OPERATIONS RESEARCH
DOI 10.1287/opre.1090.0755ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"A Shadow Simplex Method for Infinite Linear Programs" by Archis Ghate, Dushyant Sharma, and Robert L. Smith, *Operations Research*, DOI 10.1287/opre.1090.0755.

A Proofs of Technical Results

Proofs of results in the text are presented here.

A.1 Proof of Lemma 2.6

To see that the first condition is sufficient for Assumption 2.4 to hold, note that for any $x \geq 0$, $|c_j x_j| = |c_j||x_j| = |c_j|x_j \leq |c_j|u_j$ and $\sum\limits_{j=1}^{\infty} |c_j|u_j \leq ||u||_\infty \sum\limits_{j=1}^{\infty} |c_j| < \infty$ since $u \in l_\infty$ and $c \in l_1$. Similarly, to see that the second condition is sufficient, observe that for $x \geq 0$, $|c_j x_j| = |c_j||x_j| = |c_j|x_j \leq |c_j|u_j$ and $\sum\limits_{j=1}^{\infty} |c_j|u_j \leq ||c||_\infty \sum\limits_{j=1}^{\infty} |u_j| < \infty$ since $c \in l_\infty$ and $u \in l_1$.

A.2 Proof of Proposition 2.7

The proof requires three preliminary results that we now state and prove.

Lemma A.1. *The feasible region F of problem (P) is closed.*

Proof. For row i of matrix A, let $J(i)$ denote the finite (by Assumption 2.2) support set $\{j : a_{ij} \neq 0\}$. Consider sets $X^i = \{x \in R^\infty : \sum\limits_{j \in J(i)} a_{ij}x_j = b_i\}$ for $i = 1, 2, \ldots$. Notice that $F = \bigcap\limits_{i=1}^{\infty} X^i \bigcap \{x \in R^\infty : x \geq 0\}$. The set $\{x \in R^\infty : x \geq 0\}$ is closed. We show that sets X^i are closed for all i. Then since arbitrary intersections of closed sets are closed, F must be closed. Let $\{x^i(n)\}_{n=1}^{\infty}$ be a convergent sequence of points in X^i with limit $\bar{x}^i \in R^\infty$. For any integer n we have $\sum\limits_{j \in J(i)} a_{ij}x_j^i(n) = b_i$. Taking limits we obtain $\lim\limits_{n\to\infty} \sum\limits_{j \in J(i)} a_{ij}x_j^i(n) = b_i$. Hence $\sum\limits_{j \in J(i)} a_{ij}\left(\lim\limits_{n\to\infty} x_j^i(n)\right) = b_i$ since $J(i)$ is finite. Thus $\sum\limits_{j \in J(i)} a_{ij}\bar{x}_j^i = b_i$. Therefore $\bar{x}^i \in X^i$ implying X^i is closed. □

Corollary A.2. *The feasible region F of (P) is compact.*

Proof. Let $0 \leq u \in R^\infty$ be as in Assumption 2.3. Consider the set $X = \{x \in R^\infty : 0 \leq x_j \leq u_j \;\; \forall j\}$. X is compact by Tychonoff Product Theorem (Theorem 2.61 page 52 of [1]). F is closed by Lemma A.1. Assumption 2.3 implies that $F \subseteq X$. Therefore F is compact. □

Lemma A.3. *The objective function of problem (P) is continuous over its feasible region F.*

Proof. Let $\{x(n)\}_{n=1}^{\infty}$ be a convergent sequence of points in F with limit $\bar{x} \in F$. We need to show that the sequence of objective function values $\sum\limits_{j=1}^{\infty} c_j x_j(n)$ converges to $\sum\limits_{j=1}^{\infty} c_j \bar{x}_j$ as $n \to \infty$. Fix any $\epsilon > 0$. Let $0 \leq u \in R^\infty$ be as in Assumption 2.4. Since the series $\sum\limits_{i=1}^{\infty} |c_i|u_i$ of non-negative summands converges by Assumption 2.4, there exists an integer K such that the tail $\sum\limits_{i=k+1}^{\infty} |c_i|u_i < \epsilon/2$ for all $k \geq K$. Fix any such k and note that for any integer n,

$$\left|\sum_{j=1}^{\infty} c_j x_j(n) - \sum_{j=1}^{\infty} c_j \bar{x}_j\right| = \left|\sum_{j=1}^{\infty} c_j(x_j(n) - \bar{x}_j)\right| \leq \left|\sum_{j=1}^{k} c_j(x_j(n) - \bar{x}_j)\right| + \left|\sum_{j=k+1}^{\infty} c_j(x_j(n) - \bar{x}_j)\right|,$$

