thanks: Vris CHEUNG

Department of Combinatorics and Optimization

University of Waterloo

December 3, 2009

CONVEX OPT. AND ANALYSIS — Assignment 4

1 Conjugate Duality

Question 1.

(a) We first prove that if $f, -g : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ are proper convex function, then $f \odot g$ is concave. For any $y, z \in \mathbb{E}$, $\lambda \in [0, 1]$, let $z_\lambda = z + \lambda(y - z) = \lambda y + (1 - \lambda)z$. For any $x \in \mathbb{E}$, since $z_\lambda - x = \lambda(y - x) + (1 - \lambda)(z - x)$, by concavity of g we have

$$\lambda g(z - x) + (1 - \lambda)g(y - x) \leq g(z_\lambda - x)$$
$$\implies \lambda[f(x) + g(z - x)] + (1 - \lambda)[f(x) + g(y - x)] \leq f(x) + g(z_\lambda - x).$$

Taking infimum over all $x \in \mathbb{E}$,

$$\lambda \inf_x [f(x) + g(z - x)] + (1 - \lambda) \inf_x [f(x) + g(y - x)] \leq \inf_x \{\lambda[f(x) + g(z - x)] + (1 - \lambda)[f(x) + g(y - x)]\}$$
$$\leq \inf_x \{f(x) + g(z_\lambda - x)\}.$$

Therefore we have $\lambda(f \odot g)(z) + (1 - \lambda)(f \odot g)(y) \leq (f \odot g)(\lambda z + (1 - \lambda)y)$.

We go straight to Q.2; the claims in Q.1 follow easily from Q.2.

Question 2.

(a) Before we prove that $\bigodot_{i=1}^k f_i := f_1 \odot \cdots \odot f_k$ is convex, we prove the following lemmas:

Lemma 1 *$f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if the strict epigraph $epi_s(f) := \{(x, r) : f(x) < r\}$ is convex.*

Proof. If f is convex, then for all $(x, r), (y, s) \in epi_s(f)$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) < \lambda r + (1 - \lambda)s,$$

so $\lambda(x, r) + (1 - \lambda)(y, s) = (\lambda x + (1 - \lambda)y, \lambda r + (1 - \lambda)s) \in \text{epi}_s(f)$. Hence $\text{epi}_s(f)$ is convex.

Conversely, if $\text{epi}_s(f)$ is convex, then for all $x, y \in \text{dom}(f)$, $n \in \mathbb{N}$, $(x, f(x) + 1/n), (y, f(y) + 1/n) \in \text{epi}_s(f)$. By convexity of this set, for any $\lambda \in [0, 1]$, we have

$$(\lambda x + (1 - \lambda)y, \lambda f(x) + (1 - \lambda)f(y) + 1/n) = \lambda(x, f(x) + 1/n) + (1 - \lambda)(y, f(y) + 1/n) \in \text{epi}_s(f),$$

which means $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) + 1/n$. This holds for all $n \in \mathbb{N}$; taking $n \to \infty$ we have $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$, showing that f is convex. $\square$

Lemma 2 *For any finite collection of functions $f_1, \ldots, f_k : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$*

$$epi_s \left(\bigodot_{i=1}^{k} f_i \right) = \sum_{i=1}^{k} epi_s(f_i).$$

Proof. If $(x_i, r_i) \in \text{epi}_s(f_i)$ for $i = 1, \ldots, k$, then $f_i(x_i) < r_i$ for $i = 1, \ldots, k$. Summing up these k inequalities, we have

$$\bigodot_{i=1}^{k} f_i(x_1 + \cdots + x_k) = \inf_{y_1, \ldots, y_k} \left\{ \sum_{i=1}^{k} f_i(y_i) : y_1 + \cdots + y_k = x_1 + \cdots + x_k \right\} \leq \sum_{i=1}^{k} f_i(x_i) < \sum_{i=1}^{k} r_i,$$

so $(x_1 + \cdots + x_k, r_1 + \cdots + r_k) \in \text{epi}_s\left(\bigodot_{i=1}^{k} f_i\right)$. This shows that $\sum_{i=1}^{k} \text{epi}_s(f_i) \subseteq \text{epi}_s\left(\bigodot_{i=1}^{k} f_i\right)$.

Conversely, if $(x, r) \in \text{epi}_s\left(\bigodot_{i=1}^{k} f_i\right)$, then

$$\inf_{y_1, \ldots, y_n} \left\{ \sum_{i=1}^{k} f_i(x_i) : x_1 + \cdots + x_k = x \right\} < r \iff \exists x_1, \ldots, x_k \in \mathbb{E} \text{ s.t. } \sum_{i} x_i = x, \ \sum_{i=1}^{k} f_i(x_i) < r$$

For this choice of $x_1, \ldots, x_k$, fix an $\varepsilon > 0$ such that

$$\sum_{i=1}^{k-1} [f_i(x_i) + \varepsilon] + f_k(x_k) < r.$$

Then we have

$$(x_i, f_i(x_i) + \varepsilon) \in \text{epi}_s(f_i) \text{ for } i = 1, \ldots, n - 1, \quad \text{and} \quad \left(x_k, r - \sum_{i=1}^{k-1} [f_i(x_i) + \varepsilon] \right) \in \text{epi}_s(f_k).$$

Hence $(x, r) = \sum_{i=1}^{k-1} (x_i, f_i(x_i) + \varepsilon) + \left(x_k, r - \sum_{i=1}^{k-1} [f_i(x_i) + \varepsilon] \right) \in \sum_{i=1}^{k} \text{epi}_s(f_i)$. This shows that $\text{epi}_s\left(\bigodot_{i=1}^{k} f_i\right) \subseteq \sum_{i=1}^{k} \text{epi}_s(f_i)$. Therefore $\text{epi}_s\left(\bigodot_{i=1}^{k} f_i\right) = \sum_{i=1}^{k} \text{epi}_s(f_i)$. $\square$

Now observe that any function $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if $\mathrm{epi}_s (f) \in \mathbb{E} \times \mathbb{R}$ is a convex set. With this observation, we can easily see that if $f_1, \ldots, f_k : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ are all convex, then $\mathrm{epi}_s (f_1), \ldots, \mathrm{epi}_s (f_k) \in \mathbb{E} \times \mathbb{R}$ are all convex sets, so $\mathrm{epi}_s \left(\bigodot_{i=1}^k f_i \right) = \sum_{i=1}^k \mathrm{epi}_s (f_i)$ is a convex set, implying that $\bigodot_{i=1}^k f_i$ is a convex function. $\square$

Remark We saw a nice interpretation of infimal convolution, that the *strict* epigraph of $f \odot g$ is simply the Minkowski sum of those of f and g, even if these functions are not convex. Unfortunately, we don't have $\mathrm{epi}\, (f \odot g) = \mathrm{epi}\, (f) + \mathrm{epi}\, (g)$ in general. A simple example is $f, g : \mathbb{R} \to \mathbb{R}$ defined by $f(t) = \exp(t)$ and $g \equiv 0$. $f \odot g \equiv 0$, so $\mathrm{epi}\, (f \odot g) = \mathbb{R} \times \mathbb{R}_+$. By $\mathrm{epi}\, (f) + \mathrm{epi}\, (g) = \mathbb{R} \times \mathbb{R}_{++}$.

