Information Relaxations, Duality, and Convex Stochastic Dynamic Programs
(*Online Appendix*)

David B. Brown and James E. Smith

Fuqua School of Business
Duke University
Durham, NC 27708-0120

dbbrown@duke.edu, jes9@duke.edu

First version: October 15, 2013
This version: September 16, 2014

B. Detailed Derivations and Discussions

B.1. Comparisons with Brown and Smith (2011)

As discussed in §1.1, in Brown and Smith (2011), hereafter BS (2011), we used linear penalties to generate information relaxation bounds for a dynamic portfolio optimization problem with transaction costs. There, we considered an approximate model that is a portfolio optimization model that ignores transactions costs; this is a relaxation of the original model and is not difficult to solve to optimality. The value functions for this frictionless model are differentiable and hence the results of Proposition 3.1 apply and by part (iii), we can calculate bounds that certainly improve on the bound provided by the frictionless model. However, as noted in §3.1, we did not use this approach in BS (2011).

In BS (2011), we considered two kinds penalties. The first kind of penalty we considered were linear approximations of approximate value functions of the form of equation (10). However, we did not establish any theoretical properties for these penalties (other than feasibility) and did not apply the penalties as suggested in Proposition 3.1(iii). Specifically, we did use the frictionless value functions as an approximate value functions to generate a penalty of the form of (10), but we did not calculate penalties by linearizing around the optimal strategy for the frictionless system, as suggested in Proposition 3.1(iii).

In BS (2011), we also considered "gradient-based penalties" that can be related to the gradient penalties of the form of equation (10). If the gradient penalties are based on optimal value functions $\hat{V}_t$ for an approximate model that is a convex DP, following a derivation like that of equations (11)-(14), if there are no binding constraints, we can write the gradient penalty around the optimal policy $\hat{\alpha}$ for this approximate model as

$$\hat{\pi}_{\nabla}(\boldsymbol{a}) \quad = \quad \sum_{t=0}^{T} \nabla \hat{r}_t(\hat{\boldsymbol{\alpha}}_t)^\top (\boldsymbol{a}_t - \hat{\boldsymbol{\alpha}}_t) \; + \; \left(\hat{V}_{t+1}(\hat{\boldsymbol{\alpha}}_t) - \mathbb{E}\left[\hat{V}_{t+1}(\hat{\boldsymbol{\alpha}}_t) \,\middle|\, \mathcal{F}_t \right] \right) \; . \tag{41}$$

Comparing to the original definition of the gradient-based penalties in equation (10), we have replaced the terms involving differences in gradients of value functions and expected value functions with $\nabla \hat{r}_t(\hat{\boldsymbol{\alpha}}_t)$. Dual feasibility of a penalty of the form of (41) relies on the fact that the policy $\hat{\alpha}$ is optimal for the approximate model with rewards $\hat{r}_t$. When there are binding constraints, the first-order conditions (13) would include Lagrange multiplier terms and the penalty (41) would also include such terms.

The "gradient-based penalties" in BS (2011) are similar to equation (41) above. However, in BS (2011), we did not recognize that penalties of the form of equation (41) are a special case of the more general form of equation (10), in the case where the value functions are optimal value functions for an approximate system. The gradient-based penalties in BS (2011) differ from our current approach based on equation (10) in several ways. First, the gradient-based penalties in BS (2011) did not include the zero-mean control variate terms inside the parentheses in (10) or (41). However, similar control variates were used in the numerical simulations in BS (2011). Second, the method we proposed in BS (2011) for gradient-based penalties with non-differentiable approximate value functions (which was based on directional derivatives) generally does not result in tractable inner problems. We address this issue in §3.2 of this paper.

Finally, the gradient-based penalties in BS (2011) differ from our current approach when there are binding constraints. The gradient-based penalties in BS (2011) are, as proven there, dual feasible and guaranteed to improve on the value given by the relaxed model. However, the gradient-based penalties in BS (2011) have slack in that the equalities defining dual feasibility (3) need not hold with equality: the slack is due to omitting the Lagrange multiplier terms associated with these constraints (as discussed following (41)) in the gradient-based penalties in BS (2011).

We can improve on the gradient-based penalties in BS (2011) by following the approach of Proposition 3.1(iii) directly, i.e., by generating penalties using the frictionless value function with equation (10), linearizing around the policy chosen in the frictionless model. The Lagrange multipliers that were omitted in the gradient-based penalties in BS (2011) are automatically included in this approach. The bounds generated using this approach can be proven to be at least as good as the bounds given by gradient-based penalties in BS (2011) (in every scenario) and our numerical experiments suggest these improvements can be significant if the constraints (e.g., disallowing short sales or borrowing) are frequently binding.

B.2. Appendix Material for the Network Revenue Management Example

B.2.1 Proof of Proposition 4.1

Proof. For (i), the inequality in (23) follows from standard Lagrange duality results. For the equality in (23), we argue that each of restrictions (a)-(c) can be imposed on the Lagrange multipliers without affecting the minimization over $\boldsymbol{\lambda}$. To see that (b) can be imposed without loss, note that in moving from formulation (19) to (20), it is equivalent to define (20) over decision variables $\boldsymbol{a} \in \{0,1\}^{|\mathcal{L}(i)|}$ corresponding to the legs involved on each itinerary, as these are the only legs that consume capacity for a given itinerary. Omitting the other decision variables (and their corresponding coupling constraints) is equivalent to setting $\lambda_{i\ell t} = 0$ for all $\ell \notin \mathcal{L}(i)$ in (21).

To see that we can impose (a) and (c), note that the Lagrange multipliers are minimized separately for each $i \in \mathcal{I}$ and t, so we can fix i and t for the remainder of the proof.

