathbf{A} corresponding to the unknown variables \mathbf{y} are denoted as \mathbf{A}_y , and the columns corresponding to the measured variables \mathbf{x} as \mathbf{A}_x . The last column of \mathbf{A} , denoted as \mathbf{A}_z , is a column vector that contains the constraint residuals of the transformed linearized equation system \mathbf{g} evaluated with respect to $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$ and \mathbf{z} .

The rows of \mathbf{A} that contain nonzero entries in \mathbf{A}_y are denoted as \mathbf{A}_c . \mathbf{A}_c represents the coefficients of the transformed linearized equations \mathbf{g} that contain unknown variables. The rows of matrix \mathbf{A} that contain only zero entries in \mathbf{A}_y and nonzero entries in \mathbf{A}_x , are denoted as \mathbf{A}_r . \mathbf{A}_r represents the coefficients of the transformed linearized equations \mathbf{g} that contain no unknown variables but at least one measured variable. Finally, the rows of \mathbf{A} that contain only zero entries in \mathbf{A}_y and \mathbf{A}_x are denoted as \mathbf{A}_t . \mathbf{A}_t represents the residuals of the transformed linearized equations \mathbf{g} that are free of unknown and measured variables.

All other submatrices of \mathbf{A} with two index letters (\mathbf{A}_{cy} , \mathbf{A}_{cx} , \mathbf{A}_{cz} , \mathbf{A}_{rx} , \mathbf{A}_{rz} , \mathbf{A}_{tz}) are the intersection of a row matrix (\mathbf{A}_c , \mathbf{A}_r , \mathbf{A}_t) with a column matrix (\mathbf{A}_y , \mathbf{A}_x , \mathbf{A}_z). E.g., \mathbf{A}_{cy} is the intersection of the row matrix \mathbf{A}_c with the column matrix \mathbf{A}_y .

If $\mathbf{A}_{tz} \neq \mathbf{0}$ (actually the first row of $\mathbf{A}_{tz} \neq \mathbf{0}$), there exist contradictions in the constant input data. In this case, the first row of \mathbf{A}_{tz} shows the residual of a constraint $g(\mathbf{z})$ containing constant values only that should be zero per definition. These conflicts have to be resolved before it is possible to reconcile measured data or calculate unknown variables.

If $\mathbf{A}_{tz} = \mathbf{0}$, the originally given equation system includes redundant (dependent) equations that are eliminated during the Gauss-Jordan elimination, and/or a possibly found constraint $g(\mathbf{z})$ is consistent, i.e., there are no contradictions in constant input data. In both cases, these zero rows of \mathbf{A} ($= \mathbf{A}_t$) do not have to be considered any more.

If \mathbf{A}_t does not exist, all given equations are independent and constant input data are not in conflict.

If $\mathbf{A}_r \neq \mathbf{0}$ exists (this implies $\mathbf{A}_{rx} \neq \mathbf{0}$), the matrix $(\mathbf{A}_{rx} \mathbf{A}_{rz})$ can be used for data reconciliation. The constraints for data reconciliation are then reduced to

$$\mathbf{A}_{rx}(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{A}_{rz} = \mathbf{0}, \quad (10)$$

i.e., they no longer contain any unknown variables.

If \mathbf{A}_r does not exist, but there is an \mathbf{A}_x (this implies $\mathbf{A}_{cx} \neq \mathbf{0}$), given measurements cannot be reconciled.

If \mathbf{A}_x does not exist, the problem does not contain any measured variables at all. In this case there is also no \mathbf{A}_y .

The solution of minimizing the objective function (Eq. (1)) subject to the now reduced set of constraints (Eq. (10)) can be found by using the classical method of Lagrange multipliers, leading to the following equation:

$$\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{Q}\mathbf{A}_{rx}^T (\mathbf{A}_{rx}\mathbf{Q}\mathbf{A}_{rx}^T)^{-1} (\mathbf{A}_{rx}(\tilde{\mathbf{x}} - \hat{\mathbf{x}}) + \mathbf{A}_{rz}). \quad (11)$$

3.4. Example (Part 2)

After Gauss-Jordan elimination of $(\hat{\mathbf{J}}_y \quad \hat{\mathbf{J}}_x \quad \hat{\mathbf{f}})$, the resulting coefficient matrix is

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{cy} & \mathbf{A}_{cx} & \mathbf{A}_{cz} \\ \mathbf{0} & \mathbf{A}_{rx} & \mathbf{A}_{rz} \end{pmatrix} = \left(\begin{array}{ccc|ccc|c} 1 & 0 & 0 & 0 & 0 & -1 & -600 & -10 \\ 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & -10 \\ 0 & 0 & 0 & 0 & 1 & -2 & -600 & -20 \end{array} \right).$$

As \mathbf{A}_r does not exist, all given equations are independent and there are no contradiction in constant input data.

New estimates for \mathbf{x} can now be calculated by using Eq. (11):

$$\mathbf{x} = (102.4220, 302.4220, 152.4420, 0.4960)^T.$$

3.5. Theory (Part 3)

If \mathbf{A}_c does not exist (this implies there is also no \mathbf{A}_y), there are no unknown variables involved in the problem.

If $\mathbf{A}_{cy} = \mathbf{I}$, all unknown variables are observable, meaning they can be calculated ($\mathbf{A}_{cy}^* = \mathbf{A}_{cy} = \mathbf{I}$, $\mathbf{A}_{cx}^* = \mathbf{A}_{cx}$, $\mathbf{A}_{cz}^* = \mathbf{A}_{cz}$, $\mathbf{y}^* = \mathbf{y}$, $\hat{\mathbf{y}}^* = \hat{\mathbf{y}}$).

If $\mathbf{A}_{cy} \neq \mathbf{I}$, matrix \mathbf{A} must be altered in order to be able to calculate the observable unknown variables. Therefore, all rows in \mathbf{A}_c that contain more than one nonzero entry in \mathbf{A}_{cy} and all columns in \mathbf{A}_y that have nonzero entries in these rows have to be deleted ($\mathbf{A}_{cy} \rightarrow \mathbf{A}_{cy}^*$, $\mathbf{A}_{cx} \rightarrow \mathbf{A}_{cx}^*$, $\mathbf{A}_{cz} \rightarrow \mathbf{A}_{cz}^*$). The deleted columns of \mathbf{A}_{cy} refer to unobservable unknown variables (they cannot be calculated from the given data) that also have to be removed from \mathbf{y} and $\hat{\mathbf{y}}$ ($\mathbf{y} \rightarrow \mathbf{y}^*$, $\hat{\mathbf{y}} \rightarrow \hat{\mathbf{y}}^*$).

After the elimination of unobservable unknown variables the observable ones can be calculated from

$$\mathbf{I}(\mathbf{y}^* - \hat{\mathbf{y}}^*) + \mathbf{A}_{cx}^*(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{A}_{cz}^* = \mathbf{0}, \quad (12)$$

leading to

$$\mathbf{y}^* = \hat{\mathbf{y}}^* - \mathbf{A}_{cx}^*(\mathbf{x} - \hat{\mathbf{x}}) - \mathbf{A}_{cz}^*. \quad (13)$$

3.6. Example (Part 3)

Because $\mathbf{A}_{cy} \neq \mathbf{I}$ the equation system contains unobservable unknown variables that have to be eliminated. This goal can be reached by deleting row 2 (it contains more than one nonzero entry in \mathbf{A}_{cy}) and column 2 and 3 (nonzero entries in row 2 representing the unobservable unknown variable m_6 and m_7) of matrix \mathbf{A} . This leads to

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{A}_{cy}^* & \mathbf{A}_{cx}^* & \mathbf{A}_{cz}^* \\ \mathbf{0} & \mathbf{A}_{rx} & \mathbf{A}_{rz} \end{pmatrix} = \left(\begin{array}{ccc|ccc|c} 1 & 0 & 0 & -1 & -600 & -10 \\ 0 & 1 & 0 & -1 & 0 & -10 \\ 0 & 0 & 1 & -2 & -600 & -20 \end{array} \right),$$

$$\hat{\mathbf{y}}^* = (\hat{m}_4) = (150).$$

New estimates for \mathbf{y}^* can be calculated by using Eq. (13):

$$\mathbf{y}^* = (150).$$

3.7. Theory (Part 4)

If the new estimates \mathbf{x} and \mathbf{y}^* are significantly different from the initial estimates $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}^*$, respectively (the 2-norm of $\mathbf{x} - \hat{\mathbf{x}}$ or $\mathbf{y}^* - \hat{\mathbf{y}}^*$ is bigger than a chosen convergence tolerance, e.g., 10^{-10}), the procedure has to be repeated with renewed evaluation of \mathbf{J}_y , \mathbf{J}_x and \mathbf{f} , where $\hat{\mathbf{x}} = \mathbf{x}$ and $\hat{\mathbf{y}} = \mathbf{y}^*$ (\mathbf{y} is the initial $\hat{\mathbf{y}}$ updated with \mathbf{y}^* on the positions of observable unknown variables). Otherwise the iterations can be stopped and the variance-covariance matrices \mathbf{Q}_x of the reconciled variables \mathbf{x} and \mathbf{Q}_{y^*} of the observable unknown variables \mathbf{y}^* can be calculated:

$$\mathbf{Q}_x = \left(\mathbf{I} - \mathbf{Q}\mathbf{A}_{rx}^T (\mathbf{A}_{rx}\mathbf{Q}\mathbf{A}_{rx}^T)^{-1} \mathbf{A}_{rx} \right) \mathbf{Q}, \quad (14)$$

$$\mathbf{Q}_{y^*} = \mathbf{A}_{cx}^* \mathbf{Q}_x \mathbf{A}_{cx}^{*T}. \quad (15)$$

Eqs. (14) and (15) are derived by error propagation from Eqs. (11) and (13).

3.8. Example (Part 4)

Because \mathbf{x} is significantly different from $\hat{\mathbf{x}}$ (here, in the first iteration $\mathbf{y}^* = \hat{\mathbf{y}}^*$), the calculation procedure has to be repeated with

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{x} = (m_1, m_3, m_5, tc_{34})^T \\ &= (102.4220, 302.4220, 152.4420, 0.4960)^T, \end{aligned}$$

$$\hat{\mathbf{y}} = \mathbf{y} = (m_4, m_6, m_7)^T = (150, 80, 80)^T.$$

After five iterations the final solution is reached (Fig. 4):

$$\mathbf{x} = (102.4260, 302.4162, 152.4260, 0.4960)^T,$$

$$\mathbf{y}^* = (149.9903),$$

$$\mathbf{Q}_x = \begin{pmatrix} 62.1384 & 61.6419 & 62.1384 & -0.1027 \\ 61.6419 & 511.1494 & 61.6419 & 0.6481 \\ 62.1384 & 61.6419 & 62.1384 & -0.1027 \\ -0.1027 & 0.6481 & -0.1027 & 0.0014 \end{pmatrix},$$

$$\mathbf{Q}_{y^*} = (450.0040).$$

In the diagonal of the \mathbf{Q} matrices the variances of the results can be found. The standard errors are calculated by taking their square roots.

$$\sigma_x = (7.8826, 22.6086, 7.8826, 0.0377)^T,$$

$$\sigma_{y^*} = (21.2133).$$

Although the measurements were initially assumed to be independent, the \mathbf{Q}_x matrix shows that the reconciled measurements are correlated after the data reconciliation procedure due to the applied constraints. However, these correlations are not displayed in STAN.

3.9. Summary of algorithm

1. Take the measured values $\tilde{\mathbf{x}}$ as initial estimates $\hat{\mathbf{x}}$ and compute initial estimates $\hat{\mathbf{y}}$.
2. Evaluate \mathbf{J}_y , \mathbf{J}_x and \mathbf{f} with respect to $\hat{\mathbf{y}}$, $\hat{\mathbf{x}}$ and \mathbf{z} .

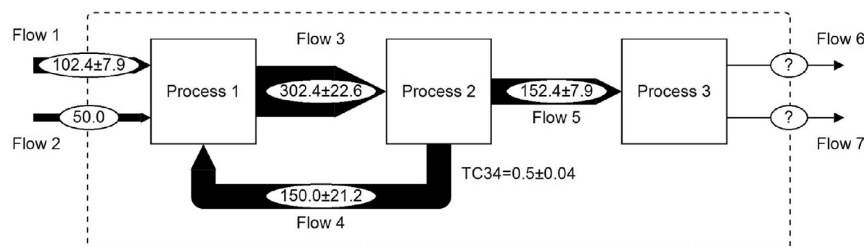


Fig. 4. Results of flowsheet example rounded to one decimal place.

3. Compute $\text{rref}(\hat{J}_y \hat{J}_x \hat{f})$.
4. Eliminate unobservable unknown variables and redundant equations.
5. Compute new estimates \mathbf{x} with Eq. (11).
6. Compute new estimates \mathbf{y}^* with Eq. (13).
7. If the new estimates \mathbf{x} and \mathbf{y}^* are significantly different from $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}^*$, respectively, set $\hat{\mathbf{x}} = \mathbf{x}$ and $\hat{\mathbf{y}} = \mathbf{y}^*$ and go to 2. Otherwise go to 8.
8. Compute the variance-covariance-matrices \mathbf{Q}_x with Eq. (14) and \mathbf{Q}_{y^*} with Eq. (15).

4. Discussion and outlook

In this paper, the nonlinear data reconciliation algorithm implemented in STAN was explained and its application demonstrated on a simple hypothetical example from MFA.

A restriction of the used weighted least squares minimization approach is the assumption of normally distributed measurement errors. In scientific models in general and in MFA models in particular, this assumption is often not valid: e.g., concentrations cannot take negative values, and transfer coefficients are restricted to the unit interval. To overcome the limitation of normality, a general framework to reconcile data with arbitrarily distributed measurement errors was introduced [10]. This framework is limited to linear constraints, but has been extended to nonlinear constraints in Ref. [11].

It was shown [12] that it is also possible to use a possibilistic approach for data reconciliation. There, the uncertainty of measurements is modelled with membership functions instead of probability density functions to account for the epistemic nature of measurements errors (that is, error due to insufficient knowledge). While the paper covers linear constraints only, the approach has been extended to nonlinear constraints in Ref. [13].

The problem of nonlinear data reconciliation can also be solved with nonlinear programming techniques, like sequential quadratic programming or reduced gradient methods. These techniques allow for a general objective function, not just one with weighted least squares, and they are able to handle inequality constraints and variable bounds. For a short review of these methods see e.g. Ref. [5].

While all of these alternative approaches definitely have their advantages, their common disadvantage is the large amount of computation time required compared to the conventional approach of weighted least squares.

In general, nonlinear data reconciliation of normally distributed input data does not result in normally distributed output data. This is only the case for linear constraints or linearized nonlinear constraints. The latter approximation, however, delivers sound results only if the uncertainties of the input data are small. If the uncertainties are large, the results of linearization can differ substantially from the precise solution.

In the weighted least squares minimization approach, the inverse of the covariance matrix \mathbf{Q} was chosen as the weight matrix because it delivers the best linear unbiased estimator of \mathbf{x} in Eq. (11). A prove of the linear case can be found in Ref. [14].

The following list contains some limitations of STAN that should be addressed/optimized in a future version:

1. While the variable classification using the Gauss-Jordan elimination is easy to understand, it is not the best way in a computational sense. Other equation-oriented approaches have been developed to reach the same goal more efficiently [5].
2. There is no equation parser implemented in STAN, thus, it is restricted to a few types of equations only: mass balances, transfer coefficient equations, linear relations between similar entities (can be used to model, e.g., losses from stocks) and concentration equations.
3. The default algorithm used in STAN (called "Cencic2012") is coded for dense matrices, thus, the speed of the calculation is reduced considerably when dealing with large models. An implementation of sparse matrices would increase the calculation speed substantially.
4. The only gross error detection test that has been yet implemented in STAN is the so called measurement test [4] that is based on measurement adjustments. A more sophisticated robust gross error detection routine would be of advantage.

Since September 2012, an alternative commercial calculation algorithm developed by J.D. Kelly is available in STAN. Originally called "Kelly2011", it was later renamed into "IAL-IMPL2013". It applies a regularization approach by assuming unknown variables to be known with a sufficient large uncertainty. Details about the algorithm can be found in Ref. [15].

Since the first version of STAN was released in 2006, a lot of MFA studies have been conducted with its help. An updated list of publications can be found under www.stan2web.net/infos/publications. Unfortunately, still a lot of recent MFA studies do not consider data uncertainties, thus, ignoring valuable information for decision makers. The author would appreciate if STAN could help to raise the awareness for the importance of uncertainties, thus, taking MFA to the next level.

Final remark: The presented nonlinear data reconciliation algorithm is of course not restricted to MFA models. It can be used for arbitrary reconciliation problems.

Acknowledgement

STAN was funded by Altstoff Recycling Austria (ARA), voestalpine, the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, and the Federal States of Austria. For more detailed information and to download the software visit www.stan2web.net.

Appendix A. Software availability

Name of software	STAN – Software for Substance Flow Analysis
Version	2.5.1302 (March 2016)
Website	www.stan2web.net
Costs	Freeware
Availability	Downloadable from www.stan2web.net (registration required)
Package size	10 MB
Language	English and German
Available since	2006
Hardware required	Intel Pentium III, 1 GHz, 512 MB RAM, 20 MB free disc space
Software required	Windows OS (minimum Windows XP with SP1), Microsoft.Net Framework 2.0 or higher
Program language	C#
Tutorial	To get a basic introduction into STAN, watch the help video on www.stan2web.net .
Model database	Free access for registered users on the website or directly from the user interface of the software. Anyone interested is invited to upload own models to share them with the community.
Developers	Oliver Cencic (Vienna University of Technology, iwr.tuwien.ac.at), Alfred Kovacs (inka software, www.inkasoft.net)
Contact address	Institute for Water Quality, Resource and Waste Management, Vienna University of Technology, Karlsplatz 13/226, A-1040 Vienna, Austria +43 1 58801 22657 (Oliver Cencic)
Phone	
Email	oliver.cencic@tuwien.ac.at

Appendix B. Notation

m	number of equations
m_c	number of equations available for calculating unknown variables (= rows of \mathbf{A}_c)
m_r	number of equations available for data reconciliation (= rows of \mathbf{A}_r)
m_t	number of redundant equations (= rows of \mathbf{A}_t)
n	number of measured variables (= columns of \mathbf{A}_x)
o	number of observable unknown variables
p	number of unknown variables (= columns of \mathbf{A}_y)
q	number of constant variables
*	superscript of vectors and matrices with removed parts due to unobservable unknown variables (in dimensions replace $m_c \rightarrow o$ and $p \rightarrow o$)
F	objective function to be minimized
\mathbf{f}	vector ($m \times 1$) of equality constraints
$\hat{\mathbf{f}}$	vector ($m \times 1$) of equality constraints evaluated at $\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z}$
\mathbf{g}	vector ($m \times 1$) of transformed equality constraints
\mathbf{x}	vector ($n \times 1$) of reconciled measurements
$\hat{\mathbf{x}}$	vector ($n \times 1$) of initial estimates of reconciled measurements
$\tilde{\mathbf{x}}$	vector ($n \times 1$) of measurements

