Completion-of-squares: revisited and extended[☆]

De-Xin Wang[a], Xi-Ren Cao[b], Li Qiu[a,1]

[a] *Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*
[b] *Shanghai Jiaotong University, Shanghai, China*

Abstract

Due to its simplicity, the *completion-of-squares* technique is quite popular in linear optimal control. However, this simple technique is limited to linear quadratic Gaussian systems. In this note, by interpreting the completion-of-squares from a new angle, we extend this technique to performance optimization of general Markov systems with the long run average criterion, leading to a new approach to performance optimization based on direct comparisons of the performance of two policies.

Keywords: Performance optimization, Performance potential, Completion-of-squares, Direct-comparison approach

1. Introduction

Completion-of-squares is one of the simplest techniques to obtain an optimal control for linear quadratic Gaussian (LQG) control problems. By completing the objective function (i.e., the system performance) in a squared form, an optimal control and the corresponding optimal value of the objective function can be obtained directly. Due to its simplicity, many real problems have been formulated under the LQG framework, e.g., the dynamic noncooperative game [1], the performance limitation problem of feedback control [2, 3], communication networks [4], and so on. However, for general Markov systems, when either "linear" or "quadratic" or "Gaussian" fails to be the case, the traditional completion-of-squares technique no longer works.

[☆]This work is supported in part by Hong Kong RGC under grant 610809.
[1]Corresponding Author. Tel: +852 2358 7067, email: eeqiu@ust.hk

In this note, by revisiting the completion-of-squares technique, we try to extend this simple method from LQG systems to general continuous-time continuous-state (CTCS) Markov systems. First, we interpret the completion-of-squares technique from a new angle, by showing that the completed square form in fact is the difference of the performance of two policies. Then, with this point of view, we extend the technique to the Markov systems, by showing that in general the Bellman optimality equation [5, 6] and the policy iteration approach [7, 8] can be directly derived from the performance difference formula between two policies. The fundamental of this idea is that performance optimization can be achieved simply by comparing the performance of two policies. Thus, this approach is named as the *direct comparison approach* [9], and the completion-of-squares method becomes its special case.

So far, many results have been obtained with the direct comparison approach for discrete-time discrete-state (DTDS) systems. For example, the nth bias optimality [10], the event-based optimization [11, 12], and the optimization of multi-chain Markov systems [13]. Recently, we have found that the optimal portfolio strategy of Markowitz's mean variance portfolio selection in a continuous-time setting [14] can also be easily obtained with the direct comparison approach. Further research on applications of this approach to other non-linear problems is going on.

The rest of the note is organized as follows. In Section 2, we state the continuous time LQG problem with the long run average criterion and introduce the operators used in dealing with the CTCS Markov systems. In Section 3, we first apply the completion-of-squares technique to solve the finite horizon LQ problem, then we show that the completed square form in fact is the difference of the performance of two policies. Then we derive the performance difference formula between two policies for the infinite horizon LQG problem with the long run average criterion. Motivated by this interpretation, in Section 4, we introduce the *direct comparison approach* to the CTCS Markov systems and show that the completion-of-squares method is its special case. Finally, some conclusion remarks are given in Section 5.

2. Problem formulation

2.1. Continuous time LQG problem

Consider an Ito process described by the following stochastic differential equation:

$$dX(t) = AX(t)dt + Bu(t)dt + DdW(t), \tag{1}$$

where $X(t)$ is an $\mathcal{R}^n$-valued state vector; $u(t)$ is an $\mathcal{R}^m$-valued control vector; $W(t)$ is an $\mathcal{R}^p$-valued standard Brownian motion; A, B and D are constant matrices with compatible sizes.

Assume that the performance function (or the cost function) f has a quadratic form:

$$f^u(X(t)) = X(t)^T Q X(t) + u(t)^T R u(t), \tag{2}$$

with Q a symmetric positive semi-definite matrix and R a symmetric positive definite matrix. The goal of control is to minimize the following long run average cost:

$$\begin{aligned}\eta^u(x) &= \lim_{L\to\infty} \frac{1}{L} E\Big\{ \int_0^L \big[X(t)^T Q X(t) + u(t)^T R u(t)\big] dt \,\Big|\, X(0) = x \Big\} \\ &= \lim_{L\to\infty} \frac{1}{L} E\Big\{ \int_0^L f^u(X(t)) dt \,\Big|\, X(0) = x \Big\}. \end{aligned} \tag{3}$$

A control (or a policy) u is a mapping from the state space to the action space, that is, at each time, if the system state is $X(t)$, we apply an action, denoted as $u(X(t))$, to the system.

