Lecture 20

Solving Dual Problems

We consider a constrained problem where, in addition to the constraint set X, there are also inequality and linear equality constraints. Specifically the minimization problem of interest has the following form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \leq 0, \ldots, g_m(x) \leq 0 \\ & a_1^T x = b_1, \ldots, a_r^T x = b_r \\ & x \in X, \end{array}$$

where $X \subset \mathbb{R}^n$, $g_j : X \to \mathbb{R}$ for all j, and $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for all i. We will also use a more compact formulation of the preceding problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \\ & Ax = b \\ & x \in X, \end{array} \tag{1}$$

where $g = [g_1, \ldots, g_m]^T$, A is an $r \times n$ matrix with rows a_i^T, and $b \in \mathbb{R}^r$ is a vector with components b_i. Throughout this section, we use the following assumption.

Assumption 1 *The set X is convex and closed. The objective f and the constraint functions $g_1, \ldots, g_m$ are convex over the set X. The optimal value f^* of problem (1) is finite.*

Solving the problem of the form (1) can be very complex due to the presence of (possibly nonlinear) inequality constraints $g(x) \leq 0$. Here, we consider the algorithms for solving the problem (1) through the use dual methods. Also, we consider solving the dual on its own right.

Consider the dual problem obtained by relaxing all the inequality and equality constraints (assigning prices to them). In this case, the dual problem is

$$\begin{array}{ll} \text{maximize} & q(\mu, \lambda) \\ \text{subject to} & \mu \geq 0, \end{array} \tag{2}$$

where the dual function $q(\mu, \lambda)$ is given by

$$q(\mu, \lambda) = \inf_{x \in X} \{f(x) + \mu^T g(x) + \lambda^T (Ax - b)\} \qquad \text{for } \mu \in \mathbb{R}^m \text{ with } \mu \geq 0 \text{ and } \lambda \in \mathbb{R}^r. \tag{3}$$

Note that the constraint $(\mu, \lambda) \in \operatorname{dom} q$ is an implicit constraint of the dual problem (silently assumed). In the dual problem, the multiplier μ is constrained to the nonnegative orthant, while the multiplier λ is a free variable. Furthermore, the dual function $q(\mu, \lambda)$ is concave, so that the dual is a constrained concave maximization problem, which is equivalent to a constrained convex minimization problem (through a sign change in the objective). Hence, if the dual function is differentiable we could apply gradient projection methods (for maximization) and solve the dual.

In some situations, a partial dual problem is considered and is still referred to as the dual. In particular, consider relaxing only the inequality constraints of the problem (1), yielding a dual problem of the form

$$\begin{array}{ll} \text{maximize} & \tilde{q}(\mu) \\ \text{subject to} & \mu \geq 0, \end{array} \tag{4}$$

where the dual function $\tilde{q}(\mu)$ is given by

$$\tilde{q}(\mu) = \inf_{Ax=b,\ x\in X} \{f(x) + \mu^T g(x)\} \qquad \text{for } \mu \in \mathbb{R}^m \text{ with } \mu \geq 0. \tag{5}$$

In this dual problem, the multiplier μ is constrained to the nonnegative orthant, while the dual function $\tilde{q}(\mu)$ is concave. To distinguish between these two different formulations of the dual problem, we will refer to problem (2)–(3) as *the dual problem* and to problem (4)–(5) as *the partial dual problem*.

The main difficulty in dealing with dual problems is the evaluation of the dual function, since it involves solving a constrained minimization problem per each value of the dual variables. The use of dual problems is the most advantageous in the situations when the dual function evaluation is "easy", i.e., when a dual solution is explicitly given. Fortunately, this is the case in many problems arising in various applications. We discuss some of them later.

In what follows, we focus on the dual problems where the minimization problem involved in the dual function evaluation has solutions. Under this assumption, for the dual function $q(\mu, \lambda)$ of Eq. (3), we have for any $\mu \geq 0$ and any λ,

$$q(\mu, \lambda) = f(x_{\mu\lambda}) + \mu^T g(x_{\mu\lambda}) + \lambda^T(Ax_{\mu\lambda} - b),$$

where $x_{\mu\lambda}$ is an optimal solution for the following problem

$$\begin{array}{ll} \text{minimize} & f(x) + \mu^T g(x) + \lambda^T(Ax - b) \\ \text{subject to} & x \in X. \end{array} \tag{6}$$

Similarly, under the assumption that the minimization problem defining the dual function $\tilde{q}(\mu)$ of Eq. (5) has optimal solutions, we have for any $\mu \geq 0$,

$$\tilde{q}(\mu) = f(x_\mu) + \mu^T g(x_\mu),$$

where x_μ is an optimal solution for the following problem

$$\begin{array}{ll} \text{minimize} & f(x) + \mu^T g(x) \\ \text{subject to} & Ax = b,\ x \in X. \end{array} \tag{7}$$

Let us now consider the relations that characterize the minimizers $x_{\mu\lambda}$ and x_μ of the problems (6) and (7) respectively, when f and all g_j are convex, and X is closed and convex.

