Duality

Theo Diamandis

January 31, 2023

In this lecture, we're going to cover duality theory. Duality gives us a set of tools to construct lower bounds on the optimal objective values of optimization problems. For convex optimization problems, these lower bounds are, in most cases, tight. As a result, we can use duality theory to certify that a feasible point for a convex optimization problem is optimal. For a nonconvex problem, we can bound the suboptimality of a locally optimal point. Much of this lecture follows [BV04, Ch. 5].

1 The Lagrange dual function

Consider an optimization problem in standard form:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array} \tag{1}$$

Let $\mathcal{D}$ denote the domain of this problem, which we assume is nonempty. We do not assume this problem is convex (yet). The idea of Lagrangian duality is to relax the constraints in (1) into penalties in the objective. The *Lagrangian* associated with (1) is

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint. Similarly, ν_i is the Lagrange multiplier associated with the ith equality constraint.

The dual function. The *dual function* associated with (1) is the minimum value of the Lagrangian over x for $\lambda \in \mathbf{R}^m$ and $\nu \in \mathbf{R}^p$:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).$$

Note that this is a pointwise infimum of a family of affine functions, so it is always concave even when the optimization problem is not. When the Lagrangian is unbounded below in x, the dual function takes on value $-\infty$. We will now restrict λ to be elementwise nonnegative. Since any point x that is feasible for the original problem will suffer no positive penalty for any $\lambda \in \mathbf{R}_+^m$, $\nu \in \mathbf{R}^p$, we have that

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D},\, f_i(x) \leq 0,\, h_i(x)=0} f_0(x) = p^\star.$$

In other words, the dual function is a lower bound on the optimal objective value.

Example: Two way partitioning problem. The two way partitioning problem attempts to partition n items into two sets. Let W_{ij} be the cost of assigning items i and j to the same set and $-W_{ij}$ be the cost of assigning them to different sets. The problem is then

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

This problem is nonconvex due to the binary constraint on x. However, we can still write the dual function

$$\begin{aligned} g(\nu) &= \inf_x \left(x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) \right) \\ &= \inf_x x^T (W + \mathbf{diag}\,(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}\,(\nu) \in \mathbf{S}_+^n \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

We know that g evaluated at any feasible ν gives a lower bound for $p^\star$. Thus, if we take $\nu = -\lambda_{\min}(W)\mathbf{1}$, we get the bound[1]

$$p^\star \geq n \lambda_{\min}(W).$$

2 The Lagrange dual problem

Since $g(\lambda, \nu)$ for $\lambda \geq 0$ gives a lower bound on $p^\star$, it is natural to try to find the best lower bound by solving the optimization problem

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

We call this problem the *Lagrange dual problem* associated with (1). In this context, λ and ν are referred to as *dual variables*. We refer to its optimal value as $d^\star$. Similarly, we refer to (1) as the *primal problem* and x as the *primal variable*. If a pair (λ, ν) is such that $\lambda \geq 0$ and $g(\lambda, \nu) \geq -\infty$, we call this pair *dual feasible*. If the primal problem is unbounded below, *i.e.*, $p^\star = -\infty$, then no such pair exists, *i.e.*, the problem (2) is infeasible.

[1] This bound can also be obtained by relaxing the constraint to $\sum_i x_i^2 = \|x\|_2^2 = n$ and recognizing the problem as an eigenvalue problem.

Weak duality. The discussion above suggests the important inequality

$$d^\star \leq p^\star,$$

referred to as *weak duality.* Weak duality holds for all optimization problems and can be useful to bound nonconvex ones. The quantity $p^\star - d^\star$ is referred to as the *duality gap.* Importantly, many nonconvex problems have zero duality gap (*e.g.,* some nonconvex QCQPs).

Strong duality. When the primal problem is convex, *i.e.,* has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array}$$

with all f_i's convex, then the duality gap is zero under mild conditions, *i.e.,*

$$d^\star = p^\star.$$

A simple condition is *Slater's condition*, which stats that there exists a strictly feasible point $x \in \textbf{relint}\, \mathcal{D}$, *i.e.,* a point such that

$$f_i(x) < 0, \quad i = 1, \ldots, m \qquad \text{and} \qquad Ax = b.$$

Usually, this holds for convex problems. For the rest of this lecture, we will assume the problem we are working with is convex unless stated otherwise.