\mathbf{y}	vector ($p \times 1$) of best estimates of unknown variables
$\hat{\mathbf{y}}$	vector ($p \times 1$) of initial estimates of unknown variables
\mathbf{z}	vector ($q \times 1$) of constant values
μ_x	vector ($n \times 1$) of true values of measured variables
$\boldsymbol{\varepsilon}$	vector ($n \times 1$) of measurement errors of measurements
$\mathbf{0}$	null vector
\mathbf{A}	coefficient matrix ($m \times (p + n + 1)$) of transformed linearized equality constraints \mathbf{g}
\mathbf{A}_c	submatrix of \mathbf{A} ($m_c \times (p + n + 1)$) for computation of unknown variables
\mathbf{A}_{cx}	submatrix of \mathbf{A} ($m_c \times n$) for computation of unknown variables
\mathbf{A}_{cy}	submatrix of \mathbf{A} ($m_c \times p$) for computation of unknown variables
\mathbf{A}_{cz}	submatrix of \mathbf{A} ($m_c \times 1$) for computation of unknown variables
\mathbf{A}_r	submatrix of \mathbf{A} ($m_r \times (p + n + 1)$) for data reconciliation
\mathbf{A}_{rx}	submatrix of \mathbf{A} ($m_r \times n$) for data reconciliation
\mathbf{A}_{rz}	submatrix of \mathbf{A} ($m_r \times 1$) for data reconciliation
\mathbf{A}_t	submatrix of \mathbf{A} ($m_t \times (p + n + 1)$) for check on contradiction in constant input data
\mathbf{A}_{tz}	submatrix of \mathbf{A} ($m_t \times 1$) for check on contradiction in constant input data
\mathbf{A}_x	submatrix of \mathbf{A} ($m \times n$) corresponding to measured variables
\mathbf{A}_y	submatrix of \mathbf{A} ($m \times p$) corresponding to unknown variables
\mathbf{A}_z	submatrix of \mathbf{A} ($m \times 1$) containing constraint residuals
\mathbf{I}	identity matrix
\mathbf{J}_x	Jacobi-matrix ($m \times n$) of measured variables
$\hat{\mathbf{J}}_x$	Jacobi-matrix ($m \times n$) of measured variables evaluated at $\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z}$
\mathbf{J}_y	Jacobi-matrix ($m \times p$) of unknown variables
$\hat{\mathbf{J}}_y$	Jacobi-matrix ($m \times p$) of unknown variables evaluated at $\hat{\mathbf{y}}, \hat{\mathbf{x}}, \mathbf{z}$
\mathbf{O}	null matrix
\mathbf{Q}	variance-covariance matrix ($n \times n$) of measurements
\mathbf{Q}_x	variance-covariance matrix ($n \times n$) of reconciled measurements
\mathbf{Q}_{y^*}	variance-covariance matrix ($o \times o$) of best estimates of observable unknown variables

References

- [1] Brunner PH, Rechberger H. Practical Handbook of Material Flow Analysis. Boca Raton, FL: CRC/Lewis; 2004.
- [2] Laner D, Rechberger H, Astrup T. Systematic evaluation of uncertainty in material flow analysis. J Ind Ecol 2014;18:859–70.
- [3] Bagajewicz MJ, Chmielewski DJ, Tanth DN. Smart Process Plants: Software and Hardware Solutions for Accurate Data and Profitable Operations. New York: McGraw-Hill; 2010.
- [4] Narasimhan S, Jordache C. Data Reconciliation & Gross Error Detection – An Intelligent Use of Process Data. Houston, TX: Gulf Publishing; 2000.
- [5] Romagnoli JA, Sánchez MC. Data Processing and Reconciliation for Chemical Process Operations. San Diego, CA: Academic; 2000.
- [6] Cencic O, Rechberger H. Material flow analysis with software STAN. J Environ Eng Manage 2008;18:3–7.
- [7] Madron F. Process Plant Performance: Measurement and Data Processing for Optimization and Retrofits. Chichester, UK: Ellis Horwood; 1992.
- [8] Weisstein EW. Echelon Form. From MathWorld – A Wolfram Web Resource. <http://mathworld.wolfram.com/EchelonForm.html>.
- [9] Weisstein EW. Elementary Row and Column Operations. From MathWorld – A Wolfram Web Resource. <http://mathworld.wolfram.com/ElementaryRowandColumnOperations.html>.
- [10] Cencic O, Frühwirth R. A general framework for data reconciliation – Part I: linear constraints. Comput Chem Eng 2015;75:196–208.
- [11] Cencic O, Frühwirth R. Data reconciliation of nonnormal observations – Part II: nonlinear constraints, correlated observations and gross error detection by robust estimation. AIChE J 2016 [submitted] for publication.

- [12] Dubois D, Fargier H, Ababou M, Guyonnet D. A fuzzy constraint-based approach to data reconciliation in material flow analysis. *Int J Gen Syst* 2014;43:787–809.
- [13] Dzubur N, Sunata O, Laner D. A fuzzy set-based approach for data reconciliation in material flow modeling. *Appl Math Model* 2016 [submitted] for publication.
- [14] Kailath T, Sayed AH, Hassibi B. *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall; 2000.
- [15] Kelly JD. A regularization approach to the reconciliation of constrained data sets. *Comput Chem Eng* 1998;22:1771–88.

Paper 2:

**A general framework for data reconciliation -
Part I: Linear constraints**

Oliver Cencic and Rudolf Frühwirth

Computers and Chemical Engineering 2015, 75

DOI: [10.1016/j.compchemeng.2014.12.004](https://doi.org/10.1016/j.compchemeng.2014.12.004)



A general framework for data reconciliation—Part I: Linear constraints

Oliver Cencic^{a,*}, Rudolf Frühwirth^b^a Institute for Water Quality, Resource and Waste Management, Vienna University of Technology, Karlsplatz 13/226, A-1040 Wien, Austria^b Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences, Nikolsdorfer Gasse 18, A-1050 Wien, Austria

ARTICLE INFO

Article history:

Received 21 April 2014

Received in revised form 3 October 2014

Accepted 2 December 2014

Available online 12 December 2014

Keywords:

Data reconciliation

Nonnormal distributions

Linear equality constraints

Markov chain Monte Carlo (MCMC)

Independence Metropolis–Hastings sampler

Material flow analysis (MFA)

ABSTRACT

This paper presents a new method, based on Bayesian reasoning, for the reconciliation of data from arbitrary probability distributions. The main idea is to restrict the joint prior probability distribution of the involved variables with model constraints to get a joint posterior probability distribution. This paper covers the case of linearly constrained variables, with the focus on equality constraints. The procedure is demonstrated with the help of three simple graphical examples. Because in general the posterior probability density function cannot be calculated analytically, it is sampled with a Markov chain Monte Carlo (MCMC) method. From this sample the density and its moments can be estimated, along with the marginal densities, moments and quantiles. The method is tested on several artificial examples from material flow analysis, using an independence Metropolis–Hastings sampler.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The goal of material flow analysis (MFA) is to model and quantify all flows and stocks of a system of interest. For this reason as much information about the system as possible is collected which comprises direct measurements when available, but more often data taken from official statistics, reports, publications, expert estimates and similar sources (Laner et al., 2014). Unfortunately these data are often in conflict with known conservation laws such as mass or energy balances, preventing the calculation of unknown quantities or parameters of the model that cannot be measured directly. The basic idea of data reconciliation (DR) is to resolve these contradictions by statistically adjusting the collected data based on the assumption that their uncertainty is described by a probability density function.

DR has been widely used in chemical engineering for more than 50 years to adjust plant measurements. Most solving techniques that have been developed in this period of time are based on a weighted least-squares minimization of the measurement adjustments subject to constraints involving reconciled, unmeasured and fixed variables (Narasimhan and Jordache, 2000; Romagnoli and Sanchez, 2000; Bagajewicz, 2010). The underlying main assumption of this approach is that of normally distributed (Gaussian)

measurement errors with zero mean (Johnston and Kramer, 1995). However, in scientific models in general and in MFA models in particular, data is often not normally distributed. If, for instance, a process model is correct, mass flows and concentrations cannot take negative values, and transfer coefficients are restricted to the unit interval.

Another example is provided by expert opinions that frequently have to be relied on in MFA due to scarce or missing data. They are often modeled by uniform, triangular or trapezoidal distributions. The more detailed the expert's knowledge about the quantity under consideration is, the more precisely the distribution can be modeled. If a sufficient number of measurements of the quantity is available, one can either fit a parametric model to the measured data or use a nonparametric model such as the empirical distribution function or the kernel estimate of the probability density function. In the following we will denote a variable as “measured” if there is prior information on the variable of any kind, which is not necessarily a proper measurement.

Although it was demonstrated in Crowe (1996) that the assumption of a normal distribution is acceptable for unknown distributions having relative standard deviations smaller than 30%, it is questionable in the context of macro-scale MFA (e.g. region, country) where relative standard deviations larger than 30% are not uncommon. In addition, the normal distribution is unsuitable to describe uncertainties with strong intrinsic asymmetry.

In the following we propose a numerical DR procedure that is also able to deal with data that cannot be modeled by normal

* Corresponding author. Tel.: +43 1 58801 22657; fax: +43 1 58801 9 22657.
E-mail address: oliver.cencic@tuwien.ac.at (O. Cencic).

distributions. In this paper we treat the case of linearly constrained variables, with the focus on equality constraints; the cases of non-linear and inequality constraints will be the subject of a subsequent paper. We start from the following assumptions:

1. There are N measured or unmeasured variables that take values in a subset $D \subseteq \mathbb{R}^N$.
2. The $I \leq N$ measured variables form an I -dimensional random variable with known joint density. The latter is called the prior density. The prior density can be either objective, i.e. the model of a measurement process, or subjective, i.e. the formalization of an expert opinion.
3. The variables are subject to linear equality constraints that define an affine subspace $S \subset \mathbb{R}^N$ of dimension $P < N$.

In Section 2 it is shown that the density of the variables conditional on the constraints is obtained by restricting their prior density to the set $D \cap S$ and normalizing the restricted density to 1. The resulting density is called the posterior density. The prior density plays a key role in the DR mechanism proposed below. No matter how it is obtained, it is good practice to study its influence on the posterior distribution.

In the case of a low-dimensional variable space, the construction of the posterior density can be demonstrated graphically. To show this, we present some simple examples.

Example 1.1. Let us assume that there are two measured variables x_1 and x_2 with the prior density $f(x_1, x_2)$ defined on $D \subseteq \mathbb{R}^2$. The constraint equation $x_1 = x_2$ defines a 1-dimensional subspace S , i.e. a line in \mathbb{R}^2 . If the prior density is restricted to points on this line and normalized to 1, the posterior density of x_1, x_2 is obtained. By computing the marginal distributions of the posterior we get the posterior densities of x_1 and x_2 , which are identical in this case. The values of $f(x_1, x_2)$ along S can be visualized by intersecting the prior density surface with the vertical plane through S .

Fig. 1 shows an instance of this problem, with x_1, x_2 independent, $f_1(x_1) = \gamma(x_1; 2, 2)$ and $f_2(x_2) = \gamma(x_2; 3, 1.5)$, where $\gamma(x; a, b)$ is the density of the Gamma distribution with parameters a and b :

$$\gamma(x; a, b) = \frac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)}.$$

Example 1.2. Let us assume that there are three measured variables x_1, x_2 and x_3 with the prior density $f(x_1, x_2, x_3)$ defined on $D \subseteq \mathbb{R}^3$. The constraint equation $x_3 = x_1 + x_2$ defines a 2-dimensional subspace S , i.e. a plane in \mathbb{R}^3 . If the prior density is restricted to points in this plane and normalized to 1, the posterior density of x_1, x_2, x_3 is obtained. By computing the marginal distributions of the posterior we get the posterior densities of x_1, x_2 and x_3 , respectively.

Fig. 2 shows an instance of this problem, with x_1, x_2, x_3 independent, $f_1(x_1) = \gamma(x_1; 2, 2)$, $f_2(x_2) = \gamma(x_2; 3, 1.5)$ and $f_3(x_3) = \gamma(x_3; 6, 1.7)$. The values of $f(x_1, x_2, x_3)$ are shown color-coded.

Example 1.3. Let us assume that there are two measured variables x_1, x_2 and one unmeasured variable x_3 . The prior density of x_1, x_2, x_3 is defined on $D \subseteq \mathbb{R}^3$, but can be written as $f(x_1, x_2)$, as it does not depend on x_3 . The rest of the procedure is the same as in Example 1.2. Due to the lack of an actual constraint the 2-dimensional prior density is not restricted by the 2-dimensional subspace S , the posterior densities of x_1 and x_2 are equal to the priors, and the posterior of x_3 is their convolution.

Fig. 3 shows an instance of this problem, with x_1, x_2 independent, $f_1(x_1) = \gamma(x_1; 2, 2)$, $f_2(x_2) = \gamma(x_2; 3, 1.5)$ and x_3 not measured. The values of $f(x_1, x_2)$ are shown color-coded. This example demonstrates that the method can also be used to calculate unknown

variables and that it even works when the measured variables cannot be reconciled.

In the case of a nonnormal prior density, the normalization constant of the restricted density cannot in general be computed analytically. In the simple examples just discussed, it can be computed numerically by a single or a double integral. For larger dimensions of S , however, numerical integration becomes cumbersome and time-consuming. We therefore propose to avoid explicit calculation of the posterior density altogether by generating a random sample from the unnormalized restricted density. This can be achieved by applying a tool that is frequently used in Bayesian statistics (O'Hagan, 1994), namely Markov chain Monte Carlo (MCMC) (Robert and Casella, 2004; Liu, 2004; Brooks et al., 2011). The method and its implementation in the context of DR is explained in Section 3. Section 4 presents the application of MCMC to four examples in MFA. Finally, Section 5 contains our conclusions and the outlook on further work.

2. Mathematical foundation

Let \mathbf{v} be a column vector of N measured or unmeasured variables. Following the notation in Madron (1992), we assume that \mathbf{v} is arranged such that $\mathbf{v} = (\mathbf{y}; \mathbf{x})$, where \mathbf{y} contains the J unmeasured variables and \mathbf{x} contains the $I = N - J$ measured variables.¹ We also may have a vector \mathbf{z} of M fixed (nonrandom) variables. DR means that \mathbf{v} is modified in such a way that it satisfies a system of constraint equations. If all K equations are linear, the constrained system can be written in the following form:

$$\mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{C}\mathbf{z} = \mathbf{0} \quad \text{or} \quad \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{c} = \mathbf{0}, \quad (1)$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are known matrices of dimension $K \times I, K \times J, K \times M$, respectively, and \mathbf{c} is a column vector of dimension $K \times 1$. We assume that

- A1. $\text{rank}(\mathbf{B}, \mathbf{A}) = \text{rank}(\mathbf{B}, \mathbf{A}, \mathbf{c})$, meaning the system is solvable;
- A2. $\text{rank}(\mathbf{B}, \mathbf{A}) = K$, meaning the model equations are linearly independent;
- A3. $\text{rank}(\mathbf{B}) = J$, meaning all unmeasured quantities are observable (they can be calculated).

If any of these assumptions is violated the underlying problems have to be resolved before being able to proceed. One way to achieve this goal is to apply the Gauss-Jordan elimination to matrix $(\mathbf{B}, \mathbf{A}, \mathbf{c})$. The result, known as the reduced row echelon form (or canonical form), serves to detect contradictions (A1), to eliminate dependent equations automatically (A2) and to classify variables, in particular to identify and eliminate unobservable unmeasured variables (A3). For detailed instructions how to proceed see Madron (1992, p. 125). There exist alternative equation-oriented approaches for variable classification (Romagnoli and Sanchez, 2000, p. 33), but in our opinion the Gauss-Jordan elimination is the easiest to understand.

We make further use of the reduced row echelon form in order to identify dependent and free variables of the system. The column numbers of the pivot elements (leading 1 in each row) denote the dependent variables, which can be unmeasured or measured ones. All other variables, which have to be measured ones, are designated as free. The outcome of this classification process depends on the initial order of the variables. Although the posterior density itself is unique, the choice of the free variables can affect its computation, so the initial order of the variables should be chosen carefully (see Section 3.2 and Example 4.4).

¹ The semicolon (comma) denotes vertical (horizontal) concatenation of matrices.

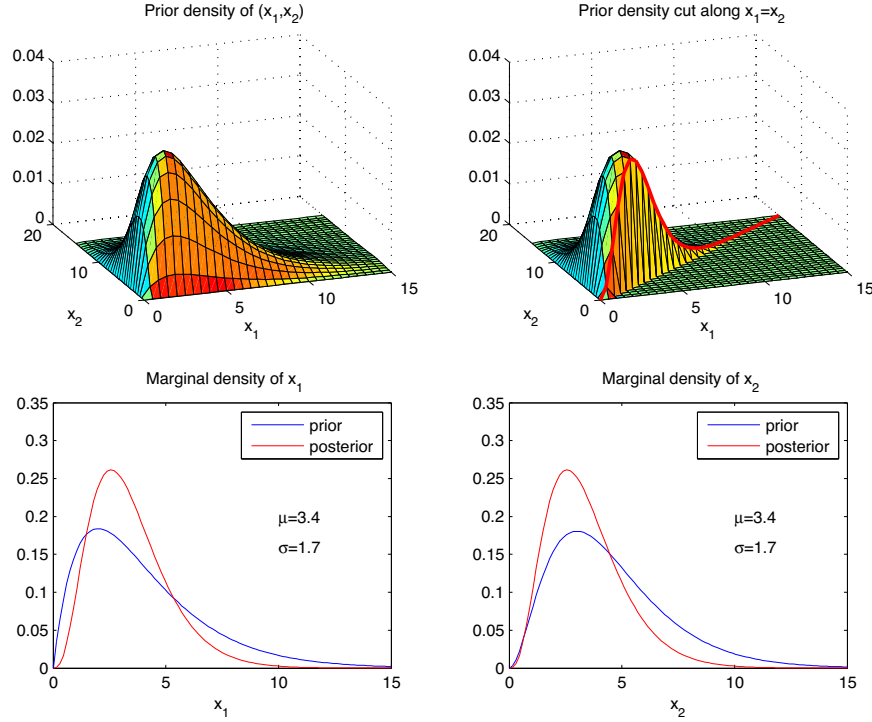


Fig. 1. Visualization of the densities in Example 1.1. Top left: prior density $f(x_1, x_2)$. Top right: prior density $f(x_1, x_2)$ cut along the line $x_1 = x_2$; the red line is the (unnormalized) restricted density. Bottom: measured (prior, blue) and reconciled (posterior, red) densities of x_1, x_2 .

After column reordering the system can be written as

$$\mathbf{I}\mathbf{v}_d + \mathbf{D}\mathbf{v}_f + \mathbf{d} = \mathbf{0}, \quad (2)$$

where \mathbf{v}_d contains the K dependent variables and \mathbf{v}_f contains the $N - K$ free variables. The number of free variables is equal to the dimension of the constraint manifold used to restrict the prior distribution. If for instance $N - K = 2$, the constraint manifold is a plane in \mathbb{R}^N . The general solution of Eq. (2) can then be written as

$$\mathbf{v}_d = -\mathbf{D}\mathbf{v}_f - \mathbf{d}. \quad (3)$$

Let us assume that all variables are independent *a priori*, so that the prior density of \mathbf{v} can be written as the product of the marginal densities:

$$f(\mathbf{v}) = \prod_{i=1}^N f_i(v_i) = f_d(\mathbf{v}_d) \cdot f_f(\mathbf{v}_f). \quad (4)$$

If a dependent variable is not measured, its marginal prior density is set to the improper prior $f_i(v_i) = 1$.

After DR, constraint Eq. (3) has to be satisfied. The posterior distribution of \mathbf{v}_d is therefore uniquely determined by the posterior distribution of \mathbf{v}_f . In order to compute the latter, we make the following affine transformation:

$$\mathbf{r} = \mathbf{v}_d + \mathbf{D}\mathbf{v}_f + \mathbf{d}, \quad \mathbf{w} = \mathbf{v}_f, \quad (5)$$

or

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{w} \end{pmatrix} = \mathbf{H} \begin{pmatrix} \mathbf{v}_d \\ \mathbf{v}_f \end{pmatrix} + \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix}, \quad \text{with } \mathbf{H} = \begin{pmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad (6)$$

where \mathbf{I} is the identity matrix and $\mathbf{0}$ is the null matrix. The joint density of $(\mathbf{r}; \mathbf{w})$ can be computed by using the following lemma.