For (c), note that we can rewrite the immediate reward in (21) as:

$$
\begin{aligned}
r_i \bar{a}_t + \sum_{\ell \in \mathcal{L}} \lambda_{i\ell t} \left(a_{\ell t} - \bar{a}_t \right) &= \sum_{\ell \in \mathcal{L}} \left(\frac{r_i}{L} a_{\ell t} + \lambda_{i\ell t} a_{\ell t} - \frac{\lambda_{i\ell t}}{L} \sum_{\ell' \in \mathcal{L}} a_{t\ell'} \right) \\
&= \sum_{\ell \in \mathcal{L}} \left(\frac{r_i}{L} + \lambda_{i\ell t} - \bar{\lambda}_{it} \right) a_{\ell t}, \qquad (42)
\end{aligned}
$$

where $\bar{\lambda}_{it}$ denotes the average of $\lambda_{i\ell t}$ over ℓ; the first equality follows from the definition of $\bar{a}_t$ and the second equality follows by rearranging terms in the summation. Notice that in (42) it is only differences in $\lambda_{i\ell t}$ from their average over legs that matters; shifting $\lambda_{i\ell t}$ by any constant across all legs does not change these differences. Thus we can always shift all $\lambda_{i\ell t}$ by a constant to ensure that $\sum_{\ell \in \mathcal{L}} \lambda_{i\ell t} = r_i$.

Finally, to see (a), note that any λ satisfying (b) and (c) reduces (42) to $\sum_{\ell \in \mathcal{L}(i)} \lambda_{i\ell t} a_{\ell t}$. If $\lambda_{i\ell t} < 0$ for some $\ell \in \mathcal{L}(i)$, we claim we can always increase this value to zero while still satisfying (a) and (c) without increasing the objective. To see this, since (b) and (c) hold, if some $\lambda_{i\ell t} < 0$, then there must exist some $\ell' \in \mathcal{L}(i)$ such that $\lambda_{i\ell' t} > 0$. Let $\epsilon_\ell = \min\{-\lambda_{i\ell t}, \lambda_{i\ell' t}\} > 0$; keeping all other values of λ fixed, set

$$
\begin{aligned}
\tilde{\lambda}_{i\ell t} &= \lambda_{i\ell t} + \epsilon_\ell \\
\tilde{\lambda}_{i\ell' t} &= \lambda_{i\ell' t} - \epsilon_\ell.
\end{aligned}
$$

Note that at least one of $\tilde{\lambda}_{i\ell t}$, $\tilde{\lambda}_{i\ell' t}$ is zero. In the Lagrangian, it is always optimal to reject a leg with $\lambda_{i\ell t} \le 0$: since $\tilde{\lambda}_{i\ell t} \le 0$, it is still optimal to reject leg ℓ with $\tilde{\lambda}_{i\ell t}$ in place of $\lambda_{i\ell t}$, and the value associated with leg ℓ does not change. On the other hand, the reward $\lambda_{i\ell' t}$ has been strictly reduced, so the value associated with leg ℓ' has either been strictly reduced (if it was optimal to accept i at ℓ') or unchanged (if it was optimal to reject i at ℓ'). Either way, the total value in the Lagrangian is no larger than it was before. This process can be continued until $\lambda_{i\ell t} = 0$, and similarly for all such legs with negative values of $\lambda_{i\ell t}$.

The proof for (24) is in Topaloglu (2009); the representation in that paper includes terms of the form $\max\{r_{i_t} - \sum_{\ell \in \mathcal{L}} \lambda_{i_t \ell t}, 0\}$, but these are zero when $\boldsymbol{\lambda} \in \Lambda$ by condition (c).

Now consider the properties of ϑ^λ mentioned in (ii). It is straightforward to see that ϑ^λ is nondecreasing in capacity. To show that ϑ^λ is piecewise linear and concave in capacity is straightforward by induction. This is clearly true at $T{+}1$, when $\vartheta^\lambda_{\ell,T+1} = 0$. Now assume it is true at $t{+}1$. We have

$$
\vartheta^\lambda_{\ell t}(c_{\ell t}, i_t) = \max_{a_{\ell t} \in A_{\ell t}(c_{\ell t}, i_t)} \left\{ \lambda_{i_t \ell t} a_{\ell t} + \mathbb{E}[\vartheta^\lambda_{\ell, t+1}(c_{\ell t} - f_{i_t \ell} a_{\ell t}, \tilde{i}_{t+1})] \right\},
$$

and since $\vartheta^\lambda_{\ell t}$ is the maximization of a piecewise linear concave function over a convex set, $\vartheta^\lambda_{\ell t}$ is piecewise linear and concave as well. $\square$

B.2.2 Gradient Selection for Penalties from Lagrangian Relaxations

The following simple result from convex analysis will be useful in our discussion of the gradient selection procedure.

Lemma B.1. *Given a convex set X and a point $\hat{\boldsymbol{x}} \in X$, a vector $\boldsymbol{d}$ satisfies $\boldsymbol{d}^\top(\boldsymbol{x} - \hat{\boldsymbol{x}}) \le 0$ for all $\boldsymbol{x} \in X$ if and only if $-\boldsymbol{d} \in \partial \mathbb{1}_X(\hat{\boldsymbol{x}})$.*

Proof. Note that

$$\begin{aligned}\partial \mathbb{1}_X(\hat{\boldsymbol{x}}) &= \left\{\boldsymbol{d} \;:\; \mathbb{1}_X(\boldsymbol{x}) \le \mathbb{1}_X(\hat{\boldsymbol{x}}) + \boldsymbol{d}^\top(\boldsymbol{x}-\hat{\boldsymbol{x}}) \text{ for all } \boldsymbol{x}\right\} \\ &= \left\{\boldsymbol{d} \;:\; \mathbb{1}_X(\boldsymbol{x}) \le \boldsymbol{d}^\top(\boldsymbol{x}-\hat{\boldsymbol{x}}) \text{ for all } \boldsymbol{x}\right\} \\ &= \left\{-\boldsymbol{d} \;:\; \boldsymbol{d}^\top(\boldsymbol{x}-\hat{\boldsymbol{x}}) \le 0 \text{ for all } \boldsymbol{x}\in X\right\},\end{aligned}$$

where the second equality uses $\hat{\boldsymbol{x}} \in X$, so that $\mathbb{1}_X(\hat{\boldsymbol{x}}) = 0$, and the third equality uses the fact that $\mathbb{1}_X(\boldsymbol{x}) = -\infty$ for any $\boldsymbol{x} \notin X$. □

