Submitted to *INFORMS Journal on Computing*
manuscript JOC-2015-03-OA-045

Online Supplement to
Parallel Non-Stationary Direct Policy Search for Risk Averse Stochastic Optimization

Somayeh Moazeni
School of Systems and Enterprises, Stevens Institute of Technology,
1 Castle Point Terrace on Hudson, Hoboken, New Jersey 07030, USA, smoazeni@stevens.edu

Warren B. Powell
Department of Operations Research and Financial Engineering, Princeton University,
Sherrerd Hall, Charlton Street, Princeton, New Jersey 08544, USA, powell@princeton.edu

Boris Defourny
Department of Industrial and Systems Engineering, Lehigh University,
200 West Packer Avenue, Bethlehem, Pennsylvania 18015, USA, defourny@lehigh.edu

Belgacem Bouzaiene-Ayari
Department of Operations Research and Financial Engineering, Princeton University,
Sherrerd Hall, Charlton Street, Princeton, New Jersey 08544, USA, belgacem@princeton.edu

Appendix A: Proof of Proposition 1

In this section, we provide a proof for Proposition 1.

Proof. The optimal Q-functions and value functions can be computed recursively, using equations (7) and (8).

For every time step t and the state variable $s \in \mathscr{S}_t$, since $\bar{x}_t(s) = X_t^\pi(s|\bar{\theta})$ is a solution of the minimization in (3), under the given assumptions, there exist a Lagrange multiplier vector

$$\bar{\lambda}_t(s) \stackrel{\text{def}}{=} \lambda_t^\pi(s|\bar{\theta}) \geq 0$$

such that the following first-order necessary optimality conditions are satisfied at $(\bar{x}_t(s), \bar{\lambda}_t(s))$,

$$\nabla_x \left(C_t\left(s, \bar{x}_t(s)\right) + \mathcal{K}_{\theta_t}\left(s, \bar{x}_t(s)\right)\right) + A(s)^\top \bar{\lambda}_t(s) = 0, \tag{A1}$$

$$A(s)\bar{x}_t(s) \leq b(s), \tag{A2}$$

$$\left(b(s) - A(s)\bar{x}_t(s)\right)^\top \bar{\lambda}_t(s) = 0. \tag{A3}$$

Applying equality (10) in equation (A1), we get

$$\nabla_x\ Q_t\left(s, \bar{x}_t(s)\right) + A(s)^\top \bar{\lambda}_t(s) = 0. \tag{A4}$$

Equation (A4) along with equalities (A2) and (A3) imply that $(\bar{x}_t(s), \bar{\lambda}_t(s))$ satisfies the KKT conditions of the minimization problem in (8). This, along with convexity and continuously differentiability of the Q-function (7), allows us to conclude that $\bar{x}_t(s)$ is an optimal point of the minimization in (8), and an optimal decision for the exact stochastic dynamic programming. $\square$

Appendix B: Proof of Proposition 2

In this section, a proof for Proposition 2 is provided.

Proof. Since the direct policy search restricts the search to the hypothesis space of policy functions $\mathcal{H}^\pi$,

$$V^*(s) \leq V^{\pi_{\theta^*}}(s). \tag{A5}$$

Since θ^* is a solution of the minimization problem (11), we have

$$\begin{aligned} V^{\pi_{\theta^*}}(s) &= \min_\theta\ \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t C\left(S_t, \pi_\theta(S_t)\right) \,\Big|\, S_0 = s\right) \text{ s.t. } \pi_\theta(s) = \arg\min_{x\in\mathscr{X}} C(s,x) + \gamma \mathbb{E}_{S'|s,x}[\theta^\top \Phi(S')], \\ &\leq \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t C\left(S_t, \pi_{\theta_{\xi,\bar{V}}}(S_t)\right) \,\Big|\, S_0 = s\right) \text{ s.t. } \pi_{\theta_{\xi,\bar{V}}}(s) = \arg\min_{x\in\mathscr{X}} C(s,x) + \gamma \mathbb{E}_{S'|s,x}[\theta_{\xi,\bar{V}}^\top \Phi(S')]. \end{aligned}$$

The right-hand-side in the above inequality is the true value of the policy $\pi_{\theta_{\xi,\bar{V}}}$ at state s. Therefore, this inequality yields

$$V^{\pi_{\theta^*}}(s) \leq V^{\pi_{\theta_{\xi,\bar{V}}}}(s). \tag{A6}$$

Combining inequalities (A11) and (A12) implies that

$$0 \leq V^{\pi_{\theta^*}}(s) - V^*(s) \leq V^{\pi_{\theta_{\xi,\bar{V}}}}(s) - V^*(s),$$

and completes the proof of (15).

