State Estimation without regularizing the initial data

Luise Blank

published in *Inverse Problems, Vol. 20, No. 5, 1357-1370, (2004)*

Abstract

Monitoring dynamical processes requires the estimation of the entire state, which is only partly accessible by measurements. Most quantities must be determined via model based state estimation, which in general is an ill–posed inverse problem. Regularization techniques have to be applied. To avoid undesired bias we omit the commonly used approach of regularizing the unknown initial data. To the authors knowledge the resulting minimization problem has not been analysed mathematically yet, which is the purpose of this paper. The first order necessary conditions are presented and the optimization problem is reduced by several variables. Hence, e.g. one otherwise necessary regularization parameter is dispensable. The influence of the regularization parameters and of the model matrices is studied in detail for linear models. It is shown that for any choice of regularization parameters the condition numbers of the evolving operators can be arbitrarily large, if the spectral radius of the system matrix is large. In case of one state only we derive additionally bounds for the perturbation of the initial data resulting from measurement errors.

Key Words: state estimation, optimal control, optimality condition, inverse problems, dynamical systems.

1 Introduction

State estimation is the task of determining from a few given, typically noisy measurements over a past time horizon all state functions, which are involved e.g. into a chemical process. The measurements correspond only to a few states. The others have to be estimated using the coupling of the states via model equations. State estimation can be viewed in general an ill–posed inverse problem. Then, regularization techniques have to be applied. In most of the literature the initial data are also regularized (see e.g.[1, 16, 17, 19]). Then the resulting optimization formulation coincides widely with a corresponding control problem, where the initial data are given. Control problems are analysed by many authors and numerical methods are available. However, in state estimation the initial data are typically unknown. Sometimes experiences provide a rough guess for the initial data, or state estimation on one horizon provides guesses for a forwardly moved time horizon, which are then used as reference values for regularization. Though precise values are not available. Hence, to avoid bias we omit the regularization of the initial data (for further discussion see [2]). To the authors knowledge there is still a lack of theoretical

investigations to this problem formulation. Thus the purpose of this paper is to analyse the effects of this problem formulation, to gain insight into the difficulties and to offer possibilities to overcome, respectively to reduce them in an analytical or numerical way.

In Section 2 we shortly deduce the problem formulation of state estimation with incorporated model error functions, discuss the ill–posedness and introduce the applied regularization.

Section 3 is concerned with the reduction of the resulting minimization problem. Based on the first oder necessary optimality conditions it turns out that the optimization problem can be formulated without the output functions, the corresponding model error functions and without output equation. In the equations for the optimality conditions also the model error function corresponding to the state equation can be eliminated. In case of linear ODE's as state equations where model error functions are present in all equations and omitting inequality constraints, the necessary conditions reduce to a differential equation of second order for the states with mixed boundary constraints. This differential equation and its corresponding weak formulation are analysed in detail in the next section.

To gain insight into the properties of the relevant operators we study in Section 4, as a start, the linear case with time independent coefficients. Of particular interest is the influence of the regularization parameters and the influence of the coefficients of the model equations.

In Section 4.1 we consider the case of one state only. Assuming fixed regularization parameters and coefficients we show that the regularized inverse problem is well posed with respect to L_2–perturbations and give bounds for the corresponding map. However, large negative system coefficients may lead to large errors in the initial data for small disturbances in the measurements. From the engineering side the more interesting case are disturbances in the L_∞–norm. These lead to a bound only dependent on the regularization parameters. Not only for the numerical approach it is essential to have bounds also on the norm and the condition number of the operator corresponding to the resulting (differential) system. We give upper and lower bounds for the norm and see that they grow to infinity for any regularization parameter, if the coefficient in the model equation goes to infinity. Furthermore, the connection to the Sturm–Liouville's eigenvalue problem yields information about the eigenvalues. An exponential eigenfunction corresponds to the smallest eigenvalue. This result serves in [5] as one essential ingredient to obtain an efficient preconditioner for the numerical solution of the resulting linear problem.

In Section 4.2 we extend the results concerning the bounds of the operator to a system with several states. Here the issue of observability comes into play. Not only a large spectral radius of the system matrix can lead to large condition numbers but also eigenvectors of the system matrix which are close to the null space of the matrix, which determines the output functions from the states.

Finally in Section 5 we summarize the conclusions.

2 Derivation of the Problem Formulation

In [2] the mathematical formulation of the estimation problem is discussed and its engineering interpretation is given in detail. In this paper we concentrate on the more mathematical approach, arguing with the mathematical features of the variables and equations. We state the model equations and their extension by model error functions and then discuss shortly the regularization of the problem.

2.1 Model Equations

The quantities of interest in process monitoring can be summarized in the state functions $\boldsymbol{x}$, the output functions $\boldsymbol{y}$, the control variables $\boldsymbol{u}$ and parameters $\boldsymbol{p}$. For the particular case of state estimation the controls $\boldsymbol{u}$ and the parameters $\boldsymbol{p}$ are known variables. Then, the mathematical description of the state equations of the process is assumed to be given by the differential algebraic equations (DAE's)

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

with a possibly time dependent matrix $\boldsymbol{G}$, a given function $\boldsymbol{f}$ and the consistent initial condition [6]

$$\boldsymbol{x}(0) = \boldsymbol{x}_0. \tag{2}$$

The output equations

$$\boldsymbol{y} - \boldsymbol{C}\boldsymbol{x} = \boldsymbol{0} \tag{3}$$

describe which states can be measured. Here $\boldsymbol{C}$ is a time independent matrix, typically an identity matrix extended by zero columns. The number of output functions n_y are typically far less than the number of involved states n_x. We require that the system is observable [7, 8, 11], i.e. the map $\mathcal{F} : \boldsymbol{x}_0 \longmapsto \boldsymbol{y}$ given by (1)–(3) is injective for any given $\boldsymbol{u}$ and $\boldsymbol{p}$. Hence given an output $\boldsymbol{y} \in Im(\mathcal{F})$ we can reconstruct uniquely $\boldsymbol{x}$ using (1) and (3).

To include possible model errors into the model equations the DAE's are extended linearly by unknown error functions $\boldsymbol{w}$ and $\boldsymbol{v}$, which results into the equations

$$\boldsymbol{G}\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}) - \boldsymbol{W}\boldsymbol{w} = \boldsymbol{0} \tag{4}$$

$$\boldsymbol{y} - \boldsymbol{C}\boldsymbol{x} - \boldsymbol{V}\boldsymbol{v} = \boldsymbol{0}. \tag{5}$$

The matrices $\boldsymbol{W}$ and $\boldsymbol{V}$ are typically identity matrices enlarged by zero rows. They reflect which equations are trusted. The above model equations (4)–(5) and (2) describe the process on a given interval (time horizon on $[0, T]$) which is assumed to be projected onto $[0, 1]$. For convenience we omit the interval $[0, 1]$ in the following. Moreover, we only assume that the equations are fulfilled in the weak sense. Consequently, the corresponding function spaces are L_2 and the Sobolev space H^1.

For simplicity if the dimension $n = n_x$ etc. is obvious in the context, we omit the exponent n and write L_2 and H^1 instead of the precise notation $(H^1)^n$.

Inequality constraints reflect, for example, safety restrictions and control limits and can be written as

$$\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}) \leq \boldsymbol{0}. \tag{6}$$

For state estimation they are typically of the form $\boldsymbol{c}_i(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}) = x_i - \bar{c}_i$ and $\boldsymbol{c}_{i+n_x}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{p}) = -x_i + \bar{c}_{i+n_x}$ for $i = 1, \ldots, n_x$.

