Proofs

EC.1. Proof of Proposition 1

Setting $u = v = 1$ in (3c) specializes to (2c). Further, (3c) guarantees consecutive support: if there exist $j_1 < j_2 < j_3$ such that $x_{j_1}, x_{j_3} > 0$, $x_{j_2} = 0$, by setting $u = j_2 - j_1, v = j_3 - j_2$, the constraint $x_{j-u}^v x_{j+v}^u \le x_j^{u+v}$ for $j = j_2$ is violated. Hence, every feasible distribution of (DMP-LC') is a feasible distribution of (DMP-LC) with the same objective value (note the objectives of both problems are identical).

On the other hand, any feasible distribution to problem (2) with support $[k, \ell]$ satisfies (3c), and by a straightforward induction starting with (2c) as a base case. To be specific, we assume that for some u and v, $x_{j-u}^v x_{j+v}^u \le x_j^{u+v}$ holds for $j \in (k, \ell), j - u \ge k, j + v \le \ell$, and we want to show that $x_{j-u}^{v+1} x_{j+v+1}^u \le x_j^{u+v+1}$. Then it suffices to establish

$$x_{j-u} x_{j+v+1}^u \le x_{j+v}^u x_j, \tag{EC.1}$$

From (2c), we have that $x_{j+v}^2 \ge x_{j+v-1} x_{j+v+1}$, $x_{j+v-1}^2 \ge x_{j+v-2} x_{j+v}, \cdots, x_{j-k}^2 \ge x_{j-k+1} x_{j-k-1}$. Putting those equalities together by applying multiplications to and canceling out the same terms on both sides give that

$$x_{j+v} x_{j-k} \ge x_{j+v+1} x_{j-k-1}, \quad 0 \le k \le j - 2. \tag{EC.2}$$

Hence, starting from $x_{j+v} x_j \ge x_{j+v+1} x_{j-1}$, multiplying it by (EC.2) with $1 \le k \le u - 1$, and (EC.1) follows as required by canceling out the same terms on both sides.

For those points j such that $j - u$ or $j + v$ is outside the support, or the middle point j outside the support, the constraint $x_{j-u}^v x_{j+v}^u \le x_j^{u+v}$ holds naturally because the left-hand side is zero for these cases. In other words, (3c) is satisfied. Hence, every feasible distribution of (DMP-LC) is a feasible distribution of (DMP-LC') with the same objective value. □

EC.2. Proof of Proposition 2

We first prove a preliminary lemma for establishing Proposition 2.

LEMMA EC.1. *The set $\{(x,y,z): x^u y^v \geq z^{u+v}, x\geq 0, y\geq 0, z\geq 0\}$ is convex for any positive integers u and v. Moreover, the set $\{(x,y,z): x^u y^v \geq (z+\epsilon)^{u+v}, x\geq 0, y\geq 0, z\geq 0\}$ is also convex for any positive integers u and v and nonnegative real ϵ.*

Proof of Lemma EC.1. For any integers u and v, let t be an integer such that $u+v\leq 2^t$. From point 11 on page 95 of Ben-Tal and Nemirovski (2001)), the set $\{(x_1,\cdots,x_{2^t},z): x_j\geq 0,\ 1\leq j\leq 2^t,\ z\leq (\Pi_{j=1}^{2^t} x_j)^{1/2^t}\}$ is conic-quadratic representable, and thus convex. Therefore, when intersecting with linear constraints, the set

$$\begin{aligned}\{(x_1,\ldots,x_{2^t},z): x_j\geq 0,\ 1\leq j\leq 2^t,\ 0\leq z\leq (\Pi_{j=1}^{2^t} x_j)^{1/2^t},\ x_j=x_k \text{ for } j,k\in[1,u]\\ x_j=x_k \text{ for } j,k\in[u+1,u+v], \text{ and } x_j=z \text{ for } j,k\in[u+v+1,2^t]\}\end{aligned} \tag{EC.3}$$

is also convex. Now, consider the image of this set under the coordinate projection onto x_1 (relabeled as x), x_{u+1} (relabeled as y), and z. The set in (EC.3) becomes

$$S := \{(x,y,z): z\leq (x^u y^v z^{(2^t-u-v)})^{1/2t}, x\geq 0, y\geq 0, z\geq 0\},$$

which is convex because the projection of a convex set is convex. Observe that the constraint $z\leq (x^u y^v z^{(2^t-u-v)})^{1/2t}$ is equivalent to $z^{u+v}\leq x^u y^v$ and so the set S is precisely the set in the statement of the lemma. The fact S is a convex set proves the result.

To show the "moreover", simply define $w=z+\epsilon$ and the set in question can be written $T=\{(x,y,w): x^u y^v\geq w^{u+v}, x\geq 0, y\geq 0, w\geq \epsilon\}$. Note that this set is the intersection of $\{(x,y,w): x^u y^v \geq w^{u+v}, x\geq 0, y\geq 0, w\geq 0\}$ (convex by the previous part of the lemma) and the convex set $\{(x,y,w): w\geq \epsilon\}$. Thus, T is convex because the intersection of convex sets is convex. $\square$

Proof of Proposition 2. First, define the set

$$S(\epsilon) \triangleq \{(x,y,z): x^u y^v \geq (z+\epsilon)^{u+v},\ x\geq 0, y\geq 0, z\geq 0\}.$$

By Lemma EC.1, $S(\epsilon)$ is convex for every $\epsilon>0$ and the sets are nested; that is, $S(\epsilon_1)\subseteq S(\epsilon_2)$ for $\epsilon_1\geq\epsilon_2$, which implies the union $\bigcup_{\epsilon>0} S(\epsilon)$ is also a convex set (the union of convex nested sets is convex).

Next, we make the following claim:

$$S \triangleq \{(x,y,z) : x^u y^v > z^{u+v},\ x \geq 0, y \geq 0, z \geq 0\} = \bigcup_{\epsilon>0} S(\epsilon). \tag{EC.4}$$

This claim suffices to prove the result because $\bigcup_{\epsilon>0} S(\epsilon)$ is convex.