Consider a linear stationary control $u_0(t) = -CX(t)$, where C is an $m \times n$ matrix. Under this policy u_0, system (1) becomes:

$$dX(t) = (A - BC)X(t)dt + DdW(t). \tag{4}$$

Assume $(A - BC)$ is a Hurwitz matrix, so that process (4) is stable with steady state distribution given by:

$$\pi(dy) = \frac{dy}{\sqrt{(2\pi)^n \det(V)}} \exp\Big(-\frac{1}{2} y^T V^{-1} y \Big), \tag{5}$$

with V the solution to:

$$(BC - A)V + V(BC - A)^T - DD^T = 0.$$

2.2. Operators for CTCS Markov systems

We review the definitions of some operators that will be used in this note [9]. Operator $\mathbb{P}_t$ maps a function $h(x)$ to another function as follows:

$$(\mathbb{P}_t h)(x) = \int_{\mathcal{R}^n} P_t(dy|x) h(y). \tag{6}$$

Obviously, we have

$$(\mathbb{P}_t h)(x) = E\Big\{ h(X(t)) \,\Big|X(0) = x\Big\}.$$

Another important operator for CTCS Markov systems is the *infinitesimal generator* $\mathbb{A}$. Specifically, for process (1), when operating on a suitable function $h(x)$, $\mathbb{A}$ takes the following form:

$$(\mathbb{A}h)(x) = (Ax + Bu)^T \frac{\partial h(x)}{\partial x} + \frac{1}{2} \operatorname{tr}\Big(\frac{\partial^2 h(x)}{\partial x^2} DD^T\Big). \tag{7}$$

The third operator ϖ maps a function $h(x)$ to a real number. Given a probability distribution $\pi(dx)$, the corresponding operator ϖ is defined as:

$$\varpi(h) = \int_{\mathcal{R}^n} h(y)\pi(dy). \tag{8}$$

For any probability distribution $\pi(dx)$, we have $\varpi(e) = 1$, where $e(x)$ is a constant function, $e(x) = 1$ for all x in space $\mathcal{R}^n$.

For the long run average criterion, as discussed in [15, 12], we assume that all policies are ergodic, meaning that there exists a steady state probability distribution π^u for each u: $\lim_{t\to\infty} P_t^u(dy|x) = e(x)\pi^u(dy)$.

Under some regularity conditions of the performance function, the order of limit and integration can be changed:

$$\begin{aligned}
\eta^u(x) &= \lim_{L\to\infty} \frac{1}{L} E\Big\{ \int_0^L f^u(X(t))dt \Big| X(0) = x\Big\} \\
&= \lim_{L\to\infty} \frac{1}{L} \int_0^L \int_{\mathcal{R}^n} f^u(y) P_t(dy|x)\, dt \\
&= \int_{\mathcal{R}^n} f^u(y)\pi^u(dy) = \varpi^u(f^u)e(x) = \bar{\eta}^u e(x),
\end{aligned} \tag{9}$$

here $\bar{\eta}^u = \varpi^u(f^u)$ is a constant number.

Lemma 1 *Assume the regularity conditions that allow the order of integration and derivation to be interchangeable. In a CTCS Markov system, for a function $h(x)$, if $\mathbb{A}^u h$ exists and is π^u-integrable, we have:*

$$\varpi^u(\mathbb{A}^u h) = 0.$$

Proof:

$$\begin{aligned}
\pi^u(\mathbb{A}^u h) &= \int_x \pi^u(dx)(\mathbb{A}^u h)(x) \\
&= \int_x \pi^u(dx)\Big\{\frac{\partial E\{h(X(t))|X(0)=x\}}{\partial t}\Big|_{t=0}\Big\} \\
&= \frac{\partial}{\partial t}\int_x \Big\{\pi^u(dx)E\{h(X(t))|X(0)=x\}\Big\}\Big|_{t=0} \\
&= \frac{\partial}{\partial t}\int_x \Big\{\pi^u(dx)(\mathbb{P}_t h(x))\Big\}\Big|_{t=0} \\
&= \frac{\partial}{\partial t}\int_x \Big\{\pi^u(dx)h(x)\Big\}\Big|_{t=0} \\
&= 0.
\end{aligned}$$

