PREFERENTIAL ESTIMATION VIA TUNING OF THE KALMAN FILTER

L. Bodizs, B. Srinivasan and D. Bonvin

Laboratoire d'Automatique
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Abstract: Estimation problems have been traditionally formulated so as to minimize the estimation error of the full state vector. However, in applications that involve the tracking of only a few unmeasured variables, it is sufficient to limit the attention along certain directions in state space. This way, it is hoped that better accuracy can be obtained along the desired directions, possibly at the cost of poorer estimates along the other directions. This problem, termed preferential estimation, is formally formulated in this paper as a least-squares minimization problem. Using calibration measurements of the preferred variables, the above mentioned problem is solved numerically via tuning of the Kalman filter. The approach is illustrated in simulation on the optimization of a penicillin fermentation process, where preferential estimation is used successfully to reduce the error in tracking a single unmeasured variable, the substrate concentration.

Keywords: State estimation, Kalman filter, Preferential estimation, Tuning via optimization, Penicillin fermentation.

1. INTRODUCTION

State estimation is an important topic within the theory of dynamic systems (Maybeck, 1979; Brown and Hwang, 1992). State estimation is a necessary component of sophisticated monitoring and control techniques since they typically require information that is too expensive or impossible to obtain from direct measurements. Estimation attempts to reconstruct the needed information from the available measurements and prior knowledge in the form of a dynamic model (Dochain, 2003; Doyle, 1998).

Traditional estimation techniques estimate the full state vector, and the main issues are how good these estimates are in terms of bias and variance (Chui and Chen, 1987). The reason for estimating all the states is due to the close link between the estimation theory and the full-state feedback literature. However, information on all the states is not always needed, especially when the objective is to track only a few unmeasured variables.

For cases where only some of the state variables have to be estimated, reduced-order estimators have been proposed in the literature. The order of such estimators typically corresponds to the difference between the system order and the number of measurements (Luenberger, 1979). These techniques have been applied succesfully to chemical reactors (Soroush, 1997) and utilized for inferential control purposes (Doyle, 1998). The main drawback is that the reduction in the number of variables to estimate is generally not significant, due to the small number of measured variables.

More important reduction in the order of the estimator can be achieved through projection methods (Marquardt, 2001). These methods rely on the fact that the most important characteristics (for a specific purpose) of the dynamics of a complex system can be described by the dominant sub-

space. Thus, the system can be described by a dynamic model whose dimension is equal to that of the dominant subspace. Significant reduction has been reported for various applications (Hahn and Edgar, 2002). However, the projection methods often need to be tailor made to an application, and there are too few guidelines for selecting the appropriate method for a particular application.

Hence, new alternatives for estimating accurately a subset of the states are welcome, and this paper aims at proposing such an alternative. The idea of the proposed method is to focus attention only along certain preferred directions, where a better accuracy is required, at the cost of possibly poorer estimation along the other directions. This problem, termed *preferential estimation*, is formulated formally in this paper. Increasing the accuracy along the preferred directions is done using one-time calibration measurements. Thus, preferential estimation draws a strong parallel with the calibration approach, and many concepts therein can be used (Martens and Naes, 1989).

The main question in preferential estimation is whether such a thing is at all possible. Can the error along certain preferred directions be reduced by increasing the errors in other directions ? This will be shown with a bioreactor example, where the parameters of a Kalman filter used to reconstruct a single state will be tuned via numerical optimization. It turns out that preferential estimation is indeed possible and can be used advantageously for tracking a desired substrate profile.

The paper is organized as follows. In Section 2, the standard full-state estimation problem and the Kalman filter are briefly reviewed. Section 3 describes the concept of preferential estimation, while Section 4 deals with the tuning of the estimator via numerical optimization. The application of preferential estimation is illustrated on a simulated batch fermentation process in Section 5. Section 6 discusses some features of preferential estimation, and Section 7 concludes the paper.