which is bounded above by

$$\sum_{j=1}^{k} |c_j||(x_j(n) - \bar{x}_j)| + \sum_{j=k+1}^{\infty} |c_j||(x_j(n) - \bar{x}_j)| \leq \sum_{j=1}^{k} |c_j||(x_j(n) - \bar{x}_j)| + \sum_{j=k+1}^{\infty} |c_j|u_j$$

because $0 \leq x_j(n) \leq u_j$ and $0 \leq \bar{x}_j \leq u_j$. The second term is strictly less than $\epsilon/2$. The first term is bounded above by $(\max\limits_{1\leq j\leq k} |(x_j(n) - \bar{x}_j)|)(\sum\limits_{j=1}^{k} |c_j|)$. The only interesting case is where $\sum\limits_{j=1}^{k} |c_j| \neq 0$. Since the sequence $\{x(n)\}_{n=1}^{\infty}$ converges to $\bar{x}$ componentwise, there exists an integer N_k large enough such that $(\max\limits_{1\leq j\leq k} |(x_j(n) - \bar{x}_j)|) < \frac{\epsilon}{\left(2\sum\limits_{j=1}^{k}|c_j|\right)}$ for all $n \geq N_k$. Therefore, $|\sum\limits_{j=1}^{\infty} c_j x_j(n) - \sum\limits_{j=1}^{\infty} c_j \bar{x}_j| < \epsilon$ for all integers $n \geq N_k$. Hence the objective function is continuous over F. (Note that an identical proof can be reproduced to reach the stronger conclusion that the objective function is continuous over $X = \{x \in R^{\infty} : 0 \leq x_j \leq u_j \ \ \forall j\}$). ☐

Note that the feasible region F of (P) is nonempty (by Assumption 2.1) convex (since it is the intersection of convex sets X^i defined in Lemma A.1 with the convex set $\{x \in R^{\infty} : x \geq 0\}$) and the product topology on R^{∞} is locally convex Hausdorff (Lemma 5.74 page 206 of [1]). Existence of an extreme point optimal solution to (P) then follows directly from Corollary A.2, Lemma A.3 and a Corollary (Corollary 7.70 page 299 of [1]) of the Bauer Maximum Principle (Theorem 7.69 page 298 of [1]).

A.3 Proof of Theorem 2.9

We use Berge's Maximum Theorem (Theorem 17.31 page 570 of [1]). Let $\mathcal{I}$ denote the set of extended positive integers $\{1, 2, \ldots\} \bigcup \{\infty\}$. Also let $X = \{x \in R^{\infty} : 0 \leq x_j \leq u_j \ \ \forall j\}$ where sequence $\{u_j\}$ is as in Assumption 2.4. Recall that $F \subseteq X$ by Assumption 2.3 and also that $F_N \subseteq X$ by the definition of projections of F in Equations (1) and (2). Now define a correspondence Ψ from $\mathcal{I}$ into X as

$$\Psi(N) = F_N \ \text{ for } \ N = 1, 2, \ldots$$

$$\Psi(\infty) = F.$$

Sets F_N are non-empty for all N since F is non-empty by Assumption 2.1. Similarly, F_N is also compact for each N implying that correspondence Ψ has non-empty compact values. Moreover, it is continuous by Lemma 2.8. Now define a function $f : \mathcal{I} \times X \to R$ as

$$f(N, x) = \sum_{i=1}^{N} c_i x_i \ \text{ for } \ N = 1, 2, \ldots, \ \ x \in X$$

$$f(\infty, x) = \sum_{i=1}^{\infty} c_i x_i \ \text{ for } \ x \in X.$$

Function f is continuous. To see this, fix $\epsilon > 0$ and suppose $x^k \to x$ in X and $N_k \to \infty$ as $k \to \infty$.

$$\begin{aligned} |f(\infty, x) - f(N_k, x^k)| &= \left|\sum_{i=1}^{\infty} c_i x_i - \sum_{i=1}^{N_k} c_i x_i^k\right| \leq \left|\sum_{i=1}^{\infty} c_i x_i - \sum_{i=1}^{\infty} c_i x_i^k\right| + \left|\sum_{i=1}^{\infty} c_i x_i^k - \sum_{i=1}^{N_k} c_i x_i^k\right| \\ &= \left|\sum_{i=1}^{\infty} c_i x_i - \sum_{i=1}^{\infty} c_i x_i^k\right| + \left|\sum_{i=N_k+1}^{\infty} c_i x_i^k\right| \leq \left|\sum_{i=1}^{\infty} c_i x_i - \sum_{i=1}^{\infty} c_i x_i^k\right| + \sum_{i=N_k+1}^{\infty} |c_i x_i^k| \\ &\leq \left|\sum_{i=1}^{\infty} c_i x_i - \sum_{i=1}^{\infty} c_i x_i^k\right| + \sum_{i=N_k+1}^{\infty} |c_i| u_i \ \text{ because } x^k \in X. \end{aligned}$$