(b) Next we prove that $\left(\bigodot_{i=1}^k f_i \right)^* = f_1^* + \cdots + f_k^*$. In fact, for any $x^* \in \mathbb{E}$,

$$
\left(\bigodot_{i=1}^k f_i \right)^* (x^*) = \sup_x \left\{ \langle x^*, x \rangle - \inf_{x_1, \ldots, x_k} \{ f_1(x_1) + \cdots + f_k(x_k) : x_1 + \cdots + x_k = x \} \right\}
$$

$$
= \sup_x \left\{ \langle x^*, x \rangle - \inf_{x_1, \ldots, x_{k-1}} \{ f_1(x_1) + \cdots + f_{k-1}(x_{k-1}) + f_k \big(x - (x_1 + \cdots + x_{k-1}) \big) \} \right\}
$$

$$
= \sup_x \sup_{x_1, \ldots, x_{k-1}} \left\{ \langle x^*, x \rangle - \big[f_1(x_1) + \cdots + f_{k-1}(x_{k-1}) + f_k \big(x - (x_1 + \cdots + x_{k-1}) \big) \big] \right\}
$$

$$
= \sum_{i=1}^k \sup_{x_i} \big[\langle x^*, x_i \rangle - f_i(x_i) \big]
$$

$$
= \sum_{i=1}^k f_i^* (x^*).
$$

Therefore, $\left(\bigodot_{i=1}^k f_i \right)^* = f_1^* + \cdots + f_k^*$.

2　Entropy Minimization

Question 1.

(a) For any $\gamma \in \mathbb{R}$, let $q(t) := \gamma t + p(t)$. Note that for $t > 0$, $q(t) = t\big[\ln t + (\gamma - 1)\big]$.

(i) q is strictly convex on $\mathbb{R}_+$.
Proof. For any $t > 0$, $q'(t) = \gamma + \ln t$ and $q''(t) = 1/t > 0$. Hence q is strictly convex on $\mathbb{R}_{++}$.

For any $t > 0$ and $\lambda \in (0, 1)$,

$$p(\lambda t) = \lambda t \ln(\lambda t) - \lambda t = \lambda(t \ln t - t) + \lambda t \ln \lambda < \lambda p(t)$$

as $\ln \lambda < 0$. This shows that for any distinct $t_1, t_2 \geq 0$ and $\lambda \in (0, 1)$, $q(\lambda t_1 + (1 - \lambda)t_2) < \lambda q(t_1) + (1 - \lambda)q(t_2)$ (since q is the sum of p and a multiple of t), that is, q is strictly convex on $\mathbb{R}_+$. $\square$

(ii) q has bounded level sets.
Proof. Fix any $\alpha \in \mathbb{R}$. For $t > \max\{1, \exp(\alpha - \gamma + 1)\}$,

$$\ln t \geq \alpha - (\gamma - 1) \quad \Longrightarrow \quad (\gamma - 1)t + t \ln t \geq \alpha t > \alpha,$$

so $\{t \in \mathbb{R} : q(t) \leq \alpha\} = \{t > 0 : \gamma t + (t \ln t - t) \leq \alpha\} \cup \{0\} \subseteq \big[0, \max\{1, \exp(\alpha - \gamma + 1)\}\big]$, that is, $\{t \in \mathbb{R} : q(t) \leq \alpha\}$ is bounded. $\square$

(iii) q is bounded below (and in fact attains global minimum on $\mathbb{R}_+$).
Proof. In the same spirit as in Lemma 1, Q6 from Assignment 3, q, as a continuous function on $\mathbb{R}_+$ with bounded level sets,[1] attains a unique minimum on $\mathbb{R}_+$. Hence q is bounded below.

Alternatively, as computed in (ai), for $t > 0$, $q'(t) = 0$ iff $t = e^{-\gamma}$. So $t = e^{-\gamma}$ is the only stationary point of the strictly convex function q; moreover,

$$q(e^{-\gamma}) = e^{-\gamma}[\ln(e^{-\gamma}) - 1 + \gamma] = -e^{-\gamma} < 0 = q(0).$$

Therefore q attains global minimum at $t = e^{-\gamma}$. $\square$

---

[1] It is well-known that q is continuous on $\mathbb{R}_{++}$; q is also continuous at zero because $\lim_{t \searrow 0} t \ln t = \lim_{t \searrow 0} t^{-1}/(-t^{-2}) = 0$ by L'hôpital's rule.)

(b) For a fixed $c \in \mathbb{R}^n$, define $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ by

$$g(x) := \sum_{i=1}^{n} q_i(x_i), \quad \text{where} \quad q_i(t) := c_i t + p(t).$$

(i) g is strictly convex on $\mathbb{R}_+^n$.

Proof. Since q_i is strictly convex on $\mathbb{R}_+$ for $i = 1, \ldots, n$ as shown in (ai), for any distinct $x, y \in \mathbb{R}_+^n$ and $\lambda \in (0,1)$, $g(\lambda x + (1-\lambda)y) = \sum_{i=1}^{n} q_i(\lambda x_i + (1-\lambda)y_i) < \sum_{i=1}^{n} \lambda q_i(x_i) + (1-\lambda)q(y_i) = \lambda g(x) + (1-\lambda)g(y)$.
$\square$

(ii) g has compact level sets.