The fifth line is due to $\pi(dx) = \int_y \pi(dy)P_t(dx|y)$. □

3. Optimal control for the LQG problem

3.1. Completion-of-squares

Researchers usually study the finite horizon deterministic Linear Quadratic (LQ) optimal control problem first, then prove that the optimal control for the LQG problem has the same form as the one for the LQ case [16]. Now, we first introduce the classical completion-of-squares method in the LQ optimal control then interpret this approach from our point of view. This view will be extended to the CTCS Markov systems later.

The system equation for a finite horizon LQ system is given by:

$$dX(t) = A(t)X(t)dt + B(t)u(t)dt, \qquad t \in [0, L]. \tag{10}$$

For simplicity, the time dependence of each matrix may not be shown explicitly in the following. Given $X(0) = x$, the control objective is to find a $u(t)$ to minimize:

$$\eta^u(x) = \int_0^L [X(t)^T Q X(t) + u(t)^T R u(t)]dt + X(L)^T F X(L) \tag{11}$$

with Q, R, and F suitable matrices. If the Riccati equation:

$$-\dot{M} = A^T M + MA + Q - MBR^{-1}B^T M, \qquad M(L) = F, \tag{12}$$

has a solution on $[0, L]$, by adding "$-x^T M(0)x$" to (11), we get:

$$\begin{aligned}\eta^u(x) - x^T M(0)x &= \int_0^L \Big(X(t)^T Q X(t) + u(t)^T R u(t) + \frac{d}{dt}(X(t)^T M X(t))\Big)dt \\ &= \int_0^L \Big(X^T Q X + u^T R u + (AX + Bu)^T M X \\ &\qquad + X^T M(AX + Bu) + X^T \dot{M} X\Big)dt \\ &= \int_0^L (u + R^{-1}B^T M X)^T R (u + R^{-1}B^T M X)dt. \qquad (13)\end{aligned}$$

From the quadratic form of (13), we immediately obtain an optimal control:

$$u^*(t) = -R(t)^{-1}B^T(t)M(t)X(t),$$

and the corresponding minimal cost:

$$\eta^{u^*}(x) = x^T M(0)x.$$

For infinite horizon LQ problems or stochastic LQG problems (both finite and infinite horizon cases), we can prove that the optimal control is also a linear feedback control given by $u^*(t) = -R^{-1}B^T M X(t)$. For details, we refer to the textbook [16].

3.2. Another viewpoint of completion-of-squares

An observation in the completion-of-squares for the LQ control is that the term "$x^T M(0)x$" in fact corresponds to the system performance of the optimal policy, thus $\eta^u(x) - x^T M(0)x = \eta^u(x) - \eta^{u^*}(x)$ actually gives us the performance difference between two policies, one of which is the optimal policy, and the other can be any policy.

This result is consistent with the authors' research experience in that performance optimization in a policy space can be based on comparisons of the performance of any two policies [12]. As we state above, for systems with special structure properties, like the LQG case, the optimal policy can be directly obtained from the performance difference between two policies. A natural extension of this result is very appealing: for a general Markov system, by simply comparing the performance of any two policies, we may obtain the Bellman optimality equation and the policy iteration approach for searching an optimal policy. This is the extension of "completion-of-squares"

technique to the general case and offers a simple explanation for general performance optimization problems: they are as simple as "completion-of-squares"!

Since it is based on direct comparisons of the performance of any two policies, we call this approach the *direct-comparison approach* [9]. As an example, we will demonstrate this approach by solving the continuous time LQG problem.