Recall from the proof of Lemma A.3 that the objective function is continuous over X. Hence $\sum\limits_{i=1}^{\infty} c_i x_i^k$ converges to $\sum\limits_{i=1}^{\infty} c_i x_i$ because $x^k \to x$ as $k \to \infty$. Thus the first term in the above upper bound can be made smaller than $\epsilon/2$ by choosing k large enough. The second term also can be made smaller than $\epsilon/2$ for k large enough by Assumption 2.4. Therefore, there exists an integer K such that for all $k \geq K$, $|f(\infty, x) - f(N_k, x^k)| < \epsilon$. Thus f is continuous.

Now define the "value function" $m : \mathcal{I} \to R$ as follows.

$$m(N) = \min_{x \in F_N} \sum_{i=1}^{N} c_i x_i = \max_{x \in \Psi(N)} -f(N, x), \quad N = 1, 2, \ldots$$

and

$$m(\infty) = \min_{x \in F} \sum_{i=1}^{\infty} c_i x_i = \max_{x \in \Psi(\infty)} -f(\infty, x).$$

From the first part of Berge's Maximum Theorem, the value function is continuous, i.e., $m(N) \to m(\infty)$ as $N \to \infty$. This proves the first claim in Theorem 2.9. For the second claim, define the "argmax" correspondence μ from $\mathcal{I}$ into X as

$$\begin{aligned} \mu(N) &= \{x \in \Psi(N) : f(N, x) = m(N)\} \equiv F_N^*, \quad N = 1, 2, \ldots \\ \mu(\infty) &= \{x \in \Psi(\infty) : f(\infty, x) = m(\infty)\} \equiv F^*. \end{aligned}$$

The second claim then follows from the second part of Berge's Maximum Theorem.

A.4 Proof of Lemma 3.6

For the "if" part, let $T_N = F_{L_N}$. Then T_N is extendable by definition of F_{L_N}. For the "only if" part, suppose $T_N \neq F_{L_N}$. Then there is some $x \in T_N$ that is not in F_{L_N}. In particular, there is no $y \in F$ whose first L_N components match with x. Thus T_N is not extendable.

A.5 Proof of Lemma 4.1

We constructively show that (DP) is feasible. Every infinite-horizon feasible state by definition has at least one feasible action. We inductively construct a feasible solution to (DP) by choosing exactly one feasible action in each state in each period and setting the corresponding z variable to a positive value so as to satisfy the equality constraints. Specifically, pick $a_1 \in A(s_1)$ and set $z(s_1, a_1) = \beta(s_1)$. Set $z(s_1, a) = 0$ for all $a \in A(s_1)$ different from a_1. Suppose choosing action a_1 in state s_1 transforms the system to state s_2 in period 2. Pick $a_2 \in A(s_2)$ and set $z(s_2, a_2) = \beta(s_2) + \alpha z(s_1, a_1) = \beta(s_2) + \alpha\beta(s_1)$. Also set $z(s_2, a) = 0$ for all $a \in A(s_2)$ different from a_2. Continuing this procedure ad infinitum yields a feasible solution to (DP) satisfying Assumption 2.1. Let Θ be a uniform upper bounded on cardinalities $|S_n|$, Λ a uniform upper bound on cardinalities $|A(s_n)|$. It is easy to see that every equality constraint has at most $\Lambda + \Theta$ variables hence Assumption 2.2 holds. We now claim that for every feasible solution z to (DP) and every period $n = 1, 2, \ldots$,

$$\sum_{s_n \in S_n} \sum_{a_n \in A(s_n)} z(s_n, a_n) = \alpha^{n-1}\beta(s_1) + \alpha^{n-2} \sum_{s_2 \in S_2} \beta(s_2) + \ldots + \alpha \sum_{s_{n-1} \in S_{n-1}} \beta(s_{n-1}) + \sum_{s_n \in S_n} \beta(s_n).$$

