ECHTZEITOPTIMIERUNG GROSSER SYSTEME

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

Regularization of dynamic data reconciliation problems by projection

Preprint 99-14

SCHWERPUNKTPROGRAMM DER DEUTSCHEN FORSCHUNGSGEMEINSCHAFT

REGULARIZATION OF DYNAMIC DATA RECONCILIATION PROBLEMS BY PROJECTION [1]

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

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

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 well–posedness of the problem. However, appropriate discretization schemes of the time–continuous problem act as regularization themselves reducing the need of problem modification. Based on this property, we suggest to successively refine 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 of the estimate. We conjecture further, that non–equidistant discretization grids are more advantageous than uniform grids.

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

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. Measurement information can be upgraded, in principle, by means of incorporating process knowledge in some mathematical model. Dynamic data reconciliation (DDR) provides a set of reconciled measurements and an estimate of unmeasured process quantities based on such a model. DDR combines state, parameter and unknown input estimation and is typically formulated as dynamic optimization problem restricted by the process model (e. g. Liebmann *et al.* 1992). In this paper, we illuminate the properties of dynamic data reconciliation, for a simple linear example on a fixed horizon, from the viewpoint of the mathematic 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). DDR problems are in general terms ill–posed and require some form of regularization. Regularization stabilizes the numerical solution and aims at minimizing the data and regularization errors in the estimate at the same time. Regularization is not explicitly addressed in the literature on DDR but implicitly always applied through either discretization of the continuous problem or by starting off with a discrete formulation, since discretization typically acts itself 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 al-

[1] This work has been supported by the "Deutsche Forschungsgemeinschaft" under grant no. MA1188/6

gorithm. The stepsize of the discretization scheme in the continuous case or the sampling time in the discrete case is of crucial importance to the estimation results and should be chosen judiciously to achieve a good compromise between data and regularization error [2]. 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, we emphasize 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.

2. PROBLEM FORMULATION

DDR problems are often formulated as optimization problems restricted to a dynamical 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. We assume that the time domain of interest has been transformed onto the unity interval $I = [0, 1]$. The DDR problem with a cost functional of least–squares type is given by

$$\min_{x_0,w} \| z - x \|^2_{L_2(I)} \tag{1}$$

$$\text{s.t.} \quad \dot{x} = -ax + u + w, \quad x(0) = x_0 \quad . \tag{2}$$

a is a time invariant known parameter, x and x_0 denote the unknown state and its initial condition, whereas u, w are known, unknown inputs respectively. z denotes an interpolated measurement function which is based upon measurement samples $z(t_i)$ of the state x corrupted by measurement errors. Note, that no penalty is present in (1).

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

3. INVERSE PROBLEMS

The estimation problem belongs to the class of inverse problems which always determines 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 dependency of the estimate on the measurement data. Suppose, we measure the nominal solution, $\tilde{z} = \tilde{x}$, for some given initial condition x_0 and nominal inputs $\tilde{u}, \tilde{w}$ such that $\tilde{w} = \dot{\tilde{z}} - a\tilde{z} - \tilde{u}$. Further suppose, that there is alternatively a corrupted measurement function $z^\delta = \tilde{x} + \delta \sin \frac{n}{\delta} t$. Since z^δ is continuously differentiable, the solution to problem (1),(2) is $w^\delta = \dot{z}^\delta - az^\delta - \tilde{u}$. It is well known that differentiation of z^δ can cause arbitrary 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 error level δ but also by frequency. 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 neither due to the special choice of the infinity norm. Rather, this holds for general L_p norms, although continuous dependency could be enforced by measuring the data error in a stronger norm containing first or higher derivatives. However, in practical situations one prefers to call an error small if the error in the function values are small and therefore some L_p norm is the natural choice. Immediately the question arises what can be done to stabilize the solution.