2.1 Examples

Note that often there are several equivalent ways to write an optimization problem, for example two equivalent formulations of ℓ_1-reguarlized regression are

$$\text{minimize} \quad \|Ax - b\|_2^2 + \lambda \|x\|_1$$

and

$$\begin{array}{ll} \text{minimize} & y^T y + \lambda \|x\|_1 \\ \text{subject to} & y = Ax - b. \end{array}$$

Equivalent formulations of the same problem can lead to very different duals. Some of these duals will be useful and some will not be. It is therefore common to transform the primal problem via transformations including

- introducing new variables and equality constraints, as above
- making explicit constraints implicit (*e.g.,* via an indicator function) or vice versa
- transforming the objective (*e.g.,* replace $f_0(x)$ with $\phi(f_0(x))$), where ϕ is increasing and convex

We will see a few examples below.

Least norm problem. Consider the standard least norm problem

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b. \end{array}$$

We assume that A is a wide matrix with full row rank. The Lagrangian is

$$L(x, \nu) = x^T x + \nu^T (Ax - b).$$

To find the dual function, we minimize over x (set the gradient equal to zero):

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \implies x = -(1/2) A^T \nu.$$

Plugging this into L, we have that

$$g(\nu) = -(1/4) \nu^T A A^T \nu - b^T \nu,$$

which is easily verified to be a concave function of ν. By the lower bound property, we have that

$$p^\star \geq -(1/4) \nu^T A A^T \nu - b^T \nu$$

for any ν. Since this problem is unconstrained, we can write the solution explicitly:

$$\nu^\star = 2(AA^T)^{-1} b.$$

Plugging this into our equation for the optimal x, we get

$$x^\star = -(1/2) A^T \nu = A^T (AA^T)^{-1} b.$$

Some of you will recognize this as the pseudoinverse of A.

Standard form LP. Recall that a 'standard form' LP is the optimization problem of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0. \end{array}$$

The Lagrangian of this problem is

$$L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

Since L is affine in x, the dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

This leads to the dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\nu \\ \text{subject to} & A^T\nu + c = \lambda \\ & \lambda \geq 0. \end{array}$$

Often, this is rewritten as

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & c \geq A^T y. \end{array}$$

(Convince yourself that this transformation is valid.)

Norm approximation problem. Recall the norm approximation problem

$$\text{minimize} \quad \|Ax - b\|$$

for some norm $\|\cdot\|$. To take the dual, we will rewrite this as

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = Ax - b. \end{array}$$

The Lagrangian is

$$L(x, y, \nu) = \|y\| + \nu^T y - \nu^T Ax + \nu^T b.$$

Minimizing over x and y, we get the dual function

$$g(\nu) = \inf_{x,y} \left(\|y\| + \nu^T y - \nu^T Ax + \nu^T b\right) = b^T\nu + \inf_y \left(\|y\| + \nu^T y\right) + \inf_x (-\nu^T Ax).$$

It is clear that we need $A^T\nu = 0$. Furthermore, since $w^T u \leq \|w\|\|u\|_*$, we need $\|\nu\|_* \leq 1$, where $\|u\|_* = \sup_{\|w\| \leq 1} w^T u$ is the dual norm of $\|\cdot\|$. Thus,

$$g(\nu) = \begin{cases} b^T\nu & A^T\nu = 0, \quad \|\nu\|_* \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$$

This gives us the dual problem

$$\begin{array}{ll} \text{maximize} & b^T\nu \\ \text{subject to} & A^T\nu = 0 \\ & \|\nu\|_* \leq 1. \end{array}$$

3 Optimality Conditions

When strong duality holds (again, this is usually the case in convex problems), we can write down a set of optimality conditions. Many optimization solvers try to solve these conditions, a set of nonlinear equations, directly. In addition, these conditions allow us to characterize how ‘bad’ a feasible point is compared to an optimal solution. Sometimes these conditions even allow us to solve optimization problems explicitly.

