Honors Contract 189HC - Math 164

Seunghyun Kim

Wednesday December 5th, 2018

Contents

1 Introduction

In Math 164, we are interested in the problems of optimization, which takes the form of minimizing (or maximizing) a given function under constraints. Since minimizing and maximizing functions are interchangeable by multiplying the function by -1, we will confine our attention to minimization problems specifically.

During the Fall 2018 quarter, the student has met with the professor of the course once a week for an hour, to discuss mathematical concepts related to convexity and the minimum-cost flow problems. For the first half of the quarter, the student learned about some of the components of convex analysis, which is desirable in finding solutions to optimization problems. For the latter half of the quarter, the student learned about minimum-cost flow problems and linear programming.

2 Convex analysis

2.1 Convex sets and convex functions

In particular, the notion of convexity is useful when minimizing a function. In convex optimization, the constraint set and the function are both convex. In his book, Galichon defines a convex set C and a convex function Φ [2]:

Definition 1. A set C is convex if for all $x, y \in C$ and $t \in [0,1]$, $tx + (1-t)y \in C$.

Definition 2. A function $\Phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if for all $x, y \in \mathbb{R}^d$ and $t \in [0,1], \Phi(tx + (1-t)y) \leq t\Phi(x) + (1-t)\Phi(y)$ and is not identically $+\infty$.

So, a line between any two points of the function would lie above the convex function. Then, finding minimizer of a convex function over convex constraint would facilitate finding solutions to our objective problem. For example, Problem 6.29 in *An Introduction to Optimization* by Edwin K. P. Chong discusses least square methods, which finds line of best fit for a given function [3]. In other words, we wish to minimize the average squared distance between points of a given function and the line of best fit.

2.2 Legendre–Fenchel Transform

If the function is not convex, one can transform the function to its convex conjugate (i.e. Legendre–Fenchel Transform) such that the new function is convex.

Definition 3. For a function Φ not identically $+\infty$, the Legendre–Fenchel transform, or convex conjugate, is $\Phi^*(y) = \sup_{x \in \mathbb{R}^d}(x^T y - \Phi(x))$.

For instance, the following example illustrates the Legendre–Fenchel transform of quadratic functions [2]:

(i) For $\Phi(x) = |x|^2/2$, one gets $\Phi^*(y) = |y|^2/2$.

(ii) For $\Phi(x) = \sum_i \lambda_i (x_i)^2/2, \lambda_i > 0$, one gets $\Phi^*(y) = \sum_i (\lambda_i)^{-1}1(y_i)^2/2$.

(iii) More generally, if $\Phi(x) = \frac{1}{2}x^T \Sigma x$, then $\Phi^*(y) = \frac{1}{2}y^T \Sigma^{-1} y$, where Σ is positive definite.

We will prove part (iii) here.

Remark. Σ has to be symmetric matrix and $\Sigma \in \mathbb{R}^{d \times d}$. We can see this by setting $f(x) = \frac{1}{2}x^T \Sigma x$. If Σ is not symmetric, then Σ could be decomposed to symmetric matrix S and anti-symmetric matrix M: $\Sigma = S + M$. So, $f(x) = \frac{1}{2}x^T Sx + \frac{1}{2}x^T Mx$. But, $\frac{1}{2}x^T Mx = 0$, since $x^TMx \in \mathbb{R}$, so $x^TMx = (x^TMx)^T$. Note, $(x^TMx)^T = x^TM^T(x^T)^T = x^TM^Tx = x^T(-M)x$. Thus, $x^TMx = -x^TMx$, so $x^TMx = 0$ and $\frac{1}{2}x^TMx = 0$. Therefore, $f(x) = \frac{1}{2}x^TSx$.