The $\supseteq$ containment in (EC.4) holds because $S(\epsilon) \subseteq S$ for all $\epsilon > 0$. Conversely, to show $\subseteq$ in (EC.4), let $(x,y,z) \in S$. Hence, we must have $x^u y^v - z^{u+v} = \delta > 0$. Next, we show there exists an $\epsilon > 0$ such that $x^u y^v = (z+\epsilon)^{u+v}$, in which case, $(x,y,z) \in S(\epsilon)$ and we are done. Taking $\epsilon = (z^{u+v} + \delta)^{\frac{1}{u+v}} - z$ does the trick. $\square$

EC.3. Proof of Lemma 1

According to Definition 2, we have the following inequality if x is an IFR distribution:

$$\frac{x_j}{\sum_{k=j}^n x_k} - \frac{x_{j+1}}{\sum_{k=j+1}^n x_k} \leq 0, \text{ for } j \in [1, n-1],$$

which is equivalent to

$$x_j \sum_{k=j+1}^n x_k - x_{j+1} \sum_{k=j}^n x_k \leq 0, \text{ for } j \in [1, n-1], \tag{EC.5}$$

whereas if $\{\bar{F}_1, \ldots, \bar{F}_n\}$ is log-concave, we have

$$\sum_{k=j-1}^n x_k \sum_{k=j+1}^n x_k - \left(\sum_{k=j}^n x_k \right)^2 \leq 0, \text{ for } j \in [2, n-1],$$

which is equivalent to

$$x_j \sum_{k=j+1}^n x_k - x_{j+1} \sum_{k=j}^n x_k \leq 0, \text{ for } j \in [1, n-2]. \tag{EC.6}$$

Inequalities (EC.5) and (EC.6) are the same except that (EC.6) does not include the case in which $j = n-1$. In this case, the inequality holds naturally: $x_{n-1} x_n - x_n (x_{n-1} + x_n) \leq 0$. Thus, the two definitions of IFR distribution are equivalent. $\square$

EC.4. Proof of Proposition 3

First, define the set

$$S(\epsilon) \triangleq \left\{(x,y,z) : \left\| \begin{pmatrix} ax \\ by+cz \end{pmatrix} \right\| \le ax+cz-\epsilon \right\}.$$

Obviously, $S(\epsilon)$ is a standard SOC and thus convex for every $\epsilon > 0$. Moreover, the sets are nested as $S(\epsilon_1) \subseteq S(\epsilon_2)$ for $\epsilon_1 \ge \epsilon_2$, which implies that the union $\bigcup_{\epsilon>0} S(\epsilon)$ is also a convex set, because the union of convex nested sets is convex. Next, we make the following claim:

$$S \triangleq \left\{(x,y,z) : \left\| \begin{pmatrix} ax \\ by+cz \end{pmatrix} \right\| < ax+cz \right\} = \bigcup_{\epsilon>0} S(\epsilon). \tag{EC.7}$$

This claim suffices to prove the result because $\bigcup_{\epsilon>0} S(\epsilon)$ is convex. The $\supseteq$ containment in (EC.7) holds since $S(\epsilon) \subseteq S$ for all $\epsilon > 0$. Conversely, to show $\subseteq$ in (EC.7), let $(x,y,z) \in S$. Hence, we must have

$$ax+cz - \left\| \begin{pmatrix} ax \\ by+cz \end{pmatrix} \right\| = \delta > 0.$$

Taking $\epsilon = \delta$, we have $(x,y,z) \in S(\epsilon)$. $\square$

EC.5. Proof of Theorem 1 and Lemma 4

Proof of Lemma 4. The fact that $\operatorname{ext} \overline{\operatorname{conv}}\, S \subseteq S$ follows immediately from Klee (1957, Theorem 3.5). Suppose, by way of contradiction, that there exists an $x \in \operatorname{ext} \overline{\operatorname{conv}}\, S$ that is not an extreme point of S. Then, there exist $y, z \in S$ with $y \neq z$ such that $x = \lambda y + (1-\lambda) z$ where $\lambda > 0$. However, since $y, z \in \overline{\operatorname{conv}}\, S$, this contradicts that $x \in \operatorname{ext} \overline{\operatorname{conv}}\, S$. The result then holds. $\square$

With Lemma 2, Lemma 3, and Lemma 4 in hand, we can now establish Theorem 1.

Proof of Theorem 1. The problem $\min\{c(x) : x \in \overline{\operatorname{conv}}\, S\}$ has an optimal extreme point solution $x^* \in \operatorname{ext} \overline{\operatorname{conv}}\, S$ by Lemma 2 and the fact that $\overline{\operatorname{conv}}\, S$ is a compact convex set by Lemma 3. Since $S \subseteq \overline{\operatorname{conv}}\, S$, $\min\{c(x) : x \in \overline{\operatorname{conv}}\, S\} \le \min\{c(x) : x \in S\}$. However, since $x^* \in S$, by Lemma 4, we have $c(x^*) = \min\{c(x) : x \in \overline{\operatorname{conv}}\, S\} \le \min\{c(x) : x \in S\} \le c(x^*)$ since x^* is optimal to the minimization over $\overline{\operatorname{conv}}\, S$ and feasible to the minimization over S. However, all inequalities must therefore be equalities and so $\min\{c(x) : x \in S\} = c(x^*)$. Since $x^* \in \operatorname{ext} S$ by Lemma 4, (8) must have an optimal extreme-point solution. $\square$

EC.6. Proof of Lemma 5

Since c is lower-semicontinuous and quasiconcave and F is compact, Theorem 1 implies that there exists an optimal extreme point solution x^*. Let F^* denote the connected component of F that contains x^* and restrict attention to those R_p that intersect F^*. Let $C_p = \mathbb{R}^n \setminus R_p$. Then, C_p is an open convex set since R_p is closed and reverse convex. Since $x^* \in F$, $x^* \notin C_p$ for all p. For all p, let y_p be such that $\operatorname{dist}(x^*, \operatorname{cl}(C_p)) = \operatorname{dist}(x^*, y_p)$; that is, y_p minimizes the distance between x^* and the closure of C_p.

Since x^* is an extreme point of F, there exists a p such that $x^* \in \operatorname{bd}(R_p)$. By definition, $\operatorname{bd}(C_p) = \operatorname{bd}(R_p)$ and so there exists a p such that $x^* \in \operatorname{bd}(C_p)$. Clearly, if $x^* \in \operatorname{bd}(C_p)$ then $\operatorname{dist}(x^*, \operatorname{cl}(C_p)) = 0$ and so $y_p = x^*$. Hence, there always exists a p such that $y_p = x^* \in \operatorname{bd}(C_p)$.

Using the vector y_p, we can define for all $p \in [1, P]$ a supporting hyperplane of $\operatorname{cl}(C_p)$ with normal α_p and right-hand side β_p that weakly separates C_p from the point x^*. These hyperplanes define the polyhedron $\hat{F} = \{x : \alpha_p^\top x \geq \beta_p,\ \text{for } p \in [1, P]\}$ so that $\alpha_p^\top x^* \geq \beta_p$ for all p. Such a choice is always possible since each hyperplane weakly separates C_p from the point x^*, which implies $x^* \in \hat{F}$. In the special case in which $y_p \neq x^*$, the hyperplane $\{x : (x^* - y_p)^\top (x - y_p) \leq 0\}$ does the trick, by the standard projection theorem. Next, we claim $\hat{F} \subseteq F^*$. Indeed, since $\alpha_p^\top x \leq \beta_p$ is a supporting hyperplane of $\operatorname{cl}(C_p)$, the set of x that satisfies $\alpha_p^\top x \geq \beta_p$ lies on the boundary of C_p or outside of C_p. Such an x lies entirely inside of R_p, which implies $\hat{F}$ is a subset of R_p for all p, and so $\hat{F} \subseteq F^*$.