Proof. Fix any $\alpha \in \mathbb{R}$. If there exists a sequence $\{y^i\}_i \subset \{x \in \mathbb{R}^n : g(x) \le \alpha\} \subseteq \mathbb{R}_+^n$ such that $\{y^i\}_i$ is unbounded, then there must exist some index $j \in \{1, \ldots, n\}$ such that $\{y_j^i\}_i \in \mathbb{R}_+$ is unbounded. By (aiii), $\{q_j(y_j^i)\}_i \in \mathbb{R}_+$ must be unbounded. But again by (aiii) this means $g(y^i) \ge q_j(y_j^i) - \sum_{k \ne j} e^{-c_k} \to +\infty$, contradicting the fact that $\{y^i\}_i \subset \{x \in \mathbb{R}^n : g(x) \le \alpha\}$. Thus $\{x \in \mathbb{R}^n : g(x) \le \alpha\}$ must be a bounded set.

Since q_i is continuous on $\mathbb{R}_+$ as mentioned in (aiii), g is continuous on $\mathbb{R}_+^n$. This shows that $\{x \in \mathbb{R}^n : g(x) \le \alpha\} \subseteq \mathbb{R}_+^n$ is a compact set. $\square$

Question 2. *For any* $\bar{x} \in \mathrm{int}(\mathbb{R}_+^n) = \mathbb{R}_{++}^n$ *and* $x \in \mathrm{bd}(\mathbb{R}_+^n) = \{x \in \mathbb{R}^n : x_i = 0 \text{ for some } i\}$, *the directional derivative* $f'(x; \bar{x} - x) = -\infty$.

Proof. Fix any j such that $x_j = 0$, (noting that $\bar{x}_i > 0$,)

$$
\begin{aligned}
\frac{p(x_j + t(\bar{x}_j - x_j)) - p(x_j)}{t} &= \frac{p(t\bar{x}_j)}{t} \\
&= \frac{(t\bar{x}_j)(\ln t\bar{x}_j - 1)}{t} \\
&= \bar{x}_j(\ln t + \ln \bar{x}_j - 1).
\end{aligned}
$$

finitely valued for all i for sufficiently small $t > 0$, we have (recalling Claim 1 from Assignment 1) By convexity of p and the fact that $p(\bar{x}_i + t(\bar{x}_i - x_i)) - p(x_i)$ is finitely valued for all i, we have (recalling Claim 1 from Assignment 1)

$$
\begin{aligned}
\frac{f(x_i + t(\bar{x}_i - x_i)) - f(x_i)}{t} &= \sum_{i=1}^{n} \frac{p(x_i + t(\bar{x}_i - x_i)) - p(x_i)}{t} \\
&\le \frac{p(x_j + t(\bar{x}_j - x_j)) - p(x_j)}{t} + \sum_{i \ne j}[p(\bar{x}_i) - p(x_i)] \\
&= \bar{x}_j(\ln t + \ln \bar{x}_j - 1) + \sum_{i \ne j}[p(\bar{x}_i) - p(x_i)] \to -\infty
\end{aligned}
$$

as $t \searrow 0$. Therefore $f'(x; \bar{x} - x) = \lim_{t \searrow 0} t^{-1}[f(x_i + t(\bar{x}_i - x_i)) - f(x_i)] = -\infty$. $\square$

Question 3. If there exists $\tilde{x} \in \mathbb{R}^n_+ (= \mathrm{dom}(f))$ such that $Ax = b$, then the level set $\{x \in \mathbb{R}^n : g(x) \leq g(\tilde{x}), \ Ax = b\} \in \mathbb{R}^n_+$ is nonempty and, by Question 2, compact. Therefore there exists

$$\bar{x} \in \arg\min_x \{g(x) : g(x) \leq g(\tilde{x}), \ Ax = b\}.$$

Obviously, $\bar{x}$ is a global solution of (PE). In fact, $\bar{x}$ is the only global solution: if $\bar{x}_1$ and $\bar{x}_2$ are distinct global solutions of (PE), then $(\bar{x}_1 + \bar{x}_2)/2$ is also feasible for (PE) and attains a strictly lower objective value than x_1 and x_2 by strict convexity of g. This contradiction indicates that there can be only one global solution of (PE).

Now we show that $\bar{x}$ must lie in $\{x \in \mathbb{R}^n_{++} : Ax = b\}$ whenever itis non-empty[2]. If $\bar{x} \notin \{x \in \mathbb{R}^n_{++} : Ax = b\}$ which is non-empty, then there exists $x \in \{x \in \mathbb{R}^n_{++} : Ax = b\}$ and $j \in \{1, \ldots, n\}$ such that $\bar{x}_j = 0$ but $x_j > 0$. Then by Q2, $f'(\bar{x}; x - \bar{x}) = -\infty$ and this means $g'(\bar{x}; x - \bar{x}) = -\infty$, so $g(\bar{x}) > g(\bar{x} + t(x - \bar{x}))$ for sufficiently small $t \in (0, 1)$,[3] contradicting the optimality of $\bar{x}$.

Next we derive $\bar{x}$ explicitly. Assume for now that A is of full row rank.By open-mapping theorem, $A(\mathbb{R}^n_{++})$ is open, so $b \in A(\mathbb{R}_{++})$ implies there exists $\delta > 0$ such that $B(b, \delta) \subset A(\mathbb{R}^n_{++}) \subseteq A(\mathbb{R}^n_+)$. Therefore $b \in \text{int} A(\mathbb{R}^n_+)$. By Corollary 3.3.11 of Borwein and Lewis, strong duality holds for the Fenchel primal-dual pair:

$$\inf_{x \in \mathbb{R}^n} \{g(x) : Ax = b\} = \sup_{\phi \in \mathbb{R}^m} \{b^T \phi - g^*(A^T \phi)\} = b^T \bar{\phi} - g^*(A^T \bar{\phi})$$

for some $\bar{\phi} \in \mathbb{R}^m$. Since $\bar{x}$ solves the primal problem, we have

$$g(\bar{x}) + g^*(A^T \bar{\phi}) = b^T \bar{\phi} = \bar{x}^T(A^T \bar{\phi}) \quad \text{(by feasibility of } \bar{x}\text{).}$$

By Fenchel-young inequality, we have that $A^T \bar{\phi} \in \partial g(\bar{x})$.But $\bar{x} > 0$, so g is indeed differentiable at $\bar{x}$, and $\partial g(\bar{x})$ contains one single element, which is $\nabla g(\bar{x}) = [\ln \bar{x}_1 + c_1, \ldots, \ln \bar{x}_n + c_n]^T$. Therefore, for $j = 1, \ldots, n$,

$$\ln \bar{x}_j + c_j = (A\bar{\phi})_j \implies \bar{x}_j = \exp(A\bar{\phi} - c)_j.$$

