THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
2017-2018 semester 2 MATH2010
week 9 tutorial

This tutorial is a brief introduction to optimization theory.

Example 1.

$$\min_{\substack{x+y\leq 4\\ x\geq 0\\ y\geq 0}} x^2 - 2xy + 3y^2 + 2x - 10y$$

Solution. Since the domain is closed and bounded and the objective function is continuous, the minimum exists. It is an interior critical point, or on the boundary.

Consider the interior. Since the function is smooth, there is no need to consider points where the function has no gradient. $\nabla f = [2x - 2y + 2, -2x + 6y - 10]^t$. Setting both entries to zero gives $(x, y) = (1, 2)$, which lies in the domain (it satisfies the defining inequalities), and where function value is -9.

Consider the boundary, which consists of the lines $y = 0$, $x = 0$ and $x + y = 2$, which can be parametrized. A boundary minimum is either a vertex or lies on an open segment, in which case it is a minimum of the function restricted on the open segment, and hence the derivative of the function value with respect to the parameter of the segment is 0, as in single-variable calculus.

On the first segment, the function is $x^2 + 2x$. Setting its derivative to 0 gives $x = -1$, but $(-1, 0)$ does not lie in the domain, and hence the function restricted to this open segment has no minimum. Similar computation shows the function restricted on the second open segment has a critical point $y = 5/3$, where the function value is $-25/3$.

Parametrizing the last segment by $y = 2 - x$, the objective function is $4x^2 - 10x$. Differentiating its derivative to 0 gives $x = 5/4$, and hence $y = 3/4$ and the function value is $-25/4$

Finally, the vertices of the triangle are also considered. The possible solutions and their function values are summariszed in the table below.

(x, y)	$f(x, y)$
$(1, 2)$	-9
$(0, 5/3)$	$-25/3$
$(5/4, 3/4)$	$-25/4$
$(4, 0)$	24
$(0, 4)$	8
$(0, 0)$	0

Comparing these vales, the global minimum is -9 and it is attained at $(1, 2)$.

Example 2.

$$\min_{\substack{\sum p_i=1\\ p_i\geq 0}} \sum p_i \log p_i$$

Solution. Note that the objective function is continuous, because $\lim_{t\to 0^+} t \log t = 0$.

The domain $\{p : \sum p_i = 1, p_i \geq 0\}$ has empty interior, so differential theory does not apply directly. Letting $q_i = p_i$ for $i < n$, and $p_n = 1 - \sum q_i$, the problem becomes

$$\min_{\substack{\sum q_i \leq 1 \\ q_i \geq 0}} \sum q_i \log q_i + (1 - \sum q_i) \log(1 - \sum q_i)$$

Consider the interior. The first-order condition gives $1 + \log q_i - (1 + \log(1 - \sum q_i)) = 0$, or equivalently, $q_i = 1 - \sum q_i$. Note that RHS is independent of i, and hence all q_i's are equal, say to q. Plugging this back in gives $q = 1 - (n-1)q$, and hence $q = 1/n$. Therefore, $p_i = 1/n$, and the objective function value is $-\log n$ there.

The boundary consists of $q_i = 0$ and $\sum q_i = 1$, or equivalently, $p_i = 0$ for some i. This is a complicated shape and admits no simple parametrization. A more subtle analysis on the boundary is needed. In fact, we have the following claim.

Claim. Fix a boundary point p. Let $I = \{i : p_i = 0\}$. Let $j \notin I$. Consider the path γ in the domain defined by

$$\gamma_i(t) = \begin{cases} t & \text{if } i \in I \\ p_j - |I|t & \text{if } i = j \\ p_i & \text{if otherwise} \end{cases}$$

The right-hand derivative of f at p along this path is $-\infty$

For instance, let $n = 4$ and $p = (0, 0.2, 0, 0.8)$. Then $I = \{1, 3\}$. Let $j = 2$. Then $\gamma(t) = (t, 0.2 - 2t, t, 0.8)$.

