Robust Filtering for Gene Expression Time Series Data with Variance Constraints (code: SI-ZW)

Guoliang Wei, Zidong Wang, Huisheng Shu, Karl Fraser and Xiaohui Liu

Abstract

In this paper, an uncertain discrete-time stochastic system is employed to represent the model for gene regulatory networks from time series data. A robust variance-constrained filtering problem is investigated for a gene expression model with stochastic disturbances and norm-bounded parameter uncertainties, where the stochastic perturbation is in the form of a scalar Gaussian white noise with constant variance and the parameter uncertainties enter both the system matrix and the output matrix. The purpose of the addressed robust filtering problem is to design a linear filter such that, for the admissible bounded uncertainties, the filtering error system is Schur stable and the individual error variance is less than the prespecified upper bound. By using the linear matrix inequality (LMI) technique, sufficient conditions are first derived for ensuring the desired filtering performance for the gene expression model. Then, the filter gain is characterized in terms of the solution to a set of LMIs, which can be easily solved by using available software packages. A simulation example is exploited for a gene expression model in order to demonstrate the effectiveness of the proposed design procedures.

Keywords

Microarray data; gene expression model; robust filtering; linear matrix inequality; stochastic disturbances; error variance constraints.

I. Introduction

DNA microarray technology [16] has made it possible to measure gene expression levels on a genomic scale, and has therefore been extensively applied to gene transcription analysis. Data thus collected promise to enhance fundamental understanding of life on the molecular level, from regulation of gene expression and gene function to cellular mechanisms, and may prove useful in medical diagnosis, treatment, and drug design. Simultaneously, the large scale of the data stimulates the development of mathematical analysis tools that are adaptable to reducing the complexity of the data to make it comprehensible. Substantial effort is being made to build models to analyze microarray data.

Clustering analysis of the gene expression data has recently drawn considerable research interest. Almost every clustering algorithm has been examined on gene expression data, such as hierarchical clustering [11] and self-organizing map [19]. However, a fundamental shortcoming of such clustering schemes is that they are based on the assumptions that there exists the correlation similarity between genes. It is also important to note that models based on clustering analysis are static and thus can not describe the dynamic evolution of

G. Wei is with the School of Information Science and Technology, Donghua University, Shanghai 200051, China.

Z. Wang, K. Fraser and X. Liu are with the Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom.

H. Shu is with the Department of Applied Mathematics, Donghua University, Shanghai 200051, China.

All correspondences concerning this paper should be addressed to Z. Wang. (Email: `Zidong.Wang@brunel.ac.uk`)

gene expression. Recently, there has been much interest to reconstruct models for gene regulatory networks from time series data [8,18], such as boolean network model [1,12], linear differential equation model [5,7,10], Bayesian model [13,14], state space model [3,17,24] and stochastic model [6,20], and a lot of algorithms have been applied to identify the system parameters such as principal component analysis (PCA) method and expectation-maximization (EM) algorithm. Examining real gene expression data, it seems clear that genes spend a lot of their time at intermediate values: gene expression levels tend to be continuous rather than binary [9]. Therefore, in recent years, linear dynamical systems described by differential equations (also called connectionist model, linear transcription model, additive regulation model, etc.) have frequently been exploited to model gene networks, where the main idea is to use an update rule based on a weighted sum of inputs.

One implicit assumption in continuous-valued models is that fluctuations in the range of single molecules can be ignored. However, molecules present at only a few copies per cell do play an important role in some biological phenomena. In that case, it may be impossible to model the behaviour of the system exactly with a *purely deterministic* model [2]. Analysis of mRNA and protein decay rates and abundances may allow us to identify those genes for which *stochastic* modelling may prove necessary [9]. Additional sources of unpredictability can include external noise, or errors on measured data, which could also be regarded as stochastic disturbances. Unfortunately, in the aforementioned linear modelling techniques, the stochastic effects have not been taken into account, mainly due to the mathematical difficulties. In this paper, one of our purposes is to model the gene expression as a linear stochastic model in order to reflect the reality in a more reasonable way.

On the other hand, it is well known that, the modelling error is unavoidable in practice. A mathematical model can by no means exactly represent the real gene network. Specifically, it is very likely that, the parameters of the model identified from the experimental data will vary from time to time, and such variations may be bounded but *unknown*. In other words, there should exist 'parameter uncertainties' in those models that are constructed from the real-time data. This leads to a new issue: given a mathematical model that contains parameter uncertainties and stochastic noises, how to estimate the variable values such that the mean square estimation error could be minimized in terms of some given criteria? This is actually a *robust data analysis* problem or, more precisely, a *robust filtering* problem. That is, the developed algorithm for estimating the real variable values should be *robust* enough to 'tolerate' the unknown modelling error within some acceptable accuracy. A robust filter should then be designed to 'remove' some unwanted errors and noises, and predict the real variable values by means of certain criteria. However, to the authors' best acknowledge, the modelling error issue has never been considered in modelling gene expression, not to mention the research on robust filtering problem. Therefore, in this paper, we will address and investigate the challenging 'robust filtering' problem in expressing gene networks in the presence of parameter uncertainties and stochastic disturbances.

In the context of robust filtering, we like to employ the idea from the constrained-variance filtering (CVF) theory, whose purpose is to design a filter such that the individual steady-state estimation error variance is not more than the prespecified upper bound. The CVF theory was first proposed in [25], and was then extended to the nonlinear case [15,26]. Comparing to the traditional Kalman filtering approach, the CVF theory provides a more straightforward technique to meet the prespecified estimation error variance constraints. In [22,23], the problem of the H_∞-norm and variance-constrained filter design has been studied for uncertain linear discrete-time observable stochastic systems, and the problem of robust filtering has been addressed in [21] for linear perturbed stochastic systems with bounded uncertainties. In the papers mentioned above, the Riccati-

equation approach has been intensively used where some parameters need to be tuned in the algorithms. In this paper, a less conservative linear matrix inequality (LMI) method will be developed and no parameter tuning is needed.