2.2 The Inverse Problem and Regularization

Given $\boldsymbol{x}_0, \boldsymbol{w}$ and $\boldsymbol{v}$ the solution $\boldsymbol{y}$ is uniquely determined by (4) and (5). This defines an operator (in case of $\boldsymbol{G} = \boldsymbol{I}$)

$$\boldsymbol{T} \begin{pmatrix} \boldsymbol{x}_0 \\ \boldsymbol{w} \\ \boldsymbol{v} \end{pmatrix} := \boldsymbol{y} \quad \text{with} \quad \boldsymbol{T} : \mathbb{R}^{n_x} \times (L_2)^{n_w} \times \left(H^1\right)^{n_v} \longrightarrow \left(H^1\right)^{n_y}. \tag{7}$$

For state estimation one wants to determine $\boldsymbol{x}_0, \boldsymbol{w}$ and $\boldsymbol{v}$ and, hence, $\boldsymbol{x}$ from given data. If the data are clean they shall coincide with $\boldsymbol{y}$. However, in general only noisy measurements are available. Assuming that the measurements taken at discrete times can be processed to signal functions $\boldsymbol{z}$, these functions are typically not in H^1 but in L_2. Hence, the output functions $\boldsymbol{y}$ cannot match the measurements $\boldsymbol{z}$ exactly. The estimation solution is given as a least squares solution, i.e. $\boldsymbol{T}^\dagger \boldsymbol{z} := \left(\boldsymbol{x}_0^T, \boldsymbol{w}^T, \boldsymbol{v}^T\right)^T$ where $\|\boldsymbol{T}\left(\boldsymbol{x}_0^T, \boldsymbol{w}^T, \boldsymbol{v}^T\right)^T - \boldsymbol{z}\|_{(L_2)^{n_y}}$ is minimal. Taking into account that each output function can be weighted separately the optimization problem is formulated as

$$\min_{\boldsymbol{x}_0, \boldsymbol{w}, \boldsymbol{v}} 1/2 \int_0^1 (\boldsymbol{y} - \boldsymbol{z})^T \boldsymbol{R}_y (\boldsymbol{y} - \boldsymbol{z})\, dt \qquad \text{with} \qquad \boldsymbol{T}\left(\boldsymbol{x}_0^T, \boldsymbol{w}^T, \boldsymbol{v}^T\right) = \boldsymbol{y} \tag{8}$$

with some time independent, positive definite matrix $\boldsymbol{R}_y$.

The generalized inverse $\boldsymbol{T}^\dagger : (L_2)^{n_y} \longrightarrow \mathbb{R}^{n_x} \times (L_2)^{n_w} \times (H^1)^{n_v}$ is an unbounded operator, since highly oscillatory functions $\boldsymbol{z} \in (H^1)^{n_y}$ with uniformly bounded L_2– or L_∞–norms can lead to unbounded L_2–, respectively, L_∞–norms of $\boldsymbol{w}$ due to the involved differential operator (for example $z(t) = \varepsilon \sin(t/\varepsilon^2)$). Additionally the problem of non–uniqueness in $\boldsymbol{w}$ and $\boldsymbol{v}$ can occur.

One possible regularization technique to overcome the ill–posedness is Tikhonov–type regularization [9, 14]. This leads to the following ansatz for the cost functional:

$$1/2 \int_0^1 (\boldsymbol{y} - \boldsymbol{z})^T \boldsymbol{R}_y (\boldsymbol{y} - \boldsymbol{z}) + \boldsymbol{v}^T \boldsymbol{R}_v \boldsymbol{v} + \boldsymbol{w}^T \boldsymbol{R}_w \boldsymbol{w}\, dt \tag{9}$$

with some time independent, positive definite matrices $\boldsymbol{R}_v$, $\boldsymbol{R}_w$. The resulting optimization problem for state estimation on a fixed horizon projected onto $[0, 1]$ is

then given by minimizing (9) with the model equations (4)–(5) and the initial condition (2) as equality constraints and the inequality constraints (6). We would like to mention that typically the initial values $\boldsymbol{x}_0$ are regularized too (for e.g. [13, 17, 20]) with estimated reference values $\boldsymbol{x}_0^{ref}$. The additional regularization of the initial values origins from the classical way of Tikhonov–regularization, which regularizes all unknowns. Moreover, in a receding horizon estimator it is used to reflect the measurements of a past time horizon into the current horizon [20], where then the regularization matrices can be adjusted according to linear filter theory. In the context of nonlinear state estimation this is not reasonable any longer [17, 20]. In addition, for the first time horizon appropriate reference values may be missing. In all cases precise values are not given. Bias may occur as the simple one-dimensional linear example $\dot{x} - \alpha x = 0, y = x, z \equiv 0$ with the solution $x(t) = x_0 \exp(\alpha t)$ shows: while omitting regularization of x_0 leads to $x \equiv 0$, which corresponds to z, regularization of x_0 with a reference value $x_0^{ref} \neq 0$ yields $x_0 = x_0^{ref} \frac{2\alpha}{\exp(2\alpha)-1+2\alpha}$ and hence $x \not\equiv z$. Therefore, as mentioned in the beginning, we omit the regularization of the initial values on purpose to avoid unnecessary bias.

3 Reduction of the Problem Formulation

Since $\boldsymbol{x}_0$ is unknown we observe that the initial condition (2) has in case of an estimation problem no information and need for the optimization problem. Hence, the equation (2) can be omitted in the problem formulation. Moreover, we do not have to distinguish between $\boldsymbol{u}$ and $\boldsymbol{p}$ for analytical reasons, since both are given. In the following we include $\boldsymbol{p}$ into $\boldsymbol{u}$.

In the next step we exploit the relations to the Lagrange multipliers corresponding to the equality constraints. Therefore, we consider in detail the first order necessary optimality conditions leading to the Euler–Lagrange equations, which are obtained by the Pontryagin's maximum principle [10, 12, 18].

Since no inequality constraints on $\boldsymbol{y}$ and $\boldsymbol{v}$ are present, we obtain —using the Lagrange multiplier $\boldsymbol{\lambda}_2$ with respect to (5)— the equations

$$\boldsymbol{y} = -\boldsymbol{R}_y{}^{-1}\boldsymbol{\lambda}_2 + \boldsymbol{z} \tag{10}$$

$$\boldsymbol{v} = \boldsymbol{R}_v{}^{-1}\boldsymbol{V}^T\boldsymbol{\lambda}_2 \tag{11}$$

and can deduce

$$\boldsymbol{v} = \boldsymbol{R}_v{}^{-1}\boldsymbol{V}^T\boldsymbol{R}_y(\boldsymbol{z} - \boldsymbol{y}). \tag{12}$$

Using (5) and (12) yields

$$\boldsymbol{y} = \left(\boldsymbol{I} + \boldsymbol{V}\boldsymbol{R}_v{}^{-1}\boldsymbol{V}^T\boldsymbol{R}_y\right)^{-1}(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{z}) + \boldsymbol{z}. \tag{13}$$

Consequently $\boldsymbol{y}$ and $\boldsymbol{v}$ can be eliminated in the formulation of the optimization problem. Let us define now $\boldsymbol{Q} := \left(\boldsymbol{R}_y{}^{-1} + \boldsymbol{V}\boldsymbol{R}_v{}^{-1}\boldsymbol{V}^T\right)^{-1}$ then the problem formulation (4)–(6), (9) is equivalent to the following minimization.