Suppose that f and all g_j are differentiable. The gradient of the Lagrangian function of Eq. (6) is

$$\nabla f(x) + \sum_{j=1}^{m} \mu_j \nabla g_j(x) + \lambda^T A.$$

Thus, by the first-order optimality conditions, $x_{\mu\lambda}$ is an optimal solution for problem (6) if and only if

$$\left(\nabla f(x_{\mu\lambda}) + \sum_{j=1}^{m} \mu_j \nabla g_j(x_{\mu\lambda}) + \lambda^T A\right)^T (x - x_{\mu\lambda}) \geq 0 \qquad \text{for all } x \in X.$$

Similarly, the gradient of the Lagrangian function of Eq. (7) is

$$\nabla f(x) + \sum_{j=1}^{m} \mu_j \nabla g_j(x).$$

By the first-order optimality conditions, x_μ is an optimal solution for problem (7) if and only if

$$\left(\nabla f(x_{\mu\lambda}) + \sum_{j=1}^{m} \mu_j \nabla g_j(x_{\mu\lambda})\right)^T (x - x_{\mu\lambda}) \geq 0 \qquad \text{for all } x \in X,\ Ax = b.$$

To this end, we discussed both the dual and the partial dual problem, since both have been traditionally used, depending on which one is more suitable for a given problem at hand. For the rest of this section, we will focus only on the dual problem (2). Analogous results hold for the partial dual.

With respect to the existence of optimal solutions for the problem (6), we consider two cases:

(1) The minimizer $x_{\mu\lambda}$ is unique for each $\mu \geq 0$ and $\lambda \in \mathbb{R}^r$.

(2) The minimizer $x_{\mu\lambda}$ is not unique.

The uniqueness of the minimizers ties closely with the differentiability of the dual function $q(\mu, \lambda)$, which we discuss next.

In some situations f of some of g_j's are not differentiable, but still the minimizers $x_{\mu\lambda}$ may be easily computed.

Example 1 *(Assignment Problem)*
The problem is to assign m jobs to n processors. If a job i is assigned to processor j, the cost is a_{ij} and it takes p_{ij} time to be completed. Each processor j has limited time t_j available for processing. The goal is to determine the minimum cost assignment of jobs to processors. Formally, the problem is

$$\textit{minimize} \qquad \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij}$$

$$
\begin{aligned}
\textit{subject to} \quad & \sum_{j=1}^{n} x_{ij} = 1 \quad \textit{for all } i, \\
& \sum_{i=1}^{n} p_{ij} x_{ij} \le t_j \quad \textit{for all } j, \\
& x_{ij} \in \{0,1\} \quad \textit{for all } i, j.
\end{aligned}
$$

By relaxing the time constraints for the processors, we obtain the dual problem

$$
\begin{aligned}
\textit{maximize} \quad & q(\mu) = \sum_{i=1}^{m} q_i(\mu) \\
\textit{subject to} \quad & \mu \ge 0,
\end{aligned}
$$

where for each i,

$$
q_i(\mu) = \min_{\substack{\sum_{j=1}^{n} x_{ij} = 1 \\ x_{ij} \in \{0,1\}}} \sum_{j=1}^{n} (a_{ij} + \mu_j p_{ij}) x_{ij} - \frac{1}{m} \sum_{j=1}^{n} t_j \mu_j .
$$

We can easily evaluate $q_i(\mu)$ for each $\mu \ge 0$. In particular, it can be seen that

$$
q_i(\mu) = a_{ij^*} + \mu_{j^*} p_{ij^*} - \frac{1}{m} \sum_{j=1}^{n} t_j \mu_j ,
$$

where the index j^* is such that

$$
a_{ij^*} + \mu_{j^*} p_{ij^*} = \min_{1 \le j \le n} \{a_{ij} + \mu_j p_{ij}\}.
$$

Here, however, in the evaluation of q_i we do not rely on differentiability.

1 Differentiable Dual Function

In this section, we discuss the dual methods for the case when the dual function is differentiable. As we will see the differentiability of the dual function is guaranteed when the minimizer of the problem (6) is unique. We formally impose this condition, as follows.

Assumption 2 *For every $\mu \ge 0$ and $\lambda \in \mathbb{R}^r$, the minimizer $x_{\mu\lambda}$ in the problem (6) exists and it is unique.*

Under convexity assumption and Assumption 2, we have the following result.

Lemma 1 *Let Assumptions 1 and 2 hold. Then, for every $\mu \geq 0$ and $\lambda \in \mathbb{R}^r$, the inequality and equality constraints of the problem (1) evaluated at the minimizer $x_{\mu\lambda}$ constitute the gradient of q at (μ, λ), i.e.,*

$$\nabla q(\mu, \lambda) = \begin{bmatrix} g_1(x_{\mu\lambda}) \\ \vdots \\ g_m(x_{\mu\lambda}) \\ a_1^T x_{\mu\lambda} - b_1 \\ \vdots \\ a_r^T x_{\mu\lambda} - b_r \end{bmatrix}.$$

Note that, in view of Lemma 1, *the differentiability of the dual function has nothing to do with the differentiability of the objective f or constraint functions g_j.* The differentiability of q is strongly related to the uniqueness of the optimizer in the problem defining the dual function value.