We now discuss how to select gradients for the gradient penalty (27) to ensure that optimal values of the inner problems are weakly tighter than V_0^λ in every scenario. Recall from the proof of Proposition 3.2 that the conditions given by (37), i.e.,

$$(\boldsymbol{\delta}_{t\text{-}1}, \mathbf{0}) \in \partial\left\{r_t(\boldsymbol{\alpha}_t) + \mathbb{1}_{A_t}(\boldsymbol{\alpha}_t)\right\} + \mathbb{E}[\, \boldsymbol{\delta}_t \mid \mathcal{F}_t \,] \tag{43}$$

for all $t = 1, \dots, T$, are sufficient to ensure a tight bound to a convex DP in every scenario. We will show how to select gradients of the Lagrangian relaxation value functions that satisfy (43). Using such penalties in the inner problems (26) will then lead to optimal values that are weakly tighter than V_0^λ in every scenario, as these inner problems also include the leg coupling constraints.

As in the proof of Proposition 3.2, we will work forward in time and show that we can calculate gradients satisfying (43) in each time period. Assume this has been done up to period t. By rearranging (43), we see that we then need to select gradients $\boldsymbol{\delta}_t$ such that

$$-\mathbb{E}[\, \boldsymbol{\delta}_t \mid \mathcal{F}_t \,] \in \partial\left\{r_t(\boldsymbol{\alpha}_t) + \mathbb{1}_{A_t}(\boldsymbol{\alpha}_t)\right\} - (\boldsymbol{\delta}_{t\text{-}1}, \mathbf{0}), \tag{44}$$

where by convention we let $\boldsymbol{\delta}_0 = \emptyset$. Proposition 3.2(ii) shows that such gradients exist.

For the Lagrangian relaxations, recall that value functions and decisions decouple by leg. It follows that we can focus on a single leg $\ell \in \mathcal{L}$ for the remainder of the discussion; from Lemma 3.1(iv), we can apply the same procedure for each leg and obtain a valid gradient selection for the overall problem. We let $i_1, \dots, i_T$ denote the itineraries in the scenario, and α_ℓ^λ denote the leg ℓ actions under the optimal policy α^λ to the Lagrangian relaxation. We let $\boldsymbol{\delta}_{t\text{-}1}^\ell$ denote the gradient selection for the leg-specific value function $\vartheta_{\ell t}^\lambda(c_{\ell t}(\boldsymbol{\alpha}_{\ell, t\text{-}1}^\lambda), i_t)$; this is a vector of length t-1 (recall by convention that we assume in this problem that the first action occurs at $t = 1$, and that the role of $t = 0$ is simply to average over the first itinerary request). We let $\delta_{t,\tau}^\ell$, for $\tau < t$, denote the τ^{th} element of gradient selection $\boldsymbol{\delta}_t$.

In the Lagrangian relaxations, the rewards are given by the Lagrange multipliers times the in-period acceptance decision, i.e., the rewards are $\lambda_{i\ell t} a_{\ell t}$ in each period. Thus we can state (44) for the Lagrangian relaxation as, given $\boldsymbol{\delta}_{t\text{-}1}^\ell$, needing to find $\boldsymbol{\delta}_t^\ell(i)$ for each itinerary i such that

$$-\mathbb{E}[\boldsymbol{\delta}_t^\ell(\tilde{i})] \;\; \in \;\; \partial\{\mathbb{1}_{A_{\ell t}}(\boldsymbol{\alpha}_{\ell t}^\lambda)\} + \lambda_{i_t \ell t}\boldsymbol{e}_t - (\boldsymbol{\delta}_{t\text{-}1}^\ell, 0), \tag{45}$$

where $A_{\ell t} = \{a_{\ell t} \in [0,1] \;:\; f_{i_t \ell} a_{\ell t} \le c_{\ell t}\}$. In (45), $\boldsymbol{e}_t$ is a vector of length t with a 1 in position t and 0 elsewhere; this reflects the fact that the time t rewards only depend directly on $a_{\ell t}$ and not earlier decisions.

The first term on the right-hand side of (45) is the differential of the characteristic function for the constraints in the Lagrangian relaxation on leg ℓ, taken at an optimal policy α_ℓ^λ to the Lagrangian relaxation. For simplicity, we will (for now) assume that the capacity constraint at leg ℓ is not binding along α_ℓ^λ through period t; in this case, only the constraints $a_{\ell t} \in [0,1]$ play a role in $\partial\{\mathbb{1}_{A_{\ell t}}(\boldsymbol{\alpha}_{\ell t}^\lambda)\}$ (recall from LemmaB.1 that zero is the only gradient for the characteristic function of a non-binding constraint).

The key to selecting gradients satisfying (45) is to recognize that the gradients $\boldsymbol{\delta}_t^\ell$ are contained in a set of dimension one. To see this, recall that leg capacity is a linear function of past actions, so:

$$\vartheta_{\ell t}^\lambda(c_{\ell t}(\boldsymbol{\alpha}_{\ell, t\text{-}1}^\lambda), i_t) \;\; = \;\; \vartheta_{\ell t}^\lambda\left(c_{\ell 0} - \sum_{\tau=1}^{t\text{-}1} f_{i_\tau \ell}\alpha_{\ell\tau}\right).$$

