# ON THE REGULARIZATION OF DYNAMIC DATA RECONCILIATION PROBLEMS [1]

**T. Binder ** L. Blank * W. Dahmen* W. Marquardt****

** Institut für Geometrie und Praktische Mathematik,*
*** Lehrstuhl für Prozesstechnik, RWTH Aachen. Germany*

to appear in
**Journal of Process Control**,
March, 2001

**Abstract:** Dynamic data reconciliation problems are discussed from the perspective of the mathematical theory of ill-posed inverse problems. Regularization is of crucial importance to obtain satisfactory estimation quality of the reconciled variables. Usually, some penalty is added to the least-squares objective to achieve a well-posed problem. However, appropriate discretization schemes of the time-continuous problem act themselves as regularization, reducing the need of problem modification. Based on this property, we suggest to refine successively the discretization of the continuous problem starting from a coarse grid, to find a suitable regularization which renders a good compromise between (measurement) data and regularization error in the estimate. In particular, our experience supports the conjecture, that non-equidistant discretization grids offer advantages over uniform grids.

**Keywords:** dynamic optimization, adaptive discretization, wavelets, dynamic data reconciliation, state estimation, input estimation, inverse problems

## 1. INTRODUCTION

Usually process measurements do not reveal complete and correct information about the process state. This is due to the limited number of process quantities accessible to measurement instruments, as well as unavoidable measurement errors caused by imperfect instrumentation and signal processing. Measurement information can be upgraded, in principle, by means of incorporating process knowledge in some mathematical model. Stationary Data Reconciliation (SDR) provides a set of reconciled measurements and an estimate of unmeasured process quantities based on a steady-state process model. Since the pioneering work of Kuehn and Davidson (1961) SDR has been studied extensively and has also been applied successfully to large scale processes in industry (Romagnoli and Sanchez, 1999; Narasimhan and Jordache, 2000). Dynamic Data Reconciliation (DDR) is a natural extension of SDR to dynamic processes and provides a set of reconciled estimates on some time horizon. DDR combines state, parameter and unknown input estimation, and is typically formulated as a dynamic optimization problem restricted by a process model. Interest in DDR was initiated by Almasy (1990), who described a data reconciliation technique for linear balance equations. Later on, several authors suggested to solve the problem using nonlinear programming methods in combination with moving horizon techniques (see e.g. Liebmann *et al.* (1992); Ramamurthi *et al.* (1993); Albuquerque and Biegler (1996)). DDR might be considered as a generalization of moving horizon state estimation schemes, where some stability and convergence results are available (Michalska and Mayne, 1995; Robertson *et al.*, 1996; Rao and Rawlings, 2000). Although DDR has received much attention recently, it seems that the combination of problem formulation and its associated

[1] This work has been supported by the "Deutsche Forschungsgemeinschaft" under grant no. MA1188/6.
Corresponding author: marquardt@lfpt.rwth-aachen.de

numerical solution still leaves quite some room for further research to improve understanding and to develop high performance numerical solution strategies.

In this paper, we want to deepen the understanding of such a combination. In particular the properties of dynamic data reconciliation are illuminated, for a simple linear example on a fixed horizon, from the viewpoint of the mathematical theory of *inverse problems* (Kirsch, 1996). An application of this theory to the estimation of reaction rates has recently been presented by Mhamdi and Marquardt (1999) where the problem is analysed in the frequency domain after a suitable model transformation. DDR problems are in general ill-posed and require some form of regularization. Regularization stabilizes the numerical solution and aims at simultaneously minimizing the data and regularization errors in the estimate. Regularization in the context of SDR has been recently addressed by Kelly (1998). In this approach, regularization is used to improve numerical stability when solving ill-conditioned algebraic equation systems. Such ill-conditioned systems might arise when unmeasured variables are present, which are hardly observable through the measured variables (see Romagnoli and Sanchez (1999), Narasimhan and Jordache (2000), for observability in SDR). Although the underlying conceptual idea of regularization is the same for SDR and DDR problems, they differ in one fundamental aspect. SDR is a problem of finite dimension whereas DDR problems involve a differential operator and are therefore of infinite dimension. With the exception of Binder *et al.* (1998), who suggested a general DDR problem formulation based on the theory of ill-posed problems, regularization is not explicitly addressed in literature on DDR but always applied implicitly through either discretization of the continuous problem or by starting off with a discrete formulation, since any appropriate discretization of a certain operator itself typically acts as regularization (Kirsch, 1996; Engl *et al.*, 1996). For example, Liebmann *et al.* (1992) and Ramamurthi *et al.* (1993) use a coarse discretization of the input variables to obtain smooth estimates as well as to resolve stability problems of the estimation algorithm. The stepsize of the discretization scheme in the continuous case or the sampling time in the discrete case are of crucial importance to the estimation results and should be chosen judiciously to achieve a good compromise between data and regularization error. However, it is not obvious how to determine a suitable discretization mesh. Here, we suggest a method which automatically generates sequences of discretization grids based on upgrades of the approximation spaces until a good compromise between data and regularization error in the estimates is accomplished. Uniform as well as non-uniform problem adapted upgrades of the approximation spaces are employed. In particular, it is emphasized that non-uniform discretization might lead to improved estimates. The method is built upon a deterministic framework and thus does not depend on statistical assumptions. The results presented apply in principle to a number of suitable discretization methods. However, the discretization employed in this work is based on wavelets, since they allow tailored problem approximations of adjustable accuracy in a natural way.

The article is organized as follows. In Section 2 we introduce the problem considered in the paper and elaborate on its properties by means of an introductory example. In particular, we argue that the problem belongs to the class of inverse problems. Regularization approaches and the associated parameter selection strategies are discussed in Section 3. Here, special emphasis is put on regularization by projection. In Section 4 we introduce the concrete choice of a discretization scheme needed to approximate the continuously formulated problem. The findings are then illustrated using a simple but realistic example (Section 5).

## 2. INVERSE PROBLEMS

### 2.1 *The Problem*

DDR problems are often formulated as optimization problems restricted to a dynamic model. In this paper we consider a very simple linear scalar problem, since it provides insight and exhibits, at the same time, the essential features of an inverse problem. It is assumed that the time domain of interest has been transformed onto the unity interval $I = [0, 1]$. The DDR problem with a least-squares cost functionl has the form

$$\min_{z_0, w} \| \, z^{\delta}(t) - z(t) \, \| \tag{1}$$

$$\text{s.t.} \quad \dot{z}(t) = az(t) + u(t) + w(t), \quad z(t_0) = z_0 \tag{2}$$

The constraint (2) represents the process model where $a$ is a time invariant known parameter; $z$ and $z_0$ denote the unknown state and its initial condition; $u$, $w$ are known and unknown inputs, respectively. $z^{\delta}(t)$, $t \in I$ denotes a measurement function which is based upon a set of discrete measurement samples $z^{\delta}(t_i)$ of the state $z$ usually taken from the process and typically corrupted by measurement error indicated by the superscript $\delta$. Note, that no penalty term is present in (1).

