ECHTZEITOPTIMIERUNG GROSSER SYSTEME

T. BINDER, L. BLANK, W. DAHMEN, W. MARQUARDT

Towards Multiscale Dynamic Data Reconciliation

Preprint 98-7

SCHWERPUNKTPROGRAMM DER DEUTSCHEN FORSCHUNGSGEMEINSCHAFT

Towards Multiscale Dynamic Data Reconciliation

T. Binder[1] , L. Blank[2] , W. Dahmen[2] , W. Marquardt[1*]

[1]Lehrstuhl für Prozeßtechnik,

[2]Institut für Geometrie und Praktische Mathematik

RWTH Aachen University of Technology,

D-52056 Aachen, Germany

Abstract

Although reconciliation of steady-state process data is routinely applied in industrial practice, the theoretical understanding of the problem and its adequate formulation in a dynamic setting is still not mature. Existing formulation approaches are based on stochastic filters, deterministic observers or mathematical programming techniques. In this contribution, we suggest a general problem formulation of dynamic data reconciliation based on the theory of ill-posed problems and their regularizations. It results in a large-scale dynamic optimization problem which requires efficient numerical solution methods in real-time under strict limitations of computational resources. We explore a novel mathematical framework for the discretization of the dynamic optimization problem and the solution of the discretized nonlinear programming problem based on multiscale approximation. The framework attempts to integrate signal processing and optimization finally leading to a fully adaptive and highly efficient numerical treatment which always provides the best possible estimate which is attainable in the allotted time period.

Keywords: dynamic data reconciliation, inverse problems, wavelets, receding horizon estimation, adaptive refinement

1 Introduction

Process measurements never reveal adequate information about the process state. This is due to the always limited number of process quantities accessible by measurement instruments as well as to unavoidable measurement errors caused by imperfect instrumentation and signal processing. Hence, there are quantitative as well as a qualitative deficiencies which cannot be removed in an industrial environment by increased instrumentation or improved maintenance procedures due to cost and technical constraints. However, imperfect and incomplete measurement information will lead to inadequately performing or even unreliable processes. Therefore, there is a significant incentive to upgrade the information content in process measurements, since any increased understanding of the process state will enable improvement of process operations, monitoring and control. The upgraded measurement information can be directly used by operators to better manually control the process, it can be fed to some operator decision support system or to some automatic control system to lead to safer process operations of increased performance.

The measurement information can be upgraded, in principle, by means of process knowledge incorporated in some mathematical model which ideally predicts process

*Correspondence author. Phone: +49-241-806712, Fax: +49-241-8888326, E-mail: marquardt@lfpt.rwth-aachen.de

behavior in some range of operating conditions. Such a model may either result from first principles or from a large set of previously recorded and reasonably correlated process data. Obviously, the model will never perfectly predict process behavior due to unmodeled dynamics, uncertain (and potentially time-varying) parameters, or unaccounted external disturbances. Model and measurements are incorporated in some algorithm to render an estimate of all measured process quantities as well as of all those unmeasured process quantities which are additionally required to fully describe the process behavior. The objective of these algorithms is therefore to remove any error from available measurements and to yield optimal estimates of the full process state as well as of unknown process parameters.

The problem stated has been treated extensively in various branches of mathematical and engineering science focusing on different problem formulations and objectives. The first notion of the problem in the context of chemical process operations has been given in 1961 by Kuehn and Davidson [26]. These authors formulated the problem of adjusting flow and temperature measurements on a crude oil distillation tower to satisfy steady-state material and energy balances. The problem of reconciling process measurements in order to obey steady-state material and energy balances as well as any other constraint has been subsequently called *data reconciliation*. The various problem formulations studied in recent years as well as the solution algorithms taking advantage of a specific problem structure are comprehensively summarized in a recent survey by Crowe [13]. Data reconciliation is nowadays routinely applied in industrial practice (see [70] for an industrial case study) relying on general-purpose commercial or in-house software packages.

A common assumption underlying all the established data reconciliation methods and software implementations is a stationary process. Therefore, the applicability of these methods is restricted to continuously operated processes during periods when the process is reasonably close to some steady-state. Since this assumption turns out to be very restrictive in many cases, the treatment of more general transient situations is an important extension of established technology as recently affirmed in an industrial assessment of the status of real-time process optimization [42].

So far, the reconciliation of dynamic process data by means of a dynamic process model has only received limited attention. Gertler and Almasy [34] seem to have been the first addressing the distribution of measurement error among process variables to satisfy dynamic mass and energy balances in a chemical engineering context. Explicit reference to *dynamic data reconciliation* (or the like) as an extension of the steady-state problem has been made only by a few researchers including Almasy [32, 34], Bagajewicz [5], Bequette [54], Biegler [2, 3], Darouach [20], Edgar [40, 41, 44, 48], as well as Himmelblau [35, 39] and coworkers. Due to the close relationship between dynamic data reconciliation and simultaneous parameter and state estimation in dynamic systems, there is a large body of related work in the chemical as well as, in particular, in the control engineering literature. One may distinguish deterministicly motivated observers (e.g. [52] for a review), stochastic filtering (e.g. [33, 37], or [52] for a review) and mathematical programming based estimators (e.g. [2, 12, 36, 55], and [52] for a review). Still, the theoretical understanding of dynamic data reconciliation, adequate formulations of the estimation problem, and the development of solution algorithms for large-scale industrial problems in process operations are still in their infancy.

This contribution does not attempt to give a comprehensive survey on the state of the art of dynamic data reconciliation. Such a survey would be quite difficult to focus due to a largely missing or at least not generally accepted problem statement on the one hand and due to the large body of literature on the closely related state and parameter estimation of dynamic systems on the other hand. The latter area has been summarized recently in an excellent survey paper by Muske and

Edgar [52]. Instead, we suggest first a general problem formulation of dynamic data reconciliation in Section 2 which should reflect the requirements of realistic (industrial) scenarios not being biased by available solution approaches and their underlying assumptions. Relevant research issues will be pointed out. Section 3 suggests a novel mathematical approach to dynamic data reconciliation based on multiscale approximation. Some details on the numerical solution techniques in the multiscale framework will be given in Section 4. An illustrating example is provided in Section 5 to demonstrate the potential as well as current limitations of the novel approach.

2 Problem formulation

This section attempts to provide a concise problem statement of dynamic data reconciliation. It is based on previous work (e.g. see Section 1) but does not introduce restricting assumptions potentially needed for the derivation of a mathematical framework for problem solution. This general problem formulation should serve as a reference for future work where several simplifying assumptions need to be introduced to either permit theoretical analysis or to derive efficient solution algorithms taking advantage of a particular structure apparent in a restricted problem class.

Dynamic data reconciliation is *defined* as a procedure to optimally rectify the measurements taken from a process under transient conditions as well as to optimally adjust potential uncertainties in a mathematical process model (such as parameters or unmodeled disturbances) to render cleaned measurements which are consistent with predicted process dynamics. Typically, besides removing or at least smoothing unavoidable measurement error, an estimate of the process state, the unknown parameters and some additional quantities representing model uncertainty is achieved simultaneously. The measurement errors include (i) small random fluctuations caused by the irreproducibility of the measurement device, (ii) potentially time-varying offsets or biases due to miscalibration, incorrect installation or drift of the sensor, and (iii) gross errors resulting from discrete events such as malfunctioning measurement or signal processing devices on the one hand or from process upsets on the other hand [44].

The problem complexity ranges from situations, where the full process state is measured and the model is perfect, to those situations usually occurring in realistic applications, where only a small number of process quantities is accessible by instrumentation, and where there are model uncertainties of various kinds. In the first case, only rectification of erroneous measurements must be achieved, whereas combined state, parameter and other uncertainty estimates must be provided by the reconciliation algorithm in addition in the latter case. Obviously, measurements and model must reveal a certain degree of redundancy in the latter case in particular, in order to render improvement in the information content of the available measurements feasible.

In the sequel, we first discuss a priori knowledge to be coded in a process model. Then, the estimation problem will be set up and discussed.

2.1 Process model

The process model under consideration can be associated with different types of process quantities. They include

- the measured input variables $\boldsymbol{u} \in (L_2)^{n_u}$, becoming dynamic forcing functions in the process model,

- the unmeasured inputs or time-varying parameters $\boldsymbol{\delta} \in (L_2)^{n_\delta}$, leading to unknown forcing functions in the process model,
- the time-invariant parameters $\boldsymbol{\theta} = [\boldsymbol{p}^T, \boldsymbol{\pi}^T]^T$ with $\boldsymbol{p} \in \mathbb{R}^{n_p}$ and $\boldsymbol{\pi} \in \mathbb{R}^{n_\pi}$ being the unknown and the known parameters respectively,
- the state variables $\boldsymbol{z} \in (L_2)^{n_z}$, uniquely representing the process state at future times for given independent initial conditions, inputs and parameters, as well as
- the measurements $\boldsymbol{y} \in (L_2)^{n_y}$, including all measured inputs $\boldsymbol{u}$ and some of the states $\boldsymbol{z}$.

All the process quantities involved in the mathematical model are formally viewed as being real functions of continuous time t. Obviously, this is often not true for the really measured quantities which are subsequently denoted by $\check{\cdot}$. In most practical situations the measurements are sampled at discrete times. Usually (but not necessarily) there is a constant sampling rate for every measurement. In order to allow for the most general situation we assume that l_i measurements at given instants $t_i \in [\tau - T, \tau]$ are taken for every measured process quantity in some time interval $[\tau - T, \tau]$ of length T with τ being the current point in time. The values of the measured quantities $\check{Y}_{i,j}, j = 1, \dots l_i, i = 1, \dots n_y$ are catenated to the vectors $\check{\boldsymbol{Y}}_i \in \mathbb{R}^{l_i}, i = 1, \dots n_y$. The discrete measurements on the time interval $[\tau - T, \tau]$ form the vector $\check{\boldsymbol{Y}} = [\check{\boldsymbol{Y}}_1^T, \check{\boldsymbol{Y}}_2^T, \dots, \check{\boldsymbol{Y}}_{n_y}^T]^T$. In contrast, $\check{y}_i \in L_2$ denotes a real function of continuous time which suitably interpolates (or extrapolates at the boundaries) the discrete measurements $\check{\boldsymbol{Y}}_i$ on the interval $[\tau - T, \tau]$. The interpolated (real) measurements form the vector $\check{\boldsymbol{y}} \in (L_2)^{n_y}$ which corresponds to the measurements $\boldsymbol{y}$ as predicted by the process model. This set-up allows the consideration of multi-rate sampling or even asynchronous sampling in a simple but formally sound manner. Furthermore, as we will see later, this continuous representation permits a more efficient numerical treatment of the estimation problem.

The mathematical model of the process is assumed to be given by the set of differential-algebraic equations

$$\boldsymbol{0} = \boldsymbol{f}(\dot{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}, \boldsymbol{p}) \tag{1}$$

and initial conditions

$$\boldsymbol{0} = \boldsymbol{b}(\dot{\boldsymbol{x}}_0, \boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{d}_0, \boldsymbol{p}) \,. \tag{2}$$

Here, $\boldsymbol{f} : (L_2(\Omega))^{n_x} \times (L_2(\Omega))^{n_x} \times (L_2(\Omega))^{n_u} \times (L_2(\Omega))^{n_d} \times \mathbb{R}^{n_p} \to (L_2(\Omega))^{n_x}$, $\Omega = [\tau - T, \tau]$ represents the n_x differential-algebraic model equations whereas $\boldsymbol{b} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_p} \to \mathbb{R}^r, r \leq n_x$ refers to the set of initial conditions required to guarantee existence of the solution. In this notation the subscript 0 denotes the quantities at the left boundary of Ω, $\tau - T$.

The differential-algebraic system often has a more special structure [47]; in particular, it is of index one and of semi-explicit form in many cases; then r initial values are assigned to a subset of the states $\boldsymbol{x}$ at $t = \tau - T$ (see [65] for a more detailed treatment). The system of equations includes n_z equations for the state variables $\boldsymbol{z}$ as well as $n_x - n_z$ equations for some of the unmeasured time-varying inputs and parameters $\boldsymbol{\delta}$. Often, simple signal models, i.e. a sequence of one or more integrators, are assumed to model these quantities (i.e. [33] or [67] for an example in the context of batch reactor monitoring) when no physical explanation can be employed as for example in catalyst decay modeling [31]. The unknowns $\boldsymbol{x} \in (L_2)^{n_x}$ denote an extended state vector containing the states $\boldsymbol{z}$ and the modeled quantities in the vector $\boldsymbol{\delta}$. The remaining unmodeled quantities in the vector $\boldsymbol{\delta}$ are collected in the vector of unmodeled disturbances $\boldsymbol{d} \in (L_2)^{n_d}$. For brevity, the known parameters $\boldsymbol{\pi}$

are not shown in the equations. For convenience, we assume that suitable reference values for the unmodeled disturbances and for the parameters are included in the model equations. Hence, the quantities $\boldsymbol{d}$ and $\boldsymbol{p}$ denote the difference between the reference values and the actual values; $\boldsymbol{d}$ and $\boldsymbol{p}$ are zero if the reference coincides with the actual value.

The number of equations is typically quite large; the Jacobian is always sparse reflecting the flowsheet topology on a coarse block matrix level and the structure of the unit model in every block matrix. More general models either including discontinuities or comprising also partial differential equations are possible [47] but will not be considered here in more detail.

The equations (1), (2) denote the nominal model of the process. Since the model is never perfect, there is always a mismatch between the real process and the model which needs to be taken into account explicitly. It should be noted that part of the model uncertainty has been already incorporated in the nominal model to the extent that it is transparent to the modeler by means of the unmodeled disturbances $\boldsymbol{d}$ as well as by the unknown constant parameters $\boldsymbol{p}$. However, usually unstructured uncertainty is present in addition. The simplest model of unstructured uncertainty is an additive time-varying forcing function $\boldsymbol{w} \in (L_2)^{n_w}$, $0 \leq n_w \leq n_x$, in the model equations and an additive constant $\boldsymbol{w}_0 \in \mathbb{R}^{n_w}$ in the initial conditions:

$$\mathbf{0} = \boldsymbol{f}(\dot{\boldsymbol{x}}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{d}, \boldsymbol{p}) + \boldsymbol{W}\boldsymbol{w}\,, \tag{3}$$

$$\mathbf{0} = \boldsymbol{b}(\dot{\boldsymbol{x}}_0, \boldsymbol{x}_0, \boldsymbol{u}_0, \boldsymbol{d}_0, \boldsymbol{p}) + \boldsymbol{W}_0\boldsymbol{w}_0\,. \tag{4}$$

The rectangular matrices $\boldsymbol{W}$ and $\boldsymbol{W}_0$ are indicator matrices containing a unit element in each column where a forcing function should be considered in a model equation and are otherwise zero. Note, that there is some arbitrariness in the uncertainty description regarding the quantities $\boldsymbol{d}$ and $\boldsymbol{w}$. However, the disturbances $\boldsymbol{d}$ may enter nonlinearly in the model due to their physical interpretation whereas the error functions $\boldsymbol{w}$ always enter linearly by assumption. Neither of the two types of forcing functions are modeled. It is the decision of the modeler, which of these uncertainty representations should be included in the model. Obviously, the number of variables in $\boldsymbol{d}$ and $\boldsymbol{w}$ should be kept at a minimum for redundancy and computational efficiency reasons.