Under the assumption that the projected Bellman operator has a fixed point such that at state s, $V^{\pi_{\theta_{\xi,\bar{V}}}}(s) = \bar{V}(s) = \Pi_\xi T\bar{V}(s)$, we have

$$\begin{aligned} V^{\pi_{\theta^*}}(s) - V^*(s) &\leq V^{\pi_{\theta_{\xi,\bar{V}}}}(s) - V^*(s) \\ &\leq |V^{\pi_{\theta_{\xi,\bar{V}}}}(s) - \theta_{\xi,\bar{V}}^\top \Phi(s) - V^*(s) + \theta_{\xi,\bar{V}}^\top \Phi(s)| \\ &\leq |V^{\pi_{\theta_{\xi,\bar{V}}}}(s) - \theta_{\xi,\bar{V}}^\top \Phi(s)| + |V^*(s) - \theta_{\xi,\bar{V}}^\top \Phi(s)|, \\ &\leq \sup_{s\in\mathscr{S}} |V^*(s) - \theta_{\xi,\bar{V}}^\top \Phi(s)| \\ &\leq \|V^* - \theta_{\xi,\bar{V}}^\top \Phi\|_\infty \\ &= \|V^* - \Pi_\xi T\bar{V}\|_\infty. \end{aligned} \tag{A7}$$

From Yu and Bertsekas (2010) (or Proposition 6.10 in Bertsekas and Tsitsiklis (1996)), we have

$$\|V^* - \Pi_\xi T\bar{V}\|_\infty \leq \frac{1}{1-\gamma}\|V^* - \Pi_\xi V^*\|_\infty. \tag{A8}$$

Inequalities (A7) and (A8) together establish (16). $\square$

Remark: When the correction function includes the expectation, the projection Π_ξ is over the functions in the hypothesis space $\mathcal{H}^{\bar{Q}}$, defined as

$$\mathcal{H}^{\bar{Q}} \stackrel{\text{def}}{=} \left\{\bar{Q} : \mathscr{S} \times \mathscr{X} \to \mathbb{R} \text{ such that } \bar{Q}(s,x) = \mathcal{K}_\theta(s,x) = \theta^\top \Phi^x(s,x), \quad \theta \in \Theta\right\}.$$

Accordingly, for any function $f : \mathscr{S} \times \mathscr{X} \to \mathbb{R}$, the projection operator is defined as

$$\Pi_\xi f \stackrel{\text{def}}{=} \arg\min_{q \in \mathcal{H}^{\bar{Q}}} \|q - f\|_\xi = \mathcal{K}_{\theta_{\xi,f}}, \tag{A9}$$

where here, $\|f\|_\xi^2 \stackrel{\text{def}}{=} \int_{\mathscr{S}\times\mathscr{X}} (f(s,x))^2 \xi(s,x) ds dx$. The fixed point of the projected Bellman operator is then iteratively computed by solving

$$\bar{Q}_{t+1} = \Pi_\xi \; T_Q \bar{Q}_t$$

where

$$(T_Q\bar{Q})(s,x) \stackrel{\text{def}}{=} \mathbb{E}_{S'|s,x}\left[\min_{x\in\mathscr{X}(S')} C(S',x) + \gamma \bar{Q}(S',x)\right].$$

and $\bar{Q}$ is the fixed point of $\Pi_\xi \; T_Q$.