This paper is concerned with the robust variance-constrained filtering problem for a discrete-time state-space gene expression model with stochastic disturbances and norm-bounded uncertainties. The stochastic perturbation is described by a scalar Gaussian white noise with constant variance. We are interested in designing a linear filter such that, for the admissible norm-bounded uncertainties and stochastic noise disturbance, the filtering error system is Schur stable and the individual error variance is less than the prespecified upper bound. By using the LMI technique, sufficient conditions are first derived to ensure the desired filtering performance for the gene expression model. Then, the filter gain is characterized in terms of the solution to a set of LMIs, which can be easily solved by using available software packages. A simulation example, in which the main model parameters are borrowed from [24], is exploited to demonstrate the effectiveness of the proposed design procedures.

Notation In this paper, $\mathbb{R}^n$ and $\mathbb{R}^{n \times m}$ denote, respectively, the n dimensional Euclidean space and the set of all $n \times m$ real matrices. I denotes the identity matrix of compatible dimension. The notation $X \geq Y$ (respectively, $X > Y$) where X and Y are symmetric matrices, means that $X - Y$ is positive semi-definite (respectively, positive definite). M^T represents the transpose of M. $(\Omega, \mathcal{F}, \{\mathcal{F}_k\}_{k \in \mathbb{N}}, \mathcal{P})$ is a complete probability space with a filtration $\{\mathcal{F}_k\}_{k \in \mathbb{N}}$ satisfying the usual conditions (i.e., the filtration contains all P-null sets and is right continuous). $\mathbb{E}\{x\}$ stands for the expectation of stochastic variable x. The shorthand diag$\{M_1, ..., M_n\}$ denotes a block diagonal matrix with diagonal blocks being the matrices $M_1, ..., M_n$. In symmetric block matrices, the symbol $*$ is used as an ellipsis for terms induced by symmetry. Matrices, if not explicitly stated, are assumed to have compatible dimensions .

II. Problem Formulation

Consider the following discrete-time state-space description of a gene expression model with norm-bounded parameter uncertainties and stochastic disturbances:

$$(\Sigma): \quad x(k+1) = (A + \Delta A(k))x(k) + B\omega(k) \tag{1}$$

$$y(k) = (C + \Delta C(k))x(k) + D\omega(k), \tag{2}$$

where $y(k) = [y_1(k), ..., y_n(k)]^T \in \mathbb{R}^n$ consists of the observation variables of the system and $y_i(k)$ $(i = 1, ..., n)$ represents the expression level of ith gene at time k, where n is the number of genes in the model. The vector $x(k) = [x_1(k), ..., x_p(k)]^T$ consists of the internal state variables of the system and $x_i(k)$ $(i = 1, ..., p)$ represents the expression value of ith internal element at time k which directly regulates the gene expression, and p is the number of the internal state variables. The matrix $A = [a_{ij}]_{p \times p} \in \mathbb{R}^{p \times p}$ is the time translation matrix of the internal state variables or the state transition matrix, which provides key information on the influences of the internal variables on each other. The matrix $C = [c_{ki}]_{n \times p} \in \mathbb{R}^{n \times p}$ is the transformation matrix between the observation variables and the internal state variables. $\omega(k)$ is assumed to be a scalar Gaussian white noise with constant variance $\alpha^2 > 0$. B and D are constant matrices quantifying the intensity of the noise. $\Delta(A)(k)$ and $\Delta(C)(k)$ are time-variant uncertain norm-bounded matrices which represent parameter perturbations and are of the form

$$\begin{bmatrix} \Delta A(k) \\ \Delta C(k) \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} F(k)N, \tag{3}$$

where $F(k)$ is an uncertain time-varying matrix bounded by

$$F(k)F^T(k) \leq I, \tag{4}$$

and M_1, M_2 and N are known constant matrices of appropriate dimensions which specify how the elements of the nominal matrices A and C are affected by the uncertain perturbed parameters in $F(k)$. ΔA and ΔC are said to be *admissible* if both (3) and (4) hold.

Remark 1: In the state-space model (1)-(2), without loss of generality, the internal state variables and observation variables are assumed to be perturbed by the same scalar Gaussian white noise through different weighted matrices. Such kind of models have been studied by many researchers, see e.g. [21–23]. We like to point out that the main results of this paper can be easily extended to the case where the state and output noises are different. Compared with the gene expression model identified in [24], there are additive noises appearing in (1)-(2), which reflects the fact that stochasticity is an inherent feature of the gene expression time series, as discussed in the introduction.

The main aim of this paper is to estimate the internal variables through noisy measurement data in the presence of modelling errors. In this paper, the state estimation vector, $\hat{x}(k)$, satisfies the following linear filter of the form:

$$\Sigma_e : \hat{x}(k+1) = A\hat{x}(k) + Ky(k), \tag{5}$$

where the matrices K is filter gain to be determined.

The steady-state estimation error covariance is defined as

$$P = \lim_{k \to \infty} P(k) := \lim_{k \to \infty} \mathbb{E}\left\{e(k)e^T(k)\right\}, \tag{6}$$

where $e(k) := x(k) - \hat{x}(k)$ denotes the error state. Then, the filtering error system is obtained as follows:

$$\Sigma_e : \quad e(k+1) = Ae(k) + (\Delta A(k) - K(C + \Delta C(k))x(k) + (B - KD)\omega(k). \tag{7}$$

Augmenting (Σ) and (Σ_e), we have

$$\Sigma_a : \xi(k+1) = (\bar{A} + \Delta\bar{A}(k))\xi(k) + \bar{B}\omega(k), \tag{8}$$

where $\xi(k) = [x^T(k) \ e^T(k)]^T$, and

$$\bar{A} = \begin{bmatrix} A & 0 \\ -KC & A \end{bmatrix}, \ \bar{B} = \begin{bmatrix} B \\ B - KD \end{bmatrix}, \ \Delta\bar{A} = \bar{M}F(k)\bar{N}, \tag{9}$$

$$\bar{M} = \begin{bmatrix} M_1 \\ M_1 - KM_2 \end{bmatrix}, \ \bar{N} = [N \ 0]. \tag{10}$$

Remark 2: A matrix is said to be Schur stable if all its eigenvalues lie within the unit circle in the complex plane. In the filter (5), the Schur stability of the matrix A is a prerequisite for the filtering error system to be mean-square stable. Since the filter (5) can't affect the state of the original system and $x(k)$ is one part of $\xi(k)$, the mean-square stability of $x(k)$ is a necessary condition of the mean-square stability of $\xi(k)$.