Lemma. Let \mathbf{t} be a vector of n random variables with joint density $f(\mathbf{t})$, \mathbf{H} a $n \times n$ matrix of full rank, and \mathbf{e} a vector of dimension n . The joint density $g(\mathbf{s})$ of $\mathbf{s} = \mathbf{H}\mathbf{t} + \mathbf{e}$ is then given by

$$g(\mathbf{s}) = \frac{f(\mathbf{t})}{|\det \mathbf{H}|} = \frac{f(\mathbf{H}^{-1}(\mathbf{s} - \mathbf{e}))}{|\det \mathbf{H}|}. \quad (7)$$

Solving Eq. (6) for \mathbf{v}_d and \mathbf{v}_f gives:

$$\mathbf{v}_d = \mathbf{r} - \mathbf{D}\mathbf{w} - \mathbf{d}, \quad \mathbf{v}_f = \mathbf{w}.$$

As $|\det \mathbf{H}| = 1$, the joint density of $(\mathbf{r}; \mathbf{w})$ is then given by

$$g(\mathbf{r}; \mathbf{w}) = f_d(\mathbf{r} - \mathbf{D}\mathbf{w} - \mathbf{d}) f_f(\mathbf{w}). \quad (8)$$

The marginal density $g_r(\mathbf{r})$ is obtained by integration over \mathbf{w} :

$$g_r(\mathbf{r}) = \int f_d(\mathbf{r} - \mathbf{D}\mathbf{w} - \mathbf{d}) f_f(\mathbf{w}) d\mathbf{w}. \quad (9)$$

The constraints are satisfied if and only if $\mathbf{r} = \mathbf{0}$. The posterior density $\pi(\mathbf{v}_f)$ of $\mathbf{v}_f = \mathbf{w}$ is therefore given by the following conditional density:

$$\pi(\mathbf{v}_f) = g(\mathbf{w} | \mathbf{r} = \mathbf{0}) = \frac{g(\mathbf{0}; \mathbf{w})}{g_r(\mathbf{0})} = \frac{f_d(-\mathbf{D}\mathbf{v}_f - \mathbf{d}) f_f(\mathbf{v}_f)}{\int f_d(-\mathbf{D}\mathbf{v}_f - \mathbf{d}) f_f(\mathbf{v}_f) d\mathbf{v}_f}. \quad (10)$$

The calculation of the posterior density $\pi(\mathbf{v}_f)$ in Eq. (10) requires the evaluation of the integral in the denominator. With the exception of very simple cases, the integral cannot be computed analytically. Numerical integration is feasible if the dimension of \mathbf{v}_f is small, but becomes cumbersome with increasing number of free variables. However, explicit calculation of the integral can be avoided altogether if a random sample is drawn from the posterior density $\pi(\mathbf{v}_f)$ by Markov chain Monte Carlo (MCMC) methods.

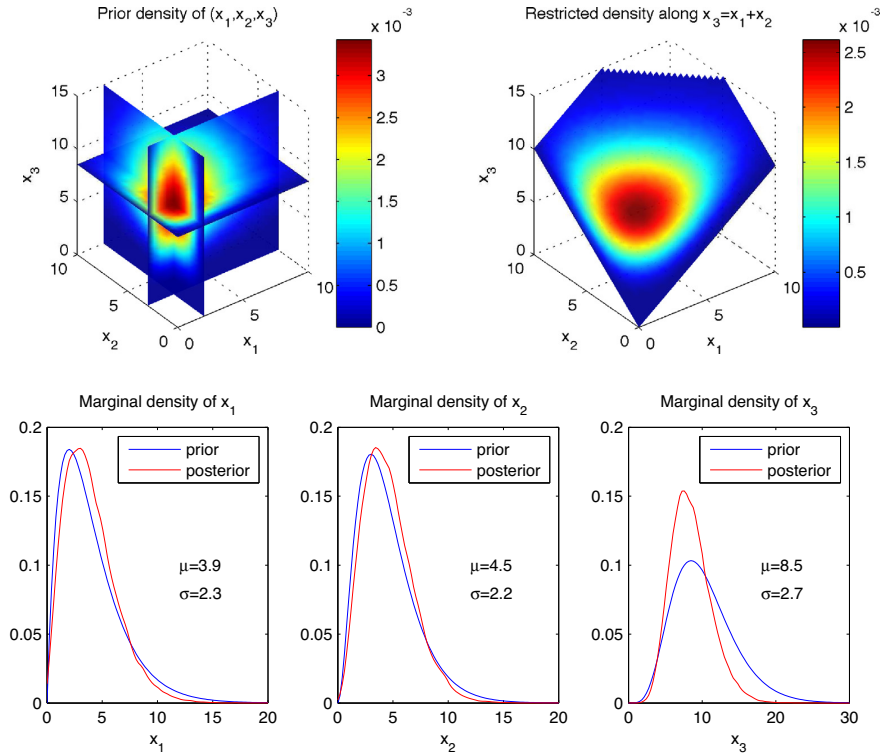


Fig. 2. Visualization of the densities in Example 1.2. Top left: prior density $f(x_1, x_2, x_3)$. Top right: prior density $f(x_1, x_2, x_3)$ on the plane $x_3 = x_1 + x_2$. Bottom: measured (prior, blue) and reconciled (posterior, red) densities of x_1, x_2, x_3 .

The sample is not independent, as each draw depends on the previous one. From this sample the posterior density and its moments can be estimated, along with the marginal densities, moments and quantiles. A general introduction to Markov chains and MCMC is contained in Robert and Casella (2004), Liu (2004), Brooks et al. (2011).

It can be proved that in the case of normally distributed observations Eq. (10) gives the same result as the traditional constrained least-squares optimization. A sketch of the proof can be found in Appendix A.

If the variables are not independent, their prior density has the general form $f(\mathbf{v}) = f(\mathbf{v}_d; \mathbf{v}_f)$. Using the same affine transformation as in Eq. (6), we obtain the posterior density of \mathbf{v}_f in the general case:

$$\pi(\mathbf{v}_f) = \frac{f(-\mathbf{D}\mathbf{v}_f - \mathbf{d}; \mathbf{v}_f)}{\int f(-\mathbf{D}\mathbf{v}_f - \mathbf{d}; \mathbf{v}_f) d\mathbf{v}_f}. \quad (11)$$

The proof in Appendix A can be modified to show that the results are again the same as in the traditional approach.

3. MCMC sampling

3.1. The sampling algorithm

The idea behind MCMC sampling is the construction of a Markov chain with an equilibrium distribution that is equal to the target

distribution $\pi(\mathbf{w})$.² A useful review of this approach is given in Tierney (1994).

The sampler presented below is a special case of the Metropolis–Hastings (MH) algorithm (Robert and Casella, 2004; Liu, 2004; Metropolis et al., 1953; Hastings, 1970). For an introductory exposition of the MH algorithm see Chib and Greenberg (1995). All MH samplers require a proposal density from which candidates for the Markov chain are drawn. The choice of the proposal density is a critical step in the construction of the sampler. If it is too narrow, it is difficult to explore the entire target distribution. If it is too wide, the acceptance rate of the proposed values is low, and the convergence to the target distribution is slow. Of course, drawing from the proposal density should be easy and fast.

We have found that the sampler best suited to the problem of DR is the independence sampler (IS) (Brooks et al., 2011; Chib and Greenberg, 1995; Liu, 1996). In the IS, the proposal values are drawn independently from a proposal density $p(\mathbf{w})$. The proposed values therefore do not depend on the current value of the Markov chain, but the accepted values do. The algorithm can be summarized as follows.

ALGORITHM IS

- IS1 Set $i = 1$, choose the sample size L and the starting value \mathbf{w}_1 .
- IS2 Draw a proposal value $\hat{\mathbf{w}}$ from the proposal density $p(\mathbf{w})$.

² In order to avoid cluttering the notation with subscripts, we write \mathbf{w} instead of \mathbf{v}_f and \mathbf{u} instead of \mathbf{v}_d from now on.

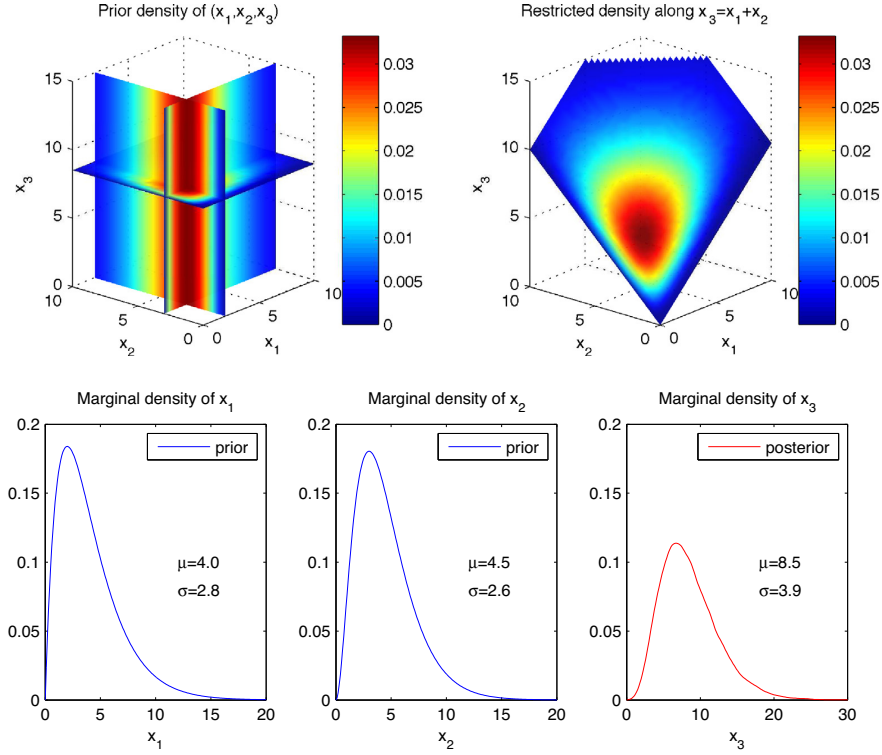


Fig. 3. Visualization of the densities in Example 1.3. Top left: prior density $f(x_1, x_2)$. Top right: prior density $f(x_1, x_2, x_3)$ on the plane $x_3 = x_1 + x_2$. Bottom: measured (prior, blue) and reconciled (posterior, red) densities of x_1, x_2, x_3 .

IS3 Compute the acceptance probability α by

$$\alpha(\mathbf{w}_i, \hat{\mathbf{w}}) = \min \left(1, \frac{\pi(\hat{\mathbf{w}})p(\mathbf{w}_i)}{\pi(\mathbf{w}_i)p(\hat{\mathbf{w}})} \right). \quad (12)$$

IS4 Draw a uniform random number $u \in [0, 1]$.

IS5 If $u \leq \alpha$, accept the proposal and set $\mathbf{w}_{i+1} = \hat{\mathbf{w}}$, otherwise set $\mathbf{w}_{i+1} = \mathbf{w}_i$.

IS6 Increase i by 1. If $i < L$, go to IS2, otherwise go to IS7.

IS7 Stop sampling.

The normalization constant of the posterior density $\pi(\mathbf{w})$ in Eq. (10) cancels out in the fraction in Eq. (12), therefore its value is irrelevant. In the context of DR, there is a natural choice of the proposal density for the IS, namely the prior density $p(\mathbf{w}) = f_{\mathbf{r}}(\mathbf{w})$. Consequently, no tuning of the proposal density is required. As an additional bonus, the acceptance probability has a particularly simple form:

$$\begin{aligned} \alpha(\mathbf{w}_i, \hat{\mathbf{w}}) &= \min \left(1, \frac{f_d(-D\hat{\mathbf{w}} - \mathbf{d})f_{\mathbf{r}}(\hat{\mathbf{w}})f_{\mathbf{r}}(\mathbf{w}_i)}{f_d(-D\mathbf{w}_i - \mathbf{d})f_{\mathbf{r}}(\mathbf{w}_i)f_{\mathbf{r}}(\hat{\mathbf{w}})} \right) \\ &= \min \left(1, \frac{f_d(-D\hat{\mathbf{w}} - \mathbf{d})}{f_d(-D\mathbf{w}_i - \mathbf{d})} \right). \end{aligned} \quad (13)$$

As all draws from the proposal density are equally valid starting points of the Markov chain, no “burn-in”, i.e. discarding an initial segment of the chain, is necessary.³

If the variables are not independent, the natural choice of the proposal density is the marginal prior density $p(\mathbf{w})$ of \mathbf{w} :

$$p(\mathbf{w}) = \int f(\mathbf{u}; \mathbf{w}) d\mathbf{u}$$

This is useful in practice only if the proposal density can be computed in closed form and is suitable for the fast generation of random draws. Otherwise an alternative proposal density has to be found.

3.2. The acceptance rate

The acceptance rate of the sampler is defined as the fraction of proposal values that are accepted in step IS5. If the acceptance rate is low, there will be many sequences of identical values in the generated chain, and the autocorrelation (see Brockwell and Davis, 2006) of the chain will be high. Such a chain contains less information about the posterior distribution as a chain with high acceptance rate and low autocorrelation. One possibility to keep the number of accepted proposal values at the same level is to estimate the acceptance rate a in a preliminary run of the sampler with sample size L_0 and to choose the final sample size L inversely proportional to a , i.e., $L = L_0/a$ with L_0 of the order of 10^5 . For a more refined way to choose L by using the autocorrelations of the chain, see Geyer (1992), where the concept of an effective sample size is introduced.

In the case of DR, the acceptance rate of the sampler is a rough indication of the extent to which the variables can be reconciled. A high acceptance rate indicates a large overlap between the prior and the posterior marginals, a low acceptance rate indicates a

³ This has been argued by C. Geyer on his website, see <http://users.stat.umn.edu/geyer/mcmc/burn.html> (accessed on 03.10.14).

small overlap. For a model with low acceptance rate, see [Example 4.4](#). It is shown there that the acceptance rate can be maximized by choosing the variables with the smallest variances as the free variables. However, even with an acceptance rate below 5% the posterior distribution in [Example 4.4](#) can be determined with sufficient precision by running the sampler for less than a minute on an off-the-shelf laptop. Given enough computing time, acceptance rates of a couple of percent can be tolerated.

If the priors have finite support it can happen that the acceptance rate is zero, showing that there is no overlap at all between the proposed values of one or several dependent variables and their prior distribution, in other words, that no DR is possible. This can be regarded as a problem, but also as an advantage of our approach: an irreconcilable discrepancy of the prior distributions is immediately visible. Determining which of the variables are responsible for the discrepancy is another matter, one which we intend to investigate in a subsequent paper.

In the IS the proposal values are drawn independently from the proposal distribution. It is therefore possible to generate a sample of proposal values in advance. In our MATLAB implementation this results in a nonnegligible gain of computational speed. It also opens up the possibility of running several Markov chains in parallel, each one with its own sample of proposal values. At the end all chains can be combined to a single posterior sample. This is particularly useful if the acceptance rate is low and the model is large.

3.3. Posterior analysis

Once a sample $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N)$ of $\mathbf{w} = \mathbf{v}_f$ has been generated, the corresponding sample $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)$ of $\mathbf{u} = \mathbf{v}_d$ can be computed by $\mathbf{u}_i = -\mathbf{D}\mathbf{w}_i - \mathbf{d}$, $i = 1, \dots, N$. From the complete sample $\mathbf{V} = (\mathbf{U}; \mathbf{W})$ posterior quantities such as means, quantiles, variances and correlations can be estimated. For a graphical representation of the posterior distributions the marginal densities can be estimated from the sample by a kernel estimator ([Silverman, 1998](#)) or obtained by smoothing the normalized frequency distribution in a histogram.

We have seen that the acceptance rate gives an indication of how well the data can be reconciled. It is, however, hard to interpret this in a quantitative way, so some formal measures of goodness are required. In DR with normal distributions there is a convenient measure of distance between the prior and the posterior distribution, namely the χ^2 -distance or its probability integral transform, defined by $u = F(\chi^2)$, where F is the cumulative distribution function of χ^2 . A value of u close to 1 indicates large discrepancies between the original and the reconciled data.

There are several ways to generalize this distance to the non-normal case. One possibility is the Mahalanobis distance ([Seber, 2004](#)) of the posterior sample from the prior distribution. This distance, however, is based on the first two moments of the sample, and therefore implicitly introduces an approximation by a normal distribution. We therefore propose to measure the goodness of the DR by computing the discrepancies between the prior and the posterior marginals. This takes into account the actual (nonnormal) shape of the marginals, and also displays the discrepancy of each individual variable. The discrepancy of the posterior sample can be computed with respect to the prior marginal, which is usually available in closed form, or with respect to a sufficiently large sample from the prior marginal.

Let us assume that we have a sample from the prior marginal distribution of v_i and a sample from the posterior marginal. A possible measure of discrepancy between the samples is the Kolmogorov–Smirnov statistic d_{KS} , i.e. the maximal absolute difference of the two empirical distribution functions ([Ross, 2009](#)). In MATLAB d_{KS} is computed by the function `kstest2`. Another possibility is $d_{BC} = 1 - BC$, where BC is the Bhattacharyya coefficient of the

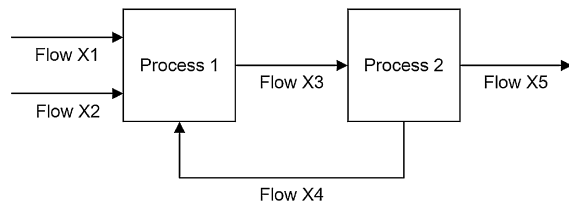


Fig. 4. Flowsheet of [Example 4.1](#) with five flows and two processes.

two samples ([Fukunaga, 1990](#)). A MATLAB function that computes BC can be obtained from the authors on request. Both measures of discrepancy lie in the interval $[0, 1]$, are equal to 0 for identical samples, and are equal to 1 for samples without overlap.

4. Examples

We illustrate the independence sampler on a couple of simple artificial examples. The MATLAB code used for these examples can be obtained from the authors on request.

Example 4.1. The first example is a small model with five flows and two processes, shown in [Fig. 4](#). Each mass flow is represented by one variable and the mass balances of the two processes form the constraints.

