OPERATIONS RESEARCH
DOI 10.1287/opre.1080.0686ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"A Stochastic Multiple-Leader Stackelberg Model: Analysis, Computation, and Application" by Victor DeMiguel and Huifu Xu, *Operations Research*, DOI 10.1287/opre.1080.0686.

Material for on-line supplement

Appendix A: Proofs for some of the results

Proof for Lemma 3.3

Part (i). For each realization of the random variable t, the reaction function $Y(X,t)$ behaves as that of a deterministic problem with multiple leaders and followers. Moreover, it does not make any difference to the followers equilibrium whether X is produced by a single leader or multiple leaders. Hence, the result follows from the analogous result for the model with deterministic demand and a single leader, which is proven in H. D. Sherali [1983, Theorem 2].

Part (ii). Let $t \in \mathcal{T}$ and $y(X,t)$ be the follower equilibrium. We consider the complementarity problem

$$y_j(X,t) \geq 0, G_j(X,y(X,t),t) \geq 0, y_j(X,t)G_j(X,y(X,t),t) = 0.$$

It suffices to show the conclusion at points where $y_j(X,t) > 0$ as $y_i(X,t)$ is a continuous non-negative function. By the complementarity condition $G_j(X,y(X,t),t) = 0$; that is,

$$-p\left(X + Y(X,t),t\right) - y_j(X,t)p'_q\left(X + Y(X,t),t\right) + c'_j(y_j(X,t)) = 0. \tag{20}$$

By using Clarke's generalized implicit function theorem [Xu 2005, Lemma 3.2], we obtain

$$\partial_X y_j(X,t) \subset \frac{\left[p'_q\left(X + Y(X,t),t\right) + p''_{qq}\left(X + Y(X,t),t\right) y_j(X,t)\right]\left(1 + \partial_X Y(X,t)\right)}{-p'_q\left(X + Y(X,t),t\right) + c''_j(y_j(X,t))}. \tag{21}$$

Note that at a point where $Y(X,t)$ is differentiable, both $\partial_X Y(X,t)$ and $\partial_X y_j(X,t)$ reduce to a singleton.

To show the conclusion, we note that part (i) of this lemma indicates that $1 + \partial_X Y(X,t) \subset (0,1]$. Part (i) of Lemma 2.4 indicates that

$$p'_q\left(X + Y(X,t),t\right) + p''_{qq}\left(X + Y(X,t),t\right) y_j(X,t) \leq 0.$$

By Assumptions 3.1 (i),

$$-p'_q\left(X + Y(X,t),t\right) + c''_j(y_j(X,t)) \geq \sigma > 0.$$

Thus every element of Clarke subdifferential $\partial_X y_j(X,t)$ is non-positive, which implies that $y_j(X,t)$ is nonincreasing with respect to X.

Part (iii). From Proposition 2.6 (iii), we know that $y_j(X,t)$ is piecewise smooth in X. Let $X^{(1)}, X^{(2)} \geq 0$ be any two points. By the mean value theorem,

$$y_j(X^{(1)},t) - y_j(X^{(2)},t) = \int_0^1 (y_j)'_X\left(X^{(2)} + \theta(X^{(1)} - X^{(2)}),t\right)(X^{(1)} - X^{(2)})d\theta.$$

Under Assumption 3.1, we have from (21) that

$$\left|(y_j)'_X\left(X^{(2)} + \theta(X^{(1)} - X^{(2)}),t\right)\right| \leq (1 + y_j^u)L_1(t)/\sigma.$$

The conclusion follows by taking $L_2(t) := L_1(t)(1 + y_j^u)/\sigma$. $\square$

Proof for Proposition 4.2

Observe that Assumptions 2.2, 2.3, 2.5, and 3.2 are satisfied and hence all conditions of Proposition 2.6 hold which means that $Y(X,t)$ is well defined.

Condition (i). We first prove the conclusion under condition (i). The method of proof is similar to that of [Sherali 1984, Theorem 3] except we use the complementarity reformulation.

Let $X(t)$ denote a point at which the supply of one or more of the followers $y_j(X,t)$ turns from strictly positive to zero as X increases. From Proposition 2.6 we have that $Y(X,t)$ is piecewise smooth with respect to X, where the only points at which $Y(X,t)$ may not be smooth are precisely the points $X(t)$. For points X at which $Y(X,t)$ is differentiable, it is easy to see, from the linearity of the demand function in X and the fact that the follower cost functions are quadratic, that $y_j(X,t)$ is linear in X for all j. Hence, $y_j(X,t)$ is convex in a neighborhood of such points.