It is well known that, if the matrix $\bar{A} + \Delta\bar{A}(k)$ is Schur stable for all admissible uncertainties, the steady-state covariance of system (Σ_a) defined by:

$$X := \lim_{k \to \infty} X(k) := \lim_{k \to \infty} \mathbb{E}[\xi(k)\xi^T(k)] := \begin{bmatrix} X_{xx} & X_{xe} \\ X_{xe}^T & P \end{bmatrix} \tag{11}$$

exists and satisfies the following discrete-time Lyapunov equation

$$X = (\bar{A} + \Delta\bar{A})X(\bar{A} + \Delta\bar{A})^T + \bar{B}\bar{B}^T. \tag{12}$$

The purpose of this paper is to study the robust variance-constrained filtering problem for the gene expression model (Σ) presented with parameter uncertainties and stochastic disturbances. That is, design the filter gain K, such that for all admissible uncertainties $\Delta A(k)$, $\Delta C(k)$ and stochastic disturbances $w(k)$, the following two requirements are *simultaneously* satisfied:

(I): The state matrix $\bar{A} + \Delta\bar{A}(k)$ is Schur stable.

(II): The steady-state error covariance P satisfies

$$[P]_{ii} \leq \sigma_i^2, \qquad i = 1, 2, ..., n, \tag{13}$$

where $[P]_{ii}$ means the ith diagonal element of P, i.e., the steady-state variance of the ith state. σ_i^2 ($i = 1, 2, ..., n$) denote the prespecified steady-state estimation error variance constraint on the ith state which can be determined by the practical requirements but should not be less than the minimal variance value obtained from the traditional H_2 estimation theory.

III. MAIN RESULTS

First, let us give the following lemmas which will be used in the proof of our main results.

Lemma 1: (Schur Complement) Given the constant matrices $\Sigma_1, \Sigma_2, \Sigma_3$ where $\Sigma_1 = \Sigma_1^T$ and $0 < \Sigma_2 = \Sigma_2^T$. Then $\Sigma_1 + \Sigma_3^T\Sigma_2^{-1}\Sigma_3 < 0$ if and only if

$$\begin{bmatrix} \Sigma_1 & \Sigma_3^T \\ \Sigma_3 & -\Sigma_2 \end{bmatrix} < 0,$$

or equivalently,

$$\begin{bmatrix} -\Sigma_2 & \Sigma_3 \\ \Sigma_3^T & \Sigma_1 \end{bmatrix} < 0.$$

Lemma 2: Let X, Y, F be real matrices of appropriate dimensions with $F^TF \leq I$. Then for any scalar $\delta > 0$, we have

$$XFY + Y^TF^TX^T \leq \delta XX^T + \delta^{-1}Y^TY.$$

Before designing the desired robust variance-constrained filter, we present the following lemma.

Lemma 3: Consider the augmented system (8). If there exists a positive definite matrix $Q > 0$ such that the following matrix inequality

$$(\bar{A} + \Delta\bar{A})Q(\bar{A} + \Delta\bar{A})^T + \bar{B}\bar{B}^T - Q < 0 \tag{14}$$

holds, then we have the following conclusions: (1) the filtering matrix $\bar{A} + \Delta\bar{A}$ remains Schur stable; (2) the steady-state error covariance P exists and satisfies $P < Q_2$ where $Q_2 := [Q]_{22}$ and $[Q]_{22} \in \mathbb{R}^{n \times n}$ is the 22-subblock of Q.

Proof: The proof of this lemma is similar to that for the results in [21–23], and is thus omitted here. ∎

In the following theorem, a robust variance-constrained filter is designed to estimate the internal state variables of gene expression model (1)-(2), and a sufficient condition is derived to guarantee the solvability of the problem.

Theorem 1: Consider the system (8) and let $\sigma_i > 0$ $(i = 1, ..., n)$ be given scalars. If there exist two positive definite matrices $X > 0$ and $Q_2 > 0$, a matrix K and a scalar $\epsilon > 0$ such that the following linear matrix inequalities

$$\begin{bmatrix} -X & 0 & XB & XA & 0 & XM_1 & 0 \\ * & -Q_2 & B - KD & -KC & AQ_2 & M_1 - KM_2 & 0 \\ * & * & -I & 0 & 0 & 0 & 0 \\ * & * & * & -X & 0 & 0 & \epsilon N^T \\ * & * & * & * & -Q_2 & 0 & 0 \\ * & * & * & * & * & -\epsilon I & 0 \\ * & * & * & * & * & * & -\epsilon I \end{bmatrix} < 0, \tag{15}$$

$$[Q_2]_{ii} < \sigma_i^2, \quad (i = 1, ..., n), \tag{16}$$

hold, then with the obtained filter parameter K, the system (8) is Schur stable and the steady-state error covariance P satisfies $[P]_{ii} \leq \sigma_i^2$ $(i = 1, 2, ..., n)$.

Proof: We first discuss the Schur stability of the system (8). From Lemma 3, it suffices to show that (14) is true.