The variables are independent and have the following prior distributions:

$$f(x_1) = \text{Tria}(90, 100, 110),$$

$$f(x_2) = \Delta(50),$$

$$f(x_3) = \text{Unif}(270, 330),$$

$$f(x_4) = \text{unknown},$$

$$f(x_5) = \text{Trap}(140, 150, 160, 170).$$

$\Delta(a)$ denotes the Dirac distribution with mass 1 at $x = a$, $\text{Unif}(a, b)$ denotes the uniform distribution in the interval $[a, b]$, $\text{Tria}(a, c, b)$ denotes the triangular distribution in the interval $[a, b]$ with maximum at c , and $\text{Trap}(a, c, d, b)$ denotes the trapezoidal distribution in the interval $[a, b]$ that is uniform in $[c, d]$. The constraints are:

$$x_1 + x_2 + x_4 = x_3,$$

$$x_3 = x_4 + x_5.$$

In order to make the representation in Eq. (1) unique, we sort the measured variables by descending variance of the prior, so that variables with smaller variance are more likely to be chosen as free variables. This in turn maximizes the acceptance rate of the sampler. After grouping the variables into unmeasured, measured and fixed variables we get $\mathbf{y} = (x_4)$, $\mathbf{x} = (x_3, x_5, x_1)^T$ and $\mathbf{z} = (x_2)$. With this definition the constraints can be written in the form

$$\mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{C}\mathbf{z} = \mathbf{0} \quad \text{or} \quad \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{c} = \mathbf{0},$$

with

$$\mathbf{B} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 50 \\ 0 \end{pmatrix}.$$

The reduced row echelon form of $(\mathbf{B}, \mathbf{A}, \mathbf{c})$ is given by

$$\text{rref}(\mathbf{B}, \mathbf{A}, \mathbf{c}) = \left(\begin{array}{ccc|cc} 1 & -1 & 0 & 1 & 50 \\ 0 & 1 & -1 & -1 & -50 \end{array} \right).$$

Rows of the reduced row echelon form with zeros in all columns corresponding to unmeasured variables indicate that DR is possible. Measured variables with nonzero entries in the corresponding

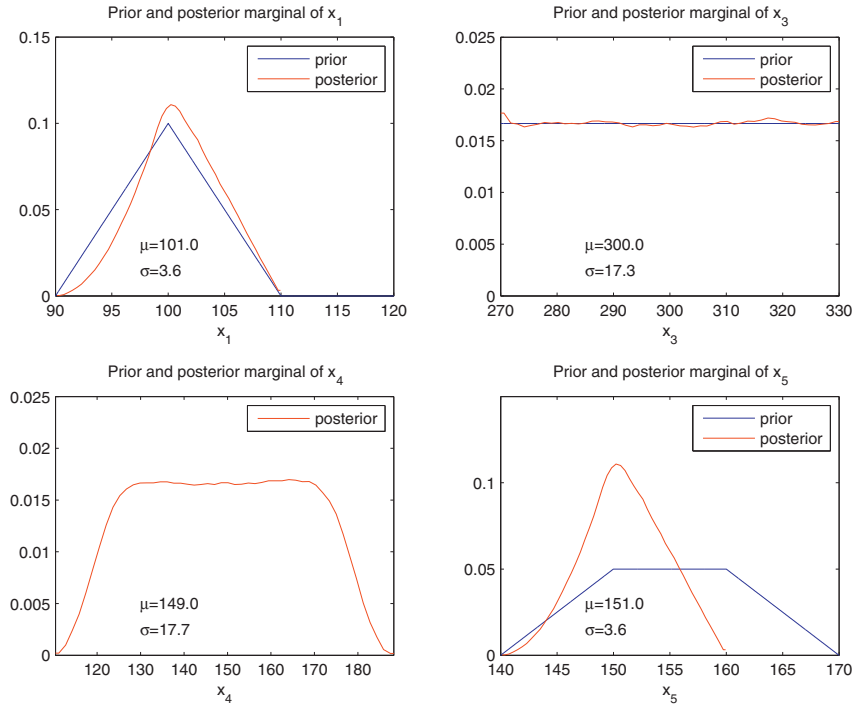


Fig. 5. Priors and smoothed marginal posteriors of x_1 , x_3 , x_4 , x_5 in Example 4.1. The improper prior of x_4 cannot be shown.

columns of these rows will be reconciled (they are redundant) while the rest of them remain unchanged (they are nonredundant). In our example x_3 is classified as nonredundant, so its posterior will be equal to its prior.

The column numbers of the pivot elements (leading 1 in each row) denote the dependent variables. All other variables are designated as free. If we regroup the variables into dependent and free ones by setting $\mathbf{v}_d = (x_4, x_5)^T$ and $\mathbf{v}_f = (x_3, x_1)^T$, the constraint equation can be written as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{v}_d + \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \mathbf{v}_f + \begin{pmatrix} 50 \\ -50 \end{pmatrix} = \mathbf{0}.$$

After rearranging the terms we get

$$\mathbf{v}_d = -\mathbf{D}\mathbf{v}_f - \mathbf{d}, \quad \text{with } \mathbf{D} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 50 \\ -50 \end{pmatrix}.$$

On a MacBook Pro with a 2.4 GHz Intel Core i5 processor the independence sampler requires about 10 s to draw a sample of size 100,000 from the posterior distribution of \mathbf{v}_f . The acceptance rate is about 86%, and the autocorrelation function (Brockwell and Davis, 2006) of the posterior samples decays very quickly. It is of the order of 0.15 at lag 1 and around 0.02 at lag 2. Fig. 5 shows the smoothed⁴ marginal posterior densities with their means and standard deviations. The latter are estimated from the posterior sample and are given with a precision that is commensurate with their standard error.

⁴ We have used the function `smooth` in the MATLAB Curve Fitting Toolbox.

A posteriori the variables are of course correlated. The posterior correlation coefficients are:

ρ	x_1	x_3	x_4	x_5
x_1	1.00	0.00	-0.20	1.00
x_3	0.00	1.00	0.98	0.00
x_4	-0.20	0.98	1.00	-0.20
x_5	1.00	0.00	-0.20	1.00

The high acceptance rate indicates a large overlap between the prior and the posterior marginals. The discrepancy as measured by d_{BC} is equal to 0.014, 0 and 0.172 for x_1 , x_3 and x_5 , respectively, while d_{KS} is equal to 0.116, 0.004 and 0.354, respectively.

Example 4.2. The second example is a small model with three flows and one process (Fig. 6). Each mass flow is represented by one variable and the mass balance of the process forms the constraint.

The measured variables are again independent and have the following prior distributions:

$$f(x_1) = \text{Unif}(10, 20),$$

$$f(x_2) = \text{Unif}(5, 15).$$

The constraint equation is

$$x_1 = x_2 + x_3.$$

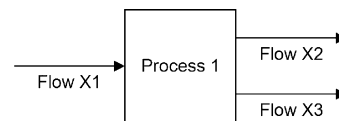


Fig. 6. Flowsheet of Example 4.2 with three flows and one process.

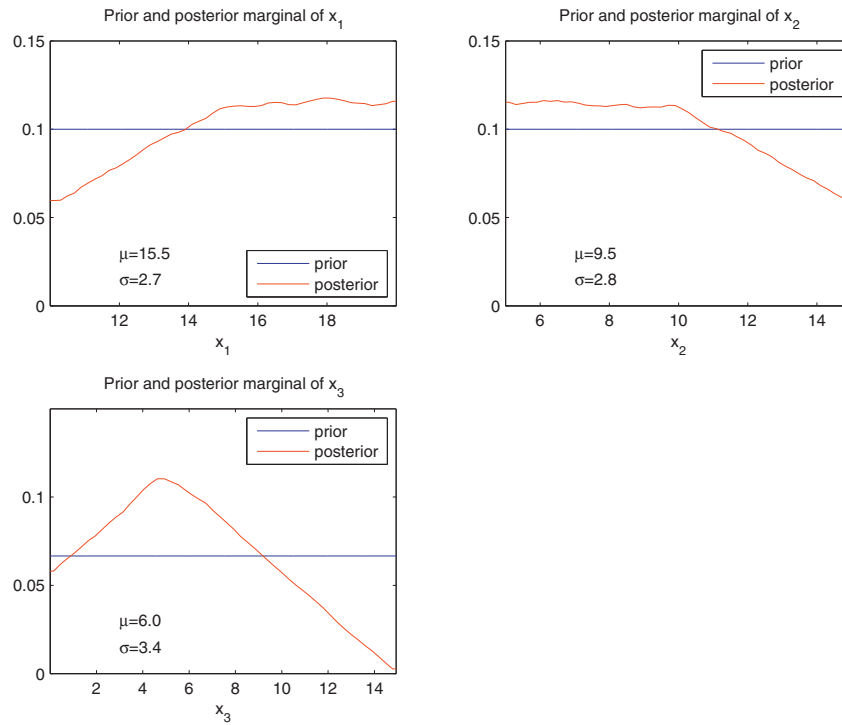


Fig. 7. Priors and smoothed marginal posteriors of x_1, x_2, x_3 in Example 4.2. The prior density of x_3 is $\text{Unif}(0, 15)$.

The only information about x_3 is that it is positive. In order to enforce positivity we use a uniform prior density that extends from 0 to the largest admissible value of x_3 , which is given by the constraint $x_3 \leq \max x_1 - \min x_2 = 20 - 5 = 15$:

$$f(x_3) = \text{Unif}(0, 15).$$

This is the distribution with maximal entropy (O'Hagan, 1994) in the admissible range. Even though x_3 is an unmeasured variable it has to be treated as a measured one because of the prior information introduced to enforce positivity. The model therefore contains only the measured variables $\mathbf{x} = (x_3, x_1, x_2)^T$, and the constraint can be written in the form

$$\mathbf{A}\mathbf{x} = \mathbf{0}, \text{ with } \mathbf{A} = \begin{pmatrix} -1 & 1 & -1 \end{pmatrix}$$

Note that the measured variables are again sorted by descending variance of the prior. The reduced row echelon form of (\mathbf{A}) is given by

$$\text{rref}(\mathbf{A}) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$$

If we regroup the variables into dependent and free ones by setting $\mathbf{v}_d = (x_3)$ and $\mathbf{v}_f = (x_1, x_2)^T$, the constraint equation can be written as

$$(1)\mathbf{v}_d + (-1 \ 1)\mathbf{v}_f = \mathbf{0}.$$

After rearranging the terms we get

$$\mathbf{v}_d = -\mathbf{D}\mathbf{v}_f, \text{ with } \mathbf{D} = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

The independence sampler requires about 11 s to draw a sample of size 100,000 from the posterior distribution of \mathbf{v}_f . The acceptance rate is about 87%. The autocorrelations decay even more quickly than in Example 4.1. Fig. 7 shows the smoothed marginal posterior densities with their means and standard deviations. The posterior correlation coefficients of (x_1, x_2, x_3) are:

ρ	x_1	x_2	x_3
x_1	1.00	0.23	0.62
x_2	0.23	1.00	-0.62
x_3	0.62	-0.62	1.00

The discrepancy between prior and posterior marginals as measured by d_{BC} is equal to 0.005, 0.005 and 0.042 for x_1, x_2 and x_3 , respectively, while d_{KS} is equal to 0.083, 0.078 and 0.196, respectively.

If there is prior information about the distribution of x_3 , it can be taken into account by a nonuniform prior distribution, for instance a triangular distribution, a scaled Beta distribution, or a truncated exponential distribution. If, for instance, the most probable value of x_3 is believed to be at $x_3 = 7$, a suitable prior is the triangular distribution $\text{Tria}(0, 7, 15)$. The corresponding posterior marginals are shown in Fig. 8. The acceptance rate drops to 65% and the discrepancies change somewhat: d_{BC} is equal to 0.014, 0.013 and 0.017 for x_1, x_2 and x_3 , respectively, while d_{KS} is equal to 0.131, 0.126 and 0.131, respectively.

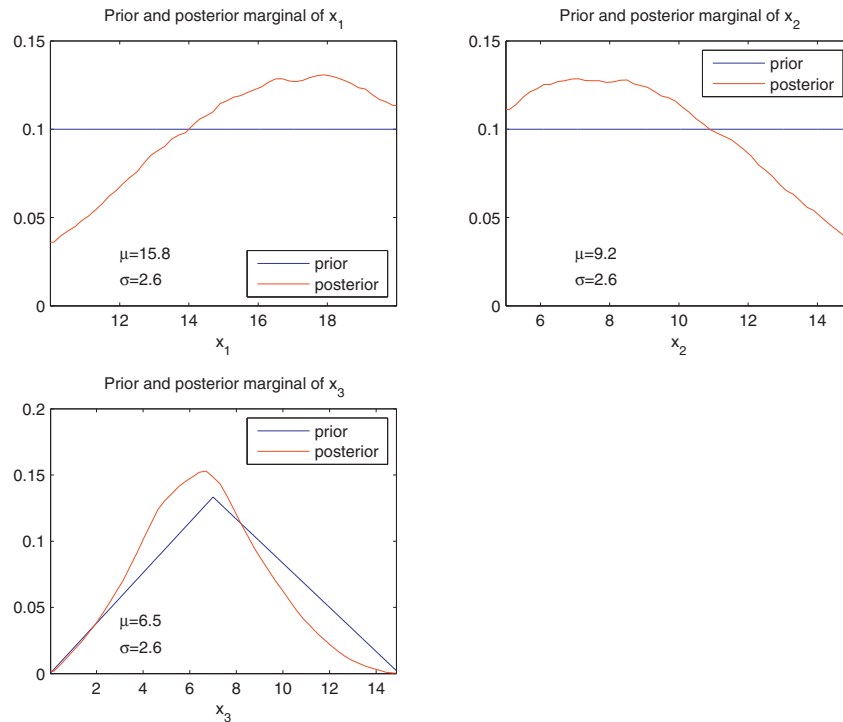


Fig. 8. Priors and smoothed marginal posteriors of x_1, x_2, x_3 in Example 4.2. The prior density of x_3 is $\text{Tri}(0, 7, 15)$.

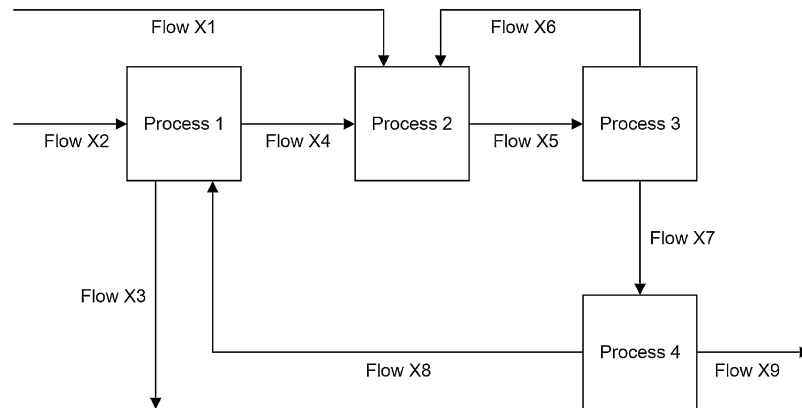


Fig. 9. Flowsheet of Example 4.3 with nine flows and four processes.

Example 4.3. The third example is a medium-sized model with nine flows and four processes (Fig. 9). Each mass flow is represented by one variable and the mass balances of the processes form the constraints. Variables x_3 and x_8 are unmeasured but assumed to be positive. Because they are observable (they can be calculated from the measured variables) they can be treated as dependent measured variables with an improper prior ($f(x) = 1$ for $x \geq 0$). All other variables are modeled by a log-normal prior with the following means m_i and standard deviations s_i :

	x_1	x_2	x_4	x_5	x_6	x_7	x_9
m	8	5	15	75	48	22	8
s	3	2	5	20	15	7	3

After sorting by descending variance we get $\mathbf{x} = (x_3, x_8, x_5, x_6, x_7, x_4, x_1, x_9, x_2)^T$. The constraints

$$x_2 + x_8 = x_3 + x_4,$$

$$x_1 + x_4 + x_6 = x_5,$$

$$x_5 = x_6 + x_7,$$

$$x_7 = x_8 + x_9,$$

can then be written in the form

$$\mathbf{Ax} = \mathbf{0},$$

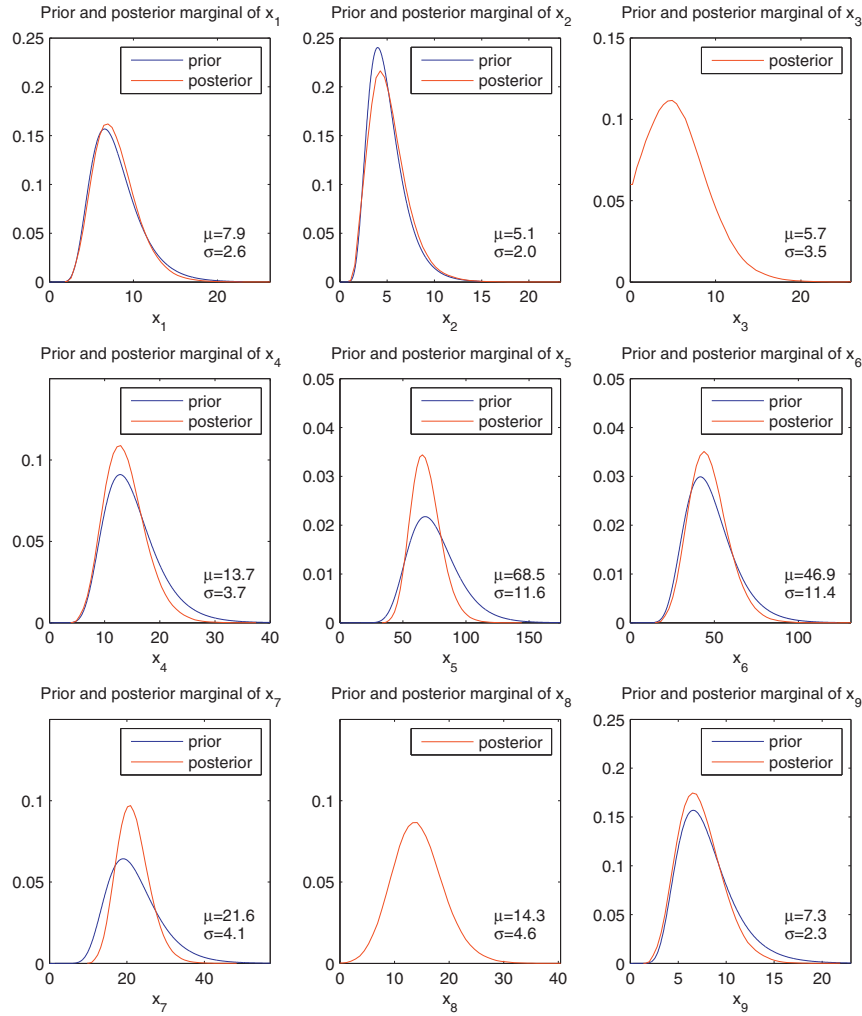


Fig. 10. Priors and smoothed marginal posteriors of all variables in Example 4.3. The improper priors of x_3 and x_8 cannot be shown.

with

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

The reduced row echelon form of \mathbf{A} is given by

$$\text{rref}(\mathbf{A}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 \end{pmatrix}.$$

The column numbers of the pivot elements (leading 1 in each row) denote the dependent variables. All other variables are designated as free. If we regroup the variables into dependent and free

ones by setting $\mathbf{v}_d = (x_3, x_8, x_5, x_7)^T$ and $\mathbf{v}_f = (x_6, x_4, x_1, x_9, x_2)^T$, the constraint equation can be written as

$$\mathbf{I}\mathbf{v}_d + \mathbf{D}\mathbf{v}_f = \mathbf{0} \quad \text{or} \quad \mathbf{v}_d = -\mathbf{D}\mathbf{v}_f,$$

with

$$\mathbf{D} = \begin{pmatrix} 0 & 0 & -1 & 1 & -1 \\ 0 & -1 & -1 & 1 & 0 \\ -1 & -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \end{pmatrix}.$$

The independence sampler requires about 12 s to draw a sample of size 100,000 from the posterior distribution of \mathbf{v}_f . The acceptance rate is about 62%. The autocorrelations decay somewhat more slowly than in the preceding examples; they are below 0.1 at lag 3. Fig. 10 shows the smoothed marginal posterior densities with their means and standard deviations. The discrepancies between prior and posterior marginals are shown in Table 1. The values for

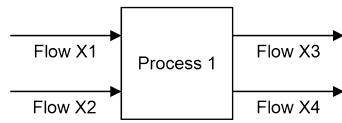


Fig. 11. Flowsheet of Example 4.4 with four flows and one process.

the actually measured variables indicate large overlaps and thus a satisfactory reconciliation.

Example 4.4. A low acceptance rate of the sampler indicates little overlap between the prior and the posterior distributions of some of the measured variables. It is, however, not always possible to identify the variable(s) causing the discrepancy. This is illustrated by the model in Fig. 11.

The model has four flows and one process. The mass balance of the process yields the following constraint equation:

$$x_1 + x_2 = x_3 + x_4.$$

All variables are measured, independent, and have the following prior distributions:

$$f(x_1) = \text{Tria}(10, 15, 20), \quad \mu = 15.0, \quad \sigma^2 = 4.1\hat{6},$$

$$f(x_2) = \text{Tria}(20, 25, 30), \quad \mu = 25.0, \quad \sigma^2 = 4.1\hat{6},$$

$$f(x_3) = \text{Tria}(10, 20, 25), \quad \mu = 18.\hat{3}, \quad \sigma^2 = 9.7\hat{2},$$

$$f(x_4) = \text{Tria}(25, 35, 45), \quad \mu = 35.0, \quad \sigma^2 = 16.\hat{6}.$$

In view of the simplicity of this example we skip the detailed calculation of **A** and **D**, as it has already been illustrated in the previous examples.