It only remains to show that $Y(X,t)$ is convex in X in a neighborhood of points $X(t)$ at which $y_j(X,t)$ turns from strictly positive to zero for one or more followers. Let $\mathcal{I}(X,t)$ denote the index set of the followers with $y_j(X,t)>0$. Then $\mathcal{I}(X(t)_-,t)\backslash\mathcal{I}(X(t)_+,t)$ is the index set of followers who turn from positive to zero at $X(t)$, where

$$\mathcal{I}(X(t)_-,t)=\lim_{\delta\to 0}\mathcal{I}(X(t)-\delta,t),\ \mathcal{I}(X(t)_+,t)=\lim_{\delta\to 0}\mathcal{I}(X(t)+\delta,t).$$

Because $y_j(X,t)$ is piecewise smooth in X (nonsmooth only at a finite number of points), we may assume that in a neighborhood of $X(t)$, $y_j(X,t)$ is differentiable except at $X(t)$. Because $p(q,t)$ is linear in X, we have from the complementarity formulation of the followers equilibrium that $G_j(X,y(X,t),t)=0$ for X in a *left* neighborhood of $X(t)$—because in this left neighborhood we have $y_j(X,t)>0$. That is

$$-\alpha(t)+\beta(t)(X+Y(X,t))+\beta(t)y_j(X,t)+c_j'(y_j(X,t))=0. \tag{22}$$

Differentiating the above equation with respect to X, we obtain

$$(\beta(t)+c_j''(y_j(X,t))(y_j)_X'(X,t)=-\beta(t)(1+Y_X'(X,t)). \tag{23}$$

Because c_j is quadratic, $c_j''(y_j(X,t))$ is a constant. To simplify notation, let η_j denote $c_j''(y_j(X,t))$. Then we have from (23)

$$(y_j)_X'(X,t)=-\frac{\beta(t)}{\beta(t)+\eta_j}(1+Y_X'(X,t)).$$

By adding this equation with respect to $j\in\mathcal{I}(X,t)$ and observing that

$$Y_X'(X,t)=\sum_{j\in\mathcal{I}(X,t)}(y_j)_X'(X,t),$$

we have

$$Y_X'(X,t)=-\sum_{j\in\mathcal{I}(X,t)}\frac{\beta(t)}{\beta(t)+\eta_j}(1+Y_X'(X,t)).$$

This can be rewritten as

$$Y'_X(X,t) = -\frac{u(X,t)}{1+u(X,t)},$$

where

$$u(X,t) := \sum_{j\in\mathcal{I}(X,t)} \frac{\beta(t)}{\beta(t)+\eta_j}.$$

Let $|\mathcal{I}(X,t)|$ denote the cardinality of the set $\mathcal{I}(X,t)$, then we obviously have $|\mathcal{I}(X(t)_-,t)| \geq |\mathcal{I}(X(t)_+,t)|$ and hence $u(X,t)$ decreases when X changes from $X(t)_-$ to $X(t)_+$, while $Y'_X(\cdot,t)$ increases. This shows the convexity of $Y(X,t)$ with respect to X.

Condition (ii). We now prove the conclusion under condition (ii). For the case where all followers have the same cost function, the production of all followers at equilibrium is identical $y_j(X,t) = y(X,t)$ for $j = 1,\cdots,N$. At a point X_1 such that $y(X_1,t) = 0$ we have that $y(X_2,t) = 0$ for all $X_2 > X_1$—this is true because from Lemma 3.3 (given that Assumptions 2.2, 2.3, 2.5, 3.1, and 3.2 hold) we have that $y(X,t)$ is nonincreasing in X and it must be nonnegative. Hence, clearly $y(X,t)$ is convex (equal to zero) for all $X \geq X_1$.

It only remains to show convexity at points X such that $y(X,t) > 0$ or at points X at which $y(X,t)$ turns from strictly positive to zero as X increases. For this points, we have from the complementarity reformulation of the followers equilibrium that

$$-\alpha(t) + \beta(t)(X + My(X,t)) + \beta(t)y(X,t) + c'(y(X,t)) = 0.$$

Differentiating the equation above with respect to X twice, we obtain

$$\beta(t)(M+1)y''_X(X,t) + c'''(y(X,t))(y'_X(X,t))^2 + c''(y(X,t))y''_X(X,t) = 0,$$

from which we derive $y''_X(X,t) \geq 0$ because by assumption $c'''(y) \leq 0$. □

Proof for Theorem 4.4

Theorem 4.3 shows that there exists an equilibrium. It only remains to show uniqueness. Because by Theorem 3.5 the leader objective functions $\phi_i(x_i, X_{-i})$ are twice continuously differentiable, we can apply Theorems 2 and 6 in Rosen [1965]. Hence, we only need to show that the Jacobian matrix of the function $g(x,r)$ is negative definite for all $x \in [0,x_1^u] \times \cdots \times [0,x_M^u]$. We proof the results in three steps.