4. REGULARIZATION

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

$$\boldsymbol{T} y = z \tag{3}$$

where $y := (w, x_0)$ and $\boldsymbol{T}$ is a bounded linear operator with domain $D(\boldsymbol{T}) \subset Y$, $y \in Y, z \in Z$. Y and Z are Hilbert spaces. Usually z does not belong to the range $R(\boldsymbol{T})$. 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 to $y^\dagger$, i. e. $\boldsymbol{T}^\dagger z = y^\dagger$, is the generalized inverse of $\boldsymbol{T}$ and reflects problem (1),(2). The inverse $\boldsymbol{T}^\dagger$ is generally unbounded such that stability is not guaranteed. Regularization can be used to stabilize the best approximate solution by introducing a family of bounded transformations $\boldsymbol{T}_\alpha^\dagger$ such that

$$\lim_{\alpha\to 0} \boldsymbol{T}_\alpha^\dagger z = \boldsymbol{T}^\dagger z,\ \forall z \quad , \tag{4}$$

where α is called the regularization parameter. Given an error bound on the measurement data $\parallel z - z^\delta \parallel \leq \delta$ and $y_\alpha^\delta := \boldsymbol{T}_\alpha^\dagger z^\delta$, the following fundamental error statement holds (Engl *et al.*, 1996; Kirsch, 1996):

$$\parallel y_\alpha^\delta - y^\dagger \parallel \leq \parallel \boldsymbol{T}_\alpha^\dagger \parallel \delta + \parallel \boldsymbol{T}_\alpha^\dagger z - y^\dagger \parallel \quad . \tag{5}$$

The bound of the error between the best approximate and estimated solution 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. The data error largely corresponds to the variance of the estimate but may contain also some bias. In contrast, the regularization error can almost solely be attributed to the bias of the estimate. Since we do not use a stochastic setting here 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. By property (4) an increase of α bounds $\boldsymbol{T}_\alpha^\dagger$ and improves stability such that for a constant error level δ the data error is decreased. On the other hand, the increasing difference between $\boldsymbol{T}_\alpha^\dagger$ and $\boldsymbol{T}^\dagger$ causes a growing approximation error.

Opposite behavior is observed for decreasing α. Naturally, we are interested to obtain sharp estimates for $\parallel y_\alpha^\delta - y^\dagger \parallel$. Therefore we have to seek for an α which minimizes the sum of data and regularization error in (5). For a variety of $\boldsymbol{T}_\alpha^\dagger$ the error bound might be accessible for *known* δ, z, although possibly being numerically expensive. However, in practise δ, z are not available and the α reflecting the best compromise can not be computed rigorously. Obviously, criteria are needed which aim to find a good compromise in the absence of concrete knowledge on the data error level as well as the exact measurement. Such criteria 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 $\parallel \boldsymbol{L} y_\alpha^\delta \parallel$ versus $\parallel z^\delta - \boldsymbol{T} y_\alpha^\delta \parallel$ in log–log scale for a large range of α. Generally, $\boldsymbol{L}$ is either the identity or a derivative operator. The curve typically has L–shape with a distinct corner point. There is theoretical evidence and substantial practical experience that the corner point reflects a good compromise between data and regularization error in the estimate.

A wide variety of regularization methods are available. For example Tikhonov regularization (Engl *et al.*, 1996) is frequently applied in engineering and minimizes the functional $\parallel \boldsymbol{T}y - z \parallel^2 + \alpha \parallel y - y_{ref} \parallel^2$ where y_{ref} denotes an a priori guess of the solution and ideally is $y^\dagger$. The associated regularization operator is given by $\boldsymbol{T}_\alpha^\dagger = (\alpha \boldsymbol{I} + \boldsymbol{T}^*\boldsymbol{T})^{-1}\boldsymbol{T}^*$ with $\boldsymbol{T}^*$ being the adjoint of $\boldsymbol{T}$. In practice, it is nontrivial to find a suitable guess y_{ref} and a wrong guess might lead to unsatisfactory results. Regularization by projection, upon we will focus next, does not depend on such a guess.