Set $X = Q_1^{-1}$. Pre- and post-multiplying (15) by $\text{diag}(Q_1, I, I, Q_1, I, I, \epsilon^{-1})$ and its transpose, we know that (15) is equivalent to

$$\begin{bmatrix} -Q_1 & 0 & B & AQ_1 & 0 & M_1 & 0 \\ * & -Q_2 & B - KD & -KCQ_1 & AQ_2 & M_1 - KM_2 & 0 \\ * & * & -I & 0 & 0 & 0 & 0 \\ * & * & * & -Q_1 & 0 & 0 & Q_1 N^T \\ * & * & * & * & -Q_2 & 0 & 0 \\ * & * & * & * & * & -\epsilon I & 0 \\ * & * & * & * & * & * & -\epsilon^{-1} I \end{bmatrix} < 0. \tag{17}$$

Letting $Q = (Q_1, \; Q_2)$ and noticing the definitions of $\bar{A}, \bar{B}, \bar{M}$ and $\bar{N}$ in (9), we can rewrite (17) as follows:

$$\begin{bmatrix} -Q & \bar{B} & \bar{A}Q & \bar{M} & 0 \\ * & -I & 0 & 0 & 0 \\ * & * & -Q & 0 & Q\bar{N}^T \\ * & * & * & -\epsilon I & 0 \\ * & * & * & * & -\epsilon^{-1} I \end{bmatrix} < 0. \tag{18}$$

By Lemma 1 (Schur Complement Lemma), (18) holds if and only if the following holds:

$$\begin{bmatrix} -Q + \epsilon^{-1} \bar{M}\bar{M}^T & \bar{B} & \bar{A}Q \\ * & -I & 0 \\ * & * & -Q + \epsilon Q\bar{N}^T \bar{N}Q \end{bmatrix} < 0, \tag{19}$$

which is, through being pre- and post-multiplied by $\text{diag}(I, I, Q^{-1})$, equivalent to

$$\Phi := \begin{bmatrix} -Q + \epsilon^{-1} \bar{M}\bar{M}^T & \bar{B} & \bar{A} \\ * & -I & 0 \\ * & * & -Q^{-1} + \epsilon \bar{N}^T \bar{N} \end{bmatrix} < 0. \tag{20}$$

So far, we have proved that $\Phi < 0$. By Lemma 2 and the definition of $\Delta \bar{A}$, we know that

$$
\begin{aligned}
\Phi &:= \begin{bmatrix} -Q & \bar{B} & \bar{A} \\ * & -I & 0 \\ * & * & -Q^{-1} \end{bmatrix} + \epsilon^{-1} \begin{bmatrix} \bar{M} \\ 0 \\ 0 \end{bmatrix} [\bar{M}^T \ 0 \ 0] + \epsilon \begin{bmatrix} 0 \\ 0 \\ \bar{N}^T \end{bmatrix} [0 \ 0 \ \bar{N}] \\
&\geq \begin{bmatrix} -Q & \bar{B} & \bar{A} \\ * & -I & 0 \\ * & * & -Q^{-1} \end{bmatrix} + \begin{bmatrix} \bar{M} \\ 0 \\ 0 \end{bmatrix} F(k) [0 \ 0 \ \bar{N}] + \begin{bmatrix} 0 \\ 0 \\ \bar{N}^T \end{bmatrix} F(k) [\bar{M}^T \ 0 \ 0] \\
&= \begin{bmatrix} -Q & \bar{B} & \bar{A} + \Delta \bar{A} \\ * & -I & 0 \\ * & * & -Q^{-1} \end{bmatrix} := \Omega.
\end{aligned}
\tag{21}
$$

Obviously, we have $\Omega < 0$, and it follows again from Lemma 1 that $\Omega < 0$ if and only if

$$
-Q + \begin{bmatrix} \bar{B} & \bar{A} + \Delta \bar{A} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \bar{B}^T \\ (\bar{A} + \Delta \bar{A})^T \end{bmatrix} < 0,
\tag{22}
$$

which is exactly (14). To this end, we can conclude that (14) is true and, from Lemma 3, the system (8) is guaranteed to be Schur stable. The second conclusion $[P]_{ii} \leq \sigma_i^2$ follows from (16) and the fact $P < Q$ immediately. This completes the proof of this theorem. ∎

Remark 3: The robust variance-constrained filter design problem is solved in Theorem 1 for the addressed linear uncertain discrete-time stochastic system. An LMI-based sufficient condition is derived for the existence of the desired filters, which ensures the asymptotic stability of the resulting filtering process and also guarantee the error precision (variance) to an acceptable degree. The feasibility of the filter design problem can be readily checked by the solvability of a set of LMIs, which can be done by resorting to the Matlab LMI toolbox. It should be mentioned that, in the past decade, LMIs have gained much attention for their computational tractability and usefulness in many areas because the so-called interior point method (see [4]) has been proven to be numerically very efficient for solving the LMIs. In next section, an illustrative example will be provided to show the application potential of the proposed techniques in gene expression time series.

IV. ILLUSTRATIVE EXAMPLES

In [24], a linear state-space system has been used to model the real-time gene expression time series data obtained from *the publicly available microarray dataset CDC15*. The system matrix A has been identified by the EM algorithm, and the matrix C can be easily obtained by the formula (4) of [24]. As discussed in the introduction, we believe it is necessary to include modelling errors and external stochastic noises in the system model so as to reflect the reality in a more proper way.

In this section, to illustrate the usefulness of the theory developed in previous sections, we present a simulation example for the gene expression model (1)-(2), where the main parameters (A and C) are identified in [24] and the uncertainties as well as noise intensity satisfy appropriate assumptions here. For the biological significance of the system states (internal variables), we refer the readers to [24], where a thorough discussion can be found. Here, our attention is focused on the design of a robust variance-constrained filter for the model with stochastic disturbance and admissible bounded uncertainties.

