Operations Research Letters 40 (2012) 521–528

Contents lists available at SciVerse ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Linear–quadratic control and information relaxations

Martin Haugh [a,*], Andrew E.B. Lim [b,c]

[a] Department of IE and OR, Columbia University, New York, NY 10027, United States

[b] NUS Business School, National University of Singapore, 119245, Singapore

[c] Department of IE and OR, University of California, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 1 May 2011
Accepted 10 July 2012
Available online 31 August 2012

Keywords:
Duality
Information relaxations
Linear–quadratic control

ABSTRACT

We apply recently developed duality methods to the classic linear–quadratic (LQ) control problem. Using value-function and gradient methods, we derive two dual optimal penalties for the LQ problem for when the control space is unconstrained. These penalties may be used to, for example, evaluate sub-optimal policies for constrained LQ problems. We also compare these penalties to the dual penalty of Davis and Zervos (1995) [7] and note that some of these duality ideas have been in circulation for some time.

1. Introduction

In this note we apply recently developed duality techniques for stochastic control problems to the classic linear–quadratic (LQ) control problem. These techniques were developed independently by [10,3] and are based on relaxing the decision-maker's information constraints. This work was motivated in part by the duality techniques developed in [6,9,8] for optimal stopping problems and the pricing of American options in particular. These duality techniques [3,10] can be used to evaluate sub-optimal policies for control problems that are too difficult to solve exactly. In particular, the sub-optimal policy may be used to compute both primal and dual bounds on the optimal value function. The primal bound can be computed by simply simulating the sub-optimal policy whereas the aforementioned duality techniques can be used with the sub-optimal policy (or indeed some other policy) to compute the dual bound. If the primal and dual bounds are close to one another then we know that the sub-optimal policy is close to optimal. We believe that these techniques will play an increasingly important role in the area of sub-optimal control and that there are many interesting related research questions to be resolved.

In this note we focus on finite horizon LQ problems and we derive two dual optimal penalties for when the control space is unconstrained. The first penalty is derived using knowledge of the optimal value function whereas the second penalty is derived using the gradient methods developed by [2] in the context of dynamic portfolio optimization under transaction costs. These penalties and others may then be used to evaluate sub-optimal policies for constrained LQ problems when it is not possible to determine the optimal policy exactly. If the controls are not too constrained then we expect the optimal unconstrained penalties to be close to optimal for the constrained problem and therefore to lead to good dual bounds. We emphasize that the derivation of these penalties is quite straightforward and is only a modest contribution to this growing literature.

We also compare these dual techniques to the work of [7] who used Lagrange multipliers to show that a stochastic LQ problem may be reduced to a deterministic LQ problem. Indeed it is easy to show that their Lagrange multipliers are also optimal dual penalties. This connection to the earlier work of [7] is not widely known and it highlights that some of these duality ideas have been in circulation for some time. In fact within the stochastic control literature the idea of relaxing the non-anticipativity constraints goes back at least to [4,5]. It has also been a feature of the stochastic programming literature where it has often been applied to stochastic programs with just a few periods. It is also interesting to note that these developments appear to mirror the development of the duality methods for solving optimal stopping problems as mentioned earlier. For this case, [6] used a dual formulation to characterize the optimal solution to the optimal stopping problem. Neither [9] nor [8] were not aware of [6] when they independently developed dual formulations of the optimal stopping problem. (The paper [6] was published as a book chapter and as a result, was not widely known until some time afterwards.) Their focus, however, was on using these dual formulations to construct good dual bounds for optimal stopping problems that were too difficult to solve exactly. This was also the focus of [3,10] who developed

* Corresponding author.
E-mail addresses: martin.b.haugh@gmail.com, mh2078@columbia.edu (M. Haugh), lim@ieor.berkeley.edu (A.E.B. Lim).

0167-6377/$ – see front matter
doi:10.1016/j.orl.2012.08.010

their dual techniques with a view to using them to evaluate sub-optimal strategies. In contrast, the seminal work of Davis and his co-authors appears to have been only on characterizing optimal solutions.

A further contribution of this note is a comparison of the three optimal dual penalties for the unconstrained LQ problem. We emphasize that while the three penalties are dual optimal, they are not actually identical. Indeed as demonstrated by [3], the penalty function constructed using the optimal value function is almost surely optimal whereas the other penalties are only optimal in expectation. (We will clarify this statement in Section 2). This observation may have significant implications when we use dual penalties to evaluate sub-optimal policies for constrained control problems in general. However, it is also worth mentioning that dual penalties constructed using value-function approximations can be quite challenging to work with and so we expect the gradient approach to often be a viable alternative.

The remainder of this paper is organized as follows. We briefly review the duality approach of [3] in Section 2 and then derive optimal dual penalties for the unconstrained LQ problem in Section 3. We review the results of [7] in Section 4 and show that the dual penalty in that section and the dual penalties derived in Section 3 only differ by a zero-mean term. We conclude in Section 5 and identify several directions for further research.

2. A review of duality based on information relaxations

We begin with a general finite horizon discrete-time dynamic program with a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. Time is indexed by $k = 0, \ldots, N$ and the evolution of information is described by the filtration $\mathbb{F} = \{\mathcal{F}_0, \ldots, \mathcal{F}_N\}$ with $\mathcal{F} = \mathcal{F}_N$. We make the usual assumption that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so that the decision-maker starts out with no information regarding the outcome of uncertainty. There is a state vector, $x_k \in S_k$, where S_k is the time-k state space. The dynamics of x_k satisfy

$$x_{k+1} = f(x_k, u_k, w_{k+1}), \quad k = 0, \ldots, N-1 \tag{1}$$

where $u_k \in U_k(x_k)$ is the control taken at time k and w_{k+1} is an $\mathcal{F}_{k+1}$-measurable random disturbance. A feasible strategy, $u := (u_0, \ldots, u_{N-1})$, is one where each individual action, $u_k \in U_k(x_k)$, is $\mathcal{F}_k$-measurable. (We note that [3] use a slightly more general formulation where they assume that $U_k = U_k(u_0, \ldots, u_{k-1})$ can depend on the entire history of past actions and states.) In particular, we require the decision-maker's strategy, u, to be $\mathcal{F}_k$-adapted. We use $\mathcal{U}_{\mathbb{F}}$ to denote the set of all such $\mathcal{F}_k$-adapted strategies. The objective is to select a feasible strategy, u, to minimize the expected total cost:

$$h(u) := g_N(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k)$$

where we assume without loss of generality that each $g(x_k, u_k)$ is $\mathcal{F}_k$-measurable. In particular, the decision-maker's problem is then given by