Complementary slackness. Assume strong duality holds, $x^\star$ is primal optimal and $(\lambda^\star, \nu^\star)$ is dual optimal. Then

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\le f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\le f_0(x^\star). \end{aligned}$$

Since $\lambda \ge 0$, this means that we must have each term in the sum be 0, *i.e.*,

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \ldots, m.$$

In other words,

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0, \quad \text{and} \quad f_i(x^\star) < 0 \implies \lambda_i^\star = 0$$

This condition is known as *complementary slackness*.

Karush-Kuhn-Tucker (KKT) conditions. The following four conditions are called the *KKT conditions*. (Here, we assume the problem is differentiable, but these can be easily generalized.) If a point (x, λ, ν) satisfies the KKT conditions, then it is an optimal point.

1. **primal feasibility:** $f_i(x) \le 0$ and $Ax = b$.
2. **dual feasibility:** $\lambda \ge 0$
3. **complementary slackness:** $\lambda_i^\star f_i(x^\star) = 0$
4. **gradient of the Lagrangian at x vanishes:**

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T \nu = 0$$

Example: water-filling. Water filling is the problem of allocating power optimally to n communication channels:

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(x_i + a_i) \\ \text{subject to} & \mathbf{1}^T x = 1 \\ & x \ge 0. \end{array}$$

We can write the optimality conditions for a primal-dual point (x, λ, ν):

$$\begin{aligned}
&\mathbf{1}^T x = 1, \quad x \geq 0 \\
&\lambda \geq 0 \\
&\lambda_i x_i = 0 \\
&\frac{1}{x_i + a_i} + \lambda_i = \nu.
\end{aligned}$$

From the equations above, we determine that

- if $\nu < 1/a_i$, then $\lambda_i = 0$ and $x_i = 1/\nu - a_i$
- if $\nu \geq 1/a_i$ then $\lambda_i = \nu - 1/a_i$ and $x_i = 0$

Thus, we can determine ν from the single variable equation

$$\mathbf{1}^T x = \sum_{i=1}^n \max\{0, 1/\nu - a_i\} = 1.$$

This has an interesting interpretation: we have n patches each with height a_i. We flood this area with a unit amount of water. The resulting height of the water is $1/\nu^\star$. Finding $\nu^\star$ (a simple 1d equation solve!) allows us to reconstruct the optimal x (an n-vector).

4 Sensitivity analysis

We consider the perturbed problem

$$\begin{array}{ll}
\text{maximize} & f_0(x) \\
\text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\
& a_i^T x - b_i = v_i, \quad i = 1, \ldots, p.
\end{array} \tag{3}$$

Let $p^\star(u, v)$ be the optimal value to this problem. We are interested in how this optimal value changes as we change u and v, *i.e.,* as we tighten and loosen constraints.

Assuming strong duality holds and $\lambda^\star$, $\nu^\star$ are dual optimal, we can use weak duality to say

$$p^\star(u, v) \geq g(\lambda^\star, \nu^\star) - u^T \lambda^\star - v^T \nu^\star = p^\star(0, 0) - u^T \lambda^\star - v^T \nu^\star.$$

This inequality allows us to use the dual variables to say how sensitive $p^\star$ is to each constraint. For example

- If $\lambda_i^\star$ is large, then $p^\star$ increases greatly if we tighten constraint i ($u_i < 0$)
- If $\lambda_i^\star$ is small, then $p^\star$ does not decrease much if we loosen constraint i ($u_i > 0$)
- If $\nu_i^\star$ is large and negative, $p^\star$ increases greatly if we take $v_i > 0$

- If $\nu_i^\star$ is small and positive, $p^\star$ does not decrease much if we take $v_i > 0$.
- And so on...

Furthermore, if $p^\star(u, v)$ is differentiable at $(0, 0)$, then we can interpret $\lambda_i^\star$ and $\nu_i^\star$ as the (negative) partial derivatives of $p^\star$ w.r.t. the ith inequality and equality constraint respectively. In essence, the dual variables give us the cost (or gain), measured in units of the objective function, of tightening or relaxing a particular constraint by a small amount.