STEP 1. *Derive an expression for the Jacobian of $g(x,r)$.* Note that under the conditions of Theorem 3.5, the functions ϕ_i are twice continuously differentiable. Moreover, note that

$$\begin{aligned}(\phi_i)'_{x_i}(x_i,X_{-i}) = &\int_{t\in\mathcal{T}} [p\left(x_i + X_{-i} + Y(x_i+X_{-i},t),t\right) \\ &+ x_i p'_q\left(x_i + X_{-i} + Y(x_i+X_{-i},t),t\right)\left(1+Y'_X(x_i+X_{-i},t)\right)]\rho(t)dt \\ &- C'_i(x_i),\end{aligned}$$

and

$$(\phi_i)''_{x_ix_j}(x_i,X_{-i}) = \begin{cases} \int_{t\in\mathcal{T}} [2p'_q(1+Y'_X) + x_i p''_{qq}(1+Y'_X)^2 + x_i p'_q Y''_X]\rho(t)dt - C''_i(x_i), & \text{if } j = i, \\ \int_{t\in\mathcal{T}} [p'_q(1+Y'_X) + x_i p''_{qq}(1+Y'_X)^2 + x_i p'_q Y''_X]\rho(t)dt, & \text{if } j \neq i. \end{cases}$$

Observe that for a given leader i, we have that $(\phi_i)''_{x_i x_k}(x_i, X_{-i}) = (\phi_i)''_{x_i x_j}(x_i, X_{-i})$ for all $j, k \neq i$. This implies that all the components of the vector $\nabla_x (\phi_i)'_{x_i}(x_i, x_{-i})$ are equal except for the i-th component.

Consequently, using the quantities ζ_i and δ_i defined in the statement of this theorem, we can obtain the following expression for the Jacobian matrix of the function $g(x, r)$

$$\nabla g(x, r) = (r_1\zeta_1, \cdots, r_M\zeta_M)^T \mathbf{e}^T + \mathrm{diag}(r_1\delta_1, \cdots, r_M\delta_M), \tag{24}$$

where $\mathrm{diag}(r_1\delta_1, \cdots, r_M\delta_M) \in \mathbb{R}^M$ is a diagonal matrix whose i-th diagonal element is $r_i\delta_i$ and $\mathbf{e} \in \mathbb{R}^M$ is a vector of ones.

Step 2. *The Jacobian matrix $\nabla g(x, r)$ is negative definite.* It suffices to show that the symmetric matrix $W = \nabla g(x, r)^T + \nabla g(x, r)$ is negative definite. We do this in three steps.

Step 2.1. *Rewriting $\nabla g(x, r)$ as the sum of a rank-two matrix and a diagonal matrix.* It is clear from (24) that

$$W = A + \Delta,$$

where $A := ae^T + ea^T$, $a := (r_1\zeta_1, \cdots, r_M\zeta_M)^T$, and $\Delta := 2\mathrm{diag}(r_1\delta_1, \cdots, r_M\delta_M)$. Note that A is a rank-two matrix and Δ is a diagonal matrix.

Step 2.2. *Characterizing the two nonzero eigenvalues of the rank-two matrix A.* Note that we can rewrite the matrix A as

$$A = (e, a)(a, e)^T.$$

Let $B := (a, e)^T(e, a)$. Then B is a 2 by 2 matrix. It is well known in algebra that the two eigenvalues of B coincide with two of the eigenvalues of A. Let λ_1, λ_2 be the two eigenvalues. Because

$$B = \begin{pmatrix} \sum_{i=1}^M a_i & M \\ \sum_{i=1}^M a_i^2 & \sum_{i=1}^M a_i \end{pmatrix}^T,$$

it is easy to derive that

$$\lambda_1 = \sum_{i=1}^M a_i + \sqrt{M \sum_{i=1}^M a_i^2} > 0, \; \lambda_2 = \sum_{i=1}^M a_i - \sqrt{M \sum_{i=1}^M a_i^2} < 0.$$

Step 2.3. *The Jacobian matrix $\nabla g(x, r)$ is negative definite.* We have just shown that the matrix A has only one positive eigenvalue equal to $\sum_{i=1}^M r_i\zeta_i + \sqrt{M \sum_{i=1}^M r_i^2 \zeta_i^2}$. Moreover, because $p'_q < 0$ and $C''_i(x_i) \geq 0$, we know that Δ is a negative definite diagonal matrix. Hence, the largest eigenvalue of W is bounded above by

$$\sum_{i=1}^M r_i\zeta_i + \sqrt{M \sum_{i=1}^M r_i^2 \zeta_i^2} + 2 \max_{i=1,\cdots,M} r_i\delta_i,$$

which is negative by assumption. Therefore W is a symmetric negative definite matrix.