### 2.2 *The Properties*

The estimation problem belongs to the class of inverse problems, that always searches for causes of observed effects (Engl *et al.*, 1996). Typically, inverse problems which involve an integral operator are not well-posed in the sense of Hadamard's definition (Kirsch, 1996). Besides the existence of a unique solution, well-posedness requires continuous dependence of the estimate on the measurement data. Suppose, we measure the nominal solution of (2), $\tilde{z}$, for some given initial condition $\tilde{z}_0$ and nominal inputs $\tilde{u}, \tilde{w}$ such that

$$\tilde{w} = \dot{\tilde{z}} - a\tilde{z} - \tilde{u}.$$

Furthermore suppose, that there is alternatively a corrupted measurement function with fixed parameters $\delta$ and and error frequency $n/\delta$ given by

$$z^\delta(t) = \tilde{z}(t) + \delta \sin \frac{n}{\delta} t.$$

Since $z^\delta$ is continuously differentiable, the solution to problem (1),(2) is

$$w^\delta = \dot{z}^\delta - a z^\delta - \tilde{u}.$$

It is well known that differentiation of $z^\delta$ can cause arbitrarily large errors in the estimation of $w$ for arbitrary small data errors, since

$$\| \tilde{z} - z^\delta \|_\infty = \delta$$

but

$$\| \dot{\tilde{z}} - \dot{z}^\delta \|_\infty = n$$

and

$$\| \tilde{w} - w^\delta \|_\infty = \| n \cos \frac{n}{\delta} t + a\delta \sin \frac{n}{\delta} t \|_\infty .$$

Hence, the errors in $\dot{z}^\delta, w^\delta$ are governed not only by the presumably small error level $\delta$ but also by the probably high frequency $\frac{n}{\delta}$. The error in the estimate clearly does not depend continuously on the data and therefore problem (1),(2) is ill-posed in the sense of Hadamard. Note, that this phenomenon is not restricted to (2), but similarly occurs for more general differential equation constraints. It is not due to the special choice of the infinity norm, but holds for general $L_p$ norms, although continuous dependence could be enforced by measuring the data error in a stronger norm containing first or higher order derivatives. However, in practical situations one often prefers to call an error small if the error in the function values is small and therefore some $L_p$ norm is the natural choice. The question immediately arises of what can be done to stabilize the solution, i.e. to reduce the high sensitivity of the estimate with respect to small corruptions of the measurement data..

## 3. REGULARIZATION

Problem (1), (2) can be generalized for subsequent analysis by replacing (2) by the linear operator equation

$$\boldsymbol{T} y = z,$$

where $y := (w, z_0)$ and $\boldsymbol{T}$ is a bounded linear operator with domain $D(\boldsymbol{T}) \subset Y$, $y \in Y, z \in Z$. In particular, for the process model (2) the linear operator $\boldsymbol{T}$ is given by

$$\boldsymbol{T} : \ (w, z_0) \to z$$

where

$$z(t) = e^{at} z_0 + \int_{t_0}^{t} e^{a(t-\tau)} (u(\tau) + w(\tau)) d\tau.$$

Usually the corrupted measurement $z^\delta$ does not belong to the range $R(\boldsymbol{T})$ since typically no $y = (w, z_0)$ exists such that

$$\boldsymbol{T} y = z^\delta. \tag{3}$$

In this case the solution of (3) is understood to be the *best approximate solution* denoted by $y^\dagger$ which satisfies (3) in a least-squares sense and at the same time is of minimal norm. The associated operator $\boldsymbol{T}^\dagger$ which maps $z^\delta$ to $y^\dagger$, i. e.

$$\boldsymbol{T}^\dagger z^\delta = y^\dagger, \tag{4}$$

is the generalized inverse of $\boldsymbol{T}$ and reflects problem (1),(2). The inverse $\boldsymbol{T}^\dagger$ is generally unbounded, so that stability in the sense of Hadamard is not guaranteed. Hence, even small errors in the measurements $z^\delta$ can lead to arbitrarily large errors in the estimate $y^\dagger$. Regularization can be used to stabilize the best approximate solution by introducing a family of bounded transformations $\boldsymbol{T}_r^\dagger$ in (4) such that

$$\lim_{r \to 0} \boldsymbol{T}_r^\dagger z^\delta = \boldsymbol{T}^\dagger z^\delta, \quad \forall z^\delta, \tag{5}$$

where $r$ is called the regularization parameter. Let $z$ denote the exact data. Given an error bound on the measurement data

$$\| z - z^\delta \| \leq \delta$$

and

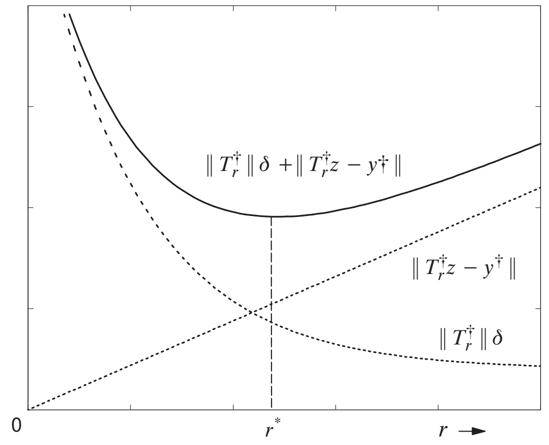


Fig. 1. Data and regularization errors (cf. eq. (6).

$$y_r^\delta := \boldsymbol{T}_r^\dagger z^\delta,$$

the following fundamental error statement holds (Kirsch, 1996; Engl *et al.*, 1996):

$$\| y_r^\delta - y^\dagger \| \leq \| \boldsymbol{T}_r^\dagger \| \delta + \| \boldsymbol{T}_r^\dagger z - y^\dagger \| . \quad (6)$$

The bound of the error between the best approximate and estimated solutions splits into two parts. The first term on the right refers to the (measurement) *data error*, while the second denotes the *regularization* or *approximation error* which would be obtained for exact data. The data and regularization errors might be associated with *variance* and *bias* as defined in a *stochastic* setting. [2] The data error largely corresponds to the variance of the estimate but may contain some bias also. In contrast, the regularization error can almost solely be attributed to the bias of the estimate. Since we do not use here any stochastic model and since the interpretation of the terms as either variance or bias would neither in a stochastic setting be straightforward, we prefer to refer to these terms as data and regularization error, respectively. Figure 1 illustrates the general behaviour of the data and regularization errors. By property (5) an increase of $r$ bounds $\boldsymbol{T}_r^\dagger$ and improves stability such that for a constant error level $\delta$ the data error is decreased. On the other hand, the increasing difference between $\boldsymbol{T}_r^\dagger$ and $\boldsymbol{T}^\dagger$ causes a growing approximation error. Opposite behavior is observed for decreasing $r$. Naturally, we are interested in obtaining sharp estimates for $\| y_r^\delta - y^\dagger \|$. Therefore we seek an $r^\star$ which minimizes the sum of data and regularization errors in (6). Parameter selection strategies to find $r^\star$ or a suitable approximation are discussed next.

[2] Regularization explicitly allows bias in the estimate to reduce variance and therefore is conceptually different from the Kalman-Bucy filter, which yields only for linear zero mean processes and white noise with Gaussian statistics bias free estimates with minimal variance (Gelb, 1978). In the general case, when nonlinearities, non-zero mean disturbances or colored noise with unknown statistics are involved, bias free estimates are not guaranteed anymore.