Consider the optimization problem

$$\begin{aligned} &\min\ c(x) \\ &\ s.t.\ \alpha_p^\top x \geq \beta_p \text{ for } p \in [1, P]. \end{aligned} \tag{EC.8}$$

Since $x^* \in \hat{F} \subseteq F^*$ and x^* is an optimal solution of the original problem, x^* is an optimal solution of (EC.8). Moreover, x^* is an extreme point of $\hat{F}$ and so at least n linearly independent constraints at x^* are tight, by the characterization of extreme points of polyhedra (Bertsimas and Tsitsiklis 1997, Theorem 2.3). Hence, at least n of the inequalities $\alpha_p^\top x \geq \beta_p$ must be tight at $x = x^*$. The points in $\hat{F}$ that satisfy $\alpha_p^\top x = \beta_p$ are boundary points of C_p since $\alpha_p^\top x = \beta_p$ is a supporting hyperplane of the convex set C_p. Since $\operatorname{bd}(R_p) = \operatorname{bd}(C_p)$, the points in $\hat{F}$ that satisfy $\alpha_p^\top x = \beta_p$ are boundary points of R_p. Hence, x^* lies on the boundary of at least n of the sets R_p. $\square$

EC.7. Proof of Theorem 2

This proof uses the following notation. For any feasible solution x to (Rev-Cvx), let $S(x)$ denote the support of x; that is, $S(x) = \{j : x_j > 0\}$. Let A^i denote the i-th row of the matrix A and A_j the j-th column. For any subset S of $[1, n]$ (e.g., the support of a feasible solution), let $A_S = [A_j]_{j \in S}$. That is, A_S is the submatrix of A consisting of the columns indexed by S. Note A_S^r denotes the r-th row of the matrix A_S and let $\mathcal{L}(S) = \text{span}(\{(A_S^1)^\top, \ldots, (A_S^m)^\top\}$ denote the span of the rows of A_S. Finally, let $\nabla f_p(x)$ denote the gradient of f_p at x, where $[\nabla f_p(x)]_S$ is the gradient of f_p restricted to the components in the subset S.

Now, for the proof. Since c is continuous and quasiconcave and F is a compact set, by Theorem 1 there exists an optimal extreme point solution. For any such optimal extreme point x^* with support $S = S(x^*)$, define

$$F_0 := \{x : Ax = b,\ x_j = 0 \text{ for } j \notin S,\ x_j > 0 \text{ for } j \in S\}.$$

Then, the feasible region F includes $\{x : f_p(x) \le 0,\ p = 1, \ldots, P\} \cap X_0$. Let $\delta_1 = \min\{x_j^* : x_j^* > 0\}$ and denote

$$F(\delta_1) := \{x : Ax = b,\ x_j = 0 \text{ for } j \notin S,\ x_j \ge \delta_1/2, \text{ for } j \in S\}.$$

Our goal is as follows. For $p = 1, \ldots, P$, we want to construct sets $\hat{F}_p$ of the form

$$\hat{F}_p := \{x : \alpha_p^\top (x - x^*) \le \beta_p\} \cap X(\delta_1), \tag{EC.9}$$

such that

$$\hat{F} := \cap_{p=1}^P \hat{F}_p = \{x : \alpha_p^\top (x - x^*) \le \beta_p,\ p = 1, \ldots, P\} \cap F(\delta_1) \tag{EC.10}$$

is a subset of F, where α_p and $\beta_p \ge 0$ will be specified later. Since $\beta_p \ge 0$, $x^* \in \hat{F}_p$ for all p, and is thus in $\hat{F}$. So, as long as $\hat{F} \subseteq F$ since x^* is an extreme point of F, it must also be an extreme point of $\hat{F}$. Note $\hat{F}$ is defined by linear equalities and inequalities and so there must exist n of them that are tight at point x^*, and we can further check which constraints are tight.

$0 \notin \mathrm{cl}(Y_p)$	$\alpha_p = \hat{\alpha}_p$, $\beta_p = \hat{\beta}_p/2$, $\hat{\alpha}_p$ and $\hat{\beta}_p$ obtained by strong separation
$0 \in \mathrm{cl}(Y_p)$	$\alpha_p = \hat{\alpha}_p$, $\beta_p = 0$, $\hat{\alpha}_p$ are obtained weak separation
$0 \in \mathrm{cl}(Y_p)$, $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$	$\alpha_p = \nabla f_p(x^*)$ and $\beta_p = 0$

Table EC.1 Specifying α_p and β_p in $\hat{X}_p$.

We now construct the $\hat{F}$ in (EC.10). Let $x_S = [x_j]_{j\in S}$ and

$$F_p = F_0 \cap \{(x_S;0) : f_p(x_S;0) > 0\}. \tag{EC.11}$$

Here, $(x_S;0)$ denotes a vector with value 0 outside of the index set S. A key property of F_p is that it admits a strong separation property useful for our arguments (see Claim EC.1 below). To describe this property, we explore a related set in a smaller subspace. Construct matrix $B \in \mathbb{R}^{|S|\times(|S|-\mathrm{rank}(A_S))}$ such that its columns span the whole null space of A_S. That is, $A_S B = 0$ and $\mathrm{rank}(B) = |S| - \mathrm{rank}(A_S)$. Then, we have that

$$\{(x_S;0) : A_S\, x_S = b\} = \{(By + x_S^*;0) : y \in \mathbb{R}^{(|S|-\mathrm{rank}(A_S))}\}. \tag{EC.12}$$

Letting

$$Y_p := \{y : By + x_S^* > 0, f_p(By + x_S^*;0) > 0\},$$

we can define the "strong separation" property of F_p as follows:

Claim EC.1. *(Strong separation) For all p, there exist $\alpha_p^\top$ and $\hat{\beta}_p > 0$ such that*

$$\begin{cases} \hat{\alpha}_p^\top (x - x^*) \geq \hat{\beta}_p > 0, \ \text{for } x \in F_p \ \text{ if } \ 0 \notin \mathrm{cl}(Y_p) \\ \hat{\alpha}_p^\top (x - x^*) > 0, \ \text{for } x \in F_p \ \text{ if } \ 0 \in \mathrm{cl}(Y_p). \end{cases} \tag{EC.13}$$

Moreover, if we further assume $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$ then $\nabla f(x^*)^\top (x - x^*) > 0$ for all $x \in F_p$.

