Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

An insight into noise covariance estimation for Kalman filter design ⋆

Simone Formentin, **Sergio Bittanti** *(IFAC Fellow)*

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, P.za L. da Vinci 32, 20133 Milano, Italy.

Abstract: In Kalman filtering applications, the variance of the estimation error is guaranteed to be minimized only if a complete description of the plant dynamics is available. While nowadays there are several established methods to find an accurate model of a physical system, the evaluation of the covariance matrices of the disturbances is still a hard task. Therefore, such matrices are usually parameterized as diagonal matrices and their entries are estimated from a preliminary set of measurements. In this paper, we analyze the properties of the Kalman filter for linear time-invariant systems and we show that, for some classes of systems, the number of design parameters can be significantly reduced. Moreover, we prove that, under some assumptions, the covariance of the process noise can be parameterized as a full matrix without increasing the complexity of the tuning procedure. The above theoretical results are tested on two numerical examples.

Keywords: Kalman filter, noise covariance estimate, state estimation.

1. INTRODUCTION

Since the celebrated paper by Kalman and Bucy [1961], Kalman filtering and prediction have become main design tools in engineering and science, in order to estimate unknown signals and unknown parameters. Specifically, Kalman filtering not only permits to use current input and output measurements to derive an estimate of the unmeasured state variables, but such an estimate is also guaranteed to minimize the variance of the estimation error, provided that the system dynamics is known and captured by a given mathematical model.

The model can be inferred from the physical laws of the underlying system or from data-based system identification methods Ljung [1999], Verhaegen and Verdult [2007]. As a matter of fact, while the system matrices can be derived from the mathematical model, the evaluation of the covariance matrices from physical considerations is a hard task, especially for the state equation noise assessment.

This is why the covariance matrices are usually estimated from a set of preliminary experimental measurements, often by resorting to trial and error procedures. Early contributions in such a direction date back to the '70s, with the proposal of an iterative process aiming at achieving an innovation signal with suitable properties, see Mehra [1970, 1972], Carew and Bélanger [1973]. Such an approach has also been further elaborated in more recent years, see *e.g.* the auto-covariance least-squares (ALS) method in Odelson et al. [2006], Åkesson et al. [2008], or the Bayesian approach using Monte Carlo simulations in Matisko and Havlena [2013].

⋆ The support by the MIUR national project "Identification and adaptive control of industrial systems" and by CNR - IEIIT is gratefully acknowledged.

Notwithstanding these recent efforts, the definition of simple and easy-to-use tuning recipes is still an open question and the industrial standard still relies on diagonal parameterization of the covariance matrices and semi-empirical tuning.

In this paper, our aim is not to present a new estimation strategy, rather we try to gather a deeper insight on how the covariance matrices affect the resulting filter, in order to understand whether their data-based optimization can be simplified.

Based on the above problem setting, we first consider the scalar case and show that the performance of the Kalman filter does not depend separately upon the variance of the process noise and that of the output noise, but only their ratio matters. This is not a new result, because a similar theorem has been already presented for frequency tracking in Bittanti and Savaresi [2000]. However, in that paper, the proof was based on a circular rationale, whereas here the proof is explicit and extended to multivariable systems under suitable assumptions on the system structure.

As a secondary result, we show that, under the assumption that the dynamic matrix of the system can be diagonalized, the covariance of the process noise can be parameterized as a full matrix, thus generally increasing the estimation performance, without increasing the number of design parameters.

The remainder of the paper is as follows. In Section 2, the problem of interest is introduced and formally posed; moreover, we state the objectives of the analysis. The main results are presented in Section 3: specifically, in Subsection 3.1, we study how many tuning knobs are really relevant in Kalman filtering, whereas in Subsection 3.2 we present a reasonable way to parameterize the covariance of the process noise as a full matrix, without increasing

the number of design parameters. The results are further discussed on two numerical examples in Section 4. The paper ends with some concluding remarks.