$$J_0(x_0) \equiv \inf_{u \in \mathcal{U}_{\mathbb{F}}} \mathbb{E}_0 \left[g_N(x_N) + \sum_{k=0}^{N-1} g(x_k, u_k) \right] \tag{2}$$

where the expectation in (2) is taken over the set of possible outcomes, $w = (w_1, \ldots, w_{N-1}) \in \Omega$. To emphasize that the total cost is random, we will often write $h(u, w)$ for $h(u)$. Letting J_k denote the time-k value function for the problem (2), the associated dynamic programming recursion is given by

$$\begin{aligned} J_N(x_N) &:= g_N(x_N) \\ J_k(x_k) &:= \inf_{u_k \in U_k(x_k)} \{ g(x_k, u_k) + \mathbb{E}_k [J_{k+1}(f(x_k, u_k, w_{k+1}))] \} \\ & k = 0, \ldots, N-1, \end{aligned} \tag{3}$$

where we write $\mathbb{E}_k[\cdot]$ for $\mathbb{E}[\cdot \mid \mathcal{F}_k]$. In practice of course it is often too difficult or time-consuming to perform the iteration in (3). This can occur, for example, if the state vector, x_k, has high dimension or if the constraints imposed on the controls are too complex or difficult to handle. In such circumstances, we must be satisfied with sub-optimal solutions or policies.

2.1. The dual formulation

We now briefly describe the dual formulation of [3], which should be consulted for further details and proofs of the results given below. Note, however that [3] focus on problems where the primal problem is a maximization problem. We have chosen to specify our primal problem as a minimization problem so that we are consistent with the usual formulation of linear–quadratic problems where the goal is to minimize expected total costs.

We say that the filtration $\mathbb{G} := \{\mathcal{G}_k\}$ is a relaxation of $\mathbb{F}$ if, for each k, $\mathcal{F}_k \subseteq \mathcal{G}_k$. We write $\mathbb{F} \subseteq \mathbb{G}$ to denote such a relaxation. For example, the perfect information filtration, $\mathbb{I} := \{\mathcal{I}_k\}$, is obtained by taking $\mathcal{I}_k = \mathcal{F}$ for all k. We let $\mathcal{U}_{\mathbb{G}}$ denote the set of all $\mathcal{G}_k$-adapted strategies. It is clear then that for any relaxation, $\mathbb{G} := \{\mathcal{G}_k\}$, we have $\mathcal{U}_{\mathbb{F}} \subseteq \mathcal{U}_{\mathbb{G}} \subseteq \mathcal{U}_{\mathbb{I}}$, so as we relax the filtration, we expand the set of feasible policies.

The set of penalties, $\mathcal{Z}$, is the set of all functions $z(u, w)$ that, like the set of costs, depend on the choice of actions, u, and the outcome, w. We define the set, $\mathcal{Z}_{\mathbb{F}}$, of dual feasible penalties to be those penalties that do not penalize temporally feasible, i.e. $\mathcal{F}_k$-adapted, strategies. In particular, we define

$$\mathcal{Z}_{\mathbb{F}} := \{ z \in \mathcal{Z} : \mathbb{E}_0 [z(u, w)] \le 0 \text{ for all } u \in \mathcal{U}_{\mathcal{F}} \}. \tag{4}$$

We then have the following version of weak duality, the proof of which follows immediately from the definition of dual feasibility in (4) and because $\mathbb{G}$ is a relaxation of $\mathbb{F}$.

Lemma 1 (*Weak Duality*)**.** *If u_F and z are primal and dual feasible respectively, i.e. $u_F \in \mathcal{U}_{\mathbb{F}}$ and $z \in \mathcal{Z}_{\mathbb{F}}$, then*

$$\mathbb{E}_0 [h(u_F, w)] \ge \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0 [h(u_G, w) + z(u_G, w)]. \tag{5}$$

Therefore any dual feasible penalty and information relaxation provide a lower bound on the optimal value function. Clearly weaker relaxations lead to weaker lower bounds as a weaker relaxation will increase the set of feasible policies over which the infimum is taken in (5). In the case of the perfect information relaxation we have $\mathbb{G} = \mathbb{I}$ and the lower bound takes the form

$$\begin{aligned} \mathbb{E}_0 [h(u_F, w)] &\ge \inf_{u \in \mathcal{U}_{\mathbb{I}}} \mathbb{E}_0 [h(u, w) + z(u, w)] \\ &= \mathbb{E}_0 \left[\inf_{u \in \mathcal{U}_{\mathbb{I}}} \{ h(u, w) + z(u, w) \} \right]. \end{aligned}$$

For a given information relaxation, we can optimize the lower, i.e., dual bound, by optimizing over the set of dual feasible penalties. This leads to the dual of the primal DP:

$$\text{Dual Problem: } \sup_{z \in \mathcal{Z}_{\mathbb{F}}} \left\{ \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0 [h(u_G, w) + z(u_G, w)] \right\}. \tag{6}$$

By weak duality, if we identify a policy, u_F, and penalty, z, that are primal and dual feasible, respectively, such that equality in (5) holds, then u_F and z must be optimal for their respective problems. Moreover, if the primal problem (2) has a finite solution, then so too has the dual problem (6), and there is no duality gap. This yields the following result.

Theorem 1 (*Strong Duality*)**.** *Let* $\mathbb{G}$ *be a relaxation of* $\mathbb{F}$*. Then*

$$\inf_{u \in \mathcal{U}_{\mathbb{F}}} \mathbb{E}_0 \left[h(u_F, w)\right] = \sup_{z \in \mathcal{Z}_{\mathbb{F}}} \left\{ \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0 \left[h(u_G, w) + z(u_G, w)\right] \right\}. \quad (7)$$

Furthermore, if the primal problem on the left is bounded, then the dual problem on the right has an optimal solution, $z^* \in \mathcal{Z}_{\mathbb{F}}$*, and there is no duality gap.*

There is also a version of complementary slackness.

Theorem 2 (*Complementary Slackness*)**.** *Let* u_F^* *and* z^* *be feasible solutions for the primal and dual problems, respectively, with information relaxation* $\mathbb{G}$*. A necessary and sufficient condition for these to be optimal solutions is that* $\mathbb{E}_0\left[z^*(u_F^*)\right] = 0$ *and*

$$\mathbb{E}_0\left[h(u_F^*, w) + z^*(u_F^*, w)\right] = \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0\left[h(u_G, w) + z^*(u_G, w)\right]. \quad (8)$$

Note that Theorem 2 implies that with an optimally chosen penalty, z^*, the decision-maker in the dual DP will be happy to choose a non-anticipative control, despite not being restricted to doing so. As shown by [3], we can also take advantage of any structural information regarding the optimal solution to the primal problem. In particular, if it is known that the optimal solution to the primal problem has a particular structure, then we can restrict ourselves to policies with the same structure when solving the dual optimization problem.