Applying Lemma 3.1(v), this implies that

$$\partial_{\boldsymbol{a}}\{\vartheta^{\lambda}_{\ell t}(c_{\ell t}(\boldsymbol{\alpha}^{\lambda}_{\ell,t\text{-}1}), i_t)\} = -\boldsymbol{f}_{\ell t}\partial_c\{\vartheta^{\lambda}_{\ell t}(c_{\ell t}(\boldsymbol{\alpha}^{\lambda}_{\ell,t\text{-}1}), i_t)\}, \tag{46}$$

where $\partial_{\boldsymbol{a}}$ denotes the differential with respect to actions up to time period t-1, ∂_c denotes the (one-dimensional) differential with respect to leg ℓ capacity, and $\boldsymbol{f}_{\ell t}$ is a vector of length t-1 with component τ being $f_{i_\tau \ell}$. Since $\vartheta^{\lambda}_{\ell t}$ is piecewise linear concave in leg capacity, its differential with respect to leg capacity at any point is the convex hull of the left and right derivatives with respect to capacity at that point; these left and right derivatives can be evaluated by taking differences of $\vartheta^{\lambda}_{\ell t}$ at adjacent values of capacity. Thus we can express all gradients $\mathbb{E}[\boldsymbol{\delta}^{\ell}_t(\tilde{i})]$ of $\mathbb{E}[\vartheta^{\lambda}_{\ell,t+1}(c_{\ell,t+1}(\boldsymbol{\alpha}^{\lambda}_{\ell t}), \tilde{i})]$ through convex combinations of these left and right derivatives for each itinerary $\tilde{i}$ in the expectation:

$$\mathbb{E}[\boldsymbol{\delta}^{\ell}_t(\tilde{i})] \quad = \quad -\boldsymbol{f}_{\ell,t+1}\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))], \tag{47}$$

where $d^-(\tilde{i})$ and $d^+(\tilde{i})$ are the left and right derivatives of $\vartheta^{\lambda}_{\ell,t+1}(c_{\ell,t+1}(\boldsymbol{\alpha}^{\lambda}_{\ell t}), \tilde{i})$ with respect to capacity, and $\gamma(\tilde{i}) \in [0,1]$ for each itinerary $\tilde{i}$. The coefficients $\gamma(\tilde{i})$ represent how much weight is applied to the right derivatives for each itinerary $\tilde{i}$ in the next-period expectation.

Plugging (47) into (45) and rearranging terms, the problem of finding the desired gradient selection reduces to finding a selection of weights $\gamma(\tilde{i})$ such that

$$\boldsymbol{f}_{\ell,t+1}\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))] + (\boldsymbol{\delta}^{\ell}_{t\text{-}1}, 0) - \lambda_{i_t \ell t}\boldsymbol{e}_t \quad \in \quad \partial\{\mathbb{1}_{A_{\ell t}}(\boldsymbol{\alpha}^{\lambda}_{\ell t})\}. \tag{48}$$

The dimension of the vectors in condition (48) is t, reflecting the fact that $\vartheta^{\lambda}_{\ell,t+1}$ depends on actions $(a_{\ell 1}, \ldots, a_{\ell t})$.

Notice that we can restrict attention to past actions $a_{\ell\tau}$ such that $\ell \in \mathcal{L}(i(\tau))$, i.e., we can restrict attention to past actions such that the itinerary $i(\tau)$ includes leg ℓ. To see this, note that for actions $a_{\ell\tau}$ with $\ell \notin \mathcal{L}(i(\tau))$, we have $f_{\ell i(\tau)} = 0$, and thus the corresponding components of both $\boldsymbol{f}_{\ell,t+1}$ and, by (46), $\boldsymbol{\delta}^{\ell}_{t\text{-}1}$ will be zero; the same reasoning applies at $\tau = t$, because $\lambda_{i_t \ell t} = 0$ is implied by $\boldsymbol{\lambda} \in \Lambda$. Thus all components on the left hand side of (48) are zero for these irrelevant past actions, and since 0 is always in the differential of the characteristic function at any feasible action, (48) automatically holds for these actions. Thus, we need only study (48) for past actions $a_{\ell\tau}$ that affect capacity on leg ℓ; for such actions we have $f_{\ell i(\tau)} > 1$. To streamline notation below, we will assume all itinerary requests consume at most one unit of capacity on leg ℓ; this is true in all examples we consider.

Treating (48) one component at a time, note that when the capacity constraint is not binding through time t on leg ℓ under α^{λ}_{ℓ}, that $A_{\ell t}$ only depends on $a_{\ell t}$, and thus the first t components of the right-hand side of (48) are zero. Thus if any itinerary $i(\tau)$ for time $\tau < t$ uses leg ℓ capacity, the corresponding component of (48) requires

$$\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))] + \delta^{\ell}_{t\text{-}1,\tau} \quad = \quad 0. \tag{49}$$

Note that by (46), $\delta^{\ell}_{t\text{-}1,\tau} = -d$ for some $d \in \partial_c\{\vartheta^{\lambda}_{\ell t}(c_{\ell t}(\boldsymbol{\alpha}^{\lambda}_{\ell,t\text{-}1}), i_t)\}$. Thus, for any $\tau < t$ such that $i(\tau)$ consumes leg ℓ capacity, (49) reduces to the same condition:

$$\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))] \quad = \quad d. \tag{50}$$

For component t of (48), note that $\partial\mathbb{1}_{\{a \geq 0\}}(0) = \{g \ : \ g \geq 0\}$. Thus if i_t uses leg ℓ capacity and it is optimal under α^{λ} to accept itinerary i_t, the t^{th} component of (48) must satisfy:

$$\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))] - \lambda_{i_t \ell t} \quad \leq \quad 0. \tag{51}$$

Similarly, $\partial\mathbb{1}_{\{a \leq 1\}}(1) = \{g \ : \ g \leq 0\}$. Thus if i_t uses leg ℓ capacity and it is optimal under α^{λ} to reject itinerary i_t, the t^{th} component of (48) must satisfy:

$$\mathbb{E}[d^-(\tilde{i}) + \gamma(\tilde{i})(d^+(\tilde{i}) - d^-(\tilde{i}))] - \lambda_{i_t \ell t} \quad \geq \quad 0. \tag{52}$$

Note that the "accept conditions" (51) require making $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))]$ small enough, and the "reject conditions" (52) require making $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))]$ large enough; finally, the conditions (50) require making $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))]$ equal to some given value d. Thus the required conditions reduce to an interval constraint, i.e., $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))] \in [L, U]$, where L and U are determined by the optimal decisions α_ℓ^λ up through time t according to conditions (50), (51), and (52); here we take $L = U = d$ if past itineraries in the scenario use leg ℓ capacity. Otherwise, we take $[L, U]$ as one of $[-\infty, \lambda_{i_t \ell t}]$, $[\lambda_{i_t \ell t}, \infty]$, or $[-\infty, \infty]$, depending on whether i_t uses leg ℓ capacity and whether it is optimal to accept or reject i_t under α^λ. Note that the relevant interval must be nonempty, since Proposition 3.2 shows that the gradient selection exists.