Proof. By definition of Legendre–Fenchel transform, we have $\Phi^*(y) = \sup_{x\in\mathbb{R}^d}(x\cdot y - \Phi(x)) = \sup_{x\in\mathbb{R}^d}(x\cdot y - \frac{1}{2}x^T\Sigma x)$. Let $g(x) = x^Ty - \frac{1}{2}x^T\Sigma x$. Then, for all h in $\mathbb{R}^d$,

$$\begin{aligned}
g(x+h) &= (x+h)\cdot y - \frac{1}{2}(x+h)^T\Sigma(x+h) \\
&= x\cdot y + h\cdot y - \frac{1}{2}x^T\Sigma x - \frac{1}{2}x^T\Sigma h - \frac{1}{2}h^T\Sigma x - \frac{1}{2}h^T\Sigma h \\
&= g(x) + h^Ty - \frac{1}{2}x^T\Sigma h - \frac{1}{2}h^T\Sigma x + o(h).
\end{aligned}$$

Note, $h^Ty = (h^Ty)^T = y^Th$, and $h^T\Sigma x = (h^T\Sigma x)^T = x^T\Sigma^Th = x^T\Sigma h$. So, $g(x+h) = g(x) + (y^T - \frac{1}{2}x^T\Sigma - \frac{1}{2}x^T\Sigma)h + o(h)$. Thus, $\nabla g(x) = y^T - x^T\Sigma$, and $D^2f(x) = \Sigma$.

Using First-Order Necessary Condition, if x^* is a local minimizer for $g(x)$, then

$$\begin{aligned}
\nabla g(x^*) = 0 &\Rightarrow y^T = (x^*)^T\Sigma \\
&\Rightarrow (y)^T = (\Sigma x^*)^T \\
&\Rightarrow y = \Sigma x^* \\
&\Rightarrow x^* = \Sigma^{-1}y.
\end{aligned}$$

Now, using Sufficient Second-Order Condition, assuming x^* is a solution to $y = \Sigma x$, since $\nabla g(x^*) = 0$, and $D^2g(x^*) = \Sigma$, which is positive definite, x^* is a local minimizer of $g(x)$. Therefore,

$$\begin{aligned}
\Phi^*(y) &= \sup_{x\in\mathbb{R}^d}(x\cdot y - \frac{1}{2}x^T\Sigma x) \\
&= (\Sigma^{-1}y)^Ty - \frac{1}{2}(\Sigma^{-1}y)^T\Sigma(\Sigma^{-1}y) \\
&= y^T(\Sigma^{-1})^Ty - \frac{1}{2}y^T(\Sigma^{-1})^Ty \\
&= \frac{1}{2}y^T(\Sigma^{-1})^Ty.
\end{aligned}$$

Then, $\Sigma^{-1} = (\Sigma^{-1})^T$, since $I = I^T$, so $\Sigma\Sigma^{-1} = I = (\Sigma\Sigma^{-1})^T = (\Sigma^{-1})^T\Sigma^T$. Also, $\Sigma^{-1}\Sigma = (\Sigma^{-1})^T\Sigma^T = (\Sigma^{-1})^T\Sigma$. So, $\Sigma^{-1}\Sigma(\Sigma^{-1}) = (\Sigma^{-1})^T\Sigma(\Sigma^{-1})$.

Therefore, $\Phi^*(y) = \frac{1}{2}y^T\Sigma^{-1}y$. ☐

Similarly, the entropy function could be transformed into its convex conjugate:

Proof. Consider

$$\Phi(x) = \begin{cases} \sum_i x_i \ln x_i, & \text{if } x \geq 0, \sum_i x_i = 1 \\ +\infty, & \text{otherwise} \end{cases}$$

We want to show $\Phi^*(y) = \ln(\sum_i e^{y_i})$.
Thus, we are solving

$$
\begin{aligned}
\Phi^*(y) &= \max_{x \in \mathbb{R}^d} (x \cdot y - \Phi(x)) \\
&= \max_{x \in \mathbb{R}^d} (x \cdot y - \sum_i x_i \ln(x_i)) \text{ subject to } x \geq 0 \text{ and } \textstyle\sum_i x_i = 1. \\
&= \max_{x \geq 0} \min_{\mu} x \cdot y - \sum_i x_i \ln(x_i) + \mu(\sum_i x_i - 1). \\
&= \min_{\mu} \max_{x \geq 0} x \cdot y - \sum_i x_i \ln(x_i) + \mu(\sum_i x_i - 1).
\end{aligned}
$$

Since $x^T y = \sum_i x_i y_i$, for all i,

$$
\begin{aligned}
&\frac{\partial}{\partial x_i}(x^T y - \sum_i x_i \ln(x_i) + \mu(\sum_i x_i - 1)) = 0 \\
&\Rightarrow y_i - \ln(x_i) - 1 + \mu = 0 \\
&\Rightarrow \ln(x_i) = y_i - 1 + \mu \\
&\Rightarrow x_i = e^{y_i - 1 + \mu}.
\end{aligned}
$$