According to Lemma 1, for the partial gradients $\nabla_\mu q(\mu, \lambda)$ and $\nabla_\lambda q(\mu, \lambda)$, we have

$$\nabla_\mu q(\mu, \lambda) = \begin{bmatrix} g_1(x_{\mu\lambda}) \\ \vdots \\ g_m(x_{\mu\lambda}) \end{bmatrix} = g(x_{\mu\lambda}),$$

$$\nabla_\lambda q(\mu, \lambda) = \begin{bmatrix} a_1^T x_{\mu\lambda} - b_1 \\ \vdots \\ a_r^T x_{\mu\lambda} - b_r \end{bmatrix} = A x_{\mu\lambda} - b.$$

To solve the dual problem, under Assumptions 1 and 2, we can now apply the projection gradient method, which is adapted to handle maximization. The projected gradient method for the dual problem has the form:

$$\begin{aligned} \mu_{k+1} &= \left[\mu_k + \alpha_k \nabla_\mu q(\mu, \lambda)\right]^+, \\ \lambda_{k+1} &= \lambda_k + \alpha_k \nabla_\lambda q(\mu, \lambda), \end{aligned}$$

where $[\cdot]^+$ denotes the projection on the nonnegative orthant $\mathbb{R}^m_+$, $\alpha_k > 0$ is the stepsize, and $\mu_0 \geq 0$ and λ_0 are initial multipliers. Note that, since the dual problem involves maximization, the gradient method takes steps along the gradients of q. We refer to this gradient method as *the dual gradient projection method.*

By using Lemma 1, we see that the method is equivalently given by

$$\mu_{k+1} = \left[\mu_k + \alpha_k\, g(x_{\mu\lambda})\right]^+, \tag{8}$$

$$\lambda_{k+1} = \lambda_k + \alpha_k\, (A x_{\mu\lambda} - b). \tag{9}$$

The dual gradient method can be used with all the stepsizes that we have discussed for the gradient methods. including the backtracking line search. However, the Polyak's stepsize and its modification, and the backtracking line search have to be suitably adjusted to account for the maximization aspect. In particular, for the dual gradient projection method the Polyak's stepsize is given by:

$$\alpha_k = \frac{q^* - q(\mu_k, \lambda_k)}{\|\nabla q(\mu_k, \lambda_k)\|^2},$$

where q^* is the optimal value of the dual problem (of course, q^* should be finite in order to use this stepsize).

Denote by x_k the minimizer $x_{\mu\lambda}$ when $(\mu, \lambda) = (\mu_k, \lambda_k)$. Note that $\|\nabla q(\mu_k, \lambda_k)\|^2 = \|g(x_k)\|^2 + \|Ax_k - b\|^2$, or equivalently

$$\|\nabla q(\mu_k, \lambda_k)\|^2 = \sum_{j=1}^{m} g_j^2(x_k) + \sum_{i=1}^{r} (a_i^T x_k - b_i)^2.$$

For the dual gradient projection method, the modified Polyak's stepsize has the following form

$$\alpha_k = \frac{\hat{q}_k - q(\mu_k, \lambda_k)}{\|\nabla q(\mu_k, \lambda_k)\|^2} \qquad \text{with} \qquad \hat{q}_k = \delta + \max_{0 \le \kappa \le k} q(\mu_\kappa, \lambda_\kappa).$$

Note that, when appropriately interpreted, all the results for the gradient projection method that we know apply to the dual maximization problem.

We next provide an example demonstrating computation of the gradient of the dual function. In particular, we revisit the Kelly's canonical utility-based network resource allocation problem (see Kelly98).

Example 2 *Consider a network consisting of a set $\mathcal{S} = \{1, \ldots, S\}$ of sources and a set $\mathcal{L} = \{1, \ldots, L\}$ of undirected links, where a link l has capacity c_l. Let $\mathcal{L}(i) \subset \mathcal{L}$ denote the set of links used by source i. The application requirements of source i is represented by a differentiable concave increasing utility function $u_i : [0, \infty) \to [0, \infty)$, i.e., each source i gains a utility $u_i(x_i)$ when it sends data at a rate x_i. Let $\mathcal{S}(l) = \{i \in \mathcal{S} \mid l \in \mathcal{L}(i)\}$ denote the set of sources that use link l. The goal of the network utility maximization problem is to allocate the source rates as the optimal solution of the problem*

$$\begin{array}{ll} \textit{maximize} & \sum_{i \in \mathcal{S}} u_i(x_i) \\ \textit{subject to} & \sum_{i \in \mathcal{S}(l)} x_i \le c_l \quad \textit{for all } l \in \mathcal{L} \\ & x_i \ge 0 \quad \textit{for all } i \in \mathcal{S}. \end{array}$$

Alternatively, we may view the problem as the minimization of differentiable convex and decreasing function $f(x) = -\sum_{i \in \mathcal{S}} u_i(x_i)$ subject to the above constraints. Note that the constraint set of the problem is compact [since $0 \le x_l \le c_l$ for all links $l \in \mathcal{L}$]. Since f is continuous over the constraint set, the optimal value f^* is finite [in fact, a unique optimal solution x^* exists]. Assumption 1 is satisfied.