2. STANDARD ESTIMATION

Consider the following discrete-time system:

$$x_{k+1} = G(x_k, u_k) + w_k, \quad x_0 = x_o \quad (1)$$
$$y_k = H(x_k) + v_k \quad (2)$$

where u are the inputs, x the states, x_o the initial values of states, w the process noise, y the measurements, and v the measurement noise. G are the nonlinear functions describing the system dynamics, and H the nonlinear measurement functions.

The standard estimation problem consists of determining the best state estimates, $\hat{x}$, given the measurements y. One of the most successful full state estimators used in practice is the Extended Kalman Filter (EKF) (Kalman, 1960; Jazwinski, 1970; Grewal and Andrews, 1993). It determines the Kalman gain matrix K_k that minimizes the a posteriori estimation error covariance P_k:

$$\min_{K_k} \quad P_k = E\langle (x_k - \hat{x}_k)(x_k - \hat{x}_k)^T \rangle \quad (3)$$
$$s.t. \ \hat{x}_k^- = G(\hat{x}_{k-1}, u_{k-1}), \quad \hat{x}_0 = E\langle x_o \rangle \quad (4)$$
$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - H(\hat{x}_k^-)) \quad (5)$$

where $\hat{x}_k^-$ is the a priori estimate, and $\hat{x}_k$ the a posteriori estimate.

The solution of Problem (3)-(5) is:

$$P_k^- = A_k P_{k-1} A_k^T + Q, \quad P_0 = E\langle x_o x_o^T \rangle$$
$$P_k = (I - K_k C_k)P_k^- \quad (6)$$
$$K_k = P_k^- C_k^T (C_k P_k^- C_k^T + R)^{-1}$$

where P_k^- is the a priori estimation error covariance, Q and R the process and measurement noise covariances, A_k the Jacobian of G with respect to $\hat{x}_k$, and C_k the Jacobian of H with respect to $\hat{x}_k^-$.

3. CONCEPT OF PREFERENTIAL ESTIMATION

In many applications, the main interest is on estimating rather accurately the values of z, a vector of much smaller dimension than x. Consider the following extension to System (1)-(2):

$$x_{k+1} = G(x_k, u_k) + w_k, \quad x_0 = x_o$$
$$y_k = H(x_k) + v_k$$
$$z_k = L(x_k), \qquad \dim(z_k) \ll \dim(x_k) \quad (7)$$

where z_k are the quantities to be estimated, and L a nonlinear function.

One way of obtaining an estimate of z is to estimate $\hat{x}$ with standard techniques and then use the function L to get $\hat{z}_k = L(\hat{x}_k)$. The drawback of this approach is that the problem formulation (3) concentrates on the accuracy of $\hat{x}$ rather than that of $\hat{z}$. In other words, the full state $\hat{x}$ is computed as accurately as possible and then used to compute $\hat{z}$. Consequently, the resulting $\hat{z}$ will inherit the accuracy of $\hat{x}$, though it could probably be estimated more accurately if the attention were placed exclusively on its estimation.

In contrast, the objective of this paper is not to maximize the accuracy of $\hat{x}$ (standard estimation,

SE), but to address directly the problem of maximizing the accuracy of $\hat{z}$ (preferential estimation, PE) regardless of the accuracy of $\hat{x}$.

Thus, the objective of PE is to focus on estimating $\hat{z}$ as accurately as possible. An optimization problem can be formulated as follows:

$$\min_{K_k} \quad J_k = E\langle (z_k - \hat{z}_k)^T (z_k - \hat{z}_k) \rangle \qquad (8)$$

$$s.t. \quad \hat{x}_k^- = G(\hat{x}_{k-1}, u_{k-1}), \quad \hat{x}_0 = E\langle x_o \rangle$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - H(\hat{x}_k^-))$$

$$\hat{z}_k = L(\hat{x}_k) \qquad (9)$$

where J_k is the mean-squared estimation error in z_k. There are two main differences between the formulations (3) and (8):

- The former deals with the states x, while the latter concerns only the preferred directions z, whose dimension is much smaller than x
- The former minimizes a matrix that represents the estimation error covariance of x, while the latter deals with a scalar that corresponds to the mean-squared estimation error in z.