If a dynamic model for all or some of the error terms $\boldsymbol{w}$ is preferred, it can be covered by the model structure suggested above. In this case, the set of model equations (3), (4) and state vector $\boldsymbol{x}$ is appropriately extended and the indicator matrices $\boldsymbol{W}$ and $\boldsymbol{W}_0$ are adjusted accordingly. Often, such dynamic models describing unstructured uncertainty are difficult to justify.

This model formulation generalizes previous work in dynamic data reconciliation such as that by Almasy [32], Edgar [44], or Biegler [2] and coworkers. The major difference is the way in which we are explicitly dealing with different types of uncertainties as well as with the measurement in general not covering the full process state in a general differential-algebraic setting. The measurements $\boldsymbol{y} = [\boldsymbol{y}_x^T, \boldsymbol{y}_u^T]^T$, $\boldsymbol{y}_x \in (L_2)^{n_{y_x}}$, $\boldsymbol{y}_u \in (L_2)^{n_{y_u}}$, $n_{y_u} = n_u$ and $n_y = n_{y_u} + n_{y_x}$, as predicted by the model are

$$\boldsymbol{y}_x = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{V}_x\boldsymbol{v}_x\,, \tag{5}$$

$$\boldsymbol{y}_u = \boldsymbol{u} + \boldsymbol{V}_u\boldsymbol{v}_u\,. \tag{6}$$

The measurement errors $\boldsymbol{v} = [\boldsymbol{v}_x^T, \boldsymbol{v}_u^T]^T$, $\boldsymbol{v}_x \in (L_2)^{n_{v_x}}$, $\boldsymbol{v}_u \in (L_2)^{n_{v_u}}$, $0 \leq n_{v_x} \leq n_{y_x}$ and $0 \leq n_{v_u} \leq n_{y_u}$, are selected by rectangular indicator matrices $\boldsymbol{v}_x$ and $\boldsymbol{V}_u$. In the differential-algebraic setting considered here, the rectangular matrix $\boldsymbol{C}$ is an indicator matrix selecting some of the state variables $\boldsymbol{x}$ without loss of generality. Nonlinear measurement functions as usually considered in nonlinear systems theory are here included as algebraic model equations in the vector $\boldsymbol{f}$ introducing auxiliary state variables.

If the model (3) is simplified to a system of pure ordinary differential equations, the measurement equation should be generalized to

$$\begin{aligned} \boldsymbol{y}_x &= \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{V}_x \boldsymbol{v}_x \,, \qquad (7) \\ \boldsymbol{y}_u &= \boldsymbol{u} + \boldsymbol{V}_u \boldsymbol{v}_u \,. \qquad (8) \end{aligned}$$

with $\boldsymbol{g} : (L_2(\Omega))^{n_x} \times (L_2(\Omega))^{n_u} \rightarrow (L_2(\Omega))^{n_{y_x}}$. This measurement model explicitly considers uncertainties in the measured inputs as well as in the measured outputs. Both types of uncertainties are present in any practical situation and are therefore accounted for. This approach has been referred to as *error–in-variables* in the context of dynamic systems by Kim et al. [40]. Our measurement model coincides in part with the literature (i.e. [3] or [55]) but differs from the formulation used for example by Edgar and coworkers in [44, 48]. These authors define the measurement errors as the *difference between the true value of the observed process quantity and the measurement* whereas here the errors $\boldsymbol{v}$ are contaminating the *outputs predicted by the model.* We prefer this formulation since the true values of the observed variables are not accessible in practice, as they always differ from the values predicted by the model.

Besides the model equations, there is additional a-priori knowledge which should be stated and used in the estimation procedure. Following Kim et al. [40] or Robertson et al. [55], lower and upper bounds can be formulated for any of the process quantities such as the states $\boldsymbol{x}$, the inputs $\boldsymbol{u}$, the disturbances $\boldsymbol{d}$, the parameters $\boldsymbol{p}$, the measurements $\boldsymbol{y}$, and the model and measurement errors $\boldsymbol{v}$ and $\boldsymbol{w}$. These additional inequality constraints are of the type

$$s_{i,min} \leq s_i \leq s_{i,max} \,, \qquad (9)$$

where s_i refers to some scalar process quantity [55]. These inequality constraints must be formulated to obtain a feasible set [52]. It should be noted that the constraints also refer to a bound on the function norm for the process quantities.

In addition to the simple bounds modeled by (9) statements on the stochastic properties of the process quantities may be included if a probabilistic interpretation of the estimation problem is pursued as done in classical filtering theory. In this case, stochastic processes with certain properties are associated with the model errors $\boldsymbol{w}$ and the measurement errors $\boldsymbol{v}$. Pragmatically, uncorrelated zero mean random variables with Gaussian probability distribution and given variance-covariance information are assumed. This assumption is crucial for reducing the analytical and computational complexity of the estimation problem; at the same time it permits an interpretation of the estimation quality and supports the choice of the estimator's tuning parameters in the linear case. In the nonlinear case, which prevails in most realistic scenarios, the advantages of using such a probabilistic interpretation largely break down. The more general situation is extremely difficult to tackle even if perfect information on the probabilistic properties of the process quantities were available. For an in-depth discussion we refer to the monographs [33] and [37] or to Robertson et al. [55] who compare stochastic filtering with mathematical programming techniques.

The general problem formulation, we are interested in, aims at avoiding any assumptions on statistical properties of the process quantities and error functions. Therefore, a deterministic rather than a stochastic setting is envisaged. This decision is based on the fact that there is no justification for simple assumptions on the probabilistic properties of the model and measurement errors in almost any real situation. Since the model errors $\boldsymbol{w}$ account for unstructured uncertainty in the process model, the signal is typically correlated in time and reveals a non-zero mean on some time interval. The signals cannot be expected to be Gaussian. Similar arguments hold for the measurement errors $\boldsymbol{v}$. In these situations the filtering approaches have shown poor performance and are extremely difficult to tune (see [52] for a summarizing discussion) aside from the loss of the theoretical foundation. We accept (at least for the moment) the loss of potentially useful information on the process statistics.

2.2 Estimation problem

For reconciliation of the measured data with the process model we may require that the degrees of freedom in the model (3)-(6) are chosen in such a way that the predicted measurements $\boldsymbol{y}$ exactly coincide with the measurements $\check{\boldsymbol{y}}$, i.e.

$$\boldsymbol{r}(t) := \boldsymbol{y}(t) - \check{\boldsymbol{y}}(t) = 0\,, \quad \forall t \in [0, \tau] \tag{10}$$

from the startup of the process at $t_0 = 0$ to the current time τ. These degrees of freedom are $\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{d}, \boldsymbol{p}$ and those r initial values $\dot{\boldsymbol{x}}_0, \boldsymbol{x}_0$ collected in $\boldsymbol{x}_0'$ for later reference to which arbitrary values can be assigned. In the sequel, these degrees of freedom will be collected in the vector $\boldsymbol{\Delta}$ for conciseness of notation.

2.2.1 Least-squares formulation

The requirement (10) would not be appropriate since the real measurements are not reflecting the measured process quantities precisely due to contamination with measurement error. A reasonable relaxed requirement is therefore a least-squares solution to equation (10), which is given by those $\hat{\boldsymbol{\Delta}}$ minimizing a cost functional Υ according to

$$\hat{\boldsymbol{\Delta}} = argmin\ \Upsilon(\boldsymbol{\Delta}) = \parallel \boldsymbol{Q}(\boldsymbol{y} - \check{\boldsymbol{y}}) \parallel \tag{P1}$$

subject to the equality constraints provided by the model (3)-(6) and a feasible set of inequality constraints of type (9) for those process quantities and error functions which can be bounded by a-priori knowledge. The weighted norm of the residual $\boldsymbol{r}$ may be defined by

$$\parallel \boldsymbol{Q}\boldsymbol{r} \parallel^2_{L_2([0,\,\tau])} = \int_0^\tau \boldsymbol{r}(t)^T\, \boldsymbol{Q}^T\boldsymbol{Q}\, \boldsymbol{r}(t)dt\,. \tag{11}$$

Then, (P1) is equivalent to a weighted least squares solution of (10) with the symmetric weighting matrix $\boldsymbol{Q}^T\boldsymbol{Q}$. A more general function norm could be used instead of the L_2-norm. An advantageous choice is the Sobolev norm for integer β in L_2 as given by

$$\parallel \boldsymbol{r} \parallel^2_{H^\beta([0,\tau])} = \parallel \boldsymbol{r}^{(0)} \parallel^2_{L_2([0,\tau])} + \dots + \parallel \boldsymbol{r}^{(\beta)} \parallel^2_{L_2([0,\tau])}\,. \tag{12}$$

Here $\boldsymbol{r}^{(\beta)}$ denotes the β-th weak derivative of the function $\boldsymbol{r}$ with respect to time and is required to be square integrable. Note that the Sobolev norm with the associated Sobolev regularity β coincides with the usual L_2-norm for the choice $\beta = 0$; employing the Sobolev norm instead of the L_2-norm in (P1) therefore generalizes the least-squares formulation.

The estimation problem (P1) is a potentially large-scale dynamic optimization problem to be solved by an appropriate technique at current time τ to provide a reconciled set of measurements $\hat{\boldsymbol{y}}$ as well as estimated $\hat{\Delta}$ after the computational delay time h at $t = \tau + h$. For on-line dynamic data reconciliation the problem must be solved repeatedly with a (not necessarily) constant sampling interval of $H \geq h$. Obviously, we would like to choose H as small as possible to get the estimate with minimal delay. The choice of H will be constrained by the algorithmic complexity and by computing resources.

This optimization problem may not be computationally tractable since the computational effort may grow unboundedly with time. On the other hand, it is also not necessary to include the measurements for $t < \tau - T_{max}$ in the reconciliation. This is due to the uncertainty in the measurements $\breve{\boldsymbol{y}}$ which dominates for a sufficiently long horizon T_{max} (much larger than the dominating process time constant) the true process measurement rendering an unfavorable signal-to-noise ratio. Hence a modified least-squares problem is formulated which reads

$$\hat{\boldsymbol{\Delta}} = argmin\, \Upsilon(\boldsymbol{\Delta}) = \| \boldsymbol{Q}(\boldsymbol{y} - \breve{\boldsymbol{y}}) \|^2_{H^\beta([\tau - T, \tau])} \tag{P2}$$

subject to the same equality and inequality constraints as in problem (P1) above. In order to capture all relevant measurement information, the length of the time interval T, the horizon, should be chosen as large as computationally tractable for good estimation quality as pointed out by various authors (i.e. [44] or [55]). As above, the problem is solved repeatedly with sampling interval $H > h$. This solution approach has been previously called *moving* or *receding horizon estimation* (i.e. [53, 55]).

2.2.2 Ill-posedness of the least-squares problem

Both estimation problems belong to the class of so-called *inverse problems* since we want to *determine causes for an observed effect* [27]. Most often these inverse problems do not fulfill Hadamard's postulates of well-posedness [28]. Any *well-posed problem in the sense of Hadamard* must possess the following properties:

1. A solution exists for all admissible data.
2. The solution is unique for all admissible data.
3. The solution depends continuously on the data.

If one of these properties is not fulfilled the problem is termed *ill-posed.* In the context of dynamic data reconciliation *data* refers to the measured inputs and states. These requirements are minimalistic. Obviously, there are stronger requirements, such as the convergence of the nonmeasured states to the true states (at least in the nominal case where there is neither measurement error nor model uncertainty) which are related to observability (see e.g. [50, 51, 53, 64] for a discussion in the context of state estimation).

Obviously, the estimation problems (P1) or (P2) are not set up in a reasonable way, if they are not at least well-posed in the Hadamard sense. A rigorous analysis of the nonlinear problem is quite difficult and, to our knowledge, not available yet. However, some qualitative considerations easily reveal the ill-posedness of (P1) and (P2).

We are not concerned with existence, since we can expect that there is a minimum of the cost functional Υ in both, (P1) and (P2), if the constraints set is feasible. Uniqueness is more of a concern. Let us look at the very simple case given by

$$\dot{x} = -ax + w\,, \quad x(0) = x_0\,, \tag{13}$$

$$y = x + v \tag{14}$$

with known parameter a, unknown initial condition x_0 and error terms v and w. The predicted output y is given by the linear operator equation

$$y(t) = e^{-at}x_0 + \int_0^t e^{-a(t-s)}w(s)ds + v(t)\,. \tag{15}$$

The cost functional (of problem (P1)) reads

$$\Upsilon(x_0, v, w) = \| (y - \breve{y}) \|^2_{H^\beta([0,\tau])} \tag{16}$$

with y taken from (15). We cannot expect to get a unique solution for the degrees of freedom since the same best $\hat{y}$ in the least-squares sense can be established by an infinite number of solutions. Remember, that there are no restrictions on the sign of the additive terms in (15). Hence, given one solution $\{\hat{y}, \hat{x}_0, \hat{v}, \hat{w}\}$, we can easily find many other solutions by arbitrarily changing x_0 or w and compensate for this influence by an adjusted v in (15). Note that there are many more subtle possibilities to get solutions with the same $\hat{y}$.

This non-uniqueness of the least-squares solution is characteristic for *many* inverse problems. An established approach to enforce uniqueness is to define the *best-approximate solution* as a *least-squares solution* of (10) where a *suitable norm* of the degrees of freedom at the solution *takes its minimum* [27]. The best-approximate solution is therefore given by the least-squares solution with the smallest possible norm of all values of degrees of freedom which include the initial conditions as well as the model and measurement uncertainties. For linear inverse problems, the best-approximate solution is unique and closely related to the Moore-Penrose (generalized) inverse of the linear operator mapping the degrees of freedom to the measurements. An alternative related approach modifies the least-squares cost functional by a regularization term penalizing the norm of the degrees of freedom, or, more generally the norm of some differential operator applied to the degrees of freedom in order to enforce uniqueness.

So far, we did not address the third requirement for a well-posed problem which asks for a continuous dependence of the solution on the data. This property guarantees that small changes in the measured data will not lead to completely different reconciled measurements and estimated degrees of freedom. Again, we rely for the moment on the results obtained for general linear inverse problems as presented in [27] and briefly summarized below.

It can be shown that the solution to the regularized problem tends to the best approximation for error free measurements if the regularization penalty approaches zero. However, for a certain fixed level of the measurement errors the difference between the regularized solution for exact data and the regularized solution for perturbed data may assume very large values if the regularization penalty approaches zero. This difference is proportional to the measurement error level and may grow without bound depending on the spectral properties of the linear operator in the equality constraints. Hence, only with non-vanishing regularization, continuous dependence of the solution on the data can be achieved.

For given regularization parameters and fixed level of the measurement error, the estimation error can be bounded by a function of the regularization penalty and the order of magnitude of the measurement error. The estimation error approaches zero for a suitable regularization adjusted to the level of measurement error as the measurement error is reduced to zero. The estimation error depends therefore only on the level of the measurement error provided the regularization penalty can be chosen in some optimal manner.

Hence, in the linear case, regularization of the least-squares problems (P1),(P2) not only renders a unique solution but also reduces the sensitivity of the solution to the data and provides a theoretical basis to bound the estimation error if the order of magnitude of the measurement error is available. Therefore, a generalization of such an approach to nonlinear problems seems to be appropriate [28] and has already been applied in system identification problems [38].