By relaxing the link capacity constraints, the dual function takes the form

$$q(\mu) \;=\; \min_{x_i \ge 0,\ i \in \mathcal{S}} \sum_{i \in \mathcal{S}} = u_i(x_i) + \sum_{l \in \mathcal{L}} \mu_l \left(\sum_{i \in \mathcal{S}(l)} x_i - c_l \right)$$

$$= \min_{x_i \geq 0,\ i \in \mathcal{S}} \sum_{i \in \mathcal{S}} \left(-u_i(x_i) + x_i \sum_{l \in \mathcal{L}(i)} \mu_l \right) - \sum_{l \in \mathcal{L}} \mu_l c_l.$$

Since the optimization problem on the right-hand side of the preceding relation is separable in the variables x_i*, the problem decomposes into subproblems for each source* i*. Letting* $\mu_i = \sum_{l \in \mathcal{L}(i)} \mu_l$ *for each* i *(i.e.,* μ_i *is the sum of the multipliers corresponding to the links used by source* i*), we can write the dual function as*

$$q(\mu) = \sum_{i \in \mathcal{S}} \min_{x_i \geq 0} \{x_i \mu_i - u_i(x_i)\} - \sum_{l \in \mathcal{L}} \mu_l c_l.$$

Hence, to evaluate the dual function, each source i *needs to solve the one-dimensional optimization problem* $\min_{x_i \geq 0} \{x_i \mu_i - u_i(x_i)\}$*. Note that* $\mu_i = 0$ *is not in the domain of the dual function [since each* u_i *is increasing, it follows that* $\min_{x_i \geq 0} \{-u_i(x_i)\} = -\infty$*]. Thus, we must have* $\mu_i > 0$ *for all* $i \in \mathcal{S}$ *for the dual function to be well defined.*

For $\mu_i > 0$*, by the first-order optimality conditions, the optimal solution* $x_i(\mu_i)$ *for the one-dimensional problem satisfies the following relation*

$$u_i'(x_i(\mu_i)) = \mu_i,$$

where $u_i'(x_i)$ *denotes the derivative of* $u_i(x_i)$*. Thus,*

$$x_i(\mu_i) = u_i'^{-1}(\mu_i),$$

where $u_i'^{-1}$ *is the inverse function of* u_i'*, which exists since* u_i *is differentiable and increasing. Hence, for each dual variable* $\mu > 0$*, the minimizer* $x(\mu) = \{\mu_i,\ i \in \mathcal{S}\}$ *in the problem of dual function evaluation exists and it is unique.*

When there is no duality gap, solving the dual problem does not yield a primal optimal solution immediately. Suppose we solve the dual and we found a dual optimal solution (μ^*, λ^*). Also, suppose we have the minimizer x^* such that

$$q(\mu^*) = \inf_{x \in X} \{f(x) + (\mu^*)^T g(x) + (\lambda^*)^T (Ax - b)\} = \{f(x^*) + (\mu^*)^T g(x^*) + (\lambda^*)^T (Ax^* - b)\}.$$

Such an x^* need not be optimal for the primal problem. For x^* to be optimal, the KKT conditions have to be satisfied by x^* and (μ^*, λ^*).

The following implication of the KKT conditions is often useful when a dual optimal solution is available.

Lemma 2 *Let Assumption 1 hold, and let strong duality hold [*$f^* = q^*$*]. Let* (μ^*, λ^*) *be an optimal multiplier pair. Then,* x^* *is a solution of the primal problem 1 if and only if*

- x^* *is primal feasible, i.e.,* $g(x^*) \leq 0,\ Ax^* = 0,\ x^* \in X$*.*
- x^* *is a minimizer of the problem (6) with* $(\mu, \lambda) = (\mu^*, \lambda^*)$*.*
- x^* *and* μ^* *satisfy the complementarity slackness, i.e.,* $(\mu^*)^T g(x^*) = 0$*.*

2 Application of Dual Methods in Communication Networks

2.1 Joint Routing and Congestion Control

Here, we are interested in the network problem when the user rates are not fixed. The congestion control adjusts the traffic rates so that the network resources are fairly shared among the users, and the network is at a reasonable operating point balancing the throughput with the delay. The model for joint routing and congestion control presented in this section is based on utility maximization framework.

The utility functions have been used in economic market models to quantify preferences of users (consumers) for certain resources (commodities). A utility function $u(x)$ "measures" the value of a resource amount x to a user. Typically, it is assumed that a utility function is concave, nondecreasing, and continuously differentiable (scalar) function, defined on the interval $[0, +\infty)$. Some of the common examples include

- *Log-function*: $u(x) = w \ln x$ for some scalar $w > 0$.
- *Power-function*: $u(x) = w\, \frac{x^{1-\alpha}}{1-\alpha}$ for some scalars $w > 0$ and $\alpha > 0$.

The derivative of u is referred to as the marginal utility per unit resource, since $u(x+\delta) = u(x) + \delta u'(x) + o(\delta)$ for a small $\delta > 0$. Due to the assumed concavity of u, the marginal utility $u'(x)$ is nonincreasing.

