STATE-SPACE ESTIMATION WITH UNCERTAIN MODELS

Ali H. Sayed and Ananth Subramanian
Adaptive Systems Laboratory
Electrical Engineering Department
University of California
Los Angeles, CA 90095, USA
{sayed,msananth}@ee.ucla.edu*

Abstract There are always discrepancies between design models and the actual physical systems or phenomena that they model. Regardless of their source, such perturbations can degrade the performance of otherwise optimal designs. This article discusses a design strategy for models with bounded perturbations. In comparison to other robust formulations, the resulting procedure performs data regularization as opposed to de-regularization. Applications in state-space estimation and adaptive filtering are discussed.

Keywords: regularization, least-squares, robust filter, adaptive filter, Kalman filter, parametric uncertainty.

1. Introduction

Many estimation and control problems rely on solving regularized least-squares formulations of the form (see, e.g., [1]-[3]):

$$\min_x \left[x^T Q x + (Ax-b)^T W (Ax-b)\right] \tag{1}$$

where x^TQx is a regularization term, $Q > 0$ and $W \geq 0$ are Hermitian weighting matrices, x is an unknown $n-$dimensional column vector, A is a known $N \times n$ data matrix, and b is a known $N \times 1$ measurement vector. The solution of (1) is

$$\hat{x} = \left[Q + A^T W A\right]^{-1} A^T W b \tag{2}$$

*in ***Total Least Squares and Errors-in-Variables Modeling, III: Analysis, Algorithms and Applications*****, S. Van Huffel and P. Lemmerling, editors, pp. 191–202, Kulwer, 2002**. This material is based on work supported by the National Science Foundation under Award No. ECS-9632919.

When the nominal data $\{A, b\}$ are subject to uncertainties, especially large ones, the performance of the optimal estimator (2) may degrade appreciably — see [3]. For example, if the actual data matrix were $(A+\delta A)$, for some unknown perturbation δA, then the estimator (2) that is designed based on A alone and without accounting for the existence of δA, can perform poorly. In [4,5], a generalization of (1) that accounts for uncertainties in $\{A, b\}$ was introduced. This generalization leads to a robust solution that performs regularization as opposed to de-regularization; a property that is useful for on-line (iterative) schemes since it avoids the need for checking existence conditions. This chapter has two objectives. First, it provides an overview of a special case of the robust formulation of [4,5] along with an illustrative application in the context of state-space estimation. Then it extends the analysis to the case of recursive least-squares filtering and discusses an application in the context of adaptive filtering. The articles [3]-[7] provide references and connections to other related works in the literature (e.g., [11]-[13]). For reasons of space, we omit this discussion here and refer interested readers to [3]-[7] and [10] and the references therein.

Notation. For a column vector z and a positive-definite matrix W, we write $\|z\|^2$ and $\|z\|_W^2$ to denote the Euclidean norm and its weighted version, namely, $z^T z$ and $z^T W z$, respectively. For a matrix C, the notation $\|C\|$ is used to denote its maximum singular value. Also, for brevity, we may sometimes write $A^T W(\cdot)$ instead of $A^T W A$ especially when the factor A admits a long expression (see, e.g., Eq. (27)).

2. Uncertain Weighted Least-Squares

Let $J(x, y)$ denote a cost function of the form $J(x, y) = x^T Q x + R(x, y)$ with

$$R(x, y) = \Big((A+\delta A)x - (b+\delta b)\Big)^T W \Big((A+\delta A)x - (b+\delta b)\Big) \tag{3}$$

Here δA denotes an $N \times n$ perturbation to A, δb denotes an $N \times 1$ perturbation to b, and $\{\delta A, \delta b\}$ are assumed to satisfy a model of the form

$$\begin{bmatrix} \delta A & \delta b \end{bmatrix} = H \Delta \begin{bmatrix} E_a & E_b \end{bmatrix} \tag{4}$$

where Δ is an arbitrary contraction, $\|\Delta\| \leq 1$, and $\{H, E_a, E_b\}$ are known quantities of appropriate dimensions (e.g., E_b is a column vector). Consider now the problem of solving

$$\hat{x} = \arg\min_{x} \max_{\{\delta A, \delta b\}} J(x, y) \tag{5}$$

subject to (4). We shall assume that H and $\{E_a, E_b\}$ are not identically zero since otherwise problem (5) trivializes to (1). The statement (5) can be interpreted as a constrained two-player game problem, with the designer trying to pick an estimate $\hat{x}$ that minimizes the cost while the opponent $\{\delta A, \delta b\}$ tries to maximize the cost. The following result is proven in [4,5].

