Research on Dual Control[1)]

Duan Li[1] Fucai Qian[2] Peilin Fu[3]

[1](*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, P.R.China*)

[2](*School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi,* 710049 *P.R.China*)

[3](*Department of Electrical Engineering, University of California, Riverside, California, U.S.A*)

(E-mail: dli@se.cuhk.edu.hk)

Abstract This paper summarizes recent progress by the authors in developing two solution frameworks for dual control. The first solution framework considers a class of dual control problems where there exists a parameter uncertainty in the observation equation of the LQG problem. An analytical active dual control law is derived by a variance minimization approach. The issue of how to determine an optimal degree of active learning is then addressed, thus achieving an optimality for this class of dual control problems. The second solution framework considers a general class of discrete-time LQG problems with unknown parameters in both state and observation equations. The best possible (partial) closed-loop feedback control law is derived by exploring the future nominal posterior probabilities, thus taking into account the effect of future learning when constructing the optimal nominal dual control.

Key words Dual control, dynamic programming, LQG control, stochastic control

1 Introduction

Feldbaum, in his pioneer work[1] about 40 years ago, pointed out that, when implementing optimal control for stochastic systems with parameter uncertainty, except for a few ideal situations, the controller usually pursues two often conflicting goals: (i) to drive the system towards a desired state, and (ii) to perform learning to reduce the systems uncertainty. Function (i) is usually featured by *caution* because when providing the control function, the controller should not use the estimated parameters blindly as if they were true, therefore avoiding an enlargement of the impact of the existing uncertainties on the cost. Function (ii) is usually accompanied by active *probing* , because in multistage control problems where observations are made on the system at each stage the controller might be able to carry out active information gathering by using the observations obtained to enhance estimation for unknown system parameters. Such a devised control, which not only affects the state of system but also affects the quality of estimation, is the so-called dual control.

The dual roles of dual control, optimization and estimation, in general situations, cannot be separated. Thus the best control must have a characteristic of appropriately distributing its energy between functions (i) and (ii). However, the complicated coupling between optimization and estimation results in, in most situations, an unattainability of an analytical law of dual control.

Although considerable progress has been made in the past 40 years in both theory and practice of dual control. Much more efforts are needed to better our understanding of dual control and to achieve satisfactory solution methods. In 2000, *IEEE Control Systems Society* listed the dual control as one of the 25 most prominent subjects in the last century which had significantly impacted the development of control theory.

Prominent features and fundamental properties of dual control have been extensively studied in the literature[2~4] etc. Previous efforts in dual control have mainly been devoting to developing certain suboptimal solution schemes, such as certainty equivalence scheme and open-loop feedback control, by bypassing the essential feature of coupling in dual control. Most resulting suboptimal control laws are of a nature of passive learning, due to that the function of future active probing of the control is purposely deprived in order to achieve an analytical attainability in the solution process. An analysis of various approximations in dual control is given in [5]. Surveys on dual control can be found in [6] and [7].

1) Supported by the Research Grants Council of Hong Kong, P.R.China under Grant CUHK 4180/03E
Received November 29, 2003; in revised form October 27, 2004

To power a control law with a property of active learning, which is indispensable in reducing reducible system's uncertainty, is a key to developing an optimal control law in dual control. Solution schemes[8~11] have been proposed in the literature to attach certain variance terms of the state or the innovation process into the objective function or use it as a constraint in order to force the control to perform active learning. These solution schemes, however, truncate the time horizon into shorter time periods of one stage, prompting a concern of possible myopic behaviors. The authors of this paper have recently proposed some novel solution methods to derive dual control laws with an active learning property and have solved the optimal dual control problem for LQG problems with unknown parameters only in the observation equation[12~15], thus advancing the state-of-the-art of dual control.

In Section 2, the dual control problem is presented and the related technical difficulties in deriving its optimal solution are discussed, serving as both the background and a motivation for the development in the later sections. In Section 3, we illustrate how an analytical optimal dual control law can be derived by a variance minimization approach for LQG problems with unknown parameters in the observation equation. In Section 4, we demonstrate that the optimal nominal dual control is the best possible (partial) closed-loop control law for general LQG problems with uncertain parameters in both state and observation equations. The paper concludes in Section 5 with a summary of the developed two solution frameworks.

2 Optimal dual control problem

We consider in this paper the following stochastic optimal control problem where parameter uncertainties exist in both the state equation and the observation equation,

$$
\begin{aligned}
&\text{(G)}\ \min \mathrm{E}\left\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N)+\sum_{k=0}^{N-1}[\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k)+\boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)]\mid I^0\right\}\\
&\quad\text{s.t. } \boldsymbol{x}(k+1)=A(k,\theta)\boldsymbol{x}(k)+B(k,\theta)\boldsymbol{u}(k)+\boldsymbol{w}(k),\ k=0,1,\cdots,N-1\\
&\qquad\ \boldsymbol{y}(k)=C(k,\theta)\boldsymbol{x}(k)+\boldsymbol{v}(k),\ k=1,2,\cdots,N
\end{aligned}
$$

where $\boldsymbol{x}(k)\in R^n$ is the state vector, $\boldsymbol{u}(k)\in R^m$ is the control vector, $\boldsymbol{y}(k)\in R^p$ is the measured output vector, and I^0 is the initial information set that includes information about the probability distribution of the initial state $\boldsymbol{x}(0)$, the statistics of the random sequences $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$, and the prior probability distribution of the unknown parameter θ. $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$ are two independent Gaussian white noises with means zero and variances σ_w^2 and σ_v^2, respectively. The initial state $\boldsymbol{x}(0)$ is assumed to be Gaussian which is independent of the process and observation noises: $\boldsymbol{x}(0)\sim N(\hat{\boldsymbol{x}}(0),P(0))$. Matrices $A(k,\theta), B(k,\theta)$ and $C(\theta,k)$ are of appropriate dimensions and depend on an unknown parameter θ. It is assumed that θ belongs to a finite set $\Theta=\{\theta_1,\theta_2,\dots,\theta_s\}$ and is a constant over the whole control horizon. The a-priori probabilities of parameter θ are given as $q_i(0)=P(\theta=\theta_i\mid I^0)$. Further, $\{Q(k)\}$ and $\{R(k)\}$ are sequences of positive semidefinite and positive definite symmetric matrices of appropriate dimensions, respectively. Define the information set at stage k to be I^k, $I^k=\{\boldsymbol{u}(0),\dots,\boldsymbol{u}(k-1),\boldsymbol{y}(1),\dots,\boldsymbol{y}(k),I^0\},\ k=0,1,\dots,N.$

The dual control problem for (G) is to find a closed-loop control, $\boldsymbol{u}(k)=\boldsymbol{f}_k(I^k), k=0,1,\dots,N-1,$ such that the performance index in (G) is minimized.