2.2.3 Regularized estimation problem

Using the regularization approach as suggested in the theory of inverse problems [28] we could consider to (heuristically) modify the cost functional in (P1) and (P2) as follows

$$\begin{aligned}\Upsilon(\boldsymbol{\Delta}) &= \| \boldsymbol{Q}(\boldsymbol{y}-\breve{\boldsymbol{y}}) \|^2_{H^{\beta_y}([t_0,\tau])} + \alpha_v \| \boldsymbol{R_v v} \|^2_{H^{\beta_v}([t_0,\tau])} \\ &\quad + \alpha_w \| \boldsymbol{R_w w} \|^2_{H^{\beta_w}([t_0,\tau])} + \alpha_d \| \boldsymbol{R_d d} \|^2_{H^{\beta_d}([t_0,\tau])} \\ &\quad + \alpha_p \| \boldsymbol{R_p p} \|^2_{L_2} + \alpha_0 \| \boldsymbol{R}_0(\boldsymbol{x}_0' - \boldsymbol{x}_{0,ref}') \|^2_{L_2} \ . \end{aligned} \tag{17}$$

Here, t_0 is either 0 for (P1) or $\tau - T$ for (P2). $\boldsymbol{R}_v, \boldsymbol{R}_w, \boldsymbol{R}_d, \boldsymbol{R}_p$ and $\boldsymbol{R}_0$ are nonsingular weighting matrices whereas $\alpha_v, \alpha_w, \alpha_d, \alpha_p$, and α_0 are nonnegative constants called regularization parameters.

Let us look at this objective function in the nominal case where there are no uncertainties either in the model or in the measurements. Then, the minimization of the cost functional should lead to $\hat{\boldsymbol{v}} = \hat{\boldsymbol{w}} = \hat{\boldsymbol{d}} = \boldsymbol{0}$, $\forall t$, $\hat{\boldsymbol{p}} = \boldsymbol{0}$ and $\breve{\boldsymbol{y}} = \hat{\boldsymbol{y}}, \forall t$, if the reference values for $\boldsymbol{d}$ and $\boldsymbol{p}$ are chosen as the actual values (which is always possible in the nominal case).

We would expect the cost functional to be zero in the nominal case. This could only be achieved if we knew the correct initial values which lead to $\breve{\boldsymbol{y}}$ which is impossible. Therefore, the cost functional will take at the minimum some positive value $\hat{\Upsilon} \geq \alpha_0 \| \boldsymbol{R}_0(\hat{\boldsymbol{x}}_0' - \boldsymbol{x}_{0,ref}') \|^2_{L_2}$. In general we cannot expect to get correct estimates by minimizing the cost functional (17), since the minimum $\hat{\Upsilon}$ might be obtained with different $\hat{\boldsymbol{x}}_0'$ and some non-vanishing and hence incorrect solution $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}, \hat{\boldsymbol{d}}$, or $\hat{\boldsymbol{p}}$. Based on this argument, we drop the regularization of the initial values in the cost functional and formulate the estimation problem as

$$\hat{\boldsymbol{\Delta}} = argmin \ \Upsilon(\boldsymbol{\Delta}) \tag{18}$$

with the cost functional

$$\begin{aligned}\Upsilon(\boldsymbol{\Delta}) &= \| \boldsymbol{Q}(\boldsymbol{y}-\breve{\boldsymbol{y}}) \|^2_{H^{\beta_y}([t_0,\tau])} \\ &\quad + \alpha_v \| \boldsymbol{R}_v \boldsymbol{v} \|^2_{H^{\beta_v}([t_0,\tau])} + \alpha_w \| \boldsymbol{R}_w \boldsymbol{w} \|^2_{H^{s_\beta}([t_0,\tau])} \\ &\quad + \alpha_d \| \boldsymbol{R}_d \boldsymbol{d} \|^2_{H^{\beta_d}([t_0,\tau])} + \alpha_p \| \boldsymbol{R}_p \boldsymbol{p} \|^2_{L_2} \end{aligned} \tag{P3}$$

subject to the equality constraints (3)-(6) and inequality constraints (9).

Note, that Rawlings and coworkers [53, 55] found in the context of nonlinear state estimation that penalizing the initial conditions does not guarantee convergence in the nominal case whereas convergence can be shown if there is no such penalty. The penalty in the initial conditions is used in [55] to optimally reflect the past measurements in a receding horizon estimator by adjusting the weight matrix in the norm according to *linear* filtering theory. This strategy is useful for short estimation horizons whereas it is not really required if the estimation horizon is taken sufficiently large [52].

For later reference in the context of the discussion of the multiscale framework in Section 5 we formulate here a simplified version of (P3) where the model (3) is substituted by a system of ordinary differential equations and the measurement model is restricted to (7). Further, we assume that the indicator matrices $\boldsymbol{W}$ and $\boldsymbol{V}_x$ are identity matrices thus allowing the error functions to contaminate every state $\boldsymbol{x}$ and measurement $\boldsymbol{y}$. $\boldsymbol{V}_u$ is assumed to be zero which corresponds to error free measured inputs. We are neither considering unknown parameters $\boldsymbol{p}$ nor disturbances $\boldsymbol{d}$. The variables with specified bounds are catenated in the vector $\boldsymbol{s}$ with upper bounds $\boldsymbol{s}_{max}$ and lower bounds $\boldsymbol{s}_{min}$. For simplicity of notation, the regularization parameters are included in the matrices $\boldsymbol{Q}, \boldsymbol{R}_v, \boldsymbol{R}_w$. In summary, these modifications yield the simplified dynamic data reconciliation problem

$$\min \| \boldsymbol{Q}(\boldsymbol{y} - \breve{\boldsymbol{y}}) \|^2_{H^{\beta_y}([t_0,\tau])} + \| \boldsymbol{R}_v \boldsymbol{v} \|^2_{H^{\beta_v}([t_0,\tau])} + \| \boldsymbol{R}_w \boldsymbol{w} \|^2_{H^{\beta_w}([t_0,\tau])} \tag{P4}$$

subject to the constraints

$$\begin{aligned} \dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{w} &= \boldsymbol{0}\,, \\ \boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{v} &= \boldsymbol{0}\,, \\ \boldsymbol{s} - \boldsymbol{s}_{max} &\leq 0\,, \\ \boldsymbol{s}_{min} - \boldsymbol{s} &\leq 0\,. \end{aligned}$$

2.2.4 Relation to previous work

Some authors (e.g. [44] or [54]) use the formulation of (P1) which without any regularization is ill-posed. Their solution does not obviously reveal amplification of measurement errors since the relatively coarse discretization of continuous time acts itself as some regularization of the estimation scheme [27]. In fact, Liebmann et al. [44] observe improved estimation quality in the context of bias estimation if they penalize the bias (corresponding to our $\boldsymbol{v}$) in the least-squares objective function, i.e. if they regularize the ill-posed problem.

The formulation suggested by Rawlings and coworkers (see [55] for example) coincides with a specialization of (P3) (e.g. their formulation does not consider our $\boldsymbol{p}$ and $\boldsymbol{d}$ variables) *if we add the constraints* (10) to (P3). Inserting this equality of real and predicted measurements in (P3) eliminates the first term in the cost functional and hence leads to the objective function used by Robertson et al. [55]. These authors therefore determine the modeled measurement errors such that the real measurements coincide exactly with the predicted measurements whereas our formulation is less restrictive at the expense of a larger number of degrees of freedom.

Finally, we want to compare our approach with the work of Michalska and Mayne [50, 51]. These authors suggest a moving horizon observer for error free measurements and a perfect model using a formulation similar to (P2). They repeatedly solve an optimization problem but do not require to reach the minimum of the cost functional. Rather, they only attempt to reduce the cost functional at each step by a given factor which assures convergence of the estimated state to the real state at a given rate as time goes to infinity under an observability condition by adjusting either the initial conditions or in addition an observer input. They motivate their approach with the limited computational resources available in real-time applications. The approach suggested here is very similar to the moving horizon observer in the nominal case if the true minimum is computed in every sampling interval.

3 Research questions

In conclusion of the preceding discussion, we want to explicitly state that there is no guarantee that the formulation of the dynamic data reconciliation problem suggested here satisfies the requirements of well-posedness nor that there is convergence of all the estimated states to the true states for the nominal model. The formulation of the dynamic data reconciliation problem (P3) is still heuristic at this point although it is a reasonable extension of (P1) or (P2) which have been studied in the literature though they are definitely ill-posed. This leads us to open research questions to be tackled before the suggested problem formulation can be safely applied to dynamic data reconciliation in practice:

1. Definition of reconcilability: reconcilability should be a property of the set of measurements and the process model alone. The discussion above shows, that we cannot avoid to include the way the estimation problem is formulated. A first suggestion is as follows:

   > A set of measurements and a given process model are called reconcilable, if there exists an algorithm which renders a unique solution for the degrees of freedom in the model for all admissible data. The solution must depend continuously on the data. In the limit of vanishing measurement error it must converge to the best-approximate solution.

2. Relationship between reconcilability and observability: Obviously, observability is a much stricter requirement than reconcilability. This system property may, however, not be of very much use in situations where there is significant error (random, systematic or gross) in the measurements.

3. Analysis of (P3) and its specializations regarding estimation quality. Here, we may distinguish convergence of the estimate to the least-squares solution for vanishing measurement error as well as convergence of the estimated degrees of freedom to the true values at least in limiting situations. This analysis may reveal necessary modifications to the suggested problem formulation. One objective of the analysis is the minimization of the number of tuning parameters and constructive guide lines for their selection.

4. Criteria and strategies to guarantee sufficient redundancy in the data and the process model to result in useful reconciled measurements and useful estimates for the degrees of freedom (see [3] for a first treatment). The selection of a set of degrees of freedom $\mathbf{\Delta}$ of minimal cardinality must also be addressed, trading off the requirements on the measurements needed for proper identification on the one and the model quality on the other hand.

5. Methods for preprocessing the measured data in order to reduce some norm of the measurement error and to provide a reasonable interpolation of the discrete measurements. The estimation quality achievable will depend on the level of measurement error as indicated by the results on linear inverse problems. Outlier removal and denoising are important requirements in this context.

6. Highly efficient numerical solution of (P3) with $t_0 = 0$ for a given set of measurements in $[0, \tau]$. This problem would be typically solved off-line to support mechanistic modeling of some (bench-scale, laboratory-scale or industrial-scale) process or to diagnose the performance of some industrial process. There are no strict limitations on computational resources. More recent contributions on numerical solution algorithms are [4, 14, 29, 61, 62, 63, 68, 69].

7. Highly efficient numerical solution of (P3) with $t_0 = \tau - T$ in a repetitive manner at a sampling interval of H for a long horizon T which takes into account the most recent measurements becoming available in $[\tau, \tau + H]$. This problem must be solved on-line in real-time obeying strict limitations of the computational resources. The numerical solution approach must take advantage of the repetitive nature and other structural properties of the problem in order to become feasible in practical implementations. A reasonable approximation to the solution of the problem must be provided within a real-time interval of H. The quality of the approximate solution should scale with H, e.g. the solution accuracy must increase with the time available for carrying out the

numerical computations. Some recent contributions addressing the solution of dynamic optimization problems on a moving horizon are [2, 4, 44, 58, 71]. It should be noted that any optimization based algorithm has to somehow compete with computationally efficient observers or filters [52].

This is a demanding list. Some of these research issues have already been addressed in previous work. Nevertheless, there are no fully satisfactory solutions available and much work remains to be done. In our work to date, we have concentrated largely on the numerical aspects (items 5, 6 and 7 of the list) as presented in the next sections.

4 Multiscale dynamic data reconciliation - the framework

The numerical solution of dynamic data reconciliation problems (P3) is typically quite challenging, in particular, if on-line applications are envisaged where a reconciled set of measurements and an estimate of the degrees of freedom must be provided in real–time ideally with minimum delay to the last available measurement. Typically, the process model restricting the minimization of the cost function is large-scale. The length of the time interval T employed in moving horizon estimation should be chosen as long as possible to improve the estimation results. Obviously, due to the increasing number of discrete measurements, the computational complexity of the estimation problem increases with T. For a given problem, algorithm and hardware, the required solution time is governed by some h which may strongly vary over time depending on the availability of the process computer and on the frequencies included in the measured (and hence) estimated quantities. Despite these variations in h, the computations must *always* be accomplished in $h \leq H$. The sampling interval H for reconciled measurements and estimated values is usually much longer than most of the sampling intervals of the measurement devices but always fixed. H is chosen such that the information demand of the operating personal or of the control system (potentially including economic set-point optimization) can be satisfied.

The majority of the available algorithms for the solution of (P3) or (P4) can be subdivided into two classes with differing solution strategies. In the so-called *sequential* (or *control-vector parameterization*) *approach* (see [1, 49, 68, 69] for an example) the degrees of freedom are approximated by a polynomial or often piecewise constant function on a fixed time mesh to result in a set of algebraic decision variables. In every iteration, the differential-algebraic equations are integrated together with the sensitivity equations in order to evaluate the objective function and to provide the required gradient information with respect to the decision variables. In contrast, in the *simultaneous approach*, the state variables as well as the degrees of freedom are discretized on a fixed time mesh approximating the dynamic optimization problem by a large-scale nonlinear programming problem (NLP) (see [2, 62] for an example). Both approaches have been used to implement moving horizon estimation algorithms with varying performance. However, due to the inherently expensive integration of the model and sensitivity equations on the one hand and the significant advances in large-scale nonlinear programming on the other, the simultaneous approach may be considered more attractive in the longer run if the (typically superior) robustness of the sequential solution approach can be met by appropriate means.

All the suggested algorithms discretize all or some process quantities on a fixed time mesh. There are only few attempts to incorporate adaptivity in the discretization scheme [46, 61] which have not been used in moving horizon algorithms to our knowledge. Always, the solution is computed in a single step. Since the computation time varies with time and can only be estimated in a very rough sense, there is by no means a guarantee that a valid solution is found within $h \leq H$. To increase the chances that this restriction is met, the user can only relax the tolerances, coarsen the grid, or simplify the process model since the available upper bound on the computing time H can typically not be increased as argued above. This approach would lead to a solution of possibly inadequate accuracy in many cases; it might be over-restrictive because the algorithm most probably will not use the full amount of available computation time to further improve the solution quality.

Multiscale theory [11, 21] seems to offer attractive alternative techniques which might push the limit of large–scale moving horizon computation forward. Such techniques have just recently been applied to moving horizon estimation and control problems [43, 60, 59]. Inspired by the adaptivity inherent to multiscale theory [15] we focus in our work on a concept which steadily refines the optimization problem starting from a coarse approximation up to arbitrary degree of detail. This concept differs from the well established single step solution approach such that the optimization problem is solved repeatedly on various discretization meshes of increasing number of decision variables and approximation quality. This simultaneous approach is based on discretizing all time-varying functions using wavelets which support refinement on various scales to resolve the global trend of the functions as well as the local detail in an appropriate manner. The suggested fully adaptive method will prompt during one cycle in the moving horizon approach a valid solution of some accuracy associated with the level of refinement at any time during the course of refinement. Moreover, the method always utilizes the available computation time H, since a solution of a coarse approximation quality on level $\ell - 1$ is always available if the time limit is reached while the computations on the current level ℓ have not been completed. In addition, preprocessing of the measured data is envisaged to be treated in an integrated manner with the solution of the optimization.

The general outline of the refinement strategy is schematically given in Figure 1. Since the measurements are sampled in discrete time an appropriate transformation into continuous time representation is necessary. As it will be seen later, this functional representation is the key for a substantially compressed representation of the measurements. Compression of the measurements is used here to avoid the substantial growth of computational complexity with an increasing number of discrete measurement samples in a different manner than explored recently in [2].