The identified and assumed model data of (1)-(2) are as follows: the noise variance is $\alpha^2 = 0.25$ and the upper bounds for the error variances are, respectively, $\sigma_1^2 = 1.2$, $\sigma_2^2 = 1.17$, $\sigma_3^2 = 1.23$, $\sigma_4^2 = 1.29$, $\sigma_5^2 = 1.18$.

Furthermore, the main model parameters are

$$
A = \begin{bmatrix}
0.4378 & -1.0077 & 0.5009 & 0.1851 & -0.1189 \\
0.6649 & 0.5244 & 0.2475 & 0.1511 & -0.1356 \\
-0.0702 & 0.1734 & 0.6794 & -0.3092 & -0.5279 \\
-0.0699 & -0.0103 & 0.1786 & 0.6163 & -0.5190 \\
0.0161 & 0.0316 & -0.07 & 0.1358 & 0.6662
\end{bmatrix}, \quad B = \begin{bmatrix}
0.01 \\
0.02 \\
0.03 \\
0.01 \\
0
\end{bmatrix},
$$

$$
M_1 = [0.04 \ 0.05 \ 0.05 \ 0.06 \ 0.04]^T, \quad N = [0.01 \ 0 \ 0 \ 0.08 \ 0.06],
$$

$$
C = \begin{bmatrix}
0.1105 & 0.2463 & 0.6724 & 0.2766 & 0.6370 \\
0.6094 & 0.3535 & 0.8414 & 0.9701 & 0.5600 \\
0.5019 & 0.1967 & 0.7177 & 1.3966 & 0.4555 \\
0.2538 & 0.5801 & 0.0968 & 1.2166 & 0.1986 \\
0.1397 & 0.0011 & -0.8519 & -0.1041 & -0.4839 \\
0.2109 & 0.5382 & 0.0654 & 1.1246 & -0.1396 \\
0.2039 & -0.1590 & -0.2573 & 0.3896 & 0.0716 \\
0.5015 & 0.3796 & 0.8491 & 1.6404 & 1.0731 \\
0.1969 & -0.2090 & -0.5357 & -0.0418 & -0.3750 \\
0.1951 & 0.4183 & 0.3949 & 1.3279 & 0.0855 \\
0.0145 & 0.0207 & -0.0449 & 0.1616 & -0.2743 \\
0.1611 & 0.5136 & 0.2014 & 0.8432 & 0.6842 \\
-0.0365 & -0.0276 & -0.5165 & 0.1408 & -0.3739 \\
-0.1269 & 0.6979 & 0.6747 & 0.6350 & -1.1011 \\
0.0731 & -0.1794 & -0.6344 & -0.3917 & -0.5290 \\
-0.0353 & -0.0344 & -0.1350 & 0.1501 & 0.6090 \\
-0.2156 & -0.3857 & -0.2954 & -1.1740 & -0.1992 \\
-0.2784 & -0.1646 & 0.4030 & -0.5321 & 0.4313 \\
-0.5334 & -0.5649 & -0.7397 & -1.3212 & -0.4794 \\
-0.3875 & -0.8582 & -0.3311 & -1.2706 & -0.4117 \\
-0.2807 & -0.8651 & -0.5047 & -1.2452 & -0.4893 \\
-0.4073 & -0.2150 & 0.4732 & -0.8055 & 0.5869 \\
-0.3765 & 0.0697 & 0.3790 & -1.5309 & 0.5855 \\
-0.2794 & 0.1961 & 0.3137 & -1.5506 & 0.9656
\end{bmatrix}, \quad D = \begin{bmatrix}
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04 \\
0.01 \\
0.04
\end{bmatrix}, \quad M_2 = \begin{bmatrix}
0.02 \\
0 \\
0.01 \\
-0.02 \\
0.01 \\
0 \\
0.02 \\
0.01 \\
0.01 \\
-0.03 \\
0.01 \\
0.013 \\
0.013 \\
0.024 \\
0.013 \\
0.004 \\
0.021 \\
0.014 \\
0.021 \\
0.004 \\
0.015 \\
0.014 \\
0.016 \\
0.014
\end{bmatrix}.
$$

Using Matlab LMI control Toolbox to solve the LMIs (15) and (16), we obtain

$$
X = \begin{bmatrix}
6.3537 & 0.1728 & 1.3092 & 1.2738 & 2.7192 \\
0.1728 & 8.6230 & -3.6033 & -2.3856 & 1.2590 \\
1.3092 & -3.6033 & 9.3660 & -1.8143 & -1.6960 \\
1.2738 & -2.3856 & -1.8143 & 9.8159 & 2.1535 \\
2.7192 & 1.2590 & -1.6960 & 2.1535 & 16.5369
\end{bmatrix},
$$

$$Q_2 = \begin{bmatrix} 1.2520 & 0.1656 & -0.1562 & -0.1317 & 0.0453 \\ 0.1656 & 1.1988 & 0.0162 & 0.0457 & 0.0431 \\ -0.1562 & 0.0162 & 1.2131 & -0.0727 & 0.3322 \\ -0.1317 & 0.0457 & -0.0727 & 1.2567 & 0.1392 \\ 0.0453 & 0.0431 & 0.3322 & 0.1392 & 1.1029 \end{bmatrix},$$