The proof of this claim is somewhat technical and so is relegated to the end of this section. We take it as given and return to constructing $\hat{F}$. According to (EC.10), it suffices to show how to construct $\hat{F}_p$ such that

$$\hat{F}_p \subseteq \{x : f_p(x) \leq 0\},\ p \in [1, P], \tag{EC.14}$$

since $F(\delta_1) \subseteq F_0$. In other words, we need to prove that $x \in \hat{F}_p$ implies $f_p(x) \leq 0$.

We show (EC.14) in two cases: (i) $0 \notin \mathrm{cl}(Y_p)$ and (ii) $0 \in \mathrm{cl}(Y_p)$. We use Table EC.1 to track some of the notation and details.

In case (i), according to Claim EC.1, there exist a $\hat{\alpha}_p \neq 0$ and $\hat{\beta}_p \neq 0$ such that

$$\hat{\alpha}_p^\top (x - x^*) \geq \hat{\beta}_p > 0 \text{ for all } x \in F_p. \tag{EC.15}$$

By letting $\alpha_p = \hat{\alpha}_p$ and $\beta_p = \frac{\hat{\beta}_p}{2}$, one has $x^* \in \hat{F}_p \neq \emptyset$. Moreover, from definition (EC.9) of $\hat{F}_p$, any $x \in \hat{F}_p$ satisfies $x \in F(\delta_1) \subseteq F_0$ and

$$\hat{\alpha}_p^\top (x - x^*) = \alpha_p^\top (x - x^*) \leq \beta_p = \hat{\beta}_p / 2.$$

Combining this statement with (EC.15) yields that $x \notin F_p$ for any $x \in \hat{F}_p$. Then, according to (EC.11), such x does not belong to F_p, simply because it violates the constraint $f_p(x) > 0$. Therefore, we can conclude $f_p(x) \leq 0$ for all $x \in \hat{F}_p$.

In case (ii), again by Claim EC.1, we have $\hat{\alpha}^\top (x - x^*) > 0$ for $x \in F_p$, where $\hat{\alpha} = \nabla f(x^*)$ if $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$. Then, we can take $\alpha_p = \nabla f_p(x^*)$ and $\beta_p = 0$ in (EC.9). Obviously, $x^* \in \hat{F}_p \neq \emptyset$ and $x \notin \hat{F}_p$ for any $x \in F_p$. Similarly, we can argue such an x does not belong to $\hat{F}_p$ due to the violation of the constraint $f_p(x) > 0$. Then, it follows that $f_p(x) \leq 0$ for all $x \in \hat{F}_p$.

So far, we have constructed $\hat{F}_p$ in the form of (EC.11) (as in Table EC.1) and $\hat{F}$ based on (EC.10). Moreover, we have shown $\hat{F} \subseteq F$. Since x^* is an extreme point of F and lies both in F and $\hat{F}$, it is an extreme point of $\hat{F}$ as well. Note $\hat{F}$ is defined by a number of linear equalities and inequalities, then there must exist n of them that are tight and linear independent at point x^* by standard theory (see e.g. Bertsimas and Tsitsiklis (1997, Theorem 2.3)).

Since A is an m by n matrix of rank m, there are $n - m$ tight constraints from

$$\begin{aligned} \alpha_p^\top (x - x^*) &\leq \beta_p \;\text{ for } p \text{ such that } 0 \notin \mathrm{cl}(Y_p) \\ {\alpha_p}^\top (x - x^*) &\leq 0 \;\text{ for } p \text{ such that } 0 \in \mathrm{cl}(Y_p) \\ x_j &= 0 \text{ for } j \notin S \\ x_j &\geq \delta_1/2 \text{ for } j \in S, \end{aligned}$$

where $\alpha_p = \nabla f_p(x^*)$ if $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$. Now, we investigate which of the above constraints are tight. First, obviously $x_j^* = 0$ is tight for all $j \notin S$, and $x_j^* \geq \delta_1 > \delta_1/2$ can not be tight for all $j \in S$. Then, for the constraint p such that $0 \notin \mathrm{cl}(Y_p)$ since $\beta_p > 0$, $\alpha_p^\top (x^* - x^*) = 0 < \beta_p$ cannot be tight. Finally, recall we have proved in the previous discussion that $f_p(x^*) \geq 0$ for all p such that $0 \in \mathrm{cl}(Y_p)$. That is, when $f_p(x^*) < 0$, it holds that

$0 \notin \mathrm{cl}(Y_p)$, and thus the corresponding constraint $\alpha_p^\top (x - x^*) \le \beta_p$ cannot be tight at x^*. In summary, all $n-m$ tight constraints come from

$$\begin{array}{r} {\alpha_p}^\top (x - x^*) \le 0 \text{ for } p \text{ such that } f_p(x^*) = 0 \text{ and } 0 \in \mathrm{cl}(Y_p) \\ x_j = 0 \text{ for } j \notin S, \end{array} \tag{EC.16}$$

which implies $n-m$ of the inequalities

$$\begin{array}{r} f_p(x^*) \le 0 \text{ for } p \in [1, P] \\ x_j^* \ge 0 \text{ for } j \in [1, n] \end{array}$$

in (Rev-Cvx) are tight. This completes the proof of Theorem 2. $\square$

The proof of Claim EC.1 relies on the following two subclaims.

Subclaim 1 *The set Y_p is a convex and open set for $p = 1, \ldots, P$.*

Proof of Subclaim 1: By assumption, $S_1 := \{x : f_p(x) > 0, x \ge 0\}$ is convex. Therefore, $S_2 := \{x : f_p(x) > 0, x \ge 0\} \cap \{x : Ax = b\}$ is also a convex set since we are intersecting S_1 with the convex set $\{x : Ax = b\}$. Moreover, the set $S_3 := S_2 \cap \{x : x_S > 0, x_{\bar{S}} = 0\}$ is again convex since $\{x : x_S > 0, x_{\bar{S}} = 0\}$ is a convex set. Finally, consider the affine map $y \mapsto (By + x_S^*, 0)$. Note that Y_p is the inverse image of this map and therefore convex.

Moreover, for any $y_1 \in Y_p$, let

$$0 < \delta = \min_{j \in \{1, \ldots, |S|\}} \{(By_1 + x_S^*)_j : f_p(By_1 + x_S^*; 0) > 0\}.$$

Since $f_p(\cdot)$ is continuous, there exists an $\epsilon > 0$ such that for any y with $\|y - y_1\|_2 \le \epsilon$, we have

$$\min_{j \in \{1, \ldots, |S|\}} \{(By + x_S^*)_j : f_p(By + x_S^*; 0) > 0\} \ge \delta/2 > 0.$$

Thus, $y \in Y_p$ and Y_p is open. This completes the proof of Subclaim 1. $\square$

Moreover, we have a "strong separation property" of Y_p described as follows.

