V-invariant methods, generalised least squares problems, and the Kalman filter

M. R. Osborne[*]   Inge Söderkvist[†]

Abstract

V-invariant methods for the generalised least squares problem extend the techniques based on orthogonal factorization for ordinary least squares to problems with multiscaled, even singular covariances. These methods are summarised briefly here, and the ability to handle multiple scales indicated. An application to a class of Kalman filter problems derived from generalised smoothing splines is considered. Evidence of severe illconditioning of the covariance matrices is demonstrated in several examples. This suggests that this is an appropriate application for the V-invariant techniques.

[*]School of Mathematical Sciences, ANU mailto:mike.osborne@maths.anu.edu.au
[†]Department of Mathematics, Luleå University of Technology

Contents

1 Introduction

The generalised least squares problem is

$$\min_{\mathbf{x}} \mathbf{r}^T V^{-1} \mathbf{r} \, ; \quad \mathbf{r} = A\mathbf{x} - \mathbf{b} \, , \tag{1}$$

where $A : R^p \to R^n$, $V : R^n \to R^n$. It will be assumed that A has its full rank $p < n$, but only that V is positive semi-definite. Typically, in data analytic situations, the covariance matrix V has the dimension n of the data set and is often large, while p, the number of independent variables in the fitted model, is relatively small. Our aim here is to present an application of V-invariant methods appropriate to a class of Kalman filter problems which have distinctly illconditioned covariances and both stable and unstable dynamics. This application forces well defined sparse structures on both the design A and the covariance V.

A class of V-invariant algorithms has been introduced by Gulliksson and Wedin [3]. Here we follow their basic development closely. Their problem of particular interest is equality constrained least squares which can be expressed formally as a generalised least squares problem with singular (and

diagonal) V. This class of problems provide a particular example of the ability of V-invariant algorithms to support a form of multi-scaling. However, there are some costs, and they point out the importance of column pivoting in this application. This proves to be collateral damage occasioned by the multiple scale capability.

Söderkvist [9] considered the Kalman Filter in the generalised least squares form [1] for the particular case of a diagonal covariance matrix V and was able to demonstrate superior numerical performance of his V-invariant methods for problems in which V possessed several distinct scales. The restriction to diagonal V is important in developing algorithms, and he has experimented with methods for reducing the problem covariance to one having this diagonal form employing both an eigenvalue decomposition using Jacobi's method ($V = Q\Lambda Q^T$) and a rank revealing Cholesky factorization with diagonal pivoting ($PVP^T = LDL^T$) [10]. In this latter case he was concerned about the influence of possible errors in small elements of D. Recently, Osborne [7] has argued that provided the number of small elements in D is $k \leq p$ then their effect is benign so that errors in them due to the rank revealing factorization are insignificant.

The plan of this paper is as follows. First the main results of [3] are summarised and an approach to deriving stability results in the case of ill-conditioned V outlined. A class of Kalman filter problems which can occur in computing smoothing splines and related objects is sketched. The structure of the generalised least squares form of this problem is then exploited in implementing V-invariant factorization algorithms. Finally, results showing that the relevant covariances can be very illconditioned are presented.

2 V-invariant transformation

We say that the transformation matrix $J : R^n \to R^n$ is V-invariant if

$$JVJ^T = V. \tag{2}$$

Let J_1 and J_2 be V-invariant. Then

- J_1^{-1}, J_2^{-1}, $J_1 J_2$ and $J_2 J_1$ are V-invariant,
- J_1^T, J_2^T are V^{-1}-invariant (V nonsingular).

If V is singular then it is assumed that it has the reduced form

$$V = \begin{bmatrix} 0 & 0 \\ 0 & V_2 \end{bmatrix}. \tag{3}$$

In this case J is V-invariant iff

$$J = \begin{bmatrix} J_{11} & 0 \\ J_{21} & J_{22} \end{bmatrix}, \quad J_{22} V_2 J_{22}^T = V_2,$$

and J_{11}, J_{22} are nonsingular.