Table 1

Discrepancy between prior and posterior marginals in Example 4.3.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
d_{BC}	0.004	0.001	–	0.016	0.059	0.013	0.057	–	0.013
d_{KS}	0.031	0.038	–	0.114	0.200	0.065	0.131	–	0.089

Table 2

Acceptance rate α as a function of the dependent variable v_d in Example 4.4.

v_d	x_1	x_2	x_3	x_4
α [%]	4.1	4.1	9.3	12.5

Table 3

Discrepancy between prior and posterior marginals in Example 4.4.

	x_1	x_2	x_3	x_4
d_{BC}	0.090	0.089	0.270	0.363
d_{KS}	0.338	0.333	0.569	0.645

Each of the four variables can be chosen as the dependent one; the other three are free. Depending on the choice of the dependent variable, the acceptance rate varies between 4% and 12%. The independence sampler requires about 36 s to draw a sample of size 500,000 from the posterior distribution of \mathbf{v}_f . The posterior densities should not and do not depend on the choice of the dependent variable, but the acceptance rate rises significantly with the variance of the dependent variable, see Table 2. Again the optimal order of the measured variables is obtained by sorting them by descending variance.

Fig. 12 shows the prior and the posterior densities with x_4 as the dependent variable, and Table 3 shows the discrepancy between the prior and the posterior marginals.

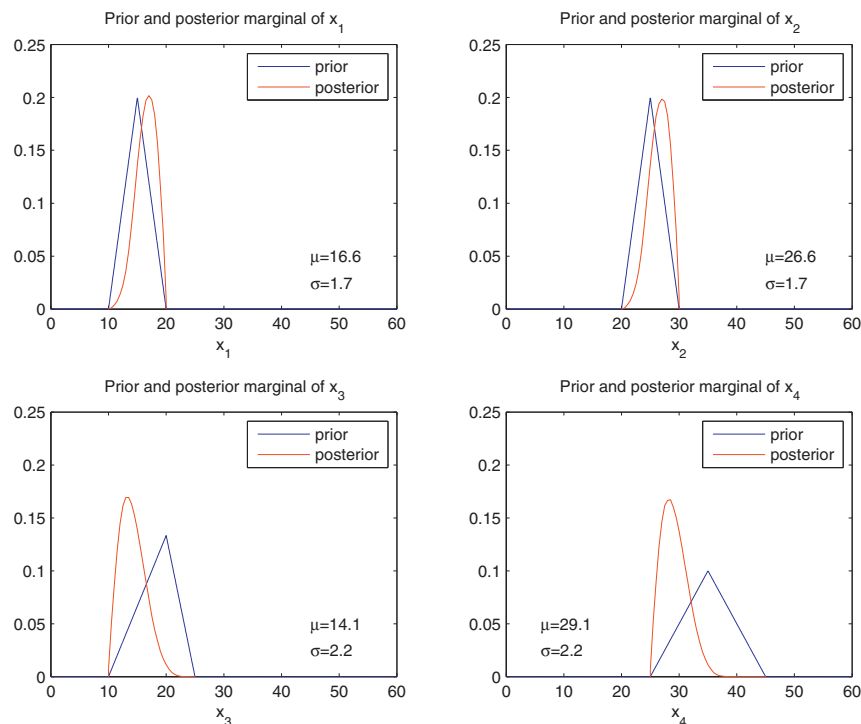


Fig. 12. Priors and smoothed marginal posteriors of all variables in Example 4.4. The dependent variable is x_4 .

A better fit and a higher acceptance rate can be achieved by shifting or rescaling anyone of the four variables. Unfortunately, in this simple case it is not possible to conclude from the posteriors which variable should be considered as the outlier. Still, it is conceivable that in a larger model with more constraints outliers can be identified. However, this issue is outside the scope of the present work and will be addressed in a subsequent paper.

5. Discussion and Outlook

We have presented a new method to reconcile nonnormally distributed data by a Markov chain Monte Carlo method. Our method of choice is the independence sampler. This sampler has several advantages. First, there is a natural proposal density that does not have to be tuned. Second, there is no need to discard an initial segment of the chain (“burn-in”). Third, the acceptance probability has an extremely simple mathematical form. Finally, the proposal values can be generated in advance, and many chains can be generated in parallel.

The acceptance rate of the sampler gives a rough indication of how well the variables can be reconciled. In the extreme case of the acceptance rate being zero it can be concluded immediately that no reconciliation is possible. We have also proposed two measures of discrepancy between the original (prior) and the reconciled (posterior) distributions of the variables. They can be used to quantify the goodness of the reconciliation.

In this paper we have dealt only with linear equality constraints. However, the presented approach is also able to solve inequality constraints by introducing slack variables. For instance, the inequality constraint $x_1 \leq x_2 + x_3$ can be transformed to the equality constraint $x_1 + x_5 = x_2 + x_3$ with an unmeasured slack variable $x_5 \geq 0$. The only restriction is that slack variables have to be classified as observable unmeasured variables in order to be treated as dependent measured variables with a proper or improper prior enforcing positivity. In fact, the first part of Example 4.2 can be interpreted as the solution to the inequality constraint $x_2 \leq x_1$ with measured x_1 and x_2, x_3 serving as the slack variable. Inequality constraints will be studied in more detail in a subsequent paper, as will be nonlinearly constrained variables.

Finally, we also intend to investigate how and to which extent the DR procedure can be robustified. This implies the automatic identification of outlying or erroneous measurements that prevent a successful reconciliation. Such measurements can then be either eliminated or given a larger uncertainty in the reconciliation process.

Acknowledgements

We thank N. Dzibur, D. Laner and H. Rechberger for useful comments on the manuscript. We also thank the anonymous referees for their valuable suggestions that have helped to improve the exposition.

Appendix A.

We sketch a proof that in the case of normally distributed observations Eq. (10) gives the same result as the traditional method, which is constrained least-squares estimation. As the unmeasured variables are known linear functions of the measured variables, the proof can be restricted to the measured variables. We denote the free variables by \mathbf{v}_f and the measured dependent variables by \mathbf{v}_d . If $\mathbf{v} = (\mathbf{v}_d; \mathbf{v}_f)$ is normally distributed, both f_d and f_f are multivariate normal density functions:

$$f_d(\mathbf{v}_d) = \varphi(\mathbf{v}_d; \boldsymbol{\mu}_d, \mathbf{V}_d), \quad f_f(\mathbf{v}_f) = \varphi(\mathbf{v}_f; \boldsymbol{\mu}_f, \mathbf{V}_f).$$

Eq. (10) states that the posterior distribution of \mathbf{v}_f is proportional to the product $f_d(-\mathbf{D}\mathbf{v}_f - \mathbf{d})f_f(\mathbf{v}_f)$. The quadratic polynomial in the exponent of the product can be rearranged as the sum of a quadratic form and a constant term. The product is therefore proportional to a multivariate normal density. The mean $\boldsymbol{\mu}_\pi$ and the covariance matrix \mathbf{V}_π can be read off the quadratic form and are the following:

$$\boldsymbol{\mu}_\pi = \mathbf{V}_\pi [\mathbf{V}_f^{-1} \boldsymbol{\mu}_f - \mathbf{D}^T \mathbf{V}_d^{-1} (\boldsymbol{\mu}_d + \mathbf{d})], \quad \mathbf{V}_\pi = [\mathbf{V}_f^{-1} + \mathbf{D}^T \mathbf{V}_d^{-1} \mathbf{D}]^{-1}.$$

The unbiased estimator of $\boldsymbol{\mu}_\pi$ is then equal to:

$$\tilde{\boldsymbol{\mu}}_\pi = \mathbf{V}_\pi [\mathbf{V}_f^{-1} \mathbf{v}_f - \mathbf{D}^T \mathbf{V}_d^{-1} (\mathbf{v}_d + \mathbf{d})]. \quad (14)$$

Next we derive the constrained least-squares estimator of $\boldsymbol{\mu} = (\boldsymbol{\mu}_d; \boldsymbol{\mu}_f)$ and show that its free component is equal to $\tilde{\boldsymbol{\mu}}_\pi$. The constraint Eq. (2) can be written in the form:

$$\mathbf{K}\boldsymbol{\mu} + \mathbf{d} = \mathbf{0}, \quad \text{with } \mathbf{K} = (\mathbf{I}, \mathbf{D}).$$

The constrained least-squares estimate of $\boldsymbol{\mu}$ is obtained by minimizing the following objective function:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = (\mathbf{v} - \boldsymbol{\mu})^T \mathbf{G}(\mathbf{v} - \boldsymbol{\mu}) + 2\boldsymbol{\lambda}^T (\mathbf{K}\boldsymbol{\mu} + \mathbf{d}),$$

where \mathbf{G} is the inverse of the covariance matrix \mathbf{V} of \mathbf{v} and $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers. Note that \mathbf{V} is a block-diagonal matrix with blocks \mathbf{V}_d and \mathbf{V}_f . Setting the gradient $\nabla \mathcal{L}$ of \mathcal{L} to zero results in the following system of linear equations:

$$\mathbf{G}\boldsymbol{\mu} + \mathbf{K}^T \boldsymbol{\lambda} = \mathbf{G}\mathbf{v},$$

$$\mathbf{K}\boldsymbol{\mu} + \mathbf{d} = \mathbf{0}.$$

Solving with respect to $\boldsymbol{\mu}$ yields the constrained minimum $\tilde{\boldsymbol{\mu}}$ of \mathcal{L} :

$$\tilde{\boldsymbol{\mu}} = \mathbf{v} + \mathbf{V}\mathbf{K}^T (\mathbf{K}\mathbf{V}\mathbf{K}^T)^{-1} (-\mathbf{d} - \mathbf{K}\mathbf{v}).$$

If the solution $\tilde{\boldsymbol{\mu}} = (\tilde{\boldsymbol{\mu}}_d; \tilde{\boldsymbol{\mu}}_f)$ is split into the dependent and the free component, we obtain for the free component:

$$\tilde{\boldsymbol{\mu}}_f = \mathbf{v}_f + \mathbf{V}_f \mathbf{D}^T (\mathbf{V}_d + \mathbf{D}\mathbf{V}_f \mathbf{D}^T)^{-1} (-\mathbf{d} - \mathbf{v}_d - \mathbf{D}\mathbf{v}_f).$$

The right hand side has the form of an update step of the Kalman filter (Jazwinski, 1970). Using the Woodbury matrix identity (Press et al., 2007), it can be transformed to the form of a weighted mean:

$$\tilde{\boldsymbol{\mu}}_f = [\mathbf{V}_f^{-1} + \mathbf{D}^T \mathbf{V}_d^{-1} \mathbf{D}]^{-1} [\mathbf{V}_f^{-1} \mathbf{v}_f + \mathbf{D}^T \mathbf{V}_d^{-1} (-\mathbf{d} - \mathbf{v}_d)]. \quad (15)$$

Obviously the estimated posterior mean in Eq. (14) and the constrained least-squares estimate in Eq. (15) are identical. This completes the proof.

References

- Brooks S, Gelman A, Jones GL, Xiao-Li Meng. In: *Handbook of Markov chain Monte Carlo*. Boca Raton: Chapman & Hall; 2011.
- Bagajewicz MJ. *Smart process plants: software and hardware solutions for accurate data and profitable operations*. McGraw-Hill: New York; 2010.
- Brockwell PJ, Davis RA. *Time series: theory and methods*. New York: Springer; 2006.
- Chib S, Greenberg E. Understanding the Metropolis–Hastings algorithm. *Am Stat* 1995;49(4):327–35.
- Crowe CM. Formulation of linear data reconciliation using information theory. *Chem Eng Sci* 1996;51(12):3359–66.
- Fukunaga K. *Introduction to statistical pattern recognition*. San Diego: Academic Press; 1990.
- Geyer C. Practical Markov chain Monte Carlo (with discussion). *Stat Sci* 1992;7:473–511.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 1970;57:97–107.
- Jazwinski A. *Stochastic processes and filtering theory* (reprinted by Dover, Mineola, 2007). New York: Academic Press; 1970.
- Johnston LPM, Kramer MA. Maximum likelihood data rectification: steady-state systems. *AIChE J* 1995;24(11):2415–26.
- Laner D, Rechberger H, Astrup T. Systematic evaluation of uncertainty in material flow analysis. *J Ind Ecol* 2014. <http://dx.doi.org/10.1111/jiec.12143> (in press).
- Liu JS. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Stat Comput* 1996;6:113–9.

- Liu JS. Monte Carlo strategies in scientific computing. New York: Springer; 2004.
- Madron F. Process plant performance. New York: Ellis Horwood; 1992.
- Metropolis N, et al. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087–91.
- Narasimhan S, Jordache C. Data reconciliation and gross error detection. An intelligent use of process data. Houston: Gulf Publishing Company; 2000.
- O'Hagan A. The advanced theory of statistics. Bayesian inference, vol. 2B. London: Arnold; 1994.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical recipes: the art of scientific computing*, section 2.7.3.0. 3rd ed. New York: Cambridge University Press; 2007.
- Robert C, Casella C. Monte Carlo statistical methods. Springer: New York; 2004.
- Romagnoli JA, Sanchez MC. Data processing and reconciliation for chemical process operations. San Diego: Academic Press; 2000.
- Ross SM. Introduction to probability and statistics for engineers and scientists. Burlington, San Diego, London: Elsevier Academic Press; 2009.
- Seber GAF. Multivariate observations. Hoboken: John Wiley & Sons; 2004.
- Silverman BW. Density estimation for statistics and data analysis. London: Chapman & Hall/CRC; 1998.
- Tierney L. Markov chains for exploring posterior distributions. *Ann Stat* 1994;22:1701–62.

Paper 3:

**Data reconciliation of nonnormal observations
with nonlinear constraints**

Oliver Cencic and Rudolf Frühwirth

Journal of Applied Statistics 2018, in press

DOI: 10.1080/02664763.2017.1421916

Data reconciliation of nonnormal observations with nonlinear constraints

Oliver Cencic^a and Rudolf Frühwirth^{b,c}

^aInstitute for Water Quality, Resource and Waste Management, TU Wien, Wien, Austria; ^bInstitute of High Energy Physics (HEPHY), Austrian Academy of Sciences, Wien, Austria; ^cInstitute of Statistics and Mathematical Methods in Economics, TU Wien, Wien, Austria

ABSTRACT

This paper presents a new method for the reconciliation of data described by arbitrary continuous probability distributions, with the focus on nonlinear constraints. The main idea, already applied to linear constraints in a previous paper, is to restrict the joint prior probability distribution of the observed variables with model constraints to get a joint posterior probability distribution. Because in general the posterior probability density function cannot be calculated analytically, it is shown that it has decisive advantages to sample from the posterior distribution by a Markov chain Monte Carlo (MCMC) method. From the resulting sample of observed and unobserved variables various characteristics of the posterior distribution can be estimated, such as the mean, the full covariance matrix, marginal posterior densities, as well as marginal moments, quantiles, and HPD intervals. The procedure is illustrated by examples from material flow analysis and chemical engineering.

ARTICLE HISTORY

Received 23 June 2017

Accepted 21 December 2017

KEYWORDS



Data reconciliation; Bayesian approach; nonlinear constraints; nonnormal distributions; Markov chain Monte Carlo

1. Introduction

Data reconciliation (DR) is a statistical method to adjust observed data that are, due to measurement or sampling errors, in conflict with known constraints such as the conservation laws of mass and energy. By imposing the constraints on the observations external information is added, and consequently the reconciled values have smaller uncertainty than the original observations. In addition, unobserved variables occurring in the constraints can be estimated, as long as the assumptions in Section 2.2 are fulfilled.

Traditionally, DR is formulated as a minimization problem where the observations are adjusted so that they obey a set of constraint equations and, at the same time, minimize an objective function. An important example of such an objective function is the sum of squared differences (weighted or unweighted) between the adjusted and the original values, resulting in a least-squares reconciliation.

In the simplest case, the observations \tilde{x} are normally distributed with unknown mean vector x and known or estimated joint covariance matrix Σ . Furthermore, the vector x is

CONTACT Oliver Cencic  oliver.cencic@tuwien.ac.at  Institute for Water Quality, Resource and Waste Management, TU Wien, Karlsplatz 13/226, A-1040 Wien, Austria

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

known to obey a system of linear equality constraints:

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{c} = \mathbf{0}, \quad (1)$$

where \mathbf{y} is a vector of unobserved variables, \mathbf{A} and \mathbf{B} are known coefficient matrices of \mathbf{x} and \mathbf{y} , respectively, and \mathbf{c} is a vector of known constants. Under certain conditions (see Appendix), a unique explicit solution can be found by constrained minimization with respect to \mathbf{x} of the following least-squares objective function:

$$F(\mathbf{x}) = (\tilde{\mathbf{x}} - \mathbf{x})^T \boldsymbol{\Sigma}^{-1} (\tilde{\mathbf{x}} - \mathbf{x}), \quad (2)$$

usually with the help of Lagrange multipliers [19]. The least-squares solution $(\hat{\mathbf{y}}; \hat{\mathbf{x}})^1$ fulfills the constraints, the reconciled vector $\hat{\mathbf{x}}$ is an unbiased estimator of \mathbf{x} , and the variances of $\hat{\mathbf{x}}$ are smaller than the variances of $\tilde{\mathbf{x}}$. As the covariance matrix of $\hat{\mathbf{x}}$ is rank deficient, its distribution is singular, i.e. concentrated on a set of Lebesgue measure zero, and does not have a pdf (cf. Section 2.3). If $\tilde{\mathbf{x}}$ follows a normal distribution, so does the solution $(\hat{\mathbf{y}}; \hat{\mathbf{x}})$; otherwise it does not.

This basic problem of DR can be generalized in two directions: first, the constraints can be extended to nonlinear and/or inequality constraints; second, the distribution of the observations $\tilde{\mathbf{x}}$ can be extended to arbitrary, but absolutely continuous, multivariate distributions, characterized by a probability density function (pdf) $f(\tilde{\mathbf{x}})$.

The first extension to nonlinear and/or inequality constraints is conceptually simple, but may pose technical difficulties. The solution can no longer be obtained in closed form, but has to be computed iteratively by suitable algorithms for the constrained minimization of $F(\mathbf{x})$, with the usual potential problems of convergence, local minima and saddle points. There is a large variety of methods that can be applied. Some examples in the engineering literature are successive linearization [10], nonlinear programming [1], quasi-Newton methods [14] and genetic algorithms [23]. For textbooks on optimization with nonlinear constraints, see, for instance, [5,6,22]. Note that with nonlinear constraints the reconciled values are not normally distributed, even if the observations $\tilde{\mathbf{x}}$ are.

The second extension to nonnormal observations can be achieved by an alternative approach to the problem of DR. As the objective function in Equation (2) uses only the first two moments of the distribution, it is not able to take into account the full information contained in the joint pdf of the observations. Consequently, the reconciliation problem cannot be fully solved by minimizing an objective function of the form of Equation (2), even if the joint pdf of the observations belongs to a parametric family. In the case of linear constraints, the constrained least-squares estimator is unbiased and a linear function of the observations, and therefore gives the correct mean and covariance matrix of the reconciled values. Their distribution, however, is not known, and it is not possible to compute quantiles or higher moments such as the skewness. This problem was solved in [3] by a Bayesian method that gives the joint (posterior) distribution of the reconciled variables under linear constraints for arbitrary (prior) distributions of the observations. In the appendix, it is proved that this method delivers the same results as the least-squares approach in the case of linear constraints and normal observations with arbitrary covariance matrix $\boldsymbol{\Sigma}$.