There exist two types of uncertainties in problem (G): i) the white noises from outside of the system, $\{\boldsymbol{w}(k)\}$ and $\{\boldsymbol{v}(k)\}$, by which the state variables and measurement variables are distorted; and ii) the unknown system parameters, θ, whose uncertainty could be caused by a changing environment or some deteriorating conditions, for example, components breakdown during system operating. The first type of uncertainty is called *irreducible* uncertainty. This kind of uncertainties can not be removed or reduced, while Kalman filter or other estimation schemes can be used to trace out the true systems state. The second type of uncertainty caused by an unknown mode of the parameter θ is called *reducible* uncertainty. This type of uncertainty can be reduced, or can be even completely removed by learning. Problem (G) differs from the classical LQG where only irreducible uncertainty exists. In dual control, a fundamental question is how much efforts of active learning should be placed to reduce the reducible uncertainty in order to achieve the optimality of the overall objective for problem (G).

When assuming $\theta=\theta_i$, the state estimate at stage k, $\hat{\boldsymbol{x}}_i(k|k)$, can be obtained by the Kalman filter:

$$\hat{\boldsymbol{x}}_i(k|k) = \hat{\boldsymbol{x}}_i(k|k-1) + F_i(k)[\boldsymbol{y}(k) - C(k,\theta_i)\hat{\boldsymbol{x}}_i(k|k-1)] \tag{1}$$

$$\hat{\boldsymbol{x}}_i(k|k-1) = A(k-1,\theta_i)\hat{\boldsymbol{x}}_i(k-1|k-1) + B(k-1,\theta_i)\boldsymbol{u}(k-1) \tag{2}$$

$$F_i(k) = P_i(k|k-1)C'(k,\theta_i)[C(k,\theta_i)P_i(k|k-1)C'(k,\theta_i) + \sigma_v^2]^{-1} \tag{3}$$

$$P_i(k|k-1) = A(k-1,\theta_i)P_i(k-1|k-1)A'(k-1,\theta_i) + \sigma_w^2 \tag{4}$$

$$P_i(k|k) = [I - F_i(k)C(k,\theta_i)]P_i(k|k-1) \tag{5}$$

with initial condition of $\hat{\boldsymbol{x}}_i(0|0) = \hat{\boldsymbol{x}}(0)$ and $P_i(0|0) = P(0)$.

Based on the observations, the posterior probability of model i at stage k, $q_i(k)$, can be calculated recursively as follows[16]

$$q_i(k) = \frac{L_i(k)}{\sum_{j=1}^{s} q_j(k-1)L_j(k)} q_i(k-1), \quad k = 1,2,\dots,N$$

with initial condition $q_i(0)$, where

$$\begin{aligned} &L_i(k) = |P_y(k|k-1,\theta_i)|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\tilde{\boldsymbol{y}}(k|k-1,\theta_i)' P_y(k|k-1,\theta_i)^{-1} \tilde{\boldsymbol{y}}(k|k-1,\theta_i)] \\ &\tilde{\boldsymbol{y}}(k|k-1,\theta_i) = \boldsymbol{y}(k) - C(k,\theta_i)\hat{\boldsymbol{x}}_i(k|k-1), \quad P_y(k|k-1,\theta_i) = C(k,\theta_i)P_i(k|k-1)C'(k,\theta_i) + \sigma_v^2 \end{aligned} \tag{6}$$

Define for $i = 1,2,\dots,s$ and $k = 1,2,\dots,N-1$,

$$J_i(k,I^k) = \mathrm{E}\{\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)|\theta_i, I^k\}, \quad J_i(N,I^N) = \mathrm{E}\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N)|\theta_i, I^N\}$$

Then the following is obvious,

$$J(k,I^k) = \mathrm{E}\{\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k) \mid I^k\Big\} = \sum_{i=1}^{s} q_i(k)J_i(k,I^k), \ k = 0,1,\dots,N-1$$

$$J(N,I^N) = \mathrm{E}\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N) \mid I^N\} = \sum_{i=1}^{s} q_i(N)J_i(N,I^N)$$

From the principle of dynamic programming, the closed-loop control that minimizes the performance index in problem (G) can be obtained by solving the following,

$$\begin{aligned} J(I^0) = &\min_{\boldsymbol{u}(0)} \mathrm{E}\Big\{ \sum_{i=1}^{s} q_i(0)J_i(0,I^0) + \min_{\boldsymbol{u}(1)} \mathrm{E}\Big\{ \sum_{i=1}^{s} q_i(1)J_i(1,I^1) + \ldots + \min_{\boldsymbol{u}(k)} \mathrm{E}\Big\{ \sum_{i=1}^{s} q_i(k)J_i(k,I^k) + \ldots + \\ &\min_{\boldsymbol{u}(N-1)} \mathrm{E}[\sum_{i=1}^{s} q_i(N-1)J_i(N-1,I^{N-1}) + \sum_{i=1}^{s} q_i(N)J_i(N,I^N)|I^{N-1}] \ldots |I^k\Big\} \ldots |I'\Big\}|I^0\Big\} \end{aligned} \tag{7}$$

Rewriting (7) in a backward recursive form yields the Bellman equation

$$J^*(k,I^k) = \min_{\boldsymbol{u}(k)} \mathrm{E}\{\boldsymbol{x}^{\mathrm{T}}(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}^{\mathrm{T}}(k)R(k)\boldsymbol{u}(k) + J^*(k+1,I^{k+1})|I^k\} \tag{8}$$

where $J^*(k,I^k)$ is the optimal cost-to-go from time k to the end. The terminal condition is $J^*(N,I^N) = \mathrm{E}\{\boldsymbol{x}^{\mathrm{T}}(N)Q(N)\boldsymbol{x}(N)|I^N\}$.