Step 3. *The SMS equilibrium is unique.* From Step 2 and Theorem 6 in Rosen [1965], we have that $\sigma(x,r)$ is diagonally strictly concave. Then, by Theorem 2 in Rosen [1965], we have that the leaders equilibrium is unique. Then, because Proposition 2.6 shows that the followers equilibrium is also unique, we have that the SMS equilibrium is unique. □

Proof for Theorem 5.1

Part (i). *We first prove the almost sure convergence of a sequence of SMS equilibria of the SAA problems to the unique SMS equilibrium.* For clarity, we break the proof of Part (i) into three steps.

Step 1. $F(x,z,\xi(\omega))$ *is Lipschitz continuous with an integrable module.*

To see this, note that by Assumptions 2.2 and 2.3 and Lemma 3.3, we have that f_i is piecewise continuously differentiable. At a point where f_i is differentiable with respect to z_i, we have

$$\begin{aligned}(f_i)'_{z_i}(x_1,\cdots,z_i,\cdots,x_M,\xi(\omega)) = {} & p\left(z_i+X_{-i}+Y(z_i+X_{-i},\xi(\omega)),\xi(\omega)\right)\\ & +z_i p'_q\left(z_i+X_{-i}+Y(z_i+X_{-i},\xi(\omega)),\xi(\omega)\right)\left(1+Y'_X(z_i+X_{-i},\xi(\omega))\right)\\ & -C'_i(z_i).\end{aligned}$$

Likewise, at a point where $f(x_1,\cdots,z_i,\cdots,x_M,\xi(\omega))$ is differentiable with respect to x_j, we have for $j=1,\cdots,i-1,i+1,\cdots,M$,

$$(f_i)'_{x_j}(x_1,\cdots,z_i,\cdots,x_M,\xi(\omega)) = z_i p'_q\left(z_i+X_{-i}+Y(z_i+X_{-i},\xi(\omega)),\xi(\omega)\right)\left(1+Y'_X(z_i+X_{-i},\xi(\omega))\right).$$

Note that by Lemma 3.3, $1+Y'_X\in[0,1]$. Under Assumptions 3.1 and 3.4,

$$|(f_i)'_{z_i}(x_1,\cdots,z_i,\cdots,x_M,\xi(\omega))| \le L_3(\xi(\omega))+x_i^u L_1(\xi(\omega))+\max_{z_i\in[0,x_i^u]}|C'_i(z_i)|$$

and

$$|(f_i)'_{x_j}(x_1,\cdots,z_i,\cdots,x_M,\xi(\omega))| \le x_i^u L_1(\xi(\omega)).$$

By the piecewise continuous differentiability of f_i, the mean-value theorem Clarke [1983] implies that there exists a nonnegative function

$$\kappa(\xi(\omega)) := L_3(\xi(\omega))+x_i^u L_1(\xi(\omega))+\max_{z_i\in[0,x_i^u]}|C'_i(z_i)|$$

such that $\mathbb{E}[\kappa(\xi(\omega))]<\infty$ and

$$|f_i(w^{(1)},\xi(\omega))-f_i(w^{(2)},\xi(\omega))| \le \kappa(\xi(\omega))\|w^{(1)}-w^{(2)}\|, \forall w^{(1)},w^{(2)}\in\mathcal{W}, \tag{25}$$

where $w^{(l)} := (x_1^{(l)},\cdots,z_i^{(l)},\cdots,x_M^{(l)})^T$, for $l=1,2$ and $\mathcal{W}:=[0,x_1^u]\times\cdots\times[0,x_M^u]$. This implies $F(x,z,\xi(\omega))$ is Lipschitz with an integrable module.

Step 2. *With probability one, the function* $\Phi_k(x,z)$ *converges to* $\Phi(x,z)$ *uniformly over* $\mathcal{W}\times\mathcal{W}$.

Because $F(x,z,\xi(\omega))$ is Lipschitz with an integrable module, by the uniform strong law of large numbers [Rubinstein and Shapiro 1993, Sections 2.6 and 6.2] we have that with probability one

$$\lim_{k\to\infty}\sup_{(x,z)\in\mathcal{W}\times\mathcal{W}}\left|\Phi_k(x,z)-\Phi(x,z)\right|=0. \tag{26}$$

STEP 3. *With probability one, the sequence $\{x^k\}$ converges to the unique SMS equilibrium x^*.*

Note that because $\mathcal{W}$ is a compact set, we know that the sequence $\{x^k\}$ has an accumulation point $\bar{x}$; that is, there exists a subsequence such that $\{x^{k_j}\} \to \bar{x}$ as $k_j \to \infty$. Then

$$|\Phi_{k_j}(x^{k_j}, z) - \Phi(\bar{x}, z)| \le |\Phi_{k_j}(x^{k_j}, z) - \Phi(x^{k_j}, z)| + |\Phi(x^{k_j}, z) - \Phi(\bar{x}, z)|.$$