2. PROBLEM FORMULATION

Consider the discrete-time linear time-invariant system

$$x(t+1) = Fx(t) + v_x(t) \tag{1a}$$
$$y(t) = Hx(t) + v_y(t) \tag{1b}$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$ and $v_x(t) \sim WN(0, Q)$, $v_y(t) \sim WN(0, R)$ are uncorrelated Gaussian noises with covariance matrices $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{p\times p}$, respectively. Moreover, $F \in \mathbb{R}^{n\times n}$ and $H \in \mathbb{R}^{p\times n}$ in (1) denote the state matrix and the output matrix. From now on, it will be assumed that the system is stabilizable and detectable.

The problem of optimal prediction of $x(t)$ using output measurements $y(t)$ is solved by the Kalman predictor

$$\hat{x}(t+1|t) = F\hat{x}(t|t-1) + K(t)\left(y(t) - H\hat{x}(t|t-1)\right) \tag{2}$$

where the time-varying Kalman gain is

$$K(t) = FP(t)H^T L(t) \tag{3a}$$
$$L(t) = \left(HP(t)H^T + R\right)^{-1} \tag{3b}$$

and $P(t) = \mathbb{E}\left[(x(t) - \hat{x}(t|t-1))(x(t) - \hat{x}(t|t-1))^T\right]$ is the covariance matrix of the one-step ahead prediction error, given by the solution of the Difference Riccati Equation (DRE)

$$P(t+1) = FP(t)F^T + Q - FP(t)H^T L(t)HP(t)F^T \tag{4}$$

with user-defined initial conditions $P(1)$ and $\hat{x}(1|0)$ (for more details, see, *e.g.*, Anderson and Moore [1979], Simon [2006], Verhaegen and Verdult [2007], Lewis et al. [2008]).

In real-world applications, Q and R capture, respectively, the process uncertainty and the noise affecting the output measurements. A common experience among engineers is that the values of the entries of matrices Q and R are unknown or very uncertain, hence difficult to tune. Another common experience is that the filtering performance is strongly dependent upon the results of such a tuning procedure. Therefore, this aspect of Kalman filter design is a time consuming trial-and-error procedure.

In some applications, it is possible to collect a set of N measurements $\mathcal{D}^N = \{x(t),\ y(t)\}_{t=1,\dots,N}$ in a preliminary tuning phase, in order to suitably design Q and R. For instance, one can solve the nonlinear optimization problem

$$\min_{Q,R}\ tr\left[\frac{1}{N}\sum_{t=1}^{N}(x(t) - \hat{x}(t|t-1))(x(t) - \hat{x}(t|t-1))^T\right] \tag{5a}$$
$$s.t.\ (1), (2), (3), (4)\ \text{hold},\ \forall t,\ \text{and}\ Q > 0,\ R > 0. \tag{5b}$$

Solving Problem (5) is generally a hard task, therefore brute force solutions, *i.e.*, gridding over a set of instances of Q and R, are sometimes employed. Moreover, since both the covariance matrices must be constrained to be positive semi-definite, Q and R are often parameterized as diagonal matrices, *i.e.* they are assumed to be of the form

$$Q = \begin{bmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_n \end{bmatrix},\ R = \begin{bmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_p \end{bmatrix}.$$

In such a form, the sign constraints for Q and R simply become $q_i > 0$, $i = 1, \dots, n$, $r_j > 0$, $j = 1, \dots, p$, thus also reducing to $n + p$ the number of parameters to tune.

However, although the diagonal parameterization of the covariance matrices works quite well in some real-world applications, it yields indeed a sub-optimal solution of the original problem. Moreover, even under the diagonal assumption, the choice of $n + p$ parameters might be a non-trivial computational task.

Summing up, it is desirable to reduce the number of tuning knobs and let the parameterization of Q, R be as general as possible. Both these problems are addressed in this paper and the related results will be presented in the next section.