Plugging back into the constraint $\sum_i x_i = 1$, $\sum_i e^{y_i - 1 + \mu} = 1$. Define a constant c to be $c := e^{-1+\mu}$. Then, $c \sum_i e^{y_i} = 1$, so $c = (\sum_i e^{y_i})^{-1}$. So, $\mu = 1 + \ln[\sum_i e^{y_i}]^{-1}$.
Then,

$$
\begin{aligned}
\Phi^*(y) &= (\sum_i x_i y_i) - (\sum_i x_i \ln(x_i)) \\
&= \frac{\sum_i e^{y_i}(y_i)}{\sum_j e^{y_j}} - \frac{\sum_i e^{y_i} \ln \frac{e^{y_i}}{\sum_j e^{y_j}}}{\sum_j e^{y_j}} \\
&= \frac{\sum_i e^{y_i}(y_i)}{\sum_j e^{y_j}} - \frac{\sum_i e^{y_i}((y_i) - \ln(\sum_j e^{y_j}))}{\sum_j e^{y_j}} \\
&= \ln(\sum_j e^{y_j}).
\end{aligned}
$$

□

2.3 Minimization of convex functions

2.3.1 Subdifferentials

However, if the function is already convex, Legendre–Fenchel transform is still useful. Specifically, we use the idea of subdifferentials, which links the transform and convexity. In particular, the following definition and propositions use subdifferentials [2]:

Definition 4. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$. Then, the subdifferential of $\Phi(x)$ is
$$\partial\Phi(x) = \{y \in \mathbb{R}^d \mid \forall \widetilde{x} \in \mathbb{R}^d, \Phi(\widetilde{x}) \geq \Phi(x) + y \cdot (\widetilde{x} - x)\}.$$

Proposition 1. (Convex envelope)

(i) $\Phi^{**} \leq \Phi$, with equality if and only if Φ is convex.

(ii) Φ^{**} is the greatest minorant of Φ, i.e. $\Phi^{**}(x) = \sup\{g(x) : g \leq \Phi \text{ and g convex}\}$.

Proposition 2. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. The following statements are equivalent:

(i) $\Phi(x) + \Phi^*(y) = x^T y$

(ii) $y \in \partial\Phi(x)$

(iii) $x \in \partial\Phi^*(y)$

Proof. Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a convex function.
(ii) $\to$ (i)
Let $y \in \partial\Phi(x)$. Our goal is to show that $\Phi(x) + \Phi^*(y) = x^T y$.
By definition, for all $\widetilde{x} \in \mathbb{R}^d$,

$$\begin{aligned}
y \in \partial\Phi(x) &\Rightarrow \Phi(\widetilde{x}) \geq \Phi(x) + y^T(\widetilde{x} - x) \\
&\Rightarrow y^T(\widetilde{x} - x) - \Phi(\widetilde{x}) \leq -\Phi(x) \\
&\Rightarrow y^T\widetilde{x} - \Phi(\widetilde{x}) \leq xy^T - \Phi(x) \\
&\Rightarrow xy^T - \Phi(x) = \sup_{\widetilde{x}} \widetilde{x}y^T - \Phi(\widetilde{x}) = \Phi^*(y) \\
&\Rightarrow \Phi^*(y) + \Phi(x) = xy^T = x^T y.
\end{aligned}$$

(i) $\to$ (ii)
Let $\Phi(x) + \Phi^*(y) = x^T y$. Our goal is to show that y $\in \partial\Phi(x)$, i.e. $\forall z \in \mathbb{R}^d, \Phi(z) \geq \Phi(x) + y \cdot (z - x)$.
By definition, $\Phi^*(y) = \sup_{x \in \mathbb{R}^d} x^T y - \Phi(x)$. So,

$$\begin{aligned}
&\Phi(x) + \Phi^*(y) = x^T y \\
&\Rightarrow 0 = \Phi(x) + \Phi^*(y) - x^T y \\
&\Rightarrow 0 \geq \Phi(x) + y^T z - x^T y - \Phi(z) \\
&\Rightarrow \Phi(z) \geq \Phi(x) + y^T z - x^T y = \Phi(x) + y^T(z - x) \\
&\Rightarrow y \in \partial\Phi(x).
\end{aligned}$$