By (26), the first term at the right hand side goes to zero uniformly with respect to z with probability one. Because $\Phi(x, z)$ is Lipschitz continuous function on the compact set $\mathcal{W} \times \mathcal{W}$, $\Phi(\cdot, z)$ is uniformly continuous on $\mathcal{W}$. Therefore the second term at the right hand side goes to zero uniformly with respect to z with probability one as $x^{k_j} \to \bar{x}$. Let $\delta > 0$ be any small positive number and k_j be sufficiently large such that

$$\sup_{z \in \mathcal{W}} |\Phi_{k_j}(x^{k_j}, z) - \Phi(\bar{x}, z)| \le \delta$$

with probability one. Let $\bar{z}$ be a global maximizer of $\Phi(\bar{x}, z)$. Observe that

$$\Phi_{k_j}(x^{k_j}, x^{k_j}) - \Phi(\bar{x}, \bar{z}) = \Phi_{k_j}(x^{k_j}, x^{k_j}) - \Phi(\bar{x}, x^{k_j}) + \Phi(\bar{x}, x^{k_j}) - \Phi(\bar{x}, \bar{z}) \le \delta,$$

where the inequality holds because $\sup_{z \in \mathcal{W}} |\Phi_{k_j}(x^{k_j}, z) - \Phi(\bar{x}, z)| \le \delta$ and $\Phi(\bar{x}, x^{k_j}) - \Phi(\bar{x}, \bar{z}) \le 0$ because $\bar{z}$ is a global maximizer. Using a symmetric argument, we can show that $\Phi_{k_j}(x^{k_j}, x^{k_j}) - \Phi(\bar{x}, \bar{z}) \ge -\delta$. Therefore

$$|\Phi_{k_j}(x^{k_j}, x^{k_j}) - \Phi(\bar{x}, \bar{z})| \le \delta,$$

and hence

$$|\Phi(\bar{x}, x^{k_j}) - \Phi(\bar{x}, \bar{z})| \le |\Phi(\bar{x}, x^{k_j}) - \Phi_{k_j}(x^{k_j}, x^{k_j})| + |\Phi_{k_j}(x^{k_j}, x^{k_j}) - \Phi(\bar{x}, \bar{z})| \le 2\delta,$$

which means x^{k_j} becomes a 2δ-global maximizer of $\Phi(\bar{x}, \cdot)$ for k_j sufficiently large. Since δ can be arbitrarily small, by driving δ to zero and k_j to infinity, we show that $\bar{x}$ is a global maximizer of $\Phi(\bar{x}, z)$ w.p.1. Under the conditions of Theorem 4.4, $\Phi(\bar{x}, z)$ is strictly concave which implies that $\Phi(\bar{x}, z)$ has a unique global maximizer, hence $\bar{x} = \bar{z}$. The uniqueness of the equilibrium (from Theorem 4.4) indicates that $\bar{x}$ must coincide with x^*.

Part (ii). *We now prove that with probability approaching* 1 *exponentially fast with the increase of the sample size, the sequence $\{x^k\}$ converges to an approximate SMS equilibrium satisfying (14).* For clarity, we break the proof into four steps.

STEP 1. *The ϵ-subdifferential, $\partial_z^\epsilon F(x, x, \xi)$ is a random set-valued mapping that is Lipschitz continuous in Hausdorff metric.*

Recall that for a concave function h, the ϵ-subdifferential of h at a point y, denoted by $\partial^\epsilon h(y)$, is the set of vectors u such that

$$u^T(z - y) \ge h(z) - h(y) - \epsilon, \forall z.$$

Because $F(x, z, \xi)$ is a concave function in z, we can define the ϵ-subdifferential of $F(x, z, \xi)$ with respect to z at point $x \in \mathcal{W}$ as above. To simplify notation we use $\mathcal{A}^\epsilon(x, \xi)$ to denote the ϵ-subdifferential, that is,

$$\mathcal{A}^\epsilon(x, \xi) := \partial_z^\epsilon F(x, x, \xi).$$

Using a similar discussion to that in [Hiriart-Urruty and Lemaréchal 1993, page 103] and [Rockafellar and Wets 1998, Theorem 14.37], we know that $\mathcal{A}^\epsilon(x,\xi)$ is measurable and hence it is a random set-valued mapping. Moreover, by [Hiriart-Urruty and Lemaréchal 1993, Theorem 4.1.3], $\mathcal{A}^\epsilon(x,\xi)$ is Hausdorff continuous with respect to x.[1] That is, there exists an integrable function $\kappa(\xi,\epsilon)>0$ such that

$$\mathbb{H}(\mathcal{A}^\epsilon(x',\xi),\mathcal{A}^\epsilon(x'',\xi))\le\kappa(\xi,\epsilon)\|x'-x''\|, \tag{27}$$

where $\kappa(\xi,\epsilon)=\max(2L(\xi),4L(\xi)^2)/\epsilon$, $L(\xi)$ is the Lipschitz constant of $F(x,z,\xi)$ with respect to variables (x,z), and $\mathbb{H}(D_1,D_2)$ is the Hausdorff distance between sets D_1 and D_2; that is,

$$\mathbb{H}(D_1,D_2):=\max\{\mathbb{D}(D_1,D_2),\mathbb{D}(D_2,D_1)\},$$

where

$$\mathbb{D}(D_1,D_2):=\sup_{x\in D_1} d(x,D_2)$$

and $d(x,D):=\inf_{x'\in D}\|x-x'\|$ is the distance from a point x to a set D.

