Systematic Perturbations of Discrete-Time Stochastic Dynamical Systems [1]

Daniel L. Kern and Floyd B. Hanson
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
851 S. Morgan Street.; M/C 249
Chicago. IL 60607-7045
e-mail: hanson@math.uic.edu or dkern@math.uic.edu

Abstract

The discrete-time stochastic optimal control problem is approximated by a variation of differential dynamic programming with systematic calculations of the perturbations due to small stochastic noise. This problem is related to the dual control aspects of stochastic optimal control problems. The motivation is to correct prior calculations for missing terms and to examine the foundations of the method. The state vector is properly expanded asymptotically, in addition to the control vector, in contrast to previous solutions. Corrections are given for the small noise expansions of the solution.

1. Introduction

Conventional discrete dynamic programming [5] is still commonly used, but aside from a few specific problem forms, Bellman's curse of dimensionality in the state space limits its use in numerical methods to those with a relatively small number of state and control variables, though the use of high performance computing permits the treatment of larger state spaces [3]. The need to search the entire state space in dynamic programming, both for deterministic and stochastic problems, can lead to large scale computational demands. Thus the method has limited usefulness to applications such as reservoir management [6, 7], groundwater quality remediation [2, 4], and others.

An alternate method, used by Kitanidis et al. [6, 7], for approximate solutions to the optimal control and cost-to-go performance index utilizes a stochastic perturbation of differential dynamic programming (DDP) [5, 3] to find an analytic solution of both the deterministic and caution (i.e., hedging or stochastic correction) terms. These terms are related to the interaction of estimation and control as found in dual control concepts [1]. This stochastically perturbed DDP method gives a formally closed form expansion of the solution and has the added advantage in dimensional computational complexity over stochastic dynamic programming. The decrease in the computational complexity is because for small noise only a small neighborhood of the deterministic trajectory needs to be searched rather than the whole state space. The caution term represents the error between the deterministic feedback control and the stochastic optimal control [6]. Care must be used in finding the caution term, and here the relative size of the caution term is not assumed before calculations.

The stochastic optimal control problem in discrete time is formulated and the reduced equations for expansions of the regular control up to quadratic order in the stochastic parameter are found in Section 2. Unlike the previous derivation in [6], both the control and the state vectors are expanded asymptotically for consistency. In addition, the expectation operator is applied to the entire cost-to-go for that stage, instead of only the cost-to-go found from the previous stage of the backwards sweep.

In order to find a caution term of the same type as Kitanidis [6], the state vector in Section 3 is not expanded until after calculations for each stage in the DDP backwards sweep. This means that the control is not the same as the consistent form found in Section 2, but second order control terms missing from [6] are included.

An Appendix gives a scalar counter-example to show that switching the order of the expectation and minimum operators is generally incorrect.

2. Control Problem and Stochastic Perturbation

The general discrete-time stochastic optimal control problem has the cost objective

$$[\mathrm{J}] = E_w\left[f_N(\vec{x}_N) + \sum_{k=1}^{N-1} c_k(\vec{x}_k, \vec{u}_{k+1})\right], \tag{1}$$

where f_N is the specified final cost and c_k is the kth stage cost for $k = 1, \ldots, N-1$. The expectation operator, $E_w = E_{\{w_k, k=2 \text{ to } N\}} = \Pi_{k=2}^{N} E_{w_k}$, denotes the expectation over independent component, discrete noise $\vec{w}_k$, and thus ensuring separability of E_w over each stage k. The goal is to minimize the cost objective subject to the linear state transition equation

$$\vec{x}_{k+1} = \Phi_k \vec{x}_k + \Psi_k \vec{u}_{k+1} + \vec{\mu}_{k+1} + \vec{w}_{k+1}, \tag{2}$$

[1] Work supported in part by the National Science Foundation Grant DMS-96-26692. This paper will appear in the *Proceedings of 37th IEEE Conference on Decision and Control*, 6 pages, December 1998

from [6], where $k = 1, \ldots, N-1$. The Φ_k and Ψ_k are known state and control coefficient matrices, respectively; the state of the system $\vec{x}_k$ is a $n \times 1$ dimension vector, the control $\vec{u}_k$ is a $m \times 1$ dimension vector, $\vec{\mu}_k$ is a known input vector of dimension $n \times 1$, and $\vec{w}_k$ is a stochastic noise vector of dimension $n \times 1$, such that

$$E_w[\vec{w}_k] = \vec{0}, \;\; E_w\left[\vec{w}_k \vec{w}_l^T\right] = Q_k \delta_{k,l}. \tag{3}$$

for $k = 1, \ldots, N$. The control constraint for the kth stage is assumed to be given component-wise as $u_{\min,k,i} \leq u_{k,i} \leq u_{\max,k,i}$ for $i = 1$ to n and $k = 2$ to N.

The optimal, expected total cost is

$$\min[\mathrm{J}] = \min_u \left[E_w \left[f_N(\vec{x}_N) + \sum_{k=1}^{N-1} c_k(\vec{x}_k, \vec{u}_{k+1}) \right] \right], \tag{4}$$

where $\min_u = \Pi_{i=2}^{N} \min_{u_i}$, satisfying the separability property of the minimization operator needed to apply the deterministic dynamic programming method, while separability of the expectation operator over the stages permits extension to stochastic dynamic programming. The final cost condition is denoted by

$$\mathrm{J}_N = f_N(\vec{x}_N) = \mathrm{J}_N^{\star}(\vec{x}_N),$$

where the "$\star$" denotes the optimal value.