By this claim, no boundary point can be the minimum. Therefore, the global minimum is $-\log n$ and it is attained at $p_i = 1/n$.

It remains to prove this claim. The proof is a straight-forward computation.

Proof.

$$\begin{aligned} f(\gamma(t)) &= |I|t \log t + (p_j - |I|t) \log(p_j - |I|t) + C \\ &= |I|t \log \frac{t}{p_j - |I|t} + [p_j \log(p_j - |I|t) + C] \end{aligned}$$

where C is independent of t.

The second term has a finite derivative at $t = 0$, so it suffices to show the derivative of the first term is $-\infty$. We show this from first principles.

$$\lim_{t \to 0^+} \frac{1}{t}(|I|t \log \frac{t}{p_j - |I|t} - 0) = \lim |I| \log \frac{t}{p_j - |I|t} = -\infty$$

The result then follows. $\square$

Remark. Alternatively, the boundary case can also be handled by mathematical induction.

Remark. Later in the course, a more systematic treatment, namely the method of Lagrange multiplier, for optimization with equality constraint will be introduced.

The next example was not covered in class because of time constraint. You are welcome to read it at your pleasure.

Example 3. Let X and y be given $N \times k$ and $N \times 1$ matrices. Suppose the columns of X are linearly independent.

$$\min_{\beta} |X\beta - y|^2$$

Remark. This is the problem of linear regression in statistics. Suppose a number of data points (x_i, y_i)'s are given and they are to be fitted by a straight line $y = mx + c$, or equivalently. Then it ought to be that

$$1c + x_i m = y_i \tag{i}$$

Let x and y be the column vectors whose entry i are x_i and y_i. Let $\mathbf{1}$ be the vector of the same dimension whose entries are all 1's. Let $X = \begin{bmatrix} \mathbf{1} & x \end{bmatrix}$.

Then equation (i) can be treated as row i of the system of equations $X\beta = y$ with unknown $\beta = \begin{bmatrix} c \\ m \end{bmatrix}$.

Such an equation typically has no solution (real data rarely lies perfectly on a straight line), but the error (more properly, the residual) can be minimized, giving rise to this example.

Solution. The domain is not bounded, but this problem is finding the point in the subspace spanned by the columns of X that is closest to y. By this geometric insight, we know this function has a minimum. The theorem that the minimum exists whenever the function is continuous and $\lim_{|x|\to\infty} f(x) = +\infty$ also guarantees the existence of the minimum.

For the gradient, since $D|x|^2 = 2x^t$ (by computing the partial derivatives), chain rule gives $D|X\beta - y|^2 = 2(X\beta - y)^t X$.

However, if you are not comfortable with chain rule, you may proceed as follows.

Letting $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$,

$$\begin{aligned}
D_k|X\beta - y|^2 &= D_k(\sum \beta_i X_{li} X_{lj} \beta_j - 2\sum y_i X_{ij} \beta_j + |y|^2) \\
&= \sum \delta_{ik} X_{li} X_{lj} \beta_j + \sum \beta_i X_{li} X_{lj} \delta_{jk} - 2\sum y_i X_{ij} \delta_{jk} \\
&= \sum X_{lk} X_{lj} \beta_j + \sum \beta_i X_{li} X_{lk} - 2\sum y_i X_{ik} \\
&= [X^t X\beta]_k + [X^t X\beta]_k - 2[]X^t y]_k \\
&= [2X^t X\beta - 2X^t y]_k
\end{aligned}$$

Either case, the first-order condition gives

$$2X^t X\beta - 2X^t y = 0 \tag{1}$$

or equivalently, $X^t X\beta = X^t y$, where β can be found by solving the linear system; $X^t X$ is invertible because $X^t X\xi = 0$ implies $|X\xi|^2 = \xi^t X^t X\xi = 0$, and hence $X\xi = 0$, and hence by linear independence of columns of X, $\xi = 0$.

Remark. Equation (1) actually says $X\beta - y$ is perpendicular to every direction in the space spanned by columns of X.