In principle, the dual control problem (G) can be solved using (7) or (8). However, the difficulty and complexity in solving (G) only begin with these equations. In fact, at stage k all the posterior probabilities in later stages are unknown. Therefore, to derive the cost-to-go functions in the stochastic dynamic programming form (8) is a formidable problem. Unless all the posterior probabilities in later stages can be found, we are not able to solve these functional equations. The curse of uncertainty of the posterior probabilities in later stages is further compounded by the required expectation operators. One probably can choose to reasonably approximate the posterior probabilities in later stages or cost-to-go functions. For example, if future learning is not considered, we can fix all the posterior probabilities in later stages at their current values and the open-loop optimal feedback control can be obtained by solving the following problem,

$$\min_{\boldsymbol{u}(k)} \sum_{i=1}^{s} q_i(k)\Big\{ \mathrm{E}\Big\{ J_i(k, I^k) + \ldots + \min_{\boldsymbol{u}(N-2)} \mathrm{E}\{J_i(N-2, I^{N-2}) + \\ \min_{\boldsymbol{u}(N-1)} \mathrm{E}[J_i(N-1, I^{N-1}) + J_i(N, I^N) | I^{N-1}] | I^{N-2}\} \ldots | I^k \Big\}\Big\}$$

The open-loop feedback control is a passive scheme that does not possess an active learning feature and does not consider any impact from the future learning, and thus can never be optimal.

3 Variance minimization method

In this section, we discuss a special case of problem (G) and achieve an optimal dual control law. Consider the following optimal control problem,

$$(\mathrm{P}) \ \min \mathrm{E}\Big\{ \boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N) + \sum_{k=0}^{N-1} [\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)] \mid I^0 \Big\}$$

$$\text{s.t. } \boldsymbol{x}(k+1) = A(k)\boldsymbol{x}(k) + B(k)\boldsymbol{u}(k) + \boldsymbol{w}(k), \ k = 0, 1, \cdots, N-1 \tag{9}$$

$$\boldsymbol{y}(k) = C(k, \theta)\boldsymbol{x}(k) + \boldsymbol{v}(k), \ k = 1, 2, \cdots, N \tag{10}$$

where $A(k)$ and $B(k)$ are known and parameter uncertainty only exists in $C(k, \theta)$.

Optimal solution to problem (P) was not identified for many years due to an analytical complexity in the cost-to-go function when applying dynamic programming. Essentially, this analytical complexity is caused by certain nonlinear terms of the state estimation which are nonseparable in the sense of dynamic programming. A passive learning algorithm (DUL) was proposed[17] by Deshpande *et al.*

3.1 A passive learning algorithm

Using the formula of the iterated expectation, minimizing the expected performance index in (P) is equivalent to

$$\min \mathrm{E}_\theta \Big\{ \mathrm{E}\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N) + \sum_{k=0}^{N-1} [\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)] \mid \theta, I^0\} \mid I^0 \Big\}$$

DUL adopts an approximation by interchanging the minimization and the first expectation,

$$\mathrm{E}_\theta \Big\{ \min \mathrm{E}\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N) + \sum_{k=0}^{N-1} [\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)] \mid \theta, I^0\} \mid I^0 \Big\}$$

which leads to an analytical form of a suboptimal control law for (P):

$$\boldsymbol{u}(k) = -\varGamma(k)\hat{\boldsymbol{x}}(k), \quad k = 0, 1, \ldots, N-1$$

where $\varGamma(k) = D(k)B'(k)S(k+1)A(k)$, $D(k) = [B'(k)S(k+1)B(k) + R(k)]^{-1}$, $S(k) = A'(k)S(k+1)A(k) + Q(k) - A'(k)S(k+1)B(k)D(k)B'(k)S(k+1)A(k)$, $k = N-1, \ldots, 1$, with the boundary condition $S(N) = Q(N)$. Assigning the current posterior probabilities as the weighting coefficients, the control in the DUL Algorithm is a weighting sum of the optimal control for s different models. It is a passive learning algorithm because the control does not probe.

Casiello and Loparo[16] later proved that this passive control law in DUL is optimal to the following modified problem of (P),

$$(\mathrm{M}_p) \ \min \mathrm{E}(\hat{J}) = \mathrm{E}\{\boldsymbol{x}'(N)Q(N)\boldsymbol{x}(N) + \sum_{k=0}^{N-1} [\boldsymbol{x}'(k)Q(k)\boldsymbol{x}(k) + \boldsymbol{u}'(k)R(k)\boldsymbol{u}(k)] - \\ \sum_{k=0}^{N-1} [\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)]'T(k)[\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)] \mid I^0\}$$

$$\text{s.t. Eq.(9) } \textit{and} \text{ Eq.(10)}$$

where $T(k) = A'(k)S(k+1)B(k)D(k)B'(k)S(k+1)A(k),\ k = N-1, \ldots, 1, 0$.

The only difference between problems (P) and (M_p) is that a summation of weighted conditional covariances is subtracted in (M_p) from the original performance index in (P). As pointed by Casiello and Loparo[16], subtracting the conditional covariance weighted by the sequence $\{T(k)\}$ balances out the function of active learning such that the passive learning controller becomes optimal in this context. Active learning is, however, indispensable for achieving an optimality in the primal problem (P) with a parameter uncertainty.

Notice that the smoothing property does not hold for $J_k(\hat{\boldsymbol{x}}(k))$ when function J_k is nonlinear, *i.e.*, $\mathrm{E}\{J_k(\mathrm{E}[\boldsymbol{x}(k+1) \mid I^{k+1}]) \mid I^k\} \neq J_k(\mathrm{E}[\boldsymbol{x}(k+1) \mid I^k])$. In other words, a nonlinear term of the state estimation is nonseparable in the sense of dynamic programming. Investigating further on (M_p), we can also conclude that subtracting the conditional covariance weighted by the sequence $\{T(k)\}$ removes these nonlinear terms involving the conditional mean of the state in the recursive equation of dynamic programming, thus making the problem tractable.