5 Example: FX exchange

In this problem, we consider the problem of optimally exchanging currencies. First, we introduce some preliminaries. Consider a single currency exchange market, for example for USD-EUR. We will tender some amount of one currency *to* the market, for which we will receive some amount of the other currency *from* the market. Let δ_i be the amount of currency i that we tender to the market and λ_i be the amount of currency i that we receive from the market. (In any reasonable trade, we expect only one δ to be non-zero.)

Forward exchange function. The *forward exchange function* describes how much of currency j, λ_j we receive by tendering δ_i units of currency i on a particular market

$$f_{ij}(\delta_i) = \lambda_j.$$

We assume that this function is concave and non-decreasing. Consider a simple example of an order book where there are offers to sell asset 2 at prices (denominated in terms of asset 1) 1 (100 units), 2 (200 units) and 3 (300 units). The function f_{12} is then piecewise linear and concave:

$$f_{12}(\delta) = \begin{cases} \delta & 0 \leq \delta \leq 100 \\ 100 + (1/2)(\delta - 100) & 100 \leq \delta \leq 300 \\ 300 + (1/3)(\delta - 300) & 300 \leq \delta \leq 600 \\ 600 & 600 \leq \delta. \end{cases}$$

It is convenient to stack the numbers δ_i and λ_j into vectors

$$\Delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \qquad \text{and} \qquad \Lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}.$$

(Again, any reasonable trade has $\Delta_i \Lambda_i = 0$ for $i = 1, 2$, since trading back and forth incurs transaction fees: $f_{21}(f_{12}(\delta_1)) < \delta_1$.) We call a trade (Δ, Λ) *feasible* if $f_{12}(\delta_1) = \lambda_2$ and $f_{21}(\delta_2) = \lambda_1$. The net trade is then $\Lambda - \Delta$, which we expect to have one negative element, corresponding to the currency tendered, and one positive element, corresponding to the currency received.

Network trade vector. Now consider a 'universe' of currencies $i = 1, \ldots, n.$ and a set of markets $k = 1, \ldots, m$. Note that a single 'exchange' likely will have multiple markets (*e.g.,* one for USD-EUR, another for USD-CAD, etc). We denote the trade with market k by Δ^k and Λ^k. The *network trade vector* is then

$$\Psi = \sum_{k=1}^{m} A^k(\Lambda^k - \Delta^k),$$

where A^k is a $n \times 2$ matrix mapping market k's local indices into the global currency indices. For example, if our universe has 3 currencies and market k trades currencies 2 and 3, then

$$A^k = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Trade utility function. Now that we have defined the preliminaries, we can introduce a utility function $U : \mathbf{R}^n \to \mathbf{R} \cup \{-\infty\}$ which gives our utility of some net trade Ψ. We assume that U is concave and increasing. Furthermore, we will use infinite values of U to encode constraints; a trade Ψ such that $U(\Psi) = -\infty$ is unacceptable to the trader. We can choose U to encode several important actions in markets. For example, if we want to exchange some amount z of currency i for as much currency j as possible, we can take

$$U(\Psi) = \begin{cases} \Psi_j & ze_i + \Psi \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

If we want to detect arbitrage in the markets (*i.e.,* find a series of trades where we can receive a non-negative amount of all currencies), we can use the utility function

$$U(\Psi) = \begin{cases} c^T\Psi & \Psi \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

where c gives the relative weighting between the currencies (*e.g.,* 'true prices' or prices on an external, infinitely liquid reference market).