5. 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 $\boldsymbol{Q}_n : Z \to Z_n$, $\boldsymbol{P}_n : Y \to Y_n$ be suitable projectors onto Z_n, Y_n, then $\boldsymbol{Q}_n \boldsymbol{T} \boldsymbol{P}_n y = \boldsymbol{Q}_n z$ reduces to a Petrov–Galerkin scheme which represents the discretized problem. Furthermore, we define $y_n^\dagger := \boldsymbol{T}_n^\dagger z_n$ as the least–squares solution of minimal norm to $\boldsymbol{T}_n y_n = z_n$ where $\boldsymbol{T}_n := \boldsymbol{Q}_n \boldsymbol{T} \boldsymbol{P}_n$, $z_n := \boldsymbol{Q}_n z$, $y_n := \boldsymbol{P}_n y$. Then, given the error bound $\parallel \boldsymbol{Q}_n(z - z^\delta) \parallel \leq \delta$ it can be shown (Engl *et al.*, 1996) that

$$\begin{aligned} \parallel y_n^\delta - y^\dagger \parallel &\leq \parallel \boldsymbol{T}_n^\dagger \parallel \delta + \parallel \boldsymbol{T}_n^\dagger z_n - y^\dagger \parallel \\ &\leq \frac{\delta}{\sigma_n} + \parallel y_n^\dagger - y^\dagger \parallel \end{aligned} \tag{6}$$

where σ_n is the smallest singular value of $\boldsymbol{T}_n$. Equation (6) is of the same structure as (5). 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 $\boldsymbol{T}_n$. The subscript n takes the role of α. Due to the spectral properties of continuous operators $\sigma_n \to 0$ as $n \to \infty$. 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 (6). Such a procedure can be obtained by application of a refinement approach suggested by the authors (Binder *et al.*, 1998). The principle idea is to successively *upgrade* the approximation 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 ℓ. 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 (Schmid and Zeitz, 1981). The stopping criterion applied is based on the aforementioned L–curve method and does not rely on assumptions on the noise level. Each approximation is represented by one point in the L–curve. As soon as a corner point is detected, the refinement can be terminated.

6. DISCRETIZATION

So far, the concrete choice of a discretization scheme has not been addressed. 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, we present two methods which refine the problem approximation uniformly as well as non–uniformly. First, we sketch some fundamental properties of wavelets (see e.g. Dahmen, 1997 for a recent survey and more details).

Wavelets

The common starting point for the construction of *wavelets* on the *interval* (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 $w_j = \sum_{k\in\mathcal{J}_j} d_{j,k}\psi_{j,k}$. 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 for f smooth on the support of $\tilde{\psi}_{j,k}$ that the values $|d_{j,k}|$ will be quite small. Therefore, functions that exhibit strong variations only locally, can be approximated very accurately by linear combinations of only relatively few wavelet basis functions corresponding to the *significant* coefficients $d_{j,k}$. Denoting by Λ 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 Ψ_Λ as a *sparse* wavelet bases in contrast to *full* wavelet bases Ψ_{Λ^J}, corresponding to index sets $\Lambda^J := \{(j,k) : j_0 - 1 \leq j \leq J-1, k \in \mathcal{J}_j\}$, 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}.

Uniform and adaptive discretization

The approximation spaces Z_n, Y_n introduced above are spanned by a wavelet basis. Specifically, the approximation for the states x and measurements z are generated by piecewise linear wavelets (so called hat functions). The inputs w, u are expanded in terms of piecewise constant (Haar) wavelets. Starting from a uniform 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 bases can also be applied since it might lead to more economical discretization grids. We employ here our own method (Binder *et al.*, 1999) to generate sparse discretization grids. As a simplification we keep in this work the spaces of z, x 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 ℓ, 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 functional. Basis functions which are associated to large sensitivities are

