The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Large Scale Optimization — Fall 2015

PROBLEM SET SEVEN SOLUTIONS

Constantine Caramanis

Matlab and Computational Assignments.

1. **MaxCut**.

   (a) The Petersen Graph: `http://en.wikipedia.org/wiki/Petersen_graph`.

   (b) Any two planar graphs of your choice (with a reasonable number of nodes).

   Solution The original integer programming is formulated as

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i<j}(1-u_i u_j)w_{ij} \\ \text{subject to} & u_i \in \{-1,1\},\ i=1,\dots,n. \end{array}$$

Relaxing the problem to an SDP, we get

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i<j}(1-v_i^\top v_j)w_{ij} \\ \text{subject to} & \|v_i\|_2 = 1,\ v_i \in \mathbb{R}^m, i=1,\dots,n. \end{array}$$

Note that v_i is not a number but a vector, so we will need further rounding off to obtain -1 or 1 for every node i after solving the SDP. For more discussion on this approach, please refer to [1] or [2]. Defining matrix X as $x_{ij} = v_i^\top v_j$, we rewrite the relaxed SDP formulation of MaxCut as

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i<j}(1-x_{ij})w_{ij} \\ \text{subject to} & x_{ii} = 1,\ i=1,\dots,n, \quad X \succeq 0, \end{array}$$

which is equivalent to

$$\begin{array}{ll} \text{minimize} & \text{Trace}(X^\top W) \\ \text{subject to} & x_{ii} = 1,\ i=1,\dots,n, \quad X \succeq 0. \end{array}$$

After we obtain a solution, which we denote by X^*, we can use the eigenvalue decomposition $X^* = V^\top D V$ to recover the vectors v_i, which is given by the ith row of $V^\top D^{1/2}$. For rounding the vectors to a number -1 or 1, we randomly select a hyperplane and separates the points v_i's using the hyperplane to determine which side each point is on.

A sample MATLAB code is given as follows.

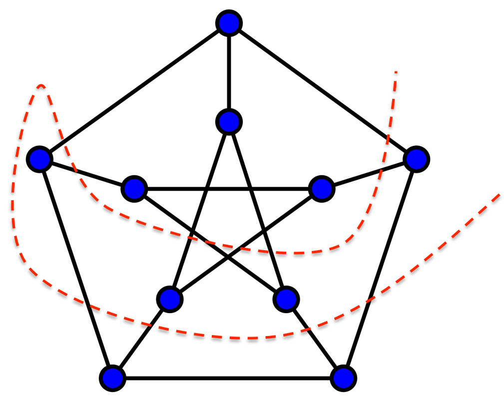

Figure 1: A maximum cut solution for the Petersen graph

```
function u = MaxCut(W)

cvx_begin sdp
        variable X(n,n) symmetric
        minimize trace(X'*W)
        subject to
                diag(X) = ones(n,1)
                X >= 0
cvx_end

[V,D] = eig(X);
u = sign((V*sqrt(D))*randn(n,1))
```

An example solution for the Petersen graph obtained from the above code is depicted in Figure 1. We can also apply the algorithm to any graphs.

Written Problems

1. *Show that the rank-constrained SDP is equivalent to the binary variable formulation of Max-Cut.*

   Solution. The rank-constrained version reads:

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i<j}(1-x_{ij})w_{ij} \\ \text{subject to} & x_{ii}=1, \quad i=1,\dots,n \\ & \operatorname{rank}(X)=1 \\ & X \succeq 0. \end{array}$$

   Recall that if $X \succeq 0$, then in particular, we know that

$$X=\sum_{i=1}^{r}\lambda_i \mathbf{v}_i\mathbf{v}_i^\top,$$

   where $\lambda_i > 0$, the $\{\mathbf{v}_i\}$ are orthonormal, and where $r = \operatorname{rank}(X)$. Therefore if $\operatorname{rank}(X) = 1$, the constraint $X \succeq 0$ now implies $X = \lambda \mathbf{v}\mathbf{v}^\top$. But now the constraint $x_{ii} = 1, \quad i = 1, \dots, n$ implies that $v_i \in \{+1, -1\}$, which is what we wanted to show.