Finding an optimal trades. Finally, we are ready to construct the problem of finding the optimal set of trades (Δ^k, Λ^k):

$$\begin{array}{ll} \text{maximize} & U(\Psi) \\ \text{subject to} & \Psi = \sum_{k=1}^{m} A^k(\Lambda^k - \Delta^k) \\ & f_{12}(\delta_1^k) = \lambda_2^k, \quad k = 1, \ldots m \\ & f_{21}(\delta_2^k) = \lambda_1^k, \quad k = 1, \ldots m \end{array}$$

First, we note that this problem is not convex because of the inequality constraints, but it can be relaxed to the equalvalent convex problem

$$
\begin{array}{ll}
\text{maximize} & U(\Psi) \\
\text{subject to} & \Psi = \sum_{k=1}^{m} A^k(\Lambda^k - \Delta^k) \\
& f_{12}(\delta_1^k) \geq \lambda_2^k, \quad k = 1, \ldots m \\
& f_{21}(\delta_2^k) \geq \lambda_1^k, \quad k = 1, \ldots m \\
& \lambda_i^k,\, \delta_i^k \geq 0, \quad i = 1, 2, \quad k = 1, \ldots m
\end{array} \tag{4}
$$

(Convince yourself that these inequalities will always be tight at a solution to the problem.) We will see that the dual of (4) has a simple structure and nice interpretation.

Dual problem. Recognize that the only constraint 'coupling' the markets together in (4) is the linear equality

$$\Psi = \sum_{k=1}^{m} A^k(\Lambda^k - \Delta^k).$$

This suggests a way to dualize the problem. We introduce the Lagrangian

$$L(\Psi, \Delta^k, \Lambda^k, \nu) = U(\Psi) + \nu^T \left(-\Psi + \sum_{k=1}^{m} A^k(\Lambda^k - \Delta^k) \right) - \sum_{k=1}^{m} I_{C^k}(\Delta^k, \Lambda^k),$$

where $I_S(x)$ is the indicator function of the set S and C^k is the set of feasible trades for market k,

$$C^k = \{(\Delta^k, \Lambda^k) \mid f_{12}(\delta_1^k) \geq \lambda_2^k,\; f_{21}(\delta_2^k) \geq \lambda_1^k \;\; \Delta^k, \Lambda^k \geq 0\}.$$

We rewrite the Langrangian slightly as

$$L(\Psi, \Delta^k, \Lambda^k, \nu) = U(\Psi) - \nu^T \Psi + \sum_{k=1}^{m} \left(({A^k}^T \nu)^T (\Lambda^k - \Delta^k) - I_{C^k}(\Delta^k, \Lambda^k) \right)$$

Now we obtain the dual function by maximizing over the primal variables:

$$
\begin{aligned}
g(\nu) &= \max_{\Psi, \Delta^k, \Lambda^k} L(\Psi, \Delta^k, \Lambda^k, \nu) \\
&= \sup_{\Psi} \left(U(\Psi) - \nu^T \Psi \right) + \sum_{k=1}^{m} \sup_{\Delta^k, \Lambda^k} \left(({A^k}^T \nu)^T (\Lambda^k - \Delta^k) - I_{C^k}(\Delta^k, \Lambda^k) \right).
\end{aligned}
$$

The first term has a closed form solution for many utility functions.[2] For example, for the arbitrage utility function, we have that

$$\begin{aligned}\sup_{\Psi}\left(U(\Psi) - \nu^T\Psi\right) &= \sup_{\Psi}\left(c^T\Psi - I(\Psi \geq 0) - \nu^T\Psi\right) \\ &= \sup_{\Psi \geq 0}\,(c-\nu)^T\Psi \\ &= \begin{cases} 0 & c - \nu \leq 0 \\ +\infty & \text{otherwise.}\end{cases}\end{aligned}$$

The second term separates over all markets. The summand can be recognized as the optimal value of the optimization problem

$$\begin{array}{ll} \text{maximize} & ({A^k}^T\nu)^T(\Lambda^k - \Delta^k) \\ \text{subject to} & f_{12}(\delta_1^k) \geq \lambda_2^k \\ & f_{21}(\delta_2^k) \geq \lambda_1^k, \qquad\qquad \Delta^k, \Lambda^k \geq 0 \end{array}$$

which is exactly the arbitrage problem for market k with 'true prices' ν (indexed by $(A^k)^T$ to convert from global to local indices). Thus, the dual problem

$$\text{minimize} \quad g(\nu)$$

is to find the optimal prices ν instead of the optimal trades Δ^k, Λ^k. This interpretation of dual variables as prices is quite common and appears even in non-financial applications (but its appearance in financial applications can be quite explicit). The fact that this dual problem is unconstrained and parallelizes over all markets considered, makes it very efficient to solve.