Theorem 1 *The problem (4)–(5) has a unique solution $\hat{x}$ that is given by (compare with (2))*

$$\hat{x} = \left[\widehat{Q} + A^T \widehat{W} A\right]^{-1} \left[A^T \widehat{W} b + \widehat{\beta} E_a^T E_b\right] \tag{6}$$

where $\{\widehat{Q}, \widehat{W}\}$ are obtained from $\{Q, W\}$ via

$$\widehat{Q} \triangleq Q + \widehat{\beta} E_a^T E_a \tag{7}$$

$$\widehat{W} \triangleq W + WH(\widehat{\beta}I - H^TWH)^{\dagger}H^TW \tag{8}$$

and the scalar $\widehat{\beta}$ is determined from the optimization

$$\widehat{\beta} = \arg \min_{\beta \geq \|H^TWH\|} G(\beta) \tag{9}$$

where the function $G(\beta)$ is defined as follows:

$$G(\beta) \triangleq x^T(\beta)Qx(\beta) \;+\; \beta\|E_a x(\beta) - E_b\|^2 \;+\; [Ax(\beta) - b]^T W(\beta)[Ax(\beta) - b] \tag{10}$$

with

$$W(\beta) \triangleq W + WH\big(\beta I - H^TWH\big)^{\dagger}H^TW \tag{11}$$

$$Q(\beta) \triangleq Q + \beta E_a^T E_a \tag{12}$$

and

$$x(\beta) \triangleq \Big[Q(\beta) + A^TW(\beta)A\Big]^{-1}\big[A^TW(\beta)b + \beta E_a^T E_b\big] \tag{13}$$

[The notation $X^{\dagger}$ denotes the pseudo-inverse of X.]

$\diamondsuit$

We shall denote the lower bound on β in (9) by $\beta_l = \|H^TWH\|$. Compared with the solution (2) of the standard regularized least-squares problem (1), we see that the expression for $\hat{x}$ in (6) is distinct in two important ways:

a) First, the weighting matrices $\{Q, W\}$ need to be replaced by corrected versions $\{\widehat{Q}, \widehat{W}\}$. These corrections are defined in terms of a scalar parameter $\widehat{\beta}$, which is obtained as the minimizing argument of a function $G(\beta)$ over the semi-open interval $[\beta_l, \infty)$.

b) Second, the right-hand side of (6) contains an additional term that is equal to $\widehat{\beta}E_a^T E_b$. This means that, with $\widehat{\beta}$ given or fixed, the $\hat{x}$ in (6) can be interpreted as the solution to a regularized least-squares problem of the form

$$\min_x \left(\begin{bmatrix} 1 & x^T \end{bmatrix} \begin{bmatrix} \widehat{\beta}\|E_b\|^2 & -\widehat{\beta}E_b^T E_a \\ -\widehat{\beta}E_a^T E_b & \widehat{Q} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} + (Ax - b)^T \widehat{W} (Ax - b) \right)$$

with a *cross-coupling* term between x and unity.

The complexity of the solution of Theorem 1 is therefore comparable to that of a standard regularized least-squares problem with the additional burden of determining the optimal scalar parameter $\widehat{\beta}$ by minimizing the cost function $G(\beta)$ over the interval $[\beta_l, \infty)$. It is shown in [5] that the function $G(\beta)$ has a unique global minimum (and no local minima) inside this interval, which means that the determination of $\widehat{\beta}$ can be pursued by employing standard search procedures without worrying about convergence to undesired local minima.

Before proceeding we note that in the applications described further ahead, it will be the case that the weighting matrix W is positive-definite so that we will always have $W(\beta) > 0$. Therefore, if we restrict the minimization in (9) to the open interval

(β_l, ∞) – i.e., if we exclude the boundary point β_l – then the pseudo-inverse operation in (11) can be replaced by normal matrix inversion, and it will hold that

$$W^{-1}(\beta) = W^{-1} - \beta^{-1}HH^T. \tag{14}$$

We now move on to apply the above results to two applications: one is in the context of state-space estimation and the other is in the context of adaptive filtering.

3. Robust State-Space Estimation

As is well-known, the Kalman filter is the optimal linear least-mean-squares state estimator for linear state-space models [1]. When the underlying model is subject to parametric uncertainties, the performance of the filter may deteriorate appreciably. The mismatch between the actual model and the assumed nominal model can be due to various factors including unmodelled dynamics and approximations during the system identification process.