Congestion control is a mechanism for adjusting the user rates $x = [x_s,\ s \in \mathcal{S}]$ fairly with respect to user utility functions $u_s,\ s \in \mathcal{S}]$. The network performance is quantified in terms of the user utilities and the cost of routing the traffic. The joint routing and congestion control problem is to determine both the user rates $x = [x_s,\ s \in \mathcal{S}]$ and the paths $v = [v_p,\ p \in \mathcal{P}]$ so as to maximize the network performance. Formally, the problem is given by

$$\begin{array}{ll} \text{maximize} & U(x) = \displaystyle\sum_{s\in\mathcal{S}} u_s(x_s) - \sum_{l\in\mathcal{L}} f_l\left(\sum_{\{p\,|\,l\in p\}} v_p\right) \\ \text{subject to} & Dv = x \\ & x \geq 0,\ v \geq 0. \end{array} \tag{10}$$

This is a concave maximization problem with a convex (in fact, polyhedral) constraint set. For this problem, the first order optimality condition the feasible vectors x and v are solution to problem (10) if and only if x and v are feasible [i.e., $Dv = x$, $x \geq 0,\ v \geq 0$] and for each $s \in \mathcal{S}$,

$$u_s'(x_s) \leq \sum_{l\in p} f_l'\left(\sum_{\{p\,|\,l\in p\}} v_p\right) \qquad \text{for all } p \in s, \tag{11}$$

with equality holding when $v_p > 0$.

We interpret the length $f_l'\left(\sum_{\{p\,|\,l\in p\}} v_p\right)$ of link l as the cost of using the link l. In view of this, the preceding relation means that, at optimal x and v, for the paths p carrying the flow of user s, the cost of any path p with $v_p > 0$ is equal to the user's marginal utility, while the cost of any path p with $v_p = 0$ is no less than the user marginal utility.

In some applications, there are explicit constraints on the link capacities, and the problem of joint routing and congestion control is given by

$$\begin{aligned} \text{maximize} \quad & U(x) = \sum_{s \in \mathcal{S}} u_s(x_s) \\ \text{subject to} \quad & Dv = x, \;\; Av \le c \\ & x \ge 0, \;\; v \ge 0, \end{aligned} \tag{12}$$

where $c = [c_l, \; l \in \mathcal{L}]$ is the vector with entries c_l representing the capacity of link l.

Consider the dual of problem (12) obtained by assigning the prices to the link constraints $Av \le c$. The dual function is given by

$$q(\mu) = \max_{Dv=x, \; x \ge 0, \; v \ge 0} \left\{ \sum_{s \in \mathcal{S}} u_s(x_s) - \mu^T Ax \right\} + \mu^T c \qquad \text{for } \mu \ge 0.$$

By the KKT conditions we have that x^* and v^* are primal optimal (and μ^* dual optimal) if and only if: x^* and v^* are feasible, $\mu^* \ge 0$, and such that they satisfy the complementarity slackness and (x^*, v^*) attains the maximum in $q(\mu^*)$. Formally, the complementarity slackness is given by

$$\mu_l^* = 0 \quad \text{only if} \quad \textstyle\sum_{\{p \,|\, l \in p\}} v_p^* < c_l. \tag{13}$$

Furthermore, (x^*, v^*) attains the maximum in $q(\mu^*)$ if and only if for each s,

$$u_s'(x_s^*) \le \sum_{l \in p} \mu_l^* \qquad \text{for all } p \in s. \tag{14}$$

with equality only when $v_p^* > 0$. This relation is similar to the relation in Eq. (11), where the cost f_l' "plays the role of" the multiplier μ_l^*.

Note that when $v_p^* > 0$, the values $\sum_{l \in p} \mu_l^*$ for $p \in s$ are the same. Denote this value by μ_s^*. By interpreting μ_s^* as the price per unit rate for user s, from (14), we have

$$u_s'(x_s^*) = \mu_s^* \;\; \text{when } \mu_s^* > 0 \qquad \text{and} \qquad u_s'(x_s^*) \le \mu_s^* \;\; \text{when } \mu_s^* = 0.$$

The preceding is the optimality condition for the following problem for user s,

$$\begin{aligned} \text{maximize} \quad & u(x_s) - \mu_s^* x_s \\ \text{subject to} \quad & x_s \ge 0. \end{aligned}$$

By introducing a new variable $w_s = \mu_s^* x_s$, we can rewrite the preceding problem as follows:

$$\begin{aligned} \text{maximize} \quad & u\left(\frac{w_s}{\mu_s^*}\right) - w_s \\ \text{subject to} \quad & w_s \ge 0. \end{aligned} \tag{15}$$

The relation $w_s = \mu_s^* x_s$ implies that $w_s / x_s = \mu_s^*$ can be interpreted as optimality condition at x^*, as follows

$$\frac{w_s^*}{x_s^*} = \mu_s^* \;\; \text{when } \mu_s^* > 0,$$

which corresponds to maximizing $w_s^* \ln x_s - \mu_x^* x_s$ over $x_s \geq 0$. This together with the feasibility of x^* and v^*, and the complementarity slackness of Eq. (13), imply by KKT conditions that w^*, x^* and v^* constitute an optimal solution to the problem

$$\begin{aligned} \text{maximize} \quad & \sum_{s \in \mathcal{S}} w_s \ln x_s \\ \text{subject to} \quad & Dv = x, \; Av \leq c \\ & x \geq 0, \; v \geq 0, \end{aligned} \tag{16}$$

Thus, by introducing a new variable $w_s = \mu_s^* x_s$ and through the use of KKT conditions, we have found that the original joint routing and congestion control problem (12) is equivalent to the set of users problems of Eq. (15) and a network problem of Eq. (16).

A similar transformation can be considered for joint routing and congestion control problem without the link capacity constraints [cf. problem (12)]. This is discussed in detail in lecture notes by Hajek for Communication Network course.