$$K = \begin{bmatrix} 0.1960 & 0.2339 & 0.2190 & -0.0149 & 0.2109 \\ -0.0888 & -0.0886 & -0.0587 & 0.0299 & -0.1188 \\ 0.1474 & 0.1784 & 0.1706 & -0.0078 & 0.1553 \\ -0.1679 & -0.1872 & -0.1570 & 0.0302 & -0.1982 \\ 0.0658 & 0.0925 & 0.1061 & 0.0136 & 0.0522 \\ 0.0175 & 0.0375 & 0.0584 & 0.0208 & -0.0033 \\ 0.2128 & 0.2630 & 0.2588 & -0.0041 & 0.2170 \\ 0.1467 & 0.1858 & 0.1889 & 0.0031 & 0.1436 \\ 0.0489 & 0.0719 & 0.0863 & 0.0144 & 0.0345 \\ -0.2687 & -0.3095 & -0.2743 & 0.0352 & -0.3040 \\ 0.0893 & 0.1133 & 0.1156 & 0.0023 & 0.0870 \\ 0.1598 & 0.2032 & 0.2077 & 0.0046 & 0.1552 \\ 0.1487 & 0.1860 & 0.1863 & 0.0002 & 0.1484 \\ 0.1832 & 0.2366 & 0.2467 & 0.0101 & 0.1731 \\ 0.0570 & 0.0823 & 0.0970 & 0.0147 & 0.0423 \\ 0.0291 & 0.0493 & 0.0665 & 0.0172 & 0.0119 \\ 0.0922 & 0.1206 & 0.1278 & 0.0071 & 0.0851 \\ 0.0514 & 0.0749 & 0.0891 & 0.0142 & 0.0372 \\ 0.1450 & 0.1828 & 0.1848 & 0.0020 & 0.1430 \\ -0.1309 & -0.1325 & -0.0911 & 0.0414 & -0.1723 \\ 0.0279 & 0.0467 & 0.0623 & 0.0157 & 0.0122 \\ 0.0455 & 0.0658 & 0.0776 & 0.0118 & 0.0336 \\ 0.0326 & 0.0393 & 0.0374 & -0.0019 & 0.0345 \\ -0.0428 & -0.0378 & -0.0168 & 0.0210 & -0.0638 \end{bmatrix}^T.$$

Let the initial states $x_i(0) = 0.3$ $(i = 1, 2, \cdots, 5)$ and $F(k) = \sin(k)$. The simulation results (trajectories of the internal variables and the estimation errors) are displayed in Fig. 1-Fig. 10, which have confirmed our theoretical results. It is shown that the state estimate traces its real value very well in the presence of modelling errors and parameter uncertainties.

V. Conclusions

In this paper, a robust filtering problem has been investigated for a linear gene expression model with stochastic disturbance and bounded uncertainties. The stochastic perturbation is in the form of a scalar Gaussian white noise with constant variance. We have designed a robust filter such that, for the admissible norm-bounded uncertainties, the filtering error system is Schur stable and the error variances are less than the prespecified upper bounds. By using the LMI technique, sufficient conditions have been derived for ensuring the desired performance of the gene expression model, and the filter gain can then be obtained directly by solving a set of linear matrix inequalities (LMIs). A simulation example, in which the main model parameters
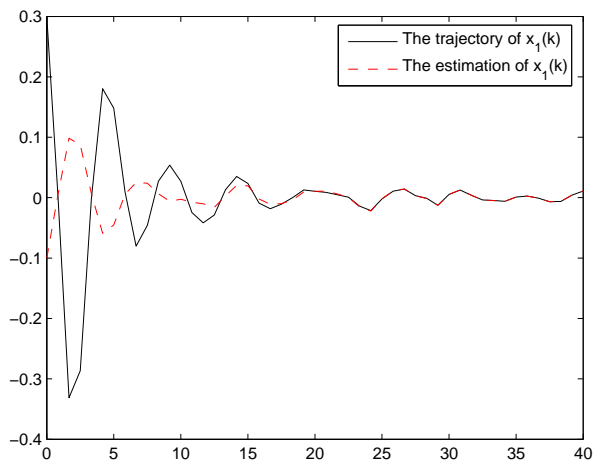
Fig. 1. The trajectory and estimation of $x_1(k)$
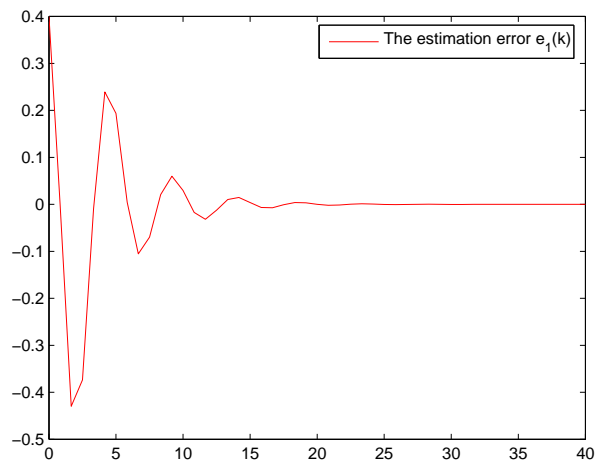