Furthermore, $\bar{\phi}$ is a Lagrange multiplier for $\bar{x}$, that is, $L(\bar{x}, \bar{\phi}) \leq L(x, \bar{\phi})$ for all $x \in \mathbb{R}^n$, where $L(x, \phi) := g(x) + \phi^T(b - Ax)$ is the Lagrangian. In fact, by Strong Fenchel duality and Fenchel-Young inequality, for any $x \in \mathbb{R}^n$,

$$
\begin{aligned}
L(\bar{x}, \bar{\phi}) &= g(\bar{x}) \\
&= \bar{\phi}^T b - g^*(A^T \bar{\phi}) \\
&\leq \bar{\phi}^T b - \left[x^T(A^T \bar{\phi}) - g(x)\right] \\
&= g(x) + \bar{\phi}^T(b - Ax) = L(x, \bar{\phi}).
\end{aligned}
$$

----

[2]More generally, $\bar{x}$ must lie in $\text{ri}\{x \in \mathbb{R}^n_+ : Ax = b\}$, which is non-empty because $\{x \in \mathbb{R}^n_+ : Ax = b\}$ is convex and assumed to be non-empty.

[3]To be more rigorous, we note that from Q.2,

$$f(\bar{x} + t(x - \bar{x})) - f(\bar{x}) \leq t \left\{ x_j(\ln t + \ln x_j - 1) + \sum_{i \neq j}[p(x_i) - p(\bar{x}_i)] \right\}$$

$$\implies g(\bar{x} + t(x - \bar{x})) - g(\bar{x}) = f(\bar{x} + t(x - \bar{x})) - f(\bar{x}) + tc^T(x - \bar{x})$$

$$\leq t \left\{ x_j \ln t + x_j(\ln x_j - 1) + \sum_{i \neq j}[p(x_i) - p(\bar{x}_i)] + c^T(x - \bar{x}) \right\}$$

which is negative for sufficiently small $t \in (0, 1)$ because $\ln t \to -\infty$ as $t \searrow 0$. This contradicts the optimality of $\bar{x}$.

This shows that $\bar{x}$ is a global minimizer of $L(\cdot, \bar{\phi})$. Note that we essentially showed that if $\bar{x}$ and $\bar{\phi}$ solve the Fenchel primal-dual problem, then $\bar{\phi}$ is a Lagrange multiplier for $\bar{x}$.

It remains to deal with the case when A is not of full row rank.

If A is of row rank $m' < m$, there exist $\tilde{A} \in \mathbb{R}^{m' \times n}, \tilde{b} \in \mathbb{R}^{m'}$ such that $\tilde{A}x = \tilde{b}$ iff $Ax = b$, and the row rank of $\tilde{A}$ equals m'. The computation above indicates that $\bar{x}_j = \exp(\tilde{A}\tilde{\phi} - c)_j$ for all j, where $\tilde{\phi} \in \mathbb{R}^{m'}$ solves the maximization problem

$$\sup_{\phi \in \mathbb{R}^{m'}} \{\tilde{b}^T \phi - g^*(\tilde{A}^T \phi)\}.$$

Then we can show (see Claim 1 below) that $\bar{\phi} \in \mathbb{R}^m$, defined by $\bar{\phi}_{\mathcal{J}} = \tilde{\phi}$ (where $\mathcal{J}$ is the index set ofrows taken from A in forming $\tilde{A}$) and $\bar{\phi}_j = 0$ for all $j \notin \mathcal{J}$, solves the maximization problem

$$\sup_{\phi \in \mathbb{R}^m} \{b^T \phi - g^*(A^T \phi)\}.$$

It follows that $\bar{x}_j = \exp(A\bar{\phi} - c)_j$ for all j. Since $(\bar{x}, \bar{\phi})$ is a Fenchel primal-dual solution pair, $\bar{\phi}$ is a Lagrange multiplier for $\bar{x}$ as explained earlier.

Claim 1 *If $\tilde{\phi} \in \mathbb{R}^m$ solves the maximization problem*

$$\sup_{\phi \in \mathbb{R}^m} \{b^T \phi - g^*(A^T \phi)\},$$

and $\alpha = A^T \mu$ and $\beta = b^T \mu$ for some $\mu \in \mathbb{R}^m$ (that is, (α, β) is a linear combination of the rows of (A, b)), then for any $\eta \in \mathbb{R}$, $\bar{\phi} := (\tilde{\phi} - \eta\mu; \eta) \in \mathbb{R}^{m+1}$ solves the maximization problem

$$\sup_{\phi \in \mathbb{R}^{m+1}} \left\{ \begin{pmatrix} b \\ \beta \end{pmatrix}^T \phi - g^* \left(\begin{bmatrix} A \\ \alpha^T \end{bmatrix}^T \phi \right) \right\}.$$

In particular, $(\tilde{\phi}, 0)$ is a solution.

Proof. It suffices to rewrite the clumsy objective function in the new optimization problem: for any $(\psi, \eta) \in \mathbb{R}^{m+1}$,

$$\begin{aligned} \begin{pmatrix} b \\ \beta \end{pmatrix}^T \begin{pmatrix} \psi \\ \eta \end{pmatrix} - g^* \left(\begin{bmatrix} A \\ \alpha^T \end{bmatrix}^T \begin{pmatrix} \psi \\ \eta \end{pmatrix} \right) &= b^T \psi + \beta\eta - g^* \left(A^T \psi + \eta\alpha \right) \\ &= b^T \psi + b^T \mu\eta - g^* \left(A^T \psi + \eta A^T \mu \right) \\ &= b^T (\psi + \eta\mu) - g^* \left(A^T (\psi + \eta\mu) \right). \end{aligned}$$

Therefore

$$\begin{aligned} \sup_{\phi \in \mathbb{R}^{m+1}} \left\{ \begin{pmatrix} b \\ \beta \end{pmatrix}^T \phi - g^* \left(\begin{bmatrix} A \\ \alpha^T \end{bmatrix}^T \phi \right) \right\} &= \sup_{(\psi, \eta) \in \mathbb{R}^{m+1}} \{b^T (\psi + \eta\mu) - g^* \left(A^T (\psi + \eta\mu) \right)\} \\ &\leq \sup_{\phi \in \mathbb{R}^m} \{b^T \phi - g^*(A^T \phi)\} \end{aligned}$$

and for any $\eta \in \mathbb{R}$, taking $\psi = \tilde{\phi} - \eta\mu$ we have

$$b^T(\psi + \eta\mu) - g^*\left(A^T(\psi + \eta\mu)\right) = b^T\tilde{\phi} - g^*(A^T\tilde{\phi}) = \sup_{\phi \in \mathbb{R}^m}\{b^T\phi - g^*(A^T\phi)\}.$$