Reduced optimization problem:

$$\min 1/2 \int_0^1 (\boldsymbol{C}\boldsymbol{x} - \boldsymbol{z})^T \boldsymbol{Q}(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{z}) + \boldsymbol{w}^T \boldsymbol{R}_w \boldsymbol{w}\, dt \tag{14}$$

s.t.

$$\begin{aligned} \boldsymbol{G}\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{W}\boldsymbol{w} &= \boldsymbol{0} && (15) \\ \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) &\leq \boldsymbol{0}. && (16) \end{aligned}$$

Considering the obtained optimization problem we can deduce:

Remark 3.1 *The inclusion of $\boldsymbol{v}$ as a model error function into the problem formulation is not necessary. We can view $\boldsymbol{C}\boldsymbol{x}$ as in (3) as outputs and, after obtaining the estimates, distinguish as in (5) which amount $\boldsymbol{y}$ of the outputs $\boldsymbol{C}\boldsymbol{x}$ is trustful and which amount is considered as error $\boldsymbol{v}$. Consequently, we can also avoid the problem of finding an appropriate regularization matrix $\boldsymbol{R}_v$.*

Exploiting the Lagrange multiplier $\boldsymbol{\lambda}$ with respect to (15) we obtain the relation

$$\boldsymbol{w} = {\boldsymbol{R}_w}^{-1} \boldsymbol{W}^T \boldsymbol{\lambda}, \tag{17}$$

which means that we can eliminate the model error function $\boldsymbol{w}$ in the optimization process. However, with this elimination we leave also the framework of the saddle point formulation. Given (17) $\boldsymbol{w}$ has the same differentiability as $\boldsymbol{\lambda}$. Defining $\boldsymbol{R} := \boldsymbol{W}{\boldsymbol{R}_w}^{-1}\boldsymbol{W}^T$ we make the following substitutions

$$\begin{aligned} \boldsymbol{W}\boldsymbol{w} &= \boldsymbol{R}\boldsymbol{\lambda} && (18) \\ \boldsymbol{w}^T \boldsymbol{R}_w \boldsymbol{w} &= \boldsymbol{\lambda}^T \boldsymbol{R} \boldsymbol{\lambda}. && (19) \end{aligned}$$

This clearly indicates that for the computation only $\boldsymbol{R}$ plays a role.

Remark 3.2 *It is not necessary to distinguish between $\boldsymbol{W}$ and $\boldsymbol{R}_w$ until one wants to determine $\boldsymbol{w}$ instead of $\boldsymbol{W}\boldsymbol{w}$.*

Necessary first order conditions:

The remaining first order necessary optimality conditions for the solution of the estimation problem are the equations

$$\begin{aligned} \boldsymbol{G}\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{R}\boldsymbol{\lambda} &= \boldsymbol{0} && (20) \\ -\frac{d}{dt}\left(\boldsymbol{G}^T \boldsymbol{\lambda}\right) - \left(\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})\right)^T \boldsymbol{\lambda} + \boldsymbol{C}^T \boldsymbol{Q} \boldsymbol{C} \boldsymbol{x} + \left(\frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u})\right)^T \boldsymbol{\nu} &= \boldsymbol{C}^T \boldsymbol{Q} \boldsymbol{z} && (21) \\ \left(\boldsymbol{G}^T \boldsymbol{\lambda}\right)(0) &= \boldsymbol{0} && (22) \\ \left(\boldsymbol{G}^T \boldsymbol{\lambda}\right)(1) &= \boldsymbol{0} && (23) \\ \boldsymbol{\nu}^T \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) &= 0 && (24) \\ \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{u}) &\leq \boldsymbol{0} && (25) \\ \boldsymbol{\nu} &\geq \boldsymbol{0} && (26) \end{aligned}$$

with unknown Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\nu}$ and unknown states $\boldsymbol{x}$. The output func-

tion $\boldsymbol{y}$ is given by (13), the model error functions $\boldsymbol{v}$ by (12) and $\boldsymbol{w}$ by (17). Equations (22) and (23) reflect the fact that no boundary conditions and no regularization of the boundary values of the states $\boldsymbol{x}$ are given.

In case $\boldsymbol{R}_w = \boldsymbol{0}$ which is for example considered in case of regularization by discretization [4], $\boldsymbol{w}$ cannot be eliminated. We obtain additionally the equation $\boldsymbol{W}^T\boldsymbol{\lambda} = \boldsymbol{0}$ and we have to substitute $\boldsymbol{R}\boldsymbol{\lambda}$ with $\boldsymbol{W}\boldsymbol{w}$ in (20). If no inequality constraints are present, (20)–(23) is a DAE system of order one, where for at most half of the variables boundary conditions are given, while for the others the boundary values are unknown. In case that the state equations are given by ODE's only, we obtain with (17)

Remark 3.3 *If $\boldsymbol{G} = \boldsymbol{I}$ then the model error functions $\boldsymbol{w}$ are in H^1 and vanish at the boundary $t = 0$ and $t = 1$. The boundary values are enforced by the described Tikhonov–regularization.*

It is questionable whether one would like to have possible model errors also at the boundary. Regularization by discretization may offer this possibility. In case of regularization of the initial values of the states $\boldsymbol{x}$ with some reference guesses $\bar{\boldsymbol{x}}_0$, which is seen mostly in the literature, the model errors are free at the boundary too. However, as mentioned, for wrong reference data undesired bias may appear in the states.

The weak formulation can play an important role in the numerical treatment of the optimization problem (e.g. applying Galerkin methods). While the weak formulation of (20) and (24) is obvious, for further references we would like to formulate at this point the weak formulation of (21) which includes the boundary conditions (22) and (23). Denoting with $\langle\,,\rangle$ the standard L_2–product, we obtain:

$$\begin{aligned}\langle \boldsymbol{G}^T\boldsymbol{\lambda}, \frac{d}{dt}\boldsymbol{\zeta}\rangle \;&-\; \langle\left(\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\right)^T\boldsymbol{\lambda},\boldsymbol{\zeta}\rangle + \langle \boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{x},\boldsymbol{\zeta}\rangle + \langle\left(\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{c}(\boldsymbol{x},\boldsymbol{u})\right)^T\boldsymbol{\nu},\boldsymbol{\zeta}\rangle \\ &=\; \langle \boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{z},\boldsymbol{\zeta}\rangle \qquad \text{for all } \boldsymbol{\zeta}\in H^1. \end{aligned} \tag{27}$$

3.1 Invertible Regularization Matrix R

If $\boldsymbol{R}$ is invertible, model error functions have to be present in each state equation. This implies that $\boldsymbol{W}$ is invertible. Hence $\boldsymbol{w}$ can be replaced in the cost functional by $\boldsymbol{W}^{-1}(\boldsymbol{G}\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}))$, which leads to an optimization problem in $\boldsymbol{x}$ with the inequality constraints (25) only.

Consequently, also $\boldsymbol{\lambda}$ can be eliminated using equation (20). It follows

$$\boldsymbol{\lambda} \;=\; \boldsymbol{R}^{-1}(\boldsymbol{G}\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})) \tag{28}$$

and we obtain a second order DAE system for the states $\boldsymbol{x}$ (see (21)):

$$-\frac{d}{dt}\left(\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{G}\frac{d}{dt}\boldsymbol{x}\right)+\frac{d}{dt}\left(\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\right)-\left(\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\right)^T\boldsymbol{R}^{-1}\boldsymbol{G}\frac{d}{dt}\boldsymbol{x}$$
$$+\left(\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\right)^T\boldsymbol{R}^{-1}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})+\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{x}+\left(\frac{\partial}{\partial\boldsymbol{x}}\boldsymbol{c}(\boldsymbol{x},\boldsymbol{u})\right)^T\boldsymbol{\nu}=\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{z} \tag{29}$$

with the boundary constraints $\boldsymbol{G}^T\boldsymbol{R}^{-1}\left(\boldsymbol{G}\dot{\boldsymbol{x}}-\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\right)=\boldsymbol{0}$ for $t=0$ and $t=1$ and the equations (24)–(26) for the inequality constraints.
In general $\boldsymbol{G}$ and $\boldsymbol{R}$ are time independent matrices, hence the differentiation in time can be interchanged in the above equation with these matrices. Moreover, we would like to remind that $\boldsymbol{R}$ and $\boldsymbol{Q}$ are symmetric and positive definite.