The reason for choosing a scalar cost function in (8) rather than a matrix as in (3) is that, except for the special case of linear systems without uncertainty, there exists no unique choice of the decision variables that minimizes every element of the matrix. Thus, a weighted sum of the various elements of the matrix is necessary in order to define a solution. The mean-squared estimation error, $E\langle (z_k - \hat{z}_k)^T (z_k - \hat{z}_k) \rangle$, which is the trace of the matrix $E\langle (z_k - \hat{z}_k)(z_k - \hat{z}_k)^T \rangle$, represents one possible weighting of the covariance matrix.

Since Problem (8) has no analytical solution, a numerical approach is needed. However, this approach requires the knowledge of z_k in order to evaluate the objective function J_k. Thus, z_k has to be measured once, during the tuning process, in order to obtain the preferential estimator. But once the estimator is tuned, measurements of z_k are no longer needed. The computation of the expectation in (8) requires numerous realizations (obtained via Monte-Carlo simulation) with known output variables y_k and desired variables z_k. Thus, a set of calibration measurements (y_k, z_k) is necessary for preferential estimation.

Hence, a parallel to calibration can be drawn, where the calibration model is tuned based on experimental data and then used for prediction (Martens and Naes, 1989). The main difference between calibration and preferential estimation are: (i) Traditional calibration methods are static, while preferential estimation deals with dynamic systems, (ii) Dynamic calibration methods such as dynamic PLS (Lakshminarayanan *et al.*, 1997),

work with black-box dynamic prediction models built just on the available calibration data. On the contrary, preferential estimation a mechanistic, first-principles or other system models (built for other purposes from a different information) along with the available calibration measurements.

The numerical solution is computationally expensive since it has to test each choice of the estimator parameters K_k over numerous realizations. Though many ideas (such as analytically computing an approximation of the cost function without taking an expectation using Monte-Carlo realizations) can be used to ease this numerical burden, a brute-force approach will be used in this paper for tuning an extended Kalman filter.

4. PREFERENTIAL ESTIMATION VIA TUNING OF EXTENDED KALMAN FILTER

In the ideal case, i.e. linear process model and white noise of known covariances (Q and R are chosen equal to the true process and noise covariances that are assumed to be known), the Kalman filter is the best unbiased estimator. In this case, the P_k obtained upon convergence reflects the true estimation error covariance, and the expectation with infinitely many realizations is the true value, i.e. $E\langle \hat{x}_k \rangle = x_k$.

However, this correspondence is lost in other cases (nonlinearities and/or model mismatch and/or non-white noise), and the Q and R matrices are simply used as tuning parameters, for which tuning rules are available (Grewal and Andrews, 1993; Chui and Chen, 1987). The objective here is typically to minimize the estimation error of $\hat{x}$ (standard estimation). However, Q and R could also be tuned so as to minimize the estimation error of $\hat{z}$ (preferential estimation).

Thus, in the context of the extended Kalman filter with Q and R being considered as tuning parameters, SE and PE can be formulated as the following optimization problems involving Q and R as decision variables:

- SE - minimizes the estimation error $\bar{J}_k = E\langle (x_k - \hat{x}_k)^T (x_k - \hat{x}_k) \rangle = \text{tr}(P_k)$

$$\min_{Q,R} \quad \bar{J}_k = E\langle (x_k - \hat{x}_k)^T (x_k - \hat{x}_k) \rangle \ (10)$$

$$s.t. \ (4) - (6)$$

- PE - minimizes the estimation error $J_k = E\langle (z_k - \hat{z}_k)^T (z_k - \hat{z}_k) \rangle$

$$\min_{Q,R} \quad J_k = E\langle (z_k - \hat{z}_k)^T (z_k - \hat{z}_k) \rangle \ (11)$$

$$s.t. \ (4) - (6), (9)$$

The data points x_k and z_k are obtained from a simulated reality. Problems (10) and (11) can

be solved using numerical optimization where the expectation is computed via Monte-Carlo simulations. An additional difficulty is that these optimization problems exhibit many local minima and, thus, a search for the global minimum is necessary.

