Lecture-25: Dual algorithm

1 Distributed algorithms: dual solution

We consider another distributed algorithm based on the dual formulation of the utility maximization problem.

Definition 1.1 (NUM). Recall that set of feasible allocations is given by $\mathcal{D} \triangleq \left\{ x \in \mathbb{R}_+^{\mathcal{S}} : y \leqslant c \right\}$. Let $U_r : \mathbb{R}_+ \to \mathbb{R}$ be a strictly concave, increasing, and continuously differentiable utility function for each source $r \in \mathcal{S}$. Then, the resource allocation problem to maximize network utility that satisfies link capacity constraints is

$$x^* \triangleq \arg\max \left\{ \sum_{r \in \mathcal{S}} U_r(x_r) : x \in \mathcal{D} \right\}. \tag{1}$$

Definition 1.2 (Lagrangian for NUM). We can define Lagrangian $L : \mathbb{R}^{\mathcal{S}} \times \mathbb{R}_+^{\mathcal{L}} \to \mathbb{R}$ for network utility problem of Definition 1.1 in terms of source rates $x \in \mathbb{R}^{\mathcal{S}}$ and Lagrange multiplier vector $p \in \mathbb{R}_+^{\mathcal{L}}$ to incorporate the constraints into the maximization, as

$$L(x,p) \triangleq \sum_{r \in \mathcal{S}} U_r(x_r) - \sum_{\ell \in \mathcal{L}} p_\ell (y_\ell - c_\ell). \tag{2}$$

Definition 1.3. Treating Lagrange multiplier vector $p \in \mathbb{R}_+^{\mathcal{L}}$ as the link price vector, for any fixed routing matrix $R \in \{0,1\}^{\mathcal{L} \times \mathcal{S}}$, we can define the the aggregate link price for each route $r \in \mathcal{S}$ as the route price vector $q \triangleq R^\top p \in \mathbb{R}_+^{\mathcal{S}}$.

Definition 1.4 (Dual for NUM). We define dual objective as $D : \mathbb{R}_+^{\mathcal{L}} \to \mathbb{R}$ for network utility problem of Definition 1.1 in terms of Lagrangian L of Definition 1.2 for any Lagrange multiplier vector $p \in \mathbb{R}_+^{\mathcal{L}}$, as

$$D(p) \triangleq \max_{x \in \mathbb{R}_+^{\mathcal{S}}} L(x,p). \tag{3}$$

Lemma 1.5. *If the inverse of derivative of utility function $(U_r')^{-1} : \mathbb{R} \to \mathbb{R}$ is positive for any positive number, then the Lagrangian $L : \mathbb{R}^{\mathcal{S}} \times \mathbb{R}_+^{\mathcal{L}} \to \mathbb{R}$ in Definition 1.2 is maximized for source rate vector $x^*(p) \in \mathbb{R}_+^{\mathcal{S}}$ in terms of link price vector $p \in \mathbb{R}_+^{\mathcal{L}}$, defined for each route $r \in \mathcal{S}$ as*

$$x_r^*(p) \triangleq (U_r')^{-1}(q_r). \tag{4}$$

Proof. Recalling that link load vector $y = Rx$ and route price vector $q = R^\top p$ for a fixed routing matrix R, source rate vector x, and link price vector p, we can rewrite the Lagrangian as

$$L(x,p) = \sum_{r \in \mathcal{S}} (U_r(x_r) - x_r q_r) + \sum_{\ell} p_\ell c_\ell.$$

The second part of the Lagrangian is independent of source rate x, and the first part tensorizes over each source rate x_r, and thus the result follows by maximizing over individual source rates. □

Remark 1. Note that we have assumed that $x_r^* = (U_r')^{-1}(q_r) > 0$ in writing down (4). This would be true, for example, if the utility function is an α-fair utility function with $\alpha > 0$.

Theorem 1.6. *Network utility maximization of Definition 1.1 is equivalent to minimizing the corresponding dual objective of Definition 1.4. That is,*

$$\max_{x \in \mathcal{D}} \sum_{r \in \mathcal{S}} U_r(x_r) = \min_{p \in \mathbb{R}_+^{\mathcal{L}}} D(p).$$

Proof. Since network utility maximization problem is a convex optimization problem, the duality gap is zero. □

1.1 Dual algorithm

As in the case of the primal problem, we would like to design an algorithm that ensures that all the source rates converge to the optimal solution. Note that, in this case, we are looking for a gradient descent (rather than the gradient ascent we saw in the primal formulation), since we would like to minimize the dual objective $D(p)$. To find the direction of the gradient, we need to know $\frac{\partial D}{\partial p_\ell}$.