3. MAIN RESULTS

In this section, we first illustrate (Subsection 3.1) some specific cases where the number of parameters can be reduced without loss of optimality. Then, in Subsection 3.2, we present some sufficient conditions under which Q and R can be set as full matrices without increasing the computational complexity. All the results will be referred only to the Kalman predictor for the sake of space. However, the theorems can be straightforwardly extended, without conceptual changes, to Kalman filtering applications (as a matter of fact, one example out of two in Section 4 is dedicated to a filtering problem).

3.1 Reducing the tuning knobs

In this first subsection, we address the problem of reducing the number of parameters when both Q and R are parameterized as diagonal matrices. To start with, consider the scalar case $n = p = 1$. The following theorem holds.

Theorem 1. (Scalar systems). Consider a scalar system (1), *i.e.* let $n = p = 1$ and $Q = q,\ R = r$, where $q, r \in \mathbb{R}^+$.

The steady-state prediction of the state (2) depends on $\lambda = r/q$ only.

Proof. To start with, notice that, in the scalar case, the expression of the Kalman gain and the DRE become, respectively,

$$K(t) = \frac{FP(t)H}{H^2P(t) + r} \tag{6}$$
$$P(t+1) = F^2P(t) + q - \frac{F^2P^2(t)H^2}{H^2P(t) + r}. \tag{7}$$

Then, consider a companion problem using a different parameterization of the noise, *i.e.* set a state prediction problem for

$$x(t+1) = Fx(t) + \tilde{v}_x(t) \tag{8a}$$
$$y(t) = Hx(t) + \tilde{v}_y(t) \tag{8b}$$

where $\tilde{v}_x(t) \sim WN(0, 1)$, $\tilde{v}_y(t) \sim WN(0, \lambda)$ are uncorrelated Gaussian noises with variance of 1 and $\lambda = r/q$, respectively. The corresponding expressions for the Kalman gain and the DRE are, respectively,

$$\tilde{K}(t) = \frac{F\tilde{P}(t)H}{H^2\tilde{P}(t) + \lambda} \tag{9}$$

$$\tilde{P}(t+1) = F^2\tilde{P}(t) + 1 - \frac{F^2\tilde{P}^2(t)H^2}{H^2\tilde{P}(t)+\lambda}, \tag{10}$$

which clearly depend on the unique tuning knob λ.

On the other hand, the companion problem can be shown to be equivalent to the original one. As a matter of fact, by multiplying (7) on both sides by q^{-1} and defining $\tilde{P} = q^{-1}P$, Equation (10) is obtained. Moreover,

$$\begin{aligned}\tilde{K}(t) &= \frac{F\tilde{P}(t)H}{H^2\tilde{P}(t)+\lambda} = \frac{q^{-1}FP(t)H}{q^{-1}H^2P(t)+\lambda} \\ &= \frac{FP(t)H}{H^2P(t)+q\lambda} = \frac{FP(t)H}{H^2P(t)+r} = K(t),\end{aligned}$$

that is, the Kalman gain given as the solution of the companion problem coincides with the optimal one, provided that the initial values are suitably tuned as $\tilde{P}(1) = q^{-1}P(1)$. In case the initial conditions do not correspond, the result is still valid for the steady-state predictor.

□

From Theorem 1, it follows that one can solve the companion problem introduced in the proof, by tuning a single parameter λ instead of q and r, and use the corresponding Kalman gain for state prediction, without loss of performance. In fact, we showed that such a gain is equivalent to the optimal gain for the original problem.

It should be here said that this idea for the scalar case is not new, as it was already presented for the frequency tracking problem in Bittanti and Savaresi [2000]. However, the structure of the proof was based on a circular argument needed for the nonlinear nature of the problem. In this paper, the proof for the linear case is given in a simpler form.

With such a modification, the previous result can be straightforwardly extended to multivariable systems with $n = p$, provided that some additional assumptions on the dynamic matrices are made. The result in the MIMO case is presented in the next Theorem.

Theorem 2. (Multivariable systems). Consider a multivariable system (1) where $n = p$. Assume that

A1. $FQ = QF$,
A2. $H = hI_n$, where $h \in \mathbb{R}$ and I_n denotes the $n \times n$ identity matrix.