Consider the direct policy search framework with the policy functions hypothesis space

$$\bar{\mathcal{H}}^\pi \stackrel{\text{def}}{=} \{\pi_\theta : \mathscr{S} \to \mathscr{X} \text{ such that } \pi_\theta(s) \in \arg\min_{x\in\mathscr{X}(s)} C(s,x) + \gamma\mathcal{K}_\theta(s,x), \, \theta \in \Theta\}. \tag{A10}$$

Let θ^* be an optimal parameter value for the direct policy search approach. Therefore,

$$V^{\pi_{\theta^*}}(s) \stackrel{\text{def}}{=} \min_\theta \; \mathbb{E}\left(\sum_{t=0}^\infty \gamma^t C\left(S_t, \pi_\theta(S_t)\right) \;\middle|\; S_0 = s\right) \text{ s.t. } \pi_\theta(s) = \arg\min_{x\in\mathscr{X}} C(s,x) + \gamma\mathcal{K}_\theta(s,x).$$

Since the direct policy search restricts the search to the hypothesis space of policy functions $\bar{\mathcal{H}}^\pi$,

$$V^*(s) \leq V^{\pi_{\theta^*}}(s). \tag{A11}$$

In addition, since θ^* is a solution of the minimization problem (11) over the policy space $\bar{\mathcal{H}}^\pi$, we have

$$V^{\pi_{\theta^*}}(s) \leq \mathbb{E}\left(\sum_{t=0}^\infty \gamma^t C\left(S_t, \pi_{\theta_{\xi,\bar{Q}}}(S_t)\right) \;\middle|\; S_0 = s\right) \text{ s.t. } \pi_{\theta_{\xi,\bar{Q}}}(s) = \arg\min_{x\in\mathscr{X}} C(s,x) + \gamma\mathcal{K}_{\theta_{\xi,\bar{Q}}}(s,x)].$$

The right-hand-side in the above inequality is the true value of the policy $\pi_{\theta_{\xi,\bar{Q}}}$ at state s. Therefore, this inequality yields

$$V^{\pi_{\theta^*}}(s) \leq V^{\pi_{\theta_{\xi,\bar{Q}}}}(s). \tag{A12}$$

Combining inequalities (A11) and (A12) implies that

$$0 \leq V^{\pi_{\theta^*}}(s) - V^*(s) \leq V^{\pi_{\theta_{\xi,\bar{Q}}}}(s) - V^*(s),$$

and completes the proof of (15). $\square$

Appendix C: Proof of Proposition 3

In this section, we provide a proof for Proposition 3. We first present several lemmas which simplify the proof.

Let $\mathcal{P}^{\text{dual}}$ denote the associated dual linear program of the following minimization problem

$$\mathcal{P}^{\text{primal}}: \qquad \min_{x_t \in \mathbb{R}^4_+} \quad C_t(S_t, x_t) \quad \text{s.t} \quad x_t \in \mathscr{X}_t(S_t). \tag{A13}$$

In the following lemmas, we characterize solutions for problem $\mathcal{P}^{\text{primal}}$. In each case, for each solution of the primal problem $\mathcal{P}^{\text{primal}}$, we exhibit a feasible point for the dual problem $\mathcal{P}^{\text{dual}}$, such that the objective values of the primal and dual problems are equal. This implies that these points are optimal for the respective problems. Throughout, we denote $\hat{D}_t \stackrel{\text{def}}{=} \tilde{D}_t - \min\{\tilde{E}_t, \tilde{D}_t\}$ and $\hat{E}_t \stackrel{\text{def}}{=} \tilde{E}_t - \min\{\tilde{E}_t, \tilde{D}_t\}$.

Lemma 1. *Let* $\underline{R}_t \leq \Delta R^D$ *and* $\eta^D \eta^C \leq 1$. *Then,* $\bar{x}_t \stackrel{\text{def}}{=} \left(0,\ \hat{D}_t,\ \eta^D \underline{R}_t\ R^{\boldsymbol{cap}},\ \hat{E}_t\right)$ *and* $\hat{x}_t \stackrel{\text{def}}{=} \left(0,\ \max\{\hat{D}_t - \eta^D \underline{R}_t\ R^{\boldsymbol{cap}}, 0\},\ \max\{\eta^D \underline{R}_t\ R^{\boldsymbol{cap}} - \hat{D}_t, 0\},\ \hat{E}_t\right)$ *are optimal points of problem* $\mathcal{P}^{\text{primal}}$ *and the achieved optimal objective value equals*

$$\tilde{P}_t \left(-\eta^D \min\{\Delta R^D, \underline{R}_t\}\ R^{\boldsymbol{cap}} - \tilde{E}_t\right). \tag{A14}$$