In this paper, the two-step numerical optimization method proposed in (Nagy *et al.*, 2001) is used:

(1) A global search method (Genetic Algorithm, GA (Houck *et al.*, 1995)) is used as pre-optimization to find the neighborhood of the global minimum, and

(2) A gradient-based local search algorithm ('fmincon' from Matlab) is used to find the global minimum.

This combined method is computationally very expensive, especially in combination with Monte-Carlo simulations. However, all these efforts are off-line and so, for the moment, the speed of solving (10) and (11) has not been given much importance.

5. PENICILLIN FERMENTATION - IMPLEMENTATION OF PREFERENTIAL ESTIMATION

The proposed estimation techniques are tested on the fed-batch penicillin fermentation process (Srinivasan *et al.*, 2003). The reactions taking place are:

$$S \rightarrow X, \quad S \xrightarrow{X} P$$
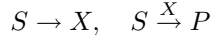
where S, X and P are the substrate, biomass and product, respectively. The specific rate of the first reaction is described by a Haldane type expression, while that of the second reaction is constant. The third-order dynamic model of the system reads:

$$\dot{c}_X = \mu(c_S)c_X - \frac{F}{V}c_X, \; c_X(0) = c_{Xo}$$

$$\dot{c}_S = -\frac{\mu(c_S)c_X}{Y_X} - \frac{\nu c_X}{Y_P} + \frac{F}{V}(c_{Sin} - c_S), \; c_S(0) = c_{So}$$

$$\dot{V} = F, \; V(0) = V_o \quad (12)$$

$$c_P = \frac{c_{Po}V_o}{V} + c_{Sin}(V - V_o)\frac{Y_P}{V} - (c_S V -$$

$$- c_{So}V_o)\frac{Y_p}{V} - (c_X V - c_{Xo}V_o)\frac{Y_P}{Y_X V}$$

$$\mu(c_S) = \frac{\mu_m c_S}{K_m + c_S + \frac{c_S^2}{K_i}}$$

where c_S, c_X and c_P are the substrate, biomass and product concentrations, respectively, V the volume, F the feed flow rate, c_{Sin} the inlet substrate concentration, ν, μ_m, K_m and K_i kinetic parameters. The numerical values are given in Table 1.

Table 1. Model parameters, initial conditions and operating bounds

μ_m	0.53	L/h
K_m	1.2	g/L
K_i	22	g/L
Y_X	0.4	
Y_P	1	
ν	0.5	L/h
c_{Sin}	20	g/L
c_{Xo}	1	g/L
c_{So}	5.138	g/L
c_{Po}	0	g/L
V_o	2	L
F_{min}	0	L/h
F_{max}	1	L/h
c_{Xmax}	3	g/L
t_f	8	h

Consider the problem of maximizing the amount of product at the given final time t_f by adjusting the feed flow rate:

$$\max_{F(t)} \quad c_P(t_f) \quad (13)$$

$$s.t. \; (12)$$

$$c_X(t) \leq c_{Xmax}, \quad F_{min} \leq F(t) \leq F_{max}$$

The solution of (13) corresponds to maximizing the rate of the first reaction $\mu(c_S)$ (Srinivasan *et al.*, 2003). This can be achieved by maintaining c_S at $c_S^* = \sqrt{(K_m K_i)} = 5.131$ g/L, the maximum of the Haldane-type expression. Therefore, the capability of tracking c_S^* accurately is vital for optimal process operation.