2.2. Using the dual formulation to construct dual bounds

In practice it is often the case that we are unable to compute the solution to the primal DP exactly. However, we can compute a lower bound on the optimal value function of the primal DP by starting with a dual feasible penalty function, $z(u, w) := \sum_{k=0}^{N-1} z_k(u, w)$, and then using this penalty function on the right-hand side of (5). In particular, we do not seek to optimize over the dual penalty but hope that the penalty function that we have chosen is sufficiently good to result in a small duality gap. We will only consider perfect information relaxations in this paper, so $\mathbb{G} = \mathbb{I}$. This is because the perfect information relaxations result in dual problems that are deterministic optimization problems which are often easy to solve. If we use other information relaxations then the resulting dual problems remain stochastic, in which case it is generally difficult to handle constraints on the control vector, u. If we use $J_{db}(x_0; z)$ to denote the resulting dual or lower bound from solving the dual problem then we see that $J_{db}(x_0; z)$ satisfies

$$J_{db} = \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0 \left[h(u_G, w) + z(u_G, w)\right] \quad (9)$$

$$= \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \mathbb{E}_0 \left[g_N(x_N) + \sum_{k=0}^{N-1} \left(g(x_k, u_k) + z_k(u_G, w)\right) \right]$$

$$= \mathbb{E}_0 \left[\inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} \left(g(x_k, u_k) + z_k(u_G, w)\right) \right\} \right]. \quad (10)$$

The optimization problem inside the expectation in (10) can be solved as a deterministic optimization problem after substituting for the x_k's using (1). An unbiased dual bound on the optimal value function, $J_0(x_0)$, can therefore be estimated by first simulating M paths of the noise process, w. If we label these paths $w^{(i)} := (w_0^{(i)}, \ldots, w_{N-1}^{(i)})$ for $i = 1, \ldots, M$, and set

$$J_{db}^{(i)}(x_0; z) := \inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \left\{ g_N(x_N) + z_N(u, w^{(i)}) + \sum_{k=0}^{N-1} \left(g(x_k, u_k) + z_k(u_G, w^{(i)})\right) \right\} \quad (11)$$

with the x_k's satisfying (1) then

$$J_{db}(x_0; z) := \frac{1}{M} \sum_{i=1}^{M} J_{db}^{(i)}(x_0; z) \quad (12)$$

is an unbiased dual or lower bound for $J_0(x_0)$.

2.3. Constructing dual penalties

We outline two methods for constructing dual feasible penalties that we will use in Section 3. We emphasize again that we only consider perfect information relaxations, so $\mathcal{G}_k = \mathcal{I}_k$ for all k.

Using value-function approximations to construct dual penalties

[3] propose taking

$$z_k(u, w) := \mathbb{E}_k\left[v_k(u, w)\right] - \mathbb{E}\left[v_k(u, w) \mid \mathcal{G}_k\right] = \mathbb{E}_k\left[v_k(u, w)\right] - v_k(u, w) \quad (13)$$

where $v_k(u, w)$ only depends on $(u_0, \ldots, u_k)$ and where (13) follows since we are using the perfect information relaxation here and so $\mathcal{G}_k = \mathcal{F}$ for all k. It is easy to see that penalties defined in this manner are dual feasible. Indeed we easily obtain that $\mathbb{E}_0\left[z_k(u_F)\right] = 0$ for all $u_F \in \mathcal{U}_{\mathbb{F}}$. [3] call the $v_k(u)$'s *generating functions* and show that if we take $v_k(u) := J_{k+1}(x_{k+1})$ where J_{k+1} is the optimal value function of the primal DP, then the corresponding penalty,

$$z(u, w) = \sum_{i=0}^{N-1} \left(\mathbb{E}_k\left[J_{k+1}(x_{k+1})\right] - J_{k+1}(x_{k+1})\right), \quad (14)$$

is optimal and results in a zero duality gap. Moreover, they show that $h(u_F^*, w) + z(u_F^*, w) = \mathbb{E}_0\left[h(u_F^*, w)\right]$ *almost surely* with this choice of penalty. This will not be true of the gradient based penalty and the penalty of [7] that we discuss in Section 4.

Note that (14) clearly implies that if we know the optimal value function to within a constant then that is enough to obtain a lower bound with a zero duality gap. More generally, this observation suggests that a good approximation to the shape of the value function should often be sufficient for obtaining a good upper bound. In practice, we do not know J_k and therefore cannot compute the dual penalty of (14). Nonetheless if we have a good approximation, say $\tilde{J}_k$, to J_k then we could use

$$\tilde{z}(u, w) := \sum_{i=0}^{N-1} \left(\mathbb{E}_k\left[\tilde{J}_{k+1}(x_{k+1})\right] - \tilde{J}_{k+1}(x_{k+1}) \right) \quad (15)$$

as a dual feasible penalty and hope to still obtain a good lower bound. This program has been implemented successfully in practice in the context of American options and indeed in the examples of [3,2]. However, the dual penalty of (15) has a number of weaknesses that can result in (11) being difficult to solve in practice. These weaknesses include:

1. It is not always the case that an approximate value function, $\tilde{J}_k(\cdot)$, is readily available. Even if a good sub-optimal policy is available, it is often the case that the value function corresponding to that sub-optimal policy is unknown. While we could simulate the sub-optimal policy and use the resulting rewards to estimate the value function, this would require additional work and we would still have to overcome problems 2 and 3 below.
2. Even if we do have $\tilde{J}_{k+1}(\cdot)$ available to us for each k, we may not be able to compute $\mathbb{E}_k\left[\tilde{J}_{k+1}(x_{k+1})\right]$ analytically and so we cannot write $\tilde{z}(u, w)$ in (15) as an analytic function of the u_i's. This in turn makes it very difficult in general to solve the optimization problem in (11).

3. Even when we can compute $\mathbb{E}_k\left[\tilde{J}_{k+1}(x_{k+1})\right]$ analytically, it may be the case that the resulting penalty, $\tilde{z}(u, w)$, causes an otherwise easily solved deterministic optimization problem to become very difficult. For example, it may be the case that (11) is convex and easy to solve if we assume a zero penalty function but that convexity is lost if we construct $\tilde{z}(u, w)$ according to (14). One possible solution to this problem is to use an approximation to $\tilde{z}(u, w)$ that is linear or otherwise convex in u. This approach has been used successfully by [2] for solving portfolio optimization problems with transaction costs but there is no guarantee that it will work in general. In particular, if the desired penalty, $\tilde{z}(u, w)$, is very non-linear in u then the linearization approach may result in poor dual bounds.