### 3.1 *Parameter selection*

Parameter selection strategies are classified in *a-priori* and *a-posteriori* methods (Engl *et al.*, 1996). The former refers to the situation where $r$ is chosen before the regularized solution is computed. This is in contrast to the latter where the selection of $r$ is based on the results of the numerical regularization algorithm. Optimal values $r^\star$ might be accessible a-priori for a variety of $\boldsymbol{T}_r^\dagger$ for *known* noise level $\delta$ and *known* smoothness of the solution $y^\dagger$, although possibly being numerically expensive. Smoothness assumptions made on $y^\dagger$ are generally of the form

$$y^\dagger = (\boldsymbol{T}^*\boldsymbol{T})^p v, \quad p > 0, \ \| v \| \leq \rho,$$

where $\boldsymbol{T}^*$ is the adjoint of $\boldsymbol{T}$. The power $p$ can be roughly interpreted as the number of weak derivatives bounded in some norm while $\rho$ is a rough measure of the magnitude of the $p$-th generalized derivative $v$ of $y^\dagger$ (Neumaier, 1998). Unfortunately, a-priori strategies rely on explicit knowledge of $\delta, p, \rho$. While the data error level $\delta$ might be accessible in some applications, the degree of smoothness $p$ and the bound $\rho$ of the exact solution $y^\dagger$ are in general at least difficult to estimate. Therefore, a-priori parameter selection strategies are clearly not applicable for most practical problems but are important for the theoretical examination on a specific regularization method.

A-posteriori strategies, which do not require the knowledge of $p$ and $\rho$ for selecting $r$ are more powerful in practice. The best-known a-posteriori strategy is based on *Morozov's discrepancy principle* (Morozov, 1984). The discrepancy principle expresses the fact that the best case to ask for is a residual in the order of $\delta$ since smaller residuals are not reliable anyway due to the noise of magnitude $\delta$ in the data. Moreover, a smaller regularization parameter destabilizes the solution. Therefore the regularization parameter $r^\star$ is chosen such that the residual (or *discrepancy*) $\| \boldsymbol{T} y_{r^\star}^\delta - z^\delta \|$ equals the *assumed* data error bound $\hat{\delta}$ including a safeguard factor $\tau \geq 1$, i.e.

$$\| \boldsymbol{T} y_{r^\star}^\delta - z^\delta \| = \tau \hat{\delta}. \quad (7)$$

Typically $r^\star$ is not determined by solving (7) directly. In practice, one selects a decreasing sequence $\{r_k\}, r_k \rightarrow 0$ and sequentially evaluates (7) until $\| \boldsymbol{T} y_{r^\star}^\delta - z^\delta \| \approx \tau \hat{\delta}$.

In the absence of reliable knowledge on the data error level $\delta$ it is necessary to consider alternative (a-posteriori) parameter choice rules that avoid knowledge of the noise level $\delta$ and determine a suitable regularization parameter $r$ on the basis of the actual performance of the regulariza-

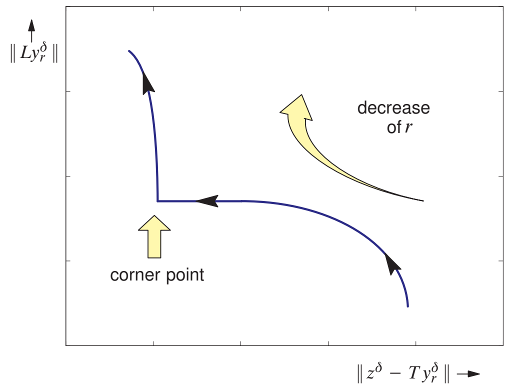


Fig. 2. Typical L-curve shape.

tion method under consideration. Such a criterion is given by the L-curve method (Hansen, 1998) which is based on an inspection of the residual norm of the computed approximation by relating it to the (semi-) norm of the approximation itself. This is done by plotting $\| \boldsymbol{L}y_r^\delta \|$ versus $\| z^\delta - \boldsymbol{T}y_r^\delta \|$ in log-log scale for a large range of a decreasing sequence $\{r_k\}, r_k \to 0$. Generally, $\boldsymbol{L}$ is either the identity or a derivative operator. The curve typically has an L-shape with a distinct corner point. It is shown qualitatively in Figure 2. There is evidence and substantial practical experience that the corner point reflects a good compromise between data and regularization error in the estimate. Note, that the residual norm in the L-curve criterion as well as in Morozov's discrepancy principle is the same such that the latter is simply a vertical line at $\tau\delta$ in the L-curve plot.

### 3.2 *Direct and indirect methods*

A wide variety of regularization methods are available which are typically classified as *direct* and *indirect* approaches. For example, *Tikhonov regularization* (Tikhonov *et al.*, 1996) is a direct method frequently applied in engineering which minimizes the functional

$$\| \boldsymbol{T}y - z \|^2 + \alpha \| y - y_{ref} \|^2 \qquad (8)$$

where $\alpha$ is the regularization parameter, i.e. $r \sim \alpha$. $y_{ref}$ denotes an a priori guess of the solution and ideally is $y^\dagger$. This approach is also known as ridge regression (Golub and van Loan, 1996) in the context of stabilization of finite dimensional statistical problems. The regularization operator of (8) is given by

$$\boldsymbol{T}_\alpha^\dagger = (\alpha \boldsymbol{I} + \boldsymbol{T}^*\boldsymbol{T})^{-1}\boldsymbol{T}^* \qquad (9)$$

with $\boldsymbol{T}^*$ being the adjoint of $\boldsymbol{T}$. The eigenvalues of $\boldsymbol{T}$ tend to zero for ill-posed problems but the eigenvalues of $\alpha\boldsymbol{I} + \boldsymbol{T}^*\boldsymbol{T}$ are bounded away from zero for $\alpha > 0$. Moreover, it can be shown for compact operators $\boldsymbol{T}$ that $\| \boldsymbol{T}_\alpha^\dagger \| \leq \frac{1}{2\sqrt{\alpha}}$ (Kirsch, 1996). Often the $L_2$-norm is chosen as penalty in the Tikhonov functional (8) but it is also common to consider stronger norms (or only semi-norms). In practice, it is nontrivial to find a suitable guess $y_{ref}$ and a wrong guess might lead to unsatisfactory results since the solution is biased.

Regularization is also achieved through indirect methods which rely on the self-regularization property of the underlying solution method. Many iterative methods exhibit such a self-regularization property in the sense that early termination of the iterative process has a regularizing effect (Kirsch, 1996). A famous example is *Landwebers iteration formula* (Engl *et al.*, 1996)

$$\boldsymbol{y}^{m+1} = (\boldsymbol{I} - \beta\boldsymbol{T}^*\boldsymbol{T})\boldsymbol{y}^m + \beta\boldsymbol{T}^* z^\delta,$$

where $\beta$ denotes a damping factor. The iteration index $m$ is the (hidden) regularization parameter $r \sim \frac{1}{m}$. As $m \to \infty$ the regularization error decreases while the data error increases. The stopping criterion is of crucial importance for the actual performance of the method. The iteration scheme can be interpreted as the steepest decent algorithm applied to the residual norm $\| \boldsymbol{T}y - z \|$. Other iterative methods yielding faster convergence rates such as the CG-method, can be applied as well (Kirsch, 1996; Hanke, 1995). *Regularization by projection*, upon which we will focus next in more detail, also reveals such a self regularization property (Plato and Vainikko, 1990).