Lemma 1.7. *Under conditions in Lemma 1.5, the gradient of the dual objective in Definition 1.4 is*

$$\nabla D(p) = -(y - c).$$

Proof. Under the conditions in Lemma 1.5, the Lagrangian is maximized for source rate $x^*(p)$ given in (4) and hence the link load vector is $y^*(p) = Rx^*(p)$. Thus, the dual objective is given by

$$D(p) = L(x^*(p), p) = \sum_{r \in \mathcal{S}} U_r(x_r^*(p)) - \sum_{\ell \in \mathcal{L}} p_\ell(y_\ell^*(p) - c_\ell).$$

Taking partial derivative with respect to p_ℓ on both sides, substituting $y_\ell^* = \sum_{r \in \mathcal{S}} R_{\ell,r} x_r^*(p)$ and $q_r = \sum_{\ell \in \mathcal{L}} R_{\ell,r} p_\ell$, we get

$$\frac{\partial}{\partial p_\ell} D(p) = \sum_{r \in \mathcal{S}} \Big(U_r'(x_r^*(p)) - q_r \Big) \frac{\partial}{\partial p_\ell} x_r^*(p) - (y_\ell^*(p) - c_\ell).$$

The result follows from substituting (4) in the above equation. □

Definition 1.8 (Dual algorithm). Consider the *dual algorithm* where the evolution of link price vector $p : \mathbb{R}_+ \to \mathbb{R}_+^{\mathcal{L}}$ is an autonomous dynamic system governed by $\dot{p} = \phi(p)$ for a map $\phi : \mathbb{R}^{\mathcal{L}} \to \mathbb{R}^{\mathcal{L}}$ such that

$$\dot{p}_\ell = \phi_\ell(p) \triangleq -h_\ell \frac{\partial}{\partial p_\ell} D(p) = h_\ell \big(\sum_{\ell \in \mathcal{L}} R_{\ell,r} (U_r')^{-1}(q_r) - c_\ell \big)_{p_\ell}^+. \tag{5}$$

where $h_\ell > 0$ is a constant and $(g(x))_x^+ \triangleq g(x) \mathbb{1}_{\{x>0\}} + (g(x) \vee 0) \mathbb{1}_{\{x=0\}}$.

Remark 2. Note that p_ℓ is not allowed to become negative because the KKT conditions impose such a condition on the Lagrange multipliers.

Remark 3. Let $h_\ell = 1$, then we observe that $\dot{p}_\ell = (y_\ell - c_\ell)_{p_\ell}^+$ where $y_\ell = \sum_{\ell \in \mathcal{L}} R_{\ell,r} x_r$. That is, the price p_ℓ has the same dynamics of the queue length at the link. The rate of increase of queue p_ℓ is the arrival rate y_ℓ, and the rate of decrease is the link capacity c_ℓ. The queue is not allowed to be negative, hence the rate of decrease is zero when queue is zero. Thus, the links do not have to compute their dual variables explicitly when $h_\ell \equiv 1$. The dual variables are simply the queue lengths.

Remark 4. The stability of this algorithm follows in a manner similar to that of the primal algorithm by considering $D(p)$ as the Lyapunov function, since the dual algorithm is simply a gradient algorithm for finding the minimum of $D(p)$.

2 Feedback delay and stability

We saw that the primal and dual algorithms are globally asymptotically stable if link prices were fed back instantaneously to sources and rate adjustments at sources were reflected instantaneously at links. Both assumptions are not true in reality due to delays. To illustrate how to analyze congestion control algorithms in the presence of delays, we will consider a simple one-link network and study the impact of feedback delay on the stability of a primal congestion controller. Consider a simple system with one link of capacity c and one source with utility function $U(x) = \ln x$ and constraint $x \leqslant c$. The price for this link is given by $f(x)$ where x is the link load. Since there is a single source and link, the route price is also $f(x)$. We choose $k(x) = \kappa x$ for some $\kappa > 0$, so that when x is close to zero the rate of increase is bounded. Consequently, the proportionally fair congestion controller derived from primal solution is