The steady-state prediction of the state (2) depends on the $n \times n$ matrix $\Lambda = RQ^{-1}$ only.

Proof. Like in the proof of Theorem 1, consider a companion problem using a different parameterization of the noise, *i.e.* set a state prediction problem for

$$x(t+1) = Fx(t) + \tilde{v}_x(t) \tag{11a}$$
$$y(t) = hx(t) + \tilde{v}_y(t) \tag{11b}$$

where $\tilde{v}_x(t) \sim WN(0, I_n)$, $\tilde{v}_y(t) \sim WN(0, \Lambda)$ are uncorrelated Gaussian noises with covariance matrices I_n and $\Lambda = RQ^{-1}$, respectively. The corresponding expressions for the Kalman gain and the DRE are, respectively,

$$\tilde{K}(t) = hF\tilde{P}(t)\tilde{L}(t) \tag{12a}$$
$$\tilde{L}(t) = \left(h^2\tilde{P}(t)+\Lambda\right)^{-1} \tag{12b}$$
$$\tilde{P}(t+1) = F\tilde{P}(t)F^T + I_n - F\tilde{P}(t)H^T\tilde{L}(t)H\tilde{P}(t)F^T \tag{13}$$

which depend on the only matrix Λ.

The companion problem can be shown to be equivalent to the original one. As a matter of fact, by post-multiplying (4) on both sides by Q^{-1}, the DRE can be rewritten, for the current case, as

$$\begin{aligned}P(t+1)Q^{-1} &= FP(t)F^TQ^{-1} + I_n + \\ &- h^2FP(t)\left(h^2P(t)+R\right)^{-1}P(t)F^TQ^{-1}.\end{aligned} \tag{14}$$

Under Assumption **A1** and defining $\tilde{P} = PQ^{-1}$, Equation (14) can be rewritten as

$$\begin{aligned}\tilde{P}(t+1) &= F\tilde{P}(t)F^T + I_n + \\ &- h^2FP(t)\left(h^2P(t)+R\right)^{-1}\tilde{P}(t)F^T \\ &= F\tilde{P}(t)F^T + I_n + \\ &- h^2FP(t)Q^{-1}\left(h^2P(t)Q^{-1}+RQ^{-1}\right)^{-1}\tilde{P}(t)F^T \\ &= F\tilde{P}(t)F^T + I_n + \\ &- h^2F\tilde{P}(t)\left(h^2\tilde{P}(t)+\Lambda\right)^{-1}\tilde{P}(t)F^T,\end{aligned} \tag{15}$$

that corresponds to (13). Moreover, using the relationship between $P(t)$ and $\tilde{P}(t)$, the matrix $\tilde{L}(t)$ given by (12) turns out to be

$$\begin{aligned}\tilde{L}(t) &= \left(h^2P(t)Q^{-1}+\Lambda\right)^{-1}, \\ &= Q\left(h^2P(t)+R\right)^{-1} = QL(t)\end{aligned} \tag{16}$$

and, as a consequence,

$$\tilde{K}(t) = hFP(t)Q^{-1}QL(t) = hFP(t)L(t) = K(t), \tag{17}$$

that is, the Kalman gain given as the solution of the companion problem coincides with the optimal one, provided that $\tilde{P}(1) = P(1)Q^{-1}$. In case the initial conditions do not correspond, the result is still valid for the steady-state predictor.

□

From Theorem 2, it follows that one can solve the companion problem introduced in the proof, by tuning a single matrix Λ instead of Q and R, and use the corresponding Kalman gain for state prediction, without loss of performance. In fact, as in the scalar case, the original and the companion problem are equivalent.