The ordinary least squares problem provides the best known example of V-invariance. Here J is orthogonal and the invariance condition becomes:

$$V = I \Rightarrow JIJ^T = I.$$

The V-invariant analogue of the Aitken-Householder elementary orthogonal transformation is the elementary reflector

$$J = I - 2\frac{V\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T V \mathbf{v}}, \quad J^2 = I. \tag{4}$$

To use (4) in matrix reduction to upper triangular form let $\mathbf{u}$ stand for the current pivotal column in the partially reduced matrix and $\mathbf{u}_2$ for the subvector to be reduced. Then $\mathbf{v}$ defining the current transformation step must be found such that

$$J \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \gamma \mathbf{e}_1 \end{bmatrix} .$$

It follows from (4) that the scale of $\mathbf{v}$ is not important, and it is convenient to set

$$V\mathbf{v} = s \begin{bmatrix} 0 \\ \mathbf{u}_2 - \gamma \mathbf{e}_1 \end{bmatrix} , \tag{5}$$

where s is a scale factor. The computation of $\mathbf{v}$ in a standard reduction to upper triangular form requires solution of this equation. However, it has a complexity comparable with the original problem unless V is readily invertible! The important special case corresponds to $V = D$ diagonal with elements ordered in increasing magnitude. To complete the specification of $\mathbf{v}$ in this case let $\dim(\mathbf{u}_1) = j - 1$,

$$D = \mathrm{diag}\{d_j,\ j = 1, \ldots, n\} = \mathrm{diag}\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k, \nu_{k+1}, \ldots, \nu_n\} ,$$
$$\varepsilon_1 \le \varepsilon_2 \le \cdots \le \varepsilon_k \ll \nu_{k+1} \le \cdots \le \nu_n ,$$
$$\mathbf{v} = \begin{bmatrix} 0 \\ \mathbf{v}_2 \end{bmatrix} = D_j \begin{bmatrix} 0 \\ D_2^{-1}(\mathbf{u}_2 - \gamma\mathbf{e}_1) \end{bmatrix} .$$

Here $s = d_j$ and the specified ordering results in the effective diagonal matrix $N_j = d_j \mathrm{diag}\{d_j, d_{j+1}, \ldots, d_n\}^{-1}$ having elements ≤ 1. The importance of the ordering of the elements of D is shown also in the calculation of γ. This computation uses V-invariance:

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}^T J^T D^{-1} J \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \gamma\mathbf{e}_1 \end{bmatrix}^T D^{-1} \begin{bmatrix} \mathbf{u}_1 \\ \gamma\mathbf{e}_1 \end{bmatrix}$$
$$\Rightarrow \mathbf{u}_2^T D_2^{-1} \mathbf{u}_2 = \gamma^2 \mathbf{e}_1^T D_2^{-1} \mathbf{e}_1 . \tag{6}$$

There are two cases to consider in (6) depending on the relative values of j and k:

j ≤ k

$$\gamma^2 = (\mathbf{u}_2)_j^2 + \sum_{s=j+1}^{k} \frac{\varepsilon_j}{\varepsilon_s}(\mathbf{u}_2)_s^2 + \sum_{s=k+1}^{n} \frac{\varepsilon_j}{\nu_s}(\mathbf{u}_2)_s^2 \,, \text{ and}$$

j > k

$$\gamma^2 = (\mathbf{u}_2)_j^2 + \sum_{s=j+1}^{n} \frac{\nu_j}{\nu_s}(\mathbf{u}_2)_s^2 \,.$$

Note that there is multiple scale behavior when $j \leq k$ and that the limit $\boldsymbol{\varepsilon} \to 0$ can be defined, for example, by setting $\varepsilon_s/\varepsilon_j = 1$, $s > j$. Also γ is the column length in the re-scaled metric defined by N_j.