These weaknesses are not to suggest that dual bounds based on penalties like (14) cannot work well in practice. Indeed as mentioned earlier, there are several applications where they have been used very successfully. We now describe gradient based penalties and we shall see that, when applicable, they can be a promising alternative to value-function based penalties in that they do not suffer from the potential weaknesses listed above.

2.3.1. *Constructing dual penalties using gradients*

[2] developed a *gradient* based dual penalty function for perfect information relaxations in the context of dynamic portfolio optimization problems with transaction costs. We will describe their gradient penalty for the more general dynamic program of (3). We define

$$z_g^*(u, w) := \nabla_u h(u^*(w))' \, (u^*(w) - u) \tag{16}$$

where $u^* = (u_0^*, \ldots, u_{N-1}^*)$ is the optimal control for the primal dynamic programming problem in (3) and $u = (u_0, \ldots, u_{N-1})$ is an arbitrary control policy. Note that we are therefore implicitly assuming that the total cost, $h(u, w)$, is differentiable in the controls, u. If we view the primal problem in (2) as an optimization problem with the entire *strategy*, u, as the decision variable, then assuming that the space of feasible strategies is convex, the first-order conditions for optimality are

$$\mathbb{E}_0\left[\nabla_u h(u^*(w))' \, (u^*(w) - u)\right] \leq 0 \tag{17}$$

which implies in particular that $z_g(u, w)$ is dual feasible. Moreover, [2] showed that when the cost function is convex the dual feasible penalty in (16) is indeed an optimal dual penalty. Note that with this choice of penalty the dual deterministic optimization problem in (11) has the form

$$\inf_{u_G \in \mathcal{U}_{\mathbb{G}}} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} \left(g(x_k, u_k) + \nabla_{u_k} h(u^*)'(u_k^*(w) - u_k) \right) \right\}. \tag{18}$$

Moreover the gradient penalty is linear in u and this suggests that the dual problem with this penalty should be no harder to solve than the deterministic version of the primal problem.

The difficulty with using (18) of course is that we do not know u^*, the optimal control policy. Indeed if we did know u^* then there would be no problem to solve. [2], however, recognized that under certain circumstances they could use $z_g(u, w)$ instead of $z_g^*(u, w)$ as their dual penalty where

$$z_g(u, w) := \nabla_u h(\tilde{u}(w))'(\tilde{u}(w) - u) \tag{19}$$

and where $\tilde{u}$ is the optimal solution to an alternative *approximate* problem. For example, in their dynamic portfolio optimization problem, [2] took $\tilde{u}$ to be the optimal control for the dynamic portfolio optimization problem *without* transactions costs. Because (i) $\tilde{u}$ is optimal for this alternative problem and (ii) the space of feasible controls for the alternative problem contains the space of feasible controls for the original problem they could still conclude that

$$\mathbb{E}_0\left[\nabla_u h(\tilde{u}(w))'(\tilde{u}(w) - u)\right] \leq 0 \tag{20}$$

for all $\mathcal{F}_k$-adapted trading strategies, so this alternative gradient penalty is also dual feasible. Moreover, intuition suggests that (19) should be similar to (16) in which case we would expect to obtain good dual bounds using (19). This was indeed the case for the problems and parameter values considered by [2].

One advantage of the gradient penalty in (18) over the value-function based penalty in (14) is that it does not require knowledge of the value function in order to be estimated; differentiability of the cost terms g and a candidate sub-optimal policy are all that we require. Additionally, the gradient penalty is linear in the controls, u, so the deterministic problem does not become harder with the introduction of the penalty. The disadvantage of the gradient approach is that it is not always applicable since it assumes that the cost function is differentiable in u. It also requires the space of feasible strategies to be convex. Finally, even if it is applicable we will see in Section 3 that the optimal gradient penalty, while dual optimal, does not result in a dual objective function that equals the primal objective function almost surely. Equality is only in expectation and this is in contrast to the case for the value-function based penalty in (14) as we discussed earlier.

In Section 3 we will compute the value-function and gradient based penalties in the context of linear–quadratic control problems. We will show that both approaches, as well as the approach of [7] in Section 4, yield penalties of the form $\sum_i \lambda_i^T u_i + z(\omega)$ where $\mathrm{E}[z(\omega)] = 0$. We will see that all three approaches yield identical λ_i's but different $z(\omega)$'s. The fact that they lead to different $z(\omega)$'s suggests that the Monte Carlo estimators of the dual bounds based on each of the three dual penalties will have different variances. The "almost sure" property of the optimal value-function based penalty, however, implies that the corresponding Monte Carlo estimator will have zero variance and therefore, from a theoretical perspective, be the best of the three bounds in terms of the computational workload. In practice, the optimal value function will not be available and so we can only conjecture that good (approximate) value-function based dual bounds will have lower variances than dual bounds based on gradient penalties. We will return again to this issue in Section 3.

3. Finite horizon linear–quadratic control problems

We now apply the duality ideas of Section 2 to constrained LQ problems. We first review the finite horizon, discrete-time LQ problem as formulated, for example, in [1]. We consider the complete information version only as it is well known that the incomplete information version can be reduced to the complete information case. Let x_k denote the n-dimensional state vector at time k. We assume that it has dynamics that satisfy

$$x_{k+1} = Ax_k + Bu_k + w_{k+1}, \quad k = 0, 1, \ldots, N-1 \tag{21}$$

where u_k is an m-dimensional vector of control variables and the w_k's are n-dimensional independent vectors of zero-mean disturbances with finite second moments. It will be useful later to observe that (21) implies

$$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^{k-1-i} \left(Bu_i + w_{i+1}\right) \quad \text{for } k = 0, \ldots, N. \tag{22}$$