Proof. Both points $\bar{x}_t$ and $\hat{x}_t$ are in the feasible set of $\mathcal{P}^{\text{primal}}$. Also, $(-\tilde{P}_t + \eta^D \eta^C \tilde{P}_t,\ 0,\ 0,\ 0,\ 0,\ -\tilde{P}_t)$ is a feasible point for the dual problem $\mathcal{P}^{\text{dual}}$ with the dual objective value $C_t(s_t, \bar{x}_t) = C_t(s_t, \hat{x}_t)$. Therefore, $\bar{x}_t$ and $\hat{x}_t$ are optimal points of the linear programming problem $\mathcal{P}^{\text{primal}}$. $\square$

Lemma 1 shows that problem $\mathcal{P}^{\text{primal}}$ may have multiple optimal points. Furthermore, given the charge level R_t, implementing either $\bar{x}_t$ or $\hat{x}_t$ implies that

$$R_{t+1} = (1 - \gamma_{\Delta t}) R_t - \min\{\Delta R^D, \underline{R}_t\} = \max\left\{(1 - \gamma_{\Delta t}) R_t - \Delta R^D,\ R^{\min}\right\}. \tag{A15}$$

In particular, $R_{t+1} = R^{\min}$ when $\underline{R}_t \leq \Delta R^D$.

Lemma 2. *In problem $\mathcal{P}^{primal}$, let $\tilde{P}_t = v \le 0$ and $\eta^D = \eta^C = 1$. Then, the point x_t^+, defined in equation (26), is optimal and $C_t(s_t, x_t^+) = \tilde{P}_t \left(\min\{\Delta R^C, \overline{R}_t\}\ R^{\boldsymbol{cap}} - \tilde{E}_t\right)$.*

Proof. The point x_t^+ is feasible in $\mathcal{P}^{\text{primal}}$. In addition, $C_t(s_t, x_t^+)$ equals the dual objective value corresponding to the dual feasible points $(0,\ v,\ v,\ 0,\ 0,\ 0)$, when $\Delta R^C \le \overline{R}_t$, and the feasible dual point $(0,\ 0,\ 0,\ 0,\ v,\ 0)$, when $\overline{R}_t \le \Delta R^C$. $\square$

Under the assumptions of Lemma 2, we have,

$$R_{t+1} = (1-\gamma_{\Delta t})R_t + \min\{\Delta R^C, \overline{R}_t\} \tag{A16}$$

Hence, $R_{t+1} = R^{\max}$, when $\overline{R}_t \le \Delta R^C$.

Now, we are ready to prove the statement of Proposition 3.

Proof. Given $V_T(S_T) = 0$ and using equation (8), we have $Q_{T-1}(S_{T-1}, x) = C_{T-1}(S_{T-1}, x)$ and

$$V_{T-1}(S_{T-1}) = \min_{x \in \mathscr{X}_{T-1}(S_{T-1})} C_{T-1}(S_{T-1}, x).$$

Therefore, at time step $t = T-1$, a myopic policy is optimal. Hence, the value function at time step $t = T-1$ is given by equation (A14), i.e.,

$$V_{T-1}(S_{T-1}) = \tilde{P}_{T-1}\left(-\underline{R}_{T-1}\ R^{\mathbf{cap}} - \tilde{E}_{T-1}\right).$$