On-line measurement of c_S is typically not available or is too expensive, while that of c_P is mostly available. So, the problem of state estimation arises. Also, since optimal process operation requires the tracking of a single state, it is of interest to formulate the estimation problem as that of minimizing the estimation error in c_S. Thus, using the notations of the earlier sections, $x_k = [c_{Xk}, c_{Sk}, V_k]^T$; $u_k = F_k$; $y_k = c_{Pk}$; $z_k = c_{Sk}$; G - the discretized version of (12) using the sampling time $\Delta t = 0.01s$. This model is used as the simulated reality, and zero-mean measurement noise (v_k) with 5% variance is considered.

The model used for estimation has the same structure as G, but with errors in the parameters and initial conditions: $\mu_m = 0.5$ L/h; $K_m = 1$ g/L; $K_i = 20$ g/L; $c_{Xo} = 0.6$ g/L; $c_{So} = 5.4$ g/L.

The estimation problem is solved in both the standard and the preferential contexts by solving numerically Problems (10) and (11). The performance criterion, the bias, and the standard deviation are presented in Table 2. The error distribution plots are given in Figure 1, while the time evolution of one realization is shown in Figure 2.

Table 2. Comparison of SE and PE over 100 realizations ($\bar{J}$ - estimation error for x; J - estimation error for z; $E\langle x - \hat{x}\rangle$ - bias; σ_x - standard deviation)

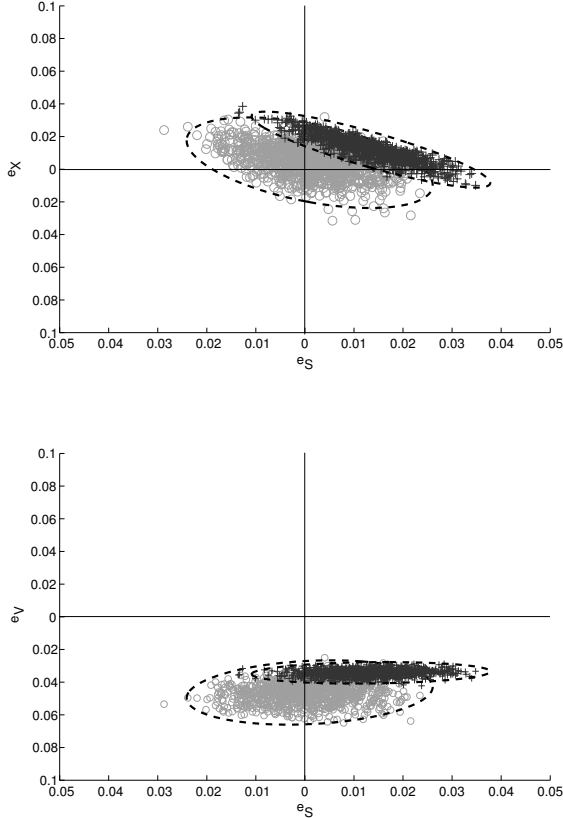|	SE	PE
$\bar{J}$	3.7171	4.7727
J	0.7850	**0.3378**
$E\langle x - \hat{x}\rangle$	0.0120 0.0135 −0.0343	0.0040 **0.0010** −0.0463
σ_x	0.0077 0.0081 0.0022	0.0092 0.0084 0.0066

Fig. 1. Error distribution plots of $[e_X, e_S, e_V]^T = x_k - \hat{x}_k$ for $1h \leq t \leq 4h$ over 100 realizations (+ (dark gray) - SE; o (light gray) - PE).

The estimation error of $\hat{z} = \hat{c}_S$ is reduced considerably in PE by reducing the bias in $\hat{c}_S$ (Table 2, Figure 2). This is also visible on the distribution plots (Figure 1), where e_S is centered around 0 in PE. The price to pay, however, is a greater bias in V and an increased variance in all three states, which is seen both in Table 2 and Figure 1.