Optimality condition. Consider the single-CFMM arbitrage problem

$$\begin{array}{ll} \text{maximize} & \nu^T(\Lambda - \Delta) \\ \text{subject to} & f_{12}(\delta_1) \geq \lambda_2 \\ & f_{21}(\delta_2) \geq \lambda_1 \\ & \Delta, \Lambda \geq 0, \end{array}$$

where we drop the superscripts for convenience. The Lagrangian is

$$L(\Delta, \Lambda, \nu, \mu, \xi, \zeta) = -\nu^T(\Lambda - \Delta) + \mu_1(\lambda_2 - f_{12}(\delta_1)) + \mu_2(\lambda_1 - f_{21}(\delta_2)) - \zeta^T\Delta - \xi^T\Lambda.$$

The optimality conditions for this problem are primal feasibility (the trades are valid), dual feasibility ($\mu \geq 0$), complementary slackness and the gradient condition

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ -\nu_1 \\ -\nu_2 \end{bmatrix} + \begin{bmatrix} \mu_1 f_{12}'(\delta_1) \\ \mu_2 f_{21}'(\delta_2) \\ \mu_2 \\ \mu_1 \end{bmatrix} + \begin{bmatrix} -\zeta \\ -\xi \end{bmatrix} = 0$$

[2] In fact, it is closely related to an important object in convex analysis called the *conjugate function*, which we did not have time to cover in this course.

Looking at the first and last equation, and using the fact that ζ, $\xi \geq 0$, we get

$$\mu_1 f'_{12}(\delta_1) \geq \nu_1 \qquad \text{and} \qquad \nu_2 \geq \mu_1,$$

allowing us to conclude

$$\nu_2 f'_{12}(\delta_1) \geq \nu_1.$$

A similar argument can be made with the second and third equation. The optimality conditions simplify to

$$\nu_1/\nu_2 \leq f'_{12}(\delta_1) \qquad \text{and} \qquad \nu_2/\nu_1 \leq f'_{21}(\delta_2).$$

We refer to f', which gives the marginal price of additional purchasing after a trade has been made, as the price impact function. This says that the prices after the trade is made must be equal to the 'true prices'. The no-trade condition is then

$$1/f'_{21}(0) \leq \nu_1/\nu_2 \leq f'_{12}(0)$$

A similar condition can be derived for the general FX exchange problem.

6 Example: Rate control

In this section, we consider the general problem of allocating network capacity. This comes up in many applications, including routing of internet packets. We will assume there are m links in the network and n 'flows', which have fixed routes. The variable $x_j \geq 0$ is the rate of flow j (for example, the bandwidth used by a particular internet connection). Each flow has an associated loss function $\ell(x_j)$, which is the 'displeasure' this user experiences with flow x_j. (Alternatively, we can cast everything in terms of utility functions and get the same results.) We assume that these loss functions are all convex and we encode nonegativity constrains by setting $\ell_j(x_j) = +\infty$ for $x_j < 0$. For example, a user may be unhappy with a slow connection but indifferent with a flow over some threshold, giving a loss function such as

$$\ell_j(x_j) = \begin{cases} +\infty & x_j < 0 \\ (x-a)^2 & 0 \leq x_j \leq a \\ 0 & a < x_j. \end{cases}$$

The traffic t_i on a particular link i is the sum of the flows over that link, and each link has some maximum capacity c_i. We introduce the routing matrix R where R_{ij} is 1 if flow j passes over link i and zero otherwise. Thus, Rx gives the total traffic on each link, and we have the constraint

$$Rx \leq c.$$

Our optimization problem is then to minimize the user losses subject to the flow constraint:

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{n} \ell_j(x_j) \\ \text{subject to} & Rx \leq c. \end{array}$$