The measurements, once in continuous representation, can as well be discretized in various approximation levels with wavelet based multiscale techniques. Since the measurements are corrupted with noise, outliers and the like, they are subject to pretreatment. The objective is in general to reduce the norm of the measurement error or more precisely to eliminate noise or outliers or to interpolate missing data due to a malfunctioning measurement device. According to the theoretical considerations on the solution of inverse problems in the presence of erroneous data, the reduction of some norm of the measurement error not only eases the numerical burden but also leads to smaller estimation errors caused by regularization and measurement error.

The approximated signal is then passed to the optimizer unit where the state variables and degrees of freedom of the optimization problem are discretized on some mesh related to the resolution used to represent the measurements. Thus we obtain an algebraic optimization problem which is solved with an appropriate NLP technique. The two units signal processing and optimization might appear at first glance to be separated. However, a discretization frame spanned by wavelet functions provides an integrated approach which enables easy and fast compression

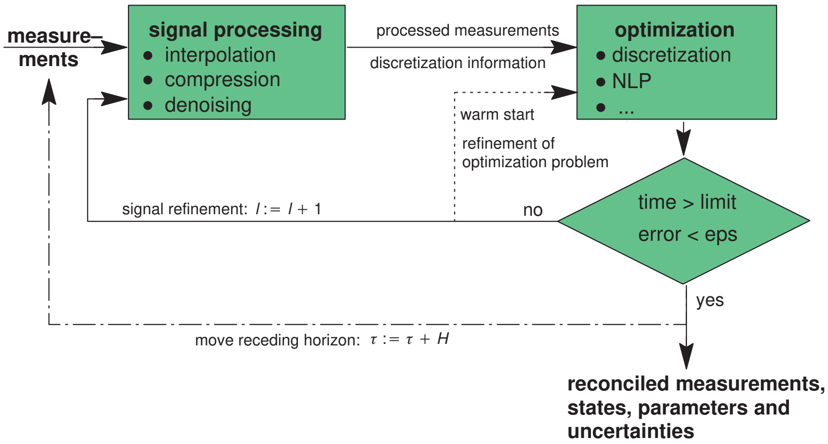

Figure 1: Schematical outline of the refinement strategy

of the problem formulation as well as a fast data exchange between signal and optimization unit.

The numerical solution of the former refinement cycle $\ell - 1$ is reused for a warm start of the optimization in the current cycle ℓ (see [44] for the computational savings with warm start with conventional discretization) avoiding significant computing loads as compared to a start from scratch. Ideally the computational complexity in each refinement step has to be closely related to the number of the additional decision variables included in the actual refinement step. An acceptable complexity is obtained, if the solution of the last approximation can be fully utilized. This includes the evaluation of the discretized equation systems including the Jacobian and Hessian matrices as well as the numerical solution of the updated equation system. For a certain problem class the wavelet discretized equation system can be completely reused while updating. In contrast, applying Runge-Kutta or orthogonal collocation on finite elements discretizations, popular for the discretization of dynamic optimization problems on moving horizons (i.e. [2]), the discrete system of equations have to be completely reformulated and solved in every refinement step.

The adaptive mesh refinement, which includes refinement of the compressed measurement functions as well as refinement of the discretization of the states and degrees of freedom leading to a discretized system of equations of steadily increasing dimension, has to ensure a small additional computational effort compared to the overall time span H available. That way the estimation accuracy is steadily improved with small increments. The procedure is repeated until either the overall approximation error is lower than a specified tolerance or the time restrictions are violated ($h = H$). Even in the latter case where the computation of the actual approximation level is aborted, a solution of the last approximation level is still available. This way, the best valid solution achievable with given computational resources can be always prompted. Finally, the horizon window is moved on, and the multiscale dynamic data reconciliation is repeated on the next time horizon.

5 Multiscale discretization and numerical solution techniques

5.1 Introduction to wavelets

The conceptional ideas of refinement in the multiscale approach have been presented in the last section. Now the ideas have to be put on a solid mathematical basis in order to make the concept concrete. A rigorous treatment of the wavelet methodology is far beyond the scope of this paper. For a more thorough treatment, the reader is referred to [11, 15, 21]. Here we introduce only the notation and the basic concepts, which are relevant for the understanding of our approach. For simplicity, the case of scalar functions is treated, which straightforwardly extends to vector functions. Let the inner product in L_2 be defined by $\langle f, g\rangle := \langle f, g\rangle_{L_2(\Omega)} = \int_0^1 f(t)g(t)dt$ with $\Omega = [0,\ 1]$ and $f, g \in L_2$. Further, we define finite dimensional function spaces spanned by a finite number of basis functions $\varphi_{j,k}$

$$S_j := span\{\varphi_{j,k}\ : k \in \Delta_j\}$$

with the scale dependent index set Δ_j. The basis functions $\varphi_{j,k}$ are derived from a single so-called scaling function $\varphi^{(m)}$. At this point we employ pragmatically a B-spline of order m [22]. In the interior of $[0,\ 1]$ the $\varphi_{j,k}$ are obtained by dyadic dilation with powers of two and integer translation while certain special modifications are necessary near the end points of the interval. The spaces S_j are to form a multiresolution sequence

$$\ldots \subset S_j \subset S_{j+1} \subset \ldots$$

whose union $\mathrm{clos}(\cup_j S_j) = L_2(\Omega)$ is dense. Due to the nestedness of the spaces S_j the basis $\mathbf{\Phi}_j = \{\varphi_{j,k} : k \in \Delta_j\}$ and $\mathbf{\Phi}_{j+1}$ (the set of functions will in the sequel also be interpreted as ordered array) form a two-scale relation which in compact form reads as

$$\mathbf{\Phi}_j^T = \mathbf{\Phi}_{j+1}^T \boldsymbol{M}_{j,0}\,.$$

In other words, each basis function on a coarse scale can be written as a linear combination of basis functions on the next finer scale. Since only finitely many fine scale basis functions appear independently of j in these linear combinations, $\boldsymbol{M}_{j,0}$ is a sparse matrix.

Let us assume that an arbitrary function $f \in S_{j+1}$ is approximated in some sense in S_j then some detail has to be added to get the more accurate approximation in S_{j+1}. This detail is embedded in the detail space

$$W_j := span\{\psi_{j,k}\ : k \in \nabla_j\}$$

such that W_j forms a complement space of S_j to S_{j+1}

$$S_{j+1} = S_j \oplus W_j\,.$$

The basis functions in $\mathbf{\Psi}_j = \{\psi_{j,k} : k \in \nabla_j\}$ of the complement space W_j are called wavelets. Since $W_j \subset S_{j+1}$ the wavelet basis on scale j can also be expressed as a linear combination by scaling functions in $\mathbf{\Phi}_{j+1}$ and hence

$$\mathbf{\Psi}_j^T = \mathbf{\Phi}_{j+1}^T \boldsymbol{M}_{j,1}$$

with $\boldsymbol{M}_{j,1}$ a sparse matrix. Repetition of the decomposition splits the space S_J into a sum of complement spaces

$$S_J = S_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \ldots \oplus W_{J-1}$$

where j_0 is some fixed coarsest scale and J an arbitrary but fixed finest scale. A thorough analysis [15] reveals that not all splittings are equally well suited. In fact, the complement spaces should be chosen in such a way that there exists a dual multiresolution of spaces

$$\tilde{S}_j := span\{\tilde{\varphi}_{j,k} \ : k \in \tilde{\Delta}_j\}$$

with corresponding complements splittings

$$\tilde{W}_j := span\{\tilde{\psi}_{j,k} \ : k \in \tilde{\nabla}_j\}.$$

They are constructed in a similar way based on a scaling function $\tilde{\varphi}^{(m,\tilde{m})}$ satisfying

$$\langle \varphi(\cdot)^{(m)}, \tilde{\varphi}^{(m,\tilde{m})}(\cdot - k)\rangle = \delta_{0,k}$$

for integer $k \in \mathbb{Z}$. Here and in the sequel, the notation $f(\cdot)$ defines a function with generalized argument $\cdot$. Moreover, the dual function spaces should possess also two-scale relations of the above form [16, 17]. For the sake of brevity we do not explicitly give the definitions for the dual function relations but will always employ a $\tilde{\cdot}$ to denote corresponding relations in the dual setting. The dual and primal single- and multiscale spaces are related through the biorthogonality property

$$\langle \boldsymbol{\Psi_j}, \tilde{\boldsymbol{\Psi}}_{\boldsymbol{j}}\rangle = \boldsymbol{I}\,, \quad \langle \boldsymbol{\Phi}_j, \tilde{\boldsymbol{\Phi}}_j\rangle = \boldsymbol{I}\,, \quad j \in \mathbb{N} \tag{19}$$

and

$$\tilde{W}_j \perp S_j\,, \quad W_j \perp \tilde{S}_j \tag{20}$$

where $\boldsymbol{I}$ denotes the identity matrix. The inner products of the functions collected in vectors are defined here and in the sequel by their inner products associated with the corresponding index set and hence form a matrix, e.g. $\langle \boldsymbol{\Psi}, \tilde{\boldsymbol{\Psi}}\rangle := (\langle \psi_{j,k}, \tilde{\psi}_{l,i}\rangle)_{(j,k)\in\nabla,\,(l,i)\in\tilde{\nabla}}$. One can show that these biorthogonality properties form essential preconditions for important stability properties of the basis, sometimes referred to a Riesz basis, explained below.

Every function $f \in L_2$ has a unique expansion

$$f = \sum_{k\in\Delta_{j_0}} \langle f, \tilde{\varphi}_{j_0,k}\rangle \varphi_{j_0,k} + \sum_{j=j_0}^{\infty} \sum_{k=\nabla_j} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k}\,. \tag{21}$$

Upon subsuming all basis functions $\varphi_{j_0,k}, \psi_{j,k}$ in an infinite array $\boldsymbol{\Psi}$ and likewise, all expansion coefficients $\langle f, \tilde{\varphi}_{j_0,k}\rangle, \langle f, \tilde{\psi}_{j,k}\rangle$ in a corresponding array $\boldsymbol{d} := \langle f, \tilde{\boldsymbol{\Psi}}\rangle$, (21) can be written in a compact form as $f = \boldsymbol{d}^T\boldsymbol{\Psi}$. Moreover, in addition to the uniqueness of such expansions a Riesz basis has the remarkable property that the (Euclidean) norm of the discrete array $\boldsymbol{d}$ that characterize the continuous objects $f \in L_2$ behaves essentially as the function norm, i.e.

$$\| f \|_{L_2}^2 \sim \| \boldsymbol{d} \|_{l_2}^2 = \sum_{j,k} |\langle f, \tilde{\psi}_{j,k}\rangle|^2\,. \tag{22}$$

Thus small variations in the coefficients result in small variations of the function and vice versa – a crucial property for numerical calculations. The basis constructed in [16] enjoys all these properties for any desired choices of m and $\tilde{m}$, $\tilde{m} \geq m$, $m + \tilde{m}$ even. These properties remain trivially valid for finite truncations $\Psi^J = \Phi_{j_0} \cup_{j=j_0}^{J-1} \Psi_j$ of Ψ. Thus every function $f \in S_J$ has two equally stable representations

$$\begin{aligned} f &= \sum_{k\in\Delta_J} \langle f, \tilde{\varphi}_{J,k}\rangle \varphi_{J,k} =: \sum_{k\in\Delta_J} c_{J,k}\varphi_{J,k} \\ &= \sum_{j=j_0-1}^{J-1} \sum_{k=\nabla_j} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k} =: \sum_{j=j_0-1}^{J-1} \sum_{k=\nabla_j} d_{j,k}\psi_{j,k} \end{aligned}$$

with $\psi_{j_0-1,1} := \varphi_{j_0,k}$ and $\nabla_{j_0-1} :=\Delta_{j_0}$. To keep the notation simple we express the above as the single- and multiscale representation of f in vector form according to

$$f(t) = \boldsymbol{c}^T \boldsymbol{\Phi}_J = \boldsymbol{d}^T \boldsymbol{\Psi}^J . \tag{23}$$

Both representations are related by the fast wavelet transformation (FWT) $\boldsymbol{c} = \boldsymbol{T}_J \boldsymbol{d}$ where $\boldsymbol{T}_J : \boldsymbol{d} \to \boldsymbol{c}$ is typically executed in cascadic fashion by subsequent applications of the two-scale matrices $\boldsymbol{M}_{j,0}, \boldsymbol{M}_{j,1}$ (see [16, 21]). When, as in the case referred to here, these matrices are sparse, it is not hard to verify that the execution of $\boldsymbol{T}_J$ requires $O(N_J)$ operations where N_J is the cardinality of the index set Δ_J and therefore denotes the dimension of the space S_J.

Let us now comment briefly on the role of the integers $m, \tilde{m}$ appearing in the above description of the basis. The m says that the spaces S_J are exact of order m, i.e. all polynomials up to degree $m-1$ are represented exactly as

$$x^r = \sum_{k\in\Delta_j} \langle (\cdot)^r, \tilde{\varphi}_{j,k}\rangle \varphi_{j,k} , \quad r = 0, \ldots, m-1 . \tag{24}$$

It is well known that polynomial exactness combined with the local supports of the $\varphi_{j,k}$ imply that smooth functions can be approximated with the order of 2^{-mj} by functions in S_j. The $\tilde{m}$ means that the dual multiresolution spaces S_j are exact of order $\tilde{m}$. On account of the biorthogonality relation (19) this in turn is equivalent to saying that

$$\begin{aligned} \int x^r \psi_{j,k}(x)dx &= 0 , \qquad r = 0, \ldots, \tilde{m}-1 , \\ \int x^r \tilde{\psi}_{j,k}(x)dx &= 0 , \qquad r = 0, \ldots, m-1 , \end{aligned} \tag{25}$$

which are often referred to as moment conditions. An immediate consequence of (25) is that one can subtract any polynomial from a function without changing its wavelet coefficients, i.e. $\langle f, \tilde{\psi}_{j,k}\rangle = \langle f - p, \tilde{\psi}_{j,k}\rangle$ where p is any polynomial of order m. Specifically, choosing p as a Taylor polynomial of order m in the support of $\tilde{\psi}_{j,k}$ shows that $\|\langle f, \tilde{\psi}_{j,k}\rangle\|$ will be quite small when f is smooth on the support of $\tilde{\psi}_{j.k}$. Consequently by far most of the wavelet coefficients in the expansion (20) are rather small. Moreover, on account of the norm equivalence (22), discarding small coefficients causes only a small error in the function. Therefore such functions may be approximated very accurately by linear combinations of only relatively very few wavelet basis functions corresponding to the large coefficients. In fact, this number may be much smaller than the dimension of the full space S_J corresponding to the highest refinement scale among the significant coefficients. Note that working with the single scale representation in S_J corresponds to a discretization based on a uniform *mesh* refinement down to width 2^{-J}. The possibly much sparser wavelet representation determined by the significant coefficients of the approximated object means to employ an individually adapted trial space which could be much smaller than S_J and thus leads to a much more economic discretization concept. Such a possibly sparse basis is denoted by

$$\boldsymbol{\Psi}_\Lambda := \{\psi_\lambda : \lambda \in \Lambda\} \tag{26}$$

where $\Lambda \subseteq \Lambda^J :=\Delta_{j_0} \cup_{j=j_0}^{J-1} \nabla_j$ is an arbitrary subset of the complete wavelet basis Λ^J. It is only assumed that Λ contains the basis functions in Δ_{j_0}, i. e. the basis functions of the coarsest single scale space S_{j_0}.

By now we have established the most important ingredients in our multiscale approach and continue with the signal processing.

5.2 Signal processing

As mentioned earlier the tasks to be performed in the signal processing part are

- transformation of the discrete measurement samples into a continuous representation,
- denoising of the measurement function,
- outlier removal,
- extrapolation of missing signal parts,
- data compression, which refers to the representation of the measurement signal on various levels of resolution.