3.3. Performance difference formula for the LQG problem

We have the following lemma for the continuous time LQG system (1):

Lemma 2 *For system (1), the system performance (3) under a linear policy* $u_0(t) = -CX(t)$ *is given by:*

$$\bar{\eta}^{u_0}(x) = \int_y \pi^{u_0}(dy) f^{u_0}(y) = \mathrm{tr}(MDD^T) \tag{14}$$

with matrix M *the solution to:*

$$(A - BC)^T M + M(A - BC) + (Q + C^T RC) = 0 \tag{15}$$

Proof: By (5), $\pi^{u_0}(dx)$ is normally distributed with mean 0 variance matrix V. Thus, by (9), we have:

$$\bar{\eta}^{u_0} = \int_y \pi^{u_0}(dy) f^{u_0}(y) = \mathrm{tr}[(Q + C^T RC)V]. \tag{16}$$

It is not hard to verify $\mathrm{tr}[(Q + C^T RC)V] = \mathrm{tr}(MDD^T)$. □

Consider a general policy $u(x)$ (not necessarily linear) and a linear policy $u_0(t) = -CX(t)$, if choosing $h(x) = x^T Mx$ in Lemma 1, we have:

Theorem 1 *The performance difference formula between a general policy* u *and a linear policy* $u_0 = -Cx$ *is given by:*

$$\bar{\eta}^u - \bar{\eta}^{u_0} = \int_x \pi^u(dx) \Big\{ [u + R^{-1}B^T Mx]^T R[u + R^{-1}B^T Mx] \\ - x^T [MBR^{-1}B^T M + C^T RC - 2C^T B^T M]x \Big\}. \tag{17}$$

Proof: Noting $\int_x \pi^u(dx)(\mathbb{A}^u h)(x) = 0$, we have

$$\begin{aligned}
\bar{\eta}^u - \bar{\eta}^{u_0} &= \int_x \pi^u(dx)\Big\{f^u(x) - \bar{\eta}^{u_0} e(x)\Big\} \\
&= \int_x \pi^u(dx)\Big\{f^u(x) + \mathbb{A}^u h(x) - \bar{\eta}^{u_0} e(x)\Big\} \\
&= \int_x \pi^u(dx)\Big\{x^T Q x + u^T R u + 2(Ax + Bu)^T M x\Big\} \\
&= \int_x \pi^u(dx)\Big\{[u + R^{-1}B^T M x]^T R[u + R^{-1}B^T M x] \\
&\qquad\qquad - x^T[MBR^{-1}B^T M + C^T R C - 2C^T B^T M]x\Big\}.
\end{aligned}$$

The fourth line above is obtained by replacing Q in the third line with $-(A - BC)^T M - M(A - BC) - C^T R C$ according to (15). □

Since $[u + R^{-1}B^T M x]^T R[u + R^{-1}B^T M x] \geq 0$, we conclude that if we choose a matrix C such that:

$$MBR^{-1}B^T M + C^T R C - 2C^T B^T M = 0, \tag{18}$$

then the performance difference between u and u_0 becomes:

$$\eta^u - \eta^{u_0} = \int_x \pi^u(dx)\Big\{[u + R^{-1}B^T M x]^T R[u + R^{-1}B^T M x]\Big\} \geq 0. \tag{19}$$

Since the above inequality holds for all policy u, we know the linear feedback control u_0 should be optimal. Solving (18) yields:

$$C = R^{-1}B^T M. \tag{20}$$

Replacing C in (15) by $R^{-1}B^T M$, we obtain the well-known algebraic Riccati equation:

$$A^T M + MA - MBR^{-1}B^T M + Q = 0. \tag{21}$$

4. Direct comparison approach to Markov systems

Although the optimal policy can be obtained directly from the quadratic form of the performance difference formula between two policies for the LQG problem, it is often impossible to find out the optimal policy directly from the performance difference between two policies for general Markov systems

due to the loss of the quadratic property; however, we will see that the performance difference formula does give us the Bellman's optimality equation for the optimal policy and the policy iteration approach to a solution of the Bellman equation. In this sense, the direct comparison approach based on the performance difference formula is an extension of the completion-of-squares method from the LQG problem to general Markov systems.

4.1. Performance difference formula

The Poisson equation for an ergodic Markov process with infinitesimal generator $\mathbb{A}$ and cost function f is given by [15, 12]:

$$-\mathbb{A}g(x) + \bar{\eta}e(x) = f(x). \tag{22}$$

The solutions to the Poisson equation differ by an additive term: if $g(x)$ is a solution to the Poisson equation, so is $g(x) + cr(x)$, with $\mathbb{A}r(x) = 0$ and c being any constant. Any solution $g(x)$ is called *performance potential*, or simply *potential*, in the direct comparison approach.