2. *Problem 7.12 from Boyd & Vandenberghe.*

   Solution. This problem can be simply formulated as

$$\begin{array}{ll} \underset{S}{\text{maximize}} & \min_i \; p_i^\top s_i \\ \text{subject to} & \mathbf{1}^\top s_i = 1, \; s_i \geq 0, \; \forall i \end{array}$$

   where $p_i^\top$ and s_i denote the ith row of P and the ith column of S, respectively. We can just choose a probability distribution s_i for each i independently so that it maximizes $p_i^\top s_i$. It is given by $s_i = e_k$ where the kth entry of p_i is the maximum among the probability values. Therefore, the solution is given by

$$S_{ji} = \begin{cases} 1, & \text{if } j = \arg\max_k P_{ik} \\ 0, & \text{otherwise} \end{cases}$$

3. *Problem 7.13 from Boyd & Vandenberghe.*

   Solution The dual problem is given by

$$\begin{array}{ll} \text{maximize} & \sum_{j:\alpha_j\in S} p_j = \mathbf{prob}(X \in S) \\ \text{subject to} & \sum_{j=1}^{m} p_j = 1, \\ & \sum_{j=1}^{m} f_i(\alpha_i)p_j = b_i, \; i=1,\dots,n, \\ & p_j \geq 0, \; j=1,\dots,m, \end{array}$$

The optimum of the dual problem is a tight upper bound on $\mathbf{prob}(X \in S)$. Since the dual problem is a linear program and has a finite optimal solution, the strong duality holds. Therefore, the optimum of the primal problem is also a tight upper bound. There exists a distribution that achieves the bound.

4. *Problem 8.8 from Boyd & Vandenberghe.*

Solution (a) LP Formulation

$$\begin{array}{ll} \text{find} & x \\ \text{subject to} & Ax \leq b, \; Fx \leq g \end{array}$$

The strong alternative The dual function is given by

$$\begin{aligned} q(\lambda, \mu) &= \inf_x \{\lambda^\top (Ax - b) + \mu^\top (Fx - g)\} \\ &= \inf_x (\lambda^\top A + \mu^\top F)x - (\lambda^\top b + \mu^\top g) \\ &= \begin{cases} -(b^\top \lambda + g^\top \mu) & \text{if } A^\top \lambda + F^\top \mu = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

where $\lambda, \mu \geq 0$. The maximum of $q(\lambda, \mu)$ is bounded above unless

$$b^\top \lambda + g^\top \mu < 0, \quad A^\top \lambda + F^\top \mu = 0$$

for some $\lambda, \mu \geq 0$ because if the condition holds we can scale up λ and μ keeping $A^\top \lambda + F^\top \mu = 0$ to get any large number of $q(\lambda, \mu)$. The strong alternative is then given by

$$\begin{array}{ll} \text{find} & \lambda, \; \mu \\ \text{subject to} & \lambda \geq 0, \; \mu \geq 0, \; b^\top \lambda + g^\top \mu < 0, \; A^\top \lambda + F^\top \mu = 0. \end{array}$$

Geometric interpretation If $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$, there is a hyperplane $\{x : \alpha^\top x = \beta\}$ separating $\mathcal{P}_1$ and $\mathcal{P}_2$.

If the strong alternative is feasible, there exists α and β such that $\alpha = \lambda^\top A = -\mu^\top F$ and $\lambda^\top b < \beta < \mu^\top g$. Then we have

$$\begin{aligned} a^\top x = \lambda^\top Ax \leq \lambda^\top b < \beta, &\quad \text{if } x \in \mathcal{P}_1 = \{x : Ax \leq b\} \\ a^\top x = -\mu^\top Fx \geq -\mu^\top g > \beta, &\quad \text{if } x \in \mathcal{P}_2 = \{x : Fx \leq g\} \end{aligned}$$

(b) LP Formulation $\mathcal{P}_1 \subseteq \mathcal{P}_2$ if and only if the solution of the following LP

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f_i^\top x \\ \text{subject to} & Ax \leq b \end{array}$$

is smaller than or equal to g_i for every i where $f_i^\top$ is the ith row of F. As the strong duality holds, the optimal solution of the dual problem

$$\begin{array}{ll} \underset{y}{\text{minimize}} & b^\top y \\ \text{subject to} & A^\top y = f_i, \; y \geq 0 \end{array}$$

is also bounded above by g_i. Therefore, we can determine whether $\mathcal{P}_1 \subseteq \mathcal{P}_2$ by solving a set of LP feasibility problems

$$\begin{array}{ll} \text{find} & y \\ \text{subject to} & b^\top y \leq g_i, \ A^\top y = f_i, \ y \geq 0 \end{array}$$

for $i = 1, \ldots, p$.