Subclaim 2 *There exist a $d_p \ne 0$ and $\hat{\beta}_p > 0$ such that*

$$\begin{cases} d_p^\top y \ge \hat{\beta}_p > 0, & \text{for } y \in Y_p \text{ if } \quad 0 \notin \mathrm{cl}(Y_p) \\ d_p^\top y > 0, & \text{for } y \in Y_p \text{ if } \quad 0 \in \mathrm{cl}(Y_p). \end{cases} \tag{EC.17}$$

Moreover, letting $g_p(y) = f_p(By + x_S^*; 0)$ and assuming $\nabla g_p(0) \ne 0$, if $0 \in \mathrm{cl}(Y_p)$ then $\nabla g_p(0)^\top y > 0$ for all $y \in Y_p$.

Proof of Subclaim 2: Note that $f_p(B\cdot 0+x_S^*;0)=f_p(x_S^*;0)=f_p(x^*)\le 0$, thus $0\notin Y_p$. Since, by Subclaim 1, Y_p is convex, $\mathrm{cl}(Y_p)$ is both closed and convex. Then if $0\notin \mathrm{cl}(Y_p)$, by the strong separation theorem for closed convex sets (see, for instance, Aliprantis and Border (2006, Corollary 5.80)), there exist $d_p\neq 0$ and $\hat{\beta}_p>0$ such that $d_p^\top y\ge\hat{\beta}_p>0$ for $y\in Y_p$. In the case of $0\in\mathrm{cl}(Y_p)$, weak separation holds. That is, there exists an $\hat{\alpha}_p\neq 0$ such that

$$\hat{\alpha}_p^\top y\ge 0^\top y=0 \text{ for all } y\in Y_p. \tag{EC.18}$$

Since Y_p is open, that weak separation becomes strict; that is, $\hat{\alpha}_p^\top y>0$ for all y. Otherwise, if there exists a $y'\in Y_p$ such that $\hat{\alpha}_p^\top y'=0$ then since Y_p is open there exists a $y''\in Y_p$ in a small neighborhood of y' such that $\hat{\alpha}_p^\top y''<0$. This violates the condition shown above that $\hat{\alpha}_p^\top y\ge 0$ for $y\in Y_p$. Together this yields (EC.17).

To establish the "moreover" part, note $g_p(y)=f_p(By+x_S^*;0)\ge 0$ for any $y\in\mathrm{cl}(Y_p)$. Hence, $g_p(0)=f_p(x_S^*;0)=f_p(x^*)\le 0$. Combining these two facts gives that $g_p(0)=0$ when $0\in\mathrm{cl}(Y_p)$. That is, 0 is a global minimizer of the problem

$$\begin{array}{ll}\min & g_p(y)\\ \text{s.t.} & y\in\mathrm{cl}(Y_p).\end{array}$$

Thus, the following optimality condition in the form of variational inequality holds: $\nabla g_p(0)^\top(y-0)\ge 0$ for $y\in\mathrm{cl}(Y_p)$, which trivially leads to $\nabla g_p(0)^\top y\ge 0$ for $y\in Y_p$. Since Y_p is open, we get strict separation $\nabla g_p(0)^\top y>0$ for all $y\in Y_p$. This completes the proof of Subclaim 2. □

Proof of Claim EC.1: We show (EC.13) holds with $\hat{\alpha}_p=(B(B^\top B)^{-1}d_p;\gamma_p)$, with d_p being defined in Subclaim 2 and any $\gamma_p\in\mathbb{R}^{n-|S|}$ and $\hat{\beta}_p$ as constructed in Subclaim 2. Indeed, for any $x\in F_p$, due to (EC.12), we can find a $y\in Y_p$ such that $x=(By+x_S^*;0)=(By;0)+x^*$. Consequently,

$$\hat{\alpha}_p^\top(x-x^*)=d_p^\top(B^\top B)^{-\top}B^\top By+\gamma_p^\top 0=d_p^\top y.$$

Then, according to Subclaim 2, (EC.13) holds.

To establish the "moreover" of Claim EC.1, observe that when $\nabla g_p(0)\neq 0$ and $0\in\mathrm{cl}(Y_p)$, by letting $\hat{\alpha}_p=(B(B^\top B)^{-1}\nabla g_p(0);\gamma_p)$ with any $\gamma_p\in\mathbb{R}^{n-|S|}$, we have $\hat{\alpha}_p^\top(x-x^*)>0$ for $x\in F_p$. The argument here is analogous to what we used when establishing (EC.13).

Now, suppose $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$ and $0 \in \mathrm{cl}(Y_p)$. We argue that

$$\nabla f(x^*)^\top (x - x^*) > 0, \text{ for } x \in F_p. \tag{EC.19}$$

First, a direct computation yields

$$\nabla g_p(0) = [B^\top\, 0]\, \nabla f_p(B\, y + x_S^*; 0)\big|_{y=0} = [B^\top\, 0]\, \nabla f_p(x_S^*; 0) = [B^\top\, 0]\, \nabla f_p(x^*) = B^\top [\nabla f_p(x^*)]_S. \tag{EC.20}$$

Since $[\nabla f_p(x^*)]_S \notin \mathcal{L}(S)$, we have $\nabla g_p(0) \neq 0$. Otherwise, due to (EC.20), $[\nabla f_p(x_S^*)]_S$ belongs to the null space of $B^\top$, which is exactly $\mathcal{L}(S)$, giving rise to a contradiction. For any $x \in F_p$, $A_S\,(x_S - x_S^*) = A_S\, x_S - A_S\, x_S^* = 0$, thus $x_S - x_S^* \in \mathrm{Null}(A_S)$. Moreover, recall that the columns of B span the whole $\mathrm{Null}(A_S)$; then there exists a $\theta \neq 0$ such that $x_S - x_S^* = B\,\theta$. Now, let $\hat{\alpha}_p = (B(B^\top B)^{-1} B^\top \nabla f_p(x_S^*); \gamma_p)$ with any $\gamma_p \in \mathbb{R}^{n-|S|}$. According to (EC.13), we have

$$\begin{aligned}
\nabla f(x^*)^\top (x - x^*) &= (\nabla f(x^*) - \hat{\alpha}_p)^\top (x - x^*) + \hat{\alpha}_p^\top (x - x^*) \\
&= \left([\nabla f(x^*)]_S - B(B^\top B)^{-1} B^\top \,[\nabla f_p(x^*)]_S\right)^\top (x_S - x_S^*) + \hat{\alpha}_p^\top (x - x^*) \\
&= [\nabla f_p(x^*)]_S^\top \left(I - B(B^\top B)^{-1} B^\top\right) B\,\theta + \hat{\alpha}_p^\top (x - x^*) \\
&= \hat{\alpha}_p^\top (x - x^*) > 0.
\end{aligned}$$

Thus, (EC.19) holds, completing the proof of Claim EC.1. $\square$

EC.8. Proof of Theorem 4

In (4), the 0-th moment constraint that $\sum_{j=1}^n x_j = 1$, together with constraint (4d), guarantees the feasible region without constraint (4c) is compact. Further, constraint (4c) can be rewritten as

$$\frac{x_j}{\sum_{k=j} x_k} - \frac{x_{j+1}}{\sum_{k=j+1} x_k} \leq 0 \text{ for } j \in [1, n-1],$$

where the LHS is continuous w.r.t. x. This gives us that the set where (4c) holds is closed, thus the feasible region is compact. By Theorem 1, problem (4) has an extreme point solution x^*. This extreme point solution corresponds to an extreme point solution y^* to problem (5) by equivalence.