In [7], a robust formulation for state-space estimation that is based on Theorem 1 has been proposed. Compared with the standard Kalman filter, which is known to minimize the regularized residual norm at each iteration, the new formulation minimizes the worst-possible regularized residual norm over the class of admissible uncertainties. In addition, compared with other robust formulations, the resulting filter performs data regularization rather than de-regularization; a property that avoids the need for existence conditions. We review the main steps below.

Thus consider a state-space model of the form

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i, \quad i \geq 0 && (15)\\ y_i &= H_i x_i + v_i && (16) \end{aligned}$$

where $\{x_0, u_i, v_i\}$ are uncorrelated zero-mean random variables with variances

$$E \begin{bmatrix} x_0 \\ u_i \\ v_i \end{bmatrix} \begin{bmatrix} x_0 \\ u_j \\ v_j \end{bmatrix}^T = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q_i\delta_{ij} & 0 \\ 0 & 0 & R_i\delta_{ij} \end{bmatrix} \tag{17}$$

that satisfy $\Pi_0 > 0$, $R_i > 0$, and $Q_i > 0$. Let further

$$\begin{aligned} \hat{x}_i &\triangleq \text{l.l.m.s. estimate of } x_i \text{ given } \{y_0, \ldots, y_{i-1}\} \\ \hat{x}_{i|i} &\triangleq \text{l.l.m.s. estimate of } x_i \text{ given } \{y_0, \ldots, y_i\} \end{aligned}$$

with corresponding error variances P_i and $P_{i|i}$, respectively. The notation l.l.m.s. stands for "linear least-mean-squares". Then $\{\hat{x}_i, \hat{x}_{i|i}\}$ can be constructed recursively via the following time- and measurement-update form of the Kalman filter (see, e.g., [1]):

$$\begin{aligned} \hat{x}_{i+1} &= F_i \hat{x}_{i|i}, \quad i \geq 0 && (18)\\ \hat{x}_{i+1|i+1} &= \hat{x}_{i+1} + P_{i+1|i+1} H_{i+1}^T R_{i+1}^{-1} e_{i+1} && (19)\\ e_{i+1} &= y_{i+1} - H_{i+1}\hat{x}_{i+1} && (20)\\ P_{i+1} &= F_i P_{i|i} F_i^T + G_i Q_i G_i^T && (21)\\ P_{i+1|i+1} &= P_{i+1} - P_{i+1} H_{i+1}^T R_{e,i+1}^{-1} H_{i+1} P_{i+1} && (22)\\ R_{e,i+1} &= R_{i+1} + H_{i+1} P_{i+1} H_{i+1}^T && (23) \end{aligned}$$

with initial conditions

$$\hat{x}_{0|0} = P_{0|0}^{-1} H_0^T R_0^{-1} y_0, \quad P_{0|0} = (\Pi_0^{-1} + H_0^T R_0^{-1} H_0)^{-1}$$

It can also be verified that these equations are equivalent to the following prediction form of the Kalman filter:

$$\hat{x}_{i+1} = F_i \hat{x}_i + K_i R_{e,i}^{-1} e_i \tag{24}$$

$$P_{i+1} = F_i P_i F_i^T + G_i Q_i G_i^T - K_i R_{e,i}^{-1} K_i^T \tag{25}$$

$$K_i = F_i P_i H_i^T, \;\; R_{e,i} = R_i + H_i P_i H_i^T \tag{26}$$

with initial conditions $\hat{x}_0 = 0$ and $P_0 = \Pi_0$.

Each step (18)–(23) of the time- and measurement-update form admits a useful deterministic interpretation as the solution to a regularized least-squares problem as follows (see, e.g., [8]). Given $\{\hat{x}_{i|i}, P_{i|i} > 0, y_{i+1}\}$, consider the problem of estimating x_i again, along with u_i, by solving

$$\min_{\{x_i, u_i\}} \left(\; (x_i - \hat{x}_{i|i})^T P_{i|i}^{-1} (\cdot) \; + \; u_i^T Q_i^{-1} u_i \; + (y_{i+1} - H_{i+1} x_{i+1})^T R_{i+1}^{-1} (\cdot) \; \right) \tag{27}$$

If we make the substitution $x_{i+1} = F_i x_i + G_i u_i$, then the cost in (27) reduces to a regularized least-squares cost of the form (1) with the identifications

$$\begin{aligned} x &\longleftarrow \text{col}\{x_i - \hat{x}_{i|i}, \; u_i\}, \quad b \longleftarrow y_{i+1} - H_{i+1} F_i \hat{x}_{i|i} \\ A &\longleftarrow H_{i+1} \begin{bmatrix} F_i & G_i \end{bmatrix}, \;\; Q \longleftarrow (P_{i|i}^{-1} \oplus Q_i^{-1}), \;\; W \longleftarrow R_{i+1}^{-1} \end{aligned}$$

The solution of this problem can be shown to lead to (18)–(23).