The tuning problem (5) then becomes

$$\min_{\Lambda} tr\left[\frac{1}{N}\sum_{t=1}^{N}(x(t)-\hat{x}(t|t-1))(x(t)-\hat{x}(t|t-1))^T\right] \tag{18a}$$
$$s.t.\ (1),(2),(3),(4)\ \text{hold},\ \forall t,\ \text{and}\ \Lambda > 0, \tag{18b}$$

where Λ is set in place of R and Q is set as the identity matrix.

Obviously, the assumptions $n = p$ and $H = hI_n$ of Theorem 2 do not make it generally applicable to any MIMO problem. However, the result can still be useful in some practical cases, *e.g.*, in those cases where all the states are measured, but the signal to noise ratio is so large that the Kalman formulas still need to be used to obtain accurate estimates. This situation will be exemplified in Section 4.

Notice also that the assumption of commutativity of Q and F is not verifiable in practice, since Q is unknown.

However, under the assumption of parameterizing Q as a diagonal matrix, the commutativity of Q and F is always satisfied whenever F is diagonal or Q can be assumed to be of the form $Q = qI_n$, where q is unknown.

In Section 4, we will present an example where the above result is useful to reduce the design complexity without jeopardizing the performance.

3.2 Diagonalization

In this subsection, we try to answer the question: "is it possible to parameterize Q as a full matrix, without increasing the number of optimization variables and without adding a more complex constraint to enforce $Q > 0$?".

The problem is certainly difficult to solve in the general case. However, we will show that, under a sufficient condition, we can find a full Q by solving a companion problem with the same number of unknowns. The above condition and the technical details are given in the following Theorem.

Theorem 3. (Diagonalization). Consider a system (1) such that

A3. $\exists$ non-singular T: $F_d = T^{-1}FT$ is diagonal and real

and let its diagonal state-space representation be

$$x_d(t+1) = F_d x_d(t) + v_{x_d}(t) \tag{19a}$$
$$y(t) = H_d x(t) + v_y(t) \tag{19b}$$

where $H_d = HT$, $x_d(t) \in \mathbb{R}^n$ and $v_{x_d}(t) \sim WN(0, Q_d)$ is the corresponding process noise with covariance matrix Q_d.

The Kalman gain for state prediction of the original system (1) is

$$K(t) = TK_d(t), \tag{20}$$

where K_d is the Kalman gain for state prediction of (19).

Proof. Consider the DRE associated with Equations (19)

$$P_d(t+1) = F_d P_d(t) F_d^T + Q_d - F_d P_d(t) H_d^T L_d(t) H_d P_d(t) F_d^T \tag{21}$$

where

$$L_d(t) = \left(H_d P_d(t) H_d^T + R\right)^{-1}. \tag{22}$$

By recalling the expressions of F_d and H_d, Equation (21) can be rewritten as

$$\begin{aligned} P_d(t+1) = {} & T^{-1}FTP_d(t)T^T F^T T^{-T} + Q_d + \\ & - T^{-1}FTP_d(t)T^T H^T L_d(t) HTP_d(t)T^T F^T T^{-T} \end{aligned} \tag{23}$$

that is, by pre-multiplying and post-multiplying both sides for T and T^T, respectively,

$$\begin{aligned} TP_d(t+1)T^T = {} & FTP_d(t)T^T F^T + TQ_d T^T + \\ & - FTP_d(t)T^T H^T L_d(t) HTP_d(t)T^T F^T \end{aligned} \tag{24}$$

Since also $L(t)$ can be rewritten as a function of the matrices in (1) as

$$L_d(t) = \left(HTP_d(t)T^T H^T + R\right)^{-1}, \tag{25}$$

by comparing (24) with (4), it is clear that P and P_d are related to each other according to the equation

$$P(t) = TP_d(t)T^T.$$

It follows that the Kalman gain associated with the diagonal system is expressed as

$$\begin{aligned} K_d(t) &= F_d P_d(t) H_d^T L_d(t) \\ &= T^{-1}FTP_d(t)T^T H \left(HTP_d(t)T^T H^T + R\right)^{-1} \\ &= T^{-1}FP(t)H \left(HP(t)H^T + R\right)^{-1} \\ &= T^{-1}K(t) \end{aligned}$$

and the thesis holds.