As before we let $\mathcal{F}_k$ denote the filtration generated by the w_k's. The objective then is to choose $\mathcal{F}_k$-adapted controls, u_k, to minimize

$$\mathrm{E}_{\mathcal{F}_0}\left[x_N' Q x_N + \sum_{k=0}^{N-1} \left(x_k' Q x_k + u_k' R u_k \right) \right]$$

where Q and R are positive semi-definite and positive definite, respectively. The optimal solution is easily seen [1] to satisfy

$$u_k^*(x_k) = L_k x_k$$

where

$$L_k := -\left(B'K_{k+1}B + R\right)^{-1} B'K_{k+1}A \tag{23}$$

and where the symmetric positive semi-definite matrices, K_k, are given recursively by the algorithm $K_N = Q$ and

$$K_k := A'\left(K_{k+1} - K_{k+1}B\left(B'K_{k+1}B + R\right)^{-1} B'K_{k+1}\right)A + Q, \quad k = N-1, \ldots, 0. \tag{24}$$

The value function then satisfies

$$\begin{aligned} J_k(x_k) &= x_k'K_kx_k + \sum_{i=k}^{N-1} \mathrm{E}\left[w_{i+1}'K_{i+1}w_{i+1}\right] \\ &= x_k'K_kx_k + \sum_{i=k}^{N-1} \quad \mathrm{trace}\ (K_{i+1}\Sigma_{i+1}) \end{aligned} \tag{25}$$

where $\Sigma_i := \mathrm{Cov}(w_i)$.

3.1. The dual penalty constructed from the optimal value function

We can use the unconstrained optimal value function to construct a dual bound. In particular let $\mathcal{G}_k$ be the perfect information relaxation of $\mathcal{F}_k$. Then we can take $z_k(u_k)$ as our dual penalty where

$$\begin{aligned} z_k(u_k) &:= \mathrm{E}\left[J_{k+1}(x_{k+1}) \mid \mathcal{F}_k\right] - \mathrm{E}\left[J_{k+1}(x_{k+1}) \mid \mathcal{G}_k\right] \\ &= \mathrm{E}\left[(Ax_k + Bu_k + w_{k+1})' K_{k+1}\big(Ax_k + Bu_k + w_{k+1}\big) \mid \mathcal{F}_k\right] + \sum_{i=k+1}^{N-1} \quad \mathrm{trace}\ (K_{i+1}\Sigma_{i+1}) \\ &\quad - \ (Ax_k + Bu_k + w_{k+1})' K_{k+1}\ (Ax_k + Bu_k + w_{k+1}) \\ &\quad - \sum_{i=k+1}^{N-1} \quad \mathrm{trace}\ (K_{i+1}\Sigma_{i+1}) \\ &= (Ax_k + Bu_k)' K_{k+1}\ (Ax_k + Bu_k) + \mathrm{trace}\ (K_{k+1}\Sigma_{k+1}) \\ &\quad - \ (Ax_k + Bu_k + w_{k+1})' K_{k+1}\ (Ax_k + Bu_k + w_{k+1}) \\ &= -w_{k+1}'K_{k+1}\ (Ax_k + Bu_k) - w_{k+1}'K_{k+1}w_{k+1} \\ &\quad - \ (Ax_k + Bu_k)' K_{k+1}w_{k+1} + \mathrm{trace}\ (K_{k+1}\Sigma_{k+1}) \\ &= -2\ (Ax_k + Bu_k)' K_{k+1}w_{k+1} - w_{k+1}'K_{k+1}w_{k+1} \\ &\quad + \mathrm{trace}\ (K_{k+1}\Sigma_{k+1})\,. \end{aligned} \tag{26, 27}$$

We know of course from the results of [3] that (27) is an optimal dual penalty for the unconstrained problem. In fact the dual objective (6) with this choice of penalty is given by

$$J_0(x_0) = \mathrm{E}_{\mathcal{F}_0}\left[\inf_{u_k}\left(x_N'Qx_N + \sum_{k=0}^{N-1}\left(x_k'Qx_k + u_k'Ru_k + 2\ (Ax_k + Bu_k)' K_{k+1}w_{k+1}\right)\right)\right] \tag{28}$$

subject to the dynamics of (21). Note that the optimization problem inside the expectation in (28) is a standard *deterministic* LQ problem and is easily solved using the usual techniques. It is easy to confirm by direct computation that the optimal control, u_k^*, in (28) is indeed non-anticipative.

Note also that the last two terms in (27) do not appear in (28) since their sum has expectation zero. While perhaps obvious, this observation emphasizes the non-uniqueness of the optimal dual penalty. Indeed let v_k be any random variable with zero expectation that does not depend on the controls or state variables. Then if $z_k(u_k)$ is an optimal dual penalty, so too is $z_k(u_k) + v_k$.

Note also that an optimal penalty is as good as any other optimal dual penalty in so far as their dual problems result in equal (and optimal) value functions as well as identifying the optimal non-anticipative control. However, the optimal dual penalty of [3] is such that any instance of the dual problem is guaranteed to equal the optimal value function *almost surely* and not just in expectation. This is not true in general of other optimal dual penalties and suggests that some (optimal) dual penalties will outperform other optimal dual penalties when Monte Carlo techniques are required to estimate the outer expectation in (10). We believe that similar observations should apply when we cannot compute the optimal dual solution but can only estimate it using sub-optimal penalty functions that are *sufficiently close* to the optimal value-function based penalty. It is true, however, that we cannot tell what "sufficiently close" means in practice. Moreover, given an approximate value-function based penalty and a gradient penalty, we cannot tell a priori which of the corresponding dual bounds will be superior and which of the two dual estimators will have a lower variance.

An alternative representation for the value-function dual penalty

Since the dual problem is deterministic, we do not need to explicitly associate $z_k(u)$ in (27) with time period k. In particular, it is the total *sum* of the dual penalties that is relevant and we now determine this sum as a function of the u_k's. This representation of the unconstrained optimal dual penalty will be useful in Section 4. Let P_{vf} denote the total penalty and let $C := \sum_{k=0}^{N-1}(\mathrm{trace}\ (K_{k+1}\Sigma_{k+1}) - w_{k+1}'K_{k+1}w_{k+1})$. Note that C has no bearing on the optimal control in any instance of the dual problem. We see that P_{vf} then satisfies