Assume now that the model (15)–(16) is uncertain, say

$$x_{i+1} = (F_i + \delta F_i) x_i + (G_i + \delta G_i) u_i \tag{28}$$

$$y_i = H_i x_i + v_i \tag{29}$$

$$\begin{bmatrix} \delta F_i & \delta G_i \end{bmatrix} = M_i \Delta_i \begin{bmatrix} E_{f,i} & E_{g,i} \end{bmatrix} \tag{30}$$

for some known matrices $\{M_i, E_{f,i}, E_{g,i}\}$ and for an arbitrary contraction Δ_i. Assume further that at step i we are given an *a priori* estimate for x_i, say $\hat{x}_{i|i}$, and a positive-definite weighting matrix $P_{i|i}$. Using y_{i+1}, we may update the estimate of x_i from $\hat{x}_{i|i}$ to $\hat{x}_{i|i+1}$ by solving

$$\min_{\{x_i, u_i\}} \max_{\{\delta F_i, \delta G_i\}} \left(\; (x_i - \hat{x}_{i|i})^T P_{i|i}^{-1} (\cdot) \; + \; u_i^T Q_i^{-1} u_i \; + \; (y_{i+1} - H_{i+1} x_{i+1})^T R_{i+1}^{-1} (\cdot) \; \right) \tag{31}$$

subject to (28)–(30). This problem can be seen to be the robust version of (27) in the same way that (3)–(5) is the robust version of (1). Now (31) can be written more compactly in the form (3)–(5) with the identifications:

$$\begin{aligned} x &\longleftarrow \text{col}\{x_i - \hat{x}_{i|i}, \; u_i\}, \quad b \longleftarrow y_{i+1} - H_{i+1} F_i \hat{x}_{i|i} \\ \delta A &\longleftarrow H_{i+1} M_i \Delta_i \begin{bmatrix} E_{f,i} & E_{g,i} \end{bmatrix} \\ \delta b &\longleftarrow -H_{i+1} M_i \Delta_i E_{f,i} \hat{x}_{i|i}, \;\; Q \longleftarrow (P_{i|i}^{-1} \oplus Q_i^{-1}) \\ W &\longleftarrow R_{i+1}^{-1}, \; H \longleftarrow H_{i+1} M_i, \; E_a \longleftarrow \begin{bmatrix} E_{f,i} & E_{g,i} \end{bmatrix} \\ E_b &\longleftarrow -E_{f,i} \hat{x}_{i|i}, \; \Delta \longleftarrow \Delta_i, \; A \longleftarrow H_{i+1} \begin{bmatrix} F_i & G_i \end{bmatrix} \end{aligned}$$

Using Theorem 1, and some considerable algebra, we arrive at the equations listed in Table 1 where we defined

$$\beta_{l,i} \triangleq \|M_i^T H_{i+1}^T R_{i+1}^{-1} H_{i+1} M_i\| \tag{32}$$

The major step in the algorithm of Table 1 is step 3, which consists of recursions that are very similar in nature to the prediction form of the Kalman filter. The main difference is that the new recursions operate on modified parameters rather than on the given nominal values. In addition, the recursion for P_i is not a standard Riccati recursion since the product $\widehat{G}_i\widehat{Q}_i\widehat{G}_i^T$ is also dependent on P_i. However, in some special cases, the recursion for P_i collapses to a Riccati recursion. In addition, alternative equivalent implementations of the robust filter of Table 1 in information form and in time- and measurement-update form are also possible (see [7]).

Assumed uncertain model. Eqs. (28)–(30).

Initial conditions: $\hat{x}_0 = 0$, $P_0 = \Pi_0$, and $\widehat{R}_0 = R_0$.

Step 1a. Using $\{\widehat{R}_i, H_i, P_i\}$ compute $\{R_{e,i}, P_{i|i}\}$:

$$R_{e,i} = \widehat{R}_i + H_i P_i H_i^T, \quad P_{i|i} = P_i - P_i H_i^T R_{e,i}^{-1} H_i P_i$$

Step 1b. If $H_{i+1}M_i = 0$, then set $\widehat{\beta}_i = 0$. Otherwise, determine $\widehat{\beta}_i$ by minimizing the corresponding $G(\beta)$ over the interval $(\beta_{l,i}, \infty)$.