If elements of J are large then this is an indicator of possible stability problems [3]! Let

$$J = I - 2\mathbf{c}\mathbf{d}^T$$

be an elementary V-invariant reflector. Then

$$\|J\|_2 = \eta + \sqrt{\eta^2 - 1}\,, \quad \eta = \|\mathbf{c}\|_2 \|\mathbf{d}\|_2 \,.$$

Here

$$\eta = \frac{\left\| D_2^{-1}\left(\mathbf{u}_2 - \gamma\mathbf{e}_1\right)\right\| \left\|\mathbf{u}_2 - \gamma\mathbf{e}_1\right\|}{\left(\mathbf{u}_2 - \gamma\mathbf{e}_1\right)^T D_2^{-1}\left(\mathbf{u}_2 - \gamma\mathbf{e}_1\right)}\,,$$

$$\Rightarrow \|J\| \geq \eta \geq \frac{\|\mathbf{u}_2\|}{2\gamma} \,.$$

That is $\|J_j\|$ will be large if

$$\left| D_j \mathbf{u}_2^T D_2^{-1}\mathbf{u}_2 \right| \ll \|\mathbf{u}_2\| \,.$$

The limit $\boldsymbol{\varepsilon} \to 0$ gives η large if

$$\|\mathbf{u}_\varepsilon\| \ll \|\mathbf{u}_2\| \,. \tag{7}$$

Here $\mathbf{u}_\varepsilon$ corresponds to $\{\varepsilon_j, \ldots, \varepsilon_k\}$ in D. The role of column pivoting is important in avoiding just this case.

Example 1 The inequality (7) is a good guide to instability in V-invariant factorizations. Let

$$D = \begin{bmatrix} \varepsilon & & \\ & \varepsilon & \\ & & 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \alpha \\ \beta \\ \nu \end{bmatrix}.$$

The transformation taking $\mathbf{w}$ to $\gamma \mathbf{e}_1$ in the limit $\varepsilon \to 0$ is

$$I - \pi \left(\begin{bmatrix} \alpha \\ \beta \\ \nu \end{bmatrix} + \theta(\pi 1)\mathbf{e}_1 \right) \left(\begin{bmatrix} \alpha \\ \beta \\ 0 \end{bmatrix} + \theta(\pi 1)\mathbf{e}_1 \right)^T,$$

where

$$\pi = \frac{1}{\pi 1 \pi 2}, \quad \theta = \mathrm{sgn}(\alpha),$$
$$\pi 1 = (\alpha^2 + \beta^2)^{1/2},$$
$$\pi 2 = |\alpha| + (\alpha^2 + \beta^2)^{1/2}.$$

Note that the term $\pi \nu$ would be large if $\alpha, \beta \ll \nu$ violating the stability condition (7).

3   Solution of the GLSQ problem

In the spirit of the Gauss-Markov theorem write the solution of (1) as a linear function of the data $\mathbf{x} = T\mathbf{b}$ where T generates the best linear unbiassed estimator and satisfies

$$\begin{bmatrix} T & \Lambda \end{bmatrix} \begin{bmatrix} D & A \\ A^T & 0 \end{bmatrix} = \begin{bmatrix} 0 & I \end{bmatrix},$$

where $V = D$ diagonal. Let $JA = [R\ 0]^T$, and J V-invariant. Under this transformation the system becomes

$$
\left[\ \left[\ \widetilde{T}_1\ \ \widetilde{T}_2\ \right]\ \ \Lambda\ \right]
\left[\begin{array}{cc}
\left[\begin{array}{cc} D_\varepsilon & 0 \\ 0 & D_{21} \end{array}\right] & 0 \\
\left[\begin{array}{cc} & \\ & D_{22} \end{array}\right] & \\
\left[\ R^T\ \ 0\ \right] & 0
\end{array}\right]
\left[\begin{array}{c} R \\ 0 \\ 0 \end{array}\right]
$$
$$
= \left[\ 0\ \ I\ \right].
$$