$$\begin{aligned} P_{vf} &= C - 2\sum_{k=0}^{N-1} u_k'B'K_{k+1}w_{k+1} - 2\sum_{k=0}^{N-1} x_k'A'K_{k+1}w_{k+1} \\ &= C - 2x_0'A'K_1w_1 - 2\sum_{k=0}^{N-1} u_k'B'K_{k+1}w_{k+1} \\ &\quad - 2\sum_{k=1}^{N-1}\left(A^kx_0 + \sum_{i=0}^{k-1} A^{k-1-i}\ (Bu_i + w_{i+1})\right)' A'K_{k+1}w_{k+1} \\ &= C_{vf} - 2\sum_{k=0}^{N-1} u_k'B'K_{k+1}w_{k+1} \\ &\quad - 2\sum_{k=1}^{N-1}\left(\sum_{i=0}^{k-1} A^{k-i}Bu_i\right)' K_{k+1}w_{k+1} \\ &= C_{vf} - 2\sum_{k=0}^{N-1} u_k'B'K_{k+1}w_{k+1} \\ &\quad - 2\sum_{i=0}^{N-2} u_i'B'\left(\sum_{k=i+1}^{N-1} \left(A^{k-i}\right)' K_{k+1}w_{k+1}\right) \\ &= C_{vf} - 2\sum_{k=0}^{N-1} u_k'B'K_{k+1}w_{k+1} \\ &\quad - 2\sum_{i=0}^{N-2} u_i'B'\left(\sum_{k=i+1}^{N-1} \left(A^{k-i}\right)' K_{k+1}w_{k+1}\right) \\ &= C_{vf} - 2\sum_{i=0}^{N-1} u_i'B'\left(\sum_{k=i}^{N-1} \left(A^{k-i}\right)' K_{k+1}w_{k+1}\right) \end{aligned} \tag{29}$$

where

$$C_{vf} := C - 2x_0' \sum_{k=0}^{N-1} (A^{k+1})' K_{k+1} w_{k+1} - 2 \sum_{k=1}^{N-1} \left(\sum_{i=0}^{k-1} A^{k-i} w_{i+1} \right)' K_{k+1} w_{k+1} \tag{30}$$

is a zero-mean term that does not depend on the u_k's. (We use the notation C_{vf} to emphasize that this term is the *constant* component of the *value-function* based dual penalty. In particular, C_{vf} does not depend on the u_i's.) The salient feature of (29) is that we have an explicit expression for the coefficient of u_i in the optimal dual penalty for the unconstrained LQ problem.

3.2. The gradient dual penalty

The gradient based optimal dual penalty is also straightforward to calculate. First, we define

$$V_0 := \sum_{i=0}^{N-1} u_i' R u_i + \sum_{i=0}^{N} x_i' Q x_i$$

which of course is the realized cost for the LQ control problem. We may then define

$$z_g(u) := \nabla_u V_0(u^*)'(u^* - u)$$

where $u^* = (u_0^*, \ldots, u_{N-1}^*)$ is the optimal control for the unconstrained problem and $u = (u_0, \ldots, u_{N-1})$ is an arbitrary control policy. By viewing the LQ problem as a convex optimization problem where the strategy, $u = (u_0, \ldots, u_{N-1})$, is the decision vector (more precisely, $u_i(x_i)$ is a decision variable for each state x_i and all $i = 0, \ldots, N-1$), we see that the first-order optimality conditions are

$$\mathrm{E}_0 \left[\nabla_u V_0(u^*)'(u^* - u) \right] \le 0. \tag{31}$$

But (31) then implies that $z_g(u)$ is dual feasible and indeed it is easy to see that $z_g(u)$ is a dual optimal penalty for the unconstrained LQ control problem. In this case we know that $u_i^* = L_i x_i^*$ where we use x_i^* to denote the trajectory of the state vector under u^*. We then see that

$$z_g(u) = \sum_{i=0}^{N-1} \nabla_{u_i} V_0(u^*)'(u_i^* - u_i)$$

where the dynamics in (21) imply

$$\begin{aligned} \nabla_{u_i} V_0(u^*)'(u_i^* - u_i) &= 2 \left[R u_i^* + B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q x_k^* \right]' (u_i^* - u_i) \\ &= 2 \left[R u_i^* + B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q x_k^* \right]' (L_i x_i^* - u_i). \end{aligned} \tag{32}$$

We can iterate $x_k^* = (A + BL_{k-1}) x_{k-1}^* + w_k$ to obtain

$$\begin{aligned} x_1^* &= (A + BL_0) x_0^* + w_1 \\ x_k^* &= \left(\prod_{j=0}^{k-1} (A + BL_j) \right) x_0 + \sum_{j=1}^{k-1} \left(\prod_{l=j}^{k-1} (A + BL_l) \right) w_j + w_k, \quad k \ge 2 \end{aligned} \tag{33}$$

and then substitute (33) into (32) to obtain an explicit expression for the gradient penalty that is linear in the u_i's. Before doing this, we have the following lemma which we will use to simplify (32).

Lemma 2. *For $i = 0, \ldots, N$ we have*

$$K_N = Q,$$

$$K_i = \sum_{j=i+1}^{N} \left(A^{j-i} \right)' Q \left(\prod_{k=i}^{j-1} (A + BL_k) \right) + Q, \quad i \le N-1 \tag{34}$$

where L_k is given by (23).

Proof. First note that when $i = N$ (34) reduces to $K_N = Q$, and when $i = N-1$

$$K_{N-1} = A'Q(A + BL_{N-1}) + Q,$$

both of which are true. Suppose now that (34) is true for $i+1$ for some $i \le N-2$. If we can show that (34) is then true for i we are done. Towards this end note that

$$K_i = A' K_{i+1} (A + BL_i) + Q \tag{35}$$

$$= A' \left[\sum_{j=i+2}^{N} \left(A^{j-(i+1)} \right)' Q \left(\prod_{k=i+1}^{j-1} (A + BL_k) \right) + Q \right] \times (A + BL_i) + Q \tag{36}$$

$$= \sum_{j=i+1}^{N} \left(A^{j-i} \right)' Q \left(\prod_{k=i}^{j-1} (A + BL_k) \right) + Q$$

where (35) follows from (23) and (24) and (36) follows from the assumption that (34) holds for $i+1$. □

We are now in a position to compare the two penalties. In particular we see that the coefficient of u_i' in each of the two penalties is given by

$$\mathrm{Coeff}_{vf}(u_i) = -2B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' \times K_k w_k \quad \text{(Value-Function Penalty)} \tag{37}$$

$$\mathrm{Coeff}_g(u_i) = -2RL_i x_i^* - 2B' \sum_{k=i+1}^{N} \times \left(A^{k-1-i} \right)' Q x_k^* \quad \text{(Gradient Penalty)}. \tag{38}$$

Note that (38) follows from (32) with $L_i x_i^*$ substituted for u_i^* and that (37) follows from (29). (We have modified the indexing in (37) so that each summation in (37) and (38) runs from $i+1$ to N.) The following lemma establishes directly that the two coefficients are identical.