The approach discussed above is precisely the approach proposed by Kelly *et. al.* The key idea in the approach is to introduce a price w_s that user s is willing to pay for his rate x_s. The price per unit rate is w_s/x_s for user s. The network receives x_s and w_s from each user $s \in \mathcal{S}$, and interprets the ratio w_s/x_s as "marginal utility" per unit of flow for user s. Thus, the network generates "surrogate utility functions"

$$\tilde{u}_s(x_s) = w_s \ln x_s \qquad \text{for each } s \in \mathcal{S}.$$

With these utility functions, the network problem of Eq.˜(12) becomes

$$\begin{aligned} \text{maximize} \quad & \sum_{s \in \mathcal{S}} w_s \ln x_s \\ \text{subject to} \quad & Dv = x, \; Av \leq c \\ & x \geq 0, \; v \geq 0. \end{aligned} \tag{17}$$

The network chooses the link prices μ_l as as optimal multipliers (link prices) for this problem. For each user s, the resulting price μ_s per unit of flow is given to user s, where $\mu_s = \sum_{l \in p} \mu_p$ for any $p \in s$ (these are the same for any $p \in s$ with $x_p > 0$). The user problem is to maximize $u_s(x_s)$ minus the pay $w_r = s$, subject to $w_s \geq 0$. Since $x_s = w_s/\mu_s$, a user problem is

$$\begin{aligned} \text{maximize} \quad & u\left(\frac{w_s}{\mu_s}\right) - w_s \\ \text{subject to} \quad & w_s \geq 0. \end{aligned} \tag{18}$$

The importance of the preceding formulations is the decomposition of the problem. The original problem of joint routing and congestion control formulated in Eq. (12) is a large optimization problem involving both user and network information. Through the use of "willingness to pay" variable and the KKT conditions, the problem is suitably decomposed into: a network problem (17) that does not require the information about user utility functions, and a set of user problems (18) that does not require any knowledge about the network (topology). Evidently, the users and the network have to exchange some information.

2.2 Rate Allocation in Communication Network

The rate allocation problem is a special case of problem (12), where the routing is fixed (i.e., v is now given, and $Dv = x$ is satisfied) and the problem is to allocate rates x_s for users $s \in \mathcal{S}$ optimally. The resulting rate allocation problem is

$$\begin{array}{ll} \text{maximize} & \sum_{s \in \mathcal{S}} u_s(x_s) \\ \text{subject to} & \sum_{s \in \mathcal{S}(l)} x_s \leq c_l \qquad \text{for all } l \in \mathcal{L} \\ & x \geq 0, \end{array}$$

where for each s, the set $\mathcal{S}(l)$ is the set of all users s whose traffic uses link l. We consider this problem with additional constraints, namely, the rate of user s is constrained within an interval $x_s \in [m_s, M_s]$, where $m_s \geq 0$ is the minimum and $M_s < \infty$ is the maximum rate for user s.

With these additional rate constraints, the rate allocation problem is given by

$$\begin{array}{ll} \text{maximize} & \sum_{s \in \mathcal{S}} u_s(x_s) \\ \text{subject to} & \sum_{s \in \mathcal{S}(l)} x_s \leq c_l \qquad \text{for all } l \in \mathcal{L} \\ & x_s \in I_s, \quad I_s = [m_s, M_s] \qquad \text{for all } s \in \mathcal{S}. \end{array} \tag{19}$$

In what follows, we discussed a dual algorithm given by Low and Lapsley 1999 for solving problem (19). It is assumed that *each utility u_s is strictly concave and increasing.* Under this assumption, the problem has an optimal solution [the constraint set is compact], and the optimal solution is unique by strict concavity of the utility functions.

The objective function of problem (19) is separable in the variables x_s, and these variables are coupled only through the link capacity constraints. Thus, by assigning prices to the link capacities, we obtain a dual problem of the form

$$\begin{array}{ll} \text{minimize} & q(\mu) \\ \text{subject to} & \mu \geq 0, \end{array} \tag{20}$$

where the dual function is

$$\begin{aligned} q(\mu) &= \max_{x_s \in I_s} \sum_{s \in \mathcal{S}} u_s(x_s) - \sum_{l \in \mathcal{L}} \mu_l \left(\sum_{s \in \mathcal{S}(l)} x_s - c_l \right) \\ &= \max_{x_s \in I_s} \sum_{s \in \mathcal{S}} \left(u_s(x_s) - x_s \sum_{l \in \mathcal{L}(s)} \mu_l \right) + \sum_{l \in \mathcal{L}} \mu_l c_l, \end{aligned}$$

where $\mathcal{L}(s)$ is the set of all links l carrying the flow of user s. Defining the variables $p_s = \sum_{l \in \mathcal{L}(s)} \mu_l$ and the functions

$$Q_s(p_s) = \max_{x_s \in I_s} \{u_s(x_s) - x_s p_s\}, \tag{21}$$

the dual function can be expressed as

$$q(\mu) = \sum_{s \in \mathcal{S}} Q_s(p_s) + \sum_{l \in \mathcal{L}} \mu_l c_l. \tag{22}$$

Given the link prices μ_l, $l \in \mathcal{L}$ and the resulting prices $p_s = \sum_{l \in \mathcal{L}(s)} \mu_l$ as seeing by the users $s \in \mathcal{S}$, for each s, the rate attaining the dual function value $Q_s(p_s)$ is denoted by $x_s(p_s)$. Note that the maximizer $x_s(p_s)$ in the problem of (21) is unique and given by

$$x_s(p_s) = P_{I_s}[u_s'(p_s)^{-1}], \tag{23}$$

where $u_s'^{-1}$ is the inverse function of the derivative u_s', and $P_{I_s}[z]$ denotes the projection on the (closed convex) interval I_s, which is in particular given by $P_{I_s}[z] = \min\{\max\{m_s, z\},\ M_s\}$.