Step 2. Compute the corrected parameters:

$$\begin{aligned}
\widehat{Q}_i^{-1} &= Q_i^{-1} + \widehat{\beta}_i E_{g,i}^T \left[I + \widehat{\beta}_i E_{f,i} P_{i|i} E_{f,i}^T\right]^{-1} E_{g,i} \\
\widehat{R}_{i+1} &= R_{i+1} - \widehat{\beta}_i^{-1} H_{i+1} M_i M_i^T H_{i+1}^T \\
\widehat{P}_{i|i} &= P_{i|i} - P_{i|i} E_{f,i}^T (\widehat{\beta}_i^{-1} I + E_{f,i} P_{i|i} E_{f,i}^T)^{-1} E_{f,i} P_{i|i} \\
\widehat{G}_i &= G_i - \widehat{\beta}_i F_i \widehat{P}_{i|i} E_{f,i}^T E_{g,i} \\
\widehat{F}_i &= (F_i - \widehat{\beta}_i \widehat{G}_i \widehat{Q}_i E_{g,i}^T E_{f,i})(I - \widehat{\beta}_i \widehat{P}_{i|i} E_{f,i}^T E_{f,i})
\end{aligned}$$

If $\widehat{\beta}_i = 0$, then simply set $\widehat{Q}_i = Q_i$, $\widehat{R}_{i+1} = R_{i+1}$, $\widehat{P}_{i|i} = P_{i|i}$, $\widehat{G}_i = G_i$, and $\widehat{F}_i = F_i$.

Step 3. Now update $\{\hat{x}_i, P_i\}$ to $\{\hat{x}_{i+1}, P_{i+1}\}$ as follows:

$$\begin{aligned}
\hat{x}_{i+1} &= \widehat{F}_i \hat{x}_i + \widehat{F}_i P_i H_i^T R_{e,i}^{-1} e_i \\
e_i &= y_i - H_i \hat{x}_i \\
P_{i+1} &= F_i P_i F_i^T - \overline{K}_i \overline{R}_{e,i}^{-1} \overline{K}_i^T + \widehat{G}_i \widehat{Q}_i \widehat{G}_i^T \\
\overline{K}_i &= F_i P_i \overline{H}_i^T, \quad \overline{R}_{e,i} = I + \overline{H}_i P_i \overline{H}_i^T
\end{aligned}$$

where $\overline{H}_i^T = \begin{bmatrix} H_i^T \widehat{R}_i^{-T/2} & \sqrt{\widehat{\beta}_i} E_{f,i}^T \end{bmatrix}$.

Table 1: Listing of the robust state-space estimation algorithm in prediction form.

Observe that the algorithm of Table 1 requires, at each iteration i, the minimization of $G(\beta)$ over $(\beta_{l,i}, \infty)$. It turns out that a reasonable approximation that avoids these repeated minimizations is to choose $\widehat{\beta}_i = (1+\alpha)\beta_{l,i}$. That is, we set $\widehat{\beta}_i$ at a multiple of the lower bound — if the lower bound is zero, we set $\widehat{\beta}_i$ to zero and replace $\widehat{\beta}_i^{-1}$ by $\widehat{\beta}_i^{\dagger}$ (which is also zero). The parameter α could be made time-variant; it serves as a "tuning" parameter that can be adjusted by the designer.

4. Robust Adaptive Filtering

In the model (28)–(30), we did not consider uncertainties in the output equation (29). We now look at this case, an application of which arises in the context of adaptive filtering.

Thus consider noisy measurements $\{d(i)\}$ that satisfy a model of the form,

$$d(i) = u_i w^o + n(i), \tag{33}$$

where w^o is an unknown $M \times 1$ column vector that we wish to estimate, u_i is a $1 \times M$ regression vector, and $n(i)$ is measurement noise. Given a collection of $(N+1)$ such data points, $\{d(i)\}_{i=0}^{N}$, and the corresponding $(N+1)$ regressors $\{u_i\}_{i=0}^{N}$, the exponentially-weighted least-squares problem estimates w^o by solving

$$\min_{w} \left[\mu\lambda^{N+1}\|w\|^2 \; + \; \sum_{i=0}^{N} \lambda^{N-i}|d(i) - u_i w|^2 \right], \tag{34}$$

where μ is a positive scalar regularization parameter and λ is a forgetting factor satisfying $0 \ll \lambda < 1$. In this way past data are exponentially weighted less than recent data.