Since dynamic programming proceeds backward through the stages from the known final condition, stage $(N-1)$ is examined first. Using the Principle of Optimality and substituting for $\vec{x}_N$ from the state transition equation gives the decomposition recursion

$$\begin{aligned} \mathrm{J}_{N-1} &= \mathrm{J}_N^{\star}(\vec{x}_N) + c_{N-1}(\vec{x}_{N-1}, \vec{u}_N) \\ &= \mathrm{J}_N^{\star}(\Phi_{N-1}\vec{x}_{N-1} + \Psi_{N-1}\vec{u}_N + \vec{\mu}_N + \vec{w}_N) \\ &\quad + c_{N-1}(\vec{x}_{N-1}, \vec{u}_N). \end{aligned} \tag{5}$$

The cost-to-go function for stage $N-1$ is

$$\begin{aligned} f_{N-1}(\vec{x}_{N-1}) &= \mathrm{J}_{N-1}^{\star}(\vec{x}_{N-1}) \\ &= \min_{u_N} [E_{w_N}[f_N(\vec{x}_N) + c_{N-1}(\vec{x}_{N-1}, \vec{u}_N)]]. \end{aligned} \tag{6}$$

By induction the cost-to-go function for stage $(k-1)$ is

$$\begin{aligned} f_{k-1}(\vec{x}_{k-1}) &= \min_{u_k} \left[E_{w_k}[f_k(\vec{x}_k) + c_{k-1}(\vec{x}_{k-1}, \vec{u}_k)] \right]. \\ &= \min_{u_k} \left[E_{w_k}[f_k(\Phi_{k-1}\vec{x}_{k-1} + \Psi_{k-1}\vec{u}_k \right. \\ &\quad \left. + \vec{\mu}_k + \vec{w}_k) + c_{k-1}(\vec{x}_{k-1}, \vec{u}_k)] \right]. \end{aligned} \tag{7}$$

Therefore, the cost-to-go at stage $(k-1)$ has to be minimized given that the cost-to-go at the kth stage is already computed.

Since the control affects the state at the kth stage according to the dynamics (2), *both* $\vec{u}_k$ and $\vec{x}_k$ are expanded to order σ^2, where σ is the stochastic covariance scaling factor, such that $\sigma^2 = \mathrm{Trace}[Q_k]$ with $0 < \sigma \ll 1$ for small stochastic noise, $\vec{w}_k = \sigma \vec{w}_k^{(1)}$, and

$$\vec{u}_k = \vec{u}_k^{(0)} + \sigma \vec{u}_k^{(1)} + \sigma^2 \vec{u}_k^{(2)} + O(\sigma^3), \tag{8}$$

$$\vec{x}_k = \vec{x}_k^{(0)} + \sigma \vec{x}_k^{(1)} + \sigma^2 \vec{x}_k^{(2)} + O(\sigma^3). \tag{9}$$

Substituting the above expansions into the transition equation (2) implies the coefficient relations

$$\vec{x}_k^{(0)} = \Phi_{k-1}\vec{x}_{k-1}^{(0)} + \Psi_{k-1}\vec{u}_k^{(0)} + \vec{\mu}_k, \tag{10}$$

$$\vec{x}_k^{(1)} = \Phi_{k-1}\vec{x}_{k-1}^{(1)} + \Psi_{k-1}\vec{u}_k^{(1)} + \vec{w}_k^{(1)}, \tag{11}$$

$$\vec{x}_k^{(2)} = \Phi_{k-1}\vec{x}_{k-1}^{(2)} + \Psi_{k-1}\vec{u}_k^{(2)}, \tag{12}$$

upon matching terms of $\mathrm{ord}(\sigma^0)$, $\mathrm{ord}(\sigma^1)$, and $\mathrm{ord}(\sigma^2)$, respectively. Here, $F(\sigma) = \mathrm{ord}(G(\sigma))$ means that $F(\sigma)$ is the *same order* as $G(\sigma)$ or that $F(\sigma) = O((G(\sigma))$, but $F(\sigma) \neq o((G(\sigma))$, as $\sigma \to 0$.

In order to properly find the minimum specified in (7), a special commutativity of the expectation and gradient, a result with proof attributed to J.M.C.Clark in [5], is needed:

$$\nabla_{u_k}[E_{w_k}[V(\vec{x}, \vec{u}_k, \vec{w}_k)]] = E_{w_k}[\nabla_{u_k}[V(\vec{x}, \vec{u}_k, \vec{w}_k)]],$$

cast in the vector-gradient notation here, under the fairly general conditions with state as a parameter:

$$\begin{aligned} E_{w_k}[V(\vec{x}, \vec{u}_k, \vec{w}_k)] &< \infty \text{ at } \vec{u}_k = \vec{a}, \\ E_{w_k}\left[\int_{\vec{a}}^{\vec{b}} |V(\vec{x}, \vec{u}_k, \vec{w}_k)| d\vec{u}_k \right] &< \infty, \end{aligned}$$

excluding zero or infinite singular conditions. This is not the same as $\min_{u_k}[E_{w_k}[V(\vec{x}_{k-1}, \vec{u}_k, \vec{w}_k)]] = E_{w_k}[\min_{u_k}[V(\vec{x}_{k-1}, \vec{u}_k, \vec{w}_k)]]$, which is generally *not true* and a scalar counter-example is given in the Appendix.

Using the commutativity of gradient and expectation operators to compute the regular or unconstrained control vector, $\vec{u}_{\mathrm{reg},k}$ from (7),

$$\begin{aligned} \vec{0} &= \nabla_{u_k}[E_{w_k}[f_k(\vec{x}_k) + c_{k-1}(\vec{x}_{k-1}, \vec{u}_k)]] \\ &= E_{w_k}[\nabla_{u_k}[f_k](\vec{x}_k) + \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}, \vec{u}_k)] \\ &= E_{w_k}\left[\Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k) + \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}, \vec{u}_k)\right], \end{aligned} \tag{13}$$

to focus on the systematic perturbation without the complications due to the corresponding constrained optimal control. In the above, the first of the following chain rules were used to convert control to state derivatives:

$$\begin{aligned} \nabla_{u_k}[f_k](\vec{x}_k) &= \Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k), \\ \nabla_{u_k}\nabla_{u_k}^T[f_k](\vec{x}_k) &= \left(\Psi_{k-1}\Psi_{k-1}^T : \nabla_{x_k}\nabla_{x_k}^T\right)[f_k](\vec{x}_k), \end{aligned} \tag{14}$$

where "$A : B$" represents the trace of the matrix product AB^T. Upon substituting expansions (8) for control and (9) for state, gives

$$\begin{aligned} \vec{0} = E_{w_k}\Big[&\Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k^{(0)} + \sigma \vec{x}_k^{(1)} + \sigma^2 \vec{x}_k^{(2)}) \\ &+ \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)} + \sigma \vec{x}_{k-1}^{(1)} + \sigma^2 \vec{x}_{k-1}^{(2)}, \\ &\vec{u}_k^{(0)} + \sigma \vec{u}_k^{(1)} + \sigma^2 \vec{u}_k^{(2)}) + O(\sigma^3)\Big]. \end{aligned} \tag{15}$$