Recent developments show that wavelets are very well suited for dealing with these tasks (e.g. [6, 7, 9, 24, 45]). We restrict the following brief presentation to interpolation, denoising and compression only. For the sake of brevity we do not attempt to give a survey on the rich literature on wavelet based signal processing. Instead we illuminate only the basic principles needed to gain a more thorough understanding of the proposed approach.

The same function spaces for signal processing as well as for the discretization of the dynamic optimization problem are used. As it turns out later, this leads to an integrated approach where the processed measurement data can be passed on fast and easily to the optimization problem on various scales of resolution. To ease notation we project the time interval $[\tau - T, \tau]$ of the moving horizon onto the unity interval $\Omega = [0, 1]$ still assuming that the measurement samples are given by $\breve{\boldsymbol{Y}} \in \Omega$. Further, we consider only one measured quantity $\breve{\boldsymbol{Y}}_i$, since the discrete measurements are processed for each single quantity separately.

5.2.1 Interpolation using a multiscale basis

The discrete, noisy, possibly non-equidistant discrete measurements $\breve{\boldsymbol{Y}}_i$ can be transformed into a continuous function by a basis spanned by scaling functions. Here, an approximation based on B-splines $\varphi^{(m)}$ is used to fit the basis used later for the discretization of the dynamic optimization problem. The continuous measurement function $\breve{y}_i$ is expanded in single scale representation in S_J where J is chosen sufficiently large to consider all l_i measurement samples $\breve{\boldsymbol{Y}}_i$ such that $\breve{Y}_{i,j} = \breve{y}_i^J(t_j) = \boldsymbol{c}_{\breve{y}_i}^T \boldsymbol{\Phi}_J(t_j), \forall t_j,\ j = 1, \ldots, l_i$. This function can then be transformed into the equivalent multiscale representation $\breve{y}_i^J = \boldsymbol{d}_{\breve{y}_i}^T \boldsymbol{\Psi}^J$ applying the fast wavelet transformation.

Of course, the wavelet interpolation of the measurement function does not limit further signal cleaning to wavelet based techniques. Rather, any other classical approach may be applied. However, we expect that the use of wavelets offers a theoretically sound and unifying way of getting a reasonably preprocessed measurement signal to feed it into the optimization algorithm. It should be noted, that any preprocessing will modify the (unknown) statistics of the measurement signal.

5.2.2 Denoising

Now $\breve{y}_i^J$ is transformed into the signal $\bar{y}_i^J$ where the over bar denotes the processed interpolated signal after denoising. To accomplish this task various methods are suggested by several authors (e.g. [10, 24, 25]). Without going into detail we only mention one common method. There, the expansion coefficients $\boldsymbol{d}_{\breve{y}_i^J}$ of the interpolated signal $\breve{y}_i^J$ are replaced by modified coefficients $\boldsymbol{d}_{\bar{y}_i^J}$ according to the shrinkage algorithm

$$\boldsymbol{d}_{\bar{y}_i^J} \quad := \quad \begin{cases} \boldsymbol{d}_{\breve{y}_i^J} - (\operatorname{sign} \boldsymbol{d}_{\breve{y}_i^J})\varepsilon & |\boldsymbol{d}_{\breve{y}_i^J}| \geq \varepsilon\,, \\ \boldsymbol{0} & |\boldsymbol{d}_{\breve{y}_i^J}| < \varepsilon\,. \end{cases}$$

This strategy interprets small wavelet coefficients as noise and reduces their magnitude by a nonlinear algorithm which leads to a smoothed (and compressed) signal. Of course, the question how to choose the threshold ε is subtle and has to be addressed with care. The more it is known about the noise and the smoothness of the signal the more rigorously ε can be determined. For example under the assumption of uncorrelated normal distributed white noise a level independent threshold ε can be derived only knowing the coefficients $\boldsymbol{d}_{\breve{y}_i^J}$. Otherwise, if little or nothing is known about the noise, ε has to be tuned in a rather heuristic manner, usually depending on location, magnitude or scale.

5.2.3 Representation of the signal in various resolutions

Keeping the refinement strategy in mind, the signal $\bar{y}_i^J$ is represented in various levels of resolutions indicated by the refinement index ℓ. Because of the norm equivalences (22), approximations of $\bar{y}_i^J$ can easily be obtained in multiscale representation retaining only those coefficients $d_{j,k},\ j,k \in \Lambda_\ell \subset \Lambda^J$ which exceed a certain threshold ε_ℓ in magnitude:

$$\bar{y}_{i,\ell} = \sum_{|d_{j,k}| \geq \varepsilon_\ell} d_{j,k}\psi_{j,k} = \boldsymbol{d}_{\bar{y}_{i,\ell}}^T \boldsymbol{\Psi}_{\Lambda_{\bar{y}_{i,\ell}}}\,.$$

The threshold ε_ℓ is chosen such that the approximation is compressed in a sense that a specified accuracy of the approximation is preserved [9, 23, 24, 45]. The signal $\bar{y}_i^J$ is first represented by a coarse approximation determined by a comparatively large threshold ε_0. The approximation accuracy is increased taking an extended set of coefficients into account which corresponds to $\varepsilon_{\ell+1} < \varepsilon_\ell$. Since the number of significant wavelet coefficients is often far less than the number of discrete measurement samples we do have a fairly compressed representation of the signal $\bar{y}_i^J$ which contributes to a reduction of the computational effort in the subsequent optimization. Since we choose the same function spaces for signal processing and optimization the information in $\Lambda_{\bar{y}_{i,\ell}}$ can be used as an information source to select an appropriate discretization of the states and degrees of freedom in the optimization problem in a truncated multiscale basis. In addition, only the expansion coefficients $\boldsymbol{d}_{\bar{y}_{i,\ell}}$ instead of the whole sampled signal have to be exchanged and therefore fast and easy data exchange is facilitated.

5.3 Optimization problem

5.3.1 Weak problem formulation

The optimization problem (P4) should be discretized in a way that the concept of adaptivity is supported. A multiscale formulation of the process model, which permits model representation of various resolutions is obtained by a Petrov–Galerkin discretization. Following this approach the weak formulation turns out to be appropriate:

$$\min \parallel \boldsymbol{Q}(\boldsymbol{y} - \bar{\boldsymbol{y}}) \parallel^2_{H^{\beta_y}(\Omega)} + \parallel \boldsymbol{R}_v \boldsymbol{v} \parallel^2_{H^{\beta_v}(\Omega)} + \parallel \boldsymbol{R}_w \boldsymbol{w} \parallel^2_{H^{\beta_w}(\Omega)} \tag{27}$$

subject to the constraints

$$\langle \dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{w}, \boldsymbol{\nu}_x \rangle \quad = \quad \boldsymbol{0} \quad \forall\, \boldsymbol{\nu}_x \in (L_2(\Omega))^{n_x}\,, \tag{28}$$

$$\langle \boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}) - \boldsymbol{v}, \boldsymbol{\nu}_y \rangle = \boldsymbol{0} \quad \forall\, \boldsymbol{\nu}_y \in (L_2(\Omega))^{n_y}\,, \tag{29}$$
$$\langle s_j - s_{max,j}, \delta_{t_{i,j}} \rangle \leq 0 \quad \forall\, j = 1 \dots n_s,\ i = 1 \dots n_{s_j}\,, \tag{30}$$
$$\langle s_{min,j} - s_j, \delta_{t_{i,j}} \rangle \leq 0 \quad \forall\, j = 1 \dots n_s,\ i = 1 \dots n_{s_j}\,. \tag{31}$$

Here $\delta_{t_{i,j}} = \delta(\cdot - t_{i,j})$ are Dirac distributions. $\boldsymbol{\delta}_{max}$ and $\boldsymbol{\delta}_{min}$ collect the Dirac functions in (30) and (31) respectively. They span together with $\boldsymbol{\nu}_x, \boldsymbol{\nu}_y$ the spaces of test functions. In this setting the j-th inequality constraint of the catenated vector $\boldsymbol{s}$ is forced to hold at the n_{s_j} mesh points $t_{i,j} \in \Omega, i = 1 \dots n_{s_j}$. For finite dimensional approximations of s_j, this mesh can always be chosen fine enough to guarantee the inequality constraints with sufficient accuracy. Note, that the weak formulation holds on every refinement level ℓ. Hence, the processed measurements as well as the states and degrees of freedom should be indicated by the refinement index ℓ which is omitted here and subsequently to ease notation.

To further simplify the notation we define $\boldsymbol{\mu} := (\boldsymbol{x}^T, \boldsymbol{y}^T, \boldsymbol{v}^T, \boldsymbol{w}^T)^T \in U$ and $\boldsymbol{\nu} := (\boldsymbol{\nu}_x^T, \boldsymbol{\nu_y}^T, \boldsymbol{\delta}_{max}^T \boldsymbol{\delta}_{min}^T)^T \in M$, where U and M are the product spaces $U := U^x \times U^y \times U^v \times U^w$ and $M := M^x \times M^y \times M^{s_{max}} \times M^{s_{min}}$.

Before a discrete problem can be formulated the trial spaces U and test spaces M have to be chosen properly. This choice can take advantage of the richness of biorthogonal multiscale function spaces. The best choice of a suitable wavelet basis and its dual would lead to discretized equations with very sparse or ideally (almost) diagonal Jacobian matrices to ease the subsequent numerical solution step and at the same time showing good approximation properties in signal processing as well as in optimization. At this point we choose basis functions from the B-spline family and its dual as already employed for signal processing for both U and M and explicitly state that our current choice is one of many other (probably more advantageous) options. Further work has to be done to find the best suitable basis.

The hat function $\varphi^{(2)} \in C^1$ and the corresponding wavelet functions $\psi^{(2,2)}$ are selected for the discretization of $\boldsymbol{x}, \boldsymbol{v}$, and $\boldsymbol{y}$ taking advantage of their smoothness and the interpolation property in the mesh points 2^{-j}. Further, in order to obtain a simple discretization structure $\boldsymbol{w}$ is discretized with Haar wavelets $\psi^{(1,1)}$ which are constructed from $\varphi^{(1)} \in C^0$. With this (first and therefore preliminary) choice the trial spaces have the form

$$\begin{aligned} U_\Lambda &:= U^x_{\Lambda_x} \times U^y_{\Lambda_y} \times U^v_{\Lambda_v} \times U^w_{\Lambda_w} \\ &= span\{\psi^{(2,2)}_{j,k} : j,k \in \Lambda_x\} \times span\{\psi^{(2,2)}_{j,k} : j,k \in \Lambda_y\} \times \\ &\quad\ span\{\psi^{(2,2)}_{j,k} : j,k \in \Lambda_v\} \times span\{\psi^{(1,1)}_{j,k} : j,k \in \Lambda_w\} \end{aligned}$$

where the finite index sets $\Lambda_x, \Lambda_y, \Lambda_v$ and Λ_w can be chosen adaptively in a specific manner for every refinement level ℓ. As mentioned already before, we only assume that the index set includes all scaling functions $\psi_{j_0-1,k}$ and selected functions on higher scales of the multiscale expansion. At this point, it is open how to select the basis functions in the index sets in order to achieve a good approximation.

5.3.2 Projection

As it will be seen during the course of this section the flexibility offered by a biorthogonal wavelet basis can be fully utilized to obtain a suitable discretization structure. To ease the notation and to avoid an explosion of the indices, we only consider a scalar component of the vector functions $\boldsymbol{x}, \boldsymbol{y}$ etc.. The extension to the vector case is straightforward.

First, we start with an inequality constraint (30). The trial spaces, defined in the last section, together with the test space $M_\Lambda^{s_{max,j}} = span\{\delta_{t_{i,j}},\ i = 1, \dots, n_{s_j}\}$

yield

$$\boldsymbol{d}_{s_j}^T \boldsymbol{\Psi}_{\Lambda_{s_j}}^{(2,2)}(t_{i,j}) - s_{min,j}(t_{i,j}) \quad \leq \quad 0,\; i = 1, \ldots, n_{s_j}\,. \tag{32}$$

Note, that the choice of the mesh $t_{i,j}, i = 1, \ldots, n_{s_j}$ is independent of the measurement samples and reflects the locations where the inequality is exactly satisfied. The inequality constraints (31) are projected in a similar manner.

In order to minimize computational work the test space M_Λ^y of the algebraic model equation (29) is chosen such that the discretization matrix for y will be the identity matrix $\boldsymbol{I}$. At this point biorthogonality of the wavelet spaces (19) comes into play. Recall that the approximation space U_Λ^y is spanned by the wavelet basis $\{\psi_{j,k}^{(2,2)} : j,k \in \Lambda_y\}$ and that there exist biorthogonal functions $\tilde{\psi}_{j,k}^{(2,2)}$ such that $\langle \psi_{j,k}^{(2,2)}, \tilde{\psi}_{i,l}^{(2,2)} \rangle = \delta_{i,j}\delta_{k,l}$. Consequently, if $M_\Lambda^y = span\{\tilde{\psi}_{i,l}^{(2,2)} : i,l \in \Lambda_y\}$ the algebraic model equation (29) is equivalent to

$$\boldsymbol{d}_y - \langle g(x_\Lambda, u_\Lambda), \tilde{\boldsymbol{\Psi}}_{\Lambda_y} \rangle - \boldsymbol{V}_\Lambda \boldsymbol{d}_v \quad = \quad \boldsymbol{0}\,, \tag{33}$$

where $\boldsymbol{V}_\Lambda = \boldsymbol{I}$ for $\Lambda_y = \Lambda_v$.

The determination of the test space M_Λ^x for the differential equations (28) requires some care. For stability reasons the test space M_Λ^x has to capture the space spanned by $\dot{\boldsymbol{\Psi}}_{\Lambda_x}$, i.e. the trial space of the differential operator in (28). Utilizing a Petrov-Galerkin approach we choose $M_\Lambda^x := span\{\psi_{j,k}^{(1,1)} : j,k \in \Lambda_x'\}$ as trial space such that $span\{\psi_{j,k}^{(1,1)} : j,k \in \Lambda_x'\} = span\{\dot{\psi}_{j,k}^{(2,2)} j,k \in \Lambda_x\}$ where Λ_x' denotes an adapted index set of the trial functions. This choice yields

$$\boldsymbol{X}_\Lambda \boldsymbol{d}_x - \boldsymbol{W}_\Lambda \boldsymbol{d}_w - \langle f(x_\Lambda, u_\Lambda), \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)} \rangle \quad = \quad 0 \tag{34}$$

with the matrices $\boldsymbol{X}_\Lambda = \langle \dot{\boldsymbol{\Psi}}_{\Lambda_x}^{(2,2)}, \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)} \rangle$ and $\boldsymbol{W}_\Lambda = \langle \boldsymbol{\Psi}_{\Lambda_w}^{(1,1)}, \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)} \rangle$. In summary, the test spaces are

$$\begin{aligned}
M_\Lambda &= M_\Lambda^x \times M_\Lambda^y \times M_\Lambda^{s_{max}} \times M_\Lambda^{s_{min}} \\
&= span\{\psi_{j,k}^{(1,1)} : j,k \in \Lambda_x'\} \times span\{\tilde{\psi}_{j,k}^{(2,2)} : j,k \in \Lambda_y\} \times \\
&\quad\; span\{\delta_{t_{i,j}} : j = 1, \ldots, n_s,\; i = 1, \ldots, n_{s_j}\} \times \\
&\quad\; span\{\delta_{t_{i,j}} : j = 1, \ldots, n_s,\; i = 1, \ldots, n_{s_j}\}
\end{aligned}$$

supporting a stable and adaptive discretization scheme.