It can be verified that, if exists, the following function satisfies the Poisson equation (22) (c.f. equation (8.22) in [15]):

$$g(x) = \lim_{L\to\infty} \int_0^L E\Big\{[f(X(t)) - \bar{\eta}]\Big|X(0) = x\Big\}dt. \tag{23}$$

(23) is called the sample path based expression of potential function $g(x)$.

Consider two policies u and u'. The corresponding processes of these two policies are denoted as $X = \{X(t), t \in [0, \infty)\}$ and $X' = \{X'(t), t \in [0, \infty)\}$, respectively. In the following, we use superscript "$'$" to denote quantities associated with X'.

Theorem 2 *Assume $g(x)$ satisfies the regularity conditions in Lemma 1. We have the following performance difference for policies u' and u:*

$$\bar{\eta}' - \bar{\eta} = \pi'\{(f' + \mathbb{A}'g) - (f + \mathbb{A}g)\}. \tag{24}$$

proof: Left-multiply both sides of the Poisson equation (22) with π', we get

$$\bar{\eta} = \pi' f + \pi'(\mathbb{A}g).$$

Noting $\pi'(\mathbb{A}'h) = 0$, we have:

$$\begin{aligned}\bar{\eta}' - \bar{\eta} &= \pi' f - \bar{\eta} + \pi'(f' - f)\\ &= -\pi'(\mathbb{A}g) + \pi'(f' - f)\\ &= \pi'\{(f' + \mathbb{A}'g) - (f + \mathbb{A}g)\}.\end{aligned}$$

(24) is called the *performance difference formula*. □

The performance difference formula indicates that the performance difference between any two policies in a Markov system can be decomposed into the product of two quantities, π' and g. The contribution of policy u' is reflected by its steady state distribution π' and the contribution of policy u is captured by its performance potential function g.

In the following, we assume for all policy u, its steady state distribution $\pi^u(x) > 0$ for all state x in state space $\mathcal{R}^n$. This assumption makes sense since the noise part of a CTCS Markov system is generally modeled by a Brownian motion, which is supported in the entire state space [9]. Thus, for a non-negative π-integrable function $h(x)$, we have

$$\pi(h) = \int_x \pi(dx)h(x) \geq 0.$$

If in addition $h(x) > 0$, for all $x \in \mathcal{B} \subset \mathcal{R}^n$, with $\mathcal{B}$ having a positive Lebesgue measure, then we have

$$\pi(h) > 0.$$

Therefore, if

$$[(f' + \mathbb{A}'g) - (f + \mathbb{A}g)](x) \geq 0, \quad \forall x \in \mathcal{R}^n \tag{25}$$

then from the performance difference formula (24), policy u performs better than, or at least the same as, policy u'. If in addition,

$$[(f' + \mathbb{A}'g) - (f + \mathbb{A}g)](x) > 0, \quad \forall x \in \mathcal{B} \subset \mathcal{R}^n, \tag{26}$$

then policy u must be better than policy u'.

Thus, based on the performance difference formula, we can identify which policy is better by only analyzing the system under one policy. This easily leads to the optimality equation and the policy iteration method.

4.2. Optimality equation and policy iteration

The next theorem follows directly from the performance difference formula (24).

Theorem 3 *For a CTCS Markov system, assume all policies are ergodic and the performance function $f^u(x)$ satisfies the regularity conditions such that $\mathbb{A}^u g^u$ exists and is π^u-integrable. A policy $\hat{u}(x)$ is optimal, if and only if*

$$(f^{\hat{u}} + \mathbb{A}^{\hat{u}} g^{\hat{u}})(x) \leq (f^u + \mathbb{A}^u g^{\hat{u}})(x), \quad \forall x \in \mathcal{R}^n, \tag{27}$$

for all policies u.

Proof: "$\Longrightarrow$" If for all policy u, we have (27), then by the performance difference formula (24) and the fact $\pi(x) > 0$, we have $\eta^{\hat{u}} \leq \eta^{u}$, for all u in the policy space, which indicates that $\hat{u}$ is optimal.