Next, using Taylor approximations, assuming sufficient differentiability of f_k and c_{k-1}, for perturbations about $\vec{x}_k^{(0)}$ for

f_k and about $(\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)})$ for c_{k-1}. Grouping together terms of order σ^3 and higher into $O(\sigma^3)$ perturbation error results in

$$
\begin{aligned}
\vec{0} \; = \; & E_{w_k}\Big[\Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) + \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)}) \\
& + \Psi_{k-1}^T \nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)}) \left(\sigma\vec{x}_k^{(1)} + \sigma^2\vec{x}_k^{(2)}\right) \\
& + \frac{1}{2}\Psi_{k-1}^T \left(\sigma\vec{x}_k^{(1)}(\sigma\vec{x}_k^{(1)})^T : \nabla_{x_k}\nabla_{x_k}^T\right) \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) \\
& + \nabla_{u_k}\nabla_{x_{k-1}}^T[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)})(\sigma\vec{x}_{k-1}^{(1)} + \sigma^2\vec{x}_{k-1}^{(2)}) \\
& + \nabla_{u_k}\nabla_{u_k}^T[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)})(\sigma\vec{u}_k^{(1)} + \sigma^2\vec{u}_k^{(2)}) \\
& + \frac{\sigma^2}{2}\left(\vec{u}_k^{(1)}(\vec{u}_k^{(1)})^T : \nabla_{u_k}\nabla_{u_k}^T\right) \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)}) \\
& + \sigma^2\left(\vec{x}_{k-1}^{(1)}(\vec{u}_k^{(1)})^T : \nabla_{x_{k-1}}\nabla_{u_k}^T\right) \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)}) \\
& + \frac{\sigma^2}{2}\left(\vec{x}_{k-1}^{(1)}(\vec{x}_{k-1}^{(1)})^T : \nabla_{x_{k-1}}\nabla_{x_{k-1}}^T\right) \\
& \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_k^{(0)}) + O(\sigma^3)\Big]. \qquad (16)
\end{aligned}
$$

Next, replacing $\vec{x}_k^{(1)}$ and $\vec{x}_k^{(2)}$ with their recursive expansion formulas in (11,12), applying the expectation operator using $\vec{w}_k = \sigma\vec{w}_k^{(1)}$, and collecting terms of the same order, so that, after much algebra, the largest order equations are

$$
\begin{aligned}
\text{ord}(\sigma^0) : \vec{0} \; &= \; \Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) + \\
& \quad \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}), \qquad (17) \\
\text{ord}(\sigma^1) : \vec{0} \; &= \; \text{G}_o\vec{x}_{k-1}^{(1)} + \text{H}_o\vec{u}_{\text{reg},k}^{(1)}, \qquad (18)
\end{aligned}
$$

where

$$
\begin{aligned}
\text{G}_o \; &\equiv \; \text{G}_o(\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}) \\
&= \; \Psi_{k-1}^T \nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)})\Phi_{k-1} \\
& \quad + \nabla_{u_k}\nabla_{x_{k-1}}^T[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}), \\
\text{H}_o \; &\equiv \; \text{H}_o(\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}) \\
&= \; \Psi_{k-1}^T \nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)})\Psi_{k-1} \\
& \quad + \nabla_{u_k}\nabla_{u_k}^T[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)})
\end{aligned}
$$

is the Hessian matrix. This notation helps to simplify the $\text{ord}(\sigma^2)$ equation to

$$
\begin{aligned}
\vec{0} \; = \; & \text{G}_o\vec{x}_{k-1}^{(2)} + \text{H}_o\vec{u}_{\text{reg},k}^{(2)} \\
& + \frac{1}{2}\Psi_{k-1}^T \Big((\Phi_{k-1}\vec{x}_{k-1}^{(1)})(\Phi_{k-1}\vec{x}_{k-1}^{(1)})^T : \\
& \nabla_{x_k}\nabla_{x_k}^T\Big) \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) \\
& + \frac{1}{2}\Psi_{k-1}^T \Big((\Psi_{k-1}\vec{u}_{\text{reg},k}^{(1)})(\Psi_{k-1}\vec{u}_{\text{reg},k}^{(1)})^T : \\
& \nabla_{x_k}\nabla_{x_k}^T\Big) \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) \\
& + \frac{1}{2}\Psi_{k-1}^T \left(\widehat{Q}_k : \nabla_{x_k}\nabla_{x_k}^T\right) \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) \\
& + \Psi_{k-1}^T \Big((\Phi_{k-1}\vec{x}_{k-1}^{(1)})(\Psi_{k-1}\vec{u}_{\text{reg},k}^{(1)})^T : \\
& \nabla_{x_k}\nabla_{x_k}^T\Big) \nabla_{x_k}[f_k](\vec{x}_k^{(0)}) \\
& + \frac{1}{2}\left(\vec{u}_{\text{reg},k}^{(1)}(\vec{u}_{\text{reg},k}^{(1)})^T : \nabla_{u_k}\nabla_{u_k}^T\right) \\
& \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}) \\
& + \left(\vec{x}_{k-1}^{(1)}(\vec{u}_{\text{reg},k}^{(1)})^T : \nabla_{x_{k-1}}\nabla_{u_k}^T\right) \\
& \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}) \\
& + \frac{1}{2}\left(\vec{x}_{k-1}^{(1)}(\vec{x}_{k-1}^{(1)})^T : \nabla_{x_{k-1}}\nabla_{x_{k-1}}^T\right) \\
& \nabla_{u_k}[c_{k-1}](\vec{x}_{k-1}^{(0)}, \vec{u}_{\text{reg},k}^{(0)}), \qquad (19)
\end{aligned}
$$