3.2 An active-learning control algorithm

The success degree of an active learning can be measured by the variance of the final state. Therefore minimizing a variance term of the final state will add a feature of active learning to the derived control law. In this subsection, we consider a modified problem of (M_p) in which a variance term at the final stage is attached to the objective function of (M_p),

$$\begin{aligned}(\mathrm{M}_a(\mu))\ \min\ &\{\mathrm{E}(\hat{J}) + \mu \mathrm{Tr}[Cov(\boldsymbol{x}(N) \mid I^0)]\} = \\ &\min\{\mathrm{E}(\hat{J}) + \mu \mathrm{E}([\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^0)]' \times [\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^0)] \mid I^0)\} \\ &\text{s.t. Eqs.(9) } \textit{and } (10)\end{aligned}$$

where $T(k) = A'(k)S(k+1)B(k)D(k)B'(k)S(k+1)A(k)$, $S(k) = A'(k)S(k+1)A(k) + Q(k) - T(k)$, $D(k) = [B'(k)S(k+1)B(k) + R(k)]^{-1}$, $S(N) = Q(N) + \mu I$. Parameter $\mu \in [0, \infty)$ is a weighting coefficient of active learning and Tr denotes the trace of a square matrix. A larger μ implies that more importance has been placed on probing (active learning and uncertainty reduction).

Problem ($M_a(\mu)$) is difficult to be solved directly, since the recursive equation of dynamic programming involves certain nonlinear terms of the state estimation (conditional mean) which introduces a nonseparability in the sense of dynamic programming. In order to overcome this difficulty, a solution scheme similar to Li and Ng[18] is adopted to embed problem ($M_a(\mu)$) into a tractable auxiliary problem which is separable in the sense of dynamic programming. Solving the auxiliary problem and investigating the relationship between the solution sets of problem ($M_a(\mu)$) and the auxiliary problem, the optimal control of problem ($M_a(\mu)$) can be identified.

The performance index of ($M_a(\mu)$) can be written as $J(J_1, J_2) = J_1 - \mu J_2' J_2$, where $J_1 = \mathrm{E}\{\hat{J} + \mu \boldsymbol{x}^{\mathrm{T}}(N)Q(N)\boldsymbol{x}(N)|I^0\}$ and $J_2 = \mathrm{E}(\boldsymbol{x}(N) \mid I^0)$. It is easy to see that the performance index in ($M_a(\mu)$), J, is a concave function of J_1 and J_2.

The following auxiliary problem is now constructed for problem ($M_a(\mu)$) with a fixed multiplier vector $\boldsymbol{\lambda} \in R^n$,

$$\begin{aligned}(\mathrm{A}(\boldsymbol{\lambda}, \mu))\ \min\ & J_1 - 2\boldsymbol{\lambda}' J_2 \\ & \text{s.t. Eq.(9) } \textit{and } \text{Eq.(10)}\end{aligned}$$

Theorem 1. Suppose that $\{\boldsymbol{u}^*(k)\}$ is an optimal control of problem ($M_a(\mu)$), then $\{\boldsymbol{u}^*(k)\}$ is also an optimal control of the auxiliary problem ($\mathrm{A}(\boldsymbol{\lambda}^*, \mu)$) where $\boldsymbol{\lambda}^*$ satisfies

$$\boldsymbol{\lambda}^* = -\frac{1}{2}\frac{\partial J(J_1, J_2)}{\partial J_2} \mid_{\{\boldsymbol{u}^*(k)\}} = \mu J_2 \mid_{\{\boldsymbol{u}*(k)\}}$$

The implication of Theorem 1 is that any optimal solution to problem ($M_a(\mu)$) is in the set of solutions to auxiliary problem ($\mathrm{A}(\boldsymbol{\lambda}, \mu)$). Note that the auxiliary problem is strictly convex with respect to $\{\boldsymbol{u}(k)\}$. Thus the solution to ($\mathrm{A}(\boldsymbol{\lambda}, \mu)$) is unique for a given $\boldsymbol{\lambda}$ and can be obtained by using dynamic programming.

Theorem 2. For a given $\boldsymbol{\lambda}$, the optimal control of the auxiliary problem ($\mathrm{A}(\boldsymbol{\lambda}, \mu)$) is

$$\boldsymbol{u}^*(k) = -\mathit{\Gamma}_1(k)\hat{\boldsymbol{x}}(k) + \mathit{\Gamma}_2(k)\boldsymbol{\lambda}, \quad k = 0, 1, \ldots, N-1 \tag{11}$$

where for $k = N-1, N-2, \ldots, 1, 0$,

$$\varGamma_1(k) = D(k)B'(k)S(k+1)A(k)$$
$$\varGamma_2(k) = D(k)B'(k)L'(k+1)$$
$$L(k) = L(k+1)[A(k) - B(k)\varGamma_1(k)]$$

with the boundary conditions $S(N) = Q(N) + \mu I$ and $L(N) = I$.

Define

$$\varPhi(k) = I - \mu\left\{\sum_{s=k+1}^{N-1}\prod_{i=s}^{N-1}[A(i) - B(i)\varGamma_1(i)]B(s-1)\varGamma_2(s-1) + B(N-1)\varGamma_2(N-1)\right\}$$

$$\varPsi(k) = \mu\prod_{i=k}^{N-1}[A(i) - B(i)\varGamma_1(i)]$$

Theorem 3. Assume that $\varPhi(0)$ is invertible. Then the optimal $\boldsymbol{\lambda}^*$ with which the optimal solution to $(\mathrm{A}(\boldsymbol{\lambda}^*, \mu))$ solves $(\mathrm{M}_a(\mu))$ is equal to $\boldsymbol{\lambda}^* = \varPhi^{-1}(0)\varPsi(0)\hat{\boldsymbol{x}}(0)$.

The optimal control of $(\mathrm{M}_a(\mu))$ can be then obtained by substituting the optimal $\boldsymbol{\lambda}$ into the optimal control law of the auxiliary problem $(\mathrm{A}(\boldsymbol{\lambda}, \mu))$,

$$\boldsymbol{u}^*(k) = -\varGamma_1(k)\hat{\boldsymbol{x}}(k) + \varGamma_2(k)\varPhi^{-1}(0)\varPsi(0)\hat{\boldsymbol{x}}(0).$$

The above optimal control law depends on both estimations of the initial state and the current state and thus is not of a full-feedback nature which is desirable in stochastic control. The optimal control law of $(\mathrm{M}_a(\mu))$ can be further improved as follows.

As proceeding to stage k, we can view stage k as the initial stage and $\hat{\boldsymbol{x}}(k)$ as the estimation of the initial state when we consider a truncated dual control problem from stage k to stage N. Based on the principle of optimality and the concept of a rolling horizon, using the same derivation scheme as in Theorem 3, the optimal value of $\boldsymbol{\lambda}$ should be $\boldsymbol{\lambda}^* = \varPhi(k)^{-1}\varPsi(k)\hat{\boldsymbol{x}}(k)$. Substituting the optimal $\boldsymbol{\lambda}^*$ into (**??**), we get the optimal full-feedback control law $\boldsymbol{u}(k) = -\varGamma(k)\hat{\boldsymbol{x}}(k)$, where $\varGamma(k) = \varGamma_1(k) - \varGamma_2(k)\varPhi(k)^{-1}\varPsi(k)$.