"$\Longleftarrow$" Suppose (27) does not hold, by the continuity property, we know there exists a policy $\tilde{u}$ and a set H with positive Lebesgue measure, in which we have:

$$(f^{\hat{u}} + \mathbb{A}^{\hat{u}} g^{\hat{u}})(x) > (f^{\tilde{u}} + \mathbb{A}^{\tilde{u}} g^{\hat{u}})(x), \quad \forall x \in H$$

Construct a policy $\bar{u}$ by:

$$\bar{u}(x) = \begin{cases} \hat{u}(x), & x \notin H \\ \tilde{u}(x), & x \in H. \end{cases}$$

By the performance difference formula for policy $\bar{u}$ and $\hat{u}$, and noting that $\pi(H) > 0$, we have:

$$\eta^{\bar{u}} < \eta^{\hat{u}},$$

which shows that $\hat{u}$ cannot be optimal. □

From (27), a policy $\hat{u}$ is optimal if and only if the following optimality equation

$$\min_{u \in \mathcal{U}} \{f^{u} + \mathbb{A}^{u} g^{\hat{u}}\} = f^{\hat{u}} + \mathbb{A}^{\hat{u}} g^{\hat{u}} = \eta^{\hat{u}} \tag{28}$$

holds for all $x \in \mathcal{R}^n$. This is the Bellman's optimality equation [5, 15].

One should note that it is generally difficult to check the sufficient and necessary condition (28) for a real system. It requires first to solve the Poisson equation (22) to get the potential function $g^{\hat{u}}$, and then to verify the relation (28) for every $x \in \mathcal{R}^n$ and every feasible action at x. However, for systems with closed-form solutions, such as the LQG problem, the condition can be verified analytically, see the discussion in the next subsection.

Next, the results in (25) and (26) lead naturally to the policy iteration method to find an optimal policy [9]. Roughly speaking, at the kth step with policy u_k, $k = 0, 1, \cdots$, we set

$$u_{k+1}(x) = \arg\{\min_{u \in \mathcal{U}} [f^{u}(x) + \mathbb{A}^{u} g^{u_k}(x)]\}, \quad x \in \mathcal{S}, \tag{29}$$

with g^{u_k} being any solution to the Poisson equation (22) for $(\mathbb{A}^{u_k}, f^{u_k})$. If at some x, $u_k(x)$ attains the minimum, we set $u_{k+1}(x) = u_k(x)$. The iteration starts with any policy u_0 and stops if u_{k+1} and u_k differ only on a set with a zero Lebesgue measure. From (25), the performance improves at every iteration; from (26), when the algorithm stops, it stops at an optimal policy.

Again, for any real Markov system, to solve the Poisson equation is computationally involved, and to implement policy iteration is as difficult as to verify the optimality equation. In addition, to find the conditions and to prove that, under these conditions, the iteration indeed stops is not an easy task [8]. This goes beyond the scope of this note.

Next, we will illustrate how the LQG problem fits the direct comparison framework.

4.3. The LQG problem

In the LQG problem with the long run average criterion, if take $u_0(x) = -Cx$ as an initial policy, we can derive its potential function:

$$g^{u_0}(x) = x^T M^{u_0} x + ce(x),$$

where c is a constant number. Consider any policy u (not necessarily linear), by (7), we have:

$$\begin{aligned}\mathbb{A}^u g^{u_0}(x) &= (Ax + Bu)^T \frac{\partial g^{u_0}(x)}{\partial x} + \frac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2 g^{u_0}(x)}{\partial x^2} DD^T\right) \\ &= (Ax + Bu)^T 2M^{u_0} x + \mathrm{tr}\left(M^{u_0} DD^T\right).\end{aligned}$$

Apply policy iteration algorithm (29), and notice that both $x^T Qx$ and $\mathrm{tr}\left(M^{u_0} DD^T\right)$ are independent on u, we have:

$$\begin{aligned}u'(x) &= arg\Big\{ \min_{all\ u} \Big[f^u(x) + \mathbb{A}^u g^{u_0}(x) \Big] \Big\} \\ &= arg\Big\{ \min_{all\ u} \Big[u^T Ru + (Ax + Bu)^T 2M^{u_0} x \Big] \Big\} \\ &= -R^{-1} B^T M^{u_0} x \\ &= -C'x\end{aligned}$$

where $C' = R^{-1} B^T M^{u_0}$.