Strong alternative The dual function for each LP feasibility problem is given by

$$\begin{aligned} q(\lambda, \mu) &= \inf_{y \geq 0} \{\lambda(b^\top y - g_i) + \mu^\top (f_i - A^\top y)\} \\ &= \inf_{y \geq 0} (\lambda b^\top - \mu^\top A^\top) y + \mu^\top f_i - \lambda g_i) \\ &= \begin{cases} f_i^\top \mu - g_i \lambda & \text{if } \lambda b - A\mu \geq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

where $\lambda \geq 0$. The maximum of $q(\lambda, \mu)$ is bounded above unless

$$f_i^\top \mu - g_i \lambda > 0, \quad \lambda b - A\mu \geq 0$$

for some $\lambda \geq 0, \mu$. The strong alternative is then given by

$$\begin{array}{ll} \text{find} & \lambda, \ \mu \\ \text{subject to} & \lambda \geq 0, \ f_i^\top \mu > g_i \lambda, \ A\mu \leq b\lambda. \end{array}$$

$\mathcal{P}_1 \subseteq \mathcal{P}_2$ if and only if the strong alternative is infeasible for every i.

Geometric interpretation Each of the strong alternatives is to find a point $x = \mu/\lambda$ such that $Ax \leq b$ and $f_i^\top x > g$. This implies that $x \in \mathcal{P}_1$ but $x \notin \mathcal{P}_2$. Therefore, $\mathcal{P}_1 \nsubseteq \mathcal{P}_2$ if and only if at least one of the strong alternatives is feasible.

$\boxed{\mathcal{P}_1 = \mathbf{conv}\{v_1, \ldots, v_K\}, \, \mathcal{P}_2 = \mathbf{conv}\{w_1, \ldots, w_L\}}$

(a) LP Formulation A point x in the intersection $\mathcal{P}_1 \cup \mathcal{P}_2$ can be written as a convex combination of each of the two groups of points, $\{v_1, \ldots, v_K\}$ and $\{w_1, \ldots, w_L\}$. An LP can be formulated as

$$\begin{array}{ll} \text{find} & x, \ \lambda, \ \mu \\ \text{subject to} & x = \sum_{i=1}^{K} \lambda_i v_i, \ \sum_{i=1}^{K} \lambda_i = 1, \ \lambda_i \geq 0 \\ & x = \sum_{j=1}^{L} \mu_j w_j, \ \sum_{j=1}^{L} \mu_j = 1, \ \mu_j \geq 0 \end{array}$$

Strong alternative The dual function is given by

$$\begin{aligned} q(a, b, \hat{a}, \hat{b}) &= \inf_{x, \lambda \geq 0, \mu \geq 0} \left\{ a^\top \left(x - \sum_i \lambda_i v_i \right) + b \left(\sum_i \lambda_i - 1 \right) + \hat{a}^\top \left(x - \sum_i \lambda_i v_i \right) + \hat{b} \left(\sum_i \lambda_i - 1 \right) \right\} \\ &= \inf_{x, \lambda \geq 0, \mu \geq 0} \left\{ (a + \hat{a})^\top x + \sum_i \lambda_i \left(b - a^\top v_i \right) + \sum_j \mu_j \left(\hat{b} - \hat{a}^\top w_j \right) - (b + \hat{b}) \right\} \\ &= \begin{cases} -(b + \hat{b}) & \text{if } a + \hat{a} = 0, \quad b - a^\top v_i \geq 0, \ \forall i, \quad \hat{b} - \hat{a}^\top w_j \geq 0, \ \forall j \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

The maximum of q is bounded above unless

$$b + \hat{b} < 0, \quad a + \hat{a} = 0, \quad a^\top v_i \le b, \ \forall i, \quad \hat{a}^\top w_j \le \hat{b}, \ \forall j$$

for some $a, \hat{a}, b, \hat{b}$. The strong alternative is then given by

$$\begin{array}{ll} \text{find} & a, \ \hat{a}, \ b, \ \hat{b} \\ \text{subject to} & b + \hat{b} < 0, \quad a + \hat{a} = 0, \quad a^\top v_i \le b, \ \forall i, \quad \hat{a}^\top w_j \le \hat{b}, \ \forall j \end{array}$$

which can be rewritten as

$$\begin{array}{ll} \text{find} & a, \ b \\ \text{subject to} & a^\top v_i \le b, \ \forall i, \quad a^\top w_j > b, \ \forall j \end{array}$$

Geometric interpretation The strong alternative finds a hyperplane which separates the two groups of the points, $\{v_1, \dots, v_K\}$ and $\{w_1, \dots, w_L\}$. Consequently, the hyperplane separates $\mathcal{P}_1$ and $\mathcal{P}_2$.