3.3 Optimal degree of active learning

An optimal full-feedback control is derived in the last subsection for problem $(\mathrm{M}_a(\mu))$ with a fixed value of μ. Recall that μ represents a degree of active learning in dual control. This subsection will address the issue of how to determine an optimal degree of active learning, which leads to an algorithm to determine the weighting coefficient μ in $(\mathrm{M}_a(\mu))$.

Our ultimate goal is to derive an optimal control law for the primal problem (P). It is evident that the solution to problem $(\mathrm{M}_a(\mu))$ will become an optimal solution to problem (P) if we select a parameter μ such that the last two terms in the performance index of $(\mathrm{M}_a(\mu))$ cancel out each other on average, *i.e.*,

$$\mu\mathrm{E}\{[\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^0)]'[\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^0)] \mid I^0\} = \mathrm{E}\left\{\sum_{k=0}^{N-1}[\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)]'T(k)[\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)] \mid I^0\right\}$$

and this cancellation should be independent of control $\boldsymbol{u}$.

Let us consider more general situations. Different time stages may require different degree of active learning. Assume that the current stage is l. The value of μ at stage l should be adjusted based on the information set I^l such that

$$\mu\mathrm{E}\{[\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^l)]'[\boldsymbol{x}(N) - \mathrm{E}(\boldsymbol{x}(N) \mid I^l)] \mid I^l\} = \mathrm{E}\left\{\sum_{k=l}^{N-1}[\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)]'T(k)[\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)] \mid I^l\right\} \tag{12}$$

The right-hand side of (12) is hard to evaluate, since it involves expected values of squares of the state estimation. We consider in the following an approximation of (12) by replacing the state estimation $\hat{\boldsymbol{x}}(k)$ by a forecasted mean of $\boldsymbol{x}(k)$ at stage l,

$$\begin{aligned}&\mu\mathrm{E}\{[\boldsymbol{x}(N)-\mathrm{E}(\boldsymbol{x}(N)\mid I^l)]'[\boldsymbol{x}(N)-\mathrm{E}(\boldsymbol{x}(N)\mid I^l)]\mid I^l\}=\\&\mathrm{E}\left\{\sum_{k=l}^{N-1}[\boldsymbol{x}(k)-\mathrm{E}(\boldsymbol{x}(k)\mid I^l)]'T(k)[\boldsymbol{x}(k)-\mathrm{E}(\boldsymbol{x}(k)\mid I^l)]\mid I^l\right\}\end{aligned}\tag{13}$$

Equation (13) can be written in the following form,

$$\mu\mathrm{Tr}[Cov(\boldsymbol{x}(N)|I^l)]=\sum_{k=l}^{N-1}\mathrm{Tr}[T(k)Cov(\boldsymbol{x}(k)|I^l)]\tag{14}$$

Assume that no new information will be collected from the current time l to time N. At time l all the information available to the controller is contained in I^l. Define

$$\begin{aligned}&\hat{\boldsymbol{x}}_i(k\mid l)=\mathrm{E}\{\boldsymbol{x}(k)|I^l,\theta=\theta_i\},\ l\leqslant k\leqslant N\\&P_i(k\mid l)=\mathrm{E}\{[\boldsymbol{x}(k)-\hat{\boldsymbol{x}}(k\mid l)][\boldsymbol{x}(k)-\hat{\boldsymbol{x}}(k\mid l)]'|I^l,\theta=\theta_i\},\ l\leqslant k\leqslant N\end{aligned}$$

Then for $l\leqslant k\leqslant N$, we have

$$\begin{aligned}&\hat{\boldsymbol{x}}_i(k\mid l)=A(k-1)\hat{\boldsymbol{x}}_i(k-1\mid l)+B(k-1)\boldsymbol{u}(k-1)\\&P_i(k\mid l)=A(k-1)P_i(k-1\mid l)A'(k-1)+\sigma_w^2\end{aligned}$$

with the initial conditions $\hat{\boldsymbol{x}}_i(l\mid l)$ and $P_i(l\mid l)$. The forecasted mean of $\boldsymbol{x}(k)$ with $l\leqslant k$ is $\mathrm{E}(\boldsymbol{x}(k)|I^l)=\sum_{i=1}^{s}q_i(l)\hat{\boldsymbol{x}}_i(k\mid l)$ and the forecasted covariance of $\boldsymbol{x}(k)$ (see [16]) is

$$\begin{aligned}&Cov(\boldsymbol{x}(k)|I^l)=\sum_{i=1}^{s}q_i(l)P_i(k\mid l)+\\&\sum_{1\leqslant i,j\leqslant s,i\neq j}q_i(l)q_j(l)[\hat{\boldsymbol{x}}_i(k\mid l)-\hat{\boldsymbol{x}}_j(k\mid l)][\hat{\boldsymbol{x}}_i(k\mid l)-\hat{\boldsymbol{x}}_j(k\mid l)]'\end{aligned}$$

Notice that the forecasted covariance of $\boldsymbol{x}(k)$, $Cov(\boldsymbol{x}(k)|I^l)$, is independent of the control applied at and after stage l, since for any $k>l$, $1\leqslant i,j\leqslant s$, and $i\neq j$,

$$\hat{\boldsymbol{x}}_i(k\mid l)-\hat{\boldsymbol{x}}_j(k\mid l)=\prod_{t=l}^{k-1}A(t)[\hat{\boldsymbol{x}}_i(l\mid l)-\hat{\boldsymbol{x}}_j(l\mid l)]$$

Let $\Delta_k=Cov(\boldsymbol{x}(k)|I^l)/\mathrm{Tr}[Cov(\boldsymbol{x}(N)|I^l)]$, then (14) can be rewritten as

$$\mu=\sum_{k=l}^{N-1}\mathrm{Tr}[T(k)\Delta_k]\tag{15}$$

In the above equation $T(k)$ is function of μ while Δ_k is independent of μ. Generally speaking, it is difficult to find an analytic solution of μ such that (15) holds. Thus we construct the following unconstrained optimization problem

$$\min_{\mu}\left(\mu-\sum_{k=l}^{N-1}\mathrm{Tr}[T(k)\Delta_k]\right)^2\tag{16}$$

and use the gradient method to search for an optimal value of μ, where the gradient can be derived as follows.