□

This result obviously does not provide any additional knowledge about the covariance of the process noise. However, from a practical point of view, it is definitely more reasonable to assume a diagonal parameterization for the diagonal companion system, than for the original one. Once Q_d is designed and the corresponding K_d is obtained, the Kalman gain for prediction of the desired state $x(t)$ is given - without additional tuning effort - by (20).

4. SIMULATION EXAMPLES

4.1 Example 1

Consider system (1), with $n = 2$ states and $p = n = 2$ output, described by the matrices

$$F = \begin{bmatrix} 0.9 & -0.4 \\ 0.2 & 0.9 \end{bmatrix}, \; Q = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix},$$
$$H = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \; R = \begin{bmatrix} 0.64 & 0 \\ 0 & 0.071 \end{bmatrix}$$

and suppose that one wants to estimate $x_1(t)$. Since $x_1(t)$ is measured via the first output $y_1(t)$, one can argue that it is sufficient to take the first term of the trivial estimator $\hat{x}(t|t) = H^{-1}y(t)$. However, notice that the signal to noise ratio of the first output is very low (the ratio between the variance of $0.5x_1(t)$ and the variance of the noise affecting the first equation is approximately 0.5), therefore the use of a Kalman filter is generally preferable to obtain an accurate estimate of $x_1(t)$.

Collect then a set of data by injecting the white noise $v_x(t) = [v_{x_1}(t),\ v_{x_2}(t)]^T$, where $v_{x_1}(t) \sim WN(0, 0.64)$ and $v_2(t) \sim WN(0, 0.071)$ and record the corresponding state and output measurements for tuning purposes.

Since F and Q commute and $H = 0.5I_2$, **A1** and **A2** hold and the result of Theorem 2 can be applied to estimate the two parameters of

$$\Lambda = RQ^{-1} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 2.56 & 0 \\ 0 & 0.2844 \end{bmatrix},$$

instead of the four parameters of the diagonal matrices Q and R. We should here remark that the result in Theorem 2 in the proposed form can be directly applied the one-step ahead predictor. However, in this case where F is invertible, the optimal filter can be simply obtained as $\hat{x}_1(t|t) = F^{-1}\hat{x}(t|t-1)$.

The result of the simulations with the filter obtained via Theorem 2 is compared to standard Kalman filtering in Figure 1 (for more details on the numerical implementation of the Kalman filter, see Grewal and Andrews [2008]). In both cases, all the required matrices are initialized as

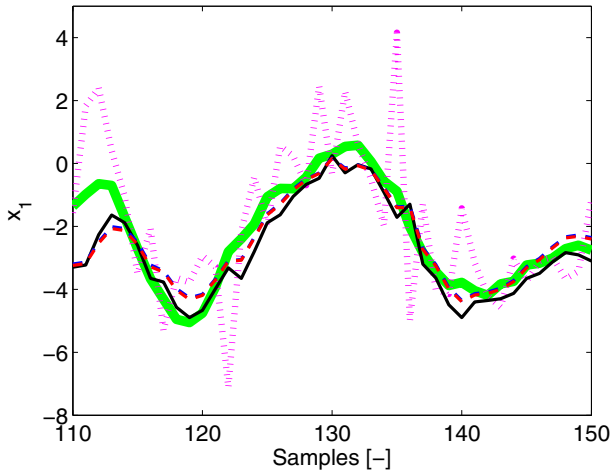

Fig. 1. Estimation of x_1 in Example 1: state trajectory (thick solid), trivial estimator (dotted), Kalman filter using the real covariance matrices Q and R (thin solid), Kalman filter with tuning of Λ, according to Theorem 2 (dash-dotted), Kalman filter with standard tuning of both Q and R (dashed). Notice that the last two lines are almost overlapping.

2×2 identity matrices, the Matlab function `fmincon` is used to solve Problems (5) and (18) and the asymptotic, time-invariant, Kalman gains are implemented.