Note that by constraint (4c), x^* must have consecutive support. Otherwise, assume there exist $1 \le j_1 < j_2 < j_3 \le n$ such that $x^*_{j_1} > 0, x^*_{j_3} > 0, x^*_{j_2} = 0$; however,

$$\frac{x^*_{j_1}}{\sum_{k=j_1}^n x^*_k} > \frac{x^*_{j_2}}{\sum_{k=j_2}^n x^*_k} = 0,$$

indicating x^* is not feasible. Without loss of generality, we can assume the support of x^* is $[1,n]$; that is, constraints (4d) are strict.

Let $\underline{x} := \min\{x^*_j : j \in [1,n]\}$ and define the following problem:

$$\max_{y\in\mathbb{R}^n} \sum_{j=1}^{n-1} f_j(y_j - y_{j+1}) + f_n y_n \tag{EC.21a}$$

$$\text{s.t.} \sum_{j=1}^{n} (w^i_j - w^i_{j-1}) y_j = q_i \quad \text{for } i \in [0,m] \tag{EC.21b}$$

$$y_{j-1} y_{j+1} \le y_j^2 \quad \text{for } j \in (1,n) \tag{EC.21c}$$

$$y_j - y_{j+1} \ge \underline{x}/2 \quad \text{for } j \in [1,n] \tag{EC.21d}$$

$$y_j \ge 0 \text{ for } j \in [1,n]. \tag{EC.21e}$$

The problem above is a restriction of (5) with given support, and plus (redundant) non-negativity constraint (EC.21e) so that it fits our reverse convex programming frameworks. Since y^* is an extreme point solution to (5), it is also an extreme point solution to (EC.21). Thus, we only need to verify the conditions of Theorem 2.

Again, the 0-th order moment constraint guarantees the feasible region is compact. As argued in Section 2.2, the constraint (EC.21c) is reverse convex relative to the nonnegative orthant, and so is the linear constraint (EC.21d), which implies all of the conditions in Theorem 2 are satisfied when applied to (EC.21).

At the extreme point solution y^*, the constraints (EC.21d) and (EC.21e) cannot be tight since $x^*_j = y^*_j - y^*_{j+1} > \underline{x}/2$ and y^*_j are all positive since the x^*_j are positive. Applying Theorem 2, we have that at least $n-m-1$ of the (EC.21c) constraints are tight at y^*, or, equivalently, there are at most $m-1$ of the (EC.21c) constraints that are not tight at y^*. These non-tight indexes can divide the interval $[1,n]$ into at most m pieces, and within each piece, we have

$$y_{j-1} y_{j+1} = y_j^2, \text{ for } j \in (u_i, v_i),$$

where u_i and v_i are the left and right endpoints of piece i of the domain. Thus, letting $\alpha_i = y_{u_i}$ and $r_i = y_{u_i+1}/y_{u_i}$ yields the solution (13). □

EC.9. Proof of Theorem 5

In (6), the zeroth moment constraint that $\sum_{j=1}^n x_j = 1$ together with constraint (6d) guarantees the feasible region without constraint (6c) is compact. Further, constraint (6c) can be rewritten as

$$j\frac{x_j}{\sum_{k=j} x_k} - (j+1)\frac{x_{j+1}}{\sum_{k=j+1} x_k} \le 0 \text{ for } j \in [1, n-1],$$

where the LHS is continuous w.r.t. x. This gives us that the set where (6c) holds is closed, thus the feasible region is compact. By Theorem 1, problem (6) has an extreme point solution x^*. This extreme point solution corresponds to an extreme point solution y^* to problem (7) by equivalence.

Note that by constraint (6c), x^* must have consecutive support. Otherwise assume there exist $1 \le j_1 < j_2 < j_3 \le n$ such that $x^*_{j_1} > 0, x^*_{j_3} > 0, x^*_{j_2} = 0$, however

$$j_1\frac{x^*_{j_1}}{\sum_{k=j_1}^n x^*_k} > j_2\frac{x^*_{j_2}}{\sum_{k=j_2}^n x^*_k} = 0,$$

indicating x^* is not feasible. Without loss of generality, we can assume the support of x^* is $[1, n]$, i.e. constraints (6d) are strict (otherwise we can redefine on the support $[k, \ell]$ and use analogous arguments).

Let $\underline{x} := \min\{x^*_j : j \in [1, n]\}$ and define the following problem:

$$\max_{y\in\mathbb{R}^n} \sum_{j=1}^{n-1} f_j(y_j - y_{j+1}) + f_n y_n \tag{EC.22a}$$

$$\text{s.t.} \sum_{j=1}^{n}(w^i_j - w^i_{j-1})y_j = q_i \quad \text{for } i \in [0, m] \tag{EC.22b}$$

$$\left\|\begin{pmatrix} j\sqrt{j-1}y_{j+1} \\ \sqrt{j-1}y_j + \frac{y_{j-1}}{2\sqrt{j-1}}\end{pmatrix}\right\| \ge j\sqrt{j-1}y_{j+1} + \frac{y_{j-1}}{2\sqrt{j-1}} \quad \text{for } j \in (1, n) \tag{EC.22c}$$

$$y_j - y_{j+1} \ge \underline{x}/2 \quad \text{for } j \in [1, n] \tag{EC.22d}$$

$$y_j \ge 0 \text{ for } j \in [1, n]. \tag{EC.22e}$$

Note that the problem above is a restriction of (7) with given support, and plus (redundant) negativity constraint (EC.22e) so that it fits our reverse convex programming frameworks.

Since y^* is an extreme point solution to (7), it is also an extreme point solution to (EC.22). Thus, we only need to verify the conditions of Theorem 2.

Again, the zeroth order moment constraint guarantees the feasible region is compact. As argued in Section 2.3, the constraint (EC.22c) is reverse convex relative to the nonnegative orthant, and so is the linear constraint (EC.22d). This implies that all of the conditions in Theorem 2 are satisfied when applied to (EC.22).