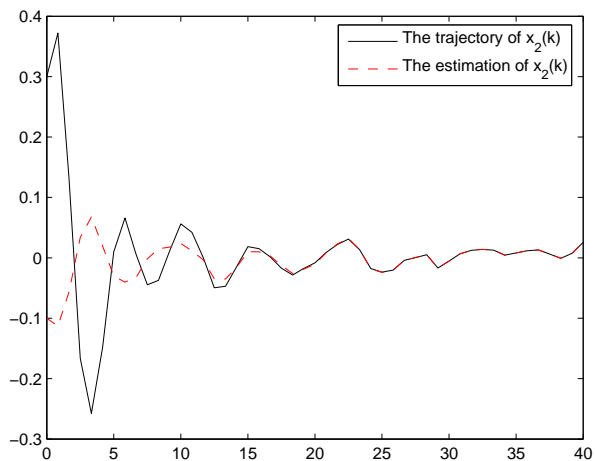Fig. 2. The estimation error $e_1(k)$

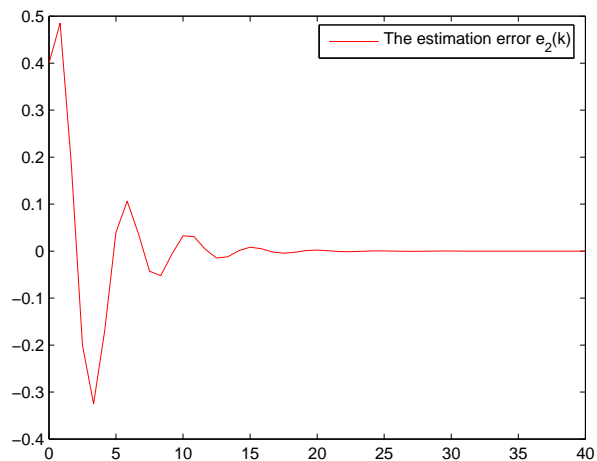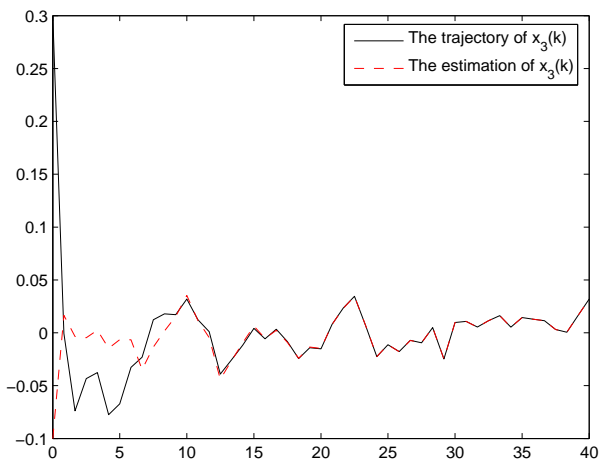Fig. 3. The trajectory and estimation of $x_2(k)$

Fig. 4. The estimation error $e_2(k)$
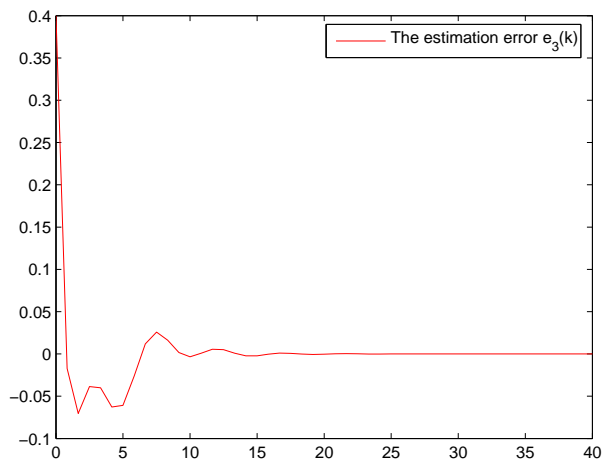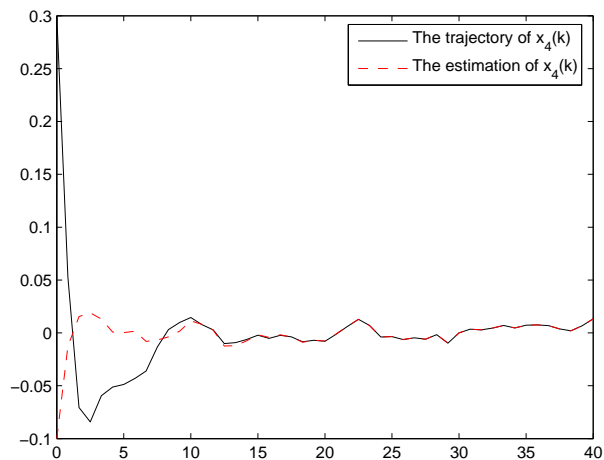

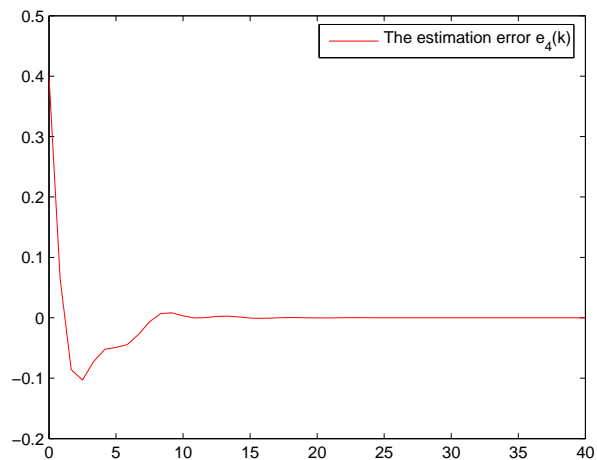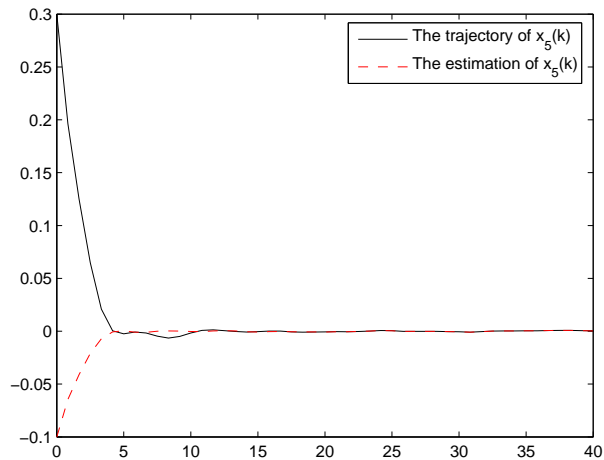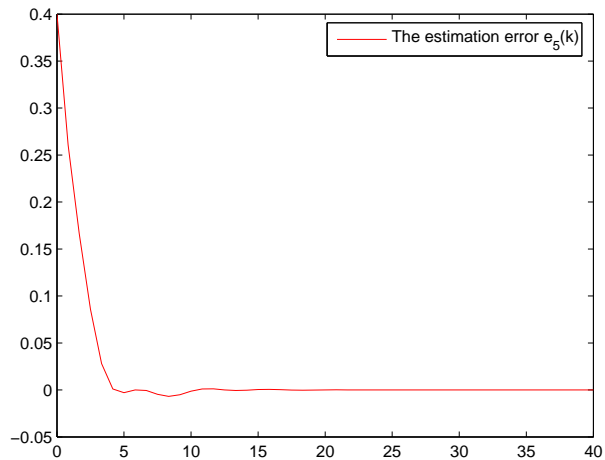Fig. 5. The trajectory and estimation of $x_3(k)$

Fig. 6. The estimation error $e_3(k)$

Fig. 7.   The trajectory and estimation of $x_4(k)$

Fig. 8.   The estimation error $e_4(k)$

Fig. 9.   The trajectory and estimation of $x_5(k)$

Fig. 10.   The estimation error $e_5(k)$

are obtained from [24], has been exploited to demonstrate the effectiveness of the proposed design procedures.