### 3.3 *Regularization by Projection*

Regularization can also be achieved indirectly by projection of (3) into finite dimensional subspaces $Z_0 \subset Z_1 \subset \ldots \subset Z$, $Y_0 \subset Y_1 \subset \ldots \subset Y$. Let $\boldsymbol{Q}_\ell : Z \to Z_\ell$, $\boldsymbol{P}_\ell : Y \to Y_\ell$ be suitable projectors onto $Z_\ell$, $Y_\ell$, then

$$\boldsymbol{Q}_\ell \boldsymbol{T} \boldsymbol{P}_\ell y = \boldsymbol{Q}_\ell z$$

reduces to a Petrov-Galerkin scheme which represents the discretized problem. The index $\ell$ might be associated with a stepsize $h \sim \frac{1}{\ell}$ for uniform subspaces which depends on the interval length and projector type. For $\ell \to \infty$ the stepsize becomes increasingly smaller, and hence $h \to 0$. Let us define

$$y_\ell^\dagger := \boldsymbol{T}_\ell^\dagger z_\ell$$

as the least-squares solution of minimal norm to

$$\boldsymbol{T}_\ell y_\ell = z_\ell$$

where

$$\boldsymbol{T}_\ell := \boldsymbol{Q}_\ell \boldsymbol{T} \boldsymbol{P}_\ell, \quad z_\ell := \boldsymbol{Q}_\ell z, \quad y_\ell := \boldsymbol{P}_\ell y.$$

Then, given the error bound $\| \boldsymbol{Q}_\ell (z - z^\delta) \| \leq \delta$, it can be shown (Engl *et al.*, 1996) that

$$\begin{aligned} \| y_\ell^\delta - y^\dagger \| &\leq \| \boldsymbol{T}_\ell^\dagger \| \delta + \| \boldsymbol{T}_\ell^\dagger z_\ell - y^\dagger \| \\ &\leq \frac{\delta}{\sigma_\ell} + \| y_\ell^\dagger - y^\dagger \|, \end{aligned} \tag{10}$$

where $\sigma_\ell$ is the smallest singular value of $\boldsymbol{T}_\ell$. Equation (10) is of the same structure as (6). The first term on the right hand side refers to the data error while the second is the regularization error. For projection methods no regularization parameter is introduced explicitly. However, there is a *hidden* regularization parameter associated to the finite dimensional approximation spaces which affects the smallest singular value of $\boldsymbol{T}_\ell$. The subscript $\ell$ takes the role of the regularization parameter $r \sim \frac{1}{\ell}$. Due to the spectral properties of continuous operators $\sigma_\ell \to 0$ as $\ell \to \infty$. This suggests to keep the problem discretization rather coarse to bound the data error. On the other hand a coarse problem discretization increases the regularization error. Again, the aforementioned compromise is needed. The question how to choose appropriate finite dimensional spaces to obtain a good compromise is of central importance. Clearly there are infinitely many choices to select the approximation spaces appropriately such that a rigorous search is impractical. Obviously, a procedure is needed which varies the approximation spaces in a systematic way to find the discretization with a suitably low bound in (10). Such a procedure can be obtained for example by application of a refinement approach suggested previously by the authors (Binder *et al.*, 1998). The principle idea is to successively *upgrade* the approximation spaces starting from a *first coarse guess* ($\ell = 0$). In each upgrade *new* basis functions are added such that the approximation $\ell + 1$ is *refined* either globally or locally. This concept yields sequences of finite dimensional optimization problems of increasing complexity which have to be solved in each refinement step $\ell$. The stopping criterion is of particular importance since the successive refinement acts more and more like differentiation as the smallest singular value decreases. This limiting behavior is known to exist also for Luenberger observers where increasingly high observer gains move the smallest eigenvalue of the observer towards $-\infty$ to result in a differentiator (Zeitz, 1983). Once a suitable refinement strategy is chosen, the a-priori and a-posteriori parameter selection strategies introduced in Section 3.1 might be applied to find the discretization with a suitably low bound in (10) (for a general discussion on their convergence properties see Plato and Vainikko (1990)). In this paper the stopping criterion applied is based on the aforementioned L-curve method since it does not rely on assumptions on the noise level $\delta$. Each approximation $\ell$ is represented by one point in the L-curve. As soon as a corner point is detected, the refinement can be terminated.

### 3.4 *Combined Regularization*

Equation (10) shows that for noisy data and severely ill-posed problems the dimension of the subspace has to be rather low to keep the total error small, since for those problems the smallest $\sigma_\ell(\boldsymbol{T}_\ell)$ decreases rapidly as $\ell$ increases. To be able to use larger dimensions the projection method might be combined with an additional regularization method. If Tikhonov's method is chosen then the functional

$$\| \boldsymbol{T}_\ell y_\ell - z_\ell \|^2 + \alpha \| y_\ell - y_{\ell,ref} \|^2 \tag{11}$$

has to be minimized. The solution of (11) is given by

$$\boldsymbol{y}_{\alpha,\ell}^\delta = \boldsymbol{T}_{\alpha,\ell}^\dagger \boldsymbol{z}_\ell^\delta,$$

where $\boldsymbol{T}_{\alpha,\ell}^\dagger = (\alpha \boldsymbol{I}_\ell + \boldsymbol{T}_\ell^* \boldsymbol{T}_\ell)^{-1} \boldsymbol{T}_\ell^*$. The generalized inverse $\boldsymbol{T}_\alpha^\dagger$ in (9) is bounded by $\| \boldsymbol{T}_\alpha^\dagger \| \leq \frac{1}{2\sqrt{\alpha}}$ which must also hold for the projected system independent of $\ell$, i. e.

$$\| \boldsymbol{T}_{\alpha,\ell}^\dagger \| \leq \frac{1}{2\sqrt{\alpha}}. \tag{12}$$

On the other hand we know that

$$\| \boldsymbol{T}_{\alpha,\ell}^\dagger \| \leq \| \boldsymbol{T}_{0,\ell}^\dagger \| \leq \frac{1}{\sigma_\ell} \quad \forall\, \alpha \geq 0. \tag{13}$$

Both inequalities (12), (13) have to be satiesfied simultaneously such that

$$\| \boldsymbol{T}_{\alpha,\ell}^\dagger \| \leq \min\left(\frac{1}{2\sqrt{\alpha}}, \frac{1}{\sigma_\ell}\right).$$

Therefore both, $\ell$ and $\alpha$, bound $\| \boldsymbol{T}_{\alpha,\ell}^\dagger \|$. If $\ell$ is chosen such that $\sigma_\ell > 2\sqrt{\alpha}$ then the discretization is the leading regularizer and the solution is hardly affected by the choice of $\alpha$. Conversely when $2\sqrt{\alpha} > \sigma_\ell$, $\boldsymbol{T}_{\alpha,\ell}$ is stabilized mainly by $\alpha$ and the discretization can be refined arbitrarily provided that $\alpha$ is sufficiently large. One should keep in mind that $\alpha$ is always acting uniformly on the solution in the domain $I = [0, 1]$. On the contrary, a specific discretization might resolve the problem in some areas in greater detail. Therefore $\alpha$ can be considered as a *global* regularizer whereas $\ell$ can regularize *globally* as well as *locally* depending on the concrete choice of the approximation spaces. If $\alpha$ dominates, then the solution is affected globally *even if* the discretization resolves the problem locally rather than globally.