We prove this claim by induction on n. The claim is true for $n = 1$ since $S_1 = \{s_1\}$ and $\sum\limits_{a_1 \in A(s_1)} z(s_1, a_1) = \beta(s_1)$ from the equality constraint since $X(s_1) = \emptyset$. Suppose the claim is true for some period n. Then the equality constraint in (DP) implies that

$$\begin{aligned} \sum_{s_{n+1} \in S_{n+1}} \sum_{a_{n+1} \in A(s_{n+1})} z(s_{n+1}, a_{n+1}) &= \sum_{s_{n+1} \in S_{n+1}} \beta(s_{n+1}) + \alpha \sum_{s_n \in S_n} \sum_{a_n \in A(s_n)} z(s_n, a_n) \\ &= \sum_{s_{n+1} \in S_{n+1}} \beta(s_{n+1}) + \alpha^n \beta(s_1) + \ldots + \alpha \sum_{s_n \in S_n} \beta(s_n), \end{aligned}$$

where the last equality follows from the inductive hypothesis. This restores our inductive hypothesis proving the claim. Non-negativity of z then implies that

$$z(s_n,a_n) \le \alpha^{n-1}\beta(s_1)+\alpha^{n-2}\sum_{s_2\in S_2}\beta(s_2)+\ldots+\alpha\sum_{s_{n-1}\in S_{n-1}}\beta(s_{n-1})+\sum_{s_n\in S_n}\beta(s_n)$$

for all $s_n\in S_n$, $a_n\in A(s_n)$ and all n. Hence Assumption 2.3 holds with

$$u(s_n,a_n) = \alpha^{n-1}\beta(s_1)+\alpha^{n-2}\sum_{s_2\in S_2}\beta(s_2)+\ldots+\alpha\sum_{s_{n-1}\in S_{n-1}}\beta(s_{n-1})+\sum_{s_n\in S_n}\beta(s_n).$$

Notice that components of u depend only on the time period and not on s_n and a_n (since s_1 is fixed). As a result, Assumption 2.4 holds since costs are in l_∞ as $0\le c_n(s_n,a_n)\le c<\infty$ and u is in l_1. To see that $u\in l_1$ note that $\sum\limits_{n=1}^{\infty}\sum\limits_{s_n\in S_n}\sum\limits_{a_n\in A(s_n)} u(s_n,a_n)$ is bounded above as

$$\begin{aligned}
&\le \Theta\Lambda\sum_{n=1}^{\infty}\left(\alpha^{n-1}\beta(s_1)+\alpha^{n-2}\sum_{s_2\in S_2}\beta(s_2)+\ldots+\sum_{s_{n-1}\in S_{n-1}}\alpha\beta(s_{n-1})+\sum_{s_n\in S_n}\beta(s_n)\right)\\
&= \Theta\Lambda\left(\beta(s_1)+\ldots+\sum_{s_n\in S_n}\beta(s_n)+\ldots\right)\left(\sum_{n=1}^{\infty}\alpha^{n-1}\right)=\left(\sum_{n=1}^{\infty}\sum_{s_n\in S_n}\beta(s_n)\right)\frac{\Theta\Lambda}{1-\alpha}<\infty.
\end{aligned}$$

Here the discount factor α, our choice of $\beta(s_n)$, and the inherent structure of our dynamic programs help us embed u in l_1. Recall that the discount factor appears in the constraints rather than with the costs in linear programming formulations of dynamic programs as in problem (DP). Thus even though the costs are discounted, the cost coefficients in the linear objective function are in l_∞, unlike say the production planning problem $(PROD)$ discussed in the paper. These same structural features will also prove helpful in deriving inequality (3) critical for our "big-M" method below.

A.6 Proof of Lemma 4.2

Let z be a feasible solution to $DP(N)$ and let s_{N+1} be any terminal state of the N-horizon truncation of our original dynamic program. Owing to our extendability of finite-horizon strategies assumption, any finite sequence of actions that terminates in s_{N+1} has an infinite-horizon feasible continuation. We append z along the state-action pairs of this continuation respecting equality constraints and non-negativity to construct a feasible solution to (DP). The detailed procedure is similar to the one used in showing that (DP) has a feasible solution and hence is omitted.

A.7 A brief outline of the "big-M" approach for dynamic programs

We modify the original non-stationary infinite-horizon dynamic program defined in Section 4 as follows. Include an artificial action $\nu(s_n)$ feasible in state $s_n\in S_n$ for $n=1,2,\ldots$. Similarly, include an artificial state Δ_n feasible in period n and a corresponding feasible action μ_n for $n=2,3,\ldots$. We set $g_n(s_n,\nu(s_n))=\Delta_{n+1}$ and $g_n(\Delta_n,\mu_n)=\Delta_{n+1}$. See Figure 5. Let $c_n(s_n,\nu(s_n))=c_n(\Delta_n,\mu_n)=M$ for some arbitrarily large number M. Notice that extendability of finite-horizon strategies holds in this "artifical" dynamic program. Let $\gamma(\Delta_n)$ be a sequence of positive numbers indexed by artificial states Δ_n such that $\sum\limits_{n=2}^{\infty}\gamma(\Delta_n)<\infty$. Then the CILP corresponding to the artificial dynamic program is given by