Theorem 4.

$$\frac{\partial T(k)}{\partial\mu}=\bar{A}'(k)\bar{A}(k)-[\bar{B}(k)\varGamma_1(k)-\bar{A}(k)]'[\bar{B}(k)\varGamma_1(k)-\bar{A}(k)]$$

where for $k = N-1, N-2, \ldots, 0$,

$$\bar{A}(k) = [\bar{B}(k+1)\Gamma_1(k+1) - \bar{A}(k+1)]A(k)$$
$$\bar{B}(k) = [\bar{B}(k+1)\Gamma_1(k+1) - \bar{A}(k+1)]B(k)$$

with boundary conditions $\bar{A}(N) = I, \bar{B}(N) = 0, \Gamma_1(N) = 0$.

3.4 On-line optimal control law

The on-line optimal control law for dual control problem (P) is given as follows.

Optimal Dual Control Algorithm for (P)**:**

Step 0. Set $k = 0$.

Step 1. Estimate $\hat{\boldsymbol{x}}_k$ based on I^k, the information set available at stage k.

Step 2. Calculate $Cov(\boldsymbol{x}(t) \mid I^k)$, $t > k$. Obtain the optimal value of μ by solving optimization problem (16) with $l = k$.

Step 3. Calculate $\{S(k)\}$, $\{T(k)\}$, $\{\Gamma_1(k)\}$, $\{\Gamma_2(k)\}$, and $\{\Gamma(k)\}$. (Note that all these matrices need to be calculated on-line, since they depend on the value of parameter μ which is adjusted on-line at every stage.)

Step 4. Calculate $\Phi(k)$ and $\Psi(k)$. Calculate the optimal control $\boldsymbol{u}(k)$ at stage k.

Step 5. Apply $\boldsymbol{u}(k)$ to the system. If $k=N-1$, stop; Otherwise, set $k=k+1$, go back to Step 1.

4 Optimal nominal dual control

We consider the general dual control problem (G) in this section. The key research issue is what is the best possible (partial) closed-loop control for (7) and what is the active learning strategy to achieve this best possible outcome. A major difficulty in solving (7) is that the optimal control cannot be determined when the future posterior probabilities are unknown, while at the same time the future posterior probabilities depend on the control applied at the early stages. In order to break up this loop, besides fixing all the posterior probabilities in later stages at their current values, a possible better solution scheme is to derive the relationship between the posterior probability and the control. A control which satisfies a deterministic version of this relationship is defined as the nominal control. The expected posterior probabilities when applying the nominal control are called nominal future posterior probabilities. Applying the nominal future posterior probabilities generated by the nominal control, the effect of future learning can be taken into account. Since in this situation, all the achievable future information is used in terms of its expected value, the control law obtained can be considered to be the best possible closed-loop control law in this sense.

Let the current time be indexed as k. For given $\boldsymbol{\lambda}^t = [\lambda_1^t, \ldots, \lambda_s^t]' \in R_+^s$, $t = k, k+1, \ldots, N$, with the currently known $\boldsymbol{\lambda}^k=[q_1(k),q_2(k),\ldots,q_s(k)]'$, consider the following optimal control problem,

$$(\mathrm{ONC}(\boldsymbol{\lambda}))\ \min \mathrm{E}\Big\{\sum_{t=k}^{N}(\sum_{i=1}^{s}\lambda_i^t J_i(t, I^t))\Big\}$$
$$\text{s.t. } \boldsymbol{x}_i(t+1) = A_i(t)\boldsymbol{x}_i(t) + B_i(t)\boldsymbol{u}(t) + \boldsymbol{w}(t),$$
$$t = k, k+1, \cdots, N-1,\ i = 1, 2, \cdots, s$$
$$\boldsymbol{y}_i(t) = C_i(t)\boldsymbol{x}_i(t) + \boldsymbol{v}(t),$$
$$t = k+1, k+2, \cdots, N,\ i = 1, 2, \cdots, s$$

where $A_i(k) = A(k,\theta_i)$, $B_i(k) = B(k,\theta_i)$, $C_i(k) = C(k,\theta_i)$, and $\boldsymbol{x}_i(k)$ and $\boldsymbol{y}_i(k)$ are the state and observation of the i^{th} fictitious system, respectively, when assuming $\theta = \theta_i$.

Let

$$X(k) = [\boldsymbol{x}_1'(k), \boldsymbol{x}_2'(k), \ldots, \boldsymbol{x}_s'(k)]',\ Y(k) = [\boldsymbol{y}_1'(k), \boldsymbol{y}_2'(k), \ldots, \boldsymbol{y}_s'(k)]'$$
$$\bar{A}(k) = diag(A_1(k), A_2(k), \ldots, A_s(k)),\ \bar{B}(k) = [B_1'(k), B_2'(k), \ldots, B_s'(k)]'$$
$$\bar{C}(k) = diag(C_1(k), C_2(k), \ldots, C_s(k)),\ \bar{Q}(k) = diag(\lambda_1^k Q(k), \lambda_2^k Q(k), \ldots, \lambda_s^k Q(k))$$
$$D_1 = [I_n, I_n, \ldots, I_n]',\ D_2 = [I_p, I_p, \ldots, I_p]'$$

where *diag* denotes a block diagonal matrix. We can obtain a compact form for the above multi-model formulation of (ONC($\boldsymbol{\lambda}$)) as follows,

$$\min \mathrm{E}\left\{X'(N)\bar{Q}(N)X(N)+\sum_{t=k}^{N-1}[X'(t)\bar{Q}(t)X(t)+\boldsymbol{u}'(t)R(t)\boldsymbol{u}(t)] \mid I^k\right\}$$
$$\text{s.t. } X(t+1)=\bar{A}(t)X(t)+\bar{B}(t)\boldsymbol{u}(t)+D_1\boldsymbol{w}(t),\ t=k,k+1,\cdots,N-1$$
$$Y(t)=\bar{C}(t)X(t)+D_2\boldsymbol{v}(t), t=k+1,\cdots,N$$