This paper extends the method of Cencic and Frühwirth [3] to the case of nonlinear constraints, which usually are present in all but the simplest models. In Section 2 the basic assumptions about nonnormal models with nonlinear constraints are spelled out, and the

posterior pdf of the free variables of the model is derived. It is shown that the joint posterior distribution of all variables, derived from the posterior distribution of the free variables, is invariant under the choice of the free variables. Section 3 presents the independence sampler that produces a Markov chain of all variables, observed and unobserved. From the Markov chain, important characteristics of the posterior distribution can be estimated, among them the mean and the covariance matrix, marginal pdfs, marginal moments and marginal quantiles. In Section 4 the sampler is applied to two examples, one from the field of material flow analysis (MFA), the other from the chemical engineering literature. Finally, Section 5 presents conclusions and gives an outlook to further work.

2. Nonnormal models with nonlinear constraints

2.1. Motivation and specification of priors

As the objective function in Equation (2) uses only the first two moments of the joint distribution of the observations, it is based on the implicit assumption of normally distributed observation errors. This assumption guarantees normally distributed reconciled observations in case of linear or linearized constraints, where in the latter case the results are only approximations. Sometimes, these approximations are not sufficient and the precise pdfs of the results are desired instead.

Additionally, there might be variables involved in the problem where the normal assumption is not appropriate. First, variables can be bounded in some ways. A typical example is found in MFA, where it is crucial to close the mass balances of goods and substances traversing different kinds of processes in order to get reliable data for subsequent evaluation. By definition, mass flows cannot take negative values, and transfer coefficients, which describe how much of the mass going into a process is transferred into specific output flows, as well as mass fractions of substances in goods are restricted to the unit interval, rendering the assumption of normality implausible. Second, modelling the uncertainty of observations by normal distributions can be too restrictive. If a sufficient number of observations of a quantity is available and the observations fail a test on normality (see, for instance [17]), one can fit a nonnormal parametric model to the observed data or construct a nonparametric model such as the empirical distribution function or the kernel estimate of the pdf. Third, if there are few or no observations of a quantity of interest, which is frequently the case in MFA, expert opinions have to be relied on. The more detailed the expert's knowledge about the quantity under consideration is, the more precisely the distribution can be modelled. Common choices in this context (with rising complexity) are uniform, triangular or trapezoidal distributions.

The data reconciliation method presented here is able to deal with arbitrary pdfs describing the uncertainty of an observation, regardless of the procedure by which it has been derived. Thus, nonnormal models can be used as they are, without any need for normal approximation. Before the constraints are imposed, the joint pdf of the observations, built from these nonnormal models, is called the 'prior' pdf. As in least-squares DR, the joint distribution of the reconciled observations is singular; there is, however, a subset of the variables (called the free variables in the following) with a nonsingular distribution and a pdf, called the 'posterior' pdf. Although Bayes' theorem is not invoked in the derivation of the posterior pdf, the terms 'prior' and 'posterior' can be justified by the fact

that the posterior pdf is a conditional density, namely the prior pdf conditioned on the constraints.

2.2. Basic assumptions

Let $\mathbf{m} = (\mathbf{y}; \mathbf{x})$ be a column vector of n variables that is arranged such that \mathbf{y} contains the n_y unobserved variables and \mathbf{x} contains the $n_x = n - n_y$ observed variables. In addition, there is a system of n_c independent constraint equations, where $n_y \leq n_c \leq n$:

$$\mathbf{G}(\mathbf{y}; \mathbf{x}) = \mathbf{0}, \quad \mathbf{G} : \mathbb{R}^n \longrightarrow \mathbb{R}^{n_c}. \quad (3)$$

The function \mathbf{G} usually also depends on a vector \mathbf{z} of known constants. In order to simplify the notation, this dependence is not shown explicitly.

The derivation of the posterior pdf of the observed variables is based on the following assumptions:

- (1) The n observed or unobserved variables take values in a subset $D \subseteq \mathbb{R}^n$.
- (2) The observations of the $n_x \leq n$ observed variables \mathbf{x} form an n_x -dimensional random variable with known joint pdf, called the prior pdf of the observed variables. As outlined in Section 2.1, the prior pdf of an observed variable can result from an individual measurement, from a collection of measurements, or from capturing expert opinion in a pdf.
- (3) The n variables are subject to n_c independent nonlinear equality constraints that define a differentiable manifold $S \subset \mathbb{R}^n$ of dimension $n - n_c$.
- (4) The n_c nonlinear equality constraints can be uniquely solved for a set of n_c dependent variables (observed or unobserved) that are differentiable functions of $n - n_c$ free variables (observed only), for each admissible value of the free variables.

There is no prior information on the unobserved variables \mathbf{y} . Their posterior distribution is computed from the posterior distribution of the free variables using the model constraints.

Assumption 4 implies that potentially involved unobservable variables have already been eliminated from the system of equations. Additional constraints such as positivity or boundedness can be incorporated directly in the prior distributions of the observed variables and need not be enforced explicitly.² Inequality constraints can be transformed into equality constraints by introducing slack variables with uninformative priors restricted to positive values. For instance, the inequality constraint $x_1 \leq x_2 x_3$ can be transformed to $x_1 + x_s = x_2 x_3$ with $x_s \geq 0$ *a priori*.

2.3. The posterior pdf with nonlinear constraints

By assumption, the nonlinear constraints are solved by representing the n_y unobserved variables \mathbf{y} and the $n_u = n_c - n_y$ dependent observed variables \mathbf{u} as functions of the $n_w = n - n_c$ chosen free observed variables $\mathbf{w} \in W$, where $W \subseteq \mathbb{R}^{n_w}$ is the domain of the free variables:

$$\mathbf{y} = \mathbf{k}(\mathbf{w}), \quad (4)$$

$$\mathbf{u} = \mathbf{h}(\mathbf{w}). \quad (5)$$

It is assumed that the function \mathbf{h} is differentiable everywhere in W , with the following Jacobian:

$$\mathbf{H}(\mathbf{w}) = \frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}}, \quad \mathbf{w} \in W. \quad (6)$$

If the constraints are sufficiently simple, the functions \mathbf{k} and \mathbf{h} can be computed explicitly. In general, however, the nonlinear constraint equations have to be solved numerically for \mathbf{y} and \mathbf{u} . The linearization method combined with Newton–Raphson iteration is a common choice in this context. If this fails, gradient-free methods such as the simplex algorithm [11] are viable alternatives to compute \mathbf{y} and \mathbf{u} for given \mathbf{w} (see Example 4.2).

The joint prior pdf does not depend on the unobserved variables \mathbf{y} and can be written as $f(\mathbf{u}; \mathbf{w})$. DR is tantamount to conditioning the prior distribution of the observed variables on the constraints in Equation (5) and to find the resulting posterior distribution. Because the unobserved variables \mathbf{y} have thus been eliminated from the DR problem, the constraint manifold S can now be considered as a n_w -dimensional differentiable manifold in \mathbb{R}^{n_x} with the following embedding:

$$\boldsymbol{\varphi}(\mathbf{w}) = (\mathbf{h}(\mathbf{w}); \mathbf{w}), \quad \mathbf{w} \in W. \quad (7)$$

The constraint manifold S has Lebesgue measure zero in \mathbb{R}^{n_x} , so the prior distribution restricted to S is singular and has no pdf with respect to the Lebesgue measure. It can, however, be characterized by a pdf with respect to the measure dS on S that is induced by the metric structure of S [15,16]. This pdf is then transformed to a pdf $\pi(\mathbf{w})$ in terms of the free variables \mathbf{w} (with respect to the Lebesgue measure in $W \subseteq \mathbb{R}^{n_w}$), which is the sought-after reconciled or posterior pdf. It is uniquely determined by the prior pdf and the set of constraints. In perfect analogy to DR by least squares, where the covariance matrix of $\hat{\mathbf{x}}$ is rank deficient, there is no joint posterior pdf of all variables. As the constraint manifold S is embedded into \mathbb{R}^{n_x} by the mapping in Equation (7), the metric tensor of the induced metric in S is equal to $T(\mathbf{w}) = \mathbf{I} + \mathbf{H}^T \mathbf{H}$, where \mathbf{H} is evaluated at \mathbf{w} [13]. The metric $T(\mathbf{w})$ induces a differential volume element in the n_w -dimensional tangent space at the point $(\mathbf{h}(\mathbf{w}); \mathbf{w})$, and thus a measure on the manifold [15,16]. The differential volume element $dS(\mathbf{w})$ of S at \mathbf{w} is then given by

$$dS(\mathbf{w}) = V(\mathbf{w}) d\mathbf{w}, \quad \text{with } V(\mathbf{w}) = \sqrt{|\mathbf{I} + \mathbf{H}^T \mathbf{H}|}. \quad (8)$$

Depending on the dimension of S , $dS(\mathbf{w})$ can actually be the differential element of a curve, a surface, a volume or a hypervolume. If S is a curve $u = h_1(w)$ in \mathbb{R}^2 or a surface $u = h_2(w_1, w_2)$ in \mathbb{R}^3 , the well-known respective expressions for the differential of the arc length and the differential of the area are recovered:

$$u = h_1(w) \implies dS = \sqrt{1 + (dh_1/dw)^2} dw, \quad (9)$$

$$u = h_2(w_1, w_2) \implies dS = \sqrt{1 + (\partial h_2 / \partial w_1)^2 + (\partial h_2 / \partial w_2)^2} dw_1 dw_2. \quad (10)$$

The prior pdf $f(\mathbf{u}; \mathbf{w})$, restricted to S and written as a function of \mathbf{w} only, is equal to $f(\mathbf{h}(\mathbf{w}); \mathbf{w})$, which in turn is proportional to the pdf $f_S(\mathbf{w})$ with respect to the measure dS

on S :

$$f_S(\mathbf{w}) = (1/C) f(\mathbf{h}(\mathbf{w}), \mathbf{w}), \quad \text{with } C = \int_S f(\mathbf{h}(\mathbf{w}), \mathbf{w}) dS(\mathbf{w}). \quad (11)$$

Finally, the posterior pdf $f_S(\mathbf{w})$ on S is transformed to the posterior pdf $\pi(\mathbf{w})$ on W , the domain of the free variables, with respect to the Lebesgue measure in W . In order to conserve probabilities, the differential posterior probability $\pi(\mathbf{w}) d\mathbf{w}$ must be equal to the differential probability $f_S(\mathbf{w}) dS(\mathbf{w})$ of the corresponding differential volume element $dS(\mathbf{w})$:

$$\pi(\mathbf{w}) d\mathbf{w} = f_S(\mathbf{w}) dS(\mathbf{w}) \propto f(\mathbf{h}(\mathbf{w}), \mathbf{w}) V(\mathbf{w}) d\mathbf{w}. \quad (12)$$

The posterior pdf of the free variables can therefore be written as follows:

$$\pi(\mathbf{w}) = \frac{f(\mathbf{h}(\mathbf{w}), \mathbf{w}) V(\mathbf{w})}{\int_W f(\mathbf{h}(\mathbf{w}), \mathbf{w}) V(\mathbf{w}) d\mathbf{w}}, \quad (13)$$

where the integral is taken over the domain W of \mathbf{w} .

If the observations of \mathbf{u} and \mathbf{w} are independent, their joint prior pdf factorizes into the marginal pdfs $f_u(\mathbf{u})$ and $f_w(\mathbf{w})$, and the posterior pdf of \mathbf{w} is equal to:

$$\pi(\mathbf{w}) = \frac{f_u(\mathbf{h}(\mathbf{w})) f_w(\mathbf{w}) V(\mathbf{w})}{\int_W f_u(\mathbf{h}(\mathbf{w})) f_w(\mathbf{w}) V(\mathbf{w}) d\mathbf{w}}. \quad (14)$$

2.4. Visualization

In the case of a low-dimensional variable space, the construction of the posterior pdf as outlined in Section 2.3 can be easily visualized, as shown by the following simple example.

Example 2.1: There are two observed variables x_1 and x_2 with the prior pdf $f(x_1, x_2)$ defined on $D \subseteq \mathbb{R}^2$. The nonlinear constraint equation $x_2 = h(x_1)$ defines a one-dimensional manifold S , i.e. a curve in \mathbb{R}^2 . The ‘volume’ element of the curve is given by the differential of the arc length:

$$dS = V(x_1) dx_1 = \sqrt{1 + h'(x_1)^2} dx_1.$$

The posterior pdf $\pi(x_1)$ is therefore equal to:

$$\pi(x_1) = \frac{f(h(x_1), x_1) \sqrt{1 + h'(x_1)^2}}{\int f(h(x_1), x_1) \sqrt{1 + h'(x_1)^2} dx_1}.$$

The values of $f(x_1, x_2)$ along S can be visualized by intersecting the prior pdf surface with the surface that is orthogonal to the x_1, x_2 -plane and contains S .

Figure 1 shows an instance of this problem, with independent observations of the variables x_1, x_2 , the constraint $x_2 = x_1^{2/3}$, and the prior pdfs $f_1(x_1) = \gamma(x_1; 2, 2)$,

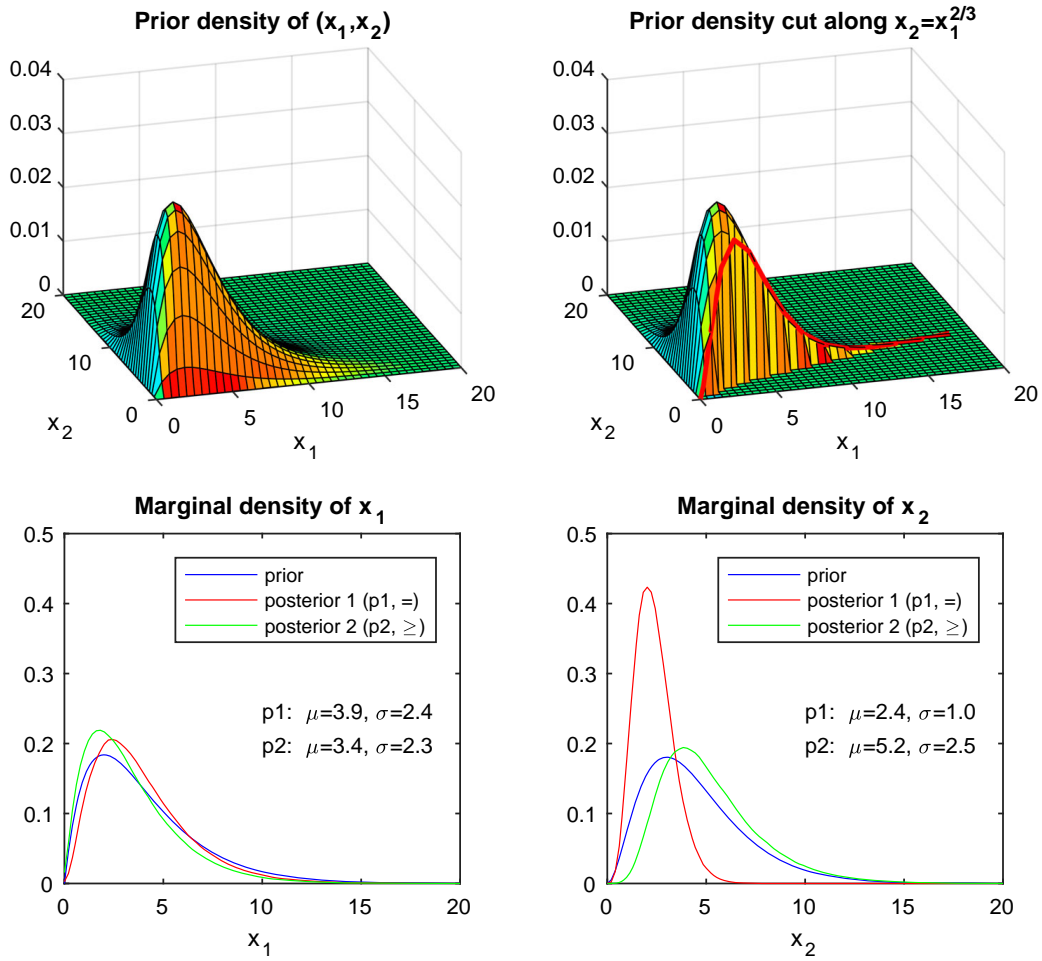


Figure 1. Visualization of the pdfs in Example 2.1. Top left: prior pdf $f(x_1, x_2)$. Top right: prior pdf $f(x_1, x_2)$ cut along the curve $x_2 = x_1^{2/3}$; the red line is the (unnormalized) restricted pdf subject to the constraint $x_2 = x_1^{2/3}$. The remaining 'mountain' can be interpreted as the (unnormalized) restricted pdf subject to the constraint $x_2 \geq x_1^{2/3}$. Bottom: observed (prior, blue) and reconciled (posterior 1, =, red; posterior 2, \geq , green) pdfs of x_1, x_2 .

$f_2(x_2) = \gamma(x_2; 3, 1.5)$, where $\gamma(x; a, b)$ is the pdf of the gamma distribution with parameters a and b :

$$\gamma(x; a, b) = \frac{x^{a-1} e^{-x/b}}{b^a \Gamma(a)} I_{[0, \infty)}(x).$$

The chosen priors induce a nonnegative constraint on the observations.

2.5. Invariance property

The posterior distribution on the constraint manifold S (see Equation (11)) is invariant under the choice of the free variables. In fact, the volume element dS is invariant under general differentiable one-to-one coordinate transformations, and therefore in particular invariant under different choices of the free variables. If \mathbf{w}_1 and \mathbf{w}_2 are two different sets of free variables, the corresponding posteriors $\pi_1(\mathbf{w}_1)$ and $\pi_2(\mathbf{w}_2)$ are related to each other

according to the transformation theorem for pdfs:

$$\pi_1(\mathbf{w}_1) d\mathbf{w}_1 = \pi_2(\mathbf{w}_2) d\mathbf{w}_2. \quad (15)$$

This invariance property is illustrated by the following example.

Example 2.2: Consider a nonlinear model with two independently observed variables x_1, x_2 and a nonlinear constraint $x_2 = h(x_1)$, where x_1 is the free variable and $h(x_1)$ is assumed to be one-to-one, i.e. invertible with the inverse function $x_1 = \eta(x_2)$. The prior pdfs of the observations are given by $f_1(x_1), f_2(x_2)$. According to Equation (13), the posterior pdfs $\pi_1(x_1), \pi_2(x_2)$ are proportional to:

$$\pi_1(x_1) \propto f_1(x_1) f_2(h(x_1)) \sqrt{1 + h'(x_1)^2}, \quad (16)$$

$$\pi_2(x_2) \propto f_1(\eta(x_2)) f_2(x_2) \sqrt{1 + \eta'(x_2)^2}. \quad (17)$$

$\pi_1(x_1)$ can be transformed to a pdf $\dot{\pi}_2(x_2)$ by applying the transformation theorem for pdfs and using the identity:

$$h'(\eta(x_2)) \eta'(x_2) = 1.$$

This yields:

$$\begin{aligned} \dot{\pi}_2(x_2) &= \pi_1(\eta(x_2)) dx_1 / dx_2 \\ &\propto f_1(\eta(x_2)) f_2(x_2) \sqrt{1 + h'(\eta(x_2))^2} \eta'(x_2) \\ &= f_1(\eta(x_2)) f_2(x_2) \sqrt{\eta'(x_2)^2 + h'(\eta(x_2))^2 \eta'(x_2)^2} \\ &= f_1(\eta(x_2)) f_2(x_2) \sqrt{1 + \eta'(x_2)^2} \\ &\propto \pi_2(x_2). \end{aligned}$$

This is equivalent to showing that:

$$V_2(x_2) = V_1(x_1) \left| \frac{dx_1}{dx_2} \right|.$$

Thus, the posterior pdfs in Equations (16) and (17), and therefore the joint distribution of all variables, are invariant under the choice of the free variable.

Remark: In the case of linear constraints, the volume element of the constraint manifold is a constant, and $V(\mathbf{w})$ cancels in the normalization of the posterior pdf (see Equations (13) and (14)). This explains why $V(\mathbf{w})$ does not appear in the posterior pdf $\pi(\mathbf{w})$ derived in [3].