Step 2. *The support function of the ϵ-subdifferential is Lipschitz continuous with an integrable module.*

Note that the ϵ-subdifferential is a random set-valued mapping. Our intention is to use [Shapiro and Xu 2008, Theorem 5.1] (which for convenience is restated in Appendix B as Lemma B.2) to show exponential convergence, but this result can only be applied to random functions. For this reason, we need to use the support function of the ϵ-subdifferential to bridge the gap.

The support function $\sigma(u,D)$ of a set D is defined as

$$\sigma(u,D):=\sup_{d\in D} p^T d.$$

Let $\mathcal{B}$ denote the unit ball in $\mathbb{R}^M$. We now show that $\sigma(u,\mathcal{A}^\epsilon(x,\xi))$ is Lipschitz continuous with respect to (u,x) and

$$|\sigma(u',\mathcal{A}^\epsilon(x',\xi))-\sigma(u'',\mathcal{A}^\epsilon(x'',\xi))|\le\tilde\kappa(\xi,\epsilon)(\|u'-u''\|+\|x'-x''\|), \tag{28}$$

for all $\xi\in\Xi$, $u',u''\in\mathcal{B}$ and $x',x''\in\mathcal{X}$, where

$$\tilde\kappa(\xi,\epsilon)\le\max(\kappa(\xi,\epsilon),\|\mathcal{A}^\epsilon(\bar x,\xi)\|+\kappa(\xi,\epsilon)\max_{x\in\mathcal{W}}\|x\|),$$

[1] In more detail, by [Hiriart-Urruty and Lemaréchal 1993, Theorem 4.1.3] we have that $\partial_z^\epsilon F(x,z,\xi)$ is Hausdorff continuous with respect to z. Moreover, the only nondifferentiable term in the definition of $F(x,z,\xi)$ is $Y(z_i+X_{-i},\xi(\omega))$, and this term depends on z_i and X_{-i} on the same manner (the sum of the two variables is the first argument of $Y(.,\xi(\omega))$). Hence, $\partial_z^\epsilon F(x,z,\xi)$ is also Hausdorff continuous with respect to x. Note that the discussion in Hiriart-Urruty and Lemaréchal [1993] is for convex function but the results are applicable to concave functions by changing the sign.

where $\bar{x}$ is any fixed point in $\mathcal{W}$ and for a compact set S, $\|S\| = \sup_{s\in S}\|s\|$. Obviously $\tilde{\kappa}(\xi,\epsilon)$ bounds $\|\mathcal{A}^\epsilon(\bar{x},\xi)\|$ and is integrable because $\mathcal{W}$ is a compact set. To show that (28) holds, note that by the Lipschitz continuity of $\mathcal{A}^\epsilon$ we have that $\mathcal{A}^\epsilon(x'',\xi)\subset\mathcal{A}^\epsilon(x',\xi)+\kappa(\xi,\epsilon)\mathcal{B}$ and hence

$$\begin{aligned}\sigma(u',\mathcal{A}^\epsilon(x',\xi))-\sigma(u'',\mathcal{A}^\epsilon(x'',\xi)) &= \sup_{a\in\mathcal{A}^\epsilon(x',\xi)}(u')^Ta-\sup_{a\in\mathcal{A}^\epsilon(x'',\xi)}(u'')^Ta\\ &\geq \sup_{a\in\mathcal{A}^\epsilon(x',\xi)}(u')^Ta-\sup_{a\in\mathcal{A}^\epsilon(x',\xi)+\kappa(\xi,\epsilon)\|x'-x''\|\mathcal{B}}(u'')^Ta\\ &\geq \sup_{a\in\mathcal{A}^\epsilon(x',\xi)}(u')^Ta-\sup_{a\in\mathcal{A}^\epsilon(x',\xi)}(u'')^Ta-\kappa(\xi,\epsilon)\|x'-x''\|\\ &\geq -\sup_{a\in\mathcal{A}^\epsilon(x',\xi)}(u''-u')^Ta-\kappa(\xi,\epsilon)\|x'-x''\|\\ &\geq -\tilde{\kappa}(\xi,\epsilon)(\|u''-u'\|+\|x'-x''\|).\end{aligned}$$

Swapping x',u' with x'',u'', we obtain (28).