## 4. DISCRETIZATION

So far, the concrete choice of a discretization scheme has not been addressed. The use of wavelets in conjunction with ill-posed problems has been recently explored by a number of authors (Donoho (1995); Dicken and Maass (1996); Freeden and Schneider (1998)). Donoho as well as Dicken *et al.* elaborate on the efficiency of the convergence rates using adaptive strategies based on wavelet-vaguelette decompositions. This special class of discretization enables a particularly fast algorithm which can, however, only be applied to a limited class of operators. The advantages of locally adapted regularization parameters for three-dimensional satellite gradiometry problems are recognized by Freeden and Schneider. In line with some of the presented arguments, we suggest the use of wavelets to span the approximation spaces since they allow tailored problem approximations to resolve global as well as local trends. In particular, two methods are presented which refine the problem approximation uniformly as well as non-uniformly. First, some fundamental properties of wavelets are sketched (see e.g. Dahmen (1997) for a recent survey and more details).

### 4.1 *Wavelets*

The common starting point for the construction of *wavelets* on an *interval* (see e.g. Dahmen *et al.* (1999)), which is the relevant setting in the present context, are multiresolution sequences of nested spaces

$$S_{j_0} \subset \ldots \subset S_j \subset S_{j+1} \subset \ldots$$

whose union is dense in $L_2(I)$. Here $j_0$ stands for some coarsest discretization scale. The spaces $S_j$ are usually spanned by compactly supported functions $\varphi_{j,k}$, $j \in \mathcal{I}_j$, like B-splines, whose supports have lengths proportional to $2^{-j}$. Thus $2^{-j}$ can be viewed as the *mesh size* for the space $S_j$. Given such a multiresolution sequence, one way of updating an approximation $f_j \in S_j$ to a function $f$ is to add some *detail information*

$$f_j + w_j = f_{j+1}$$

where $w_j$ belongs to some *complement* $W_j$ of the space $S_j$ in the next finer space, i.e., $S_{j+1} = S_j \oplus W_j$. With a basis $\{\psi_{j,k}; k \in \mathcal{J}_j\}$ for $W_j$ the detail $w_j$ can be written as

$$w_j = \sum_{k \in \mathcal{J}_j} d_{j,k}\psi_{j,k}.$$

Note that $\psi_{j,k}$ is a linear combination of the $\varphi_{j+1,l}$. With successive decomposition of $S_{j+1}$ into

$$S_{j+1} = S_{j_0} \oplus_{k=j_0}^{j} W_k$$

the function $f_{j+1}$ is associated with a *multiscale representation*

$$f_{j+1} = \sum_{k \in \mathcal{I}_{j_0}} d_{j_0-1,k}\varphi_{j_0,k} + \sum_{l=j_0}^{j} \sum_{k \in \mathcal{J}_l} d_{l,k}\psi_{l,k}.$$

Under certain assumptions on the complements $W_j$ the complement basis functions $\psi_{j,k}$ are called *wavelets*. The whole collection

$$\{\varphi_{j_0,k} : k \in \mathcal{I}_{j_0}\} \cup \{\psi_{j,k} : k \in \mathcal{J}_j, j = j_0, j_{0+1}, \ldots\}$$

is a *basis* for $L_2(I)$. Wavelets are constructed such that the Euclidean norm of the wavelet coefficients of $f$ is always proportional to the $L_2$ norm of $f$. Thus discarding small coefficients $d_{j,k}$ will cause only small changes in approximate representations of $f$. Moreover, it can be shown that the values $|d_{j,k}|$ will be quite small when $f$ is smooth on the support of $\tilde{\psi}_{j,k}$. Therefore, functions that exhibit only strong variations locally, can be approximated very accurately by linear combinations of relatively few wavelet basis functions corresponding to the *significant* coefficients $d_{j,k}$. Denoting by $\Lambda$ the set of the corresponding indices $(j,k)$, the finite basis

$$\Psi_\Lambda := \{\psi_{j,k} : (j,k) \in \Lambda\}$$

is expected to lead to a particularly *adapted* approximation to $f$. We refer to $\Psi_\Lambda$ as a *sparse* wavelet basis in contrast to a *full* wavelet basis $\Psi_{\Lambda^J}$, corresponding to index sets

$$\Lambda^J := \{(j,k) : j_0 - 1 \le j \le J-1, k \in \mathcal{J}_j\}$$

and obtained by cutting off all frequencies above scale $J$. The former may be associated with a possibly highly *nonuniform adapted mesh* while the latter induces a *uniform* discretization of mesh size $2^{-J}$.

### 4.2 *Uniform and adaptive discretization*

In our realizations the approximation spaces $Z_\ell, Y_\ell$ introduced above are spanned by a wavelet basis. Specifically, the approximation for the states $z$ and measurements $z^\delta$ are generated by piecewise linear wavelets. The inputs $w, u$ are expanded in terms of piecewise constant (Haar) wavelets. Starting from a uniform coarse grid $\ell = 0$ the upgrades of the approximation spaces can be either accomplished uniformly or non-uniformly. For uniform adaptation, in each upgrade the complete next scale is added to the approximation, e.g. the basis is $\Psi^\ell_{\Lambda^J}$ with $J := j_0 + \ell$. Alternatively, a sparse basis can also be applied since it might lead to more economical discretizations. Here our own method has been employed (Binder *et al.*, 2000) to generate sparse discretization grids. The basic idea followed in this approach is to successively identify trial functions with large directional sensitivities and thus use them for an improved problem approximation. As a simplification we keep in this work the spaces of $z, z^\delta$ constant at a very fine resolution and vary only the approximation spaces for the input $w$. Based on the computed estimates at the refinement stage $\ell$, candidate sets of potential basis functions for the upcoming upgrade $\ell+1$ are specified which are not yet used in the current approximation. The adaptive refinement of $w$ is based on a sensitivity evaluation of the Lagrange function with respect to the functions contained in the candidate sets since it can be expected that large absolute values indicate substantial impact on the reduction of the cost function. Basis functions which are associated to large sensitivities are used for an upgraded approximation at refinement stage $\ell+1$. As candidate sets, simply the local neighbors are taken with respect to scale and translation of the current approximation.

## 5. EXAMPLE

Our example is built upon calorimetric estimation (Schuler and Schmidt, 1992) applied to a nonisothermal CSTR where a first order reaction $A \to B$ takes place. The nonlinear model of the reactor and a temperature controller is simulated rigorously assuming a disturbance in the feed concentration and step changes in the inlet temperature. Given noisy temperature measurements as well as the transients of inlet and cooling temperature $T_0$ and $T_{cool}$, the task is to find reconciled values of the heat of reaction $Q_{rea}$ without knowledge of the reaction kinetics. The estimation model (2) is here the energy balance equation

$$\dot{T} = \frac{q}{V}(T_0 - T) + Q_{rea} - \frac{UA}{(\rho c_p V)}(T - T_{cool}) \quad (14)$$

with flow rate $q = 10\ cm^3/s$, vessel volume $V = 10^3\ cm^3$, heat transfer coefficient $U = 5\ 10^{-4}\ cal/(cm^2 sK)$, heat transfer area $10\ cm^2$, liquid density $\rho = 0.001\ g/cm^3$, and heat capacity $c_p = 1\ cal/(gK)$. The nominal solution of (14) denoted by $T_{nom}$ is corrupted by uniformly distributed noise of magnitude 1 to be used as simulated perturbed measurements $T_{noise}$. $T_{nom}$, $T_{noise}$ are shown in Figure 3. The reconciliation problem is of least-squares form and given by