This result shows that if the original policy is linear, then the improved policy constructed by the policy iteration algorithm is also linear. Obviously, if $C' = C$, the linear policy u_0 satisfies the optimality equation (28) and thus is optimal. By letting $C = R^{-1} B^T M$, we may simplify equation (15):

$$\begin{aligned}0 &= (BC - A)^T M + M(BC - A) - (Q + C^T RC) \\ &= (BR^{-1} B^T M - A)^T M + M(BR^{-1} B^T M - A) - (Q + (R^{-1} B^T M)^T B^T M) \\ &= -A^T M - MA + MBR^{-1} B^T M - Q.\end{aligned}$$

Now, equation (15) becomes:

$$A^T M + MA - MBR^{-1}B^T M + Q = 0, \tag{30}$$

which is the algebraic Riccati equation.

5. Conclusion

In this note, we first show that, in the completion-of-squares technique for linear quadratic Gaussian systems, the completed square term in fact is the difference of the performance of two policies. The optimal policy can be easily obtained with the form of the completed square. We show that this intuitive approach can be extended to the performance optimization problems of general Markov systems. We find that the performance difference formula plays a similar role as the completion-of-squares in the LQG problem. From the performance difference formula, we can easily derive (with no dynamic programming) Bellman's optimality equation and the policy iteration method. This leads to the "direct comparison approach", and the completion-of-squares for the LQG problem becomes a special case of this new approach.

Acknowledgement

The authors are grateful to the associate editor and three anonymous reviewers for their valuable comments which improved the quality of this note.

[1] T. Basar and P. Bernhard, *H-infinite Optimal Control and Related minimax Design Problems – A Dynamic Game Approach, Second Edition* , Birkhauser Boston, 1995.

[2] L. Qiu and E. Davision, "Performance limitations of non-minimum phase systems in the servomechanism problem," *Automatica*, vol. 29, pp. 337–349, 1993.

[3] L. Qiu and J. Chen, "Time domain characterizations of performance limitations of feedback control," *Learning, Control, and Hybrid Systems, Y. Yamamoto and S. Hara, editors, Springer-Verlag*, pp. 397–415, 1998.

[4] V. Gupta, B. Hassibi, and R. M. Murray, "Optimal LQG control across packet-dropping links," *Systems and Control Letters*, vol. 56, pp. 439–446, 2007.

[5] R. E. Bellman, *Dynamic Programming*, Courier Dover Publications, 2003.

[6] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. I, II, Belmont, MA: Athena Scientific, 2007.

[7] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, 1994.

[8] S. P. Meyn, "The policy iteration algorithm for average reward Markov decision processes with general state space," *IEEE Transactions on Automatic Control*, vol. 42, 1663–1680, 1997.

[9] X. R. Cao, D. X. Wang, T. Lu and Y. F. Xu, "Stochastic control via direct comparison," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 21, pp. 11–38, 2011.

[10] X. R. Cao and J. Y. Zhang, "The nth-order bias optimality for multichain Markov decision processes," *IEEE Transactions on Automatic Control*, vol. 53, pp. 496–508, 2008.

[11] X. R. Cao and J. Y. Zhang, "Event-based optimization of Markov systems," *IEEE Transactions on Automatic Control*, vol. 53, pp. 1076–1082, 2008.

[12] X. R. Cao, *Stochastic Learning and Optimization – A Sensitivity-Based Approach*, Springer, 2007.

[13] X. R. Cao and X. P. Guo, "A unified approach to Markov decision problems and performance sensitivity analysis with discounted and average criteria: multichain cases", *Automatica*, vol. 40, pp. 1749–1759, 2004.

[14] X. Y. Zhou and D. Li, "Continuous-time mean-variance portfolio selection: a stochastic LQ framework," *Applied Mathematics and Optimization*, Vol. 42, pp. 19–33, 2000.

[15] S. P. Meyn, *Control Techniques for Complex Networks*, Cambridge, 2008

[16] M. Green and D. J. N. Limebeer, *Linear Robust Control*. Prentice Hall, New Jersey, 1995.