where $\widehat{Q}_k = Q_k/\sigma^2$. Since the state depends on the control from (2), $\vec{x}_k^{(j)} \to \vec{x}_{\text{reg},k}^{(j)}$, but the state "reg" subscript is suppressed here to focus on getting the regular control. We can clearly solve for the critical (regular) control $\vec{u}_{\text{reg},k}$ through $\text{ord}(\sigma^2)$ provided (17) is solvable for $\vec{u}_{\text{reg},k}^{(0)}$ and H_o is invertible.

3. General Caution Term for Cost-To-Go

Kitanidis does not expand the state vector asymptotically in [6] consistently as above, so when the deterministic and caution terms of the cost-to-go are found, different solution for the controls are found. However, the expansion of the control vector induces an expansion of the state vector according to the transition equation (2) for each stage k, so that for consistency the state vector should be expanded with the control vector for each $(k-1)$st stage as well as the kth stage.

This section will follow the dual expansion of state and control of only the kth stage in examining the asymptotics of the cost-to-go performance index. The section initially looks at the case for a general cost-to-go. Note that $\vec{v}_k$ will be used to denote the control vector in this section to distinguish it from the control $\vec{u}_k$ used above.

3.1. Stage $(k-1)$

As above, there is a backward sweep through the stages. The objective to be minimized and the state transition equation are the same as before:

$$
[\text{J}] = E\left[f_N(\vec{x}_N) + \sum_{k=1}^{N-1} c_k(\vec{x}_k, \vec{v}_{k+1})\right], \qquad (20)
$$

$$
\vec{x}_k = \Phi_{k-1}\vec{x}_{k-1} + \Psi_{k-1}\vec{v}_k + \vec{\mu}_k + \vec{w}_k, \qquad (21)
$$

for $k = 2, \ldots, N-1$, except $\vec{v}_k$ is now used for the control. Since the first stage $(k = 1)$ involves an initial condition, it will be treated separately. The cost-to-go for stage $(k-1)$ is

$$
f_{k-1}(\vec{x}_{k-1}) = \min_{v_k}[E_{w_k}[f_k(\vec{x}_k) + c_{k-1}(\vec{x}_{k-1}, \vec{v}_k)]]. \qquad (22)
$$

The critical control vector satisfies

$$
\begin{aligned}
\vec{0} \; &= \; \nabla_{v_k}\left[E_{w_k}[f_k(\vec{x}_k) + c_{k-1}(\vec{x}_{k-1}, \vec{v}_k)]\right] \qquad (23) \\
&= \; E_{w_k}[\Psi_{k-1}^T \nabla_{x_k}[f_k](\vec{x}_k) + \nabla_{v_k}[c_{k-1}](\vec{x}_{k-1}, \vec{v}_k)].
\end{aligned}
$$

Expanding $\vec{v}_k$ and $\vec{x}_k$ asymptotically up to order σ^2,

$$
\begin{aligned}
\vec{v}_k \; &= \; \vec{v}_k^{(0)} + \sigma\vec{v}_k^{(1)} + \sigma^2\vec{v}_k^{(2)} + O(\sigma^3) \qquad (24) \\
\vec{x}_k \; &= \; \vec{x}_k^{(0)} + \sigma\vec{x}_k^{(1)} + \sigma^2\vec{x}_k^{(2)} + O(\sigma^3), \qquad (25)
\end{aligned}
$$

holding $\vec{x}_{k-1}$ temporarily unexpanded. Substituting this into the previous equation yields

$$\begin{aligned}\vec{0} &= E_{w_k}\Big[\Psi_{k-1}^T\nabla_{x_k}[f_k](\vec{x}_k^{(0)}+\sigma\vec{x}_k^{(1)}+\sigma^2\vec{x}_k^{(2))})\\ &\quad +\nabla_{v_k}[c_{k-1}](\vec{x}_{k-1},\vec{v}_k^{(0)}+\sigma\vec{v}_k^{(1)}+\sigma^2\vec{v}_k^{(2)})\\ &\quad +O(\sigma^3)\Big]. \end{aligned} \tag{26}$$

Using a Taylor approximation about $\vec{x}_k^{(0)}$ for f_k, and about $\vec{v}_k^{(0)}$ for c_{k-1}, assuming sufficient differentiability of f_k and c_{k-1},

$$\begin{aligned}\vec{0} &= E_{w_k}\Big[\Psi_{k-1}^T\nabla_{x_k}[f_k](\vec{x}_k^{(0)})\\ &\quad +\Psi_{k-1}^T\nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)})\left(\sigma\vec{x}_k^{(1)}!+\sigma^2\vec{x}_k^{(2)}\right)\\ &\quad +\frac{\sigma^2}{2}\Psi_{k-1}^T\left(\vec{x}_k^{(1)}(\vec{x}_k^{(1)})^T:\nabla_{x_k}\nabla_{x_k}^T\right)\\ &\quad \nabla_{x_k}[f_k](\vec{x}_k^{(0)})+\nabla_{v_k}[c_{k-1}](\vec{x}_{k-1}^{(0)},\vec{v}_k^{(0)})\\ &\quad +\nabla_{v_k}\nabla_{v_k}^T[c_{k-1}](\vec{x}_{k-1}^{(0)},\vec{v}_k^{(0)})\left(\sigma\vec{v}_k^{(1)}+\sigma^2\vec{v}_k^{(2)}\right)\\ &\quad +\frac{\sigma^2}{2}\left(\vec{v}_k^{(1)}(\vec{v}_k^{(1)})^T:\nabla_{v_k}\nabla_{v_k}^T\right)\\ &\quad \nabla v_k[c_{k-1}](\vec{x}_{k-1},\vec{v}_k^{(0)})+O(\sigma^3)\Big]. \end{aligned} \tag{27}$$