(b) LP Formulation $\mathcal{P}_1 \subseteq \mathcal{P}_2$ if and only if $v_i \in \mathbf{conv}\{w_1, \dots, w_L\}$ for every $i = 1, \dots, K$. A set of LP feasibility problems can be formulated as

$$\begin{array}{ll} \text{find} & \mu \\ \text{subject to} & v_i = \sum_{j=1}^L \mu_j w_j, \ \ \sum_{j=1}^L \mu_j = 1, \ \ \mu \ge 0 \end{array}$$

for $i = 1, \dots, K$.

Strong alternative The dual functions of the LP feasibility problems are given by

$$\begin{aligned} q_i(a, b) &= \inf_{\mu \ge 0} \left\{ a^\top \left(v_i - \sum_j \mu_j w_j \right) + b \left(\sum_j \mu_j - 1 \right) \right\} \\ &= \inf_{\mu \ge 0} \left\{ a^\top v_i - b + \sum_j \mu_j (b - a^\top w_j) \right\} \\ &= \begin{cases} a^\top v_i - b & \text{if } b - a^\top w_j \ge 0, \ \forall j \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

for $i = 1, \dots, K$. The maximum of q_i is bounded above unless

$$a^\top v_i - b > 0, \quad b - a^\top w_j \ge 0, \ \forall j$$

for some a, b. The strong alternatives are given by

$$\begin{array}{ll} \text{find} & a, \ b \\ \text{subject to} & a^\top v_i > b, \quad a^\top w_j \le b, \ \forall j \end{array}$$

for $i = 1, \dots, K$. $\mathcal{P}_1 \subseteq \mathcal{P}_2$ if and only if the strong alternative is infeasible for every i.

Geometric interpretation The set of strong alternatives check whether v_i can be separated from $\mathbf{conv}\{w_1, \dots, w_L\}$ by a hyperplane. If one of the strong alternatives is feasible, then we can tell that $\mathcal{P}_1 \not\subseteq \mathcal{P}_2$.

5. *Problem 8.9 from Boyd & Vandenberghe.*

   Solution As discussed in Section 8.3.3, a matrix D is a Euclidean distance matrix if and only if

$$D_{ii} = 0,\ i = 1, \ldots, n, \quad D_{ij} \geq 0,\ i, j = 1, \ldots, n, \quad (I - (1/n)\mathbf{1}\mathbf{1}^\top)D(I - (1/n)\mathbf{1}\mathbf{1}^\top) \preceq 0$$

   Since the constraint is a set of linear equalities, linear inequalities, and an linear matrix inequality, the set of Euclidean distance matrix is convex.

   The problem can be formulated as

$$\begin{array}{ll} \text{minimize} & \sum_{i,j} (\sqrt{D_{ij}} - \hat{d}_{ij})^2 = \sum_{i,j} \left(D_{ij} - 2\hat{d}_{ij}\sqrt{D_{ij}} + \hat{d}_{ij}^2 \right) \\ \text{subject to} & D_{ii} = 0,\ i = 1, \ldots, n, \quad D_{ij} \geq 0,\ i, j = 1, \ldots, n, \\ & (I - (1/n)\mathbf{1}\mathbf{1}^\top)D(I - (1/n)\mathbf{1}\mathbf{1}^\top) \preceq 0 \end{array}$$

   Since $-\sqrt{D_{ij}}$ is a concave function of D_{ij}, the objective function is convex. The optimization problem is a semidefinite programming.