We now consider a dual algorithm for rate allocation problem (19). In what follows, we assume that the problem is feasible and that each utility function u_s is strictly concave, twice differentiable and increasing on the interval I_s. Furthermore, for each s, the curvature of u_s is bounded away from zero on the interval I_s, i.e.,

$$-u_s''(z) \geq \frac{1}{a_s} > 0 \qquad \text{for all } z \in I_s \text{ and some } a_s > 0.$$

Under this condition, the rate problem has a unique optimal solution. Furthermore, note that the strong duality holds for the primal problem (19) and the dual problem (20)–(22). (recall the strong duality result for convex objective and linear constraints).

As noted earlier, under the preceding assumptions, the maximizer $x_s(p_s)$ in problem (21) exists and it is unique, implying that the dual function $q(\mu)$ is differentiable with the partial derivatives given by

$$\frac{\partial q(\mu)}{\partial \mu_l} = c_l - \sum_{s \in \mathcal{S}(l)} x_s(p_s).$$

Let $\mu_l(k)$ be the link prices at a given time k, and $\mu(k)$ be the vector of these prices. Let $x_s(k)$ be the maximizer given by Eq. (23) for $p_s = \sum_{l \in \mathcal{L}(s)} \mu_l(k)$. Consider the following gradient projection method for minimizing $q(\mu)$:

$$\mu_l(k+1) = \left[\mu_l(k) - \alpha_k \frac{\partial q(\mu(k))}{\partial \mu_l}\right]^+ \qquad \text{for all } l \in \mathcal{L},$$

where $\alpha_k > 0$ is a stepsize. Equivalently, the method is given by

$$\mu_l(k+1) = \left[\mu_l(k) + \alpha_k \left(\sum_{s \in \mathcal{S}(l)} x_s(k) - c_l\right)\right]^+ \qquad \text{for all } l \in \mathcal{L}. \tag{24}$$

Note that, given the aggregate rate $\sum_{s \in \mathcal{S}(l)} x_s(k)$ of the traffic through link l, the iterations of algorithm (24) are completely distributed over the links, and can be implemented by individual links using local information only.

By interpreting the set $\mathcal{L}$ of links and the set $\mathcal{S}$ of users as processors in a distributed system, the dual problem can be solved. In particular, given the link prices $\mu_l(k)$, the aggregate link price $\sum_{l \in \mathcal{L}(s)} \mu_l(k)$ is communicated to user s. Each user s evaluates its corresponding dual function $Q_s(p_s)$ of Eq. (21) [i.e., user s computes the maximizer $x_s(p_s)$]. Each user s communicates its rate $x_s(p_s)$ to links $l \in \mathcal{L}(s)$ [the links carrying the flow of user s]. Every link l updates its price $\mu_l(k)$ according to the gradient projection algorithm [cf. Eq. (24)]. The updated aggregate link prices $\sum_{l \in \mathcal{L}(s)} \mu_l(k+1)$ are communicated to users, and the process is repeated. We formally summarize these steps in the following algorithm.

Dual Gradient Projection Algorithm. At times $k = 1, \ldots$, each link $l \in \mathcal{L}$ performs the following steps:

1. Receives the rates $x_s(k)$ from users $s \in \mathcal{S}(l)$ using the link.

2. Updates its price

$$\mu_l(k+1) = \left[\mu_l(k) + \alpha_k \left(\sum_{s \in \mathcal{S}(l)} x_s(k) - c_l \right) \right]^+.$$

3. Communicates the new price $\mu_l(k+1)$ to all users $s \in \mathcal{S}(l)$ using the link l.

At times $k = 1, \ldots$, each user $s \in \mathcal{S}$ performs the following steps:

1. Receives the aggregate price $p_s(k) = \sum_{l \in \mathcal{L}(s)} \mu_l(k)$ [sum of link prices over the links carrying its flow].

2. Computes its new rate by $x_s(k+1) = x_s(p_s(k)$ [i.e., determines the maximizer in $Q_s(p_s(k))$].

3. Communicates the new rate $x_s(k+1)$ to all links $l \in \mathcal{L}(s)$ [the links in its flow path].

In the preceding, we have not specified the stepsize α_k. Under the assumption that the second derivative of each utility u_s is bounded away from zero by $1/a_s$, it can be seen that the gradient of the dual function is Lipschitz continuous, i.e.,

$$\|\nabla q(\mu) - \nabla q(\tilde{\mu})\| \leq L \, \|\mu - \tilde{\mu}\| \qquad \text{for all } \mu, \tilde{\mu} \geq 0.$$

with constant L given by

$$L = \max_{s \in \mathcal{S}} a_s \; \max_{s \in \mathcal{S}} |\mathcal{L}(s)| \; \max_{l \in \mathcal{L}} |S(l)|. \tag{25}$$

We next discuss a convergence result for the method. We use x to denote the vector of user rates $[x_s, \; s \in \mathcal{S}]$ and μ to denote the vector of link prices $[\mu_l, \; l \in \mathcal{L}]$. We assume that the method is started with initial rates $x_s(0) \in I_s$ for all s and initial prices $\mu_l(0) \geq 0$ for all l. The constant stepsize can be used, as seen from the following theorem, which is established in a paper by Low and Lapsley 1999.