From (24) and (21),

$$\vec{x}_k^{(0)} = \Phi_{k-1}\vec{x}_{k-1}+\Psi_{k-1}\vec{v}_k^{(0)}+\vec{\mu}_k, \tag{28}$$

$$\vec{x}_k^{(1)} = \Psi_{k-1}\vec{v}_k^{(1)}+\vec{w}_k, \tag{29}$$

$$\vec{x}_k^{(2)} = \Psi_{k-1}\vec{v}_k^{(2)}. \tag{30}$$

Replacing $\vec{x}_k^{(1)}$ and $\vec{x}_k^{(2)}$ with their recursive expansion formulas above, and applying the expectation operator as in Section 2 yields

$$\begin{aligned}\vec{0} &= \Psi_{k-1}^T\nabla_{x_k}[f_k](\vec{x}_k^{(0)})+\nabla_{v_k}[c_{k-1}](\vec{x}_{k-1}^{(0)},\vec{v}_k^{(0)})\\ &\quad +\sigma\Psi_{k-1}^T\nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)})\Psi_{k-1}\vec{v}_k^{(1)}\\ &\quad +\sigma^2\Psi_{k-1}^T\nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_k^{(0)})\Psi_{k-1}\vec{v}_k^{(2)}\\ &\quad +\frac{\sigma^2}{2}\Psi_{k-1}^T\Big((\Psi_{k-1}\vec{v_k}^{(1)})(\Psi_{k-1}\vec{v_k}^{(1)}):\\ &\quad \nabla_{x_k}\nabla_{x_k}^T\Big)\nabla_{x_k}[f_k](\vec{x}_k^{0)})\\ &\quad +\frac{\sigma^2}{2}\Psi_{k-1}^T\left(\widehat{Q}_k:\nabla_{x_k}\nabla_{x_k}^T\right)\nabla_{x_k}[f_k](\vec{x}_k^{0)})\\ &\quad +\nabla_{v_k}\nabla_{v_k}^T[c_{k-1}](\vec{x}_{k-1}^{(0)},\vec{v}_k^{(0)})\left(\sigma\vec{v}_k^{(1)}+\sigma^2\vec{v}_k^{(2)}\right)\\ &\quad +\frac{\sigma^2}{2}\left(\vec{v}_k^{(1)}(\vec{v}_k^{(1)})^T:\nabla_{v_k}\nabla_{v_k}^T\right)\\ &\quad \nabla_{v_k}[c_{k-1}](\vec{x}_{k-1},\vec{v}_k^{(0)})+O(\sigma^3), \end{aligned} \tag{31}$$

where $\widehat{Q}_k = Q_k/\sigma^2$ and $Q_k = E_{v_k}[\vec{v}_k\vec{v}_k^T]$, here.

Collecting terms of like order, results in

$$\begin{aligned}\text{ord}(\sigma^0):\vec{0} &= \Psi_{k-1}^T\nabla_{x_k}[f_k](\vec{x}_k)\\ &\quad +\nabla_{v_k}[c_{k-1}](\vec{x}_{k-1}^{(0)},\vec{v}_k^{(0)}), \end{aligned} \tag{32}$$

$$\text{ord}(\sigma^1):\vec{0} = \text{H}_o\vec{v}_{\text{reg},k}^{(1)}, \tag{33}$$

where here

$$\begin{aligned}\text{H}_o &\equiv \text{H}_{o,\text{reg}}(\vec{x}_{k-1},\vec{v}_{\text{reg},k}^{(0)})\\ &= \Psi_{k-1}^T\nabla_{x_k}\nabla_{x_k}^T[f_k](\vec{x}_{\text{reg},k}^{(0)})\Psi_{k-1}\\ &\quad +\nabla_{v_k}\nabla_{x_{k-1}}^T[c_{k-1}](\vec{x}_{k-1},\vec{v}_{\text{reg},k}^{(0)}),\\ \vec{x}_{\text{reg},k}^{(0)} &= \Phi_{k-1}\vec{x}_{k-1}+\Psi_{k-1}\vec{v}_{\text{reg},k}^{(0)}+\vec{\mu}_k. \end{aligned}$$

The ord(σ^0) equation implicitly defines a solution for $\vec{v}_{\text{reg},k}^{(0)}$. Since the Hessian H_o must be positive definite, $\vec{v}_{\text{reg},k}^{(1)} = \vec{0}$. This simplifies the ord(σ^2) equation to the form:

$$\begin{aligned}\vec{0} &= \text{H}_o\vec{v}_{\text{reg},k}^{(2)}+\frac{1}{2}\Psi_{k-1}^T\left(\widehat{Q}_k:\nabla_{x_k}\nabla_{x_k}^T\right)\\ &\quad \nabla_{x_k}[f_k](\vec{x}_{\text{reg},k}^{(0)}), \end{aligned} \tag{34}$$

so that

$$\begin{aligned}\vec{v}_{\text{reg},k}^{(2)} &= -\text{H}_o^{-1}\Big(\frac{1}{2}\Psi_{k-1}^T\left(\widehat{Q}_k:\nabla_{x_k}\nabla_{x_k}^T\right)\\ &\quad \nabla_{x_k}[f_k](\vec{x}_{\text{reg},k}^{(0)})\Big), \end{aligned} \tag{35}$$

provided H_o is invertible.