   Once we obtained an optimum D^* from the semidefinite programming, we then compute the vectors $x_1, \ldots, x_n$ by using eigenvalue decomposition of a Gram matrix

$$G = -\frac{1}{2}(I - (1/n)\mathbf{1}\mathbf{1}^\top)D^*(I - (1/n)\mathbf{1}\mathbf{1}^\top) = X^\top \Lambda X = (\Lambda^{1/2}X)^\top(\Lambda^{1/2}X)$$

   The columns of $(\Lambda^{1/2}X)$ will be estimated vectors in $\mathbb{R}^n$. However, the dimension k can be smaller than n if there is a zero eigenvalue. The coordinate corresponding to the zero eigenvalue will always be equal to zero for every estimated vector. This coordinate can be removed, so the dimension k becomes smaller. There is at least one zero eigenvalue because the column and row spaces of G are orthogonal to span$\{\mathbf{1}\}$. There is also a possibility that some eigenvalues relatively close to zero and their corresponding coordinates are induced by noise or error. We can remove these coordinates to get a smaller dimension.

   We see that this procedure will not have k greater than $n - 1$. This is natural because just as many as $n - 1$ dimensions are enough to describe any set of distances among n points.

6. *Problem 8.23 from Boyd & Vandenberghe.*

   Solution (a) If t^* is positive, we have

$$\begin{aligned} a^{*\top} x_i &\geq b^* + t^* > b^*,\ i = 1, \ldots, N \\ a^{*\top} y_i &\leq b^* - t^* < b^*,\ j = 1, \ldots, M \end{aligned}$$

   where (a^*, b^*) is the optimal solution achieving t^*. The function $f(x) = a^{*\top}x - b^*$ separates the two sets of points.

   Conversely, if the two sets of points can be linearly separated, there exists a linear separator $f(x) = a^\top x - b$ such that

$$\begin{aligned} a^\top x_i &> b,\ i = 1, \ldots, N \\ a^\top y_i &< b,\ j = 1, \ldots, M \end{aligned}$$

Suppose $t = \min\{\inf_i(a^\top x_i - b), \inf_i(b - a^\top y_i)\} > 0$. Since (a, b, t) is a feasible solution of (8.23), t^*, which is greater than or equal to t, is also positive.

Tightness of $\|a\|_2 \le 1$ Suppose that the optimal solution (a^*, b^*, t^*) satisfies $\|a^*\| < 1$. Defining $\hat{a} = a^*/\|a^*\|_2$, we get

$$\hat{a}^\top x_i - \frac{b^*}{\|a^*\|_2} \ge \frac{t^*}{\|a^*\|_2},\ i = 1, \dots, N$$
$$\hat{a}^\top y_i - \frac{b^*}{\|a^*\|_2} \le -\frac{t^*}{\|a^*\|_2},\ i = 1, \dots, M$$

which implies that $(\hat{a}, b^*/\|a^*\|_2, t^*/\|a^*\|_2)$ is also a feasible solution. Since $t^*/\|a^*\|_2 > t^*$, t^* is not the optimum. Contradiction. This proves that $\|a^*\|_2 = 1$.

(b) As the constraint $\|a\|_2 \le 1$ is tight, we can replace it by $\|a\|_2 = 1$. Using the change of variables, the problem (8.23) can be written as

$$\begin{array}{ll} \text{maximize} & \|a\|_2/\|\tilde{a}\|_2 \\ \text{subject to} & \tilde{a}^\top x_i - \tilde{b} \ge 1,\ i = 1, \dots, N \\ & \tilde{a}^\top x_i - \tilde{b} \le -1,\ i = 1, \dots, M \\ & \|a\|_2 = 1 \end{array}$$

Since maximization of $\|a\|_2/\|\tilde{a}\|_2$ for $\|a\|_2 = 1$ is equivalent to minimization of $\|\tilde{a}\|_2$, this is equivalent to the QP.

7. *Problem 8.24 from Boyd & Vandenberghe.*

Solution The problem can be formulated as

$$\begin{array}{ll} \underset{a,b,\rho}{\text{maximize}} & \rho \\ \text{subject to} & (a+u)^\top x_i \ge b,\ \forall u \in \{u : \|u\|_2 \le \rho\},\ i = 1, \dots, N \\ & (a+u)^\top y_i \le b,\ \forall u \in \{u : \|u\|_2 \le \rho\},\ i = 1, \dots, M \\ & \|a\|_2 \le 1. \end{array}$$

The discriminative constraints can be written instead as

$$a^\top x_i + \left(\begin{array}{rl} \min_u & u^\top x_i \\ \text{s.t.} & \|u\|_2 \le \rho \end{array} \right) \ge b,\ i = 1, \dots, N$$
$$a^\top y_i + \left(\begin{array}{rl} \max_u & u^\top y_i \\ \text{s.t.} & \|u\|_2 \le \rho \end{array} \right) \le b,\ i = 1, \dots, M$$