Linear ODE's without inequality constraints

In the linear ODE case, i.e. $\boldsymbol{G}\dot{\boldsymbol{x}}-\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})=\dot{\boldsymbol{x}}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{u}$ without inequality constraints where we additionally summarized all given parameters, controls and their perhaps nonlinear influence to a given function $\boldsymbol{u}$, the optimization problem is equivalent to the unconstraint minimization problem

$$min\left\{\|\boldsymbol{Q}^{1/2}\left(\boldsymbol{C}\boldsymbol{x}-\boldsymbol{z}\right)\|_{L_2}^2+\|\boldsymbol{R}^{-1/2}\left(\dot{\boldsymbol{x}}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{u}\right)\|_{L_2}^2\right\}. \tag{30}$$

As necessary first order conditions we obtain the second order differential equation (see also (29))

$$\begin{aligned}&-\boldsymbol{R}^{-1}\ddot{\boldsymbol{x}}+\left(\boldsymbol{R}^{-1}\boldsymbol{A}-\boldsymbol{A}^T\boldsymbol{R}^{-1}\right)\dot{\boldsymbol{x}}+\left(\boldsymbol{A}^T\boldsymbol{R}^{-1}\boldsymbol{A}+\boldsymbol{R}^{-1}\dot{\boldsymbol{A}}+\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\right)\boldsymbol{x}\\ =\;&\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{z}-\boldsymbol{R}^{-1}\dot{\boldsymbol{u}}-\boldsymbol{A}^T\boldsymbol{R}^{-1}\boldsymbol{u}\end{aligned} \tag{31}$$

with the boundary conditions $\dot{\boldsymbol{x}}-\boldsymbol{A}\boldsymbol{x}=\boldsymbol{u}$ at $t=0$ and $t=1$.

While the strong formulation does not look symmetric, the weak formulation reflects the symmetry given by the formulation (30).

$$\begin{aligned}&\langle\dot{\boldsymbol{\zeta}}-\boldsymbol{A}\boldsymbol{\zeta},\boldsymbol{R}^{-1}\left(\dot{\boldsymbol{x}}-\boldsymbol{A}\boldsymbol{x}\right)\rangle+\langle\boldsymbol{C}\boldsymbol{\zeta},\boldsymbol{Q}\boldsymbol{C}\boldsymbol{x}\rangle\\ =\;&\langle\boldsymbol{C}\boldsymbol{\zeta},\boldsymbol{Q}\boldsymbol{z}\rangle+\langle\dot{\boldsymbol{\zeta}}-\boldsymbol{A}\boldsymbol{\zeta},\boldsymbol{R}^{-1}\boldsymbol{u}\rangle\qquad\text{for all }\boldsymbol{\zeta}\in H^1.\end{aligned} \tag{32}$$

4 Properties of the involved operators

To gain more insight in the difficulties for the solution process we consider the linear case, with time independent coefficients and without inequality constraints in more detail. It enhances already some of the main features we face for state estimation if the initial data are not regularized in the problem formulation. We first consider a model with one state only. In this case we can determine the solution explicitly and analyse its dependency on perturbation in the signal $\boldsymbol{z}$. For the numerical approach

however, equation (32) or equivalent formulations are used. Hence, the properties of the operator corresponding to (32) is analysed in more detail. In the second subsection we then can extend most ideas to systems with several states taking into account the issue of observability.

4.1 Linear model with one state function

With $n_x = 1$, $\boldsymbol{A} = \alpha$, $\boldsymbol{C} = \delta$, $\boldsymbol{W} = 1$, $\boldsymbol{R}_w = r$ and $\boldsymbol{Q} = q$ above differential equations reduce to

$$\begin{aligned} -r\ddot{x} + \left(r\alpha^2 + q\delta^2\right) x &= q\delta z - r\dot{u} - r\alpha u \qquad (33)\\ \text{s.t.} \quad \dot{x}(1) - \alpha x(1) = u(1) \quad &\text{and} \quad \dot{x}(0) - \alpha x(0) = u(0). \end{aligned}$$

The solution x and the model error function w are given with $\zeta := \sqrt{\alpha^2 + \frac{q\delta^2}{r}}$ by:

$$\begin{aligned} x_0 = \frac{-1}{\sinh(\zeta)} \int_0^1 \{&\sinh(\zeta - \zeta s)u(s) \qquad (34)\\ &+\frac{1}{\delta}\left[\alpha \sinh(\zeta - \zeta s) - \zeta \cosh(\zeta - \zeta s)\right] z(s)\Big\}\, ds \end{aligned}$$

$$\begin{aligned} x(t) = \frac{1}{\zeta}\left(\zeta \cosh(\zeta t) + \alpha \sinh(\zeta t)\right) x_0 \qquad (35)\\ +\frac{1}{\zeta}\int_0^t \left\{\left(\zeta \cosh(\zeta(t-s)) + \alpha \sinh\left(\zeta(t-s)\right)\right) u(s) - \frac{q\delta}{r}\sinh\left(\zeta(t-s)\right) z(s)\right\} ds \end{aligned}$$

$$\begin{aligned} w(t) = \frac{q\delta^2}{\zeta r}\sinh(\zeta t) x_0 \qquad (36)\\ +\frac{q\delta^2}{\zeta r}\int_0^t \left\{\sinh\left(\zeta(t-s)\right) u(s) + \frac{1}{\delta}\left(\alpha \sinh\left(\zeta(t-s)\right) - \zeta \cosh(\zeta(t-s))\right) z(s)\right\} ds. \end{aligned}$$

One typically has $\delta = 1$ meaning the state x corresponds to the measurements z in the output equation (5). The parameters influencing the solution are the regularization parameters and the coefficient α of the ODE.

Theorem 4.1 *Assuming a linear model with one state function and time independent coefficients and omitting inequality constraints we obtain with above notation:*

1. *The regularized problem formulation (14)–(16) of state estimation, i.e. given z determining x_0 and w, is well posed.*

2. *Small perturbation of z in the L_2–norm can lead to unbounded error propagation in the initial data x_0 independently of r and q for $\alpha \longrightarrow -\infty$.*

3. *For perturbations of z in the L_∞–norm we have bounds independently of α:*

$$|x_0^\varepsilon| \le \frac{2}{|\delta|}\|z^\varepsilon\|_{L_\infty} \quad \textit{and} \quad |w^\varepsilon(t)| \le \frac{2q\delta}{r\zeta(1-e^{-2\zeta})}\|z^\varepsilon\|_{L_\infty} \le \frac{2\sqrt{q/r}}{1-e^{-2\delta\sqrt{q/r}}}\|z^\varepsilon\|_{L_\infty}$$

where we used the abbreviation $z^\varepsilon = z - \bar{z}$ etc. for the perturbation.

Proof: Let us denote $f(t,s) := \zeta\cosh(\zeta(t-s)) - \alpha\sinh(\zeta(t-s))$ then $f > 0$ for all t, s and

$$\|\frac{1}{\sinh(\zeta)}f(1,.)\|_{L_2} = \left[\frac{q\delta^2}{2r\sinh^2(\zeta)} + \frac{\zeta^2+\alpha^2}{2\zeta}\frac{\cosh(\zeta)}{\sinh(\zeta)} - \alpha\right]^{1/2}.$$

Hence $\|\frac{1}{\sinh(\zeta)}f(1,.)\|_{L_2} \approx \sqrt{2|\alpha|}$ as $\alpha \longrightarrow -\infty$.
Moreover we have

$$\|f(.,.)\|^2_{L_2((0,1)\times(0,1))} = \frac{q\delta^2}{2r} + (1 - \frac{q\delta^2}{2(\alpha^2 r + q\delta^2)})\sinh^2(\zeta).$$

1.) With the Hölder–inequality we have

$$|x_0^\epsilon| \le \|\frac{1}{\delta\sinh(\zeta)}f(1,.)z^\epsilon\|_{L_1} \le \|\frac{1}{\delta\sinh(\zeta)}f(1,.)\|_{L_2}\|z^\epsilon\|_{L_2} \le c\|z^\epsilon\|_{L_2}$$

for fixed r, q, δ and α. Also we obtain the bound

$$\|w^\epsilon\|_{L_2} \le \left(\|\frac{1}{\delta\sinh(\zeta)}f(1,.)\|_{L_2}\|\sinh(\zeta\cdot)\|_{L_2} + \|\frac{1}{\delta}f(.,.)\|_{L_2((0,1)\times(0,1))}\right)\frac{q\delta^2}{r\zeta}\|z^\epsilon\|_{L_2}.$$