3. MCMC sampling

In general, the integral in the denominator of the right-hand side of Equations (13) and (14) cannot be computed analytically. In Example 2.1, the normalizing constant could be computed numerically by a single integral over the curve $S : x_2 = x_1^{2/3}$. For higher dimensions

of \mathbf{S} , however, numerical integration becomes cumbersome and time consuming. It also has the distinctive disadvantage that only the posterior pdf $\pi(\mathbf{w})$ of the free variables \mathbf{w} is immediately available, and that nothing is known about the distribution of the dependent variables \mathbf{y} and \mathbf{u} .

Generating a random sample from the posterior distribution in Equations (13) and (14) by means of Markov chain Monte Carlo (MCMC) methods [2,8,18] has two advantages:

- (1) There is no need to compute the normalization constant.
- (2) The corresponding sample values of the dependent variables \mathbf{y} and \mathbf{u} can be computed by Equations (4) and (5).

It is therefore straightforward to estimate posterior marginals, expectations, variances and covariances of *all* variables from the full sample.

It was argued in [3] that the sampler best suited to the context of DR is the independence sampler [2,4,7], in which the proposal values $\dot{\mathbf{w}}$ are drawn independently from a proposal pdf $p(\mathbf{w})$. The acceptance probability of the sampler is given by

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{\pi(\dot{\mathbf{w}}) p(\mathbf{w}_i)}{\pi(\mathbf{w}_i) p(\dot{\mathbf{w}})} \right). \quad (18)$$

In the case of independent observations \mathbf{u} and \mathbf{w} , this is equivalent to

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{f_u(\mathbf{h}(\dot{\mathbf{w}})) f_w(\dot{\mathbf{w}}) V(\dot{\mathbf{w}}) p(\mathbf{w}_i)}{f_u(\mathbf{h}(\mathbf{w}_i)) f_w(\mathbf{w}_i) V(\mathbf{w}_i) p(\dot{\mathbf{w}})} \right). \quad (19)$$

Note that the normalizing constant of $\pi(\mathbf{w})$ cancels in Equations (18) and (19), so there is no need to compute it. If the proposal pdf is chosen as $p(\mathbf{w}) = f_w(\mathbf{w})$, Equation (19) reduces to

$$\alpha(\mathbf{w}_i, \dot{\mathbf{w}}) = \min \left(1, \frac{f_u(\mathbf{h}(\dot{\mathbf{w}})) V(\dot{\mathbf{w}})}{f_u(\mathbf{h}(\mathbf{w}_i)) V(\mathbf{w}_i)} \right). \quad (20)$$

The algorithm for independent observations is summarized in the box Algorithm 1. For more details about the implementation of the sampler, see [3, Section 3]. In the general case of correlated observations, the acceptance probability has to be computed according to Equation (18), with a suitable proposal pdf $p(\mathbf{w})$.

As all draws from the proposal pdf are equally valid starting points of the Markov chain, no ‘burn-in’, i.e. discarding an initial segment of the chain, is necessary.³

ALGORITHM 1. ALGORITHM IS-NL

- (IS-NL1). Choose the sample size L , set $i = 1$.
- (IS-NL2). Generate the starting value \mathbf{w}_1 by drawing independent random numbers from the prior marginal distributions of the free variables. Repeat (IS-NL2) until $f_u(\mathbf{h}(\mathbf{w}_1)) > 0$.
- (IS-NL3). Generate a proposal value $\dot{\mathbf{w}}$ by drawing independent random numbers from the prior marginal distributions of the free variables.

- (IS-NL4). Compute the acceptance probability α according to Eq. (20).
 (IS-NL5). Draw a uniform random number $\xi \in [0, 1]$.
 (IS-NL6). If $\xi \leq \alpha$, accept the proposal and set $\mathbf{w}_{i+1} = \dot{\mathbf{w}}$, otherwise set $\mathbf{w}_{i+1} = \mathbf{w}_i$.
 (IS-NL7). Increase i by 1. If $i < L$, go to (IS-NL3), otherwise stop sampling.

Remark: If a large number of variables is involved, the numerator and/or the denominator of Equation (20) may fall below machine precision. This can be circumvented by taking the logarithms of both sides of the equations and accepting when $\log \xi \leq \log \alpha$.

The proposal values $\dot{\mathbf{W}} = (\dot{\mathbf{w}}_1, \dot{\mathbf{w}}_2, \dots, \dot{\mathbf{w}}_L)$ and the corresponding values of $\dot{\mathbf{U}} = (\dot{\mathbf{u}}_1, \dot{\mathbf{u}}_2, \dots, \dot{\mathbf{u}}_L)$ and $\dot{\mathbf{Y}} = (\dot{\mathbf{y}}_1, \dot{\mathbf{y}}_2, \dots, \dot{\mathbf{y}}_L)$ can be precomputed via $\dot{\mathbf{u}}_i = \mathbf{h}(\dot{\mathbf{w}}_i)$ and $\dot{\mathbf{y}}_i = \mathbf{k}(\dot{\mathbf{w}}_i)$, respectively, for $i = 1, \dots, L$ (see Equations (4) and (5)). As the proposal values are independent, their computation can be easily parallelized. After the sampler has finished, the final Markov chain $(\mathbf{Y}; \mathbf{U}; \mathbf{W})$ can then be used to estimate posterior quantities such as means, variances, quantiles and correlations of all variables, both observed and unobserved. For a graphical representation of the posterior distributions, the marginal pdfs can be estimated from the Markov chain by a kernel estimator [20] or obtained by smoothing the normalized frequency distribution in a histogram. For more details about the acceptance rate and the posterior analysis, see [3].

By construction, all individual elements $\mathbf{m}_i = (\mathbf{y}_i; \mathbf{u}_i; \mathbf{w}_i)$ of the Markov chain satisfy the nonlinear constraints, but the sample mean in general does not. If a representative value of the posterior distribution satisfying the constraints is required, the element of the Markov chain with the smallest distance from the sample mean, the sample median or the sample mode can be selected. The multivariate sample mode is, however, notoriously difficult to compute, and there are different ways to define a multivariate median [12]. Thus, the element closest to the sample mean turns out to be virtually the only practical possibility.

4. Applications

The two examples in this section show the application of the independence sampler to MFA and chemical engineering. The MATLAB code used for the examples can be obtained from the authors on request.

In both examples we assume that there is independent prior information for each of the observed variables. The prior pdf displays the initial knowledge about the variable. This knowledge can be objective, being the model of a measurement process, or it can be subjective, being the expression of an expert opinion via a pdf. The mode of the prior pdf represents the most probable or most credible value of the variable before imposing the constraints. Similarly, the mode of the posterior pdf represents the most probable or credible value after imposing the constraints. Note that even if the observations are assumed to be independent *a priori*, all variables are correlated *a posteriori*. A comparison of the prior and the posterior marginal pdfs shows that the uncertainty of the observed variables is reduced: standard deviations become smaller, HPD intervals become shorter. Both examples additionally contain unobserved variables. Although they have no prior pdfs, their posterior marginal pdfs are estimated from the Markov chain of the free variables.

This demonstrates the two essential features of data reconciliation: improving the precision of the observed variables, and enabling the estimation of unobserved variables.

Example 4.1: This example is based on a simple problem in MFA. Figure 2 shows a graphical representation of material flows. Process 1 splits the input flow (1) into two output flows (2 and 3). Each flow i is characterized by three variables: the mass g_i of the transported good, the mass s_i of a specific substance contained in the good, and the mass fraction $c_i = s_i/g_i$ of the substance in the good. A typical example of this model could be the splitting of plastic waste into a heavy and a light fraction, where the cadmium (= substance) mass flow and its mass fraction in each of the goods is of interest.

Since the masses of both the goods and the substance are conserved in the process, and the mass of the substance equals the mass of the good times its mass fraction in each flow, the reconciled values of all variables have to fulfil the following minimal set of constraints:

$$g_1 = g_2 + g_3, \quad s_1 = s_2 + s_3, \quad s_1 = g_1 c_1, \quad s_2 = g_2 c_2, \quad s_3 = g_3 c_3.$$

There are six observed variables $\mathbf{x} = (g_1; g_2, g_3; c_1; c_2; c_3)$, and three unobserved variables $\mathbf{y} = (s_1; s_2; s_3)$. As the variables g_1, g_2, g_3 cannot take negative values, their uncertainty is modelled by independent lognormal priors. The variables c_1, c_2, c_3 are restricted to the interval $[0, 1]$, and their uncertainty is modelled by independent beta priors. The modes m_i and standard deviations σ_i of the priors as well as the mean values are given in Table 1.

The choice of nonnegative priors guarantees that also the unobserved mass flows s_i of the substance will automatically take nonnegative values only. Thus, it is not necessary to enforce their positivity.

In this example, the functional dependencies $\mathbf{y} = \mathbf{k}(\mathbf{w})$ and $\mathbf{u} = \mathbf{h}(\mathbf{w})$ (see Equations (4) and (5)) can be stated explicitly. The number of free variables to be selected from the observed variables \mathbf{x} is equal to $n - n_c = 9 - 5 = 4$. If for instance $\mathbf{w} = (g_2; g_3; c_2; c_3)$, $\mathbf{y} = (s_1; s_2; s_3)$ and $\mathbf{u} = (g_1; c_1)$, then:

$$\mathbf{y} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} g_2 c_2 + g_3 c_3 \\ g_2 c_2 \\ g_3 c_3 \end{pmatrix},$$

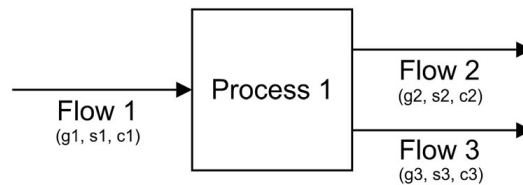


Figure 2. Flowsheet of Example 4.1 with three flows and one process.

Table 1. Modes, standard deviations and means of the prior distributions of the observed variables in Example 4.1.

	g_1	g_2	g_3	c_1	c_2	c_3
m_i	15.00	8.00	5.00	0.30	0.20	0.50
σ_i	5.00	3.00	2.00	0.03	0.02	0.05
μ_i	16.99	9.28	5.89	0.30	0.20	0.50

$$\mathbf{u} = \begin{pmatrix} g_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} g_2 + g_3 \\ (g_2 c_2 + g_3 c_3)/(g_2 + g_3) \end{pmatrix}.$$

The independence sampler (see box Algorithm 1) was run for $L = 200,000$ iterations. The acceptance rate of the sampler is about 62%. Figure 3 shows the smoothed marginal posterior pdfs of $\mathbf{q} = (g_1; g_2; g_3; c_1; c_2; c_3; s_1; s_2; s_3)$ with their modes, means and standard deviations. The gain in precision due to DR is most clearly seen in the posterior standard deviations of the mass flows g_i of the goods, which are much smaller than the prior ones. In addition, there is valuable information about the unobserved mass flows of the substance, which can be obtained only by DR.

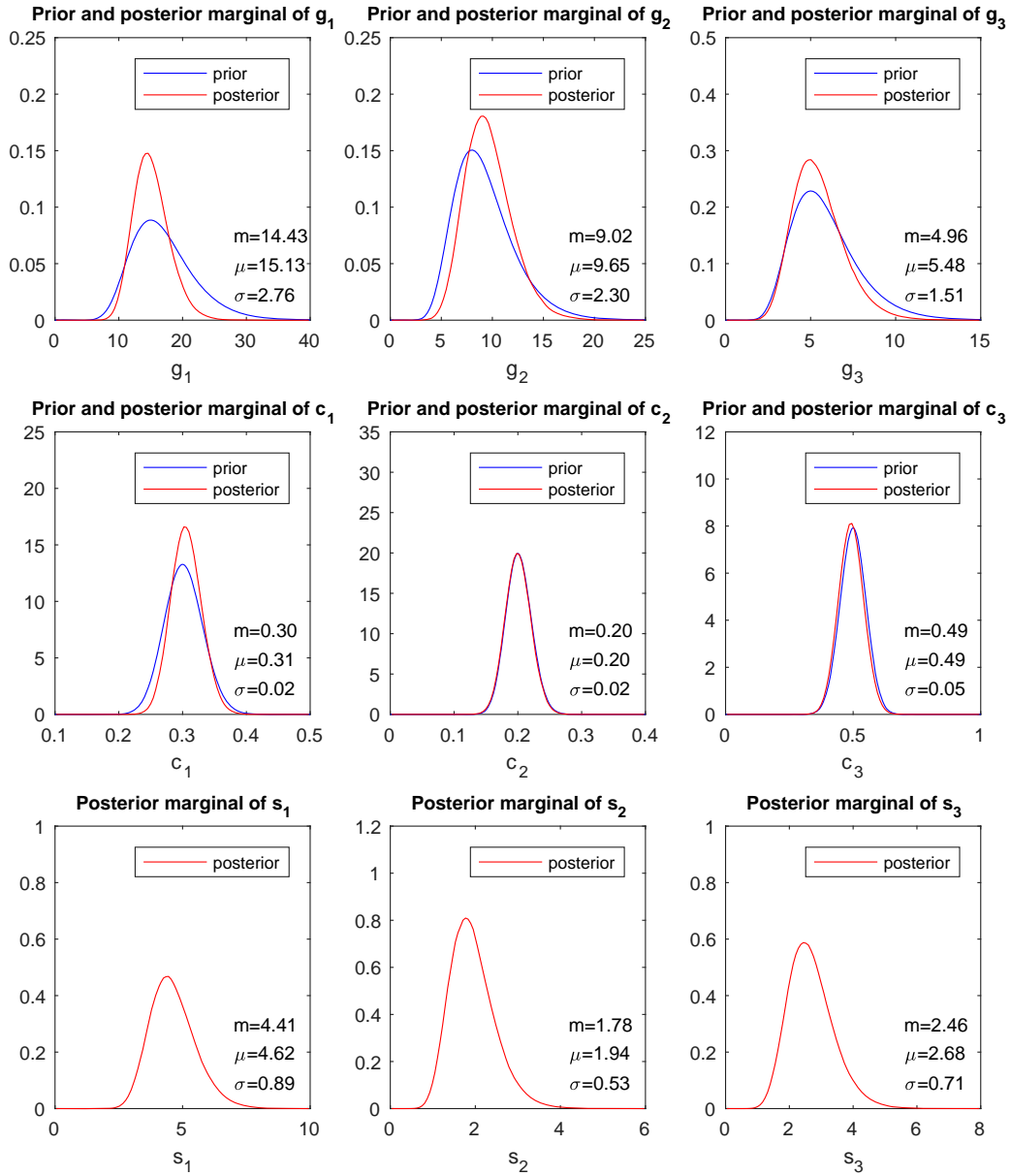


Figure 3. Priors and smoothed marginal posteriors of $g_1, g_2, g_3, c_1, c_2, c_3, s_1, s_2, s_3$ in Example 4.1, with mode m , mean μ and standard deviation σ .

Table 2. Sample mean \bar{q} and sample value q_{\min} that is closest to the sample mean as measured by the L_2 distance.

Variable	g_1	g_2	g_3	c_1	c_2	c_3	s_1	s_2	s_3
\bar{q}	15.13	9.65	5.48	0.31	0.20	0.49	4.62	1.94	2.68
q_{\min}	15.12	9.65	5.46	0.31	0.20	0.50	4.67	1.94	2.73

The sample mean \bar{q} and the sample value q_{\min} that is closest to the sample mean, as measured by the L_2 distance, are given in Table 2.

Example 4.2: The second example uses the system of nonlinear constraints analysed in [14], which is a commonly used benchmark example⁴ for nonlinear data reconciliation with normally distributed observation errors:

$$G_1(\mathbf{y}; \mathbf{x}) = 0.5x_1^2 - 0.7x_2 + x_3y_1 + x_2^2y_1y_2 + 2x_3y_3^2 - 255.8 = 0,$$

$$G_2(\mathbf{y}; \mathbf{x}) = x_1 - 2x_2 + 3x_1x_3 - 2x_2y_1 - x_2y_2y_3 + 111.2 = 0,$$

$$G_3(\mathbf{y}; \mathbf{x}) = x_3y_1 - x_1 + 3x_2 + x_1y_2 - x_3y_3^{0.5} - 33.57 = 0,$$

$$G_4(\mathbf{y}; \mathbf{x}) = x_4 - x_1 - x_3^2 + y_2 + 3y_3 = 0,$$

$$G_5(\mathbf{y}; \mathbf{x}) = x_5 - 2x_3y_2y_3 = 0,$$

$$G_6(\mathbf{y}; \mathbf{x}) = 2x_1 + x_2x_3y_1 + y_2 - y_3 - 126.6 = 0.$$

The variables $\mathbf{x} = (x_1; x_2; x_3; x_4; x_5)$ are observed with the following values:

$$\tilde{\mathbf{x}} = (4.4; 5.5; 1.7; 1.6; 5)$$

and the covariance matrix $\Sigma = \mathbf{I}$. Although the context of the model is not stated in [14], we assume that all variables have to be nonnegative. The prior of the observation x_i is assumed to be the lognormal pdf with mean \tilde{x}_i and standard deviation $\sigma_i = 1$, for $i = 1, \dots, 5$. Assuming the observations to be independent, the joint prior is the product of the individual priors. The unobserved variables $\mathbf{y} = (y_1; y_2; y_3)$ can be forced to be nonnegative by assigning them an improper constant prior on $[0, \infty)$; however, this turns out to be unnecessary.

If the two free variables are chosen as $\mathbf{w} = (x_1; x_2)$, the three dependent observed variables are equal to $\mathbf{u} = (x_3; x_4; x_5)$. For given \mathbf{w} , the functions $\mathbf{u} = \mathbf{h}(\mathbf{w})$ and $\mathbf{y} = \mathbf{k}(\mathbf{w})$ (see Equations (4) and (5)) are computed by minimizing $\|\mathbf{G}(\mathbf{y}; \mathbf{u}; \mathbf{w})\|_2$, using the simplex algorithm [11] with restart. The starting point of the minimization is equal to $(\tilde{\mathbf{y}}; \tilde{\mathbf{x}})$, where

$$\tilde{\mathbf{y}} = \arg \min_{\mathbf{y}} \|\mathbf{G}(\mathbf{y}; \tilde{\mathbf{x}})\|_2.$$

The Jacobian $\mathbf{H}(\mathbf{w}) = \partial \mathbf{h}(\mathbf{w}) / \partial \mathbf{w}$ is computed by numerical differentiation.

The independence sampler was run for $L = 200,000$ iterations, with an acceptance rate of about 13%. The posterior pdfs are shown in Figure 4. Clear deviations from normality can be observed. The reconciled values of \mathbf{x} and \mathbf{y} given in [14] are very close to the posterior modes. The posterior standard deviations of the observed values are significantly smaller than 1, showing again the gain in precision due to data reconciliation.