Notice that, although the result is suboptimal with respect to the ideal filter using real Q and R (a local minimum is achieved with the above initialization), the standard and the proposed strategies almost coincide, and the corresponding estimates in Figure 1 are practically overlapping. This fact can be verified also numerically, by looking at the achieved value of $\Lambda = RQ^{-1}$ and the mean squared error (MSE) between the considered predictors and the ideal one (see Table 1).

Figure 1 also shows that the trivial estimation would be an unfortunate choice, since the variance of the noise on $y_1(t)$ is evidently too large for an accurate filtering of $x_1(t)$.

	Standard	**Reduced**
λ_1	6.6258	6.7916
λ_2	0.9338	0.9143
MSE	0.16144	0.038773

Table 1. Mean squared errors (MSE) with respect to the ideal Kalman predictor and entries of Λ.

Finally, notice that, on a Intel i7 processor at 2.67 GHz, the computation time using `fmincon` for the standard parameterization is 1.99 s, whereas for the reduced parameterization becomes 1.48 s. In other words, the computational time is decreased of 35%. This difference is obviously not crucial in the proposed toy example, but might become significant when the dimension of the system is large.

4.2 Example 2

Consider the system (1), with $n = 2$ states and $p = 1$ output, described by the matrices

$$F = \begin{bmatrix} -0.5 & 0.5 \\ -0.25 & 0.95 \end{bmatrix}, \; Q = \begin{bmatrix} 1.059 & 1.054 \\ 1.054 & 1.051 \end{bmatrix},$$

$$H = [1 \; 1.5], \; R = 0.01.$$

From the above selection of Q, it is evident that the process noises on the two state equations are highly correlated each other. Therefore, a diagonal parameterization of Q is far from being suitable in this case.

However, since F can be shown to be equivalent to the real diagonal matrix

$$F_d = \begin{bmatrix} -0.4079 & 0 \\ 0 & 0.8579 \end{bmatrix},$$

by means of the transformation $F_d = T^{-1}FT$, where

$$T = \begin{bmatrix} -0.9835 & -0.3455 \\ -0.1811 & -0.9384 \end{bmatrix},$$

i.e., **A3** holds, the result in Theorem 3 can be applied to find a full matrix Q without increasing the complexity of the optimization problem.

Specifically, let the covariance matrix for the process noise affecting the diagonal system be Q_d. Collect a set of data by injecting the white noise $v_x(t) = [v_{x_1}(t), \; v_{x_2}(t)]^T$, where $v_{x_2}(t) = v_{x_1}(t) + \epsilon(t)$, $v_{x_1}(t) \sim WN(0,1)$ and $\epsilon(t) \sim WN(0, 0.0025)$.

Problem 5 can then be solved using data, by simply parameterizing Q_d as a diagonal matrix, as usually done in practice. The gain of the Kalman predictor for the original system is given by $K(t) = TK_d(t)$, where $K_d(t)$ is the Kalman gain for the diagonal system, directly coming from the solution of the Riccati equation using Q_d and R_d.

In Figure 2, the performance of such a predictor is shown for steady-state prediction of the second state variable $x_2(t)$. In the same Figure, the ideal Kalman predictor using the real covariance matrices Q and R and the standard tuning optimizing over diagonal matrices are illustrated. In both the proposed and the standard method, the Matlab function `fmincon` has been used to solve Problem (5). Moreover, in both cases, the same initialization values are used. Specifically, the initial value for the covariance of the process noise is $Q_0 = I_2$, where I_2 is the 2×2 identity matrix, whereas the variance of the output noise has been initialized at its real value, since here it is not the critical task.

	Diagonal parameterization	**Proposed method**
MSE x_1	0.072551	0.25275
MSE x_2	0.44257	0.34181

Table 2. Mean squared error (MSE) between the ideal Kalman predictor and the standard strategy, compared to MSE between ideal Kalman predictor and the proposed method.