Theorem 1 *Assume that each utility function u_s is strictly concave, twice differentiable and increasing on the interval I_s. Furthermore, assume that for each s, the curvature of u_s is bounded away from zero on the interval I_s [i.e., $-u_s''(z) \geq 1/a_s > 0$ for all $z \in I_s$]. Also, assume that the constant stepsize $\alpha_k = \alpha$ is used in the dual gradient projection method, where $0 < \alpha < \frac{2}{L}$ and L as given in Eq. (25). Then, every accumulation point (x^*, μ^*) of the sequence $\{(x(k), \mu(k))\}$ generated by the dual gradient projection algorithm is primal-dual optimal.*

3 Non-Differentiable Dual Function

We focus now on the dual problems where the minimization problem involved in the dual function evaluation has multiple solutions. Under this assumption, for the dual function $q(\mu, \lambda)$ of Eq. (3), we have for any $\mu \geq 0$ and any λ,

$$q(\mu, \lambda) = f(x_{\mu\lambda}) + \mu^T g(x_{\mu\lambda}) + \lambda^T (Ax_{\mu\lambda} - b),$$

where $x_{\mu\lambda}$ is an optimal solution for the following problem

$$\begin{array}{ll} \text{minimize} & f(x) + \mu^T g(x) + \lambda^T (Ax - b) \\ \text{subject to} & x \in X. \end{array} \tag{26}$$

In the presence of multiple solutions $x_{\mu\lambda}$, the dual function is non-differentiable.

At a given $(\hat{\mu}, \hat{\lambda})$, let $x_{\hat{\mu}\hat{\lambda}}$ be a solution to (26) with $(\mu, \lambda) = (\hat{\mu}, \hat{l})$. Let $(\mu, \lambda) \in \operatorname{dom} q$ be arbitrary. We have

$$\begin{aligned} q(\mu, \lambda) &= \inf_{x \in X} \{f(x) + \mu^T g(x) + \lambda^T (Ax - b)\} \\ &\leq f(x_{\hat{\mu}\hat{\lambda}}) + \mu^T g(x_{\hat{\mu}\hat{\lambda}}) + \lambda^T (Ax_{\hat{\mu}\hat{\lambda}} - b) \\ &= f(x_{\hat{\mu}\hat{\lambda}}) + \hat{\mu}^T g(x_{\hat{\mu}\hat{\lambda}}) + \hat{\lambda}^T (Ax_{\hat{\mu}\hat{\lambda}} - b) + (\mu - \hat{\mu})^T g(x_{\hat{\mu}\hat{\lambda}}) + (\lambda^T - \hat{\lambda})(Ax_{\hat{\mu}\hat{\lambda}} - b). \end{aligned}$$

Since $x_{\hat{\mu}\hat{\lambda}}$ is the minimizer at which $q(\hat{\mu}, \hat{\lambda})$ is attained, it follows for any $(\mu, \lambda) \in \operatorname{dom} q$,

$$q(\mu, \lambda) \leq q(\hat{\mu}, \hat{\lambda}) + (\mu - \hat{\mu})^T g(x_{\hat{\mu}\hat{\lambda}}) + (\lambda^T - \hat{\lambda})(Ax_{\hat{\mu}\hat{\lambda}} - b).$$

If we multiply the preceding relation with (-1), we would have

$$-q(\hat{\mu}, \hat{\lambda}) - (\mu - \hat{\mu})^T g(x_{\hat{\mu}\hat{\lambda}}) - (\lambda^T - \hat{\lambda})(Ax_{\hat{\mu}\hat{\lambda}} - b) \leq -q(\mu, \lambda) \qquad \text{for all } (\mu, \lambda) \in \operatorname{dom}(-q),$$

showing that $-\Big(g(x_{\hat{\mu}\hat{\lambda}}), Ax_{\hat{\mu}\hat{\lambda}} - b\Big)$ is a subgradient of convex function $-q$ at $(\hat{\mu}, \hat{\lambda})$. Equivalently, $\Big(g(x_{\hat{\mu}\hat{\lambda}}), Ax_{\hat{\mu}\hat{\lambda}} - b\Big)$ is a subgradient of the concave function q at $(\hat{\mu}, \hat{\lambda})$.

Thus, we can solve the dual problem by using subgradient method (adapted to maximization). The method has the form

$$\mu_{k+1} = [\mu_k + \alpha_k \, g(x_k)]^+ ,$$

$$\lambda_{k+1} = \lambda_k + \alpha_k \, (Ax_k - b),$$

where x_k is any minimizer of the problem in (26) with $(\mu, \lambda) = (\mu_k, \lambda_k)$. Hence, we can apply any of the stepsize rules discussed for the subgradient methods.