Since $\vartheta_{\ell,t+1}^\lambda$ is concave, choosing $\gamma(\tilde{i}) = 1$, i.e., all right capacity derivatives for all itineraries $\tilde{i}$ minimizes the expected value $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))]$; conversely, choosing $\gamma(\tilde{i}) = 0$, i.e., all left capacity derivatives for all itineraries $\tilde{i}$ maximizes this expected value.

There are various ways to select $\gamma(\tilde{i})$ to ensure $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))]$ is in the relevant interval. In our examples, we search for the desired weights by starting with all right derivatives, i.e., $\gamma(\tilde{i}) = 0$ for all itineraries $\tilde{i}$; if the expected value with this choice falls outside the relevant interval (i.e., if $\mathbb{E}[d^+(\tilde{i})] < L$), then we can increase the weight on some left derivatives to increase the expected value.

There are also various ways to shift towards left derivatives. In our examples, we sort the itineraries in order of how much changes in weights $\gamma(\tilde{i})$ from 1 to 0 increases the expected value (i.e., in order of how much switching to left derivatives for each itinerary increases the expected value). We then change as few values of the weights from 1 to 0 as necessary to ensure $\mathbb{E}[d^-(\tilde{i})+\gamma(\tilde{i})(d^+(\tilde{i})-d^-(\tilde{i}))] \geq L$, placing a fractional value in the "last" needed $\gamma(\tilde{i})$ if necessary. Other ways of choosing the weights $\gamma(\tilde{i})$ are of course possible and may be more effective; regardless, any such choice in this problem ensures the inner problem values (26) will be weakly smaller than V_0^λ.

When the capacity constraint is binding, we have $c_{\ell,t+1}(\boldsymbol{\alpha}_{\ell t}^\lambda) = 0$. In this case, we can capture the effect of the capacity constraint by taking left derivatives with respect to capacity at $\vartheta_{\ell,t+1}^\lambda(c_{\ell,t+1}(\boldsymbol{\alpha}_{\ell t}^\lambda), \tilde{i})$ to be sufficiently large. In theory we have $\vartheta_{\ell,t+1}^\lambda(c_{\ell,t+1}(\boldsymbol{\alpha}_{\ell t}^\lambda), \tilde{i}) = -\infty$ for any negative capacity values; in our examples we have found it sufficient to take the left derivative at zero capacity to be some multiple (greater than one) of the right derivative at zero capacity. Otherwise, we proceed as described above in those cases where the capacity constraint is binding.

Note that in the case with reoptimization, there is no guarantee that we can select gradients satisfying the above conditions, due to the fact that the value functions are changing at selected points in time along each scenario, as well as the fact that we take gradients around the Lagrangian heuristic, which is in general not optimal for the Lagrangian relaxation. However, we can apply the same type of procedure as described above to try to ensure the expected value of the next-period gradients are contained within the relevant interval $[L, U]$. In our examples, we apply a similar procedure that starts with all right derivatives and mixes in as few left derivatives as necessary to meet the interval condition. In the event that the expected value with all left derivatives falls below L, which can occur with reoptimization, we simply use all left derivatives to make the expected value as close to L as possible.

B.3. Appendix Material for the Lost-Sales Model

B.3.1 Proof of Proposition 5.1

Proof. Part (i). The proof is by induction. We first consider the terminal case and show that this terminal value function is $L^\natural$-convex. To streamline the notation, we let $\boldsymbol{z} = (z_0, \ldots, z_{t+L})$ let $g(\boldsymbol{z}) = J_{T+L+1}(\boldsymbol{z})$. We want to prove that $g(\boldsymbol{z})$ is $L^\natural$-convex, i.e., that $\hat{g}(\epsilon, \boldsymbol{z}) = g(\boldsymbol{z} - \epsilon\mathbf{1})$ is submodular in $(\epsilon, \boldsymbol{z})$ on $\mathbb{Z}^1 \times \mathbb{Z}^{T+L+1}$. Using (30), we can write $\hat{g}(\epsilon, \boldsymbol{z})$ as

$$\begin{aligned}
&\hat{g}(\epsilon, \boldsymbol{z}) = g(\boldsymbol{z} - \epsilon\mathbf{1}) \\
&\qquad \min_{\boldsymbol{s}} \quad \sum_{t=0}^{T+L} \gamma^t \left(\gamma^L c(z_t - z_{t\text{-}1}) \ + \ h(z_{t-L} - \epsilon - s_t) \ + \ p(d_t - s_t + s_{t\text{-}1})\right) + (\gamma^L c + L)\epsilon \qquad (53) \\
&\text{subject to} \quad s_t \geq s_{t\text{-}1} \qquad \text{for } t = 0, \ldots, T+L \\
&\qquad\qquad\quad\ s_t \leq z_{t-L} - \epsilon \qquad \text{for } t = 0, \ldots, T+L\ .
\end{aligned}$$

(Here the $(\gamma^L c + L)\epsilon$ term in the objective reflects our convention that $z_\tau = 0$ when $\tau < 0$.) Defining $s'_t = s_t + \epsilon$ for $t \geq 0$ and taking $s'_{-1} = 0$, we can rewrite this as

$$\begin{aligned}
&\hat{g}(\epsilon, \boldsymbol{z}) = \\
&\qquad \min_{\boldsymbol{s}'} \quad \sum_{t=0}^{T+L} \gamma^t \left(\gamma^L c(z_t - z_{t\text{-}1}) \ + \ h(z_{t-L} - s'_t) \ + \ p(d_t - s'_t + s'_{t\text{-}1}) \right) + (\gamma^L c + L - p)\epsilon \\
&\text{subject to} \quad s'_0 \geq \epsilon \\
&\qquad\qquad\quad s'_t \geq s'_{t\text{-}1} \qquad \text{for } t = 1, \ldots, T+L \\
&\qquad\qquad\quad s'_t \leq z_{t-L} \qquad \text{for } t = 0, \ldots, T+L \ .
\end{aligned} \tag{54}$$