At the extreme point solution y^*, the constraints (EC.22d) and (EC.22e) cannot be tight since $x_j^* = y_j^* - y_{j+1}^* > \underline{x}/2$ and y_j^* are all positive since the x_j^* are positive. Applying Theorem 2, we have that at least $n-m-1$ of the (EC.22c) constraints are tight at y^*, or equivalently there are at most $m-1$ of the (EC.22c) constraints that are not tight at y^*. These non-tight indexes can divide the interval $[1,n]$ into at most m pieces, and within each piece, we have

$$\left\| \begin{pmatrix} j\sqrt{j-1}y_{j+1} \\ \sqrt{j-1}y_j + \frac{y_{j-1}}{2\sqrt{j-1}} \end{pmatrix} \right\| = j\sqrt{j-1}y_{j+1} + \frac{y_{j-1}}{2\sqrt{j-1}}, \text{ for } j \in (u_i, v_i)$$

where u_i and v_i are the left and right endpoints of piece i of the domain. Note the equalities above are equivalent to

$$(j-1)\frac{y_{j-1}-y_j}{y_{j-1}} = j\frac{y_j - y_{j+1}}{y_j}, \text{ for } j \in (u_i, v_i).$$

Letting

$$r_i := j\frac{y_j - y_{j+1}}{y_j}, \text{ for } j \in (u_i, v_i),$$

we have

$$y_j = y_{j-1}(1 - \frac{r_i}{j-1}), \text{ for } j \in (u_i, v_i].$$

using the fact that $j\frac{y_j - y_{j+1}}{y_j} = r$ is equivalent condition $y_{j+1} = y_j(1-\frac{r}{j})$. Thus, letting $\alpha_i = y_{u_i}$ yields the solution (14). $\square$

EC.10. Proof of Lemma 6

If the a_j are all nonnegative or nonpositive, then $\phi(z)$ is monotone and has at most one root. Otherwise, there is an m such that $a_j \le 0$ when $j \le m$ and $a_j \ge 0$ when $j > m$. Denote $\phi_1(z) := -\sum_{j=1}^m a_j z^{i_j}$ and $\phi_2(z) = \sum_{j=m+1}^M a_j z^{i_j}$. Obviously, $\phi(z) = \phi_2(z) - \phi_1(z)$. Suppose $z_0 > 0$ is a root of $\phi(z)$, from non-negativity of ϕ_1 and ϕ_2 we have $\phi_2(z_0) = \phi_1(z_0) \neq 0$. Given any $z_1 > z_0$, due to (23) we have that

$$\phi_2(z_1) = \sum_{j=m+1}^{M} a_j \left(\frac{z_1}{z_0}\right)^{i_j} (z_0)^{i_j} \ge \left(\frac{z_1}{z_0}\right)^{i_{m+1}} \phi_2(z_0)$$
$$\phi_1(z_1) = \sum_{j=1}^{m} -a_j \left(\frac{z_1}{z_0}\right)^{i_j} (z_0)^{i_j} \le \left(\frac{z_1}{z_0}\right)^{i_m} \phi_1(z_0).$$

Combining these two inequalities yields

$$\phi(z_1) \ge \left(\frac{z_1}{z_0}\right)^{i_{m+1}} \phi_2(z_0) - \left(\frac{z_1}{z_0}\right)^{i_m} \phi_1(z_0) > \left(\frac{z_1}{z_0}\right)^{i_m} (\phi_2(z_0) - \phi_1(z_0)) = 0. \qquad \text{(EC.23)}$$

Similarly, for any $z_2 < z_0$, it holds that $\phi(z_2) < \left(\frac{z_2}{z_0}\right)^{i_m} (\phi_2(z_0) - \phi_1(z_0)) = 0$.

Consequently, z_0 is the only root. Moreover, when z_0 is not a root and satisfies $\phi(z_0) \ge 0$, then according to (EC.23) $\phi(z_1) > \left(\frac{z_1}{z_0}\right)^{i_m} (\phi_2(z_0) - \phi_1(z_0)) > \phi_2(z_0) - \phi_1(z_0) = \phi(z_0)$ implying that $\phi(z)$ is monotonically increasing on $\{z \mid \phi(z) \ge 0\}$. $\square$

EC.11. Proof of Lemma 7

It suffices to provide a uniform bound on $|\frac{\partial g_2(\alpha,\beta)}{\partial \beta}|/|\frac{\partial g_1(\alpha,\beta)}{\partial \beta}|$. When $a \le 0$, since the distribution has positive mass on both sides of a, it has positive mass at $X < a$. This combined with (24) implies that $\mathbb{E}[X \cdot (a-X)\mathbb{1}_{X<a}] \le \mathbb{E}[-(X-a)^2\mathbb{1}_{X<a}] < 0$. By further invoking (26) and (22), we have that

$$\begin{aligned}
\frac{|\frac{\partial g_2(\alpha,\beta)}{\partial \beta}|}{|\frac{\partial g_1(\alpha,\beta)}{\partial \beta}|} = \frac{|\mathbb{E}[(1-X^2)(a-X)\mathbb{1}_{X<a}]|}{|\mathbb{E}[X \cdot (a-X)\mathbb{1}_{X<a}]|} &\le \frac{|\mathbb{E}[(1-X^2)(a-X)\mathbb{1}_{X<a}]|}{\mathbb{E}[(X-a)^2\mathbb{1}_{X<a}]} \\
&\le \max_{x:x=a-j\epsilon, j=1,2,\cdots,\tilde{k}} \frac{|(1-x^2)(a-x)|}{(x-a)^2} \\
&\le \frac{\max_{x:x=a-j\epsilon, j=1,2,\cdots,\tilde{k}} |(1-x^2)|}{\min_{x:x=a-j\epsilon, j=1,2,\cdots,\tilde{k}} (a-x)} \\
&\le \frac{1+D^2}{\epsilon},
\end{aligned}$$

where the second inequality is due to the fractional linear function $\frac{y}{z}$ is quasi-convex and thus $\frac{\sum_i \alpha_i y_i}{\sum_i \alpha_i z_i} \leq \max_i\{\frac{y_i}{z_i}\}$ for $\sum_i \alpha_i = 1$ and $\alpha_i \geq 0$.
Similarly, when $a > 0$ and $\mathbb{E}X = 0$ implies that $\mathbb{E}\left[X \cdot \mathbb{1}_{X<a}\right] + \mathbb{E}\left[X \cdot \mathbb{1}_{X\geq a}\right] = 0$. Therefore,

$$a\,\mathbb{E}\left[X \cdot \mathbb{1}_{X<a}\right] = -a\mathbb{E}\left[X \cdot \mathbb{1}_{X\geq a}\right] \leq -a^2 P(X \geq a),$$

and due to (24)

$$\mathbb{E}\left[X \cdot (a - X)\mathbb{1}_{X<a}\right] \leq \mathbb{E}\left[-X^2\,\mathbb{1}_{X<a}\right] - a^2 P(X \geq a) \leq \mathbb{E}\left[\min(X,a)^2\right] < 0,$$

where the last inequality follows from the fact that X has positive mass at both a and $X < a$. Since the domain of X is bounded with diameter D, we also have that for any $b > 0$

$$\begin{aligned} 1 = \mathbb{E}X^2 = \mathbb{E}[X^2\mathbb{1}_{|X|<b}] + \mathbb{E}[X^2\mathbb{1}_{|X|\geq b}] &\leq b^2\mathbb{E}[\mathbb{1}_{|X|<b}] + D^2\mathbb{E}[\mathbb{1}_{|X|\geq b}] \\ &= b^2(1 - P(|X| \geq b)) + D^2 P(|X| \geq b). \end{aligned}$$