REFERENCES

[1]   T. Akutsu, S. Miyano, S. Kuhara, Identification of genetic networks from a small number of gene expression patterns under the boolean network model, *Proc. of Pacific Symposium on Biocomputing*, vol. 4, pp. 17-28, 1999.

[2]   A. Arkin, J. Ross and H. H. McAdams, Stochastic kinetic analysis of development pathway bifurcation in phage lmbdainfected Escherichia coli cells, *Genetics*, vol. 149, pp. 1633-1648.

[3]   M. Beal, E. Falciani, et al., A Bayesian approach to reconstructing genetic regulatory networks with hidden factor, *Bioinformatics*, October, 2004.

[4]   S. Boyd, L. EI Ghaoui, E. Feron and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. Philadephia: SIAM, 1994.

[5]   T. Chen, H. He, G. Church, Modeling gene expression with differential equations, *Proc. of Pacific Symposium on Biocomputing*, vol. 4, pp. 29-40, 1999.

[6]   D. Cook, A. Gerber, S. Tapscott, Modeling stochastic gene expression: implications for haploinsufficiency, *Proc. of the National Academy of Science*, USA, vol. 95, pp. 15641-15646, 1998.

[7] M. de Hoon, S. Imoto, et al., Infering gene regulatory networks from time-ordered gene expression data of bacillus using differential equations, *Proc. of Pacific Symposium on Biocomputing*, vol. 8, pp. 17-28, 2003.

[8] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, *J. Comput Biol.*, vol. 9, no. 1, pp. 67-103, 2002.

[9] P. D'haeseleer, S. Liang and R. Somogyi, Genetic network inference: from co-expression clustering to reverse engineering, *Bioinformatics*, vol. 16, no. 8, pp. 707-726, 2000.

[10] P. D'haeseleer, X. Wen, S. Fuhrman and R. Somogyi, Linear modeling of mRNA expression levels during CNS development and injury, *Proc. of Pacific Symposium on Biocomputing*, vol. 4, pp. 41-52, 1999.

[11] M. Eisen, P. Spellman, P. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, *Proc. of the National Academy of Science*, USA, vol. 95, pp. 14863-14868, 1998.

[12] S. Huang, Gene expression profilling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery, *Journal of Molecular Medicine*, vol. 77, pp. 469-480, 1999.

[13] P. Kellam, X. Liu, et al., A framework for modeling virus gene expression data, *Intelligent Data Analysis*, vol. 6, pp. 265-279, 2002.

[14] K. Murphy, S. Mian, Modeling gene expression data using dynamic Bayesian networks, *Technical Report*, University of California: Berkeley, 1999.

[15] W. Nanacara, E. Yaz, linear and nonlinear estimation with uncertain observtions,*In Proc. Amer. Contr. Conf.*, Maryland, 1994.

[16] A. Pease, et al., Light-generated oligonucleotide arrays for rapid DNA sequense analysis, *Proc. Natl. Acad. Sci. USA*, vol. 91, pp. 5022-5026, 1994.

[17] C. Rangel, J. Angus, et al., Modeling T-cell activation using gene expression profilling and state space models, *Bioinformatics*, Vol. 20, No. 9, pp. 1361-1372, 2004.

[18] P. Smolen, D. Baxter and J. Byrne, Mathematical modeling of genetics networks review, *Neuron*, vol. 26, No. 3, pp. 567-580, 2000.

[19] P. Tamayo, et al., Interpreting patterns of gene expression with self-organizing maps:methods and application to hematopoietic differentiation, *Proc. of the National Academy of Science*, USA, vol. 96, pp. 2907-2912, 1999.

[20] T. Tian, K. Burrage, stochastic neural network models for gene regulatory networks, *Proc. IEEE Congress on Evolutionary Computation*, pp. 162-169, 2003.

[21] Z. Wang, Robust state estimation for perturbed systems with error variance and circuits pole constraints: the Continuous-Time Case, *Int. J. Control*, vol. 73, no. 4, pp. 303-311, 2000.

[22] Z. Wang and X. Liu, On designing H_∞ filters with circular pole and error variance constraints, *IEEE Transactions Circuits and systems-II: Analog Digital Signal Processing*, vol. 50, no. 3, pp. 139-143, 2003.

[23] Z. Wang, H. Unbehauen, Robust H_2/H_∞ state estimation for systems with error variance constraints: the continuous-time case, *IEEE Transactions on Automatic Control*, vol. 44, no. 5, pp. 1061-1065, 1999.

[24] F. Wu, W. Zhang, A. Kusalik, Modeling gene expression from microarray expression data with state-space equations, *Proc. Pacific Symposium on Biocomputing*, vol. 9, pp. 581-592, 2004.

[25] E. Yaz, R. Skelton, Continuous and discrete state estimation with error variance assignment, *In Proc. IEEE Conf. Decision Contr.*, Brighton, U.K., 1991, pp. 3091-3092.

[26] E. Yaz, W. Nanacara, Nonlinear estimation by covariance assignment, *In Proc. 12th IFAC World Congr.*, Sydney, Australia, vol. 6, pp. 87-90, 1993.