Therefore,

$$\begin{aligned}
Q_{T-2}(S_{T-2}, x) &= C_{T-2}(S_{T-2}, x) + \mathbb{E}_{T-2}\left[V_{T-1}(S_{T-1})\right] \\
&= \tilde{P}_{T-2} e^\top x - \mathbb{E}_{T-2}[\tilde{P}_{T-1}\tilde{E}_{T-1}] - \mathbb{E}_{T-2}[\tilde{P}_{T-1}]\underline{R}_{T-1}\ R^{\mathbf{cap}} \\
&= \left(\tilde{P}_{T-2} - \mathbb{E}_{T-2}[\tilde{P}_{T-1}]\right) e^\top x - \mathbb{E}_{T-2}[\tilde{P}_{T-1}]\left(\underline{R}_{T-2} R^{\mathbf{cap}} + \tilde{E}_{T-2} - \tilde{D}_{T-2}\right) \\
&\quad - \mathbb{E}_{T-2}\left[\tilde{P}_{T-1}\tilde{E}_{T-1}\right],
\end{aligned}$$

which confirms that equation (28) holds for $t = T-2$, with $B_{T-2} = 0$.

An optimal decision at time step $t = T-2$ is then computed as,

$$x^*_{T-2} = \arg\min_{x \in \mathscr{X}_{T-2}(S_{T-2})} Q_{T-2}(S_{T-2}, x) = \arg\min_{x \in \mathscr{X}_{T-2}(S_{T-2})} \left(\tilde{P}_{T-2} - \mathbb{E}_{T-2}[\tilde{P}_{T-1}]\right) e^\top x.$$

Thus, using Lemmas 1 and 2, x^*_{T-2} is the myopic policy if $\tilde{P}_{T-2} \ge \mathbb{E}_{T-2}[\tilde{P}_{T-1}]$, and is given by x_t^+, otherwise.

Now assume that $Q_{t+1}(S_{t+1}, x)$ is as in equation (28) for some $t+1 \le T-2$. In the following, we show that $Q_t(S_t, x)$ will also take a form as in (28) and x_t^* is determined by the sign of $\tilde{P}_t \le \mathbb{E}_t[\tilde{P}_{t+1}]$.

From equation (7), $Q_t(S_t, x)$ at time step t equals

$$Q_t(S_t, x) = C_t(S_t, x) + \mathbb{E}_t\left[V_{t+1}(S_{t+1})\right] = \tilde{P}_t e^\top x + \mathbb{E}_t\left[Q_{t+1}(S_{t+1}, x^*_{t+1})\right].$$

From the induction hypothesis, the second term of this equation is equal to

$$\begin{aligned}\mathbb{E}_t\left[Q_{t+1}(S_{t+1},x^*_{t+1})\right] = &\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}\left[\tilde{P}_{t+2}\right]\right)e^\top x^*_{t+1}\right]\\ &-\mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\tilde{P}_{t+2}\right]\ \left(\underline{R}_{t+1}R^{\mathbf{cap}}+\tilde{E}_{t+1}-\tilde{D}_{t+1}\right)\right]\\ &+\mathbb{E}_t\left[B_{t+1}\right](R^{\max}-R^{\min})R^{\mathbf{cap}}-\mathbb{E}_t\left[\sum_{i=t+2}^{T-1}E_{t+1}\left[\tilde{P}_i\tilde{E}_i\right]\right]. \end{aligned} \tag{A17}$$

Depending on the sign of $(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}])$,

$$\begin{aligned}&e^\top x^*_{t+1} = -\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}}, \quad \text{or}\\ &e^\top x^*_{t+1} = -\tilde{E}_{t+1}+\bar{R}_{t+1}\ R^{\mathbf{cap}} = -\tilde{E}_{t+1}+(R^{\max}-R^{\min}-\underline{R}_{t+1})\ R^{\mathbf{cap}}\end{aligned}$$

Therefore, the conditional expectation involving x^*_{t+1} in equation (A17) can be simplified as below,

$$\begin{aligned}&\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)e^\top x^*_{t+1}\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})+(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\right]\\ &\quad+\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}}\right)\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\right]\\ &\quad+\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(-\tilde{E}_{t+1}-(\underline{R}_t\ R^{\mathbf{cap}}+e^\top x+\tilde{E}_t-\tilde{D}_t)\right)\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\right]\\ &\quad-\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\tilde{E}_{t+1}\right]-\mathbb{E}_t\left[\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\left(\underline{R}_t\ R^{\mathbf{cap}}+e^\top x+\tilde{E}_t-\tilde{D}_t\right). \end{aligned}\tag{A18}$$