$$(DP_M)\quad \min\sum_{n=1}^{\infty}\sum_{s_n\in S_n}\sum_{a_n\in A(s_n)}c_n(s_n,a_n)z(s_n,a_n)+\sum_{n=1}^{\infty}\sum_{s_n\in S_n}My(s_n,\nu(s_n))+\sum_{n=1}^{\infty}Mw(\Delta_n,\mu_n)$$

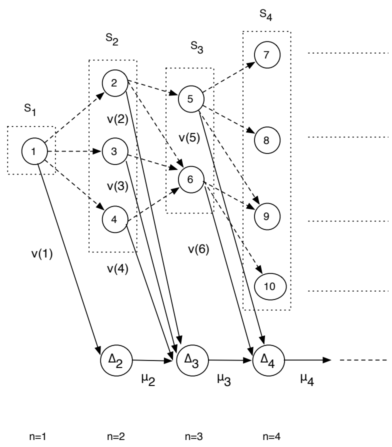

Figure 5: A portion of the dynamic programming network for the artificial dynamic program corresponding to the dynamic program in Figure 3. The artificial states and actions are shown with solid arrows for emphasis. Artificial action labels are included next to the arrows.

$$\begin{aligned}
y(s_n,\nu(s_n)) + \sum_{a_n \in A(s_n)} z(s_n,a_n) - \alpha \sum_{(s_{n-1},a_{n-1})\in X(s_n)} z(s_{n-1},a_{n-1}) &= \beta(s_n), \;\; s_n \in S_n, \; \forall n, \\
w(\Delta_n,\mu_n) - \alpha w(\Delta_{n-1},\mu_{n-1}) - \alpha \sum_{s_{n-1}\in S_{n-1}} y(s_{n-1},\nu(s_{n-1})) &= \gamma(\Delta_n), \;\; n = 2,3,\ldots \\
z(s_n,a_n) \geq 0, \; s_n \in S_n, \; a_n \in A(s_n), \; \forall n, \;\; y(s_n,\nu(s_n)) &\geq 0, \; s_n \in S_n, \; \forall n \\
w(\Delta_n,\mu_n) &\geq 0, \; n = 2,3\ldots.
\end{aligned}$$

Note that the variables in (DP_M) include the original variables z in (DP) as well as the artificial variables y and w. By replicating the proof of Lemma 4.1 one can confirm that (DP_M) satisfies Assumptions 2.1-2.4. Hence (DP_M) has an optimal solution. More importantly, all finite-horizon truncations of (DP_M) are extendable by an argument similar to Lemma 4.2 because finite-horizon strategies in the artificial dynamic program are extendable. Notice that if solution z is feasible to (DP) then it is also feasible to (DP_M) by setting $y(s_n,\nu(s_n)) = 0$ for all $s_n \in S_n$, $n = 1,2,\ldots$, and $w(\Delta_n,\mu_n) = \gamma(\Delta_n) + \sum_{i=2}^{n-1} \alpha^{n-i}\gamma(\Delta_i) \equiv \hat{w}(\Delta_n,\mu_n)$ for $n = 2,3,\ldots$. Suppose $\bar{z}$, $\bar{y}$, $\bar{w}$ is an optimal solution to (DP_M) and suppose $\bar{y}(s_n,\nu(s_n)) > 0$ for some s_n and $\nu(s_n)$. Then $\bar{w}(\Delta_n,\mu_n) \geq \hat{w}(\Delta_n,\mu_n)$ for all n. Then as in the "big-M" method for finite-dimensional linear programs (Exercise 3.26 of [11]) it is easy to prove that (DP) is infeasible. For if it is not, and if z is feasible to (DP), then $\sum_{n=1}^{\infty}\sum_{s_n\in S_n}\sum_{a_n\in A(s_n)} c_n(s_n,a_n)z(s_n,a_n) + \sum_{n=1}^{\infty} M\hat{w}(\Delta_n,\mu_n)$ is at least

$$\sum_{n=1}^{\infty}\sum_{s_n\in S_n}\sum_{a_n\in A(s_n)} c_n(s_n,a_n)\bar{z}(s_n,a_n) + \sum_{n=1}^{\infty}\sum_{s_n\in S_n} M\bar{y}(s_n,\nu(s_n)) + \sum_{n=1}^{\infty} M\bar{w}(\Delta_n,\mu_n),$$

