E0 230 : Computational Methods for Optimization
Tutorial 4

1. Suppose f is a stritly convex function, and C is a compact convex set. What can you say about x^*, where
$$x^* = \arg\max_C f(x)?$$
What if $f(\cdot)$ is only convex, and not strictly convex?

Solution: Recall the Krein-Millmann theorem: (informally) A compact convex set can be represented as the convex hull of its extreme points. Let's represent the set of extreme points of C as ∂C. We claim that $x^* \in \partial C$. We'll show this by contradiction. First, let's suppose $w^* \in C \backslash \partial C$ (that is, w^* isn't an extreme point). Suppose we have $y_1, y_2 \in \partial C$ such that there exists some $\lambda \in (0, 1)$ such that $x^* = \lambda y_1 + (1-\lambda) y_2$. Then, by the convexity of f, we have

$$\begin{aligned} f(x^*) &= f(\lambda y_1 + (1-\lambda) y_2) \\ &< \lambda f(y_1) + (1-\lambda) f(y_2) \\ &< \max\{f(y_1), f(y_2)\}. \end{aligned}$$

Obviously, this cannot hold as $f(x^*) \geq f(x)\ \forall\ x \in C$. Thus, we prove the statement by contradiction.

If the function is not strictly convex, then the maximum can be attained at a non-extreme point of C. However, suppose that there exist $y_1, y_2 \in \partial C$ such that there exists a $\lambda \in (0, 1)$ where $x^* = \lambda y_1 + (1-\lambda) y_2$. Then, by the convexity of f, we have every point on the line segment between x^* and $\max\{f(y_1), f(y_2)\}$ is also a maximum of f.

If the function is not strictly convex, we may have interior points which are equivalent to the maximum achieved at the extreme points. Suppose we have $x^* in \partial C$ and $y \in C \backslash \partial C$ such that $f(y) = f(x^*) = \max_C f(x)$. Then, by the convexity of f, we may have that every point on the line segment between x^* and y is also a maximum (though not necessarily so).

2. Consider the problem

$$
\begin{aligned}
\underset{w,b,\{s_i\}_{i=1}^M}{\arg\min} \quad & \frac{1}{2}\|w\|^2 + C\sum_i s_i \\
\text{such that} \quad & y_i(w^T x_i + b) \geq 1 - s_i \qquad i \in \{1, ..., M\} \\
& s_i \geq 0 \qquad\qquad\qquad\quad\;\; i \in \{1, ..., M\}
\end{aligned}
$$

where $C > 0,\ y_i \in \{\pm 1\},\ x_i \in \mathbb{R}^n$ are given scalars. Find the dual of this problem.

Solution: First, we write the Lagrangian:

$$
\mathcal{L}(w, b, s, \lambda, \rho) \;= \frac{1}{2}\|w\|^2 + C\sum_i s_i - \sum_i \lambda_i(y_i(w^T x_i + b) - 1 + s_i) - \sum_i \rho_i s_i.
$$

This is clearly a convex function in (w, bs). The dual is given by

$$
g(\lambda, \rho) = \inf_{w,b,s} \mathcal{L}(w, b, s, \lambda, \rho).
$$

Thus, to find a minumum, we simply need to set the gradient w.r.t. w, b, and s to 0, and substitute the value back into the Lagrangian.

$$
\begin{aligned}
\nabla_w \mathcal{L} &= w - \sum_i \lambda_i y_i x_i = 0 \Rightarrow w = \sum_i \lambda_i y_i x_i \\
\nabla_b \mathcal{L} &= -\sum_i \lambda_i y_i = 0 \\
\nabla_s \mathcal{L} &= (CI - \operatorname{diag}(\rho + \lambda))\mathbf{1} = 0.
\end{aligned}
$$

Furthermore, we have $\rho_i = C - \lambda_i \geq 0 \Rightarrow 0 \leq \lambda_i \leq C$. Substituting these values back into the Lagrangian, we get

$$
\begin{aligned}
g(\lambda, \rho) &= \frac{1}{2}\|\sum_i \lambda_i y_i x_i)\|^2 - \sum_i \lambda_i((y_i x_i^T(\sum_j \lambda_j y_j x_j) + y_i b - 1) + \sum_i s_i(C - \lambda_i - \rho_i) \\
&= \frac{1}{2}\sum_i\sum_j \lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle - \sum_i\sum_j \lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle - b\sum_i \lambda_i y_i + \sum_i \lambda_i \\
&= -\frac{1}{2}\sum_i\sum_j \lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle + \sum_i \lambda_i
\end{aligned}
$$

Thus, the dual problem becomes

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\sum_i\sum_j \lambda_i\lambda_j y_i y_j \langle x_i, x_j\rangle + \sum_i \lambda_i \\
\text{s.t.}\;\; & 0 \leq \lambda_i \leq C.
\end{aligned}
$$

3. Consider the problem

$$m_\infty = \arg\min_m \sum_i \|x_i - m\|_\infty.$$

Can you reformulate this problem as a linear program? If so, what is the dual?