2.) Choosing $z^\epsilon(s) = \frac{1}{\sqrt{|\alpha|}}\varepsilon\delta\frac{1}{\sinh(\zeta)}f(1,s)$ we have $\|z^\epsilon\|_{L_2} \le c\varepsilon$ with c independent of α. On the other hand

$$x_0^\varepsilon = \frac{1}{\sqrt{|\alpha|}}\varepsilon\|\frac{1}{\sinh(\zeta)}f(1,.)\|^2_{L_2} \approx 2\varepsilon\sqrt{|\alpha|} \quad \text{as } \alpha \longrightarrow -\infty.$$

3.) For the bound of w^ε we use the following transformation

$$\begin{aligned} w^\varepsilon(t) &= \frac{q\delta}{r\zeta\sinh(\zeta)}\left\{\sinh(\zeta t)\int_t^1 f(1,s)z^\varepsilon(s)ds \right. \\ &\qquad \left. + \int_0^t \{\sinh(\zeta t)f(1,s) - \sinh(\zeta)f(t,s)\}z^\varepsilon(s)ds\right\} \\ &= \frac{q\delta}{r\zeta\sinh(\zeta)}\left\{\sinh(\zeta t)\int_t^1 f(1,s)z^\varepsilon(s)ds - \sinh(\zeta-\zeta t)\int_0^t f(0,s)z^\varepsilon(s)ds\right\}, \end{aligned}$$

$$\begin{aligned} |w^\varepsilon(t)| &\le \frac{q\delta}{r\zeta}\|z^\varepsilon\|_{L_\infty}\left\{\frac{\sinh(\zeta t)}{\sinh(\zeta)}\left(\sinh(\zeta-\zeta t) - \frac{\alpha}{\zeta}(\cosh(\zeta-\zeta t)-1)\right)\right. \\ &\qquad \left. + \frac{\sinh(\zeta-\zeta t)}{\sinh(\zeta)}\left(\sinh(\zeta t) + \frac{\alpha}{\zeta}(\cosh(\zeta t)-1)\right)\right\} \\ &\le \frac{q\delta}{r\zeta}\|z^\varepsilon\|_{L_\infty}\frac{2e^\zeta}{e^\zeta - e^{-\zeta}}. \end{aligned}$$

Moreover we can conclude

$$|x_0^\varepsilon| \leq \|\frac{1}{\delta \sinh(\zeta)} f(1,.)\|_{L_1} \|z^\varepsilon\|_{L_\infty} < \frac{2}{|\delta|} \|z^\varepsilon\|_{L_\infty}.$$

q.e.d.

For the numerical approach not only the map $z \longmapsto (x_0, w)$ is of interest but also the properties of the discretization matrix corresponding to the equation (32), respectively to (33). Let us define the symmetric bilinear form $a : H^1 \times H^1 \to \mathbb{R}$

$$a(\xi, x) \;=\; r\langle \dot{\xi} - \alpha\xi, \dot{x} - \alpha x\rangle + q\delta^2 \langle \xi, x\rangle. \tag{37}$$

Obviously, a is positive definite if $r, q\delta^2 > 0$. Given the operator $\boldsymbol{\mathcal{S}} : H^1 \longrightarrow (H^1)'$ defined as $(\xi, \boldsymbol{\mathcal{S}} x) := a(\xi, x)$, the equation (32) is equivalent to $(\xi, \boldsymbol{\mathcal{S}} x) = q\delta\langle \xi, z\rangle + r\langle \dot{\xi}, u\rangle - r\alpha\langle \xi, u\rangle$. For any function $g \in (H^1)'$ we set $u = 0$ and $z = 1/(q\delta)g$. Then $x = \boldsymbol{\mathcal{S}}^{-1} g$ is given by (34) and (35). Furthermore, we obtain the following estimates for the norm of $\boldsymbol{\mathcal{S}}$ and its inverse. Here we employ the H^1–norm $\|x\|_{H^1}^2 = \|x\|_{L_2}^2 + \|\dot{x}\|_{L_2}^2$.

Theorem 4.2 *The operator $\boldsymbol{\mathcal{S}}$ is a linear isomorphism and we have:*

$$\min(4r, q\delta^2) \leq \quad \|\boldsymbol{\mathcal{S}}\|_{H^1 \to (H^1)'} \quad \leq 2r \max(1, \alpha^2) + q\delta^2, \tag{38}$$

$$\frac{c(\alpha)}{q\delta^2} \leq \quad \|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)' \to H^1} \quad \leq \max\left\{\frac{2}{r}, \frac{2\alpha^2 + 1}{q\delta^2}\right\}, \tag{39}$$

with

$$c(\alpha) \approx |\alpha|^{3/2} \sqrt{\frac{e^2 - 1}{2(e^2 + 1)}} \quad \textit{for large } |\alpha|. \tag{40}$$

Hence cond$(\boldsymbol{\mathcal{S}}) = \|\boldsymbol{\mathcal{S}}\|_{H^1 \to (H^1)'} \|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)' \to H^1}$ is bounded but tends to infinity with $|\alpha| \to \infty$ for fixed regularization parameters r and q.

Proof: The bilinear form a is continuous and coercive since

$$a(\xi, x) \leq r\|\dot{\xi} - \alpha\xi\|_{L_2} \|\dot{x} - \alpha x\|_{L_2} + q\delta^2 \|\xi\|_{L_2} \|x\|_{L_2} \leq C_0 \|\xi\|_{H^1} \|x\|_{H^1}$$

with $C_0 = 2r \max(1, \alpha^2) + q\delta^2$ and the parallelogram identity yields

$$\begin{aligned} c_0 \|x\|_{H^1}^2 \;&=\; c_0 \|\dot{x} - \alpha x + \alpha x\|_{L_2}^2 + c_0 \|x\|_{L_2}^2 \\ &\leq\; 2c_0 \|\dot{x} - \alpha x\|_{L_2}^2 + c_0(2\alpha^2 + 1)\|x\|_{L_2}^2 \leq a(x, x) \end{aligned}$$

with $c_0 = \min\left\{\frac{r}{2}, \frac{q\delta^2}{2\alpha^2+1}\right\}$. Hence, due to the theorem of Lax Milgram the operator $\boldsymbol{\mathcal{S}} : H^1 \longrightarrow (H^1)'$ is a linear isomorphism and we have the upper bounds of (38) and (39).

For the lower bound in (38) we set $\bar{x} = \exp(-\alpha t)$. Then

$$\begin{aligned} \|\boldsymbol{\mathcal{S}}\|_{H^1 \to (H^1)'} &\geq \|\boldsymbol{\mathcal{S}}\bar{x}\|_{(H^1)'} / \|\bar{x}\|_{H^1} \geq a(\bar{x}, \bar{x}) / \|\bar{x}\|_{H^1}^2 = \\ &= (4r\|\dot{\bar{x}}\|_{L_2}^2 + q\delta^2 \|\bar{x}\|_{L_2}^2) / \|\bar{x}\|_{H^1}^2 \geq \min(4r, q\delta^2). \end{aligned}$$

For the lower bound in (39) we consider the function $\bar{x} = \exp(\alpha t)$. Obviously it holds $\boldsymbol{\mathcal{S}}\bar{x} = q\delta^2 \bar{x}$ and hence $\|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)' \to H^1} \geq \frac{c(\alpha)}{q\delta^2}$ with $c(\alpha) := \|\bar{x}\|_{H^1} / \|\bar{x}\|_{(H^1)'}$.