5.3.3 Discretization of the cost functional

For the actual computations, the cost functional in (P4) has to be substituted by a discrete norm which can be computed utilizing only the wavelet coefficients. Straightforward substitution of the trial functions into (27) yields the discretized cost functional

$$\acute{e} = (\boldsymbol{d}_y - \boldsymbol{d}_{\bar{y}})^T \acute{\boldsymbol{Q}}_\Lambda (\boldsymbol{d}_y - \boldsymbol{d}_{\bar{y}}) + \boldsymbol{d}_v^T \acute{\boldsymbol{R}}_{\Lambda_v} \boldsymbol{d}_v + \boldsymbol{d}_w^T \acute{\boldsymbol{R}}_{\Lambda_w} \boldsymbol{d}_w \tag{35}$$

with $\acute{\boldsymbol{Q}}_{\Lambda_y} = \boldsymbol{Q}^T \boldsymbol{Q} \otimes \langle \boldsymbol{\Psi}_{\Lambda_y}, \boldsymbol{\Psi}_{\Lambda_y} \rangle$ and analogously for $\acute{\boldsymbol{R}}_{\Lambda_v}, \acute{\boldsymbol{R}}_{\Lambda_v}$. For the moment the Sobolev regularity is specialized to $\beta = 0$; i.e. differentiation of the function is avoided. The matrices in (35) exhibit sparse block finger structure which is typical for wavelet discretized systems. Applying the norm equivalences (22), the block finger structure can be simplified to a block diagonal structure due to biorthogonality of the wavelets and their duals. To allow for a more general cost functional involving Sobolev regularities $\beta \geq 0$ and thus weighting smoothness properties of

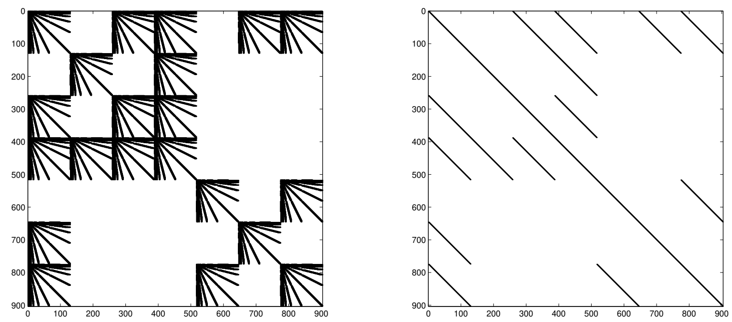

Figure 2: Structure of the discretization matrix $\acute{\boldsymbol{Q}}_{\Lambda_y}$ (left) and $\boldsymbol{Q}_{\Lambda_y}$ (right).

the functions, we use another important property of wavelet bases which extends the Riesz basis property (22) to the Sobolev space H^β. In fact, for a certain range of regularity parameters β, depending only on the regularity of $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$, one has the norm equivalence

$$\| f \|^2_{H^\beta(\Omega)} \sim \sum_{j,k} 2^{2\beta j} |\langle f, \tilde{\psi}_{j,k} \rangle|^2, \quad f \in C^\beta(\Omega)\,.$$

Thus the weighting of the smoothness of the functions in the cost functional can be easily obtained through a simple level dependent weighting of the wavelet coefficients. Therefore (35) can be replaced by

$$e = (\boldsymbol{d}_y - \boldsymbol{d}_{\bar{y}})^T \boldsymbol{Q}_\Lambda (\boldsymbol{d}_y - \boldsymbol{d}_{\bar{y}}) + \boldsymbol{d}_v^T \boldsymbol{R}_{\Lambda_v} \boldsymbol{d}_v + \boldsymbol{d}_w^T \boldsymbol{R}_{\Lambda_w} \boldsymbol{d}_w \tag{36}$$

where $\boldsymbol{Q}_{\Lambda_y} = \boldsymbol{Q} \otimes \boldsymbol{D}^\beta_{\Lambda_y}$ with $\left(\boldsymbol{D}^\beta_{\Lambda_y}\right)_{(j,k),(i,l)} = 2^{2\beta j} \delta_{j,i} \delta_{k,l}$ and analogously for $\boldsymbol{R}_{\Lambda_v}, \boldsymbol{R}_{\Lambda_w}$. Note, that $\boldsymbol{Q}_\Lambda, \boldsymbol{R}_{v_\Lambda}, \boldsymbol{R}_{w_\Lambda}$ in (36) are matrices with block diagonal structure and do not reveal the block finger structure of the direct discretization approach (35). For illustration the discretization matrices $\acute{\boldsymbol{Q}}_{\Lambda_y}$ and $\boldsymbol{Q}_{\Lambda_y}$ are shown in Figure 2 for a non-diagonal $\boldsymbol{Q}$ where the nonzero entries are represented by dots.

5.4 Solution of discretized optimization problem

The discretization employing biorthogonal multiscale approximation and test spaces results in a potentially large number of simple inequalities and a large-scale sparse nonlinear equation system (32)–(34) constraining a quadratic cost functional (36). The resulting NLP can be written in the following compact form

$$\min e(\boldsymbol{\mu}_\Lambda) \tag{37}$$

subject to the equality and inequality constraints

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{\mu}_\Lambda) &= \boldsymbol{0}\,, \\ \boldsymbol{h}(\boldsymbol{\mu}_\Lambda) &\leq \boldsymbol{0}\,. \end{aligned}$$

This large-scale NLP reveals a lot of structure which must be exploited to take full advantage of the multiscale approach. At this point, only preliminary results and explorative codes have been developed to assess some of the design issues. In the following we present a straightforward application of a general purpose NLP solver and give some remarks on the research questions to be tackled when tailoring an NLP solver to the particular problem structure.

5.4.1 Solution of the NLP with a general purpose solver

In principle, (37) can be solved with any available nonlinear programming technique. However, since the function evaluations are quite expensive we are interested in solution techniques which are known to take as few iterations as possible. Here, sequential quadratic programming (SQP) or interior point SQP methods seem to be more advantageous than other alternatives like gradient or generalized Gauss-Newton methods.

Hence, we have decided to employ a general purpose SQP solver which is suitable for large-scale problems. Our current implementation employs the SQP solver RND developed by Biegler and coworkers [46, 56, 57]. Only first order information needs to be provided in addition to the residuals since the Hessian is approximated by means of the BFGS update formula. The SQP method solves the QP subproblem

$$\min[\boldsymbol{b}_\Lambda(\boldsymbol{\mu}_\Lambda^i)^T + \frac{1}{2}\Delta\boldsymbol{\mu}_\Lambda^{i\,T}\boldsymbol{\mathcal{A}}_\Lambda(\boldsymbol{\mu}_\Lambda^i)\,\Delta\boldsymbol{\mu}_\Lambda^i] \tag{38}$$

subject to

$$\begin{aligned} \boldsymbol{g}_\Lambda(\boldsymbol{\mu}_\Lambda^i) + \boldsymbol{\mathcal{G}}_\Lambda(\boldsymbol{\mu}_\Lambda^i)\Delta\boldsymbol{\mu}_\Lambda^i &= 0 \\ \boldsymbol{h}_\Lambda(\boldsymbol{\mu}_\Lambda^i) + \boldsymbol{\mathcal{H}}_\Lambda(\boldsymbol{\mu}_\Lambda^i)\Delta\boldsymbol{\mu}_\Lambda^i &\leq 0 \end{aligned}$$

in every iteration i to generate the search direction $\Delta\boldsymbol{\mu}_\Lambda^i$. Here, $\boldsymbol{\mathcal{A}}_\Lambda$ is the Hessian of the Lagrangian and $\boldsymbol{b}_\Lambda = \nabla e_\Lambda, \boldsymbol{\mathcal{G}}_\Lambda = \nabla \boldsymbol{g}_\Lambda^T, \boldsymbol{\mathcal{H}}_\Lambda = \nabla \boldsymbol{h}_\Lambda^T$ are defined as the local derivatives of the cost functional and constraints respectively. In an application of a general purpose code, the user has to provide the function evaluations $e(\boldsymbol{\mu}_\Lambda^i), \boldsymbol{g}(\boldsymbol{\mu}_\Lambda^i), \boldsymbol{h}(\boldsymbol{\mu}_\Lambda^i)$ as well as the first order information $\boldsymbol{\mathcal{G}}(\boldsymbol{\mu}_\Lambda^i), \boldsymbol{\mathcal{H}}(\boldsymbol{\mu}_\Lambda^i), \boldsymbol{b}(\boldsymbol{\mu}_\Lambda^i)$ at the current iterate $\boldsymbol{\mu}_\Lambda^i$.

In order to make this information available, inner products of wavelet functions have to be computed. For example, consider equation (34) which is included in the set of equality constraints of the NLP (37):

$$\boldsymbol{X}_\Lambda \boldsymbol{d}_x^i - \boldsymbol{W}_\Lambda \boldsymbol{d}_w^i - \langle f(x_\Lambda^i, u_\Lambda), \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)}\rangle \quad = \quad 0\,.$$

Note that $x_\Lambda^i = \boldsymbol{d}_x^{i\,T}\boldsymbol{\Psi}_\Lambda$, $w_\Lambda^i = \boldsymbol{d}_w^{i\,T}\boldsymbol{\Psi}_\Lambda$ are part of $\boldsymbol{\mu}_\Lambda^i$. Due to the linearity, $\boldsymbol{X}_\Lambda$ and $\boldsymbol{W}_\Lambda$ can be computed prior to the optimization run with known numerical techniques [18]. However, the evaluation of the nonlinearity is much more difficult. To date, there are no methods available which allow an efficient evaluation of this expression. The objective of current research is to find methods for the evaluation of nonlinear functions of arguments given as a wavelet expansion which avoid the complexity of the finest mesh (determined by the highest resolution in Λ, briefly referred to as full single scale dimension) but require only an effort proportional to the dimension of the adapted wavelet basis.

Pragmatically, the nonlinearity is evaluated by a method which is of the complexity of the single scale dimension at this stage of our investigations. For this purpose, the function f is evaluated pointwise first leading to a representation in terms of the single scale basis functions which are dual to the generators of $\tilde{\boldsymbol{\Psi}}_{\Lambda_x'}^{(1,1)}$. Then a fast wavelet transformation is carried out to result in

$$f(\boldsymbol{d}_x^{i\,T}\boldsymbol{\Psi}_{\Lambda_x}, \boldsymbol{d}_u^T\boldsymbol{\Psi}_{\Lambda_u}) \simeq \boldsymbol{d}_f^{i\,T}\tilde{\boldsymbol{\Psi}}_{\Lambda_x'}^{(1,1)} \tag{39}$$

which is a multiscale approximation of f. Hence, the inner product simplifies to

$$\langle f(\boldsymbol{d}_x^{i\,T}\boldsymbol{\Psi}_{\Lambda_x}, \boldsymbol{d}_u^T\boldsymbol{\Psi}_{\Lambda_u}), \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)}\rangle \simeq \langle\tilde{\boldsymbol{\Psi}}_{\Lambda_x'}^{(1,1)}, \boldsymbol{\Psi}_{\Lambda_x'}^{(1,1)}\rangle \boldsymbol{d}_f^i = \boldsymbol{d}_f^i\,.$$

5.4.2 Refinement of discretization

The exactness of the solution to problem (P4) is governed by the choice of the trial space U_Λ. The key element of the proposed strategy is the successive refinement. Let Λ_ℓ denote the "old" index set which gives rise to the trial space U_{Λ_ℓ}, while the "new" trial functions, added in the refinement step, are given by the set Γ_ℓ. The optimization problem has been solved already in resolution U_{Λ_ℓ} using the SQP approach. The solution is then refined taking the additional trial functions selected by Γ_ℓ into account. Hence the resulting trial space after one refinement step is $U_{\Lambda_{\ell+1}}$ which is determined by the index set $\Lambda_{\ell+1} = \Lambda_\ell \cup \Gamma_\ell$.

The update procedure ensures that at any time the solution to the last approximated problem is available. Furthermore, this solution can also be used as an initial guess for the refined problem. Since the moment condition (25) implies decreasing wavelet coefficients with increasing scales for smooth functions, the zero augmented solution vector of resolution U_{Λ_ℓ} is a good initial guess for the refined problem (37) in resolution $U_{\Lambda_{\ell+1}}$. Hence, $\boldsymbol{\mu}^0_{\Lambda_{\ell+1}} = [\boldsymbol{\mu}^T_{\Lambda_\ell}, \mathbf{0}^T]^T$ is used as an initial guess and the changes in every entry of the old vector are expected to become smaller and smaller.

In principle, the trial functions spanning U_{Λ_ℓ} can be chosen arbitrarily. The index set Λ_ℓ might therefore contain *all* wavelet functions up to a scale J, which corresponds to an equidistant mesh. Alternatively, in an adaptive approach Λ_ℓ would consist of a strict subset of all wavelet indices up to scale J. Thus, when progressing up to higher discretization levels, the trial spaces are augmented possibly only by those subsets of the complements bases $\Psi_j, j < J$, that within some tolerance recapture most of the searched object. This provides a higher resolution for all measurements, state variables and degrees of freedom only at places where needed due to the individual behavior of these functions. The number of trial functions needed to get a certain accuracy in a nonuniform adaptive scheme is often significantly lower compared to a uniform refinement depending on the properties of the function to be approximated.

The question immediately arises how to determine Γ_ℓ in an adaptive refinement approach. Ultimately, an error estimator will be constructed on the basis of the theoretical multiscale framework such that those trial functions are selected which guarantee the best possible reduction of the approximation error. At this point, no such error estimator is available instead we employ a heuristic adaptation strategy introduced in [30]. Given the index set Λ_ℓ and the solution $\boldsymbol{\mu}_{\Lambda_\ell}$ in refinement level ℓ we search for the expansion coefficients with the largest absolute values of $\boldsymbol{\mu}_{\Lambda_\ell}$. Because of the norm equivalence property (22) these coefficients have the biggest impact on the quality of the approximation. The multiscale property together with smoothness assumptions suggest that neighbors of trial functions with large coefficients are excellent candidates for refinement. Here the wavelet functions $\psi_{j,k-1}$ and $\psi_{j,k+1}$ as well as the functions of the higher scale $\psi_{j+1,2k}$ are referred to as neighbors to the of $\psi_{j,k}$ and therefore correspond to the local neighborhood involving scale and translation. Typically not all the neighbors of the large coefficients not yet contained in Λ_ℓ are selected and added into Γ_ℓ. Rather, the additional basis functions are chosen to ensure reduction of the approximation error on the one side and a sufficiently small computation time h which is balanced against the cycle time H such that several update steps can be executed.

5.4.3 Tailored NLP algorithm

In order to establish efficiency in the algorithmical treatment the strategy for the solution of the NLP (37) presented so far is not sufficient. Rather, a tailored NLP algorithm need to be developed to take full advantage of our framework. Preliminary theoretical as well as experimental studies indicate, that there may be significant gains in efficiency if the following issues are considered during the development of a tailored NLP solver:

- exploitation of the sparse discretization structure,
- application of wavelet matrix compression techniques,
- tailored linear algebra solvers,
- efficient preconditioning of the linear systems in particular if iterative solvers are employed,
- updating of the Hessian and Jacobian, and
- the treatment of the inequality constraints.

In the following some more detailed remarks are given in order to render our current ideas a little bit more transparent.