Solution: We can rewrite the problem as follows:

$$m_\infty = \arg\min_{t,m} \sum_i t_i \text{ such that } t_i = \max_j |x_{i,j} - m_j|.$$

From this, we get

$$\begin{aligned}
m_\infty = \arg\min_{t,m} &\quad \sum_i t_i \\
\text{s.t.} &\quad |x_{i,j} - m_j| \leq t_i \;\forall i, j \\
\Rightarrow m_\infty = \arg\min_{t,m} &\quad \sum_i t_i \\
\text{s.t.} &\quad x_{i,j} - m_j \leq t_i \;\forall i, j \\
&\quad m_j - x_{i,j} \leq t_i \;\forall i, j.
\end{aligned}$$

To find the dual, we first define the Lagrangian:

$$\begin{aligned}
\mathcal{L}(t, m, \lambda, \rho) &= \sum_i t_i + \sum_{i,j} \lambda_{i,j}(x_{i,j} - m_j - t_i) - \sum_{i,j} \rho_{i,j}(x_{i,j} - m_j + t_i) \\
&= \sum_{i,j} (\lambda_{ij} - \rho_i j) x_{i,j} + \sum_i t_i \left(1 + \sum_j \lambda_{ij} - \rho_{ij}\right) + \sum_j m_j \left(\sum_i \rho_{ij} - \lambda_{ij}\right)
\end{aligned}$$

The dual function is

$$g(\lambda, \rho) = \inf_{t,m} \mathcal{L}(t, m, \lambda, \rho).$$

Since the Lagrangian is affine, we see that the dual function becomes

$$g(\lambda, \rho) = \begin{cases} \sum_{i,j} (\lambda_{ij} - \rho_{ij}) x_{i,j} & 1 + \sum_j \lambda_{ij} - \rho_{ij} = 0 \;\forall i \\ & \sum_i \rho_{ij} - \lambda_{ij} = 0 \;\forall j \\ & \\ -\infty & \text{otherwise} \end{cases}$$

Thus, the dual problem becomes

$$\begin{aligned}
\arg\max_{\lambda,\rho} \quad & \sum_{i,j}(\lambda_{ij} - \rho_{ij})x_{i,j} \\
\text{subject to: } \quad & 1 + \sum_j \lambda_{ij} - \rho_{ij} = 0 \ \forall i \\
& \sum_j \rho_{ij} - \lambda_{ij} = 0 \ \forall j \\
& \lambda_{ij}, \ \rho_{ij} \geq 0 \ \forall \ i, j
\end{aligned}$$

4. Consider the problem

$$\begin{aligned}
\min \quad & -x - y \\
\text{such that: } \quad & x \leq 2, \ y \leq 1, \ 2x + 3y \leq 6, \ x, y \geq 0.
\end{aligned}$$

What are the vertices of the feasible set, and what are the values of the cost function at these points? Next, use the simplex algorithm to solve this problem. Lastly, describe the solution if we changed the cost to $-\frac{1}{3}x - \frac{1}{2}y$?

Solution: First, we see that the vertices of the feasible set are $(0,0)$, $(0,1)$, $(2,0)$, $(2,2/3)$, and $(1,3/2)$. The minimum is attained at $(x,y) = (2,2/3)$ and the value of the cost is $-8/3$.

Next, to use the simplex method, we convert the LP to standard form. This gives us

$$\begin{aligned}
\min \quad & -x - y \\
\text{such that: } \quad & x + t_1 = 2 \\
& y + t_2 = 1 \\
& 2x + 3y + t_3 = 6 \\
& x, y, t_1, t_2, t_3 \geq 0.
\end{aligned}$$

We write our problem data as:

$$c = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 3 & 0 & 0 & 1 \end{bmatrix}, \ b = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$$

We pick the first 3 columns as the basis for our BFS. Thus, we get

$$c_B = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}, \ c_N = \begin{bmatrix} 0 \\ 0 \end{bmatrix} B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix}, \ N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We can check the optimality of our BFS if

$$c_{test}^T = c^T - c_B^T B^{-1} A \geq 0.$$

We get

$$c_{test}^T = \begin{bmatrix} 0 & 0 & 0 & -0.5000 & 0.5000 \end{bmatrix}.$$

So we see that we can choose $j = 4$ as our descent direction. We now need to choose a stepsize. First, let's find $\tilde{A}_j = B^{-1}A_j = [-1.5, 1, 1.5]^T$. Obviously, we need not worry about the first coordinate. But we do need to think about the second coordinate. So next, define

$$\alpha = \min_{(\tilde{A}_j)_l > 0} \frac{(x_B)_l}{(\tilde{A}_j)_l} = \min\{1, 1/3\} = 1/3.$$

With this, we get

$$x_{new} = [x_B - \alpha A_j, x_N]^T + \alpha e_4 = \begin{bmatrix} 2.0000 & 0.6667 & 0 & 0.3333 & 0 \end{bmatrix}.$$

With this, we see our new basis $B = [1, 2, 4]$. We obtain $c_B,\ c_N$ and test for optimality, giving us

$$c_{test}^T = \begin{bmatrix} 0 & 0 & 0.3333 & 0 & 0.3333 \end{bmatrix}.$$

Thus, the BFS with basis B is optimal. We get $x_{OPT} = B^{-1} * b = [2, 2/3, 1/3]^T$. So we see that $(x_1, x_2) = (2, 2/3)$ which matches what we found previously with an exhaustive search.