As expected, the method proposed herein provides better performance in case the original Q is not diagonal. This fact concerns the whole state vector, as shown in Table 2, where the mean squared error (MSE) between the standard version of the Kalman predictor and the ideal one with the real Q is compared with the MSE between the ideal predictor and the proposed one.

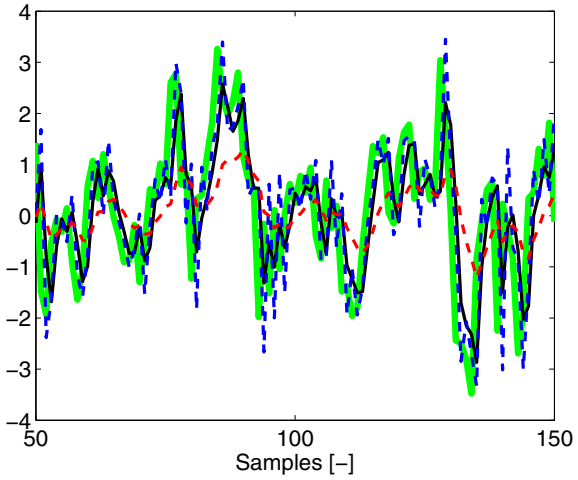

Fig. 2. Prediction of x_2 in Example 2: state trajectory (thick solid), Kalman prediction using the real covariance matrices Q and R (thin solid), Kalman prediction using the result of Theorem 3 (dash-dotted), Kalman prediction using diagonal matrices (dashed).

The performance can be obviously shown to be the same in case the real Q is diagonal and local minima do not occur.

5. CONCLUSIONS

In this paper, we addressed the problem of tuning the process and noise covariance matrices in Kalman formulas for prediction problems. Specifically, we considered two subproblems: the issue of decreasing the number of optimization variables when the matrices are parameterized as diagonal matrices and the issue of parameterizing them as full matrices. For the first problem, we considered the scalar case, by reformulating the result in Bittanti and Savaresi [2000] and we extended it to the MIMO case for the specific setting where $n = p$. Concerning the second problem, we showed that, under the assumption of a diagonalizable state matrix, it is possible to find a full covariance matrix for process noise, without increasing the complexity of the design issue. The results showed to be effective on two numerical examples, where implementation issues were also discussed in detail.

Future work will deal with a number of real-world applications, as well as the extension to nonlinear filtering problems.

REFERENCES

B.M. Åkesson, J.B. Jørgensen, N.K. Poulsen, and S. Jørgensen. A generalized autocovariance least-squares method for kalman filter tuning. *Journal of Process control*, 18(7):769–779, 2008.

B. Anderson and J. Moore. *Optimal filtering*. Prentice-Hall, 1979.

S. Bittanti and S.M. Savaresi. On the parametrization and design of an extended kalman filter frequency tracker. *IEEE Transactions on Automatic Control*, 45(9):1718–1724, 2000.

B. Carew and P. Bélanger. Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 18(6):582–587, 1973.

M.S. Grewal and A.P. Andrews. *Kalman filtering: theory and practice using MATLAB*. John Wiley & Sons, 2008.

R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83 (3):95–108, 1961.

F.L. Lewis, L. Xie, and D. Popa. *Optimal and robust estimation: with an introduction to stochastic control theory*. CRC Press, 2008.

L. Ljung. System identification: theory for the user. 1999.

P. Matisko and V. Havlena. Noise covariance estimation for kalman filter tuning using bayesian approach and monte carlo. *International Journal of Adaptive Control and Signal Processing*, 27(11):957–973, 2013.

R. Mehra. On the identification of variances and adaptive kalman filtering. *IEEE Transactions on Automatic Control*, 15(2):175–184, 1970.

R. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17(5):693–698, 1972.

B.J. Odelson, M.R. Rajamani, and J.B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, 2006.

D. Simon. *Optimal state estimation: Kalman, H-infinity, and nonlinear approaches*. Wiley – Interscience, 2006.

M. Verhaegen and V. Verdult. *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.