The involved norms of $\bar{x}$ are:

$$\begin{aligned} \|\bar{x}\|_{H^1}^2 &= (1+\alpha^2)\|\bar{x}\|_{L_2}^2 = \frac{(1+\alpha^2)}{(2\alpha)}(e^{2\alpha} - 1), \\ \|\bar{x}\|_{(H^1)'}^2 &= \frac{(2\alpha+1)(\alpha-1)^2(e^{2\alpha}-1)(e^1 - e^{-1}) + 4\alpha^3 e^{-1}(e^\alpha - e^1)^2}{2\alpha(\alpha+1)^2(\alpha-1)^2(e^1 - e^{-1})}. \end{aligned}$$

Here $\|\bar{x}\|_{(H^1)'} = \sup_{\|v\|_{H^1}^2 = 1} \int_0^1 v(t) e^{\alpha t} dt$ is determined by the Langrange technique for optimization problems.
The function v has to fulfill $\ddot{v} - v = -\frac{1}{2k} e^{\alpha t}$ with $\dot{v}(0) = \dot{v}(1) = 0$, which is true for $v(t) = \frac{1}{2k(\alpha^2-1)} \frac{1}{e^1 - e^{-1}} \{\alpha e^t (e^\alpha - e^{-1}) + \alpha e^{-t}(e^\alpha - e^1) - e^{t\alpha}(e^1 - e^1)\}$. Furthermore, k is given by $\|v\|_{H^1}^2 = 1$. Then $\|\bar{x}\|_{(H^1)'} = 2k\langle v, v - \ddot{v}\rangle = 2k\|v\|_{H^1}^2 = 2k$. For large $|\alpha|$ we conclude $c(\alpha) \approx |\alpha|^{3/2} \sqrt{\frac{e^2-1}{2(e^2+1)}}$. q.e.d.

This indicates clearly that regularization with r is sufficient to obtain a bounded inverse $\boldsymbol{\mathcal{S}}^{-1}$. However, for large $|\alpha|$ the norm is still very large independently of r.

Obviously $\exp(\alpha t)$ is an eigenfunction for $\boldsymbol{\mathcal{S}}$ with the corresponding eigenvalue $q\delta^2$. The next Lemma gives more information about the eigenvalues and eigenfunctions of $\boldsymbol{\mathcal{S}}$.

Lemma 4.1 *The operator $\boldsymbol{\mathcal{S}}$ has infinite, countable many eigenvalues $\hat{\lambda}_i$ and $\hat{\lambda}_i$ behaves like i^2 for large i. The corresponding eigenfunctions $\hat{x}_i$ have i zeros in $[0,1]$. Considering the operator $\boldsymbol{\mathcal{S}}$ for $r = 1$ and $q\delta^2 = 0$ with its eigenvalues $\tilde{\lambda}_i(\alpha)$, we have $\hat{\lambda}_i = r\tilde{\lambda}_i(\alpha) + q\delta^2$ and the eigenfunctions $\hat{x}_i$ depend only on α.*

The smallest eigenvalue is $\hat{\lambda}_0 = q\delta^2$ and the eigenfunction is given by $\hat{x}_0 = \bar{x} = \exp(\alpha t)$.

Proof: The operator $\boldsymbol{\mathcal{S}}$ is defined correspondingly to (33). Hence we obtain the Sturm–Liouville's eigenvalue problem (see e.g. [21]) $r\ddot{x} - (r\alpha^2 + q\delta^2)\, x + \lambda x = 0$ with $\dot{x}(1) - \alpha x(1) = \dot{x}(0) - \alpha x(0) = 0$. This yields the first statements.

Since $\bar{x}$ has no zeros, we can conclude that it is the eigenfunction to the smallest eigenvalue $\hat{\lambda}_0$, which therefore equals $q\delta^2$. q.e.d.

Consequently, the parameters q and δ determine the smallest eigenvalue, r controls the gap between the eigenvalues and α determines the behaviour of the eigenvalues

for growing i and gives the corresponding eigenfunctions. Moreover, $\exp(\alpha t)$ is the eigenfunction to the smallest eigenvalue. This is one essential ingredient to construct an efficient preconditioner of the corresponding discretized system (see [5]) and leads to a drastically reduced condition number of the discretized system.

4.2 Linear model with several state functions

In case of a linear model with constant coefficients with several state functions we define correspondingly to (32) and (37) the symmetric bilinear form

$$a(\boldsymbol{\zeta}, \boldsymbol{x}) \;=\; \langle \dot{\boldsymbol{\zeta}} - \boldsymbol{A}\boldsymbol{\zeta}, \boldsymbol{R}^{-1}(\dot{\boldsymbol{x}} - \boldsymbol{A}\boldsymbol{x})\rangle + \langle \boldsymbol{C}\boldsymbol{\zeta}, \boldsymbol{Q}\boldsymbol{C}\boldsymbol{x}\rangle \tag{41}$$

and the operator $\boldsymbol{\mathcal{S}}$ given by $(\boldsymbol{\zeta}, \boldsymbol{\mathcal{S}}\boldsymbol{x}) := a(\boldsymbol{\zeta}, \boldsymbol{x})$. We cannot extend in a straight forward way the results concerning the eigenvalues. However, we can extend some ideas of Theorem 4.2 taking into account the observability of the system.

Proposition 4.1 *The symmetric bilinear form $a(\boldsymbol{\zeta}, \boldsymbol{x})$ is continuous and coercive.*

Proof: We obtain $a(\boldsymbol{\zeta}, \boldsymbol{x}) \le \left(2\|\boldsymbol{R}\|^{-1}\max(1, \|\boldsymbol{A}\|^2) + \|\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\|\right)\|\boldsymbol{\zeta}\|_{H^1}\|\boldsymbol{x}\|_{H^1}$ as in Theorem 4.2 which yields continuity. To show that a is coercive we first proof positive definiteness. Given $a(\boldsymbol{x}, \boldsymbol{x}) = 0$ then $\boldsymbol{x}$ is the solution of the system $\dot{\boldsymbol{x}} - \boldsymbol{A}\boldsymbol{x} \equiv \boldsymbol{0}$, $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $(\boldsymbol{y} \equiv)\boldsymbol{C}\boldsymbol{x} \equiv \boldsymbol{0}$. Since the system is observable $\boldsymbol{y} \equiv \boldsymbol{0}$ yields $\boldsymbol{x}_0 = \boldsymbol{0}$ and consequently $\boldsymbol{x} \equiv \boldsymbol{0}$. Hence, $a(\boldsymbol{x}, \boldsymbol{x}) > 0$ for $\boldsymbol{x} \not\equiv 0$. Assuming that a is not coercive, there exists a sequence $\{\boldsymbol{x}_n\} \in H^1$ with $\|\boldsymbol{x}_n\|_{H^1} = 1$ such that $\lim_{n\to\infty} a(\boldsymbol{x}_n, \boldsymbol{x}_n) = 0$. Since $\{\boldsymbol{x}_n\}$ is bounded it has a subsequence which we call for simplicity again $\boldsymbol{x}_n$ and which is weakly convergent in H^1, strongly convergent in L_2 to $\boldsymbol{0} \not\equiv \boldsymbol{x} \in H^1$ and $\|\dot{\boldsymbol{x}}\|_{L_2} \le \lim_{n\to\infty}\|\dot{\boldsymbol{x}}_n\|_{L_2}$. Hence, we obtain the contradiction $0 = \lim_{n\to\infty} a(\boldsymbol{x}_n, \boldsymbol{x}_n) \ge a(\boldsymbol{x}, \boldsymbol{x}) > 0$ and coercivity holds. q.e.d.

Corollary 4.1 *For any $\boldsymbol{z}, \boldsymbol{u} \in L_2$ the solution $\boldsymbol{x}$ of (32) determines the unique solution of the minimization problem (30).*

For the corresponding operator $\boldsymbol{\mathcal{S}}$ and its inverse we obtain the following specific estimates depending on the model matrices and the regularization matrices.