owing to feasibility of z, $y = 0$, $\hat{w}$ and optimality of $\bar{z}$, $\bar{y}$ and $\bar{w}$ to (DP_M) respectively. Therefore, $\sum_{n=1}^{\infty}\sum_{s_n\in S_n}\sum_{a_n\in A(s_n)} c_n(s_n,a_n)z(s_n,a_n)$ is bounded below as

$$\geq \sum_{n=1}^{\infty}\sum_{s_n\in S_n}\sum_{a_n\in A(s_n)} c_n(s_n,a_n)\bar{z}(s_n,a_n) + \sum_{n=1}^{\infty}\sum_{s_n\in S_n} M\bar{y}(s_n,\nu_n) + \sum_{n=1}^{\infty} M\left(\bar{w}(\Delta_n,\mu_n) - \hat{w}(\Delta_n,\mu_n)\right)$$

$$\geq \sum_{n=1}^{\infty}\sum_{s_n\in S_n} M\bar{y}(s_n,\nu_n) + \sum_{n=1}^{\infty} M\left(\bar{w}(\Delta_n,\mu_n) - \hat{w}(\Delta_n,\mu_n)\right),$$

where the second inequality uses non-negativity of $c_n(s_n, a_n)$ and $\bar{z}(s_n, a_n)$. Let $n \geq 1$ be the smallest period for which there exist s_n and $\nu(s_n)$ such that $\bar{y}(s_n, \nu(s_n)) > 0$. This implies that $\bar{w}(\Delta_{n+1}, \mu_{n+1}) - \hat{w}(\Delta_{n+1}, \mu_{n+1}) > \alpha\bar{y}(s_n, \nu(s_n))$. Then using an upper bound from Lemma 4.1 on the left hand side we get $\frac{c\Theta\Lambda}{1-\alpha}\left(\sum\limits_{n=1}^{\infty}\sum\limits_{s_n\in S_n}\beta(s_n)\right) \geq M(\bar{y}(s_n, \nu(s_n)) + \alpha\bar{y}(s_n, \nu(s_n)))$, i. e.,

$$\frac{\frac{c\Theta\Lambda}{1-\alpha}\left(\sum\limits_{n=1}^{\infty}\sum\limits_{s_n\in S_n}\beta(s_n)\right)}{\bar{y}(s_n, \nu(s_n)) + \alpha\bar{y}(s_n, \nu(s_n))} \geq M, \tag{3}$$

which is a contradiction since M is arbitrarily large. Thus Assumption 2.1 implies $\bar{y} = 0$. This also implies that $\bar{w} = \hat{w}$. Moreover, in that case, $\bar{z}$ is in fact optimal to (DP). In summary, (DP_M) satisfies all assumptions required for Shadow Simplex and is equivalent to (DP). This also shows that our assumption that finite-horizon strategies are extendable is without loss of generality.

A.8 Proof of Lemma 4.3

First recall that a point in a convex set is its extreme point if it cannot be expressed as a convex combination of two other distinct points in the convex set. Let z be as in the hypothesis of the lemma. Suppose by way of contradiction that z is not an extreme point. Then there exist feasible solutions x and y and a fraction $0 < \lambda < 1$ such that $z = \lambda x + (1-\lambda)y$. We show that $x = y = z$. Non-negativity of x and y implies that for all states s_n and actions $a_n \in A(s_n)$ for which $z(s_n, a_n) = 0$, $x(s_n, a_n) = y(s_n, a_n) = 0$. Then one can confirm using simple algebra starting at state s_1 and working inductively that under this restriction the equality constraints in (DP) have a unique solution, namely, z. This implies $x = y = z$.