(i) → (iii)
Let $\Phi(x) + \Phi^*(y) = x^T y$. Our goal is to show that x $\in \partial\Phi^*(y)$, i.e. $\forall z \in \mathbb{R}^d, \Phi^*(z) \geq \Phi^*(y) + x \cdot (z - y)$. Then,

$$\begin{aligned}
\Phi^*(y) + x^T(z - y) &= \Phi^*(y) + x^T z - x^T y \\
&= \Phi^*(y) + x^T z - \Phi(x) - \Phi^*(y) \\
&= x^T z - \Phi(x) \leq \sup_{x \in \mathbb{R}^d} x^T z - \Phi(x) = \Phi^*(z).
\end{aligned}$$

So, $\Phi^*(z) \geq \Phi^*(y) + x^T(z - y)$.
(iii) → (i)
Let x $\in \partial\Phi^*(y)$. Our goal is to show that $\Phi(x) + \Phi^*(y) = x^T y$. By definition, for all z in $\mathbb{R}^d, \Phi^*(z) = \sup_{x \in \mathbb{R}^d} x^T z - \Phi(x)$,

$$\begin{aligned}
&\Phi^*(z) \geq x^T z - \Phi(x) \geq \Phi^*(y) + x^T(z - y) \\
&\Rightarrow \Phi^*(y) + x^T(z - y) = x^T z - \Phi(x) \\
&\Rightarrow \Phi^*(y) - x^T y = -\Phi(x) \\
&\Rightarrow x^T y = \Phi(x) + \Phi^*(y).
\end{aligned}$$

Thus, **(i)** $\iff$ **(ii)** $\iff$ **(iii)**. □

2.3.2 Fenchel's Duality Theorem

Furthermore, if there are two functions of interest, one of which is convex and the other concave, both not identically infinity, then Fenchel's Duality Theorem states that finding the smallest vertical distance between the two functions is equivalent to finding the largest distance between the tangents (with same slope) of the two functions, i.e. conversion of $\inf_x f(x) - g(x)$ to finding supremum of conjugate functions. Here, f is convex function on $\mathbb{R}^n$, and g is a concave function on $\mathbb{R}^n$. In the case that our original problem of finding infimum is hard to solve, this Duality Theorem conversion might facilitate finding the solution if it would be easier to find supremum of the conjugate functions.

Equivalently, since the negative of the concave function $g(x)$, $-g(x)$, could be interpreted as convex function $g(x)$, and that infimum (or supremum) could be approximately equal to minimum (or maximum) if the value is in the objective function, we can look at the problem $\min_x f(x) + g(x)$. Since $g(x)$ is convex, $g(x) = g^{**}(x) = \sup_{y \in \mathbb{R}^d} y^T x - g^*(y)$

by **Proposition 1**. Then, our problem becomes

$$\begin{aligned}
\min_x f(x) + g(x) &= \min_x f(x) + \sup_y y^T x - g^*(y) \\
&= \min_x \max_y f(x) + y^T x - g^*(y) \\
&= \max_y \min_x f(x) + y^T x - g^*(y) \\
&= \max_{\widetilde{y}} \min_x f(x) + (-\widetilde{y})^T x - g^*(-\widetilde{y}), \text{ letting } \widetilde{y} = -y \\
&= \max_{\widetilde{y}} -f^*(\widetilde{y}) - g^*(-\widetilde{y}) \\
&= \max_y -f^*(y) - g^*(-y).
\end{aligned}$$

3 Applications

3.1 Linear programming

A linear program, in standard form, is $\min c^T x$ subject to $Ax = b$, $x \geq 0$, with $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $m < n$, rank $A = m$, $b \geq 0$. Here, A, b, and c are known and fixed. If the linear program is not in a standard form, we can introduce surplus variables or slack variables to convert original inequality constraints to equality constraints in standard form.