Theorem 4.3 *The operator $\boldsymbol{\mathcal{S}} : H^1 \to (H^1)'$ is a linear isomorphism and has a bounded inverse. Furthermore, we have*

$$\|\boldsymbol{\mathcal{S}}\|_{H^1\to(H^1)'} \le 2\|R\|^{-1}\max(1, \|\boldsymbol{A}\|^2) + \|\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\|, \tag{42}$$

and

$$\max\{k_S(\boldsymbol{v}) | \boldsymbol{v} \text{ eigenvector of } \boldsymbol{A} \text{ with } \|v\|_{l_2} = 1\} \;\le\; \|\boldsymbol{\mathcal{S}}\|_{H^1\to(H^1)'}, \tag{43}$$

$$\max\{c(\alpha)k_{S^{-1}}(\boldsymbol{v}) | \alpha, \boldsymbol{v} \text{ with } \boldsymbol{A}\boldsymbol{v} = \alpha\boldsymbol{v} \text{ with } \|v\|_{l_2} = 1\} \;\le\; \|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)'\to H^1}, \tag{44}$$

with

$$k_S(\boldsymbol{v}) = \min\{4(\boldsymbol{v}^T\boldsymbol{R}^{-1}\boldsymbol{v}), (\boldsymbol{v}^t\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v})\} > 0, \tag{45}$$

$$\infty > k_{S^{-1}}(\boldsymbol{v}) = 1/\|\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v}\|_{l_2} > 1/\|\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\| > 0 \tag{46}$$

and $c(\alpha)$ given as in Theorem 4.2.

If the spectral radius $\rho(\boldsymbol{A})$ of $\boldsymbol{A}$ is larger than 1, we can choose

$$k_S(\boldsymbol{v}) = 2\boldsymbol{v}^T\boldsymbol{R}^{-1}\boldsymbol{v} \geq 2/\|\boldsymbol{R}\|. \tag{47}$$

For any fixed regularization parameters $\boldsymbol{R}$ and $\boldsymbol{Q}$, $cond(\boldsymbol{\mathcal{S}}) \geq c(\alpha)k_S(\boldsymbol{v})k_{S^{-1}}(\boldsymbol{v})$ can be arbitrarily large, if the spectral radius $\rho(\boldsymbol{A})$ of $\boldsymbol{A}$ is large or if there is an eigenvector $\boldsymbol{v}$ of $\boldsymbol{A}$ which is close to the null space of $\boldsymbol{C}$.

Proof: Due to the Riesz representation theorem and Proposition 4.1 $\boldsymbol{\mathcal{S}}$ is an linear isomorphism with an bounded inverse and the estimation (42) holds. For the other estimates let α be an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{v}$ a normalized corresponding eigenvector. Then, defining $\bar{\boldsymbol{x}} = exp(-\alpha t)\boldsymbol{v}$ we obtain

$$a(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}) = 4\|\dot{\bar{\boldsymbol{x}}}\|^2_{L_2}(\boldsymbol{v}^T\boldsymbol{R}^{-1}\boldsymbol{v}) + \|\bar{\boldsymbol{x}}\|^2_{L_2}(\boldsymbol{v}^t\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v}).$$

Following the steps in Theorem 4.2 we have $k_S(\boldsymbol{v}) < \|\boldsymbol{\mathcal{S}}\|_{H^1\to(H^1)'}$ for any eigenvector $\boldsymbol{v}$. To show $k_S(\boldsymbol{v}) > 0$ we use the observability condition: We consider the model equations $\dot{\boldsymbol{x}} - \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$, which is required to be observable. Hence $\boldsymbol{x}_0 \mapsto \boldsymbol{y}$ is injective. Therefore defining $\boldsymbol{x}_0 := \boldsymbol{v} \neq \boldsymbol{0}$ we follow $exp(\alpha t)\boldsymbol{C}\boldsymbol{v} = \boldsymbol{y} \not\equiv \boldsymbol{0}$ and thus $\boldsymbol{C}\boldsymbol{v} \neq 0$. Then $\boldsymbol{v}^t\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v} = \|\boldsymbol{Q}^{1/2}\boldsymbol{C}\boldsymbol{v}\|^2 \geq \|\boldsymbol{C}\boldsymbol{v}\|^2/\|\boldsymbol{Q}^{-1}\| > 0$ and $k_S(\boldsymbol{v}) > 0$. Since $\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v} \neq 0$ we also have $\infty > k_{S^{-1}}(\boldsymbol{v})$.

If the spectral radius of $\boldsymbol{A}$ is larger or equal than 1, there exists an eigenvalue $\alpha > 1$ and we can estimate

$$\|\boldsymbol{\mathcal{S}}\|_{H^1\to(H^1)'} \geq \frac{a(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}})}{\|\bar{\boldsymbol{x}}\|^2_{H^1}} = \frac{4\alpha^2(\boldsymbol{v}^T\boldsymbol{R}^{-1}\boldsymbol{v}) + (\boldsymbol{v}^t\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v})}{(1+\alpha^2)} \geq 2\boldsymbol{v}^T\boldsymbol{R}^{-1}\boldsymbol{v}$$

which gives (47).

For estimation (44) we define $\bar{\boldsymbol{x}} = \exp(\alpha t)\boldsymbol{v}$, which yields $(\boldsymbol{\xi}, \boldsymbol{\mathcal{S}}\bar{\boldsymbol{x}}) = \langle\boldsymbol{\xi}, \boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\bar{\boldsymbol{x}}\rangle$. Thus we can follow for any $\alpha, \boldsymbol{v}$

$$\frac{1}{\|\boldsymbol{C}^T\boldsymbol{Q}\boldsymbol{C}\boldsymbol{v}\|_{l^2}}\, c(\alpha) \leq \|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)'\to H^1}.$$

q.e.d.

While the boundedness of the inverse $\boldsymbol{\mathcal{S}}^{-1}$ is known, a specified upper bound is missing. However, in an analogous manner to the case of one state function we obtain:

Proposition 4.2 *For an invertible matrix $\boldsymbol{C}$ the inverse operator $\boldsymbol{\mathcal{S}}^{-1}$ is bounded by*

$$\|\boldsymbol{\mathcal{S}}^{-1}\|_{(H^1)'\to H^1} \leq \max\{2\|\boldsymbol{R}\|, (2\|\boldsymbol{A}\|^2+1)\|\boldsymbol{Q}^{-1}\|\|\boldsymbol{C}^{-1}\|^2\}.$$

Due to the boundedness of the inverse and the compact inbedding $H^1 \hookrightarrow C^0$ we have:

Corollary 4.2 *The regularized problem formulation (9) with time independent, linear ODE's as constraints and no inequality constraints is well posed, i.e.* $\max\{|x_0|, \|\boldsymbol{w}\|_{L_2}\} \leq c\|\boldsymbol{z}\|_{L_2}$. *Furthermore we have* $\|\boldsymbol{x}\|_{C^0} \leq c\|\boldsymbol{z}\|_{L_2} \leq c\|\boldsymbol{z}\|_{L_\infty}$ *with a generic constant c.*

As for one state only, Theorem 4.3 shows that independent of the regularization of $\boldsymbol{w}$ we face for a large spectral radius $\rho(\boldsymbol{A})$ a high condition number. This influences, for example, strongly the condition number relevant for the efficiency of the iterative solver for the discretized system corresponding to (32), see for example [3, 5], and the accuracy of the solution with respect to perturbations.

To be more precise concerning the constants involved in the estimates in [3], there we considered the not completely reduced system for $\boldsymbol{x}$, $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}} = -\boldsymbol{\lambda}_2$ (see (10)) which yields the saddle point problem

$$\left(\boldsymbol{\nu}, \tilde{\mathcal{L}}\begin{pmatrix}\boldsymbol{\lambda}\\ \hat{\boldsymbol{\lambda}}\\ \boldsymbol{x}\end{pmatrix}\right) := \begin{pmatrix}\langle \boldsymbol{Q}^{-1}\hat{\boldsymbol{\lambda}} + \boldsymbol{C}\boldsymbol{x}, \boldsymbol{\nu}_1\rangle\\ \langle \boldsymbol{R}\boldsymbol{\lambda} - \dot{\boldsymbol{x}} + \boldsymbol{A}\boldsymbol{x}, \boldsymbol{\nu}_2\rangle\\ \langle \boldsymbol{\lambda}, -\dot{\boldsymbol{\nu}}_3 + \boldsymbol{A}\boldsymbol{\nu}_3\rangle + \langle \boldsymbol{C}^T\hat{\boldsymbol{\lambda}}, \boldsymbol{\nu}_3\rangle\end{pmatrix} = \langle \begin{pmatrix}\boldsymbol{z}\\ -\boldsymbol{u}\\ 0\end{pmatrix}, \boldsymbol{\nu}\rangle. \quad (48)$$

The first equation reflects (10) together with (13). The relevant constants are c_7 and c_8 with

$$c_7(\|\boldsymbol{\lambda}\|_{L_2} + \|\hat{\boldsymbol{\lambda}}\|_{L_2} + \|\boldsymbol{x}\|_{H^1}) \leq \|\tilde{\mathcal{L}}(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}, \boldsymbol{x})\|_{L_2\times L_2\times (H^1)'} \leq c_8(\|\boldsymbol{\lambda}\|_{L_2} + \|\hat{\boldsymbol{\lambda}}\|_{L_2} + \|\boldsymbol{x}\|_{H^1})$$

(see [3] equation (36)).