STEP 3. *We apply [Shapiro and Xu 2008, Theorem 5.1] to obtain the exponential convergence.*

We have shown that $\sigma(u,\mathcal{A}(x,\xi))$ is Lipschitz continuous with an integrable module. Moreover, because $\mathcal{A}^\epsilon(x,\xi)$ is bounded by an integrable function, which is also denoted by $\tilde{\kappa}(\xi,\epsilon)$, we have for $\|u\|=1$

$$|\sigma(u,\mathcal{A}^\epsilon(x,\xi))-\mathbb{E}[\sigma(u,\mathcal{A}^\epsilon(x,\xi))]|\leq\|\mathcal{A}^\epsilon(x,\xi)\|+\mathbb{E}[\|\mathcal{A}^\epsilon(x,\xi)\|]\leq\tilde{\kappa}(\xi,\epsilon)+\mathbb{E}[\tilde{\kappa}(\xi,\epsilon)].$$

Because ξ has a bounded support set, then the moment generating function of $\tilde{\kappa}(\xi,\epsilon)$ (that is, the function $\mathbb{E}[e^{t\tilde{\kappa}(\xi,\epsilon)}]$) is finite valued for t close to zero. Hence, from [Shapiro and Xu 2008, Theorem 5.1],[13] we have that for any $\alpha>0$, there exist positive constants $C(\alpha)>0$ and $\beta(\alpha)>0$ such that for k sufficiently large we have that

$$Prob\left(\sup_{\|u\|=1,x\in\mathcal{W}}\left|\frac{1}{k}\sum_{l=1}^k\sigma(u,\mathcal{A}^\epsilon(x,\xi^l))-\mathbb{E}[\sigma(u,\mathcal{A}^\epsilon(x,\xi))]\right|\geq\alpha\right)\leq C(\alpha)e^{-\beta(\alpha)k}. \tag{29}$$

Moreover, because $\mathcal{A}^\epsilon(x,\xi^l)$, $l=1,\cdots,k$, is a convex set, then we have (see for instance Artstein and Vitale [1974])

$$\sigma(u,\mathcal{A}_k^\epsilon(x))=\frac{1}{k}\sum_{l=1}^k\sigma(u,\mathcal{A}^\epsilon(x,\xi^l)),$$

where $\mathcal{A}_k^\epsilon(x)$ is the average of the Minkowski sum of sets $\mathcal{A}^\epsilon(x,\xi^l)$; that is,

$$\mathcal{A}_k^\epsilon(x)=\frac{1}{k}\sum_{l=1}^k\mathcal{A}^\epsilon(x,\xi^l).$$

Furthermore, by [Papageorgiou 1985, Proposition 3.4]

$$\sigma(u,\mathbb{E}[\mathcal{A}^\epsilon(x,\xi)])=\mathbb{E}[\sigma(u,\mathcal{A}^\epsilon(x,\xi))].$$

Therefore, (29) can be rewritten as

$$Prob\left(\sup_{\|u\|=1,x\in\mathcal{W}}|\sigma(u,\mathcal{A}_k^\epsilon(x))-\mathbb{E}[\sigma(u,\mathcal{A}^\epsilon(x,\xi))])|\geq\alpha\right)\leq C(\alpha)e^{-\beta(\alpha)k}. \tag{30}$$

Finally, by [Hiriart-Urruty and Lemaréchal 1993, Theorem 4.1.3], for compact convex sets D_1, D_2

$$\mathbb{H}(D_1, D_2) = \max_{\|u\|=1} \{|\sigma(u, D_1) - \sigma(u, D_2)|\}.$$

Hence, from (30) we have that

$$Prob\left(\sup_{x\in\mathcal{W}} \mathbb{H}(\mathcal{A}_k^\epsilon(x), \mathbb{E}[\mathcal{A}^\epsilon(x,\xi)]) \geq \alpha\right) \leq C(\alpha)e^{-\beta(\alpha)k}. \quad (31)$$

STEP 4. *With probability approaching* 1 *exponentially fast with the increase of the sample size,* x^k *converges to an approximate SMS equilibrium satisfying (14).*

Because $\partial_z\Phi_k(x^k, x^k) \subset \mathcal{A}_k^\epsilon(x^k)$, it follows by (14) that x^k satisfies

$$0 \in \mathcal{A}_k^\epsilon(x^k) + \mathcal{N}_{\mathcal{W}}(x^k). \quad (32)$$

From (31) we then have that with probability $1 - e^{-\beta(\alpha)k}$, x^k satisfies

$$0 \in \mathbb{E}[\mathcal{A}^\epsilon(x^k,\xi)] + \mathcal{N}_{\mathcal{W}}(x^k) + \alpha\mathcal{B} \quad (33)$$

where $\mathcal{B}$ denotes a unit ball. Following the discussion of ϵ-subdifferentials in [Hiriart-Urruty and Lemaréchal 1993, pages 130-131], we have that

$$\mathcal{A}^\epsilon(x^k,\xi) \subset \bigcup_{z\in x^k+\sqrt{\epsilon}\mathcal{B}} \partial_z F(x^k, z, \xi).$$