This shows that $(\psi; \eta)$ solves $\sup_{(\psi,\eta) \in \mathbb{R}^{m+1}}\left\{b^T(\psi + \eta\mu) - g^*\left(A^T(\psi + \eta\mu)\right)\right\}$. $\square$

3 Bonus Questions

3.1 DAD problem

First, recall that vec: $\mathbb{R}^{n \times n} \to \mathbb{R}^{n^2}$ defined by $\mathrm{vec}(M) = (M_1; M_2; \ldots; M_n)$ where $M_1, \ldots, M_n$ are the columns of M (that is, $\mathrm{vec}(M)$ stacks up the columns of M), is an isomorphism between $\mathbb{R}^{n \times n}$ and $\mathbb{R}^{n^2}$. In particular, if $\mathcal{J} \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ is non-empty and $\mathcal{L}$ is the subspace of $\mathbb{R}^{n \times n}$ consisting of matrices whose (i, j)-th entry is zero whenever $(i, j) \notin \mathcal{J}$, this isomorphism indicates that $\mathcal{L}$ is isomorphic to $\mathbb{R}^{|\mathcal{J}|}$ (by eliminating all the zero entries).

Thanks to this trivial isomorphism, the results in Q.2 can be extended to the case of matrices: given any such $\mathcal{J}$ and $\mathcal{L}$, and given a linear map $\mathcal{A} : \mathbb{R}^{n \times n} \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{n \times n}$, the entropy minimization problem

$$(PE') \qquad \min_{X \in \mathcal{L}} F(X) + \langle C, X \rangle$$

$$\text{s.t.} \quad \mathcal{A}(X) = b,$$

$$\text{where} \quad F(X) := \sum_{i,j} p(X_{ij}) \left(= \sum_{(i,j) \in \mathcal{J}} p(X_{ij}) \quad \text{for } X \in \mathcal{L} \right)$$

has the following properties:

1. The objective function is strictly convex on $\mathcal{L}_{++} := \{X \in \mathcal{L} : X_{ij} > 0 \ \forall (i, j) \in \mathcal{J}\}$.
   The objective function has compact level sets.

2. For any $X \in \mathrm{int}(\mathcal{L}_+)$ and $\bar{X} \in \mathrm{bd}(\mathcal{L}_+)$, $F'(\bar{X}; X - \bar{X}) = -\infty$.
   (Here $\mathcal{L}_+ := \{X \in \mathcal{L} : X_{ij} \geq 0 \ \forall (i, j) \in \mathcal{J}\}$.)

3. Whenever there exists an $X \in \mathcal{L}_+$ such that $\mathcal{A}(X) = b$, (PE') has a unique solution $\bar{X} \in \mathcal{L}$.
   Moreover, if there exists an $X \in \mathcal{L}_{++}$, $\bar{X}$ can be determined explicitly:

   $$\bar{X}_{ij} = \exp(\mathcal{A}^*\bar{\phi} - C) \quad \forall (i, j) \in \mathcal{J},$$

   where $\bar{\phi}$ is a solution to the dual problem

   $$\max_{\phi \in \mathbb{R}^m}\left\{\langle \phi, b \rangle - \left(F + \langle C, \cdot \rangle\right)^*(\mathcal{A}^*\phi)\right\}.$$

(a) *Necessity:* Given a matrix $A \in \mathbb{R}^{n \times n}$ with doubly stochastic pattern, let $\mathcal{J} := \{(i_1, j_1), (i_2, j_2), \ldots, (i_r, j_r)\}$ consist of indices of positive entries of A. Consider the subspace $\mathcal{L}$ of $\mathbb{R}^{n \times n}$ consisting of matrices whose (i, j)-th entry is zero whenever $(i, j) \notin \mathcal{J}$. The interior of $\mathcal{L}$ consists of matrices whose (i, j)-th entry are non-zero if and only if $(i, j) \in \mathcal{J}$.

Define $C \in \mathbb{R}^{n \times n}$ by

$$C_{ij} := \begin{cases} -\log A_{ij} & \text{if } A_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} .$$

Define $D_i := (0, \ldots, \mathbb{1}, \ldots, 0) \in \mathbb{R}^{n \times n}$ (that is, the i-column of D consists of all ones and all other columns are zero columns) and $F_i := D_i^T$ for $i = 1, \ldots, n$.

Consider the following minimization problem:

$$\min_{X \in \mathcal{L}} \quad F(X) + \langle C, X \rangle$$
$$\text{s.t.} \quad \langle D_i, X \rangle = 1 \quad \text{for } i = 1, \ldots, n$$
$$\langle F_i, X \rangle = 1 \quad \text{for } i = 1, \ldots, n,$$

where $\langle U, V \rangle := \text{tr}(U^T V)$ for matrices U and V of the same dimension.

Since A has doubly stochastic pattern, there exists $X \in \mathcal{L}_{++}$ such that X is doubly stochastic. Then X is feasible, so there exists a unique solution $\bar{X}$ for the given problem: there exists $(\bar{\phi}, \bar{\psi}) \in \mathbb{R}^{n+n}$ (which is a Lagrange multiplier of the minimization problem) such that for all $(i, j) \in \mathcal{J}$,[4]

$$\bar{X}_{ij} = \exp\big((\mathcal{D}^* \bar{\phi})_{ij} + (\mathcal{F}^* \bar{\psi})_{ij} - C_{ij}\big)$$
$$= \exp\big((\mathbb{J}\text{Diag}(\phi))_{ij} + (\text{Diag}(\psi)\mathbb{J})_{ij} - C_{ij}\big)$$
$$= \exp\big((\phi_j + \psi_i + \log A_{ij})\big)$$
$$= \exp(\phi_j)\exp(\psi_i)A_{ij}.$$

Therefore $\bar{X} = \Psi A \Phi$, where $\Psi = \text{Diag}(\exp(\psi_1), \ldots, \exp(\psi_n))$ and $\Phi = \text{Diag}(\exp(\phi_1), \ldots, \exp(\phi_n))$. By its feasibility for the minimization problem, $\bar{X}$ is doubly stochastic. This proves that if A has doubly stochastic pattern, then there exists diagonal matrices Ψ and Φ of strictly positive diagonal entries such that $\Psi A \Phi$ is doubly stochastic.