The solution of the generalised least squares problem can now be written down. This gives:

$$
\widetilde{T}_1 = R^{-1}, \quad \widetilde{T}_2 = 0,
$$
$$
\Lambda = R^{-1} \left[\begin{array}{cc} D_\varepsilon & 0 \\ 0 & D_{21} \end{array}\right] R^{-T},
$$
$$
\mathbf{x} = \left[\ R^{-1}\ \ 0\ \right] J\mathbf{b}.
$$

The connection with the solution of the ordinary least squares problem will be recognised immediately.

If V is not diagonal then start with an LDL^T factorization of V and rewrite the problem (1) by setting $L^{-1}\mathbf{r} = \widetilde{\mathbf{r}} = D^{1/2}\mathbf{s}$ to obtain

$$
\min_{\mathbf{x}} \mathbf{s}^T \mathbf{s}; \quad D^{1/2}\mathbf{s} = L^{-1}A\mathbf{x} - L^{-1}\mathbf{b}.
$$

This has been implemented by making a rank-revealing Cholesky factorization of V:

$$
PVP^T \rightarrow L\mathrm{diag}\{d_n, d_{n-1}, \ldots, d_1\}\,L^T,
$$

where the the lower triangular matrix L has unit diagonal, and diagonal pivoting ensures both

$$
d_n \geq d_{n-1} \geq \cdots \geq d_1 \tag{8}
$$

and

$$
L_{ij} \leq 1, \quad i > j. \tag{9}
$$

Our code terminates if a sufficiently small (or negative) d_i is encountered. A key point is that any illconditioning in V is largely forced into D by this factorization as a consequence of the bound (9) on the subdiagonal elements of L. Conditions for success [4, e.g.] correspond to the assumptions made already on D. However, a further permutation step is needed to reverse the order of the elements in the computed D to stabilize the V-invariant transformation. This means the diagonal elements computed last in the Cholesky algorithm enter first into the factorization process. It would be expected that these small d_i could have high relative error. It is an important result that these errors are not significant. This is demonstrated in [7]. The result can be made plausible by considering the case

$$D = \operatorname{diag}\{0, \ldots, 0, d_{k+1}, \ldots, d_n\}, \quad k < p,$$

which gives the equality constrained problem

$$\min_{\mathbf{x}} \mathbf{s}^T \mathbf{s}; \quad \begin{bmatrix} 0 \\ & D_2^{1/2} \end{bmatrix} \mathbf{s} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}.$$

This is the limiting problem associated with the penalised objective

$$\min_{\mathbf{x}} \left\{ \mathbf{r}_2^T D_2^{-1} \mathbf{r}_2 + \lambda \mathbf{r}_1^T \mathbf{r}_1 \right\}; \quad \mathbf{r} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix},$$

which has the alternative form

$$\min_{\mathbf{x}} \mathbf{s}^T \mathbf{s}; \quad \begin{bmatrix} \lambda^{-1/2} I \\ & D_2^{1/2} \end{bmatrix} \mathbf{s} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}.$$

From the theory of penalty functions [2] expect

$$\|\mathbf{x}(\lambda) - \mathbf{x}(\infty)\| = O(1/\lambda), \quad \lambda \to \infty.$$

The argument is developed by showing that a certain Jacobian matrix corresponding to the $\lambda = \infty$ limit is nonsingular, and that the large λ case is a small perturbation of this.