The optimal solutions are given by $u^* = -\rho x_i/\|x_i\|_2$ and $u^* = \rho y_i/\|y_i\|_2$. Then we get a linear program

$$\begin{array}{ll} \underset{a,b,\rho}{\text{maximize}} & \rho \\ \text{subject to} & a^\top x_i + \rho\|x_i\|_2 \ge b,\ i = 1, \dots, N \\ & a^\top y_i + \rho\|y_i\|_2 \le b,\ i = 1, \dots, M \\ & \|a\|_2 \le 1. \end{array}$$

which finds a and b that maximize the weight error margin.

8. *Problem 8.25 from Boyd & Vandenberghe.*

   Solution A separating ellipsoid is specified by $A \succ 0$, b, and c such that

$$x_i^\top A x_i + b^\top x_i + c < 0,\ \ i = 1, \ldots, N$$
$$y_i^\top A y_i + b^\top y_i + c > 0,\ \ i = 1, \ldots, M$$

   To minimize the condition number of A, we can lower-bound the minimum eigenvalue and minimize the maximum eigenvalue. This can be described as

$$\begin{array}{ll} \underset{A,b,c,\alpha}{\text{minimize}} & \alpha \\ \text{subject to} & x_i^\top A x_i + b^\top x_i + c < 0,\ \ i = 1, \ldots, N \\ & y_i^\top A y_i + b^\top y_i + c > 0,\ \ i = 1, \ldots, M \\ & I \preceq A \preceq \alpha I \end{array}$$

   which is an SDP.

9. *Optional: 8.2 from Boyd & Vandenberghe.*

   Solution

   (a) A Chebyshev set is nonempty by definition. To show that it is closed, suppose that a Chebyshev set C is not closed, i.e., there is a limit point $x \notin C$ such that for any $\epsilon > 0$, some point $c \in C$ satisfies $\|x - c\| < \epsilon$. Then we cannot define $P_C(x)$, which contradicts the definition. A Chebyshev set is closed.

   (b) It is sufficient to show that for any sequence $\{x_n\}_{n=1}^\infty$ converging to x, i.e., $\lim_{n\to\infty} x_n = x$, the sequence $\{P_C(x_n)\}_{n=1}^\infty$ converges to $P_C(x)$, i.e., $\lim_{n\to\infty} P_C(x_n) = P_C(x)$. We have

$$\|x - P_C(x_n)\| - \|x_n - x\| \leq \|x_n - P_C(x_n)\| \leq \|x_n - P_C(x)\| \leq \|x_n - x\| + \|x - P_C(x)\|$$

   where the first and the third inequality follow by the triangle inequality, and the second inequality follows from that $P_C(x_n)$ is the closest point in C to x_n. We then take the limit to get

$$\limsup_{n\to\infty} \|x - P_C(x_n)\| \leq \|x - P_C(x)\|$$

   It follows from the continuity of the function $f(c) = \|x - c\|$ that for any converging subsequence $\{P_C(x_{n_t})\}_{t=1}^\infty$

$$\|x - \lim_{t\to\infty} P_C(x_{n_t})\| \leq \|x - P_C(x)\|$$

   Since C is closed, the limit point $\lim_{t\to\infty} P_C(x_{n_t})$ is also in C. There are no other points in C closer to x than $P_C(x)$, so we should have $\lim_{t\to\infty} P_C(x_{n_t}) = P_C(x)$. This means that the sequence $\{P_C(x_n)\}_{n=1}^\infty$ also converges to $P_C(x)$. P_C is continuous.

   (c) For any $c \in C$, we have

$$\|x_0 - P_C(x_0)\| \leq \|x_0 - c\| \leq \|x_0 - x\| + \|x - c\|$$

where the first inequality follows from the definition of $P_C(x_0)$, and the second inequality is the triangle inequality. Then we have for any $c \in C$,

$$\|x - c\| \geq \|x_0 - P_C(x_0)\| - \|x_0 - x\| = \|x - P_C(x_0)\|$$

where the equality holds because x lies on the line segment between x_0 and $P_C(x_0)$. This shows that $P_C(x) = P_C(x_0)$.