$$\min_{T(0), Q_{rea}} \int_{t=0s}^{512s} (T_{noise} - T)^2 d\tau \quad (15)$$

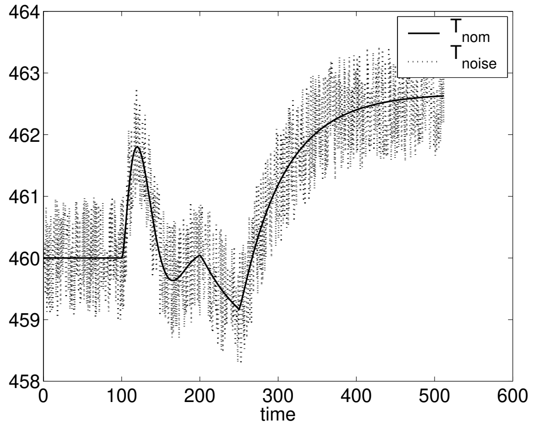


Fig. 3. Nominal and corrupted temperatures

subject to (14). The formulation matches the general problem given by (1), (2) and is discretized as outlined before. The approximation spaces for $T_{nom}, T_{noise}$ are kept constant at scale $J = 9$ while we employ uniform as well as sparse wavelet bases $\Psi^\ell_{\Lambda^J}$ and $\Psi^\ell_\Lambda$ to approximate $Q_{rea}$. The refinement is started from an initial uniform mesh $\ell = 0$ at the approximation scale $J = 1$ and is then gradually upgraded. Estimates obtained employing a uniform basis at the refinement steps $\ell \in \{3, 4, 5\}$ are shown in Figure 4, which corresponds to approximation scale $J \in \{4, 5, 6\}$. The observed differentiation effect due to a decreasing smallest singular value $\sigma_\ell$ can be seen clearly. The $L_2$-error of $Q_{rea,estim}$ and $Q_{rea,nom}$ vs. refinement index $\ell$ are shown for all upgrades in Figure 5. However, in practice, there is no information about the nominal values or the noise level $\delta$. The L-curve in Figure 6 is used instead to detect a good compromise between data and regularization error. The first derivative of the estimate is taken as semi-norm $\| \boldsymbol{L} Q_{rea,estim} \|_{L_2}$. As residual norm we use $\| T_{noise,J} - \boldsymbol{T}_J Q^\ell_{rea,estim} \|_{L_2}$ where the measurements and continuous operator are evaluated at high resolution (scale $J = 9$). Since the example is a simulated experiment where the noise level is available, we can evaluate $\delta = \| T_{nom} - T_{noise} \|_{L_2}$ to obtain a sharp $L_2$-estimate. This value is included in the L-curve plot for $\tau = 1$ with Morozov's method at $\hat{\delta} = \delta = 0.47$ as a straight line. According to Morozov's discrepancy principle $\tau\delta$ is the value which should be achieved as residual error norm by any suitable regularization strategy. As one can see in Figure 6, the refinement level $\ell = 4$ associated to the corner point of the uniform discretization coincides quite nicely with Morozov's criterion and with the best achievable estimates taken from Figure 5.

The $L_2$ error and the L-curve are shown as well for non-uniform refinement in Figure 5 and Figure 6. In each refinement step only one trial function with the largest sensitivity has been added, but this can easily be enlarged to obtain faster conver-

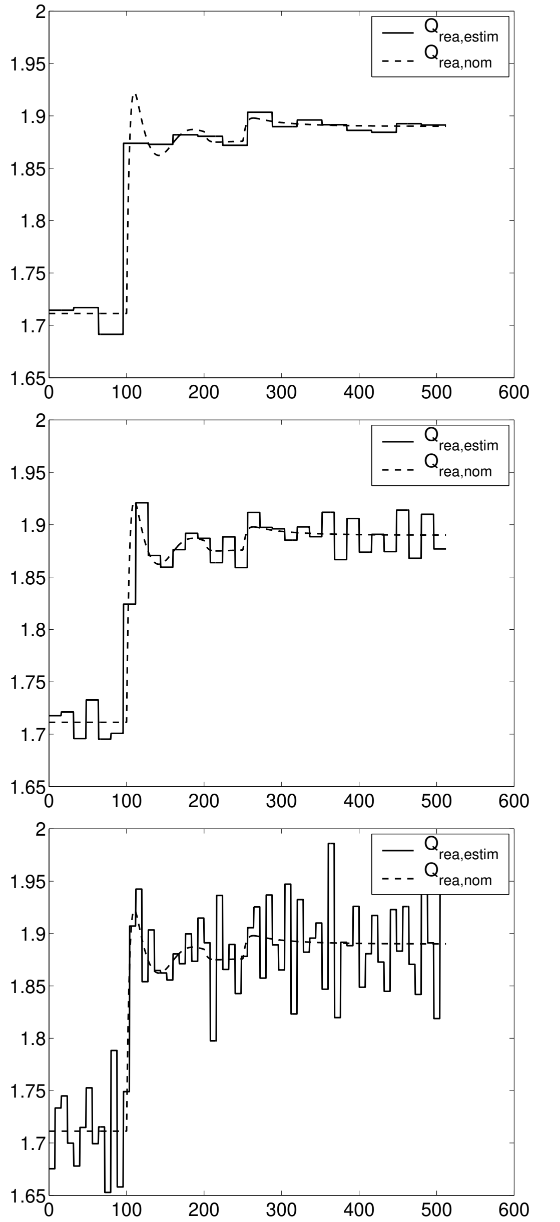


Fig. 4. $Q_{rea,estim}$ on uniform mesh: $\ell = 3,\ J = 4$ (top); $\ell = 4,\ J = 5$ (middle); $\ell = 5,\ J = 6$ (bottom).

gence rates. Again, the corner point of the L-curve coincides well with the minimal $L_2$ error. However, a range of discretizations corresponds to the points on the curve near its corner. The $L_2$ error reveals that an approximation quality is achieved in the non-uniform case which can not be obtained using uniform discretization. This is also seen by comparing Figure 4 to Figure 7 where $Q^{\ell}_{rea,estim}$ is shown at upgrade step $\ell = 6$ and $\ell = 8$. In such adaptive discretizations the sharp transient can be resolved using small stepsizes while the flat areas are discretized with larger stepsizes and therefore keeps the noise in the estimates low. On the contrary, uniform discretization either resolves the sharp transient at the cost of an increased noise in the flat areas or does not resolve the sharp transient, while the noise in the flat areas are kept

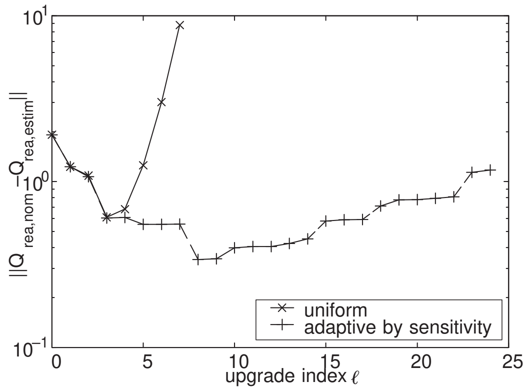


Fig. 5. $L_2$ error of the estimates.