4   g-splines and the Kalman filter

Generalised smoothing splines having the representation $\mathcal{E}\{\mathbf{h}^T\mathbf{x}(t) \mid y_1, \ldots, y_n\}$ can be defined by means of the class of stochastic differential equations:

$$d\mathbf{x} = M\mathbf{x}\, dt + \sigma\sqrt{\lambda}\mathbf{b}\, dw \,, \tag{10}$$

where $M : R^m \to R^m$ and w is a unit scale Wiener process, in conjunction with the observation process

$$\mathbf{h}^T\mathbf{x}(t_i) + \varepsilon_i = y_i \,, \quad \varepsilon_i \sim N(0, \sigma^2) \,, \tag{11}$$

where the errors ε_i are assumed independent. Here the vector $\mathbf{h}$ must satisfy identifiability conditions and, together with $\mathbf{b}$, determines the smoothness of the resulting spline [6]. Let the fundamental matrix $X(t, \xi)$ be defined by

$$\frac{dX}{dt} = MX \,, \quad X(\xi, \xi) = I \,.$$

Variation of parameters gives the discrete dynamics equation

$$\mathbf{x}_{i+1} = X_i\mathbf{x}_i + \sigma\sqrt{\lambda}\mathbf{u}_i \,, \tag{12}$$

where

$$\mathbf{u}_i = \int_{t_i}^{t_{i+1}} X(t_{i+1}, s)\mathbf{b}\frac{dw}{ds}\, ds \,,$$
$$\mathbf{u}_i \sim N(0, \sigma^2\lambda R_i)) \,.$$

This system—dynamics + observation equations—leads via the Kalman filter subject to a diffuse prior to the generalised least squares problem [1]

$$\min_{\mathbf{x}} \left\{\mathbf{r}_1^T R^{-1}\mathbf{r}_1 + \mathbf{r}_2^T V^{-1}\mathbf{r}_2\right\} \,, \tag{13}$$

where

$$
\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \begin{bmatrix} -X_1 & I & & & \\ & -X_2 & I & & \\ & & \ddots & & \\ & & & -X_{n-1} & I \\ \mathbf{h}^T & & & & \\ & \mathbf{h}^T & & & \\ & & \ddots & & \\ & & & & \mathbf{h}^T \end{bmatrix} \mathbf{x} - \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \tag{14}
$$

$$
R = \sigma^2 \lambda \operatorname{diag}\{R_1, R_2, \ldots, R_{n-1}\}, \quad V = \sigma^2 I, \tag{15}
$$

and

$$
R_i = \int_{t_i}^{t_{i+1}} X(t_{i+1}, s) \mathbf{b}\mathbf{b}^T X(t_{i+1}, s)^T \, ds. \tag{16}
$$

Paige and Saunders [8] suggested reducing this problem to an ordinary least squares problem by making a Cholesky factorization of the covariance blocks R_i. These are then used to rescale the dynamics equations to produce a problem with unit diagonal covariance that can be reduced by orthogonal transformations. This leads to their famous information filter. However, Sőderkvist [9] showed there are potential problems arising from illconditioned R_i. The alternative we consider is to use a V-invariant factorization starting with a rank revealing Cholesky applied to the R_i corresponding to successive column blocks, followed by within column block sorting of the diagonal elements d_i and division by potentially much better behaved lower triangular matrices L_i. Straight forward application of this procedure under the ordering given in (14) results in accumulating fill. This is seen readily by

considering the first few steps. These give:

$$
\begin{bmatrix}
-L_1^{-1}X_1 & L_1^{-1} & & \\
& -L_2^{-1}X_2 & -L_2^{-1} & \\
\vdots & \vdots & \vdots & \\
\mathbf{h}^T & & & \\
& \mathbf{h}^T & & \\
& & \mathbf{h}^T &
\end{bmatrix}
\rightarrow
\begin{bmatrix}
U_1 & W_1 & & \\
& -L_2^{-1}X_2 & L_2^{-1} & \\
\vdots & \vdots & \vdots & \\
& \mathbf{z}_{11}^T & & \\
& \mathbf{h}^T & & \\
& & & \mathbf{h}^T
\end{bmatrix}
$$

$$
\rightarrow
\begin{bmatrix}
U_1 & W_1 & \\
& U_2 & W_2 \\
\vdots & \vdots & \vdots \\
& & \mathbf{z}_{21}^T \\
& & \mathbf{z}_{22}^T \\
& & \mathbf{h}^T
\end{bmatrix}.
$$

The ordering used in the Paige and Saunders information filter generates less direct fill. However, fill can be controlled to a total of $m + 1$ rows in the next to pivotal block column by orthogonal transformations which are in this context V-invariant as the blocks affected are associated with covariances proportional to the unit matrix being derived from the second term in (13).