Let w_N denote the optimal solution of (34), and let w_{N-1} denote the solution to a least-squares problem similar to (34) with data up to time $N-1$ (and with $\mu\lambda^{N+1}$ replaced by $\mu\lambda^{N}$). The well-known recursive least-squares (RLS) algorithm allows us to update w_{N-1} to w_N as follows:

$$\begin{aligned} w_N &= w_{N-1} + g_N \epsilon(N) & (35)\\ \epsilon(N) &= d(N) - u_N w_{N-1} & (36)\\ g_N &= \lambda^{-1}\overline{P}_{N-1}u_N^T \gamma(N) & (37)\\ \gamma^{-1}(N) &= 1 + \lambda^{-1}u_N \overline{P}_{N-1}u_N^T & (38)\\ \overline{P}_N &= \lambda^{-1}\overline{P}_{N-1} - g_N \gamma^{-1}(N) g_N^T & (39) \end{aligned}$$

with initial conditions $w_{-1} = 0$ and $\overline{P}_{-1} = \mu^{-1}I$. .

Now there is an intrinsic relation between exponentially-weighted RLS and Kalman filtering. More specifically, it was shown in [9] that the RLS equations (35)–(39) can be obtained directly from the Kalman filtering equations that correspond to a particular state-space model. This model is constructed as follows. Define the scaled variables

$$y(i) \triangleq d(i)/\sqrt{\lambda^i}, \qquad x_i \triangleq w^o/\sqrt{\lambda^i} \tag{40}$$

and introduce the state-space model

$$\begin{aligned} x_{i+1} &= \lambda^{-1/2}x_i, \;\; i \geq 0, \\ y(i) &= u_i x_i + v(i), \end{aligned} \tag{41}$$

where $\{x_0, v(i)\}$ are taken as uncorrelated zero-mean random variables with variances $\mu^{-1}\lambda^{-1}I$ and δ_{ij}, respectively. Here δ_{ij} is the Kronecker delta function that is equal to unity when $i = j$ and zero elsewhere. The Kalman filter equations in this case collapse to the following:

$$\hat{x}_{i+1} = \lambda^{-1/2}\hat{x}_i + \lambda^{-1/2}P_i u_i^T R_{e,i}^{-1}[y(i) - u_i\hat{x}_i]$$

where

$$P_{i+1} = \lambda^{-1}P_i - K_i R_{e,i}^{-1} K_i^T, \quad K_i = \lambda^{-1/2}P_i u_i^T, \quad R_{e,i} = 1 + u_i P_i u_i^T$$

with initial conditions $\hat{x}_0 = 0$ and $P_0 = \mu^{-1}\lambda^{-1}I$. It can be verified that the RLS variables $\{\overline{P}_i, w_i\}$ and the Kalman filtering variables $\{P_i, \hat{x}_{i+1}\}$ are related as follows:

$$\hat{x}_{i+1} = w_i/\sqrt{\lambda^{(i+1)}}, \qquad \overline{P}_i = \lambda P_{i+1}$$

When the regressors $\{u_i\}$ are subject to uncertainties, we can instead consider the model

$$x_{i+1} = \lambda^{-1/2}x_i, \quad i \geq 0, \tag{42}$$
$$y(i) = (u_i + \delta u_i)x_i + v(i). \tag{43}$$

This would correspond to assuming that the observations $\{d(i)\}$ in (33) actually arise from the perturbed linear regression model

$$d(i) = (u_i + \delta u_i)w^o + n(i). \tag{44}$$

The perturbations $\{\delta u_i\}$ are now row vectors and they are modeled as

$$\delta u_i = M_i \Delta_i E_{u,i} \tag{45}$$

where M_i is a known scaling factor, Δ_i is an unknown scalar that is bounded by unity (i.e., $|\Delta_i| \leq 1$), and $E_{u,i}$ is a known row vector. The model (44)–(45) allows us to account for uncertainties in the regression vectors u_i in different ways. Consider, for example, the case when M_i is a positive constant, say $M_i = \eta$, and choose

$$E_{u,i} = [1\ 1\ \ldots\ 1] \tag{46}$$

Then these choices for $\{M_i, E_{u,i}\}$ amount to assuming that each entry of u_i is perturbed by an amount that is bounded by η. We can also vary the degree of uncertainty in the individual entries of u_i by assigning different values to the entries of $E_{u,i}$. In addition, the fact that M_i and $E_{u,i}$ are allowed to change with time offers the designer the opportunity to modify the specification of the uncertainty with time.

