Nonlinear optimization

Problem formulation and solution

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ be convex functions. Consider the following optimization problem.

$$\min_{x_i} F(x_1, \ldots, x_n) \tag{1}$$

subject to

$$\begin{array}{rcl} h_1(x_1, \ldots, x_n) & = & c_1, \\ h_m(x_1, \ldots, x_n) & = & c_m \\ \vdots & & \vdots \end{array}$$

If the functions are strictly convex, there exists a unique solution to this problem. Otherwise, there may be more than one solution; however they all produce the same value of F. This problem is solved by creating a Lagrangian function

$$L(x_1, \ldots, x_n) = F(x_1, \ldots, x_n) + \sum_{i=1}^{m} \lambda_i (h_i(x_1, \ldots, x_n) - c_i)$$

where $\{\lambda_i\}$ are a set of Lagrange multipliers, one for each constraint. Note that the optimization problem becomes,

$$\min_{x_i} L(x_1, \ldots, x_n) \tag{2}$$

where the optimal values of $\{x_i\}$ are now functions of the Lagrange multipliers. The Lagrange multipliers $\{\lambda_i\}$ are chosen so that the m constraints listed above are satisfied.

The optimum values $\{x_i^*\}$ and $\{\lambda_i\}$ satisfy the following sets of equations,

$$\begin{array}{rcl} \partial L / \partial x_i |_{x_i = x_i^*(\{\lambda_i\})} & = & 0, \quad i = 1, \ldots, n \\ h_i(x_1^*(\{\lambda_i\}), \ldots, x_n^*(\{\lambda_i\})) & = & c_i, \quad i = 1, \ldots, m \end{array}$$

Consider the problem with an additional k inequality constraints,

$$\min_{x_i} F(x_1, \ldots, x_n) \tag{3}$$

subject to

$$\begin{array}{rcl} h_1(x_1,\ldots,x_n) & = & c_1, \\ \vdots & & \vdots \\ h_m(x_1,\ldots,x_n) & = & c_m, \\ g_1(x_1,\ldots,x_n) & \leq & d_1, \\ \vdots & & \vdots \\ g_k(x_1,\ldots,x_n) & \leq & d_k \end{array}$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex. If we had an oracle that could tell us that the optimal solutions would satisfy $l \leq k$ of the constraints as equalities and the remaining constraints as strict inequalities, then the problem would reduce to a version of the original problem. Specifically, if the first l constraints are satisfied as equalities, the problem becomes

$$\min_{x_i} F(x_1,\ldots,x_n) \tag{4}$$

subject to

$$\begin{array}{rcl} h_1(x_1,\ldots,x_n) & = & c_1, \\ \vdots & & \vdots \\ h_m(x_1,\ldots,x_n) & = & c_m, \\ g_1(x_1,\ldots,x_n) & = & d_1, \\ \vdots & & \vdots \\ g_k(x_1,\ldots,x_n) & = & d_l \end{array}$$

The challenge posed by problems with inequality constraints is determining which constraints are active and which are not. Efficient means of making this determination are very much problem dependent.

A simple routing problem

Consider the following simple routing formulation where there is a single source and single destination. The source generates traffic at a rate of r packets per second, each consisting on average of $1/\mu$ bits. There are n links connecting the source to the destination labeled $i-1,\ldots,n$. Link i has a capacity of C_i bits per second. We are interested in allocating traffic from the source to the n links so as to minimize the average packet delay. Let r_i denote the traffic rate assigned to link i. The minimum average delay routing algorithm can be formulated as

$$\min_{r_i} \frac{1}{r} \sum_{i=1}^{n} F_i(r_i)$$

subject to

$$\begin{aligned} \sum_{i=1}^{n} r_i &= r, \\ r_i &\geq 0, i=1,\ldots n, \\ r_i &\leq C_i \mu, i=1,\ldots n \end{aligned}$$

where $F_i(r_i)$ is the average number of packets on link i when the offered traffic rate is r_i, $i = 1,\ldots,n$. We assume that $\{F_i(r_i)\}$ are increasing convex functions.