3.1.1. Caution Term: In this subsection, the "reg" notation is dropped for the sake of simplicity. Taking Taylor approximations about $\vec{x}_k^{(0)}$ for $f(\vec{x}_k)$ and about $\vec{v}_k^{(0)}$ for $c_{k-1}(\vec{x}_{k-1},\vec{v}_k)$, then substituting into (22),

$$\begin{aligned}f_{k-1}(\vec{x}_{k-1}) &= f_k(\vec{x}_k^{(0)})+\sigma\nabla_{x_k}[f_k](\vec{x}_k^{(0)})\Psi_{k-1}\vec{v}_k^{(1)}\\ &\quad +\sigma^2\nabla_{x_k}[f_k](\vec{x}_k^{(0)})\Psi_{k-1}\vec{v}_k^{(2)}\\ &\quad +\frac{\sigma^2}{2}\Big((\Psi_{k-1}\vec{v}_k^{(1)})(\Phi_{k-1}\vec{v}_k^{(1)})^T:\\ &\quad \nabla_{x_k}\nabla_{x_k}^T\Big)\nabla_{x_k}[f_k](\vec{x}_k^{(0)})\\ &\quad +\frac{\sigma^2}{2}\left(\widehat{Q}_k:\nabla_{x_k}\nabla_{x_k}^T\right)[f_k](\vec{x}_k^{(0)})\\ &\quad +c_{k-1}(\vec{x}_{k-1},\vec{v}_k^{(0)})\\ &\quad +\nabla_{v_k}[c_{k-1}](\vec{x}_{k-1},\vec{v}_k^{(0)})\left(\sigma\vec{v}_k^{(1)}+\sigma^2\vec{v}_k^{(2)}\right)\\ &\quad +\frac{\sigma^2}{2}\left(\vec{v}_k^{(1)}(\vec{v}_k^{(1)})^T:\nabla_{v_k}\nabla_{v_k}^T\right)\\ &\quad [c_{k-1}](\vec{x}_{k-1},\vec{v}_k^{(0)})+O(\sigma^3), \end{aligned} \tag{36}$$

where use has been made of the recursive expansion state formulas (29,30) for fixed $\vec{x}_{k-1}$.

Let

$$f_{k-1}(\vec{x}_{k-1}) = \left(D_{k-1}+\sigma Z_{k-1}+\sigma^2 S_{k-1}\right)(\vec{x}_{k-1}), \tag{37}$$

which is the Kitanidis [6] decomposition of the $(k-1)$st stage cost into deterministic D_{k-1} and caution term S_{k-1}, but here with an extra Z_{k-1} term linear in the scaling parameter σ for generality. Matching terms of the same order with those of

(36) and using the fact that $\vec{v}_k^{(1)} = \vec{0}$,

$$\begin{aligned} &\mathrm{ord}(\sigma^0): \\ D_{k-1}(\vec{x}_{k-1}) &= D_k(\vec{x}_k^{(0)}) + c_{k-1}(\vec{x}_{k-1}, \vec{v}_k^{(0)}), \quad (38) \\ &\mathrm{ord}(\sigma^1): \\ Z_{k-1}(\vec{x}_{k-1}) &= Z_k(\vec{x}_k^{(0)}), \quad (39) \\ &\mathrm{ord}(\sigma^2): \\ S_{k-1}(\vec{x}_{k-1}) &= S_k(\vec{x}_k^{(0)}) + \frac{1}{2}\left(\widehat{Q}_k : \nabla_{x_k}\nabla_{x_k}^T\right)[D_k](\vec{x}_k^{(0)}) \\ &+ \Big(\nabla_{x_k}^T[D_k](\vec{x}_k^{(0)})\Psi_{k-1} \\ &+ \nabla_{v_k}^T[c_{k-1}](\vec{x}_{k-1}, \vec{v}_k^{(0)})\Big)\vec{v}_k^{(2)}. \quad (40) \end{aligned}$$

Since $Z_{k-1}(\vec{x}_{k-1}) = Z_k(\vec{x}_k^{(0)})$, $Z_{k-1}(\vec{x}_{k-1}) = 0$ by the purely deterministic final condition $f_N(\vec{x}_N^{(0)}) = D_N(\vec{x}_N^{(0)})$, as assumed in [6]. The deterministic term (38) which corresponds to Kitanidis [6] result if $D_k(\vec{x}_k^{(0)})$ is replaced by $D_k(\Phi_{k-1}\vec{x}_{k-1}^{(0)} + \Psi_{k-1}\vec{u}_k^{(0)} + \vec{\mu}_k)$. However, the second order control terms proportional to $\vec{v}_k^{(2)}$ in (40) are *missing* in [6] entirely, due to error, not involving any extra assumptions here..

3.2. Stage One

A similar series of calculations occurs for this stage. However, $\vec{x}_1$, which is fixed as an initial condition, is now involved. This causes some terms present in the calculations of an non-initial stage to disappear.

The cost-to-go for this stage is

$$f_1(\vec{x}_1) = \min_{v_2}\left[E_{w_2}[f_2(\vec{x}_2) + c_1(\vec{x}_1, \vec{v}_2)]\right]. \quad (41)$$

The goal is to find the critical control vector satisfying

$$\begin{aligned} \vec{0} &= \nabla_{v_2}\left[E_{w_2}[f_2(\vec{x}_2) + c_1(\vec{x}_1, \vec{v}_2)]\right] \quad (42) \\ &= E_{w_2}\left[\Psi_1^T\nabla_{x_2}[f_2](\vec{x}_2^{(0)}) + \nabla_{v_2}[c_1](\vec{x}_1^{(0)}, \vec{v}_2^{(0)})\right], \end{aligned}$$

where the commutativity of the gradient and the expectation has been used again. Substituting the asymptotic expansions for $\vec{v}_2$ and $\vec{x}_2$ up to order σ^2 into (42), recalling that $\vec{x}_1$ is fixed, gives small

$$\begin{aligned} \vec{0} &= E_{w_2}\Big[\Psi_1^T\nabla_{x_2}[f_2](\vec{x}_2^{(0)} + \sigma\vec{x}_2^{(1)} + \sigma^2\vec{x}_2^{(2)}) \quad (43) \\ &+ \nabla_{v_2}[c_1](\vec{x}_1, \vec{v}_2^{(0)} + \sigma\vec{v}_2^{(1)} + \sigma^2\vec{v}_2^{(2)}) + O(\sigma^3)\Big] \end{aligned}$$