Depending on the sign of $(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}])$ and optimal decision x^*_{t+1}, we get either

$$\begin{aligned}&\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)|\tilde{P}_{t+1}\geq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left((-\tilde{E}_{t+1}+\bar{R}_{t+1}\ R^{\mathbf{cap}})-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)|\tilde{P}_{t+1}\geq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)|\tilde{P}_{t+1}\geq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right](R^{\max}-R^{\min})R^{\mathbf{cap}}. \end{aligned}\tag{A19}$$

or

$$\begin{aligned}&\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\ \mid\ \tilde{P}_{t+1}\leq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\\ &=\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\times 0\ \mid\ \tilde{P}_{t+1}\leq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]=0. \end{aligned}\tag{A20}$$

Therefore, the first term in equation (A18) is equal to,

$$\begin{aligned}&\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(e^\top x^*_{t+1}-(-\tilde{E}_{t+1}-\underline{R}_{t+1}\ R^{\mathbf{cap}})\right)\right]\\ &=\Pr\left[\tilde{P}_{t+1}\geq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\ \mid\ \tilde{P}_t\right]\mathbb{E}_t\left[\left(\tilde{P}_{t+1}-\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)|\tilde{P}_{t+1}\geq\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right](R^{\max}-R^{\min})R^{\mathbf{cap}}. \end{aligned}\tag{A21}$$

By plugging (A21) into (A18) and then (A17), we get

$$\begin{aligned}Q_t(S_t,x) = {} & \tilde{P}_t e^\top x \\ & + \Pr\left[\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}] \mid \tilde{P}_t\right] \mathbb{E}_t\left[\left(\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right) |\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right] (R^{\max} - R^{\min}) R^{\mathbf{cap}} \\ & + \mathbb{E}_t\left[\left(\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right)\left(-\tilde{E}_{t+1}\right)\right] \\ & - \mathbb{E}_t\left[\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\left(\underline{R}_t\ R^{\mathbf{cap}} + e^\top x + \tilde{E}_t - \tilde{D}_t\right) \\ & - \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\tilde{P}_{t+2}\right]\ \left(\underline{R}_{t+1} R^{\mathbf{cap}} + \tilde{E}_{t+1}\right)\right] \\ & + \mathbb{E}_t\left[B_{t+1}\right](R^{\max} - R^{\min}) R^{\mathbf{cap}} - \mathbb{E}_t\left[\sum_{i=t+2}^{T-1} E_{t+1}\left[\tilde{P}_i \tilde{E}_i\right]\right]. \end{aligned} \tag{A22}$$

Moving the $e^\top x$ terms together, cancelling out the term $\mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\tilde{P}_{t+2}\right]\ \tilde{E}_{t+1}\right]$, and extending the last term to $t+1$, we get,

$$\begin{aligned}Q_t(S_t,x) = {} & \left(\tilde{P}_t - \mathbb{E}_t\left[\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\right) e^\top x \\ & + \Pr\left[\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}] \mid \tilde{P}_t\right] \mathbb{E}_t\left[\left(\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right) |\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right] (R^{\max} - R^{\min}) R^{\mathbf{cap}} \\ & - \mathbb{E}_t\left[\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\left(\underline{R}_t\ R^{\mathbf{cap}} + \tilde{E}_t - \tilde{D}_t\right) \\ & - \mathbb{E}_t\left[\mathbb{E}_{t+1}\left[\tilde{P}_{t+2}\right]\ \underline{R}_{t+1} R^{\mathbf{cap}}\right] + \mathbb{E}_t\left[B_{t+1}\right](R^{\max} - R^{\min}) R^{\mathbf{cap}} - \sum_{i=t+1}^{T-1} \mathbb{E}_t\left[\tilde{P}_i \tilde{E}_i\right]. \end{aligned} \tag{A23}$$

Note that

$$\mathbb{E}_t\left[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\ \underline{R}_{t+1} R^{\mathbf{cap}}\right] = \mathbb{E}_t\left[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\ \left(\underline{R}_t R^{\mathbf{cap}} + \tilde{E}_t - \tilde{D}_t\right) + \mathbb{E}_t\left[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right] e^\top x. \tag{A24}$$