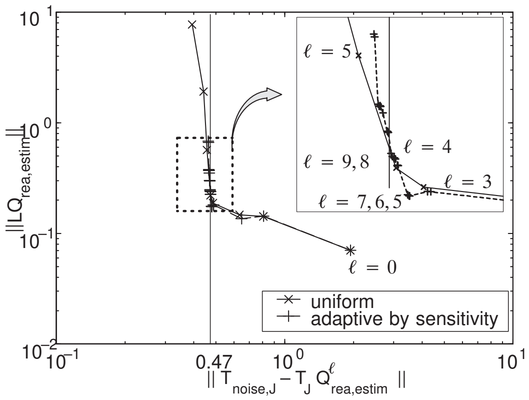


Fig. 6. L-curve and Morozov's discrepancy value.

low. Note, that for the non-uniform refinement in Figure 7 we can not attribute a uniform approximation scale $J$, rather we would have to refer to the complete index set $\Lambda_\ell$ to characterise the discretization.

Next the behaviour is examined if regularization by projection is applied in combination with Tikhonov's method. The cost function (15) is augmented by the penalty

$$\alpha \int\limits_{t=0s}^{512s} (Q_{rea,estim} - Q_{rea,ref})^2 d\tau.$$

$Q_{rea,ref}$ is typically not known but should be guessed as well as possible since a wrong guess might introduce a large bias. As a guess a constant profile $Q_{rea,ref} = 1.8\frac{K}{s}$ is assumed. The reconciliation problem is solved for each combination of $\alpha \in \{0, 5, 25, 50, 75, 100, 250, 500\}$ and uniform meshes at resolution $J \in \{1, 2, 3, 4, 5, 6, 7\}$. Each point in Figure 8 reflects one combination of $\alpha$ and $J$. The discretizations scales $J = 1, \ldots, 7$ starting from $\ell = 0$ form one curve for a specific choice of $\alpha$. The curve for $\alpha = 0$ is the same as in Figure 6 for uniform discretization. At $\alpha \sim 50$ a qualitative change in the behaviour occurs. For $\alpha \gtrsim 50$ the Tikhonov regularization is dominating and an improved approximation does not affect the stability of $\boldsymbol{T}^{\dagger}_{n,\alpha}$. For large values of $\alpha$ there

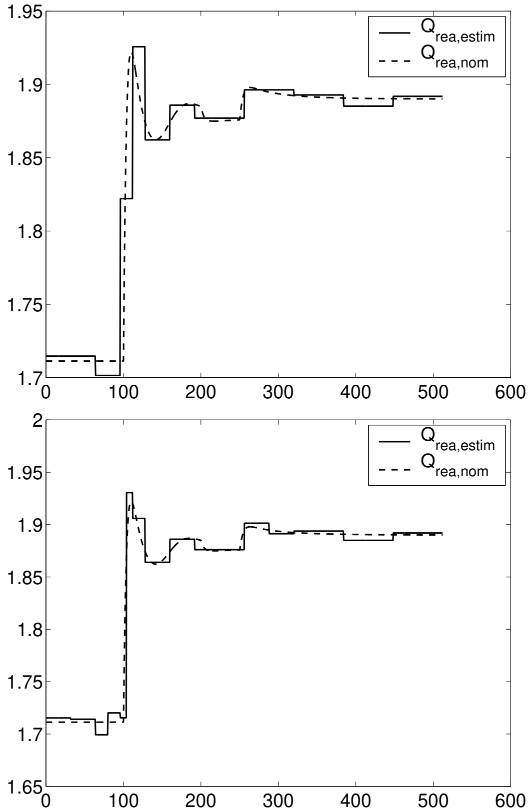


Fig. 7. $Q_{rea,estim}$ on adaptive mesh: $\ell = 6$ (top); $\ell = 8$ (bottom).

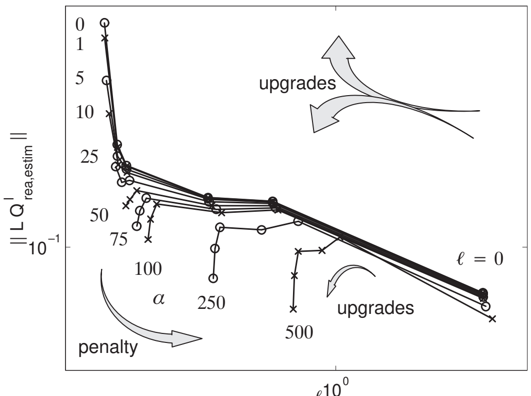


Fig. 8. Tikhonov regularization for different uniform discretizations.

is clearly an over-regularization and $Q_{rea,estim}$ is dominated by the choice of $Q_{rea,ref}$ due to the penalty $\alpha$. However, for $\alpha \lesssim 50$ the discretization destabilizes the estimates and there is an increase of the noise in the estimates. These curves reveal the expected L-curve behaviour for the sequence of refined discretizations. Note, that if the combinations $J$ and $\alpha$ are virtually redrawn such that all $\{\alpha_k\}, k = 1, \ldots 8$ for a selected $\ell$ are connected, then the resulting curves would also reveal L-shape (not shown).

The transient $Q_{rea,estim}$ obtained for $\alpha = 25,\ J = 6$ and $\alpha = 50,\ J = 7$, which are close to the value suggested by Morozov's discrepancy principle, are

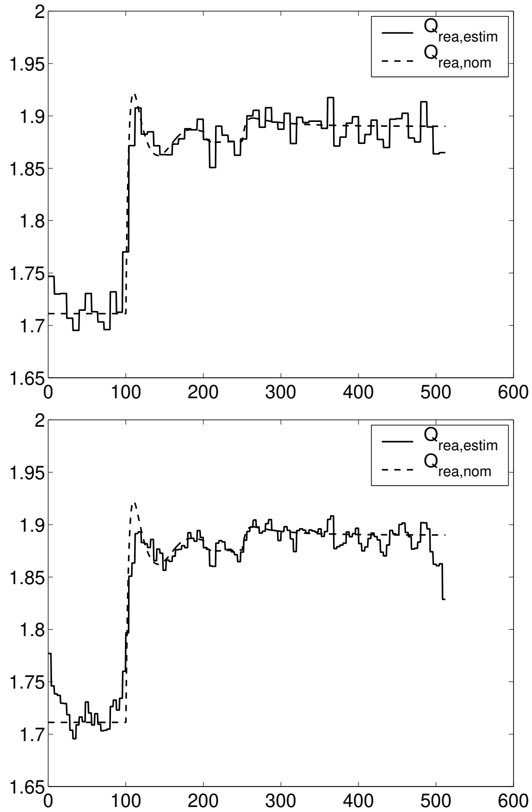


Fig. 9. $Q_{rea,estim}$ on uniform mesh: $\alpha = 25$ $J = 6$ (top); $\alpha = 50$ $J = 7$ (bottom).

shown in Figure 9. The noise in the estimates is relatively small but still larger compared to the estimates shown in Figure 7. As mentioned before, the Tikhonov regularization acts uniformly and does not allow a locally adapted regularization. Therefore the fast transient at $t \approx 110s$ is not resolved at an acceptable noise level in the flat areas of the estimate. This is also the case if the problem is discretized non-uniformly (not shown). To illustrate the effect of over-regularization the transient $\alpha = 250,\ J = 7$ is shown in Figure 10 where the estimates are smooth but have a large bias. Note, that the estimates reveal in particular significant bias near the boundaries which is due to the (wrong) guess $Q_{rea,ref} = 1.8\frac{K}{s}$. The bias gets worse penetrating also into the interior region for bad guesses such as $Q_{rea,ref} = 0\frac{K}{s}$. Therefore, a good guess of the solution is crucial for the performance. An improved guess of the reference function $Q_{rea,ref}$ might be obtained by an estimation of $Q_{rea}$ on a suitable (possibly non-uniform) grid with $\alpha = 0$. This estimate can then be used as reference function $Q_{rea,ref} = Q^{\star}_{rea,estim}$ and a higher resolution might be obtained applying Tikhonov's method with an appropriately chosen $\alpha$.