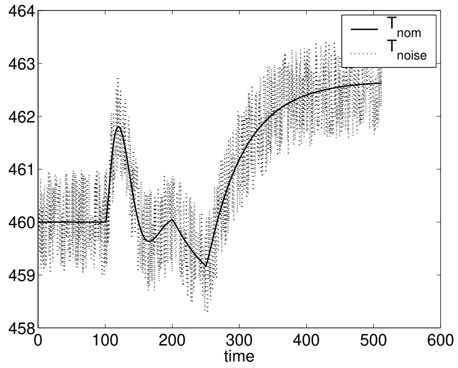

Figure 1. Nominal and corrupted temperatures

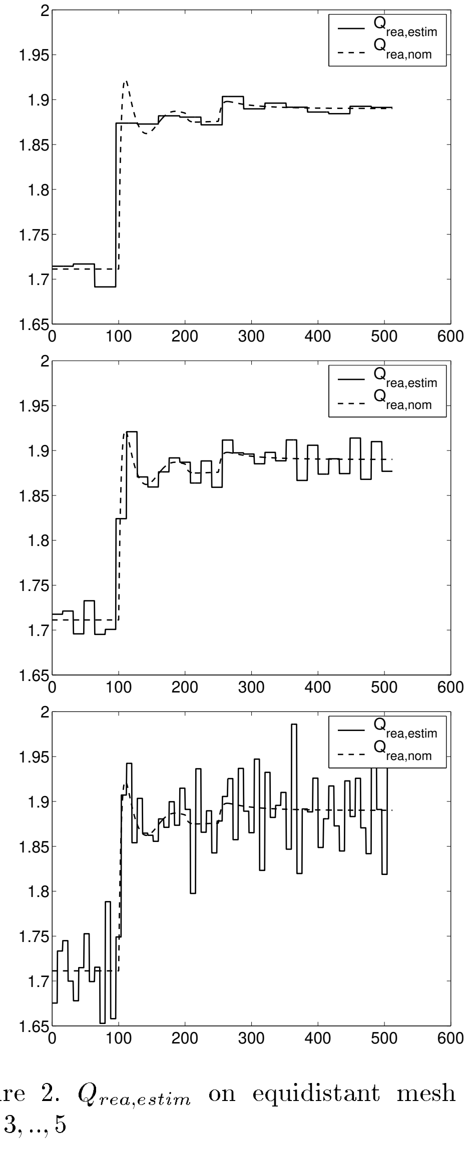

Figure 2. $Q_{rea,estim}$ on equidistant mesh $\ell = 3,..,5$

used for an upgraded approximation at refinement stage $\ell + 1$. As candidate sets, we simply take the local neighbors of the current approximation.

7. EXAMPLE

Our example is built upon calorimetric estimation (Schuler and Schmidt, 1992) applied to a nonisothermal CSTR where a first order reaction $A \rightarrow B$ 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

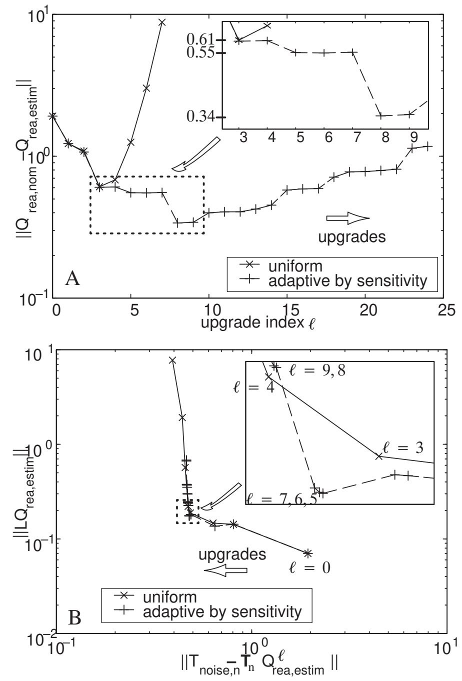

Figure 3. Plot A: L_2 error of the estimates. Plot B: L–curve.