Plugging equation (A24) into (A23), canceling out the term $\mathbb{E}_t\left[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\left(\underline{R}_t\ R^{\mathbf{cap}} + \tilde{E}_t - \tilde{D}_t\right)$ and moving the $e^\top x$ to the first line, we arrive at

$$\begin{aligned}Q_t(S_t,x) = {} & \left(\tilde{P}_t - \mathbb{E}_t\left[\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right] - \mathbb{E}_t\left[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right]\right) e^\top x \\ & + \Pr\left[\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}] \mid \tilde{P}_t\right] \mathbb{E}_t\left[\left(\tilde{P}_{t+1} - \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right) |\tilde{P}_{t+1} \geq \mathbb{E}_{t+1}[\tilde{P}_{t+2}]\right] (R^{\max} - R^{\min}) R^{\mathbf{cap}} \\ & - \mathbb{E}_t\left[\tilde{P}_{t+1}\right]\left(\underline{R}_t\ R^{\mathbf{cap}} + \tilde{E}_t - \tilde{D}_t\right) + \mathbb{E}_t\left[B_{t+1}\right](R^{\max} - R^{\min}) R^{\mathbf{cap}} - \sum_{i=t+1}^{T-1} \mathbb{E}_t\left[\tilde{P}_i \tilde{E}_i\right]. \end{aligned}$$

We now cancel the term $\mathbb{E}_t[\mathbb{E}_{t+1}[\tilde{P}_{t+2}]]$ in the first line, and collect the second line with the forth line. Thus, we get to equation (28), where B_t is as in equation (29). This confirms the validity of equation (28) for time step t.

In addition, an optimal decision at time step t, x_t^*, can be computed by

$$\arg\min_{x \in \mathscr{X}_t} Q_t(S_t,x) = \arg\min_{x \in \mathscr{X}_t} \left(\tilde{P}_t - \mathbb{E}_t[\tilde{P}_{t+1}]\right) e^\top x.$$

Thus, depending on the sign of $\tilde{P}_t - \mathbb{E}_t[\tilde{P}_{t+1}]$, it is the myopic policy or x_t^+. $\square$

Appendix D: Data

This document describes the method to obtain the data used in the computational experiment presented in Section 5 of the paper. The parameter estimation methods are analyzed in greater detail in (Moazeni et al. 2015).

Electricity Price Data and Load Data

Our numerical study uses publicly available market operations data from the website of the New York Independent System Operator (NYISO) at `http://www.nyiso.com/public/markets_operations/`.

For prices, we use five-minute real-time zonal locational marginal prices (in \$/MWh), referred to as Real-Time Market LBMP, for the New York City zone (N.Y.C.). For demand, we use five-minute real time zonal load (in MWh) referred to as RTD Actual Load, for the New York City zone (N.Y.C.). The time range for both data sets goes from 12 am of January 1, 2007 to $11:55$ pm of December 31, 2011. We replace the negative prices with \$1/MWh, and take the average of 12 measurements from five-minute prices for each hour to derive hourly prices.

Deseasonalization of the Price and Load Data

The estimation of the deterministic seasonal trend for the price is carried out as follows: (1) For every hour of a day, P_t^{hour} is set as the mean of the price of that hour over all days in our data set. Then for every hour, the computed P_t^{hour} is subtracted from the hourly prices to derive hourly residual prices. (2) For each day in a week, P_t^{day} is computed as the mean of the average hourly residual prices per day over all the weeks in the historical data set. Then for each hour this value is subtracted from the hourly residual prices to obtain updated hourly residual prices. (3) For each month of the year, P_t^{month} is the mean of the average updated hourly residual prices per month over the five years 2007 to 2011. This value is then subtracted from each updated hourly residual price to attain deseasonalized hourly prices.

The same three-step deseasonalization procedure is carried out to obtain deseasonalized load levels. The estimated values of the seasonality parameters are reported in Tables 1, 2, and 3.