It is often the case that $\lim_{r_i \to C_i\mu} F_i(r_i) = \infty$ in which case, the constraint that $r_i \leq C_i\mu$ is not needed and the problem becomes

$$\min_{r_i} \frac{1}{r} \sum_{i=1}^{n} F_i(r_i)$$

subject to

$$\begin{aligned} \sum_{i=1}^{n} r_i &= r, \\ r_i &\geq 0, i=1,\ldots n \end{aligned}$$

Note the presence of the inequality constraints. For now, let us assume that they are not active and can be ignored; the problem becomes

$$\min_{r_i} \frac{1}{r} \sum_{i=1}^{n} F_i(r_i)$$

subject to

$$\sum_{i=1}^{n} r_i = r$$

When we form the Lagrangian, the problem becomes

$$\min_{r_i} L(r_1, \ldots, r_n) = \frac{1}{r} \sum_{i=1}^{n} F_i(r_i) - \lambda(\sum_{i=1}^{n} r_i - r).$$

and the optimum solution satisfies

$$\frac{\partial L}{\partial r_i} = r_i \, F_i'(r_i) - \lambda = 0, \quad i = 1, \ldots, n$$

coupled with

$$\sum_{i=1}^{n} r_i = r$$

Note that the optimality condition requires that $F_i'(r_i)$ to be the same for all links $i = 1, \ldots, n$, i.e., $F_i'(r_i) = \lambda$. Let $\{r_i^*\}$ be the solution to the above equations.

If we assume that packets arrive according to a Poisson process and packet lengths are exponentially distributed with mean μ, we have $F_i(r_i) = r_i/(\mu C - r_i)$ and the optimality condition is

$$F_i'(r_i) = \frac{C_i \mu}{(C_i \mu - r_i)^2} = \lambda, \quad i = 1, \ldots, n$$

which has a solution

$$r_i^* = C_i \mu - \sqrt{\frac{C_i \mu}{\lambda}}, \quad i = 1, \ldots, n \tag{5}$$

The multiplier λ is obtained by substituting the above expression for r_i^* in the constraint $\sum r_i^* = r$,

$$\mu \sum_{i=1}^{n} C_i - \frac{\sum_{i=1}^{n} \sqrt{C_i \mu}}{\sqrt{\lambda}} = r$$

with solution

$$\frac{1}{\sqrt{\lambda}} = \frac{\mu \sum_{i=1}^{n} C_i - r}{\sum_{i=1}^{n} \sqrt{C_i \mu}}.$$

Substitution into (5) produces

$$r_i^* = C_i\mu - (\sum_{i=1}^{n} C_i\mu - r)\frac{\sqrt{C_i\mu}}{\sum_{i=1}^{n}\sqrt{C_i\mu}}$$

Consider the case where some of the constraints may be active, i.e., $r_i^* = 0$ for some i. In this case $\{r_i^*\}$ is the optimum solution if and only if

$$\begin{array}{rcl} F_i'(r_i^*) & = & \lambda, \quad r_i^* > 0, \\ F_i'(r_i^*) & \geq & \lambda, \quad r_i^* = 0 \end{array} \tag{6}$$

Let L_c be the set of links such that $r_i^* = 0$ and L_u the set of links for which $r_i^* > 0$. Order the links such that $F_1'(0) \leq F_2'(0) \leq \cdots \leq F_n'(0)$. The $\{r_i^*\}$ exhibits the following property.

Property 1 *If $i < j$ and $i \in L_c$, then $j \in L_c$.*

This is easily seen through contradiction. Suppose that $r_j^* > 0$. Then $F_j'(r_j^*) > F_j'(0)$ because F_i is conves.. But $F_j'(0) \geq F_i'(0)$; hence $F_j'(r_j^*) > F_i'(0)$ which according to (6) violates the condition for $r_j^* > 0$ and $r_i^* - 0$.

Consequently, $L_u = \{1, 2, \ldots, j^*\}$ and $L_u = \{j^* + 1, \ldots, n\}$ for some $1 \leq j^* \leq n$. j^* can be found using a binary search algorithm based on the following test. Suppose you guess $j' < j^*$. Then the link rates will violate (6) in that $F_i'(r_i^*) > F_{j+1}'(0)$. SImilarly, if you guess $j' > j^*$, the link rate r_j^* will be computed to be less than zero, $r_j^* < 0$.