(d) We will show that for every point $x = \theta x_0 + (1-\theta)P_C(x_0)$ with $\theta > 1$, x_0 lies on the line segment between x and $P_C(x)$. Then (c) implies that $P_C(x) = P_C(x_0)$.

We use Brouwer's fixed point theorem: For any continuous function f that maps a sphere to the sphere itself, there exists a fixed point x where $f(x) = x$.

Fix a radius $r > 0$, and define a function

$$f(x) = x_0 + r\frac{x_0 - P_C(x)}{\|x_0 - P_C(x)\|}$$

on the sphere $x_0 + rB$ where $B = \{x : \|x\| \leq 1\}$. Note that $f(x) = x$ means that x_0 lies on the line segment between x and $P_C(x)$, and $\|x - x_0\| = r$.

As $f(x)$ is continuous, the fixed point theorem implies that there exists a point x where $f(x) = x$ for any $r > 0$. It follow from (c) that $P_C(x_0) = P_C(x)$. This means that x lies on the line passing x_0 and $P_C(x_0)$. It holds for any $r = \|x - x_0\|$, so every point on the line $x = \theta x_0 + (1-\theta)P_C(x_0)$ with $\theta > 1$ has its projection $P_C(x) = P_C(x_0)$.

(e) Suppose that C is not convex, so there are points $c_1, c_2 \in C$ such that $x_0 \triangleq \frac{1}{2}(c_1 + c_2) \notin C$. Since $P_C(x_0)$ is a unique point in C, the point can neither be c_1 nor c_2. Now consider a point $x = x_0 + \theta(x_0 - P_C(x_0))$ where θ is sufficiently large. It follows from (c) that $P_C(x) = P_C(x_0)$, which leads to

$$\|x - P_C(x)\|_2 < \min\{\|x - c_1\|, \|x_0 - c_2\|\},. \tag{1}$$

We will show that (1) is a contradiction.

First we claim that

$$\min\{\|x - c_1\|^2, \|x - c_2\|^2\} \leq \|x - x_0\|^2 + \|x_0 - c_2\|^2 \tag{2}$$

This can be proved as follows. WLOG, suppose $\|x - c_1\| \geq \|x - c_2\|$.

$$\begin{aligned}
4\|x - x_0\|^2 + 4\|x_0 - c_2\|^2 &= 4\|(x - (c_1 + c_2)/2)\|^2 + 4\|(c_1 - c_2)/2\|^2 \\
&= \|(x - c_1) + (x - c_2)\|^2 + \|c_1 - c_2\|^2 \\
&= \|x - c_1\|^2 + 2(x - c_1)^\top(x - c_2) + \|x - c_2\|^2 + \|c_1 - c_2\|^2 \\
&= \|x - c_1\|^2 + 2(x - c_1)^\top(c_1 - c_2) + \|c_1 - c_2\|^2 + \|x - c_2\|^2 + 2(x - c_1)^\top(x - c_1) \\
&= \|(x - c_1) + (c_1 - c_2)\|^2 + \|x - c_2\|^2 + 2(x - c_1)^\top(x - c_1) \\
&= \|x - c_2\|^2 + \|x - c_2\|^2 + 2\|x - c_1\|^2 \\
&\geq 4\|x - c_2\|^2 = 4\min\{\|x - c_1\|^2, \|x - c_2\|^2\}
\end{aligned}$$

Also, if $\|x - x_0\|$ is sufficiently large, we have

$$\begin{aligned} \|x - x_0\|^2 + \|x_0 - c_2\|^2 &< \|x - x_0\|^2 + 2\|x - x_0\| \cdot \|x_0 - P_C(x_0)\| + \|x_0 - P_C(x_0)\|^2 \\ &= \|x - x_0\|^2 + 2(x - x_0)^\top (x_0 - P_C(x_0)) + \|x_0 - P_C(x_0)\|^2 \\ &= \|x - x_0 + x_0 - P_C(x_0)\|^2 \\ &= \|x - P_C(x_0)\|^2 \\ &= \|x - P_C(x)\|^2. \end{aligned} \tag{3}$$

Putting (2) and (3) together, we obtain

$$\min\{\|x - c_1\|, \|x - c_2\|\} < \|x - P_C(x)\|,$$

which contradicts (1). Therefore, C is a convex set.

10. *Optional: 8.13 from Boyd & Vandenberghe.*

Solution The simple structure of the simplex allows us to find the symbolic form of the minimum volume ellipsoid (Löwner-John ellipsoid) and show that the shrunk ellipsoid is included in the simplex.