Now $\hat{g}(\epsilon, \boldsymbol{z})$ can be shown to be submodular by using the result that the partial minimization of a submodular function over a sublattice results in a submodular function (Topkis (1998)[1] Theorem 2.7.6). Here the objective function is linear in $(\boldsymbol{s}', \epsilon, \boldsymbol{z})$ and hence submodular in $(\boldsymbol{s}', \epsilon, \boldsymbol{z})$. The set of $(\boldsymbol{s}', \epsilon, \boldsymbol{z})$ satisfying the constraints is a sublattice because each constraint involves just two variables that appear with positive signs on opposite sides of the constraint inequalities (Topkis 1998, Example 2.2.7(b)). Then, we can apply the partial minimization result to conclude that, when minimizing $\boldsymbol{s}'$ over a section of this sublattice, the result, namely $\hat{g}(\epsilon, \boldsymbol{z})$, is submodular in $(\epsilon, \boldsymbol{z})$. Thus $g(\boldsymbol{z})$ is $L^\natural$-convex, completing the proof for the terminal case.

For the inductive step, we make the dependence of the value functions J_t on demands explicit and assume that $J_{t+1}(\boldsymbol{z}_t; d_0, \ldots, d_t)$ is $L^\natural$-convex in $\boldsymbol{z}_t$ for all demand sequences $(d_0, \ldots, d_t)$; we then show the same is true for period t. Since a weighted average of $L^\natural$-convex functions is $L^\natural$-convex, fixing the demand sequence $(d_0, \ldots, d_{t\text{-}1})$ and taking expectations over the period-t demand $\tilde{d}_t$, we know that

$$h_{t+1}(\boldsymbol{z}_t; d_0, \ldots, d_{t\text{-}1}) = \mathbb{E}\Big[\, J_{t+1}(\boldsymbol{z}_t; d_0, \ldots, d_{t\text{-}1}, \tilde{d}_t) \,\Big]$$

is $L^\natural$-convex in $\boldsymbol{z}_t$.

Now we use an argument like that for the terminal case (partial minimization of a submodular function over a sublattice results in a submodular function) to complete the proof. Suppressing demands again, from (31), we have

$$J_t(\boldsymbol{z}_{t\text{-}1} - \epsilon \mathbf{1}) \quad = \quad \min_{\{z'_t : z'_t \geq z_{t\text{-}1}\}} h_{t+1}((\boldsymbol{z}_{t\text{-}1}, z'_t) - \epsilon(\mathbf{1}, 1)) \tag{55}$$

where $\mathbf{1}$ denotes a t-vector of ones and $z'_t = z_t + \epsilon$. The objective function in (55) is submodular in $(\epsilon, \boldsymbol{z}_{t\text{-}1}, z'_t)$ because h_{t+1} is $L^\natural$-convex. The constraint in (55) involves two variables and they appear with positive signs on opposite sides of the constraint inequalities and thus, as above, the feasible set forms a sublattice. Choosing z'_t from the sublattice (or section of the sublattice) to minimize this objective then results in a function $J_t(\boldsymbol{z}_{t\text{-}1} - \epsilon\mathbf{1})$ that is submodular in $(\epsilon, \boldsymbol{z}_{t\text{-}1})$. Thus $J_t(\boldsymbol{z}_{t\text{-}1})$ is $L^\natural$-convex.

Part (ii). Let $\hat{g}(z_0, \boldsymbol{z}) = g(\boldsymbol{z} - z_0 \mathbf{1})$, as in the definition of $L^\natural$-convexity. First note that all gradients of $\partial \hat{g}(0, \boldsymbol{z})$ must be of the form $(-\mathbf{1}'\boldsymbol{x}, \boldsymbol{x})$ where $\boldsymbol{x} \in \mathbb{R}^n$. This implies the gradient $\boldsymbol{x} \in \mathbb{R}^n$ is in $\partial g(\boldsymbol{z})$ if and only if $(-\mathbf{1}'\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{n+1}$ is in $\partial \hat{g}(0, \boldsymbol{z})$.

Let $\hat{N} = \{0\} \cup N$ and let $\hat{\rho}$ be the difference function for $\hat{g}(\hat{X})$ at the point $(0, \boldsymbol{z})$ defined for any $\hat{X} \subset \hat{N}$ as follows:

(a) When $0 \notin \hat{X}$ and $X = \hat{X}$, we have $\hat{\rho}(\hat{X}) = g(\boldsymbol{z} + \mathbb{1}(X)) - g(\boldsymbol{z})$
(b) When $0 \in \hat{X}$ and $X = \hat{X} - \{0\}$, we have $\hat{\rho}(\hat{X}) = g(\boldsymbol{z} + \mathbb{1}(X) - \mathbb{1}(N)) - g(\boldsymbol{z}) = g(\boldsymbol{z} - \mathbb{1}(\bar{X})) - g(\boldsymbol{z})$, where $\bar{X} = N - X$ is the complement of X in N.

Murota (2003; Theorem 7.43) shows that difference function $\hat{\rho}$ is submodular and the differential of $\hat{g}$ is

$$\begin{aligned}
&\partial \hat{g}(z_0, \boldsymbol{z}) = \mathbf{B}(\hat{\rho}) \\
&\qquad = \left\{ \hat{x} = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} : \sum_{i \in \hat{X}} \hat{x}_i \leq \hat{\rho}(\hat{X}) \text{ for all } \hat{X} \subseteq \hat{N} \text{ and } \sum_{i \in \hat{N}} \hat{x}_i = \hat{\rho}(\hat{N}) \right\}.
\end{aligned} \tag{56}$$

[1] Topkis, D.M. 1998. *Supermodularity and Complementarity*, Princeton University Press.

where $\mathbf{B}(\hat{\rho})$ is referred to as the base polyhedron associated with $\hat{\rho}$. Note that when $\hat{X} = \hat{N}$, $\hat{\rho}(\hat{N}) = g(\boldsymbol{z}+\mathbb{1}(N)-\mathbb{1}(N))-g(\boldsymbol{z}) = g(\boldsymbol{z})-g(\boldsymbol{z}) = 0$. Thus the condition in (56) that $\sum_{i\in\hat{N}} \hat{x}_i = \hat{\rho}(\hat{N})$ is automatically satisfied for all gradients of $\hat{g}(0, \boldsymbol{z})$ (and $\sum_{i\in\hat{N}} \hat{x}_i = 0$ as noted earlier).