Using Taylor approximations about $\vec{x}_2^{(0)}$ for f_2 and about $\vec{v}_2^{(0)}$ for c_1, then substituting state transition expansions $\vec{x}_2^{(1)}$ and $\vec{x}_2^{(2)}$, and finally collecting terms of like order leads to

$$\begin{aligned} \mathrm{ord}(\sigma^0): \vec{0} &= \Psi_1^T\nabla_{x_2}[f_2](\vec{x}_2^{(0)}) + \nabla_{v_2}[c_1](\vec{x}_1^{(0)}, \vec{v}_2^{(0)}), \quad (44) \\ \mathrm{ord}(\sigma^1): \vec{0} &= \Psi_1^T\nabla_{x_2}\nabla_{x_2}^T[f_2](\vec{x}_2^{(0)})\left(\Psi_1\vec{v}_2^{(1)}\right) \\ &+ \nabla_{v_2}\nabla_{v_2}^T[c_1](\vec{x}_1^{(0)}, \vec{v}_2^{(0)})\vec{v}_2^{(1)}. \quad (45) \end{aligned}$$

Here (45) is equivalent to $\mathrm{H}_o\vec{v}_{\mathrm{reg},2}^{(1)} = \vec{0}$, where H_o is the Hessian. The Hessian needs to be positive definite since the cost-to-go is being minimized, hence $\vec{v}_{\mathrm{reg},2}^{(1)} = \vec{0}$. The order (σ^2) equation then simplifies to,

$$\begin{aligned} \vec{0} &= \mathrm{H}_o\vec{v}_{\mathrm{reg},2}^{(2)} + \Psi_1^T\left(\widehat{Q}_2 : \nabla_{x_2}\nabla_{x_2}^T\right)\nabla_{x_2}[f_2](\vec{x}_2^{(0)}), \\ \vec{v}_{\mathrm{reg},2}^{(2)} &= -\mathrm{H}_o^{-1}\Psi_1^T\left(\widehat{Q}_2 : \nabla_{x_2}\nabla_{x_2}^T\right)\nabla_{x_2}[f_2](\vec{x}_2^{(0)}), \quad (46) \end{aligned}$$

completing the solution for $\vec{v}_{\mathrm{reg},2}$. We can now find

$$\begin{aligned} \vec{x}_{\mathrm{reg},2} &= \Phi_1\vec{x}_1 + \Psi_1\vec{v}_{\mathrm{reg},2} + \vec{\mu}_2 + \vec{w}_2, \text{ and} \quad (47) \\ \vec{x}_{\mathrm{reg},3} &= \Phi_2\vec{x}_{\mathrm{reg},2} + \Psi_2\vec{v}_{\mathrm{reg},3} + \vec{\mu}_3 + \vec{w}_3. \quad (48) \end{aligned}$$

It is interesting that the derivation of the control is identical with that done previously at this stage. However, since the cost-to-go for stage two does not match, the control from stage one to stage two will be in error.

3.2.1. Recursion for Caution Term:

The cost-to-go is found by putting $\vec{x}_{\mathrm{reg},2}$ and $\vec{v}_{\mathrm{reg},2}$ into (41). The cost-to-go becomes a function of $\vec{x}_1$, fixed as an initial condition, $\vec{w}_2^{(1)}$ and $\vec{x}_2^{(0)}$ (note that $\vec{x}_2^{(0)}$ is by definition a function of $\vec{x}_1$).

The vectors $\vec{v}_2$ and $\vec{x}_2$ were found previously. Recalling the definition (37), expanding

$$f_1(\vec{x}_1) = \min_{v_2}\left[E_{w_2}[f_2(\vec{x}_2) + c_1(\vec{x}_1, \vec{v}_2)]\right], \quad (49)$$

by using appropriate Taylor approximations and substitutions from the state transition equation, and then matching terms of like order yields

$$\begin{aligned} &\mathrm{ord}(\sigma^0): \\ D_1(\vec{x}_1) &= D_2(\vec{x}_2^{(0)}) + c_1(\vec{x}_1, \vec{v}_2^{(0)}), \quad (50) \\ &\mathrm{ord}(\sigma^1): \\ Z_1(\vec{x}_1) &= Z_2(\vec{x}_2^{(0)}) + \Big(\nabla_{x_2}^T[D_2](\vec{x}_2^{(0)})\Psi_1 \\ &+ \nabla_{v_2}^T[c_1](\vec{x}_1, \vec{v}_2^{(0)})\Big)\vec{v}_2^{(1)}, \quad (51) \\ &\mathrm{ord}(\sigma^2): \\ S_1(\vec{x}_1) &= S_2(\vec{x}_2^{(0)}) + \frac{1}{2}\left(\widehat{Q} : \nabla_{x_2}\nabla_{x_2}^T\right)[D_2](\vec{x}_2^{(0)}) \quad (52) \\ &+ \left(\nabla_{x_2}^T[D_2](\vec{x}_2^{(0)})\Psi_1 + \nabla_{v_2}^T[c_1]\left(\vec{x}_1, \vec{v}_2^{(0)}\right)\right)\vec{v}_2^{(2)}. \end{aligned}$$

In simplifying (52), the fact that $\vec{v}_2^{(1)} = \vec{0}$ was used, leading to $Z_1(\vec{x}_1) = \vec{0}$. Therefore, an $\mathrm{ord}(\sigma^2)$ caution term again appears as the correction to the deterministic term, $D_1((\vec{x}_1)$, but with an extra correction proportional to the second order control vector $\vec{v}_2^{(2)}$.