Lemma 3. $\mathrm{Coeff}_{vf}(u_i) = \mathrm{Coeff}_g(u_i)$ *for $i = 0, \ldots, N-1$.*

Proof. First note that (33) can be restated more generally as

$$\begin{aligned} x_{i+1}^* &= (A + BL_i) x_i^* + w_{i+1}, \\ x_k^* &= \left(\prod_{j=i}^{k-1} (A + BL_j) \right) x_i^* + \sum_{j=i+1}^{k-1} \left(\prod_{l=j}^{k-1} (A + BL_l) \right) w_j + w_k, \\ &\quad \text{for } k \ge i+2. \end{aligned} \tag{39}$$

We can then use (39) to substitute for x_k^* in (38) to obtain

$$-\frac{1}{2} \mathrm{Coeff}_g(u_i) = \left[RL_i + B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q \left(\prod_{j=i}^{k-1} (A + BL_j) \right) \right] x_i^* + B' \left\{ \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q w_k + \sum_{k=i+2}^{N} \left(A^{k-1-i} \right)' Q \right.$$

$$\times \sum_{j=i+1}^{k-1} \left(\prod_{l=j}^{k-1} (A + B L_l) \right) w_j \Bigg\}$$

$$= \left[R L_i + B' K_{i+1} (A + B L_i) \right] x_i^* \quad \text{by (34)}$$

$$+ B' \Bigg\{ \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q w_k + \sum_{k=i+2}^{N} \sum_{j=i+1}^{k-1} \left(A^{k-1-i} \right)' Q$$

$$\times \left(\prod_{l=j}^{k-1} \left(A + B L_l \right) \right) w_j \Bigg\}$$

$$= B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q w_k$$

$$+ B' \sum_{j=i+1}^{N} \left[\sum_{k=j+1}^{N} \left(A^{k-1-i} \right)' Q \left(\prod_{l=j}^{k-1} (A + B L_l) \right) \right] w_j \tag{40}$$

$$= B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q w_k + B' \sum_{j=i+1}^{N} \left(A^{j-1-i} \right)'$$

$$\times \left[\sum_{k=j+1}^{N} \left(A^{k-j} \right)' Q \left(\prod_{l=j}^{k-1} (A + B L_l) \right) \right] w_j$$

$$= B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q w_k + B' \sum_{j=i+1}^{N} \left(A^{j-1-i} \right)'$$

$$\times \left(K_j - Q \right) w_j \quad \text{(by (34) again)}$$

$$= B' \sum_{j=i}^{N-1} \left(A^{j-i} \right)' K_{j+1} w_{j+1} \tag{41}$$

$$= -\frac{1}{2} \quad \text{Coeff}_{vf}(u_i)$$

where (40) follows by changing the order of the double summation and noting that $RL_i + B'K_{i+1}(A + B L_i) = 0$ by the definition of L_i in (23). □

Of course Lemma 3 is not at all surprising since the value-function and gradient penalties are both dual optimal. Indeed the more interesting question is that of whether or not the constant terms in each of the two penalties are equal. If we use C_g to denote the constant component of the gradient penalty, then using (32) and summing over i we see that is given by

$$C_g = 2 \sum_{i=0}^{N-1} \left[Ru_i^* + B' \sum_{k=i+1}^{N} \left(A^{k-1-i} \right)' Q x_k^* \right]' L_i x_i^* \tag{42}$$

$$= 2 \sum_{i=0}^{N-1} \left[B' \sum_{j=i}^{N-1} \left(A^{j-i} \right)' K_{j+1} w_{j+1} \right] L_i x_i^*. \tag{43}$$

Note that we have used (41) to substitute for the term inside the square brackets in (42). We recall that the corresponding constant component of the value-function based penalty, C_{vf}, is given by (30). It is clear that C_{vf} and C_g are different. For example, C_{vf} contains the term $\sum_{k=0}^{N-1}$ trace $(K_{k+1} \Sigma_{k+1})$ and no such term appears in C_g. This observation demonstrates that in general the value-function based penalty and the gradient penalty do not coincide when the latter is actually defined.

4. The Davis and Zervos approach

While we have derived the optimal value-function and gradient penalties in Section 3 using the recent results of [3,2], it turns out that these penalties are very closely related to the work of [7] which we will now describe. [7] consider the three classic cases of discrete-time LQ problems: the deterministic, stochastic full information and stochastic partial information cases. They show that the solution to the deterministic problem can be used to solve the two stochastic versions of the problem after including appropriate Lagrange multiplier terms in the objective function. We will show below that the Lagrange multiplier terms of [7] in the stochastic full information case also constitute a dual optimal penalty. (We will not consider the stochastic partial information case as the ideas are identical and of course, it is well known that the partial information case can be reduced to the full information case by expanding the state space.) Indeed, the only differences between the [7] penalty and our two earlier penalties are terms that have zero expectation that do not depend on the u_i's. In particular, [6] prove the following theorem, which is a combination of Theorem 2 in [7] and the analysis that they provide in "Case 2" immediately following their Theorem 2. ([7] included a cross-term of the form $x_k' T u_k$ in their objective function but we will omit this term without any loss of generality so that we can compare their penalty with our penalty in (29). They also assume that $A_k = A$, $B_k = B$, $Q_k = Q$ and $R_k = R$ for all k and we will maintain this assumption in this subsection, again without loss of generality.)

Theorem 3. *Consider the linear system model*

$$x_{i+1} = Ax_i + Bu_i + w_{i+1} \quad i = 0, \ldots, N-1 \tag{44}$$

where $w = (w_1, \ldots, w_N)$ is a sequence of independent zero-mean random vectors and $u = (u_0, \ldots, u_{N-1})$ is the control sequence. Let

$$J(u, \lambda) = \mathrm{E}_{\mathcal{F}_0} \left[x_N' Q x_N + \sum_{i=0}^{N-1} \left(x_i' Q x_i + u_i' R u_i + 2\lambda_i' u_i \right) \right] \tag{45}$$

be the cost associated with the pair (u, λ) and let

$$J_d(u, w, \lambda) = x_N' Q x_N + \sum_{i=0}^{N-1} \left(x_i' Q x_i + u_i' R u_i + 2\lambda_i' u_i \right) \tag{46}$$

be the cost associated with (u, w, λ), so $J(u, \lambda) = \mathrm{E}[J_d(u, w, \lambda)]$. Assume the matrices Q and R are symmetric positive semi-definite and symmetric positive definite, respectively, and $\lambda = (\lambda_0, \ldots, \lambda_{N-1})$ is a given sequence of vectors. Suppose λ is chosen such that

$$\lambda_i = -B' K_{i+1} w_{i+1} - B' \beta_{i+1}, \quad \textit{for } i = 0, \ldots, N-1 \tag{47}$$

where K_{i+1} satisfies (24) and where β_i satisfies

$$\beta_i = A' \beta_{i+1} + A' K_{i+1} w_{i+1}, \quad \beta_N = 0. \tag{48}$$

Then: (i) *$u_i^*(x_i) = L_i x_i$ where L_i is given by (23) is the optimal non-anticipative control vector that minimizes (45);* (ii) *$u_i^*(x_i) = L_i x_i$ is also the optimal control that minimizes the deterministic objective function of (46) where w is known in advance. Moreover, this choice of λ is almost surely the unique one for which the minimizer of $J_d(u, w, \lambda)$ is non-anticipative and for which the Lagrange multiplier terms in (45) disappear.*