Due to the reduced bias in $\hat{c}_S$, the tracking of c_S^* based on PE is more accurate, as illustrated on the time evolution of one realization in Figure 3. Hence, a slightly higher amount of product at final time can be obtained: $c_P(t_f)_{PE} = 2.428$ g/L $> c_P(t_f)_{SE} = 2.419$ g/L.
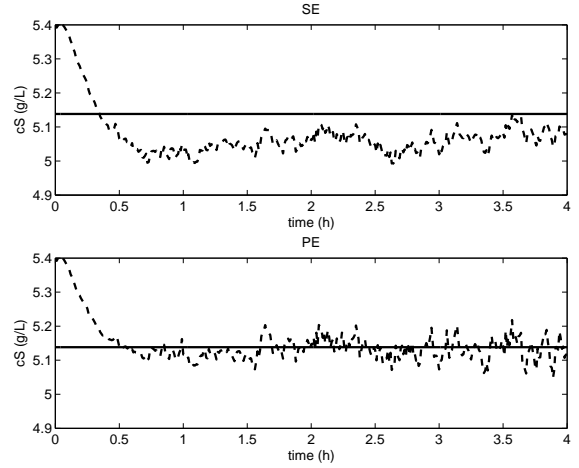

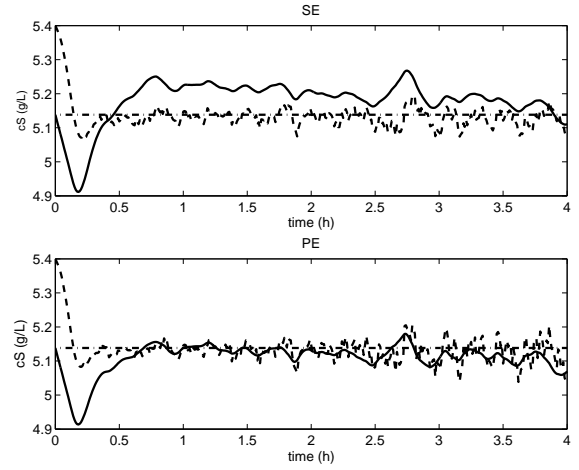Fig. 2. Estimation of c_S using SE and PE (dashed - $\hat{c}_S$; continuous - c_S)

Fig. 3. Tracking of c_S^* using a P controller and $\hat{c}_S$ based on SE and PE (dashed-dot - c_S^*; dashed - $\hat{c}_S$; continuous - c_S)

6. DISCUSSION

The previous section has shown that it is possible to reduce the mean-squared estimation error in z, the preferred direction. The mean-squared estimation error has two parts:

$$\text{tr } E\langle e_{z_k} e_{z_k}^T\rangle = \text{tr } E\langle \bar{e}_{z_k} \bar{e}_{z_k}^T\rangle + \text{tr } E\langle (e_{z_k} - \bar{e}_{z_k})(e_{z_k} - \bar{e}_{z_k})^T\rangle \quad (14)$$

where $e_{z_k} = z_k - \hat{z}_k$, and $\bar{e}_{z_k}$ is its mean over various time instants and realizations. $\text{tr } E\langle \bar{e}_{z_k} \bar{e}_{z_k}^T\rangle$ is the bias and $E\langle (e_{z_k} - \bar{e}_{z_k})(e_{z_k} - \bar{e}_{z_k})^T\rangle$ is the covariance.

It is possible to reduce the mean-squared estimation error by a tradeoff between the bias and variance parts of (14). In the bioreactor example, the bias part of $\hat{z}$ was reduced and the variance slightly increased to achieve the desired goal. In some sense, PE uses the bias-variance tradeoff that is well-known in statistical literature (Hoerl

and Kennard, 1970). Note that the tradeoff used here is the opposite of that used in partial least-squares techniques, where the bias is increased to reduce the variance (Martens and Naes, 1989).

In addition, tradeoffs between the different directions can be observed. For example, the bias along V is increased in order to reduce the bias along c_S. Thus, in the context of PE, bias-bias tradeoffs between different directions can also exist.