A.9 Proof of Lemma 4.4

Suppose z is a feasible solution to (DP) and there exists a state s_n (called a "bad" state) with (at least) two actions $a_n \in A(s_n)$ and $b_n \in A(s_n)$ such that $z(s_n, a_n) > 0$ and $z(s_n, b_n) > 0$. Let $s_{n+1} = g_n(s_n, a_n)$ and $t_{n+1} = g_n(s_n, b_n)$ and note that $s_{n+1} \neq t_{n+1}$ since two distinct actions lead to two distinct states. In this proof, we consider two types of state-action sequences whose existence follows from the structure of problem (DP). The first is of the form $\{(s_r, a_r)\}_{r=n+1}^{\infty}$ starting at s_{n+1} such that for all r, $s_r \in S_r$, $a_r \in A(s_r)$, $s_{r+1} = g_r(s_r, a_r)$ and $z(s_r, a_r) > 0$. The set of all such sequences is denoted $\Omega(s_n, a_n)$. The second is of the form $\{(t_r, b_r)\}_{r=n+1}^{\infty}$ starting at t_{n+1} such that for all r, $t_r \in S_r$, $b_r \in A(t_r)$, $t_{r+1} = g_r(t_r, b_r)$ and $z(t_r, b_r) > 0$. The set of all such sequences is denoted $\Omega(s_n, b_n)$. We say that a sequence from $\Omega(s_n, a_n)$ "passes through" a particular state, or this particular state "belongs to" the sequence if it is included in a state-action pair that is in the sequence. Similarly for sequences in $\Omega(s_n, b_n)$. We also use this terminology for state-action pairs. We consider two possible cases and in both these show that it is possible to construct two distinct solutions x and y feasible to (DP) such that $z = (x + y)/2$ and hence z cannot be an extreme point to complete the proof by contrapositive.

Case 1: There exists a bad state s_n for which a sequence from $\Omega(s_n, a_n)$ passes through a state that also belongs to some sequence in $\Omega(s_n, b_n)$.

Let s_{n+k} be the first state that these two sequences have in common for some $k > 1$, that is, $s_{n+k} = t_{n+k}$, and $s_{n+j} \neq t_{n+j}$ for $j = 2, 3 \ldots, k-1$. Specifically, let $\epsilon(a_n) > 0$ be the largest amount that can be subtracted from $z(s_n, a_n)$, reducing values of $z(s_{n+1}, a_{n+1}), \ldots, z(s_{n+k-1}, a_{n+k-1})$ in order to satisfy the equality constraints in (DP) at the same time forcing these variables to be non-negative. Similarly, let $\epsilon(b_n) > 0$ be the largest amount that can be subtracted from $z(s_n, b_n)$, reducing values of $z(t_{n+1}, b_{n+1}), \ldots, z(s_{n+k-1}, b_{n+k-1})$ in order to satisfy the equality constraints in (DP) at the same time forcing these variables to be non-negative. Set $\epsilon = \min\{\epsilon(a_n), \epsilon(b_n)\}$. Let x be the feasible solution formed from z by subtracting ϵ from $z(s_n, a_n)$, reducing values of $z(s_{n+1}, a_{n+1}), \ldots, z(s_{n+k-1}, a_{n+k-1})$, adding ϵ to $z(s_n, b_n)$, and increasing values of $z(t_{n+1}, b_{n+1})$, $\ldots$, $z(t_{n+k-1}, b_{n+k-1})$ to satisfy equality constraints. Similarly, let y be the feasible solution formed from z by adding ϵ to $z(s_n, a_n)$, increasing values of $z(s_{n+1}, a_{n+1}), \ldots, z(s_{n+k-1}, a_{n+k-1})$, subtracting ϵ from $z(s_n, b_n)$, and reducing values of $z(t_{n+1}, b_{n+1}), \ldots, z(t_{n+k-1}, b_{n+k-1})$ to satisfy equality constraints. Then it is easy to check that $z = (x + y)/2$ and hence z is not an extreme point.

Case 2: There is no bad state s_n for which a sequence from $\Omega(s_n, a_n)$ and a sequence from $\Omega(s_n, b_n)$ both pass through a common state.

Consider any bad state s_n and corresponding actions $a_n \in A(s_n)$ and $b_n \in A(s_n)$ such that $z(s_n, a_n) > 0$, and $z(s_n, b_n) > 0$. For any sequence $\zeta \in \Omega(s_n, a_n)$, and any period $N \geq n+1$, ζ_N denotes the state-action pair that ζ passes through in period N and thus we can write $z(\zeta_N)$. Define $\phi^z_N(s_n, a_n) = \sum\limits_{\zeta \in \Omega(s_n, a_n)} z(\zeta_N)$.