Parameters of the Stochastic Processes for Price, Load, and Wind

We use the maximum likelihood method to estimate the parameters in $\tilde{Y}_t^P$ and $\tilde{Y}_t^D$. The estimated parameters are reported in Table 4. The initial values of the stochastic processes are set to the values corresponding to the deseasonalized demand and price at hour 12 am of January 1, 2007. Our numerical studies assume that the system owner is only responsible to serve 25% of the demand

Table 1 Hour-of-Day Seasonality Parameters

Hour		D_t^{hour} [MWh]	P_t^{hour} [$/MWh]
0	[12 am , 1 am)	5159.62	52.92
1	[1 am , 2 am)	4943.20	47.81
2	[2 am , 3 am)	4819.20	43.88
3	[3 am , 4 am)	4781.16	42.30
4	[4 am , 5 am)	4885.42	44.07
5	[5 am , 6 am)	5225.77	48.49
6	[6 am , 7 am)	5713.78	57.95
7	[7 am , 8 am)	6160.71	61.54
8	[8 am , 9 am)	6515.21	66.08
9	[9 am , 10 am)	6756.23	71.13
10	[10 am , 11 am)	6898.87	72.75
11	[11 am , 12 pm)	6973.70	73.00
12	[12 pm , 1 pm)	7006.82	75.05
13	[1 pm , 2 pm)	7015.28	77.30
14	[2 pm , 3 pm)	7017.51	80.85
15	[3 pm , 4 pm)	7029.42	81.30
16	[4 pm , 5 pm)	7036.35	85.72
17	[5 pm , 6 pm)	6974.70	88.22
18	[6 pm , 7 pm)	6881.89	82.30
19	[7 pm , 8 pm)	6773.92	80.02
20	[8 pm , 9 pm)	6595.16	77.30
21	[9 pm , 10 pm)	6309.24	68.12
22	[10 pm , 11 pm)	5918.03	61.52
23	[11 pm , 12 am)	5492.11	56.51

Table 2 Day-of-Week Seasonality Parameters

Day	D_t^{day} [MWh]	P_t^{day} [$/MWh]
Monday	174.19	2.43
Tuesday	266.76	2.49
Wednesday	263.84	3.42
Thursday	224.28	1.17
Friday	146.66	0.34
Saturday	-468.64	-3.65
Sunday	-592.24	-5.45

generated by the calibrated model to NYISO load data, as in (Moazeni et al. 2015). Parameter values for the wind energy E_t, are reported in Table 4 and are identical to those described in (Moazeni et al. 2015).

Table 3 Month-of-Year Seasonality Parameters

Month	$D_t^{\mathbf{month}}$ **[MWh]**	$P_t^{\mathbf{month}}$ **[\$/MWh]**
January	-221.78	10.29
February	-253.70	6.04
March	-520.57	-2.71
April	-682.52	-0.97
May	-454.68	1.69
June	659.15	10.29
July	1454.98	13.99
August	1147.22	-0.79
September	297.15	-8.56
October	-524.26	-14.23
November	-627.55	-15.38
December	-307.61	0.23

Table 4 Parameters for Energy Price, Energy Demand, and Wind Energy Models

Parameters for $\tilde{Y}_t^P$	Parameters for $\tilde{Y}_t^D$	Parameters for $\tilde{Y}_t^E$ and $\tilde{W}_t$
$\mu_P = 4.35$	$\phi_D = 0.97$	$\phi_E = 0.95$
$\beta_P = 37.48$	$\sigma_D = 138.08$	$\sigma_E = 0.9$
$\sigma_P = 2.08$	$\tilde{Y}_0^D = -63.63$	$\mu_E = 3$
$\mu_J = 0.03$		$\tilde{Y}_0^E = 0$
$\sigma_J = 0.41$		
$\lambda_J = 0.27$		
$\tilde{Y}_0^P = -5.88$		

References

Bertsekas, D. P. and J. Tsitsiklis (1996). *Neuro-Dynamic Programming.* Athena Scientific, 1st edition.

Moazeni, S., W. B. Powell, and A. H. Hajimiragha (2015). Mean-Conditional Value-at-Risk optimal energy storage operation in the presence of transaction costs. *IEEE Transactions on Power Systems* 30(3), 1222–1232.

Yu, H. and D. P. Bertsekas (2010). Error bounds for approximations from projected linear equations. *Mathematics of Operations Research* 35(2), 306–329.