Define $\hat{X}(t)=[\hat{\boldsymbol{x}}_1'(t|t),\hat{\boldsymbol{x}}_2'(t|t),\ldots,\hat{\boldsymbol{x}}_s'(t|t)]'$, then the optimal solution to (ONC($\boldsymbol{\lambda}$)) can be derived by using dynamic programming,

$$\boldsymbol{u}^*(t)=-\Gamma(t)\hat{X}(t) \tag{17}$$

where for $t=k,k+1,\ldots,N-1$

$$\Gamma(t)=-G^{-1}(t)\bar{B}'(t)S(t+1)\bar{A}(t),\ \mathrm{G}(t)=\bar{B}'(t)S(t+1)\bar{B}(t)+R(t)$$
$$S(t)=\bar{A}'(t)S(t+1)\bar{A}(t)+\bar{Q}(t)-\Gamma'(t)\mathrm{G}(t)\Gamma(t)$$

with boundary condition $S(N)=\bar{Q}(N)$. Note that the optimal control, $\{\boldsymbol{u}^*(t)\}_{t=k}^{N-1}$, is linear in $\hat{X}$ and is $\boldsymbol{\lambda}$ dependent.

At stage k, the true observation $\boldsymbol{y}(k)$ is known, therefore $\hat{\boldsymbol{x}}_i(k|k)$ can be obtained by the Kalman filter given in (1) to (5). Since future observations can not be known in advance, a predicted nominal state trajectory $\{\hat{\boldsymbol{x}}_i^*(t)\}_{t=k+1}^N$ and a predicted nominal observation trajectory $\{\hat{\boldsymbol{y}}_i^*(t)\}_{t=k+1}^N$, can be calculated by setting all random variables at their expected values, *i.e.*

$$\hat{\boldsymbol{x}}_i^*(t+1)=A_i(t)\hat{\boldsymbol{x}}_i^*(t)+B_i(t)\boldsymbol{u}^*(t),\ t=k,k+1,\ldots,N-1$$
$$\hat{\boldsymbol{y}}_i^*(t)=C_i(t)\hat{\boldsymbol{x}}_i^*(t),\ t=k+1,k+2,\ldots,N$$

with initial condition $\hat{\boldsymbol{x}}_i^*(k)=\hat{\boldsymbol{x}}_i(k|k)$.

For $t=k+1,k+2,\ldots,N$, let $\hat{X}(t)\doteq[\hat{\boldsymbol{x}}_1^*(t)',\hat{\boldsymbol{x}}_2^*(t)',\ldots,\hat{\boldsymbol{x}}_s^*(t)']'$, and then substituting $\hat{X}(t)$ back into Eq.(17), we can calculate a predicted nominal control.

Comparing problem (ONC($\boldsymbol{\lambda}$)) with the closed-loop control problem (7) at stage k, it is easy to verify that if the values of λ_i^t stand for the posterior probabilities at every stage, the optimal control of problem (ONC($\boldsymbol{\lambda}$)) is also optimal to problem (G) at stage k. However, those posterior probabilities at the later stages are unattainable. A feasible way is to use the nominal posterior probabilities generated by the nominal control instead. The control law achieved under this framework is referred to as the optimal nominal dual control to the original problem.

Define $\hat{\boldsymbol{y}}^*(t)=\sum_{i=1}^s\lambda_i^t\hat{\boldsymbol{y}}_i^*(t)$ for $t=k+1,k+2,\ldots,N$. Using the Bayes formula, the predicted nominal posterior probability of mode i at stage k, $i=1,2,\ldots,s$, satisfy the following recursive equation:

$$\tilde{q}_i(t)=\frac{L_i(t)}{\sum_{j=1}^s\tilde{q}_j(t-1)L_j(t)}\tilde{q}_i(t-1),\ t=k+1,k+2,\ldots,N \tag{18}$$

with initial condition $q_i(k)$, where $L_i(t)$ is the same as in Eq. (**??**), except in the present case,

$$\tilde{y}(t|t-1,\theta_i)=\hat{y}^*(t)-\hat{y}_i^*(t).$$

Note from (18) that $\tilde{q}_i(t)$ is a function of $\boldsymbol{\lambda}^k,\boldsymbol{\lambda}^{k+1},\ldots,\boldsymbol{\lambda}^N$. Furthermore, the weighting coefficient λ_i^t should be equal to the nominal posterior probability $\tilde{q}_i(t)$ for all $t=k+1,k+2,\ldots,N$, in generating the nominal control. We thus construct the following optimization problem at stage k to find out the optimal $\boldsymbol{\lambda}^t$, $t=k+1,\ldots,N$,

$$\text{(O)}\ \min\sum_{t=k+1}^N\sum_{i=1}^s(\lambda_i^t-\tilde{q}_i(t))^2$$
$$\text{s.t.}\ \sum_{i=1}^s\lambda_i^t=1,\ \textit{and all}\ \lambda_i^t\geqslant 0,\quad t=k+1,\ldots,N$$

The gradient method can be used to search for the optimal value of λ_i^t. Although the derivative of each $\tilde{q}_i(t)$ with respect to λ_j^k is very involved, it can be obtained, see [13] for details.

The above derived optimal nominal dual control law can be implemented via the following algorithm.

Optimal Nominal Dual Control Algorithm for (G)**:**

Step 0. Set $k = 0$.

Step 1. Estimate $\hat{\boldsymbol{x}}_k$ based on I^k, the information set available at stage k. Calculate the posterior probabilities at stage k, $q_i(k)$ for $i = 1, 2, \ldots, s$.

Step 2. Let $\lambda_i^k = q_i(k)$. Use the gradient method to solve problem (O) in order to find out the optimal values of λ_i^t, $i = 1, 2, \ldots, s$, $t = k+1, k+2, \ldots, N$.

Step 3. Calculate matrices $\{\Gamma(t)\}$, $\{\mathrm{G}(t)\}$, $\{S(t)\}$, $t = k, k+1, \ldots, N$. Calculate the optimal nominal control $\boldsymbol{u}^*(k) = -\Gamma(k)\hat{X}(k)$.

Step 4. Apply $\boldsymbol{u}^*(k)$ to the system. If $k = N - 1$, stop; Otherwise, set $k = k + 1$, go back to Step 1.

5 Conclusions

In dual control, we need to consider the information flows in both directions. The current control affects the future probability description of the reducible uncertainty in the system. At the same time, the future evolution of the systems uncertainty will also influence the current control to be determined. One important fact is that the optimal dual control not only relies on the current information, but also depends on the impact of the future learning. Such a coupling relationship results in an intractable problem setting, thus a challenge in developing a solution method. The challenge is to break up the coupling loop.