The discretization matrices in the QP subproblem (38) exhibit a block sparse structure with sparse block matrices. The block structure reflects the structure of the cost functional and the constraints. Each block matrix reveals the particular wavelet discretization structure depending on the choice of the test and trial spaces. A proper combination may lead to more favorable patterns than those we have in our current approach. At the moment, some of the blocks of the matrices in (38) possess the wavelet typical finger structure while others are diagonal. The finger structure has $O(N \log N)$ non zero entries, where N denotes the dimension of each block. With known matrix compression techniques [19] the block finger structure can be represented by only $O(N)$ entries without significant loss of accuracy.

The underlying linear system of the QP subproblem can be solved either by factorization or by iterative techniques. Only the latter may allow for a reasonable order of complexity roughly proportional to the number of variables. Fast convergence of the iterative algorithm is a key issue. Convergence acceleration can be established through efficient preconditioning. For $J \rightarrow \infty$ diagonal preconditioners exist [8, 19] in a sense that the condition number will be independent from the number of trial functions used to discretize the system. However, more work remains to be done to develop efficient iterative solution approaches for the problem considered here.

Since we do not assume that the second order information of the model is available the Hessian is approximated using established techniques. During the course of updating the approximation to the Hessian $\boldsymbol{\mathcal{A}}_{\Lambda_{\ell+1}}$ can also be warm started augmenting the Hessian $\boldsymbol{\mathcal{A}}_{\Lambda_\ell}$ obtained from the old approximation such that positive definiteness is guaranteed. Although the elements of the Jacobian $\boldsymbol{\mathcal{G}}_{\Lambda_{\ell+1}}$ and Hessian $\boldsymbol{\mathcal{A}}_{\Lambda_{\ell+1}}$ change in each QP iteration, the parts which refer to the old trial functions given through the set Λ_ℓ stay almost constant. Therefore they only have to be updated rather than computed in each iteration step. However, this observation is based on the assumption that the mesh is already chosen fine enough such that the coarse scale components of $\boldsymbol{\mu}^i_{\Lambda_\ell}$ remain nearly unchanged when augmenting Λ_ℓ.

6 Case-studies

In this section we illustrate the principles of the multiscale approach to dynamic reconciliation using a toy example. We consider a nonisothermal continuous stirred tank reactor (CSTR) with an exothermic first order reaction. The model equations can be written in dimensionless form [66]

$$\begin{aligned}\dot{x}_1 &= -x_1 + Da(1-x_1)\exp(\frac{x_2}{1+x_2/\gamma}) + u_1\,, \\ \dot{x}_2 &= -(1+\beta)x_2 + BDa(1-x_1)\exp(\frac{x_2}{1+x_2/\gamma}) + \beta u_2\,.\end{aligned} \tag{40}$$

Here x_1, x_2 are defined as the dimensionless concentration and temperature in the reactor whereas u_1, u_2 are the associated feed concentration and cooling temperature respectively. Da is the Damkohler number, β is the dimensionless heat transfer coefficient, and B is the adiabatic temperature rise. Note, that the time in the above formulation is as well normalized. The process parameters used in this example are $(Da = 0.162, B = 14.0, \beta = 3.0, \gamma = \infty)$. In our numerical experiments, we consider only one fixed horizon of the length $[0,\ T]$ with T = 10 and project this horizon on the normalized interval $\Omega = [0,\ 1]$. We do not yet consider a moving time window. The "true" (or nominal) process state variables are simulated by integration of the model equations with $u_1 = u_2 = 0$ starting with the initial conditions $\boldsymbol{x}_0 = [0.35, 0.9]^T$. In this setting the dimensionless temperature x_2 is considered to be the measured variable. The measurements $\breve{\boldsymbol{Y}}$ which are corrupted by normally distributed uncorrelated noise are sampled on an equidistant mesh with step size 2^{-J}, $J = 7$. The goal of the optimization problem is to obtain reconciled estimates $\hat{x}_1, \hat{x}_2$ of concentration and temperature from the temperature measurements $\breve{\boldsymbol{Y}}$ for some $\boldsymbol{Q}, \boldsymbol{R}_v, \boldsymbol{R}_w$ utilizing the model (40).

The test and trial spaces are chosen as outlined in Section 5.3.1 and 5.3.2. In the first step, we interpolate the measurement samples $\breve{\boldsymbol{Y}}$ by means of B–splines $\varphi^{(2)}$ to obtain the continuous representation $\breve{y}^7$. Since this B–spline is a piecewise linear function the samples are interpolated also by a piecewise linear function. The measurement function $\breve{y}^7$ then is transformed into the multiscale representation and denoised with an appropriate chosen threshold parameter ε. The nominal temperature x_2, the measurement samples $\breve{\boldsymbol{Y}}$, the interpolated measurements $\breve{y}^7$ and the processed measurements $\bar{y}^7$ are shown in Figure 3. Occasionally sharp spikes can be seen in the denoised signal. With the use of higher order trial functions, i.e. quadratic ($m = 3$) or cubic B–splines ($m = 4$), the representation will be smoother.

In the example considered here, the optimization problem is formulated and solved starting from a coarse approximation $\ell = 0$ with three subsequent refinement steps ($\ell = 0, \ldots, 3$) employing uniform refinement: Each quantity $\bar{y}, \boldsymbol{x}, y, v, \boldsymbol{w}$ in the optimization problem (P4) is discretized by wavelet functions of the index set $\Lambda_\ell^{J_\ell}$ with $J_\ell = 2, \ldots, 5$. This choice represents all trial functions of $S_1 \oplus W_1 \oplus \ldots \oplus W_{J_\ell - 1}$ and is equivalent to an equidistant mesh with step size 2^{-J_ℓ}. To ease notation we ommit the index J_ℓ in $\Lambda_\ell^{J_\ell}$. Figure 4 reveals the processed measurement function $\bar{y}_{\Lambda_\ell} = \boldsymbol{d}_{\bar{y}_\ell}^T \boldsymbol{\Psi}_{\Lambda_{\bar{y}_\ell}}$ which are used as input to the optimization problem. The reconciled states $\hat{x}_{i,\Lambda_\ell} = \boldsymbol{d}_{\hat{x}_{i,\ell}}^T \boldsymbol{\Psi}_{\Lambda_{\hat{x}_{i,\ell}}}$ of the optimization based on the choice of $\boldsymbol{Q} = \boldsymbol{R}_v = \boldsymbol{R}_w = \boldsymbol{I}$ are shown in the Figures 5-6. Note that starting from a very coarse approximation where each quantity has only 5 degrees of freedom the approximation accuracy is steadily refined by halving the mesh size in each refinement step.

During the course of refinement the change of the shape of the reconciled quantities is quite remarkable. By contrast, the coefficients $\boldsymbol{d}_{\hat{x}_{i,\ell}}$ in the expansion stay almost constant. The coefficients of $\boldsymbol{d}_{\hat{x}_{i,\ell}}$ which correspond to the first nine wavelet trial functions (i.e. the functions of the trial space $S_1 \oplus W_1 \oplus W_2$) are shown in Figure 7 during refinement of the approximation. Recall that in the case of uniform refinement the wavelets in Λ_ℓ, i.e. the basis function for the coarse approximation, are also contained in $\Lambda_{\ell+k}, k \geq 0$. The wavelet functions that span W_2 (the last four wavelet functions in the graph) are not contained in the initial coarse approximation ($\ell = 0$) and therefore take the value zero. As it can be seen, the changes in the

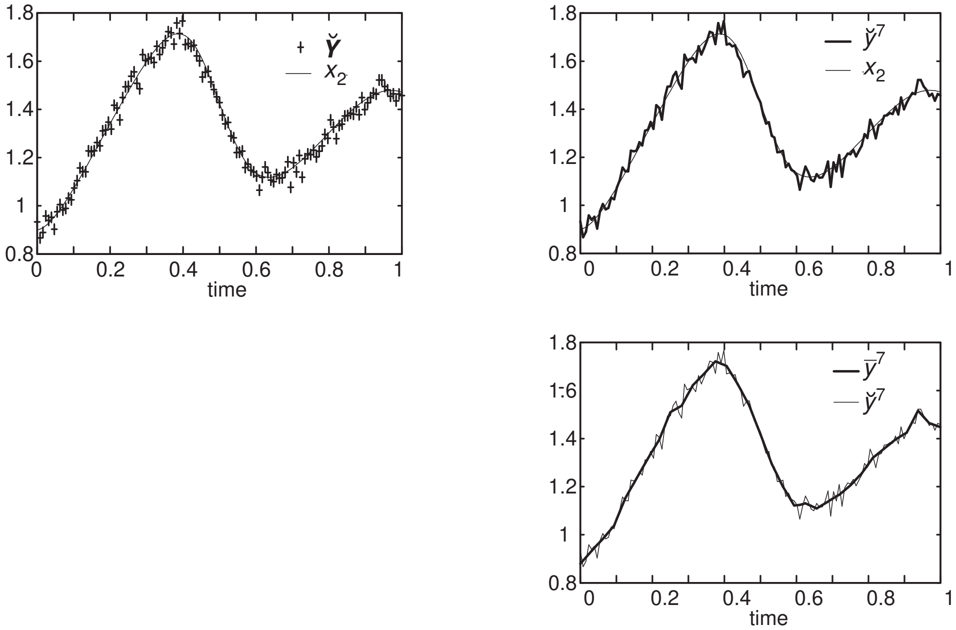

Figure 3: Measurements, original and processed measurements.

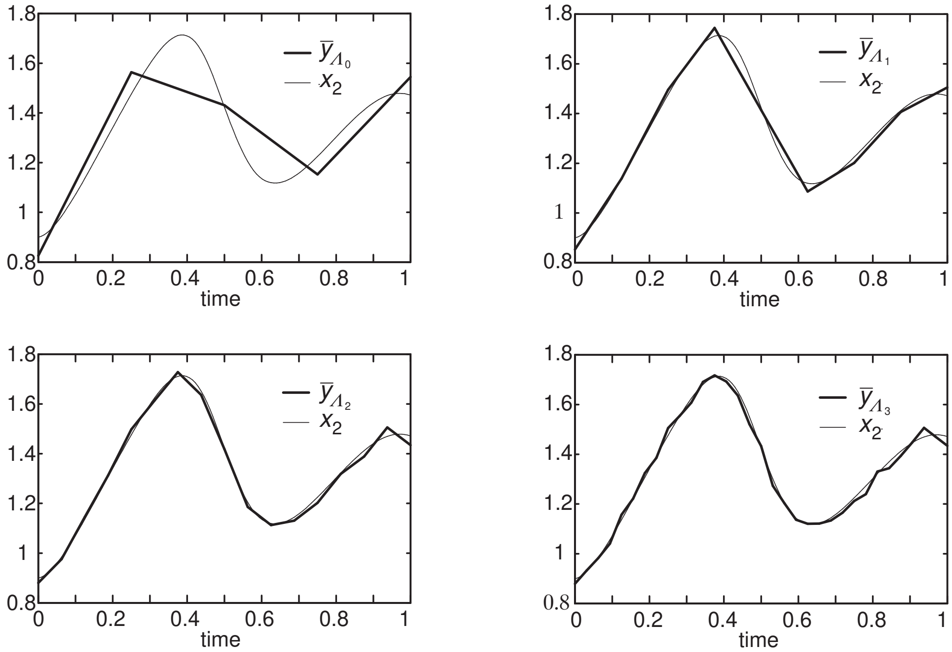

Figure 4: Processed measurement function on various scales.

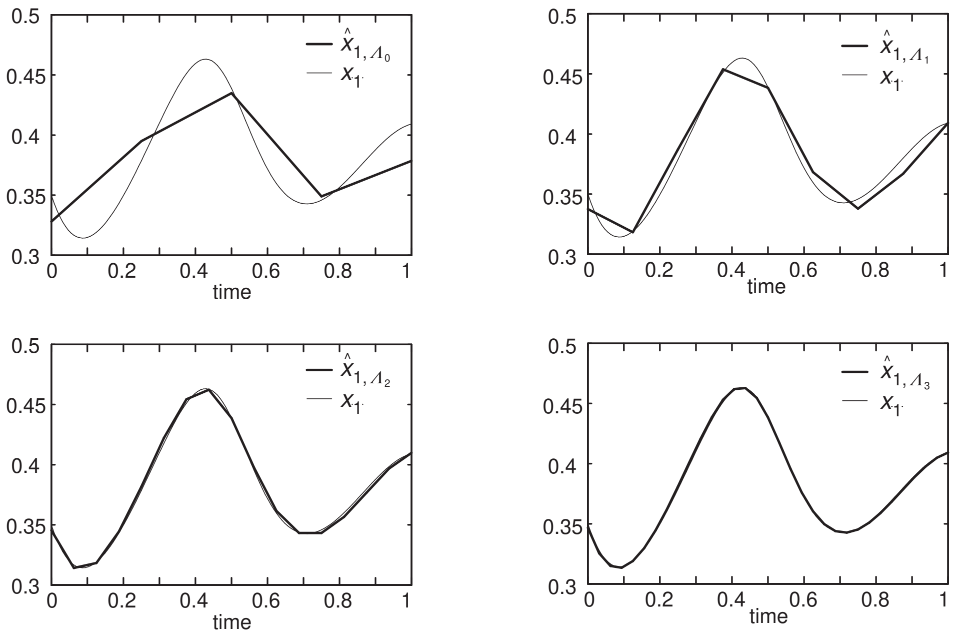

Figure 5: Refinements and original of state x_1.

coefficients are hardly noticeable. Moreover, the wavelet coefficients $\boldsymbol{d}_{\hat{x}_{i,\ell}}$ decrease in magnitude with increasing scale. As mentioned before, this is due to the moment condition and the smoothness of the approximated function. Again, levelwise refinement is only one strategy offered by the discretization scheme presented. It leads to a large number of degrees of freedom after few refinement steps. On the other hand, there is no need to increase the resolution uniformly. In this example, however, the levelwise refinement strategy is very favorable due to the nature of the approximated process quantities and a nonuniform adaptive strategy does not result in significant reduction of the effort. In other examples we have studied, a significant reduction can be achieved by nonuniform refinement as outlined in Section 5.4.2.

7 Conclusions

We have sketched a novel treatment of dynamic data reconciliation problems. In the first part, it suggests a purely deterministic formulation of the problem based on the theory of inverse problems. At this point, the formulation is still partly heuristic and based on theoretical evidence gained from the understanding of general linear inverse problems. A theoretical analysis of the estimation problem for the linear as well as the nonlinear data reconciliation problem is subject to future work in order to show the appropriateness of the suggested formulation. In particular, an assessment of the estimation quality is essential.