Using the upper semicontinuity of $\partial_z F(x^k, \cdot, \xi)$, we have that for $\eta > 0$, there exists ϵ such that for all $z \in x^k + \sqrt{\epsilon}\mathcal{B}$, we have that

$$\partial_z F(x^k, z, \xi) \subset \partial_z F(x^k, x^k, \xi) + \frac{1}{2}\eta\mathcal{B}, \forall\xi.$$

The uniformity w.r.t. ξ is due to the fact that both $p(q,t)$ and $p'_q(q,t)$ and $Y_X(X,t)$ are uniformly continuous w.r.t. q and X by Assumption 3.1, Lemma 3.3 and the boundedness of support set of $\xi(\omega)$. Therefore

$$\mathbb{E}[\partial_z F(x^k, z, \xi)] \subset \mathbb{E}[\partial_z F(x^k, x^k, \xi)] + \frac{1}{2}\eta\mathcal{B}.$$

Note that from Lemma 3.3 it is easy to see that $F(x^k, z, \xi)$ is piecewise twice continuously differentiable. Hence

$$\mathbb{E}[\partial_z F(x^k, x^k, \xi)] = \mathbb{E}[\nabla_z F(x^k, x^k, \xi)] = \nabla_z\Phi(x^k, x^k),$$

where the last equality is due to the Lebesgue dominated convergence theorem because $\nabla_z F(x^k, x^k, \xi)$ is bounded by an integrable random variable. By setting ϵ sufficiently small in the first place and $\alpha = \frac{1}{2}\eta$, we have

$$\mathbb{E}[\mathcal{A}^\epsilon(x^k,\xi)] \subset \nabla_z\Phi(x^k, x^k) + \eta\mathcal{B}.$$

Combining this with (33), we obtain (14). □

Appendix B: Statement of frequently used results from the literature

Lemma B.1. ([Ruszczyński and Shapiro 2003, Proposition 2]) Let $h(x,\xi): \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and ξ is a random variable. Suppose that: (a) $\mathbb{E}[h(x,\xi)]$ is well defined, (b) $h(\cdot,\xi)$ is differentiable at x w.p.1, (c) there exists an integrable function $\kappa(\xi) > 0$ such that

$$\|h(x_1,\xi) - h(x_2,\xi)\| \le \kappa(\xi)\|x_1 - x_2\|$$

holds w.p.1. Then $\mathbb{E}[h(x,\xi)]$ is differentiable and

$$\nabla_x \mathbb{E}[h(x,\xi)] = \mathbb{E}[\nabla_x h(x,\xi)].$$

Theorem B.2. ([Shapiro and Xu 2008, Theorem 5.1]) Let $h(x,\xi): \mathcal{X} \times \Xi \to \mathbb{R}$ be a random real valued function and $f(x) = \mathbb{E}[h(x,\xi)]$. Let $\xi^1,...,\xi^N$ be an iid sample of the random vector ξ, and consider the corresponding sample average function $\hat{f}_N(x) := \frac{1}{N}\sum_{j=1}^N h(x,\xi^j)$.

We discuss now uniform exponential rates of convergence of $\hat{f}_N(x)$ to $f(x)$. We denote by

$$M_x(t) := \mathbb{E}\left\{e^{t[h(x,\xi)-f(x)]}\right\}$$

the moment generating function of the random variable $h(x,\xi) - f(x)$. Let us make the following assumptions.

Suppose that: (C1) For every $x \in \mathcal{X}$ the moment generating function

$$M_x(t) := \mathbb{E}\left\{e^{t[h(x,\xi)-f(x)]}\right\}$$

is finite valued for all t in a neighborhood of zero; (C2) there exists a (measurable) function $\kappa: \Xi \to \mathbb{R}_+$ and constant $\gamma > 0$ such that

$$|h(x',\xi) - h(x,\xi)| \le \kappa(\xi)\|x' - x\|^\gamma$$

for all $\xi \in \Xi$ and all $x', x \in \mathcal{X}$; (C3) the moment generating function $M_\kappa(t)$ of $\kappa(\xi)$ is finite valued for all t in a neighborhood of zero. Then for any $\mathbf{e} > 0$ there exist positive constants $C = C(\mathbf{e})$ and $\beta = \beta(\mathbf{e})$, independent of N, such that

$$Prob\left\{\sup_{x\in\mathcal{X}} \left|\hat{f}_N(x) - f(x)\right| \ge \mathbf{e}\right\} \le C(\mathbf{e})e^{-N\beta(\mathbf{e})}.$$