Two innovative solution methodological frameworks have been introduced and summarized in this paper. Both solution methodological frameworks offer dual control laws with an active learning feature. The first methodological framework can be applied to situations where parameter uncertainty only exists in the observation equation. This framework can be described via its three major steps: i) Appropriately modifying the performance index in the original problem generates a tractable modified problem and obtains a corresponding analytical suboptimal control law. ii) Attaching a variance term of the final state to the modified performance index forces an addition of a learning feature to an active dual control law, while at the same time retaining the tractability of the modified problem. iii) Calculating an optimal degree of the active learning converts the performance index in the modified problem back to the original performance index on average, thus making the active dual control law optimal to the original dual control problem. The second methodological framework can be applied to general dual control problems where the parameter uncertainty exists in both state and observation equations. Adopting the concept of the nominal posterior probability enables us to calculate the future learning impact on the optimal determination of the current control, thus forming the best possible partial closed-loop feedback control. Prominent numerical results have been observed when performing these two solution methodological frameworks in various testing problems, while comparing with other existing methods in the literature.

References

1 Feldbaum A A. *Optimal Control Systems*, New York: Academic, 1965

2 Bar-Shalom Y. Stochastic dynamic programming: Caution and probing, *IEEE Transactions on Automatic Control*, 1981, **AC-26**: 1184~1194

3 Tse E, Bar-Shalom Y, Meier L III. Wide-sense adaptive dual control for nonlinear stochastic systems, *IEEE Transactions on Automatic Control*, 1973, **AC-18**: 98~108

4 Bar-Shalom Y, Tse E. Dual effect, certainty equivalence, and separation in stochastic control, *IEEE Transactions on Automatic Control*, 1974, **AC-19**: 494~500

5 Lindoff B, Holst J, Wittenmark B. Analysis of approximations of dual control, *International Journal of Adaptive Control and Signal Processing*, 1999, **13**: 593~620

6 Wittenmark B. Stochastic adaptive control methods: A survey, *International Journal of Control*, 1975, **21**: 705~730

7 Filatov N M, Unbehauen H. Survey of adaptive dual control methods, *IEE Proceedings Control Theory and Applications*, 2000, **147**: 118~128

8 Alster J, Langer P R B. A technique for dual adaptive control, *Automatica*, 1974, **10**: 627∼634
9 Wittenmark B. An active suboptimal dual controller for systems with stochastic parameters, *Automatica Control Theory Appl.*, 1975, **3**: 13∼19
10 Milito R, Padilla C S, Padilla R A, Cadorin D. An innovations approach to dual control, *IEEE Transactions on Automatic Control*, 1982, **AC-27**: 132∼137
11 Ishihara T, Abe K I, Takeda H. Extensions of innovations dual control, *International Journal of Systems Sciences*, 1988, **19**: 653∼667
12 Li D, Qian F C, Fu P L. Variance minimization approach for a class of dual control problems, *IEEE Transactions on Automatic Control*, 2002, **AC-47**: 2010∼2020
13 Li D, Fu P L, Qian F C. Optimal normal dual control for discrete-time LQG problem with unknown parameters, In: Proceedings of the Society of Instrument and Control Engineers (SICE) Annual Conference, Fukul, Japan, 2003, 2147∼2150
14 Fu P L, Li D, Qian F C. Active dual control for linear-quadratic Gaussian system with unknown parameters, In: Proceedings of the 15th World Congress of IFAC, Barcelona, Spain, 2002
15 Li D, Qian F C, Fu P L. Variance minimization in stochastic systems, *Stochastic Models and Optimization*, D. D. Yao, H. Q. Zhang and X. Y. Zhou (Eds.), Springer-Verlag, 2002
16 Casiello F, Loparo K A. Optimal policies for passive learning controllers. *Automatica*, 1989, **25**: 757∼763
17 Deshpande J G, Upadhyay T N, Lainiotis D G. Adaptive control of linear stochastic systems, *Automatica*, 1973, **9**: 107∼115
18 Li D, Ng W-L. Optimal dynamic portfolio selection: Multi-period mean variance formulation, *Mathematical Finance*, 2000, **10**: 387∼406

Duan Li Graduated from Fudan University, Shanghai, P.R.China, in 1977 and received his master degree in automatic control from Shanghai Jiaotong University, China in 1982, and his Ph.D. degree in systems engineering from Case Western Reserve University, Cleveland, OH, U.S.A., in 1987. From 1987 to 1994, he was a Faculty Member at the University of Virginia, Charlottesville. He joined the Department of Systems Engineering and Engineering Management, the Chinese University of Hong Kong, P.R.China, in December 1994, where he is currently a Professor and Department Chairman.

Duan Li's research interests include optimization and control, and he has authored and coauthored over 140 technical papers in these areas. He is an Associate Editor of IEEE Transactions on Automatic Control, was a Member of the Editorial Board of Control-Theory and Advanced Technology, and was an Associate Editor of Information and Decision Technologies. He also served as a Guest Editor for several special issues of Journal of Global Optimization, IIE Transactions on Operations Engineering, Optimization and Engineering, Control-Theory and Advanced Technology and Reliability Engineering and System Safety.

Fucai Qian Received his bachelor degree in mathematics from Shaanxi Normal University in 1984, his master degree in mathematics from Northwest University in 1988, and his Ph.D. degree in systems engineering from Xi'an Jiaotong University in 1998. He was a postdoctoral fellow at Chinese University of Hong Kong P.R.China in 1999. From 1988 to 1998 he was a lecturer at Xi'an Petroleum Institute. Since 1999 he has been with School of Automation and Information Engineering, Xi'an University of Technology where he is currently Professor. His current research interests include optimal control theory, dynamic programming, large-scale systems, and stochastic estimation, identification and control.

Peilin FU Received her bachelor degree in automatic control from the Qingdao University of Science and Technology (Previously Qingdao Institute of Chemical Technology) in 1996, her master degree in electronic engineering from the Ocean University of Qingdao in 1999, and her Ph.D. degree in Systems Engineering and Engineering Management from the Chinese University of Hong Kong in 2003. She was a postdoctoral fellow at the Hong Kong University of Science and Technology from 2003 to 2004. She is currently a Lecturer at University of California, Riverside. Her research interests include stochastic control theory, operations research and biological sequence analysis.