We can then divide the subsets of $\hat{N}$ into those subsets that contain 0 and those that do not contain 0 (as in (a) and (b) above) to rewrite the constraints in (56):

(a) For the cases where $0 \notin \hat{X}$ and $X = \hat{X}$, the constraint in (56) is:

$$\sum_{i\in\hat{X}} \hat{x}_i = \sum_{i\in X} \hat{x}_i \leq \hat{\rho}(\hat{X}) = g(\boldsymbol{z} + \mathbb{1}(X)) - g(\boldsymbol{z}) \quad \text{for all } X \subseteq N.$$

(b) For the cases where $0 \in \hat{X}$ and $X = \hat{X} - \{0\}$, taking $\bar{X} = N - X$ to be the complement of X in N, the constraint in (56) is:

$$\begin{aligned}\sum_{i\in\hat{X}} \hat{x}_i &= \hat{x}_0 + \sum_{i\in X} \hat{x}_i = -\sum_{i\in N} \hat{x}_i + \sum_{i\in X} \hat{x}_i = -\sum_{i\in\bar{X}} \hat{x}_i \\ &\leq \hat{\rho}(\hat{X}) = g(\boldsymbol{z} + \mathbb{1}(X) - \mathbb{1}(N)) - g(\boldsymbol{z}) = g(\boldsymbol{z} - \mathbb{1}(\bar{X})) - g(\boldsymbol{z}) \quad \text{for all } X \subseteq N\end{aligned}$$

or, equivalently,

$$\sum_{i\in X} \hat{x}_i \geq g(\boldsymbol{z}) - g(\boldsymbol{z} - \mathbb{1}(X)) \quad \text{for all } X \subseteq N.$$

Thus, combining (a) and (b), we can represent (56) as in Proposition 5.1(ii). □

B.3.2 Gradient Selections for Penalties in the Lost-Sales Model

As discussed in §5.4, we take the approximate value functions $\hat{V}_t$ in the gradient penalty (15) to be the value function (30)-(31), but with the horizon T set to $t+L$. We calculate these values by adding the previously incurred costs to the myopic cost-to-go function (32). As in the myopic heuristic, we augment the terminal value (30) to include a residual term $\rho(-)$ to value leftover inventory. As this is a DP value function for a lost-sales model (albeit with a different time horizon), these value functions will be $L^\natural$-convex and we can use the characterization of the gradients provided in Proposition 5.1(ii). In our numerical examples, we will also consider gradients based on the exact value function in cases where the lead times are short enough to solve the model exactly.

Our description of the gradient selection procedure will follow the notation of the proof of Proposition 5.1(ii). For a given period t, we take $\boldsymbol{z}$ to be the cumulative order quantity $\boldsymbol{z}_{t\text{-}1}$ and the function $g(\boldsymbol{z})$ to be the approximate or exact value function, $\hat{V}_t$, as discussed above. The dimension n of the proof corresponds to t in the example. Also, as in the proof, we take $\hat{g}(z_0, \boldsymbol{z}) = g(\boldsymbol{z} - z_0\mathbf{1})$. The differential $\partial\hat{g}(0, \boldsymbol{z})$ is given by the base polyhedron $\mathbf{B}(\hat{\rho})$ in (56). As discussed in the proof, $\boldsymbol{x} \in \mathbb{R}^n$ is in $\partial g(\boldsymbol{z})$ if and only if $(-\mathbf{1}'\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}^{n+1}$ is in $\partial\hat{g}(0, \boldsymbol{z})$. Thus, there is a one-to-one correspondence between the extreme points of $\partial g(\boldsymbol{z})$ and those of $\partial\hat{g}(0, \boldsymbol{z})$.

We can generate all extreme points of the base polyhedron $\mathbf{B}(\hat{\rho})$ using the greedy algorithm (see, e.g., Murota (2003)). The greedy algorithm proceeds as follows: Pick any sequence of elements of $\hat{N} = \{0, 1, \ldots, n\}$, and let $\sigma(i)$ denote the index of the i^{th} element in this sequence. Now consider the increasing sequence of subsets

$$\emptyset = S_{-1} \subset S_0 \subset \ldots \subset S_n = \hat{N}$$

where elements of $\hat{N}$ are added to the subsets one-at-a-time according to the chosen sequence of indices; i.e., such that $S_i = S_{i-1} \cup \{\sigma(i)\}$. Then define the vector $\hat{\boldsymbol{x}}$ with elements

$$\hat{x}_{\sigma(i)} = \hat{\rho}(S_i) - \hat{\rho}(S_{i-1}) = \hat{g}((0, \boldsymbol{z}) + \mathbb{1}(S_i)) - \hat{g}((0, \boldsymbol{z}) + \mathbb{1}(S_{i-1})), \qquad \text{for } i = 0, \ldots, n. \tag{57}$$

If the elements of $\hat{\boldsymbol{x}}$ are all finite, $\hat{\boldsymbol{x}}$ is an extreme point of $\mathbf{B}(\hat{\rho})$ and maximizes the weighted combination of elements $\sum_{i=0}^{n} w_i\hat{x}_i$ over $\mathbf{B}(\hat{\rho})$ whenever the weights w_i are nonincreasing according to the chosen sequence. If some elements of $\hat{\boldsymbol{x}}$ are infinite, then this linear programming problem is unbounded and we can use the function evaluations in (57) to identify extreme rays instead of extreme points (see, e.g., Fujishige (2005)[2] for more discussion). Thus, using the greedy algorithm, we can identify an extreme point (or extreme ray) of the differential of $\hat{V}_t$ by evaluating the approximate valuation function $\hat{V}_t$ a total of $t+1$ times. The set of all gradients (i.e., the differential) is then given by the convex hull of these extreme points (and rays).