5. Consider the polytope $S = \{x : c_i^T x + b_i \leq 0,\ i = 1, ..., M\}$. What is the projection of a point z onto $\{x : c_1^T x + b_1 \leq 0\}$. What is the projection onto S? Can you think of an algorithm that uses the first result to efficiently compute the projection of a point onto S?

Solution: The projection of the z onto the hyperplane solves

$$P_C(z) = \arg\min \ \frac{1}{2}\|x - z\|^2 \quad \text{such that} \ c_1^T x + b_1 \leq 0$$

can be solved using the KKT condtions:

$$x - z + \lambda c_1 = 0 \Rightarrow x = z - \lambda c_1 \tag{1}$$
$$\lambda \geq 0 \tag{2}$$
$$c_1^T x + b_1 \leq 0 \tag{3}$$
$$\lambda(c_1^T x + b_1) = 0 \tag{4}$$

If $\lambda = 0$, then the constraint is inactive, and we see that $x = z$. For $\lambda > 0$, we have $c_1^T x = -b$. Takign the inner product of the gradient of the Lagrangian with c_1, we get

$$c_1^T x + \lambda c_1^T c_1 - c_1^T z = 0$$
$$\Rightarrow \lambda = \frac{c_1^T z + b}{c_1^T c_1}$$

Thus, we get $x = z - \frac{c_1^T z+b}{c_1^T c_1} c_1$.

This gives us a close form solution to the distance from a point to a hyperplane. To find the projection of a point onto a polytope, we can use the following algorithm:

1. Initialize z
2. Compute $I(z) = \{i : c_i^T z + b > 0\}$
3. if $I(z) \equiv \emptyset$, $P_S(z) = z$. Else,
   (a) let $z_0 = z$,
   (b) for each $j = 1, 2, ...|I(z)|$, $z_j = P_{S_j}(z_{j-1}) = \frac{c_j^T z_{j-1}+b_j}{c_j^T c_j} \|c_j\|$.
4. $x = z_{|I(z)|}$

6. **(Optional)** Consider a convex set S that is the intersection of 2 convex sets C and D. Suppose we want to compute the projection of a point onto S but can't do so efficiently; however, we can efficiently compute projections onto C and D. Find an algorithm to compute the projection of a point onto S using the projections onto C and D. Show that this algorithm converges.

Solution: Algorithm:

1. Initialize: $x_0 = P_C(z)$
2. while $\|x_{k+1} - y_{k+1}\| \geq 0$, Compute:
   (a) $y_k = P_C(x_k)$
   (b) $x_{k+1} = P_D(y_k)$
3. Output $w = y_{k+1} = P_S(z)$

To show convergence, we simply need to show that $\|y_k - w\| \leq \|x_k - w\|$. We'll rely on the inequality

$$\langle x - P_C(x), P_C(x) - z\rangle \geq 0 \text{ for all } z \in C.$$

From this, we get

$$\langle x_k - y_k, y_k - w\rangle \geq 0 \text{ for all } s \in D,$$

where $w = P_S(z)$. Next, we have

$$\begin{aligned}
\|x_k - w\|^2 &= \|x_k - y_k + y_k - w\|^2 \\
&= \|x_k - y_k\|^2 + \|y_k - w\|^2 + y_k - w\|^2 + 2\langle x - P_C(x), P_C(x) - z\rangle \\
&\geq \|x_k - y_k\|^2 + \|y_k - w\|^2 \;\Rightarrow \|y_k - w\|^2 \leq \|x_k - w\|^2 - \|y_k - x_k\|^2.
\end{aligned}$$

Using the same technique, we can show that $\|x_{k+1} - w\|^2 \leq \|y_k - w\|^2 - \|x_{k+1} - y_k\|^2$. Thus, it follows that $\|x_k - w\| \leq \|x_0 - w\|$ and $\|y_k - w\| \leq \|x_0 - w\|$ for all k. Since this implies the sequence $\{x_k\}$ is bounded and decreasing, it must have an accumulation point, say $x^* \in C$. Then, since the sequence $\|x_0 - w\|, \|y_0 - w\|, ...$ is decreasing, it follows that $\|x_k - y_k\|$ and $\|x_{k+1} - y_k\|$ also decrease to 0 as k increases. Thus, since $y_k \in D$ and D closed, it follows that $\min_{z \in D} \|x_k - z\| \to 0$; this means that x_k converges to a point in D. However, we have $x_k \to x^* \in C$. Thus, $x^* \in C \cap D \equiv S$.