In order to compare Theorem 3 with our earlier results, we need to compare the penalty terms in the three approaches. But first note that [7] do not include a constant term like C_{vf} or C_g, though it is possible they simply omitted it since it has no bearing on the optimal controls, and so we can immediately conclude that the [7] penalty is different to the two earlier penalties. The following lemma shows, however, that the coefficient of u_i in the penalty term in (45), i.e. $2\lambda_i$, is equal to $\text{Coeff}_{vf}(u_i)$ as given in (37).

Lemma 4. *For $i = 0, \ldots, N-1$, we have*

$$\lambda_i = -B' \sum_{k=i}^{N-1} \left(A^{k-i}\right)' K_{k+1} w_{k+1}$$

so the coefficient of u_i in (29) is equal to the coefficient of u_i in (45). In particular the Lagrangian terms of [7] in (45) are also dual optimal in the framework of BSS.

Proof. First note that we can iterate (48) to obtain

$$\beta_{i+1} = \sum_{k=i+1}^{N-1} \left(A^{k-i}\right)' K_{k+1} w_{k+1}. \tag{49}$$

We can then substitute (49) into (47) to obtain

$$\begin{aligned} \lambda_i &= -B' \left(K_{i+1} w_{i+1} + \beta_{i+1}\right) \\ &= -B' \left(K_{i+1} w_{i+1} + \sum_{k=i+1}^{N-1} \left(A^{k-i}\right)' K_{k+1} w_{k+1}\right) \\ &= -B' \sum_{k=i}^{N-1} \left(A^{k-i}\right)' K_{k+1} w_{k+1} \end{aligned}$$

as desired. □

Note that while the Lagrangian terms of [7] are dual optimal in the framework of BSS, they do not result in zero-variance dual bounds. This was also the case with the gradient based penalty, and as suggested earlier, this observation suggests that penalties based on value-function approximations may be more efficient than other penalties when dual bounds need to be computed using Monte Carlo methods.

When we consider the results of this section and the earlier developments in the optimal stopping literature that we mentioned in Section 1, it becomes clear then that many of the ideas behind the information relaxation duality theory of [3,10] have been around for some time and in particular, since the work of [6,7]. (It should also be mentioned that the idea of relaxing the non-anticipativity constraints has been well known in the stochastic programming literature.) This is not to say, however, that [3,10] are somehow redundant. On the contrary, they have unified these ideas in a discrete-time framework and demonstrated that the dual problem can be used successfully for evaluating sub-optimal strategies when it is not practically feasible to construct optimal policies. This of course parallels the earlier literature on optimal stopping problems. The results of [3] also apply to information relaxations that are more general than the perfect information relaxation. Moreover, their optimal dual penalty in the case of the perfect information relaxation is a *zero-variance* penalty which may be particularly useful when evaluating sub-optimal strategies via Monte Carlo methods.

5. Conclusions and further research

There are several directions for future research that are particularly interesting. First, we would like to consider constrained LQ problems and compare the dual bounds corresponding to each of the three penalties. Of course these penalties are only optimal for unconstrained LQ problems and they may not produce good dual bounds when the constraints are frequently binding. When that is the case, it would be necessary to construct other dual feasible penalties, possibly using good sub-optimal policies for the constrained problem. This has already been done for optimal stopping problems and other problems. See for example [3,8,9] among others.

A particularly interesting direction for future research is in comparing the efficiency of value-function based penalties with gradient penalties. We know from [3] that the former are almost surely optimal when the optimal value function is used. Of course the optimal value function is never available in practice and so *approximate* value functions must be used. The question then arises as to whether penalties constructed using approximate value functions are more efficient or have a lower variance than corresponding gradient penalties. Finally, variance reduction methods should be of considerable use when computing dual bounds. For example, the optimal value function of the unconstrained problem (when it is available analytically) should be a good control variate and indeed such a control variate was used by Brown and Smith [2]. More generally, however, the dual instances of these problems can often be very computationally demanding and constructing good variance reduction methods should be of considerable value.

Acknowledgments

This research was supported by the NSF CAREER Award DMI-0348746 and the NSF Grants CMMI-0500503, CMMI-1031637 and CMMI-1201085. The opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Support from the Coleman Fung Chair in Financial Modeling is also acknowledged.

References

[1] D.P. Bertsekas, Dynamic Programming and Optimal Control: Volume One, Athena Scientific, Belmont, Massachusetts, 2000.
[2] D.B. Brown, J.E. Smith, Dynamic portfolio optimization with transaction costs: heuristics and dual bounds, Working Paper, Duke University, 2010.
[3] D.B. Brown, J.E. Smith, P. Sun, Information relaxations and duality in stochastic dynamic programs, Operations Research 58 (4) (2010) 785–801.
[4] M.H.A. Davis, Anticipative LQG control, IMA Journal of Mathematical Control and Information 6 (1989) 259–265.
[5] M.H.A. Davis, Anticipative LQG control II, in: Applied Stochastic Analysis, 1991, pp. 205–214.
[6] M.H.A. Davis, I. Karatzas, A deterministic approach to optimal stopping, in: F. Kelly (Ed.), Probability, Statistics and Optimization: A Tribute to Peter Whittle, J. Wiley and Sons, New York, Chichester, 1994, pp. 455–466.
[7] M.H.A. Davis, M. Zervos, A new proof of the discrete-time LQG optimal control theorems, IEEE Transactions on Automatic Control 40 (8) (1995) 1450–1453.
[8] M.B. Haugh, L. Kogan, Pricing American options: a duality approach, Operations Research 52 (2) (2004) 258–270.
[9] L.C.G. Rogers, Monte-Carlo valuation of American options, Mathematical Finance 12 (2002) 271–286.
[10] L.C.G. Rogers, Pathwise stochastic optimal control, SIAM Journal on Control and Optimization 46 (3) (2007) 1116–1132.