In the ideal case (linear model, no model mismatch and zero-mean gaussian white noise), the bias is zero. In this case, the Kalman filter is the best unbiased estimator (Section 4), and all the elements of $E\langle (x_k - \hat{x}_k)(x_k - \hat{x}_k)^T \rangle$ are minimal. However, it should be noted that, in certain cases, a biased estimator may result in a smaller mean-squared estimation error. This issue has been extensively studied in the context of static systems and calibration (principle component regression, partial least squares, etc.), while there has been little study on biased estimation techniques for dynamic systems. This is one of the issues preferential estimation addresses, but it clearly calls for a different estimator structure, i.e. simply tuning the parameters of a Kalman filter will not do it.

In other cases (nonlinear model and/or model mismatch and/or non-white noise), the bias is non-zero. This provides much more potential for preferential estimation as was seen in the bioreactor example, and various tradeoffs can be exploited in this context. In minimizing the mean-squared estimation error, the tradeoff between the bias and the variance of a given variable (direction) naturally arises. Either the variance can be reduced by increasing the bias (as in usual bias-variance tradeoff) or the bias can be reduced by increasing the variance (as in the example of Section 5). Which option to choose depends upon the relative magnitudes of the bias and variance. In the example considered in this paper, the bias was more important (modeling error, errors in initial conditions) and, consequently, the bias in $\hat{c}_S$ was reduced and its variance increased.

7. CONCLUSION

This paper has formulated the problem of preferential estimation, where the issue is more accurate estimation along a few preferred directions in state space. PE was implemented on a penicillin fermentation example by tuning a Kalman filter via numerical optimization. It was shown that PE exploits the bias-variance and bias-bias tradeoffs. Searching for a systematic way of influencing these tradeoffs, instead of the numerical optimization method used in this paper, constitutes a promising future research direction.

REFERENCES

Brown, R. G. and P. Y. C. Hwang (1992). *Introduction to Random Signals and Applied Kalman Filtering*. 2nd ed.. John Wiley.

Chui, C. K. and G. Chen (1987). *Kalman Filtering with Real-Time Applications*. Springer-Verlag.

Dochain, D. (2003). State and parameter estimation in chemical and biochemical processes : a survey. *To appear in J. Process Control*.

Doyle, F. J. (1998). Nonlinear inferential control for process applications. *J. Process Contr.* **8**, 339–353.

Grewal, M. S. and A. P. Andrews (1993). *Kalman Filtering. Theory and Practice*. Prentice Hall.

Hahn, J. and T. F. Edgar (2002). An improved method for nonlinear model reduction using balancing of empirical gramians. *Comp. Chem. Eng.* **26**, 1379–1397.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**(1), 55–67.

Houck, C., J. Joines and M. Kay (1995). A Genetic Algorithm for Function Optimization: A Matlab Implementation. Technical Report NCSU-IE-TR-95-09. North Carolina State University, Raleigh, NC.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME-Journal of Basic Engineering* pp. 35–45.

Lakshminarayanan, S., S. L. Shah and K. Nandakumar (1997). Modeling and control of multivariable processes: a dynamic PLS approach. *AIChE J.* **43**, 2307–2322.

Luenberger, D. G. (1979). *Introduction to Dynamic Systems*. John Wiley, New York.

Marquardt, W. (2001). Fundamental modeling and model reduction for optimization based control of transient processes. In: *Chemical Process Control-6*. Tucson, Arizona.

Martens, H. and T. Naes (1989). *Multivariate Calibration*. John Wiley.

Maybeck, P. S. (1979). *Stochastic Models, Estimation and Control*. Vol. 1. Academic Press.

Nagy, Z., S. Agachi, F. Allgower, R. Findeisen, M. Diehl, H. G. Bock and J. P. Schloder (2001). Using genetic algorithm in robust nonlinear model predictive control. In: *ESCAPE-11*. Kolding, Denmark.

Soroush, M. (1997). Nonlinear state-observer design with application to reactors. *Chem. Eng. Science* **52**, 387–404.

Srinivasan, B., S. Palanki and D. Bonvin (2003). Dynamic optimization of batch processes: I. Characterization of the nominal solution. *Comp. Chem. Eng.* **27**, 1–26.