As in the state-space context of Sec. 3, assume now that at step i we are given an *a priori* estimate for the state x_i of (42)–(43). We shall denote this initial estimate by $\hat{x}_{i|i}$. Assume further that we are also given a positive-definite weighting matrix $P_{i|i}$, along with the observation at time $(i+1)$, i.e., $y(i+1)$. Using this initial information, we may update the estimate of x_i from $\hat{x}_{i|i}$ to $\hat{x}_{i|i+1}$ by solving

$$\min_{\{x_i\}} \max_{\delta u_i} \left(\|x_i - \hat{x}_{i|i}\|^2_{P_{i|i}^{-1}} + \|y(i+1) - (u_{i+1} + \delta u_{i+1})x_{i+1}\|^2 \right) \tag{47}$$

subject to (42)–(45). This problem can be written more compactly in the form (4)–(5) with the identifications:

$$x \longleftarrow x_i - \hat{x}_{i|i}, \qquad b \longleftarrow y(i+1) - \lambda^{-1/2} u_{i+1}\hat{x}_{i|i} \tag{48}$$

$$A \longleftarrow \lambda^{-1/2} u_{i+1}, \qquad \delta A \longleftarrow \lambda^{-1/2} M_{i+1}\Delta_{i+1}E_{u,i+1} \tag{49}$$

$$\delta b \longleftarrow -\lambda^{-1/2} M_{i+1}\Delta_{i+1}E_{u,i+1}\hat{x}_{i|i}, \qquad Q \longleftarrow P_{i|i}^{-1} \tag{50}$$

$$W \longleftarrow 1, \qquad H \longleftarrow M_{i+1} \tag{51}$$

$$E_a \longleftarrow \lambda^{-1/2} E_{u,i+1}, \quad E_b \longleftarrow -\lambda^{-1/2} E_{u,i+1}\hat{x}_{i|i}, \quad \Delta \longleftarrow \Delta_{i+1} \tag{52}$$

Using Theorem 1 and the correspondences (40) between the original variables and the state-space variables, we arrive at the equations listed in Table 2.

<u>Assumed uncertain model</u>. Eqs. (44)–(45).

<u>Initial conditions</u>: $w_{-1} = 0$, $P_0 = \frac{1}{\mu\lambda} I$, and $\widehat{R}_0 = 1$.

<u>Step 1a.</u> Using $\{\widehat{R}_i, u_i, P_i\}$ compute $\{R_{e,i}, P_{i|i}\}$:

$$R_{e,i} = \widehat{R}_i + u_i P_i u_i^T, \qquad P_{i|i} = P_i - P_i u_i^T R_{e,i}^{-1} u_i P_i$$

<u>Step 1b.</u> If $u_{i+1}M_{i+1} = 0$, then set $\widehat{\beta}_i = 0$. Otherwise, construct $G(\beta)$ of (10) with the identifications (48)–(52) and determine $\widehat{\beta}_i$ by minimizing $G(\beta)$ over the interval $\widehat{\beta}_i > \beta_{l,i} = M_{i+1}^2$.

<u>Step 2.</u> Compute the corrected parameters

$$\begin{aligned} \widehat{R}_{i+1} &= 1 - M_{i+1}^2/\widehat{\beta}_i \\ \widehat{P}_{i|i} &= P_{i|i} - \lambda^{-1} P_{i|i} E_{u,i+1}^T (\widehat{\beta}_i^{-1} I + \lambda^{-1} E_{u,i+1} P_{i|i} E_{u,i+1}^T)^{-1} E_{u,i+1} P_{i|i} \\ \widehat{F}_i &= (I - \widehat{\beta}_i \lambda^{-1} \widehat{P}_{i|i} E_{u,i+1}^T E_{u,i+1}) \end{aligned}$$

If $\widehat{\beta}_i = 0$, then simply set $\widehat{P}_{i|i} = P_{i|i}$, $\widehat{R}_{i+1} = 1$, and $\widehat{F}_i = I$.

<u>Step 3.</u> Update $\{w_i, P_i\}$ as follows:

$$\begin{aligned} w_i &= \widehat{F}_i w_{i-1} + \widehat{F}_i P_i u_i^T R_{e,i}^{-1} e(i) \\ e(i) &= d(i) - u_i w_{i-1} \\ \overline{K}_i &= \lambda^{-1/2} P_i \overline{u}_i^T, \quad \overline{R}_{e,i} = 1 + \overline{u}_i P_i \overline{u}_i^T \\ P_{i+1} &= \lambda^{-1} P_i - \overline{K}_i \overline{R}_{e,i}^{-1} \overline{K}_i^T \end{aligned}$$

where

$$\overline{u}_i \triangleq \begin{bmatrix} u_i/\sqrt{\widehat{R}_i} \\ \sqrt{\widehat{\beta}_i/\lambda}\, E_{u,i+1} \end{bmatrix}$$

Table 2: Listing of the robust adaptive algorithm.