Sufficiency: Conversely, for any $A \in \mathbb{R}_+^{n \times n}$ such that $X := \text{Diag}(\psi)A\text{Diag}(\phi)$ is doubly stochastic for some $\psi, \phi \in \mathbb{R}_{++}^n$, since $X_{ij} = \psi_i \phi_j A_{ij}$, $X_{ij} > 0$ if and only if $A_{ij} > 0$. This shows that A has doubly stochastic pattern.

---

[4]If we write $\mathcal{D}(X) := [\langle D_i, X \rangle]_i$ and $\mathcal{F}(X) := [\langle F_i, X \rangle]_i$ for $i = 1, \ldots, n$, for any $\phi \in \mathbb{R}^n$, $\mathcal{D}^*(\phi) = \mathbb{J}\text{Diag}(\phi)$ and $\mathcal{F}^*(\phi) = \text{Diag}(\phi)\mathbb{J}$, where $\mathbb{J}$ is the $n \times n$ matrix of all ones.

3.2 Fenchel duality *vs.* Lagrangian duality

For any (finite dimensional) inner product space $\mathbb{E}$ and vector space Y with a partial-order-inducing closed convex cone K, proper convex functions $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ and K-convex function $g : \mathbb{E} \to Y$, we ask whether it is true that in the abstract convex program

$$\inf_{x \in \mathbb{E}} \ f(x) \quad \text{s.t.} \quad g(x) \preceq_K 0,$$

strong Fenchel duality implies strong Lagrangian duality, or vice verse. As it turns out, neither way is true.

Strong Fenchel duality does not imply strong Lagrangian duality

Consider the convex program

$$\min_{x \in \mathbb{R}} \ x \quad \text{s.t.} \quad x^2 \leq 0.$$

(Note that $\bar{x} = 0$ as the only feasible point is the optimal solution.) The Lagrangian is given by

$$L(x, \mu) = x + \mu x^2 = \begin{cases} \mu \left(x - \frac{1}{2\mu} \right)^2 - \frac{1}{4\mu} & \text{if } \mu > 0 \\ x & \text{if } \mu = 0 \end{cases}$$

$$\implies \quad \min_x L(x, \mu) = \begin{cases} -\frac{1}{4\mu} & \text{if } \mu > 0 \\ -\infty & \text{if } \mu = 0 \end{cases}.$$

The Lagrangian dual

$$\sup_{\mu \geq 0} \min_x \ L(x, \mu) = \sup_{\mu > 0} \left\{ -\frac{1}{4\mu} \right\}$$

has an optimal value 0 which is not attained by any feasible μ. Hence strong Lagrangian duality fails.

On the other hand, strong Fenchel duality holds: rewrite the convex program as

$$\min_x f(x) + h(x) := x + \delta_{\mathbb{R}_-}(x^2).$$

Now we compute the conjugates of f and h: for any $t \in \mathbb{R}$,

$$f^*(t) = \sup_x (tx - x) = \sup(t - 1)x = \begin{cases} 0 & \text{if } t = 1 \\ +\infty & \text{otherwise} \end{cases}$$

$$h^*(t) = \sup_x [tx - h(x)] = \sup \left\{ tx : x^2 \leq 0 \right\} = 0$$

Therefore the Fenchel dual problem is given by

$$\sup_t [-f^*(t) - h^*(-t)] \ = \ -\inf_t f^*(t) \ = \ -f^*(1) \ = \ 0,$$

so strong Fenchel duality holds.

Strong Lagrangian duality does not imply strong Fenchel duality

Consider the convex program

$$\min_{x \in \mathbb{R}} \ f(x) \quad \text{s.t.} \quad \delta_0(x) := \delta_{\{0\}}(x) \le 0,$$

where $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is the entropy function

$$f(x) := \begin{cases} x \log x - x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{if } x < 0 \end{cases} \quad .$$

The optimal solution is $\bar{x} = 0$ (as this is the only feasible point). The Lagrangian is given by

$$L(x, \mu) = f(x) + \mu \delta_0(x) = \begin{cases} \mu \delta_0(x) & \text{if } \mu > 0 \\ f(x) & \text{if } \mu = 0 \end{cases}$$

$$\implies \quad \min_x L(x, \mu) = \begin{cases} 0 & \text{if } \mu > 0 \\ -1 & \text{if } \mu = 0 \end{cases} \quad .$$

It follows that the Lagrangian dual

$$\max_{\mu \ge 0} \min_x \ L(x, \mu)$$

attains its maximum at any $\mu > 0$ (meaning that any $\mu > 0$ is a Lagrange multiplier). Therefore, strong Lagrangian duality holds.

Now we write down the Fenchel conjugates relevant for the Fenchel dual: as in the computation in entropy maximization problem, for any $t \in \mathbb{R}$,

$$f^*(t) \ = \ \sup_x [tx - f(x)] \ = \ -\inf_x [f(x) - tx] \ = \ -[f(e^t) - te^t] \ = \ e^t.$$

Letting $h(x) := \delta_{\mathbb{R}_-}(\delta_0(x))$, we have that for any $t \in \mathbb{R}$,

$$h^*(t) \ = \ \sup_x [tx - h(x)] \ = \ \sup_x \{tx : \delta_0(x) \le 0\} \ = \ \sup_x \{tx : x = 0\} \ = \ 0.$$

Therefore the Fenchel dual problem is given by

$$\sup_t [-f^*(t) - h^*(-t)] \ = \ -\inf_t f^*(t) \ = \ 0,$$

so the duality gap is zero, but obviously the dual optimality is not attained. Thus strong Fenchel duality does not hold.

Remark In both examples, the duality gap is zero, and one of the strong dualities fail because the dual does not attain the optimal value.

We notice that in both examples we have an obscure constraint that amounts to restrict $x = 0$. In both cases, the conjugate h^* that takes care of the constraint is identically zero, so whether strong Fenchel duality holds reduces to the question of whether the conjugate of the objective function attains global minimum. This marks the difference between the Fenchel duals of the first and the second problem. This seems to indicate that whether strong Fenchel duality holds or not depends at least partially on the global behavior of the Fenchel conjugates of relevant functions.

As for the well-studied Lagrangian duality, it appears that whether strong duality holds or not depends more on the way the constraint is formulated. For instance, in the second example where strong Lagrangian duality holds, if we replace the constraint by $x^2 \leq 0$, strong Lagrangian duality would fail for the same reason as in the first example.