Suppose without loss of generality that $z(s_n, a_n) \geq z(s_n, b_n)$. Note that $\phi^z_N(s_n, a_n) \geq \alpha^{(N-n)} z(s_n, a_n)$, which in turn is at least $\alpha^{(N-n)} z(s_n, b_n)$, and also that $\phi^z_N(s_n, b_n) \geq \alpha^{(N-n)} z(s_n, b_n)$. Let x be a new feasible solution formed as follows. $x(s_n, a_n) = z(s_n, a_n) - z(s_n, b_n)$ and $x(\zeta_N) = z(\zeta_N) - \epsilon(\zeta_N)$ for some $0 \leq \epsilon(\zeta_N) \leq z(\zeta_N)$ for all $\zeta_N \in \Omega(s_n, a_n)$ for all $N \geq n+1$. These $\epsilon(\zeta_N)$ are chosen so that $\phi^\epsilon_N(s_n, a_n) = \alpha^{(N-n)} z(s_n, b_n)$ satisfying (DP) equality constraints at every state that sequences ζ pass through. Moreover, $x(s_n, b_n) = 2z(s_n, b_n)$ and $x(\xi_N) = z(\xi_N) + \delta(\xi_N)$ for some $0 \leq \delta(\zeta_N) \leq z(\xi_N)$ for all $\xi_N \in \Omega(s_n, b_n)$ for all $N \geq n+1$. These $\delta(\xi_N)$ are chosen so that $\phi^\delta_N(s_n, b_n) = \alpha^{(N-n)} z(s_n, b_n)$ satisfying (DP) equality constraints at every state that sequences ξ pass through. Similarly, y is a new feasible solution formed so that $y(s_n, a_n) = z(s_n, a_n) + z(s_n, b_n)$ and $y(\zeta_N) = z(\zeta_N) + \epsilon(\zeta_N)$. Moreover, $y(s_n, b_n) = 0$ and $y(\xi_N) = z(\xi_N) - \delta(\xi_N)$. All the other components of x and y are equal to the corresponding components of z implying $z = (x + y)/2$.

A.10 Proof of Lemma 5.1

We first show constructively that (MDP) has a feasible solution. The procedure is similar to problem (DP). The first equality constraint is simply $\sum\limits_{a_1 \in A(s_1)} z(s_1, a_1) = \beta(s_1)$ since $X(s_1) = \emptyset$. Thus we arbitrarily choose one action $a_1 \in A(s_1)$ and set $z(s_1, a_1) = \beta(s_1)$. For all other actions in $a \in A(s_1)$, we set $z(s_1, a) = 0$. Then, the equality constraint corresponding to every $s_2 \in S_2$ reduces to $\sum\limits_{a_2 \in A(s_2)} z(s_2, a_2) = \alpha p_1(s_2|s_1, a_1) z(s_1, a_1) + \beta(s_2)$. We therefore choose an arbitrary action $a_2 \in A(s_2)$ and set $z(s_2, a_2) = \alpha p_1(s_2|s_1, a_1) z(s_1, a_1) + \beta(s_2)$. That is, set $z(s_2, a_2) = \alpha p_1(s_2|s_1, a_1)\beta(s_1) + \beta(s_2)$. For all other actions $a \in A(s_2)$ we set $z(s_2, a) = 0$. Continuing this way we can construct a feasible solution to (MDP) thus satisfying Assumption 2.1. Every equality constraint has at most $\Lambda + \Lambda\Theta$ variables hence Assumption 2.2 holds. Similar to problem (DP) we now claim that for any n, $\sum\limits_{s_n \in S_n} \sum\limits_{a_n \in A(s_n)} z(s_n, a_n) = \alpha^{n-1}\beta(s_1) + \alpha^{n-2} \sum\limits_{s_2 \in S_2} \beta(s_2) + \ldots + \sum\limits_{s_n \in S_n} \beta(s_n)$. We prove this by induction. The claim is true for $n = 1$ as $S_1 = \{s_1\}$ and $X(s_1) = \emptyset$ implying that $\sum\limits_{a_1 \in A(s_1)} z(s_1, a_1) = \beta(s_1)$. Now suppose it is true for some period n. Then owing to the equality constraint, $\sum\limits_{s_{n+1} \in S_{n+1}} \sum\limits_{a_{n+1} \in A(s_{n+1})} z(s_{n+1}, a_{n+1})$ equals

$$\sum_{s_{n+1} \in S_{n+1}} \beta(s_{n+1}) + \alpha \sum_{s_{n+1} \in S_{n+1}} \sum_{s_n \in X(s_{n+1})} \sum_{a_n \in \mathcal{X}(s_n, s_{n+1})} p_n(s_{n+1}|s_n, a_n) z(s_n, a_n)$$

$$= \sum_{s_{n+1} \in S_{n+1}} \beta(s_{n+1}) + \alpha \sum_{s_n \in S_n} \sum_{a_n \in A(s_n)} z(s_n, a_n) = \sum_{s_{n+1} \in S_{n+1}} \beta(s_{n+1}) + \alpha^n \beta(s_1) + \ldots + \alpha \sum_{s_n \in S_n} \beta(s_n),$$

where the last equality follows from the inductive hypothesis. This restores the inductive hypothesis proving our claim. The rest of the proof is identical to (DP) and hence is omitted.

A.11 Proof of Lemma 5.2

Similar to the (DP) case hence omitted.