To illustrate the developed filter, we choose an implementation of order 5 with $E_{u,i} = [2.0 \ \ 1.5 \ \ 2.5 \ \ 1.0 \ \ 1.3]$, $M_i = 2.5$ for all i, and Δ_i selected uniformly within $[-1, 1]$. We also set $\lambda = 0.95$. The figure shows the average squared weight-error curves (averaged over 100 experiments) for the conventional RLS algorithm and for the robust algorithm. The improvement in performance will be more noticeable for larger uncertainties.

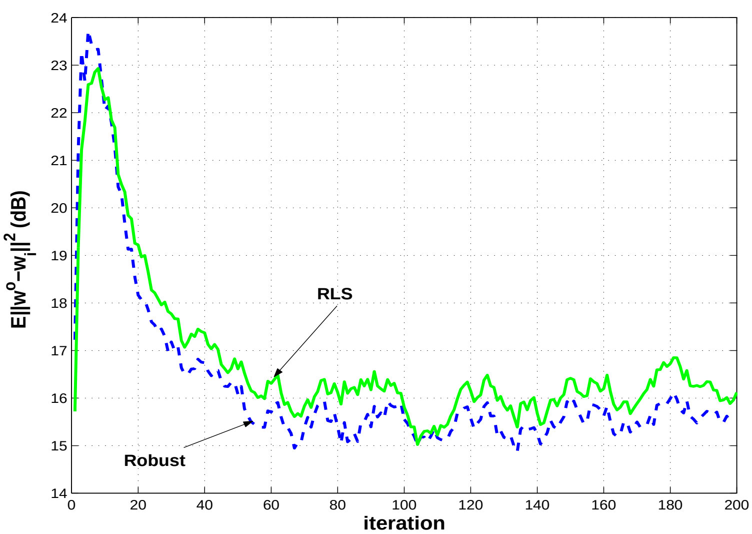

Figure 1. Averaged squared weight-error curves for RLS (solid line) and the robust filter (dashed line) for $\lambda = 0.95$.

5. Concluding Remarks

The robust design procedure described in this chapter performs a local optimization step at each iteration. The optimization is not global and the parameter $\widehat{\beta}$ becomes a tuning parameter that could be adjusted by the designer. At present we are investigating designs with weighting matrices related to actual error covariance matrices, in addition to allowing for stochastic modeling of the uncertainties. These issues will be discussed elsewhere.

References

[1] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation.* Prentice-Hall, NJ, 2000.

[2] B. Hassibi, A. H. Sayed, and T. Kailath. *Indefinite Quadratic Estimation and Control: A Unified Approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Theories.* SIAM, PA, 1999.

[3] A. H. Sayed, V. H. Nascimento, and S. Chandrasekaran. Estimation and control with bounded data uncertainties. *Linear Algebra and Its Applications*, vol. 284, pp. 259–306, Nov. 1998.

[4] A. H. Sayed and V. H. Nascimento. Design criteria for uncertain models with structured and unstructured uncertainties, in *Robustness in Identification and Control*, A. Garulli, A. Tesi, and A. Vicino, editors, vol. 245, pp. 159–173, Springer Verlag, London, 1999.

[5] A. H. Sayed, V. H. Nascimento, and F. A. M. Cipparrone. A regularized robust design criterion for uncertain data. To appear in *SIAM J. Matrix Anal. Appl.*.

[6] A. H. Sayed and S. Chandrasekaran. Parameter estimation in the presence of multiple sources and levels of uncertainties, *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 680–692, Mar. 2000.

[7] A. H. Sayed. A framework for state space estimation with uncertain models. *IEEE Trans. Automat. Contr.*, vol. 46, no. 7, pp. 998–1013, July 2001.

[8] A. E. Bryson and Y.-C. Ho. *Applied Optimal Control: Optimization, Estimation, and Control.* Taylor & Francis, revised printing, 1975.

[9] A. H. Sayed and T. Kailath. A state-space approach to adaptive RLS filtering. *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60, July 1994.

[10] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed. Parameter estimation in the presence of bounded data uncertainties. *SIAM. J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[11] I. R. Petersen and A. V. Savkin. *Robust Kalman Filtering for Signals and Systems with Large Uncertainties.* Birkhauser, Boston, 1999.

[12] L. Xie, Y. C. Soh, and C. E. de Souza. Robust Kalman filtering for uncertain discrete-time systems. *IEEE Trans. Automat. Contr.*, vol. 39, no. 6, pp. 1310–1314, 1994.

[13] L. E. Ghaoui and H. Hebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, vol. 18, no.4, pp. 1035–1064, 1997.