The second part suggests a new framework to the numerical treatment of dynamic data reconciliation. It is based on multiscale techniques for the discretization of the dynamic optimization problem to result in a nonlinear programming problem as well as for the solution of the underlying system of equations. The presentation is limited to an outline of the general ideas underlying the multiscale framework. The technical realization of the conceptional ideas is still at the very early stage of research. Hence, all the techniques presented at this point must be viewed as being of preliminary nature. These first results demonstrate, however, convincingly

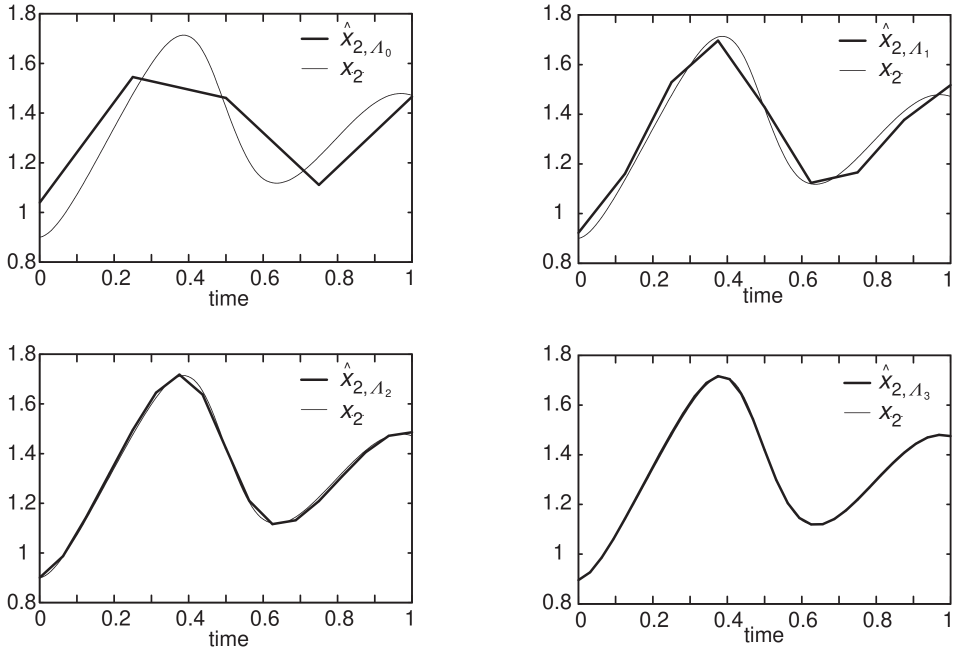

Figure 6: Refinements and original of state x_2.

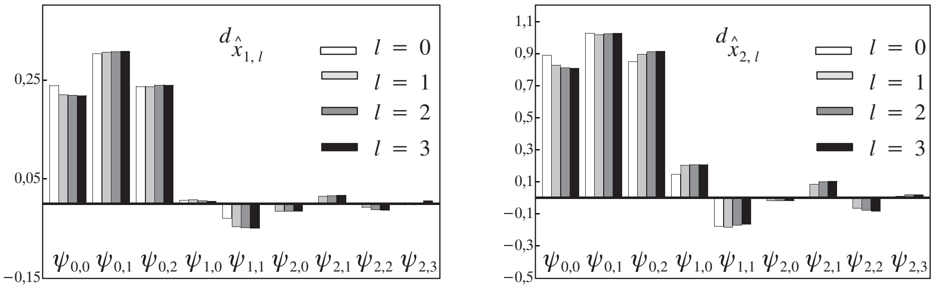

Figure 7: Wavelet coefficients during refinements of $\hat{x}_1$ and $\hat{x}_2$.

the potential of the approach to finally result in a highly efficient numerical technique which is able to provide the best possible estimate for given computational resources.

Obviously, much work remains to be done. We will concentrate on the one hand on a theoretical analysis of the linear reconciliation problem first. We are interested in bridging the gap between the theory of inverse problems and observer theory, both employing a deterministic setting for the estimation. The numerical research will focus on improved multiscale discretization schemes which can be integrated with signal processing and on the development of tailored methods for the solution of the mathematical programming problem.

ACKNOWLEDGEMENT

The authors acknowledge the contribution by A. M'hamdi. The work presented here was supported by the "Deutsche Forschungsgemeinschaft" under grant no. MA1188/6 which is also gratefully acknowledged.

References

[1] O. Abel, A. Helbig, and W. Marquardt. Optimization approaches to control-integrated design of industrial batch reactors. NATO ASI on Nonlinear Model Based Process Control. This volume.

[2] J.S. Albuquerque and L.T. Biegler. Decomposition algorithmus for on–line estimation with nonlinear DAE models. *Comput. Chem. Eng.*, 19:1031–1039, 1995.

[3] J.S. Albuquerque and L.T. Biegler. Data reconcilliation and gross–error detection for dynamic systems. *AIChE J.*, 42(10):2841–2856, 1996.

[4] J.S. Albuquerque, V. Gopal, G.H. Stauss, L.T. Biegler, and B.E. Ydstie. Interior point SQP strategies for structured process optimization problems. *Comput. Chem. Eng.*, 21(Suppl.):853–859, 1997.

[5] M.J. Bagajewicz and Q. Jiang. Integral approach to plant linear dynamic reconciliation. *AIChE J.*, 43(10):2546–2558, 1997.

[6] B. Bakshi and G. Stephanopoulos. Representation of process trends: III multi–scale extraction of trends from process data. *Comput. Chem. Eng.*, 18(4), 1994.

[7] B. Bakshi and G. Stephanopoulos. Compression of chemical process data by functional approximation and feature extraction. *AIChE J.*, 42(2):477–492, 1996.

[8] J.H. Bramble and J.E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations for elliptic problems. *Math. Comp.*, 50:1–17, 1988.

[9] A.G. Bruce, D.L. Donoho, H.Y. Gao, and R.D. Martin. Smoothing and robust wavelet analysis. In *Proceedings on Computational Statistics. 11th Symposium, Berlin*, volume 4, pages 531–547. COMPSTAT, 1994.

[10] A. Chambolle, R. DeVore, N.Y. Lee, and J. Lucier. Nonlinear wavelet image processing: Varational problems, compression, and noise removal through wavelet shrinkage. 1996.

[11] C.K. Chui. *An Introduction to Wavelets.* Academic Press, Boston, 1992.

[12] H. Cox. On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Auto. Cont.*, pages 5–12, January 1964.

[13] C.M. Crowe. Data reconciliation – progress and challage. *J. Proc. Cont.*, 6:89–98, 1996.

[14] S.A. Dadebo and K.B. McAuley. Dynamic optimization of constrained chemical engineering problems using dynamic programming. *Comput. Chem. Eng.*, 19(5):513–525, 1995.

[15] W. Dahmen. Wavelet and Multiscale Methods for Operator Equations. *Acta Numerica*, pages 55–228, 1997.

[16] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline-wavelets on the interval – stability and moment conditions. IGPM–Report 129, RWTH Aachen, 1996.

[17] W. Dahmen, A. Kunoth, and K. Urban. Wavelets in Numerical Analysis and their Quantitative Properties. In A.L̃e Méhauté, C. Rabut, and L.L. Schumaker, editors, *Surface Fitting and Multiresolution Methods*, pages 93–130. Vanderbilt University Press, 1997.

[18] W. Dahmen and C.A. Micchelli. Using the refinement equation for evaluating integrals of wavelets. *SIAM J. Numer. Anal.*, 30(2):507–537, 1993.

[19] W. Dahmen, S. Prössdorf, and R. Schneider. Multiscale methods for pseudo-differential equations on smooth manifolds. In C.K. Chui, L. Montefusco, and L. Puccio, editors, *Proceedings of the International Conference on Wavelets: Theory, Algorithms, and Applications*, pages 385–424. Academic Press, San Diego, 1994.

[20] M. Darouach and M. Zasadzinski. Data reconciliation in generalized linear dynamic systems. *AIChE J.*, 37:193–201, 1991.

[21] I. Daubechies. *Ten Lectures on Wavelets.* Society for Industrial and Applied Math., Philadelphia, 1992.

[22] C. de Boor. *A Practical Guide to Splines.* Springer-Verlag, New York, 1978.

[23] R.A. DeVore, B. Jawerth, and B. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Information Theory*, 38(2):719–746, 1992.

[24] D. Donoho and I. Johnstone. Ideal spatial adaption by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[25] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Assoc.*, 90:301–369, 1995.

[26] D.R.Kuehn and H.Davidson. Computer control. *Chemical Engineering Progress*, 57(6):44–47, 1961.

[27] H.W. Engl. Regularization methods for the stable solution of inverse problems. *Surv. Math. Ind.*, 3:71–143, 1993.

[28] H.W. Engl, M.H. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers, 1996.

[29] W.F. Feehery and P.I. Barton. Dynamic simulation and optimization with inequality path constraints. *Comput. Chem. Eng.*, 20(Suppl.):S707–S712, 1996.

[30] J. Fröhlich and K. Schneider. An adaptive wavelet–vaguelette algorithm for the solution of nonlinear PDEs. *Preprint SC. ZIB*, (95-28), 1997.

[31] G.F. Froment and K.B. Bischoff. *Chemical Reactor Analysis and Design.* John Wiley & Sons, 1990.

[32] G.A.Almasy. Principles of dynamic balancing. *AIChE J.*, 36:1321–1330, 1990.

[33] A. Gelb. *Applied Optimal Estimation.* MIT Press, 1974.

[34] J. Gertler and G.A. Almasy. Balance calculation through dynamic system modeling. *Automatica*, 9:79–85, 1973.

[35] D.M. Himmelblau and T.W. Karjala. Rectification of data in a dynamic process using artificial neural networks. *Comput. Chem. Eng.*, 20:805–812, 1996.

[36] S.-S. Jang, B. Joseph, and H. Mukal. Comparison of two approaches to on-line parameter and state estimation of nonlinear systems. *Ind. Eng. Chem. Proc. Des. Dev.*, 25:809–814, 1986.

[37] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press Inc., 1970.

[38] T. Johansen. On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica*, 33(3):441–446, 1997.

[39] T.W. Karjala and D.M. Himmelblau. Dynamic data rectification by recurrent neural networks vs. traditional methods. *AIChE J.*, 40(11):1865–1875, 1994.

[40] I.-W. Kim, M.J. Liebmann, and T.F.Edgar. A sequential error-in variables method for non-linear dynamic systems. *Comput. Chem. Eng.*, 15:663–670, 1991.

[41] I.-W. Kim, S.W. Park, and T.F.Edgar. Data reconciliation for input-output models in nonlinear dynamic systems. *Korean J. Chem. Engng*, 13:211–215, 1996.

[42] J.L.A. van Koolen. Plant operation in the future. *Comput. Chem. Eng.*, 18:S477–S481, 1994.

[43] J. Lee, Y. Chikkula, Z. Yu, and J. Kantor. Improving computational efficiency of model predictive control algorithm using wavelet transformation. *Int. J. Control*, 61(4):859–883, 1995.

[44] M.J. Liebmann, T.F. Edgar, and L.S. Lasdon. Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Comput. Chem. Eng.*, 16:963–986, 1992.

[45] L.C. Lin and C.-C. Kuo. Signal extrapolation in noisy data with wavelet representation. In *Proceedings of the SPIE*, volume 2028, pages 156–167. The International Society for Optical Engineering, 1993.

[46] J.S. Logsdon and L.T. Biegler. Decomposition strategies for large-scale dynamic optimization problems. *Chem. Eng. Sci.*, 47(4):851–864, 1992.

[47] W. Marquardt. Numerical methods for the simulation of differential-algebraic process models. In R. Berber, editor, *Methods of Model-Based Control*, NATO-ASI Series, pages 42–79. Kluwer Press, 1995.

[48] K.F. McBrayer, T.F Edgar, and L.S. Lasdon. Bias detection and estimation in dynamic data reconciliation. *J. Proc. Cont.*, 5, 1995.

[49] A. M'hamdi, A. Helbig, O. Abel, and W. Marquardt. Newton-type receding horizon control and state estimation - a case study. In J.J. Gertler, J.B. Cruz, and M. Peshkin, editors, *Preprints of the 13th World Congress of IFAC*, volume M, pages 121–126, San Francisco, 30 June–5 July 1996.

[50] H. Michalska and D.Q. Mayne. Moving horizon observers. In *IFAC Symposium NOLCOS'92*, Bordeaux, France, 1992.

[51] H. Michalska and D.Q. Mayne. Moving horizon observers and observer–based control. *IEEE Trans. Aut. Contr. AC-40*, pages 995–1006, 1995.

[52] K.R. Muske and T.F. Edgar. Nonlinear state estimation. In M.A. Henson and D.E. Seborg, editors, *Nonlinear Process Control*. Prentice Hall, 1997.

[53] K.R. Muske and J.B. Rawlings. Nonlinear moving horizon state estimation. In R. Berber, editor, *Methods of Model-Based Control*, NATO-ASI Series, pages 349–365. Kluwer Press, 1995.

[54] Y. Ramamurthi, P.B. Sistu, and B.W. Bequette. Control-relevant dynamic data reconciliation and parameter estimation. *Comput. Chem. Eng.*, 17(1):41–59, 1993.

[55] D. Robertson, J.H. Lee, and J.B. Rawlings. A moving horizon-based approach for least-squares estimation. *AIChE J.*, 42(8):2209–2223, 1996.

[56] C. Schmid and L.T. Biegler. Quadratic programming methods for reduced hessian SQP. *Comput. Chem. Eng.*, 18(9):817–832, 1994.

[57] C. Schmid, J. Nocedal, and L.T. Biegler. A reduced hessian method for large–scale constrained optimization. *SIAM J. Optimization*, 5(2):314, 1995.

[58] P.B. Sistu, R.S. Gopinath, and J.B. Rawlings. Computional issues in nonlinear predictive control. *Comput. Chem. Eng.*, 17:361, 1993.

[59] G. Stephanopoulos, M. Dyer, and O. Karsligil. Multi–scale aspects in linear and nonlinear estimation and control. NATO ASI on Nonlinear Model Based Process Control. This volume.

[60] G. Stephanopoulos, M. Dyer, and O. Karsligil. Multi–scale modeling, estimation and control of processing systems. *Comput. Chem. Eng.*, 21(Suppl.):S797–S803, 1997.

[61] O. von Stryk. *Numerische Lösung optimaler Steuerungsprobleme: Diskretisierung, Parameteroptimierung und Berechnung der adjungierten Variablen.* Reihe 8: Meß-, Steuerungs- und Regelungstechnik Nr. 441. VDI-Verlag, Düsseldorf, 1995.

[62] P. Tanartkit and L.T. Biegler. Stable decomposition for dynamic optimization. *Ind. Eng. Chem. Res.*, 34(4):1253–1266, 1995.

[63] D. Tieu, W.R. Cluett, and A. Penlidis. A comparison of collocation methods for solving dynamic optimization problems. *Comput. Chem. Eng.*, 19(4):375–381, 1995.

[64] M.L. Tyler and M. Morari. Stability of constrained moving horizon estimation schemes. In R. Berber, editor, *Nonlinear Model Based Process Control*, NATO-ASI Series. Kluwer Press, 1998.

[65] J. Unger, A. Kröner, and W. Marquardt. Structural analysis of differential-algebraic equation systems - theory and applications. *Comput. Chem. Eng.*, 19(8):867–882, 1995.

[66] A. Uppal, W. H. Ray, and A. B. Poore. On the dynamic behaviour of continuous stirred tank reactors. *Chem. Eng. Sci.*, 29:967–985, 1974.

[67] P. de Vallière and D. Bonvin. Application of estimation techniques to batch reactors - III: Modelling refinements which improve the quality of state and parameter estimation. *Comput. Chem. Eng.*, 14(7):799–808, 1990.

[68] V.S. Vassiliadis, R.W.H. Sargent, and C.C. Pantelides. Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints. *Ind. Eng. Chem. Res.*, 33(9):2111–2122, 1994.

[69] V.S. Vassiliadis, R.W.H. Sargent, and C.C. Pantelides. Solution of a class of multistage dynamic optimization problems. 2. Problems with path constraints. *Ind. Eng. Chem. Res.*, 33(9):2123–2133, 1994.

[70] G.H. Weiss, J.A. Romagnoli, and K.A. Islam. Data reconciliation - an industrial case study. *Comput. Chem. Eng.*, 20:1441–1449, 1996.

[71] S.J. Wright. *Primal–Dual Interior–Point Methods.* SIAM, 1996.