For $\bar{\boldsymbol{x}} = \exp(\alpha t)\boldsymbol{v}$ we have $\|\tilde{\mathcal{L}}(\boldsymbol{0}, \boldsymbol{0}, \bar{\boldsymbol{x}})\|_{L_2\times L_2\times (H^1)'} = \|\boldsymbol{C}\boldsymbol{v}\|\|\exp(\alpha t)\|_{L_2}$. Hence we can follow

$$c_7 \leq \|\boldsymbol{C}\boldsymbol{v}\|\|\exp(\alpha t)\|_{L_2}/\|\exp(\alpha t)\|_{H^1} = \|\boldsymbol{C}\boldsymbol{v}\|/\sqrt{1+\alpha^2}.$$

For $\bar{\boldsymbol{x}} = \exp(-\alpha t)\boldsymbol{v}$ we have $\|\tilde{\mathcal{L}}(\boldsymbol{0}, \boldsymbol{0}, \bar{\boldsymbol{x}})\|^2_{L_2\times L_2\times (H^1)'} = (\|\boldsymbol{C}\boldsymbol{v}\|^2 + 4\alpha^2)\|\exp(-\alpha t)\|_{L_2}$ and

$$c_8 \geq \sqrt{\|\boldsymbol{C}\boldsymbol{v}\|^2 + 4\alpha^2}/\sqrt{1+\alpha^2}.$$

Consequently

$$cond(\tilde{\mathcal{L}}) = c_8/c_7 \geq \sqrt{1 + 4\frac{\alpha^2}{\|\boldsymbol{C}\boldsymbol{v}\|^2}}$$

for all eigenvalues α and corresponding eigenvectors $\boldsymbol{v}$. Hence independent of the regularization parameters $cond(\tilde{\mathcal{L}})$ tends to infinity if the spectral radius of $\boldsymbol{A}$ tends to infinity or if there is an eigenvector which tends to the null space of $\boldsymbol{C}$.

Finally, we would like to bring to the readers attention the influence of a long time horizon. As mentioned in the beginning state estimation uses measurements of a past time horizon $[0, T]$, which is then projected onto $[0, 1]$. Given a fixed model equation $\dot{\boldsymbol{x}}(t) - \boldsymbol{A}\boldsymbol{x}(t) - \boldsymbol{u}(t) = \boldsymbol{0}$ for $t \in [0, T]$ the projection onto $[0, 1]$ leads to $\dot{\tilde{\boldsymbol{x}}}(s) - T\boldsymbol{A}\tilde{\boldsymbol{x}}(s) - \tilde{\boldsymbol{u}}(s) = \boldsymbol{0}$ for $s \in [0, 1]$. Hence in the estimation for $cond(\boldsymbol{\mathcal{S}})$ the spectral radius $T\rho(\boldsymbol{A})$ is relevant.

Remark 4.1 *While long horizons may be favourable for stability reasons concerning the confidence into the obtained measurements, from the point of view of the resulting system long horizons should be avoided, since they increase the condition number.*

Some first approaches for the choice of an appropriate length of the time horizon are known under stochastical assumptions on the underlying measurement errors and for linear systems [15, 17, 19]. However this is still a fairly open research problem, in particular if stochastical assumptions cannot be assured in the applications or nonlinearity occurs. The above remark ensures the importance of closing this gap and illuminates an additional issue which should to be considered for the choice of the horizon length.

5 Conclusion

The formulation of state estimation as an optimization problem including possible model error functions but without regularization of the initial data offers solutions without bias. It turned out that the inclusion of model error functions into the output equations is not necessary for the computation process. In this way the otherwise difficult task of choosing an appropriate regularization parameter for these model error function can be avoided. Moreover, one can reduce the resulting optimization problem without disadvantage to the unknown states and the model error functions in the state equations.

For linear ODE's with time independent coefficients and without inequality constraints it is shown that the condition numbers of the evolving operators can be arbitrarily large independently of the chosen regularization parameters, if the spectral radius of the system matrix is large or an eigenvector of the system matrix is close to the null space of the matrix, which depicts some states as measurable. These condition numbers influence for example essentially the numerical solution process. Furthermore, in the one dimensional case the eigenfunction to the smallest eigenvalue of the operator is derived, which is one important ingredient to obtain an efficient preconditioner for the numerical approach in [5].

References

[1] F. Allgöwer, T.A. Badgwell, J.S. Qin, J.B. Rawlings, and S.J. Wright. Nonlinear predictive control and moving horizon estimation – an introductory overview. In P.M. Frank, editor, *Advances in Control: Highlights of ECC 1999*, pages 391–449. Springer–Verlag, 1999.

[2] T. Binder, L. Blank, W. Dahmen, and W. Marquardt. Towards multiscale dynamic and data reconciliation. In R. Berber and C. Kravaris, editors, *Nonlinear Model Based Process Control, NATO ASI*, pages 623–665. Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.

[3] T. Binder, L. Blank, W. Dahmen, and W. Marquardt. Iterative algorithms for multiscale state estimation, part 1: Concept. *J. Optim. Theory Appl.*, 111(3):501–527, 2001.

[4] T. Binder, L. Blank, W. Dahmen, and W. Marquardt. On the regularization of dynamic data reconciliation problems. *Journal of Process Control*, 12:557–567, 2002.

[5] L. Blank. Preconditioning via a Schur complement method: an application in state estimation. to be published in SIAM J. Sci. Comput., 2003.

[6] K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. SIAM, Philadelphia, 1996.

[7] M.J. Corless and A.E. Frazho. *Linear Systems and Control, An Operator Perspective*. Marcel Dekker, Inc., New York, Basel, 2003.

[8] R.F. Curtain and H. Zwart. *An introduction to infinite–dimensional linear systems theory*. Springer–Verlag, New York, 1995.

[9] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, The Netherlands, 1996.

[10] R.F. Hartl, S.P. Sethi, and R.G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, 37(2):181–218, 1995.

[11] M.A. Henson and D.E. Seborg. *Nonlinear Process Control*. Prentice Hall, Upper Saddle River, 1997.

[12] M.R. Hestenes. *Calculus of Variations and Optimal Control Theory*. John Wiley & Sons, Inc., New York, 1966.

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

[14] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems.* Springer–Verlag, 1996.

[15] H. Michalska and D.Q. Mayne. Moving horizon observers and observer–based control. *IEEE Trans. Aut. Contr.*, 40(6):995–1006, 1995.

[16] K.R. Muske and T.F. Edgar. Nonlinear state estimation. In M.A. Henson and D.E. Seborg, editors, *Nonlinear Process Control*, pages 311–711. Prentice Hall, 1997.

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

[18] L.S. Pontryagin, V.G. Boltyanski, R.V. Gamkrelidze, and E.F. Miscenko. *The Mathematical Theory of Optimal Processes.* Wiley, 1962.

[19] C.V. Rao and J.B. Rawlings. Constrained process monitoring: Moving–horizon approach. *AIChE J.*, 48(1):97–109, 2002.

[20] 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.

[21] W. Walter. *Gewöhnliche Differentialgleichungen.* Springer–Verlag, Berlin, 1976.