4. Conclusions

Finding the DDP approximation for the stochastic regular control, and hence the stochastic optimal control, requires an accurate expansion of both control and state vectors. The order of the expectation and minimum operators in general cannot be exchanged during the calculations. While general recursive forms exist for the deterministic and caution terms of the cost-to-go as they are defined by Kitanidis [6], the control and state of the system are not the same as the stochastically perturbed DDP approximation found here and missing terms are corrected. The systematic treatment given here can be used to achieve accurate numerical approximations up to ord(σ^2) in various applications that require a number of state and control variables. Future work will include numerical results for concrete f_N and c_k costs, as well as filtering for partial observations.

Acknowledgements. The authors are grateful to Professor C. A. Shoemaker of Cornell University for suggesting that we explore the foundations of discrete dynamic programming under small stochastic perturbations and for an interesting semester sabbatical that the second author spent in the School of Civil and Environmental Engineering, and to Karen R. Matthews, as well as John J. Westman, who helped in this project. We also thank the reviewers, who gave thoughtful and constructive criticisms.

Appendix: Interchange of Minimization and Expectation: Scalar Counter-Example

In this appendix, the validity of the interchange of minimization and expectation operations is examined, by looking at a simple quadratic cost model where the two operations are performed in opposite order, with minimization first, then perturbation expansion, and lastly expectation. However, *this order does not give the same answer* as the original order.

The simple scalar quadratic cost counter-example is

$$f(x,u;w) = a_0+a_1u+\frac{1}{2}a_2u^2+(b_0+b_1u)w+c_0w^2$$

with $b_1 \neq 0$ and $a_2 > 0$, ensuring a unique minimum. For simplicity, scalar variables and unconstrained control are assumed. The example could be made simpler, but at the sacrifice of realism.

When the expectation is properly performed first as specified in the problem formulation,

$$\min_u[E_w[f(x,u;w)]] = \min_u[a_0+a_1u+\frac{1}{2}a_2u^2+c_0\overline{w^2}],$$

where $E_w[w] = 0$ and $E_w[w^2] \equiv \overline{w^2}$ has been used. To find the local minimum, the derivative is set equal to zero, giving

$$\frac{\partial}{\partial u}[E_w[f(x,u;w)]] = a_1+a_2u = 0,$$

so that $u_{\rm reg} = -a_1/a_2$, assuming unconstrained control. Substituting the unconstrained control, $u_{\rm reg}$, back into the original equation and taking the minimum results in

$$\min_u[E_w[f(x,u;w)]] = a_0-\frac{1}{2}\frac{(a_1)^2}{a_2}+c_0\overline{w^2}.$$

However, if the minimization is done first, setting the derivative of $f(x,u;w)$ equal to zero gives

$$\begin{aligned}\frac{\partial}{\partial u}[f(x,u;w)] &= a_1+a_2u+b_1w = 0\\ \Rightarrow \widehat{u}_{\rm reg} &= \widehat{u}_{\rm reg}(w) = -\frac{a_1+b_1w}{a_2},\end{aligned}$$

noting that this control is noise dependent, which greatly complicates the calculation of an optimal control subject to constraints. Taking the minimum of $f(x,u;w)$:

$$\begin{aligned}\min_u[f(x,u;w)] &= a_0-\frac{a_1(a_1+b_1w)}{a_2}\\ +\frac{1}{2}a_2\frac{(a_1+b_1w)^2}{(a_2)^2} &+ c_0w^2+\left(b_0-b_1\frac{a_1+b_1w}{a_2}\right)w.\end{aligned}$$

Now taking the expectation of this gives us

$$E_w[\min_u[f(x,u;w)]] = a_0-\frac{1}{2}\frac{(a_1)^2}{a_2}+c_0\overline{w^2}-\frac{1}{2}\frac{b_1^2\overline{w^2}}{a_2}.$$

The extra term, $[-b_1^2\overline{w^2}/(2a_2)]$, above, with $b_1 \neq 0$ and $a_2 > 0$, means that

$$\min_u[E_w[f(x,u;w)]] \neq E\left[\min_u[f(x,u;w)]\right],$$

in general, so that the minimization operation is not interchangeable with the expectation operation. However, according to a result of Clark reported in [5] (pp. 194-195), the order of differentiation and expectation can be interchanged under fairly general conditions. Here, for example,

$$\frac{\partial}{\partial u}[E_w[f(x,u;w)]] = E_w\left[\frac{\partial}{\partial u}[f(x,u;w)]\right],$$

so that in the correct form the control u is a variable independent of noise variable w.

References

[1] Y. Bar-Shalom, "Stochastic Dynamic Programming: Caution and Probing," *Trans. Autom. Control*, vol. AC-26, pp. 1184-1195, October 1981.

[2] T. Culver and C.A. Shoemaker, "Optimal Control for Groundwater Remediation by Differential Dynamic Programming with Quasi-Newton Approximation," *Water Resources Research*, vol. 29, pp. 823-831, April 1993.

[3] F.B. Hanson, "Techniques in Computational Stochastic Dynamic Programming," in *Control and Dynamic Systems*, vol. 76, Academic Press, pp. 103-162, April 1996.

[4] F.B. Hanson and J.J. Westman, "Computational Stochastic Dynamic Programming Problems: Groundwater Quality Remediation," *Proc. 33rd IEEE Conference on Decision and Control*, vol. 1, pp. 455-460, December 1994.

[5] D.H. Jacobson and D.Q. Mayne, *Differential Dynamic Programming*, American Elsevier Publishing Company, Inc., 1970.

[6] P.K. Kitanidis, "A First-Order Approximation to Stochastic Optimal Control of Reservoirs," *Stochastic Hydrology and Hydraulics*, vol. 1, pp. 169-184, 1987.

[7] P.K. Kitanidis and R. Andricevic : "Optimization of the Pumping Schedule in Aquifer Remediation Under Uncertainty," *Water Resources Research*, vol. 26, pp. 875-885, May 1990.