Step 1 Since the simplex $\mathbf{conv}\{0, e_1, \ldots, e_n\}$ is symmetric around the line $\{x : x = z\mathbf{1}, z \in \mathbb{R}\}$, the ellipsoid center must be on the line. Let $c\mathbf{1}$ denote the center. Therefore, the ellipsoid can be expressed as

$$\{x : (x - c\mathbf{1})^\top A(x - c\mathbf{1}) \le 1\}.$$

for some $A \in \mathbb{S}^n_+$ and $c \in \mathbb{R}$.

Step 2 Now consider the hyperplane $\{x : \mathbf{1}^\top x = 0\}$, which is orthogonal to the line $\{x : x = z\mathbf{1}, z \in \mathbb{R}\}$. We see that the projection of the ellipsoid should include the projection of the simplex, which is the $(n-1)$-dimensional tetrahedron. The smallest ellipsoid on the hyperplane that includes the tetrahedron is the $(n-1)$-dimensional sphere. This implies A has eigenvalue r_1 with the eigenvector $\frac{1}{\sqrt{n}}\mathbf{1}$ (orthogonal to the hyperplane) and eigenvalue r_2 with the rest of the eigenvectors (on the hyperplane). This can be expressed as

$$A = \frac{r_1}{n}\mathbf{1}\mathbf{1}^\top + r_2\left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right)$$

Step 3 The minimum volume ellipsoid should have the extreme points of the simplex on the boundary. Plugging each of points $\{0, e_1, \ldots, e_n\}$ into the inequality, we get

$$\begin{aligned} &(0 - c\mathbf{1})^\top A(0 - c\mathbf{1}) = c^2 r_1 n = 1, \\ &(e_i - c\mathbf{1})^\top A(e_i - c\mathbf{1}) = c^2 r_1 n + r_1/n + r_2(1 - 1/n) - 2r_1 c = 1. \end{aligned}$$

Solving the two equations, we obtain

$$r_1 = \frac{1}{nc^2}, \quad r_2 = \frac{2nc - 1}{(n-1)nc^2}$$

Step 4 The volume of the ellipsoid is proportional to

$$\log\det A^{-1} = \log\left(r_1 r_2^{n-1}\right)^{-1} = -\log r_1 - (n-1)\log r_2$$

The minimum volume should have the zero derivative, so we have

$$r_1 = \frac{(n+1)^2}{n}, \quad r_2 = \frac{n+1}{n}, \quad c = \frac{1}{n+1}.$$

Step 5 Now let us shrink the ellipsoid by a factor of n. We have

$$\hat{A} = nr_1\mathbf{1}\mathbf{1}^\top + n^2 r_2 \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\right) = n(n+1)(\mathbf{1}\mathbf{1}^\top + I)$$

The shrunk ellipsoid $\{x : (x - c\mathbf{1})^\top \hat{A}(x - c\mathbf{1}) \leq 1\}$ can be simply expressed as the set of x such that

$$nx^\top x + n(\mathbf{1}^\top x - 1)^2 \leq 1.$$

To have the shrunk ellipsoid included in the simplex, every point in the ellipsoid should have nonnegative entries, and the sum of the entries should be smaller than one. Thus, we solve the following optimization problems

$$\begin{array}{ll} \text{minimize} & e_i^\top x \\ \text{subject to} & nx^\top x + n(\mathbf{1}^\top x - 1)^2 \leq 1 \end{array} \tag{4}$$

for every i, and

$$\begin{array}{ll} \text{maximize} & \mathbf{1}^\top x \\ \text{subject to} & nx^\top x + n(\mathbf{1}^\top x - 1)^2 \leq 1. \end{array} \tag{5}$$

Solving (4) yields $\frac{1}{n}(\mathbf{1} - e_i)$, which ensures that any point in the ellipsoid cannot have negative number at the ith entry. Solving (5) results in $\frac{1}{n}\mathbf{1}$, which guarantees that every point is under the hyperplane $\{x : \mathbf{1}^\top x = 1\}$. This shows that the shrunk ellipsoid is in the simplex.

References

[1] Goemans, M. X. and Williamson, D. P. (1995). Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming, Journal of the Association for Computing Machinery, 42(6), 1115-1145.

[2] Example 3, `http://en.wikipedia.org/wiki/Semidefinite_programming`