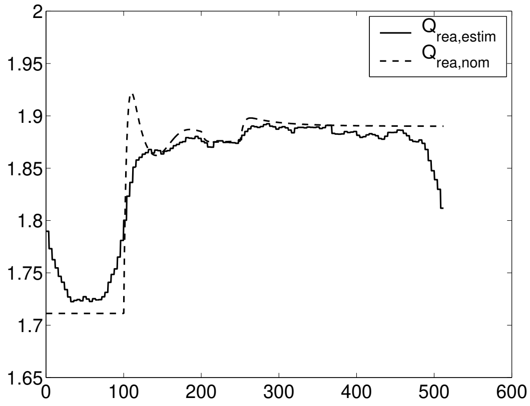


Fig. 10. $Q_{rea,estim}$ on uniform mesh: $\alpha = 250$, $J = 7$.

## 6. CONCLUSIONS

DDR has been discussed from the viewpoint of inverse problems. The DDR problem is ill-posed and requires regularization of some form. A well known method is Tikohonov regularization which adds a suitable penalty to the cost function. However, regularization might also be obtained by discretization. Discretization indirectly regularizes the DDR problem and has to be chosen appropriately to minimize data and regularization error in the estimates. Based on successive upgrades a novel method is suggested to automatically generate a hierarchy of discretization grids until a best compromise between data and regularization error is achieved. Problem adapted uniform as well as non-uniform upgrades have been considered. In particular, it turned out for the specific example, that non-uniform approximation leads to estimates that cannot be achieved employing a uniform grid. The regularization framework has been presented for linear models on a fixed time window and serves as a starting point for moving horizon approaches, where a similar behavior is expected. The method should carry over to control problems which are dual to estimation problems. The analysis of ill-posed problems can be also extended to problems with nonlinear models. However, in the nonlinear case the concept of ill-posedness can not be motivated by singular values and therefore the obtained estimates are valid only qualitatively. The regularization approach has been illustrated using a model with one unknown input. The problem of quantitatively assessing the best compromise becomes increasingly difficult for a growing number of unknown input functions due to the combinatorial increase of parameter variations. Here suitable, low-dimensional strategies have to be developed to effectively choose the regularization parameters.

## 7. REFERENCES


Albuquerque, J.S. and L.T.Biegler (1996). Data reconciliation and gross-error detection for dynamic systems. *AIChE J.* **42**(10), 2841–2856.

Almasy, G. A. (1990). Principles of dynamic balancing. *AIChE J.* **36**(9), 1321–1330.

Binder, T., L. Blank, W. Dahmen and W. Marquardt (1998). Towards multiscale dynamic data reconciliation. In: *Nonlinear Model Based Process Control* (R. Berber and C. Kravaris, Eds.). pp. 623–665. NATO–ASI Series. Kluwer. Dordrecht.

Binder, T., L. Blank, W. Dahmen and W. Marquardt (2000). Grid refinement in multiscale dynamic optimization. In: *European Symposium on Computer Aided Process Engineering - 10* (S. Pierucci, Ed.). pp. 31–36. Elsevier. Amsterdam.

Dahmen, W. (1997). Wavelet and multiscale methods for operator equations. In: *Acta Numerica*. pp. 55–228. University Press. Cambridge.

Dahmen, W., A. Kunoth and K. Urban (1999). Biorthogonal spline-wavelets on the interval–stability and moment conditions. *Appl. Comp. Harmon. Anal.* **6**, 132–196.

Dicken, V. and P. Maass (1996). Wavelet–Galerkin methods for ill-posed problems. *J. Inv. Ill-Posed Problems* **4**(3), 203–222.

Donoho, D. (1995). Nonlinear solution of linear inverse problems by Wavelet–Vaguelette decomposition. *Appl. Comp. Harmon. Anal.* **2**, 101–126.

Engl, H.W., M.H. Hanke and A. Neubauer (1996). *Regularization of Inverse Problems.* Kluwer. Dordrecht.

Freeden, W. and F. Schneider (1998). Regularization wavelets and multiresolution. *J. Inv. Ill-Posed Problems* **14**, 225–243.

Gelb, A. (1978). *Applied Optimal Estimation.* MIT Press. Cambridge.

Golub, G. and C. van Loan (1996). *Matrix Computations.* John Hopkins Press. London.

Hanke, M. (1995). *Conjugate gradient type methods for ill–posed problems.* Longman Scientific and Technical. Harlow, Essex.

Hansen, C. (1998). *Rank-Deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion.* SIAM. Philadelphia.

Kelly, J. (1998). A regularization approach to the reconciliation of constrained data sets. *Comput. Chem. Eng.* **12**(12), 1771–1778.

Kirsch, A. (1996). *An Introduction to the Mathematical Theory of Inverse Problems.* Springer.

Kuehn, D. R. and H. Davidson (1961). Computer control II: Mathematics of control. *Chem. Eng. Prog.* **57**(6), 16531–1672.

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

Mhamdi, A. and W. Marquardt (1999). An inversion approach to the estimation of reaction rates in chemical reactors. In: *Proceedings of European Control Conference 1999*. pp. F1004–1. Karlsruhe, 31 Aug. – 3 Sept.

Michalska, H. and D. Mayne (1995). Moving horizon observers-based control. *IEEE Transactions on Automatic Control* **40**(6), 995–1006.

Morozov, V. A. (1984). *Methods for solving incorrectly posed problems*. Springer. New York.

Narasimhan, S. and C. Jordache (2000). *Data Reconciliation and Gross Error Detection*. Gulf Publishing Company. Houston, Texas.

Neumaier, A. (1998). Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Rev.* **40**, 636–666.

Plato, R. and G. Vainikko (1990). On the regularization of projection methods for solving ill-posed problems. *Numer. Math* **57**, 63–79.

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

Rao, C.V. and J.B. Rawlings (2000). Nonlinear horizon state estimation. In: *Nonlinear Model Predictive Control* (F. Allgöwer and A. Chen, Eds.). pp. 45–69. Birkhäuser Verlag. Basel.

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

Romagnoli, J. A. and M. C. Sanchez (1999). *Data Processing and Reconciliation for Chemical Process Operation*. Academic Press. San Diego.

Schuler, H. and C. Schmidt (1992). Calorimetric state estimators for chemical reactor diagnosis and control: review of methods and applications. *Chem. Eng. Sci.* **47**, 899–915.

Tikhonov, A.N., A.V. Goncharsky, V.V. Stepanov and A.G. Yagola (1996). *Numerical Methods for the Solution of Ill–posed Problems*. Kluwer Academic Publishers. New York.

Zeitz, M. (1983). Transfer characteristics of extremely fast state feedback and observer systems. *Int. J. Systems Science* **14**, 167–177.