This indicates that the concept of strong Lagrangian duality is rather "algebraic representation sensitive" (which is indeed subtlely hinted at if we think of the common constraint qualifications such as LICQ and MFCQ which are more algebraic than geometric in nature). On the other hand, Fenchel duality *seems* to be more immune from such sensitivity.

4 MATLAB

The difference map algorithm is destinated to solve the *set intersection problem*: given two sets $A, B \in \mathbb{R}^n$, find an $x \in \mathbb{R}^n$ such that $x \in A \cap B$. The algorithm relies on the ability to compute P_A and P_B, as can be seen in the following:

Algorithm 1 Difference map algorithm

1: Inputs: non-empty sets $A, B \in \mathbb{R}^n$, and $x \in \mathbb{R}^n$;
2: Set tolerance $\varepsilon > 0$; $\Delta \leftarrow 2\varepsilon$; $z \leftarrow x$;
3: **while** $\Delta > \varepsilon$ **do**
4:　　Compute $P_A(x)$ and $P_B(x)$;
5:　　Choose $\beta \in \{-1, 1\}$;
6:　　Compute

$$\begin{aligned} f_A(x) &= P_A(x) - \beta[P_A(x) - x], \\ f_B(x) &= P_B(x) + \beta[P_B(x) - x], \\ \Omega &= P_A\big(f_B(x)\big) - P_B\big(f_A(x)\big); \end{aligned}$$

7:　　$x \leftarrow x + \beta\Omega$, $z \leftarrow P_B(f_A(x))$, $\Delta \leftarrow \|\Omega\|$;
8: **end while**
　　RETURN z .

β on Line 6 is "determined through experimentation". When the while loop terminates, we get a point x such that

$$x \approx x + P_A\big(f_B(x)\big) - P_B\big(f_A(x)\big), \quad \text{i.e. } P_A\big(f_B(x)\big) \approx P_B\big(f_A(x)\big) \in A \cap B.$$

In this way we get a fixed point by computing $P_B(f_A(x))$, for instance.

As can be seen from the algorithm, the most expensive and difficult step is to compute the projections on sets A and B. It is not always possible if one of A and B is non-convex (the problem could be ill-posed in that case); but if A and B are both closed convex, the projection maps are well-defined and could be easily computed for special classes of convex sets.

Recall that $\bar{y} = P_A(x)$ if and only if $y \in A$ solves the minimization problem

$$\min_y \left\{ \frac{1}{2}\|x - y\|^2 : y \in A \right\}.$$

We see that if A is convex and if we have a nice representation of A, then the problem is actually tractable.

Implementing the algorithm...

Different ways of choosing β lead to slightly different versions of the difference map algorithm. One way is to *always* choose $\beta = 1$, as in Algorithm 2; another way is to compute the result for both $\beta = 1$ and $\beta = -1$, and then take whichever one that makes a greater progress, as in Algorithm 3. Naturally, Algorithm 3 takes significantly longer to terminate because more projections are done. As a heuristic, Algorithm 2 is probably a better one to go for, unless further information is given about the convex sets of interest.

Algorithm 2 Difference map algorithm, Version 1

1: Inputs: non-empty sets $A, B \in \mathbb{R}^n$, and $x \in \mathbb{R}^n$;
2: Set tolerance $\varepsilon > 0$, $\Delta \leftarrow 2\varepsilon$;
3: **while** $\Delta > \varepsilon$ **do**
4:     Compute $P_A(x)$ and $P_B(x)$;
5:     $x \leftarrow x + \big[P_A\big(2P_B(x) - x\big) - P_B(x) \big]$;
6:     $\Delta \leftarrow \big\| P_A\big(2P_B(x) - x\big) - P_B(x) \big\|$;
7: **end while**
    RETURN $P_B(x)$.

Algorithm 3 Difference map algorithm, Version 2

1: Inputs: non-empty sets $A, B \in \mathbb{R}^n$, and $x \in \mathbb{R}^n$;
2: Set tolerance $\varepsilon > 0$, $\Delta \leftarrow 2\varepsilon$, $z \leftarrow x$;
3: **while** $\Delta > \varepsilon$ **do**
4:     Compute $P_A(x)$ and $P_B(x)$;
5:     Compute:
6:     $\circ$ $f_A^+(x), f_B^+(x), \Omega^+$ corresponding to $\beta = 1$, and
7:     $\circ$ $f_A^-(x), f_B^-(x), \Omega^-$ corresponding to $\beta = -1$:

$$
\begin{array}{rcll}
f_A^+(x) = & x & = P_A(x) - [P_A(x) - x] \\
f_B^+(x) = & 2P_B(x) - x & = P_B(x) + [P_B(x) - x] \\
f_A^-(x) = & 2P_A(x) - x & = P_A(x) - (-1)[P_A(x) - x] \\
f_B^-(x) = & x & = P_B(x) + (-1)[P_B(x) - x]
\end{array}
$$

$$
\begin{aligned}
\Omega^+ &= P_A\big(f_B^+(x)\big) - P_B\big(f_A^+(x)\big) \\
\Omega^- &= P_A\big(f_B^-(x)\big) - P_B\big(f_A^-(x)\big).
\end{aligned}
$$

8:     **if** $\|\Omega^+\| \geq \|\Omega^-\|$ **then**
9:         $x \leftarrow x + \Omega^+$, $z \leftarrow P_B\big(f_A^+(x)\big)$, $\Delta \leftarrow \|\Omega^+\|$;
10:     **else**
11:         $x \leftarrow x - \Omega^-$, $z \leftarrow P_B\big(f_A^-(x)\big)$, $\Delta \leftarrow \|\Omega^-\|$;
12:     **end if**
13: **end while**
    RETURN z

Computational results : closest correlation matrix problem

We apply the difference map algorithm on the problem of projecting a random $n \times n$ matrix onto the set of symmetric positive semidefinite matrices with diagonal of all ones, by taking A as the PSD cone and B as the set of symmetric matrices with diagonal of all ones. The computed projection is then compared to the best approximation of X by measuring the distances of X to its projection and the distance from X to the set $A \cap B$. We use the operator norm (that is, the largest singular value) of matrix in the computation.

Nine sets of data were used in the numerical experiment. The largest matrix size that CVX can handle is 32 by 32, so our input matrices ranges from size 5 to 32.

The numerical result does not indicate significant difference between the computed results of Algorithm 2 and 3. As expected, Algorithm 3 takes much longer time to terminate. Interestingly, the output is far from being the best approximation of X as we can see from the distance of X to $A \cap B$.