We consider two different approaches for choosing gradient selections in the lost-sales problem. In the first approach, we use a simple 50-50 convex combination of two "extreme" extreme points of the differential, corresponding to taking sequences $(0, 1, \ldots, n)$ and $(n, \ldots, 1, 0)$ in the greedy algorithm. If one of these sequences leads to infinite values when taking difference (57), we place 100% weight on the other point. (In the lost-sales problem, only the first of these two extreme sequences, $(0, 1, \ldots, n)$, can have infinite values.) These gradients are easy to compute but generally will not satisfy the first-order condition of equation (16).

In the second approach for selecting gradients, we use a more sophisticated procedure where we try to find an element of the differential that satisfies condition (16). Here we work forward in time as discussed following Proposition 3.2, attempting to select a gradient $\boldsymbol{\delta}_t$ for period t such that $\mathbb{E}[\,\boldsymbol{\delta}_t \,|\, \mathcal{F}_t\,]$ matches $(\boldsymbol{\delta}_{t\text{-}1}, 0)$ where $\boldsymbol{\delta}_{t\text{-}1}$ is the selected gradient for the previous period. To use the greedy algorithm, we work with the function $\hat{g}(z_0, \boldsymbol{z}) = g(\boldsymbol{z} - z_0\mathbf{1})$ where $g(\boldsymbol{z}) = \mathbb{E}\Big[\,\hat{V}_t(\boldsymbol{z})\,\Big|\,\mathcal{F}_t\,\Big]$ with differential $\partial\hat{g}(0, \boldsymbol{z})$ given by the base polyhedron $\mathbf{B}(\hat{\rho})$ in (56). The goal then is to find a point in $\partial\hat{g}(0, \boldsymbol{z}) = \mathbf{B}(\hat{\rho})$ equal to $\boldsymbol{x} \equiv (-\mathbf{1}'\boldsymbol{\delta}_{t\text{-}1}, \boldsymbol{\delta}_{t\text{-}1}, 0)$.

To find such a point, we construct $\mathbf{B}(\hat{\rho})$ by sequentially generating extreme points or extreme rays using the greedy algorithm, stopping when the desired point $\boldsymbol{x}$ is inside the convex hull of identified extreme points (and rays) or when we know that $\boldsymbol{x}$ lies outside of $\mathbf{B}(\hat{\rho})$. We focus on the case where $\mathbf{B}(\hat{\rho})$ contains no extreme rays (i.e., is compact):

(i) We begin with $i = 0$, taking the initial set $\mathbf{B}_0$ of extreme points to be the two points used in the simple 50-50 approach described above, i.e., corresponding to the sequences $(0, 1, \ldots, n)$ and $(n, \ldots, 1, 0)$ in the greedy algorithm. Let $\Sigma_0 = \{(0, 1, \ldots, n)\} \cup \{(n, \ldots, 1, 0)\}$ denote the set of sequences corresponding to the extreme points in $\mathbf{B}_0$.

(ii) Given $\mathbf{B}_i$, we solve a linear least squares optimization problem to find the point $\boldsymbol{y}$ in the convex hull of $\mathbf{B}_i$ (written $conv(\mathbf{B}_i)$) that is closest (in the sense of two-norm) to $\boldsymbol{x}$.

(iii) If $\boldsymbol{x}$ is in $conv(\mathbf{B}_i)$ (i.e., the minimum distance is zero), then the weights in the convex combination define a gradient selection satisfying the desired conditions. Stop.

(iv) If $\boldsymbol{x}$ is not in $conv(\mathbf{B}_i)$, then the difference $\boldsymbol{x} - \boldsymbol{y}$ defines a hyperplane that separates $\boldsymbol{x}$ from $conv(\mathbf{B}_i)$. (This follows from the "projection theorem" and is a standard step in proofs of the separating hyperplane theorem; see e.g., Bertsekas, Nedić, and Ozdaglar (2003; Theorem 2.3.1).)

(v) We then sort the components of $\boldsymbol{x} - \boldsymbol{y}$ in descending order and let the sort order define an sequence σ_i on the index set $\hat{N}$ for use in the greedy algorithm.

(vi) If σ_i is in Σ_i, then Stop: $\boldsymbol{x}$ is not in $\mathbf{B}(\hat{\rho})$ and $\boldsymbol{y}$ is the closest point to $\mathbf{B}(\hat{\rho})$. (In this case, the hyperplane $\boldsymbol{x} - \boldsymbol{y}$ separates the whole base polyhedron $\mathbf{B}(\hat{\rho})$ from the desired point $\boldsymbol{x}$.)

(vii) Otherwise, let $\boldsymbol{g}_i$ be the extreme point of $\partial\hat{g}(z_0, \boldsymbol{z})$ given by the greedy algorithm with sequence σ_i; let $\Sigma_{i+1} = \Sigma_i \cup \{\sigma_i\}$ and $\mathbf{B}_{i+1} = \mathbf{B}_i \cup \{\boldsymbol{g}_i\}$. Set $i = i + 1$ and continue to step (ii) above.

As there are a finite number of extreme points (corresponding to the $(n + 1)!$ possible sequences of elements in $\hat{N}$), this algorithm will stop eventually. In our setting, this algorithm usually stops in fewer than n steps, far fewer this worst-case theoretical possibility. Of course, we can limit the number of steps taken in the algorithm by stopping after, say, 100 iterations. In this case, when we stop, we take the current $\boldsymbol{y}$ to define the gradient selection. This will be a valid gradient selection, but will not satisfy the desired condition (16) exactly. When we are working with the true value function, we can be sure that the algorithm, if not stopped early, will find a gradient selection satisfying (16) exactly. If this is not the case, we have no such guarantee, even if the algorithm considers all extreme points.

If infinite differences are encountered in the greedy algorithm, we augment the set of extreme points $\mathbf{B}_i$ to include the identified extreme rays. In this case, we proceed in a similar manner but, rather than looking for the desired point in the convex hull of the extreme points, we look for the desired point in the convex hull of the extreme points plus the cone generated by the extreme rays.

[2]Fujishige, S. 2005. *Submodular Functions and Optimization (Second Edition)*, Annals of Discrete Mathematics, 58. Elsevier: Amsterdam.