The orthogonal transformations are first applied at step m

$$
\begin{bmatrix}
\cdots & \cdots & \cdots \\
U_m & W_m & \\
& -L_{m+1}^{-1}X_{m+1} & L_{m+1}^{-1} \\
\cdots & \cdots & \cdots \\
& \mathbf{z}_{m1}^T & \\
& \mathbf{z}_{m2}^T & \\
& \vdots & \\
& \mathbf{z}_{mm}^T & \\
& \mathbf{h}^T &
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\cdots & \cdots & \cdots \\
U_m & W_m & \\
& -L_{m+1}^{-1}X_{m+1} & L_{m+1}^{-1} \\
\cdots & \cdots & \cdots \\
& Z_m & \\
& 0 &
\end{bmatrix},
$$

where $Z_m : R^m \to R^m$. It is convenient to carry out the transformation in two steps in order to compute auxiliary quantities such as the innovations $\mathbf{x}_{i|i-1}$. These are:

$$
\begin{bmatrix}
\mathbf{z}_{i1}^T \\
\vdots \\
\mathbf{z}_{im}^T \\
\mathbf{h}^T
\end{bmatrix}
\rightarrow
\begin{bmatrix}
Z_i^1 \\
\mathbf{h}^T
\end{bmatrix}
\rightarrow
\begin{bmatrix}
Z_i \\
0
\end{bmatrix}.
$$

Our initial implementation has proved satisfactory in early experiments. In particular, there has been good control over the magnitudes occurring. The information filter has been used previously in problems of this kind [5], but it is not able to cope with the severe illconditioning that can occur in the R_i matrices [9]. This illconditioning is directly related to the smoothness of the spline. This is illustrated in the following examples.

Example 2 Quintic splines. This corresponds to the case

$$
M = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad
\mathbf{h} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad
\mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},
$$

with **h** and **b** chosen for maximum smoothness. The covariance matrix blocks are readily computed:

$$R_i = \delta \begin{bmatrix} \frac{\delta^4}{20} & \frac{\delta^3}{8} & \frac{\delta^3}{6} \\ \frac{\delta^3}{8} & \frac{\delta^2}{3} & \frac{\delta}{2} \\ \frac{\delta^3}{6} & \frac{\delta}{2} & 1 \end{bmatrix}.$$

The rank revealing Cholesky gives

$$PR_iP^T = \delta \begin{bmatrix} 1 & & \\ \frac{\delta}{2} & 1 & \\ \frac{\delta^2}{6} & -\frac{\delta}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & \frac{\delta^2}{12} & \\ & & \frac{\delta^4}{720} \end{bmatrix} \begin{bmatrix} 1 & \frac{\delta}{2} & \frac{\delta^2}{6} \\ & 1 & -\frac{\delta}{2} \\ & & 1 \end{bmatrix}.$$

Note the small elements in the component blocks D_i of D. However, there are $(n-1)(m-1)$ of these all told while the design is $R^{nm} \to R^{nm+n}$ so the conditions for the solubility of the generalised least squares problem can be satisfied.