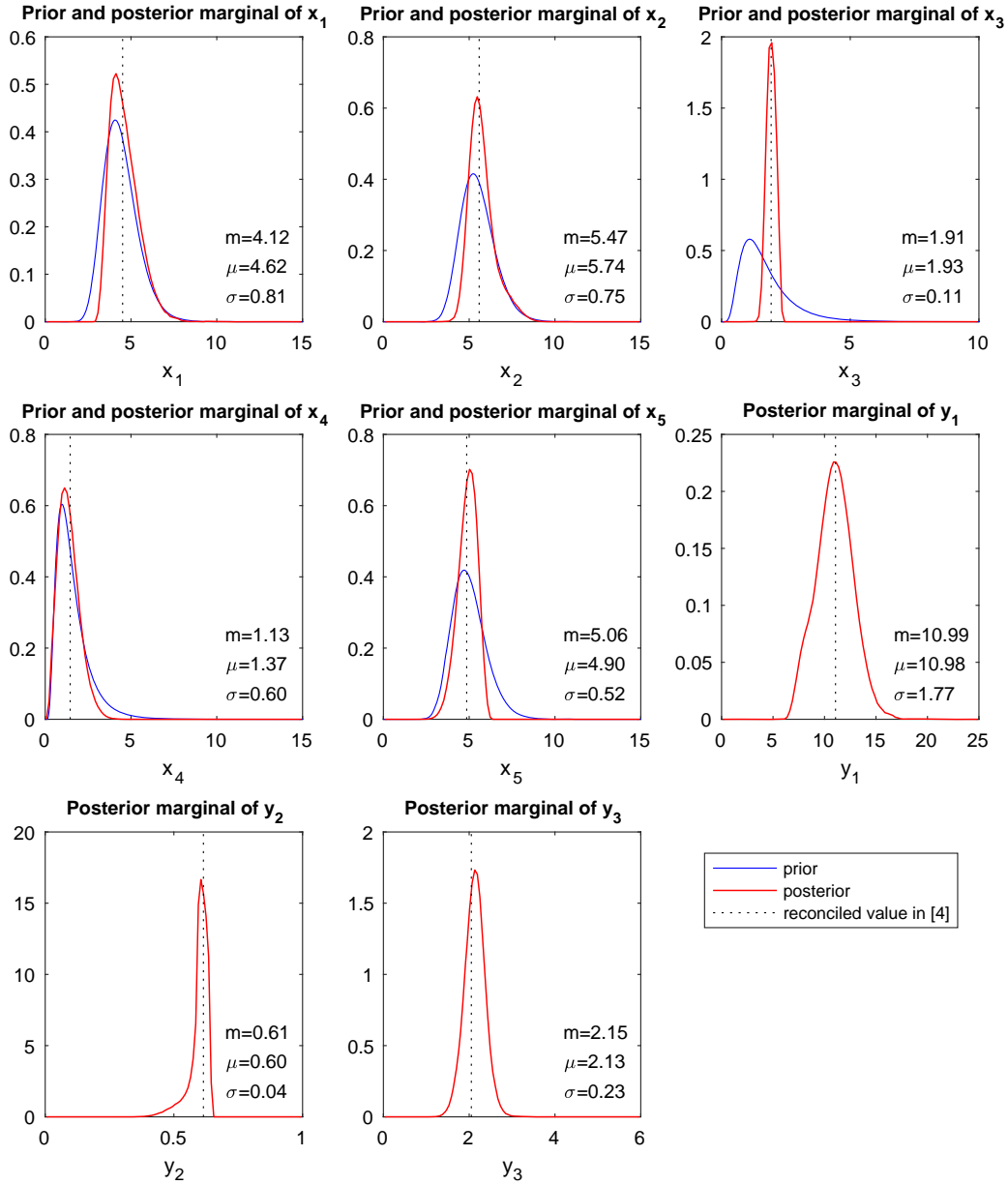


Figure 4. Priors and smoothed marginal posteriors of $x_1, x_2, x_3, x_4, x_5, y_1, y_2, y_3$ in Example 4.2, with mode m , mean μ and standard deviation σ . The reconciled values given in [14] are also shown.

5. Conclusions and outlook

In this paper, the method for reconciling nonnormally distributed data by a Markov chain Monte Carlo method presented in [3] was extended to nonlinear constraints for uncorrelated and correlated observations. A crucial difference to the posterior $\pi(\mathbf{w})$ derived for the linear case (Equation (A10)) is the presence of $V(\mathbf{w})$ (Equation (8)) in the corresponding posterior for the nonlinear case (Equation (13)). If $V(\mathbf{w})$ is neglected, the posterior is no longer invariant under the choice of the free variables.

The method presented here has several advantages. First, arbitrary continuous pdfs can be used to describe the uncertainty of the observations. Second, even nonparametric

estimators of the pdf are allowed, provided that it is possible to draw a random sample from them. Third, not only means, variances and covariances of observed and unobserved variables can be computed *a posteriori*, but also various other characteristics of the marginal posteriors, such as the mode, skewness, quantiles, and HPD intervals.

In some systems of nonlinear constraints not all choices of free variables lead to feasible solutions for the dependent variables. If this is the case, we propose to permute the positions of the free variables randomly and repeat the calculations until a feasible solution is reached. In large and complex models this may lead to an unacceptable combinatorial overhead. A general procedure for finding a suitable set of free variables therefore remains an important open question for further research.

The procedure developed in this paper for DR with nonlinear constraints can also be applied to linear constraints. Thus, it is not necessary to assume improper priors for unobserved variables, as was proposed in [3].

Beyond their statistical uncertainty, the observations may also be corrupted by gross errors such as biased observations or faulty readings. Work is already in progress to develop a robustification of the DR procedure by automatic identification of outlying or erroneous observations that prevent a successful reconciliation.

In subsequent work, the method will be applied to a more extensive real-world example in order to compare the results to alternative approaches such as classical weighted least squares and fuzzy sets.

Notes

1. The comma (semicolon) denotes horizontal (vertical) concatenation of vectors and matrices (MATLAB convention).
2. See, however, the discussion of constraints versus priors in [21].
3. This has been argued by C. Geyer on his website, see <http://users.stat.umn.edu/geyer/mcmc/burn.html>. Last accessed on January 2, 2018.
4. See 'Citing literature' on <http://onlinelibrary.wiley.com/doi/10.1002/aic.690340521/abstract>. Last accessed on January 2, 2018.

Acknowledgments

We thank Nada Dzibur, Manuel Hahn, David Laner and Oliver Schwab for their useful comments on the manuscript. We also thank the editor and the anonymous reviewers for their comments, suggestions and corrections, which have greatly improved the manuscript. The authors acknowledge the TU Wien University Library for financial support through its Open Access Funding Program.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] V.B. Barbosa Jr, M.R.M. Wolf, and R.M. Fo, *Development of data reconciliation for dynamic nonlinear system: Application the polymerization reactor*, Comput. Chem. Eng. 24 (2000), pp. 501–506.
- [2] S. Brooks, A. Gelman, G.L. Jones, and X.-L. Meng (eds.), *Handbook of Markov Chain Monte Carlo*, Chapman & Hall, Boca Raton, FL, 2011.

- [3] O. Cencic and R. Frühwirth, *A general framework for data reconciliation – Part I: Linear constraints*, Comput. Chem. Eng. 75 (2015), pp. 196–208.
- [4] S. Chib and E. Greenberg, *Understanding the metropolis–hastings algorithm*, Am. Stat. 49 (1995), pp. 327–335.
- [5] L.R. Foulds, *Optimization Techniques: An Introduction*, Springer, New York, 1981.
- [6] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*, Academic Press, London, 2000.
- [7] J.S. Liu, *Metropolized independent sampling with comparisons to rejection sampling and importance sampling*, Stat. Comput. 6 (1996), pp. 113–119.
- [8] J.S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2004.
- [9] F. Madron, *Process Plant Performance*, Ellis Horwood, New York, 1992.
- [10] F. Madron, V. Veverka, and V. Venecek, *Statistical analysis of material balance of a chemical reactor*, AIChE J. 23 (1977), pp. 482–486.
- [11] J.A. Nelder and R. Mead, *A simplex method for function minimization*, Comput. J. 7 (1965), pp. 308–313.
- [12] H. Oja, *Multivariate median*, in *Robustness and Complex Data Structures*, C. Becker, R. Fried, and S. Kuhnt, eds., Springer, Berlin, 2013.
- [13] B. O’Neill, *Semi-Riemannian Geometry*, Academic Press, San Diego, CA, 1983.
- [14] C.C.D. Pai and G.D. Fisher, *Application of Broyden’s method to reconciliation of nonlinearly constrained data*, AIChE J. 34 (1988), pp. 873–876.
- [15] X. Pennec, *Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements*, In Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP ’99), Vol. 1. Bogaziçi University, Istanbul, Turkey, 1999, pp. 194–198. Available at <http://www-sop.inria.fr/asclepios/Publications/Xavier.Pennec/Pennec.NSIP99.pdf> (last accessed on January 2, 2018).
- [16] X. Pennec, *Probabilities and statistics on Riemannian manifolds: A geometric approach*, Tech. Rep. RR-5093, INRIA, January 2004. Available at <https://hal.inria.fr/inria-00071490/document> (last accessed on January 2, 2018).
- [17] W.-D. Richter, L. Strelec, H. Ahmadinezhad, and M. Stehlik, *Geometric aspects of robust testing for normality and sphericity*, Stoch. Anal. Appl. 35 (2017), pp. 511–532.
- [18] C. Robert and C. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.
- [19] J.A. Romagnoli and M.C. Sanchez, *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press, San Diego, 2000.
- [20] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, London, 1998.
- [21] P.B. Stark, *Constraints versus Priors*, SIAM/ASA J Uncertain Quantif. 3 (2015), pp. 586–598.
- [22] P. Whittle, *Optimization under Constraints: Theory and Applications of Nonlinear Programming*, Wiley-Interscience, London, 1971.
- [23] W. Wongrat, T. Srinophakun, and P. Srinophakun, *Modified genetic algorithm for nonlinear data reconciliation*, Comput. Chem. Eng. 29 (2005), pp. 1059–1067.

Appendix. Proof of equivalence with linear constraints

For the case of normally distributed independent observations and linear constraints, it was proved in [3] that the unbiased estimate of the expectation μ_π of $\pi(\mathbf{w})$ is equal to the least-squares solution $\hat{\mathbf{w}}$. Here, the proof is generalized to correlated observations.

Consider a system of n_c linear constraint equations:

$$\mathbf{G}(\mathbf{y}; \mathbf{x}) = \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} + \mathbf{c} = \mathbf{0}, \quad (\text{A1})$$

where \mathbf{x} is the vector of observed variables of dimension $n_x \times 1$, \mathbf{y} is the vector of unobserved variables of dimension $n_y \times 1$, \mathbf{c} is a constant vector of dimension $n_c \times 1$, and \mathbf{A} and \mathbf{B} are known matrices of dimension $n_c \times n_x$ and $n_c \times n_y$, respectively. The system is assumed to have the following properties:

(A1) $\text{rank}(\mathbf{B}, \mathbf{A}) = \text{rank}(\mathbf{B}, \mathbf{A}, \mathbf{c})$, meaning the system has at least one solution;

- (A2) $\text{rank}(\mathbf{B}, \mathbf{A}) = n_c$, meaning the system equations are linearly independent;
 (A3) $\text{rank}(\mathbf{B}) = n_y$, meaning all unobserved quantities can be calculated.

If any of these assumptions are violated, the underlying problems have to be resolved before proceeding. This can be achieved, e.g. by transforming the matrix $(\mathbf{B}, \mathbf{A}, \mathbf{c})$ to the reduced row echelon form (RREF). This allows to detect contradictions (A1), to eliminate dependent equations automatically (A2), and to classify variables, in particular to identify and eliminate unobservable unobserved variables (A3). For detailed instructions how to proceed see, for instance [9, p. 125].

After the reduction to the RREF, the observed variables \mathbf{x} can be split into $n_w = n - n_c$ free variables \mathbf{w} and $n_u = n_c - n_y$ dependent variables \mathbf{u} , which are linear functions of \mathbf{w} :

$$\mathbf{G}(\mathbf{y}; \mathbf{x}) = (\mathbf{B} \quad \mathbf{A} \quad \mathbf{c}) \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{O} & \mathbf{E} & \mathbf{e} \\ \mathbf{O} & \mathbf{I} & \mathbf{D} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{w} \\ 1 \end{pmatrix} = \mathbf{0} \quad (\text{A2})$$

Remark: To reach the RREF displayed in Equation (A2), a reordering of variables in \mathbf{x} may be necessary, leading to $\mathbf{v} = (\mathbf{u}; \mathbf{w})$.

Due to the reduction to the RREF, the n_y unobserved variables \mathbf{y} , which are linear functions of \mathbf{w} only, can be eliminated from the DR problem. Thus, the system of constraint equations can be simplified to

$$\mathbf{u} + \mathbf{D}\mathbf{w} + \mathbf{d} = \mathbf{0}, \quad (\text{A3})$$

which the reconciled values $\hat{\mathbf{u}}$ and $\hat{\mathbf{w}}$ have to fulfil after DR:

$$\hat{\mathbf{u}} + \mathbf{D}\hat{\mathbf{w}} + \mathbf{d} = \mathbf{0} \quad \text{or} \quad \hat{\mathbf{u}} = -\mathbf{D}\hat{\mathbf{w}} - \mathbf{d}. \quad (\text{A4})$$

The reconciled values $\hat{\mathbf{w}}$ can be used to compute $\hat{\mathbf{y}}$ via:

$$\hat{\mathbf{y}} + \mathbf{E}\hat{\mathbf{w}} + \mathbf{e} = \mathbf{0} \quad \text{or} \quad \hat{\mathbf{y}} = -\mathbf{E}\hat{\mathbf{w}} - \mathbf{e}. \quad (\text{A5})$$

The least-squares solution of the linear DR problem can be found by the standard method of Lagrange multipliers [19]:

$$\hat{\mathbf{v}} = \tilde{\mathbf{v}} + \Sigma \mathbf{K}^T (\mathbf{K} \Sigma \mathbf{K}^T)^{-1} (-\mathbf{d} - \mathbf{K} \tilde{\mathbf{v}}), \quad (\text{A6})$$

where $\tilde{\mathbf{v}} = (\tilde{\mathbf{u}}; \tilde{\mathbf{w}})$ is the vector of observations, Σ is the joint covariance matrix of $\tilde{\mathbf{v}}$, and $\mathbf{K} = (\mathbf{I}, \mathbf{D})$. The solution $\hat{\mathbf{v}}$ can be written as $\hat{\mathbf{v}} = (\hat{\mathbf{u}}; \hat{\mathbf{w}})$, with

$$\hat{\mathbf{w}} = \tilde{\mathbf{w}} + (\Sigma_{wu} + \Sigma_{ww} \mathbf{D}^T) \mathbf{H}^{-1} (-\mathbf{d} - \tilde{\mathbf{u}} - \mathbf{D} \tilde{\mathbf{w}}), \quad (\text{A7})$$

where

$$\mathbf{H} = \mathbf{K} \Sigma \mathbf{K}^T = \Sigma_{uu} + \mathbf{D} \Sigma_{wu} + \Sigma_{uw} \mathbf{D}^T + \mathbf{D} \Sigma_{ww} \mathbf{D}^T, \quad (\text{A8})$$

and Σ is partitioned in the following way:

$$\Sigma = (\Sigma_{uu}, \Sigma_{uw}; \Sigma_{wu}, \Sigma_{ww}). \quad (\text{A9})$$

In [3] it has been shown that with linear constraints the reconciled or posterior pdf $\pi(\mathbf{w})$ of the free variables can also be obtained by restricting the joint pdf of the observations to the constraint manifold. Assume that the joint pdf $f(\mathbf{u}; \mathbf{w})$ of the observations is a normal pdf with mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_u; \boldsymbol{\mu}_w)$ and covariance matrix Σ . Then the posterior $\pi(\mathbf{w})$, given by

$$\pi(\mathbf{w}) = \frac{f(-\mathbf{D}\mathbf{w} - \mathbf{d}; \mathbf{w})}{\int f(-\mathbf{D}\mathbf{w} - \mathbf{d}; \mathbf{w}) d\mathbf{w}}, \quad (\text{A10})$$

is a normal pdf too. If the inverse covariance matrix $\Gamma = \Sigma^{-1}$ is partitioned as follows:

$$\Gamma = (\Gamma_{uu}, \Gamma_{uw}; \Gamma_{wu}, \Gamma_{ww}), \quad (\text{A11})$$

the quadratic form $Q(\mathbf{w})$ in the exponent of $\pi(\mathbf{w})$:

$$Q(\mathbf{w}) = [(-D\mathbf{w} - \mathbf{d}; \mathbf{w}) - (\boldsymbol{\mu}_u; \boldsymbol{\mu}_w)]^T \boldsymbol{\Gamma} [(-D\mathbf{w} - \mathbf{d}; \mathbf{w}) - (\boldsymbol{\mu}_u; \boldsymbol{\mu}_w)], \quad (\text{A12})$$

can be written in the following form:

$$\begin{aligned} Q(\mathbf{w}) &= \mathbf{w}^T (D^T \boldsymbol{\Gamma}_{uu} D - D^T \boldsymbol{\Gamma}_{uw} - \boldsymbol{\Gamma}_{wu} D + \boldsymbol{\Gamma}_{ww}) \mathbf{w} \\ &\quad - 2\mathbf{w}^T [(\boldsymbol{\Gamma}_{ww} - D^T \boldsymbol{\Gamma}_{uw}) \boldsymbol{\mu}_w + (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu})(\mathbf{d} + \boldsymbol{\mu}_u)] + \text{const.} \end{aligned} \quad (\text{A13})$$

$\pi(\mathbf{w})$ is therefore the pdf of the normal distribution with inverse covariance matrix \mathbf{C} and mean $\boldsymbol{\mu}_\pi$ given by

$$\mathbf{C} = D^T \boldsymbol{\Gamma}_{uu} D - D^T \boldsymbol{\Gamma}_{uw} - \boldsymbol{\Gamma}_{wu} D + \boldsymbol{\Gamma}_{ww}, \quad (\text{A14})$$

$$\boldsymbol{\mu}_\pi = \mathbf{C}^{-1} [(\boldsymbol{\Gamma}_{ww} - D^T \boldsymbol{\Gamma}_{uw}) \boldsymbol{\mu}_w + (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu})(\mathbf{d} + \boldsymbol{\mu}_u)]. \quad (\text{A15})$$

$\boldsymbol{\mu}_\pi$ can be expanded and rearranged in the following form:

$$\boldsymbol{\mu}_\pi = \boldsymbol{\mu}_w + \mathbf{C}^{-1} (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu})(\mathbf{d} + \boldsymbol{\mu}_u + D\boldsymbol{\mu}_w). \quad (\text{A16})$$

Its unbiased estimator $\hat{\boldsymbol{\mu}}_\pi$ is obtained by replacing $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_w$ by the observations $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{w}}$, respectively:

$$\hat{\boldsymbol{\mu}}_\pi = \tilde{\mathbf{w}} + \mathbf{C}^{-1} (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu})(\mathbf{d} + \tilde{\mathbf{u}} + D\tilde{\mathbf{w}}). \quad (\text{A17})$$

A comparison with $\hat{\mathbf{w}}$ in Equation (A7) shows that

$$\begin{aligned} \hat{\boldsymbol{\mu}}_\pi = \hat{\mathbf{w}} &\iff \mathbf{C}^{-1} (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu}) = -(\boldsymbol{\Sigma}_{wu} + \boldsymbol{\Sigma}_{ww} D^T) \mathbf{H}^{-1} \\ &\iff (\boldsymbol{\Gamma}_{wu} - D^T \boldsymbol{\Gamma}_{uu}) \mathbf{H} + \mathbf{C} (\boldsymbol{\Sigma}_{wu} + \boldsymbol{\Sigma}_{ww} D^T) = \mathbf{O}. \end{aligned} \quad (\text{A18})$$

Expansion of the products while taking into account the symmetry of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ along with the relations:

$$\begin{aligned} \boldsymbol{\Sigma}_{uu} \boldsymbol{\Gamma}_{uu} + \boldsymbol{\Sigma}_{uw} \boldsymbol{\Gamma}_{wu} &= \mathbf{I}, \\ \boldsymbol{\Sigma}_{uu} \boldsymbol{\Gamma}_{uw} + \boldsymbol{\Sigma}_{uw} \boldsymbol{\Gamma}_{ww} &= \mathbf{O}, \\ \boldsymbol{\Sigma}_{wu} \boldsymbol{\Gamma}_{uu} + \boldsymbol{\Sigma}_{ww} \boldsymbol{\Gamma}_{wu} &= \mathbf{O}, \\ \boldsymbol{\Sigma}_{wu} \boldsymbol{\Gamma}_{uw} + \boldsymbol{\Sigma}_{ww} \boldsymbol{\Gamma}_{ww} &= \mathbf{I} \end{aligned} \quad (\text{A19})$$

shows that indeed $\hat{\boldsymbol{\mu}}_\pi = \hat{\mathbf{w}}$.