That is $P(|X| \geq b) \geq \frac{1-b^2}{D^2-b^2}$. Furthermore, note that for any $|X| \geq b$ we have $|\min(X,a)| \geq \min(b,a)$. Then,

$$\begin{aligned} \mathbb{E}\left[\min(X,a)^2\right] &= \mathbb{E}\left[\min(X,a)^2 \cdot \mathbb{1}_{|X|\geq b}\right] + \mathbb{E}\left[\min(X,a)^2 \cdot \mathbb{1}_{|X|<b}\right] \\ &\geq \mathbb{E}\left[\min(X,a)^2 \cdot \mathbb{1}_{|X|\geq b}\right] \\ &\geq \min(b,a)^2 P(|X| \geq b) \\ &\geq \min(b,a)^2 \frac{1-b^2}{D^2-b^2}. \end{aligned}$$

Since $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$, the upper bound D of $|X|$ is greater than 1. Therefore, when $a > 0$ by taking $b = 0.5$, we have $D^2 - b^2 > 0$ and when $a > 0$ it holds that

$$\frac{|\frac{\partial g_2(\alpha,\beta)}{\partial\beta}|}{|\frac{\partial g_1(\alpha,\beta)}{\partial\beta}|} = \frac{|\mathbb{E}\left[(1-X^2)(a-X)\mathbb{1}_{X<a}\right]|}{|\mathbb{E}\left[X\cdot(a-X)\mathbb{1}_{X<a}\right]|} \leq \frac{|\mathbb{E}\left[(1-X^2)(a-X)\mathbb{1}_{X<a}\right]|}{\mathbb{E}\left[\min(X,a)^2\right]} \leq \frac{(1+D^2)(a+D)}{\min(0.5,a)^2\frac{0.75}{D^2-0.25}}.$$

Combining the bound on $a \leq 0$, we conclude that

$$\frac{|\frac{\partial g_2(\alpha,\beta)}{\partial\beta}|}{|\frac{\partial g_1(\alpha,\beta)}{\partial\beta}|} \leq \begin{cases} \frac{(1+D^2)(a+D)(D^2-0.25)}{(0.75)\min(0.5,a)^2}, & a > 0 \\ \frac{1+D^2}{\epsilon}, & a \leq 0 \end{cases}.$$

That is, the absolute value of $|\frac{\delta}{\delta\beta}g_1(\alpha,\beta)|$ is uniformly bounded below. $\square$

EC.12. Proof of Theorem 7

Since ϕ' is a continuous function, there exists $\delta > 0$ such that $|\phi'(z) - \phi'(\gamma)| \leq \frac{1}{4}|\phi'(\gamma)|$ whenever $|z - \gamma| \leq \delta$. Consider a Newton step with $|z_k - \gamma| \leq \delta$ and

$$\begin{aligned} z_{k+1} - \gamma = z_k - \gamma - \frac{\phi(z_k) - \phi(\gamma)}{\phi'(z_k)} &= \frac{\phi(\gamma) - \phi(z_k) - \phi'(z_k)(\gamma - z_k)}{\phi'(z_k)} \\ &= \frac{(\phi'(\gamma_k) - \phi'(z_k))(\gamma - z_k)}{\phi'(z_k)}, \end{aligned}$$

where γ_k is a point between γ and z_k. Note that $|\phi'(\gamma_k) - \phi'(z_k)| = |\phi'(\gamma_k) - \phi'(\gamma) + \phi'(\gamma) - \phi'(z_k)| \leq \frac{1}{2}|\phi'(\gamma)|$ and $|\phi'(z_k)| \geq \frac{3}{4}|\phi'(\gamma)|$. Therefore,

$$|z_{k+1} - \gamma| \leq \frac{|(\phi'(\gamma_k) - \phi'(z_k)| \cdot |\gamma - z_k|}{|\phi'(z_k)|} \leq \frac{2}{3}|z_k - \gamma| \leq \delta.$$

Moreover, we have

$$\lim_{k\to\infty} \frac{|z_{k+1} - \gamma|}{|z_k - \gamma|} = \lim_{k\to\infty} \frac{|\phi'(\gamma_k) - \phi'(z_k)|}{|\phi'(z_k)|} = 0,$$

and we conclude that $\{z_k\}$ is superlinear convergent. □

EC.13. Proof of Theorem 8

According to (25) and (21), we have that $|\frac{\partial g_2(\alpha,\beta)}{\partial \alpha}| \leq \frac{(1+D^2)(a+D)}{\rho\cdot\epsilon}$ and $|\frac{\partial g_1(\alpha,\beta)}{\partial \alpha}| \leq \frac{(a+D)D}{\rho\cdot\epsilon}$ respectively. Finally, by invoking (27) we obtain an upper bound of gradient of $g_2(\cdot, h(\cdot))$:

$$\begin{aligned} \left|\frac{\mathrm{d}g_2(\alpha, h(\alpha))}{\mathrm{d}\alpha}\right| &= \left|\frac{\partial g_2(\alpha,\beta)}{\partial \alpha} + \frac{\partial g_2(\alpha,\beta)}{\partial \beta} \cdot \frac{-\frac{\partial g_1(\alpha,\beta)}{\partial\alpha}}{\frac{\partial g_1(\alpha,\beta)}{\partial\beta}}\right| \\ &\leq \left|\frac{\partial g_2(\alpha,\beta)}{\partial \alpha}\right| + \left|\frac{\partial g_2(\alpha,\beta)}{\partial \beta}\right| \cdot \left|\frac{-\frac{\partial g_1(\alpha,\beta)}{\partial\alpha}}{\frac{\partial g_1(\alpha,\beta)}{\partial\beta}}\right| \\ &\leq \begin{cases} \frac{(1+D^2)(a+D)}{\rho\epsilon}\left(1 + \frac{D(a+D)(D^2-0.25)}{(0.75)\min(0.5,a)^2}\right), & a > 0 \\ \frac{(1+D^2)(a+D)}{\rho\epsilon}\left(1 + \frac{D}{\epsilon}\right), & a \leq 0 \end{cases}. \end{aligned}$$

□