Example 3 Tension smoothing splines. Here the spline is constructed using exponentials rather than polynomials. This corresponds to an example with potentially unstable dynamics. For one and two parameter splines we have

$$M = \begin{bmatrix} 0 & 1 \\ \alpha^2 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ \alpha^2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \beta^2 & 0 \end{bmatrix}.$$

Smoothness is maximized by the choice $\mathbf{h} = \mathbf{e}_1$, $\mathbf{b} = \mathbf{e}_m$. Again the covariances are illconditioned leading to very small elements in the component blocks D_i. These particular tension spline examples are not very unstable, but instability does not appear to be the major driver for the small elements in D_i. The following table gives the diagonal elements produced by the rank revealing Cholesky decomposition for two different meshes on $0 \leq t \leq 1$. They are the same for each block of D.

TABLE 1: Tension splines provide a potentially unstable system

$\alpha = 1$	
$n = 11$	$D_i = \{8.3 - 5, 1.0 - 1\}$
$n = 51$	$D_i = \{6.7 - 7, 2.0 - 2\}$
$\alpha = 1, \quad \beta = 2$	
$n = 11$	$D_i = \{9.9 - 13, 1.4 - 8, 8.3 - 5, 1.0 - 1\}$
$n = 51$	$D_i = \{0.0, 4.4 - 12, 6.7 - 7, 2.0 - 2\}$

TABLE 2: A stable example from chemical kinetics

$k_1 = 1, \quad k_2 = 2$	
$n = 11$	$D_i = \{5.5 - 8, 6.8 - 5, 9.1 - 2\}$
$n = 51$	$D_i = \{1.8 - 11, 6.4 - 7, 2.0 - 2\}$

Example 4 A stable example is provided by the simple chemical reaction $A \to B \to C$ with rates k_1 and k_2. Here the differential equation is

$$\frac{d}{dt} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} .$$

Well-posedness of the estimation problem requires $(\mathbf{h})_3 \neq 0$. Maximum smoothness of the g-spline is achieved with $\mathbf{b} = \mathbf{e}_1$, $\mathbf{h} = \mathbf{e}_3$. Table 2 gives the diagonal elements of the rank revealing Cholesky for the same meshes as above. Again small elements are produced.

References

[1] D. B. Duncan and S. D. Horn, *Linear dynamic recursive estimation from the viewpoint of regression analysis*, J. Amer. Statist. Assoc. **67** (1972), 816–821. C234, C241

[2] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley and Sons, INC., 1968. C240

[3] M. Gulliksson and P. Wedin, *Modifying the QR-decomposition to constrained and weighted linear least squares*, SIAM J. Matrix Anal. Appl. **13** (1992), no. 4, 1298–1313. C233, C234, C237

[4] N. J. Higham, *Accuracy and stability of numerical algorithms*, SIAM, 1996. C240

[5] M. R. Osborne and Tania Prvan, *On algorithms for generalised smoothing splines*, J. Austral. Math. Soc. B **29** (1988), 322–341. C244

[6] M. R. Osborne and Tania Prvan, *Smoothness and conditioning in generalised smoothing spline calculations*, J. Austral. Math. Soc. B **30** (1988), 43–56. C241

[7] M. R. Osborne, *V-invariant methods for generalised least squares problems*, Computational Science - ICCS03,part 3, Springer Lecture Notes in Computer Science (Peter M. A. Sloot, David Abramson, Alexander V. Bogdanov, Jack J. Dongarra, Albert Y. Zomaya, and Yuriy E. Gorbachev, eds.), Springer, 2003, pp. 673–682. C234, C240

[8] C. C. Paige and M. A. Saunders, *Least squares estimation of discrete linear dynamic systems using orthogonal transformations*, SIAM J. Numer. Anal. **14** (1977), 180–193. C242

[9] I. Söderkvist, *An algorithm for Kalman filtering and smoothing*, Computational Techniques and Applications: CTAC95 (Robert L. May and Alan K. Easton, eds.), World Scientific Publishing Co., 1996, pp. 709–716. C234, C242, C244

[10] I. Söderkvist, *On algorithms for generalised least squares problems with ill-conditioned covariance matrices*, Computational Statistics **11**, pp. 303–313. C234