Problem	Size	Time (s) (Ver. 1)	Time (s) (Ver. 2)	Overheads (Ver.1)	Overheads (Ver. 2)	Dist(X,Z) (Ver.1)	Dist(X,Z) (Ver.1)	Dist(X,$A \cap B$)
1	5	2.2758	4.6190	8	8	1.5361	1.5361	1.0856
2	5	1.9589	4.0655	7	7	3.1145	3.1145	1.6935
3	5	1.9392	4.0448	7	7	1.2609	1.2609	0.9959
4	10	3.1817	6.4658	9	9	6.4111	6.9103	1.9858
5	10	3.1945	6.4168	10	10	7.7248	7.7248	2.1469
6	10	2.7114	5.6836	9	9	6.6505	6.9725	2.4085
7	20	8.5419	18.1181	11	11	15.9036	15.9036	2.9235
8	30	22.9039	48.8439	14	14	25.8676	25.8676	3.7527
9	32	24.1903	53.9457	14	14	29.4057	29.4057	4.1762

5-by-5 matrix inputs and their results:

Problem	X					Output				
1	1.7904	1.0917	0.69350	1.4154	0.60587	1.0000	0.74833	0.58381	0.75657	0.56528
	1.0917	0.27173	0.81778	1.3663	0.42120	0.74833	1.0000	0.64252	0.79714	0.58172
	0.69350	0.81778	1.7373	0.85626	0.50337	0.58381	0.64252	1.0000	0.63732	0.48269
	1.4154	1.3663	0.85626	1.3627	0.78234	0.75657	0.79714	0.63732	1.0000	0.62083
	0.60587	0.42120	0.50337	0.78234	0.13864	0.56528	0.58172	0.48269	0.62083	1.0000
2	1.7059	0.45636	0.28667	1.3925	0.68911	1.0000	0.31936	0.22027	0.40188	0.32422
	0.45636	0.73692	0.58237	1.7528	1.6990	0.31936	1.0000	0.35863	0.55281	0.54961
	0.28667	0.58237	0.31853	1.1620	0.52296	0.22027	0.35863	1.0000	0.41103	0.32826
	1.3925	1.7528	1.1620	1.9610	1.4308	0.40188	0.55281	0.41103	1.0000	0.50767
	0.68911	1.6990	0.52296	1.4308	1.3380	0.32422	0.54961	0.32826	0.50767	1.0000
3	1.5442	0.55581	0.54504	1.4249	1.5048	1.0000	0.52555	0.54667	0.73060	0.73459
	0.55581	0.0040511	0.94756	0.66358	0.53595	0.52555	1.0000	0.62698	0.56668	0.51651
	0.54504	0.94756	0.81503	0.79213	0.78669	0.54667	0.62698	1.0000	0.59218	0.57277
	1.4249	0.66358	0.79213	0.62212	0.57944	0.73060	0.56668	0.59218	1.0000	0.61805
	1.5048	0.53595	0.78669	0.57944	0.79959	0.73459	0.51651	0.57277	0.61805	1.0000

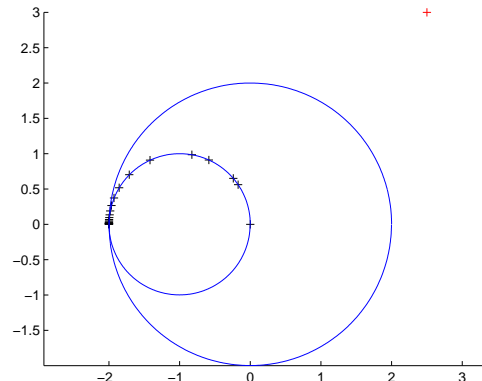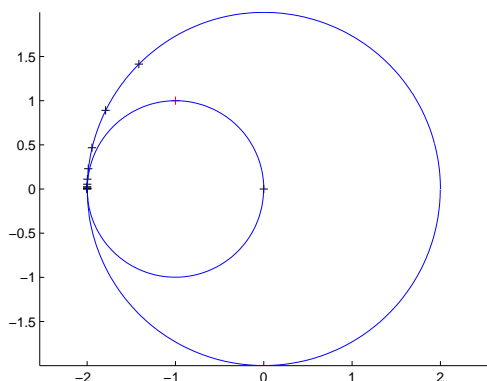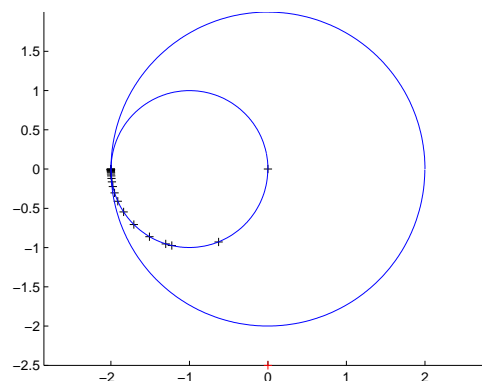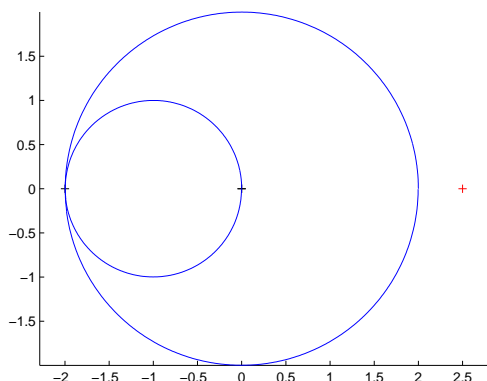## Computational results : projection on two circles

We apply the difference map algorithm on the problem of projecting a point $x \in \mathbb{R}^2$ onto the intersection of

$$A := \{y : \|y - (1,0)\| = 1\} \quad \text{and} \quad B := \{y : \|y\| = 2\}.$$

Since $A \cap B = \{(-2,0)\}$, the program should return something close to $(-2,0)$.

The choice of β is based on the position of the initial point x: $\beta = 1$ if $x_1 \geq 0$ and $\beta = -1$ otherwise. Four different initial points were tried in the numerical experiment, where the projections are explicitly computed without invoking CVX.

Problem	Time (s)	Overheads	Initial point	Result
1	9.4006e-5	3	(2.5, 0)	(-2, 0)
2	4.7338e-4	39	(0, -2.5)	(-2.0000, -1.2313e-5)
3	0.0048	16	(-1, 1)	(-2.0000, 3.6409e-5)
4	0.0020	37	(2.5, 3)	(-2.0000, 1.2025e-5)

Submitted on December 3, 2009.