$$\dot{x} = k(x) \Big(\frac{1}{x} - f(x) \Big) = \kappa(1 - x f(x)). \tag{6}$$

Suppose that $f(x)$ is the loss probability or marking probability when the arrival rate at the link is x. Then, the congestion control algorithm can be interpreted as follows: increase x at rate κ and decrease it proportional to the rate at which packets are marked, with κ being the proportionality constant. In reality, there is a delay from the time at which a packet is released at the source to the time at which it reaches the link, called the forward delay, and a delay for any feedback to reach the source from the link, called the backward delay. So, one cannot implement the above congestion controller in the form (6). Let us denote the forward delay by T_f and the backward delay by T_b. Taking both delays into consideration and denoting the round-trip time (RTT) as $T \triangleq T_f + T_b$, we have

$$y(t) = x(t - T_f), \qquad q(t) = p(t - T_b) = f(y(t - T_b)) = f(x(t - T_f - T_b)) = f(x(t - T)).$$

The congestion controller becomes the following delay differential equation,

$$\dot{x} = \kappa(1 - x(t - T)f(x(t - T))). \tag{7}$$

Let $\hat{x}$ be the unique equilibrium point of this delay differential equation, i.e., $1 - \hat{x}f(\hat{x}) = 0$. To analyze this system, we will assume that $x(t)$ is close to the equilibrium and derive conditions under which the system is asymptotically stable. As such we define the perturbation for any time $t \in \mathbb{R}_+$ as

$$z(t) \triangleq x(t) - \hat{x},$$

and find the conditions under which $\lim_{\to\infty} z(t) = 0$. To this end, we will linearize the system around the equilibrium point and derive conditions under which the resulting linear delay differential equation is asymptotically stable. Clearly, we would like to study the global asymptotic stability of the system instead of just the asymptotic stability assuming that the system is close to its equilibrium, but typically studying the global asymptotic stability is a much harder problem when there are many links and many sources.

2.1 Linearization

To study the asymptotic stability of (7), we substitute $x(t) = \hat{x} + z(t)$ and linearize the delay differential equation by substituting the Taylor series expansion $f(\hat{x} + z(t - T)) \approx f(\hat{x}) + f'(\hat{x})z(t - T)$ upto the first order linear terms in the neighborhood of $\hat{x}$. The rationale is that, since x is close to $\hat{x}$, perturbation $z = x - \hat{x}$ is close to zero and z dominates z^2, z^3, etc. We obtain the linearized differential equation

$$\dot{z} = \kappa(1 - (\hat{x} + z(t - T))f(\hat{x} + z(t - T)) \approx \kappa(1 - (\hat{x} + z(t - T))(f(\hat{x}) + f'(\hat{x})z(t - T)).$$

Using the equilibrium condition $1 - \hat{x}f(\hat{x}) = 0$ and dropping the z^2 terms for the same reasons, we obtain

$$\dot{z} \approx -\kappa(f(\hat{x}) + \hat{x}f'(\hat{x}))z(t - T). \tag{8}$$

To understand the stability of this linear delay differential equation, we introduce the following theorem, which characterizes the necessary and sufficient condition for a general linear delay differential equation to be asymptotically stable.

Theorem 2.1. *Consider a system governed by the following linear delay differential equation*

$$\dot{x}(t) = -ax(t) - bx(t - T),$$

where the initial condition $x(t)$ for time $t \in [-T, 0]$, is specified. For any choice of the initial condition, $\lim_{t\to\infty} x(t) =$ 0 the unique equilibrium of the system, if and only if there exists a χ such that

$$\chi \in (0, \frac{\pi}{T}), \qquad a = -c\cos\chi T, \qquad c = \frac{\chi}{\sin\chi T}, \qquad b \in [-a, c].$$

If $a = 0$, the condition simplifies to $b \in [0, \frac{\pi}{2T}]$.

Proposition 2.2. *The linearized version of the proportionally fair controller, given in* (8)*, is asymptotically stable if and only if*

$$\kappa T(f(\hat{x}) + \hat{x}f'(\hat{x})) \leqslant \frac{\pi}{2}. \tag{9}$$

Equation (9) suggests that the parameter κ should be chosen inversely proportional to T. This means that the congestion control algorithm should react more slowly when the feedback delay is large. This is very intuitive since, when T is large, the algorithm is reacting to events that occurred a long time ago and thus should change the source rate x very slowly in response to such old information. We considered the problem of deriving delay differential equation models for congestion control protocols operating on a single link accessed by a single source. The modeling part can be easily extended to general networks with an arbitrary number of links and sources. However, the stability analysis of the resulting equations is considerably more challenging than in the case of the simple single-source, single-link model.