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

with flow rate $q = 10cm^3/s$, vessel volume $V = 10^3cm^3$, heat transfer coefficient $U = 510^{-4}cal/(cm^2sK)$, liquid density $\rho = 0.001g/cm^3$, and heat capacity $c_p = 1cal/(gK)$. The measurement temperature is taken to be the nominal temperature corrupted by uniformly distributed noise. T_{nom},T_{noise} are shown in Figure 1. The reconciliation problem is discretized as outlined before. The approximation spaces for T_{nom}, T_{noise} are kept constant at scale $J = 9$ while we employ uniform $\Psi^{\ell}_{\Lambda^J}$ as well as sparse wavelet bases Ψ^{ℓ}_{Λ} to approximate Q_{rea}. The refinement is started from an initial uniform mesh $\ell = 0$ at scale $J = 1$ and is then gradually upgraded. Estimates obtained employing uniform bases at scale $J = 3, \ldots, 5$ are shown in Figure 2. The observed differentiative effect can be seen nicely. The L_2 error of $Q_{rea,estim}$ and $Q_{rea,nom}$ vs. refinement index ℓ are shown for all upgrades in Figure 3:A. However, in practise, we have no information about the nominal values. The L–curve is used instead to detect a

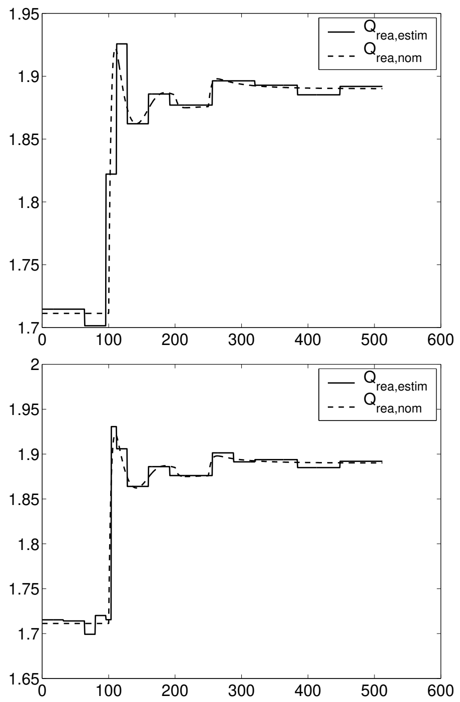

Figure 4. $Q_{rea,estim}$ on adaptive mesh $\ell = 6, 8$

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 $\| T_{noise,n} - \boldsymbol{T}_n Q^{\ell}_{rea,estim} \|_{L_2}$ where the interpolated measurement and continuous operator are evaluated at high resolution (Scale 9). As one can see in Figure 3:B, the refinement level ℓ associated to the corner point of the uniform discretization coincides quite nicely with the best achievable estimates. The L–curve and the L_2 error are shown as well in Figure 3 for non–uniform refinement. 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 2 to Figure 4 where $Q^{\ell}_{rea,estim}$ is shown at upgrade step $\ell = 6, 8$. For uniform discretization either the sharp transient can not be resolved or the flat areas of the estimate reveal significant noise.

CONCLUSIONS

We have discussed the DDR from the viewpoint of inverse problems. The problem is ill–posed and requires regularization of some form. Discretization indirectly regularizes the DDR problem and has to be chosen appropriately to minimize data and regularization error in the estimates. A novel method, based on successive upgrades 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 results for a fixed time window serve as a starting point for moving horizon approaches where similar behavior is expected. The method should carry over to control problems which are dual to estimation problems.

8. REFERENCES

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.). NATO–ASI Series. Kluwer. Dordrecht. pp. 623–665.

Binder, T., L. Blank, W. Dahmen and W. Marquardt (1999). An adaptive multiscale approach to real-time dynamic optimization. LPT Report. RWTH Aachen.

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

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.

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

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

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

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.* Karlsruhe.

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.

Schmid, W. and M. Zeitz (1981). Limiting procedure observer – differentiator. *Regelungstechnik* **29**(8), 270–274.

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.