3.1.1 Duality

In some cases, it might be desirable to convert the linear program, or the primal problem, to its corresponding dual problem. To do so, we need to introduce Lagrange multipliers and apply min-max principle to find the equivalent dual problem. Generally, if $F(x, \lambda)$ is convex in x and concave in λ for convex sets X and Λ, then $\min_{x \in X} \max_{\lambda \in \Lambda} F(x, \lambda) = \max_{\lambda \in \Lambda} \min_{x \in X} F(x, \lambda)$. Since the functions in linear program are linear and must necessarily satisfy the assumption, min-max principle must hold. So, if the solution λ^* to the dual problem can be found, the information may help us solve for x^*, the solution to our original primal problem.

For example, let the primal problem be our linear program in standard form. Introducing Lagrange multiplier λ and using min-max principle, we can rewrite this problem as a dual problem:

$$\begin{aligned}\min_{Ax=b, x\geq 0} c^T x &= \min_{x\geq 0}\max_{\lambda} c^T x + \lambda^T(b - Ax)\\ &= \max_{\lambda}\min_{x\geq 0} c^T x + \lambda^T(b - Ax)\\ &= \max_{\lambda}\min_{x\geq 0} b^T\lambda + (c - A^T\lambda)^T x\\ &= \max_{c - A^T\lambda\geq 0} b^T\lambda\\ &= \max_{c\geq A^T\lambda} b^T\lambda.\end{aligned}$$

3.2 Minimum-cost flow problem

A further application of the above linear programming problem is the minimum-cost flow problem. The min-cost problem, as will be abbreviated here, is looking for a solution that will minimize cost of moving one product (supply) to another node (demand) under constraints. Mathematically, the general min-cost flow problem is [1]:

$\min \sum_{(i,j)\in A} a_{ij} xij$ subject to:

$\sum_{j|(i,j)\in A} x_{ij} - \sum_{j|(j,i)\in A} x_{ji} = s_i \ \forall i \in \mathbb{N}$, $b_{ij} \leq x_{ij} \leq c_{ij} \forall (i,j) \in A$.

Here, i and j represent supply and demand nodes, respectively, and a_{ij}, b_{ij}, c_{ij}, and s_i are known scalars, where a_{ij} =cost of (i,j) and

$$s_i = \begin{cases}\text{supply of node } i, & \text{if } s_i > 0\\ \text{demand of node } i, & \text{if } s_i < 0\end{cases}$$

3.2.1 Assignment problem

In particular, the min-cost problem could equivalently be changed to an assignment problem. For an assignment problem, given same n number of people (e.g. workers) and number of jobs, we want to maximize total benefit of matching the two while completing all jobs. So here, a_{ij} = benefit of matching person i with object j. So, the assignment problem is of the form:

$\max \sum_{(i,j)\in A} a_{ij} x_{ij}$ subject to:

$\sum_{j|(i,j)\in A} x_{ij} = 1$ for $i \in \{1, ..., n\}$,

$\sum_{i|(i,j)\in A} x_{ij} = 1$ for $j \in \{1, ..., n\}$, with $0 \leq x_{ij} \leq 1 \ \forall (i,j) \in A$.

3.2.2 Auction algorithm

An auction algorithm is a specific example of an assignment problem or a min-cost problem. We are still solving to maximize the benefit of matching n persons to n objects on a one-to-one basis, but here, we assign i to j based on the value of j. So now, i wants to find an object j, among n objects, that offers maximum value when the price of object h is subtracted from the benefit of matching person i with object h. So, the selected object increases in price by how much i bids for the object.

4 Conclusion

In this paper, we looked at various ways to solve optimization problems. This process would be easier if we restrict our attention to convex functions subject to convex sets. To achieve our ultimate goal to solve minimization problem, it might be easier to convert our objective function to its convex conjugate, using ideas of Legendre–Fenchel Transform, subdifferentials, and Fenchel's Duality Theorem.

In everyday life, such optimization problems could be applied to linear programming problems, further to minimum-cost flow problems. The subsequent assignment problem and auction algorithm show specific examples of how the idea of optimization could be used when trying to maximize benefit of matching nodes, with the latter involving further the bidding process.

References

[1] Dimitri P. Bertsekas. *Network optimization: continuous and discrete models.* Citeseer, 1998.

[2] Alfred Galichon. *Optimal transport methods in economics.* Princeton University Press, 2016.

[3] E. K.P. Chong and S. Zak. *An Introduction to Optimization*, 4th Edition. Wiley.