Dual problem. The Lagrangian for the rate control problem is

$$L(x,\lambda) = \sum_{j=1}^{n} \ell_j(x_j) + \lambda^T(Rx - c) = -\lambda^T c + \sum_{j=1}^{n} \left(\ell_j(x_j) + (r_j^T \lambda)x_j\right),$$

where r_j is the jth column of R. The dual function is then

$$g(\lambda) = -\lambda^T c + \sum_{j=1}^{n} \inf_{x_j} \left(\ell_j(x_j) + (r_j^T \lambda)x_j\right).$$

Recall that we can interpret λ_i as the 'price' for link i. Since r_j is a $\{0,1\}^m$ vector with 1 in the ith entry if flow j passes over link i, the quantity $r_j^T \lambda$ is the total price paid by flow j for using the network. The term inside the parentheses is closely related to a special objective in convex optimization called the *Fenchel conjugate function*. For some function f, the Fenchel conjugate is defined as

$$f^*(y) = \sup_x \left(y^T x - f(x)\right).$$

This function has a number of nice properties (see [BV04, §3.3]). Importantly, it has a closed form expression for many common functions. We can rewrite the dual function as

$$-\lambda^T c - \sum_{j=1}^{n} \ell_j^*(-r_j^T \lambda).$$

The dual problem is then

$$\begin{array}{ll} \text{maximize} & -\lambda^T c - \sum_{j=1}^{n} \ell_j^*(-r_j^T \lambda) \\ \text{subject to} & \lambda \geq 0. \end{array}$$

It turns out that this problem has a special structure that leads to both a nice interpretation and a parallelizable algorithm.

Solving the dual problem. We will solve the dual problem by running (projected) gradient ascent.[3] That is, after every gradient step on the variable λ, we will project it back onto the feasible region, $\lambda \geq 0$. The gradient of the dual function is

$$\nabla g(\lambda) = -c + R\bar{x},$$

where

$$\bar{x}_j = \text{argmin} \left(\ell_j(x_j) + (r_j^T \lambda)x_j\right)$$

[3]An easy extension is to consider another method that only requires gradient evaluations such as L-BFGS-B, which is usually much faster practice.

In other words, x_j is the flow value that minimizes the loss incurred by this flow plus the cost given by the current link prices λ. The gradient update is

$$\lambda^{k+1} = \left(\lambda^k + \alpha\left(-c + R\bar{x}^k\right)\right)_+,$$

where α is the step size parameter or 'learning rate'. This update does what we expect. If we are over capacity on a link (*i.e.,* $c_i < (R\bar{x}^k)_i$), then we increase the price. If we are under capacity on a link, we decrease the price.

We can interpret this update as a decentralized algorithm. The network setting some prices λ^k then distributes these prices to all users. The the users then submit the flows $\bar{x}$ they would choose at these prices. The network then calculates the gradient, *i.e.,* the additional capacity needed assuming these flows, $R\bar{x}^k - c$, and updates the prices accordingly. This process repeats until convergence (which we know happens because the problem is convex). The network can run this process without knowing anything about the individual user loss functions! And the algorithm is entirely decentralized—most of the compute is done independently by users in parallel.[4]

It turns out that many network flow algorithms, including TCP, can be interpreted as this type of dual decomposition algorithm applied to the rate control problem with a particular loss function. This optimization framework provides a principled approach to designing these algorithms. Instead of focusing on a particular method and its properties, we can focus on the loss function. The method above then will converge to a solution that minimizes this loss function. For more, check out [LL99; Low03].

Negative prices. Consider the closely related problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n \ell_i(t_i) \\ \text{subject to} & Rx = t, \end{array}$$

with variables $t \in \mathbf{R}^m$ and $x \in \mathbf{R}^n$. Here, our objective is a function of the traffic instead of the rates. For example, the network may have some 'ideal rate' of traffic on each edge and deviation from this ideal rate is penalized. This loss function could look something like

$$\ell_i(t_i) = \begin{cases} (t_i - t_i^\star)^2 & t_i \neq t_i^\star \text{ and } 0 \leq t_i \\ +\infty & t_i < 0. \end{cases}$$

The equality constraint means that, in general, we may have negative dual prices ν_i. We can interpret a negative price as a subsidy; if the network wants a traffic $t_i^\star > \bar{t}_i$ on some link, it can subsidize users to put more traffic on that link. (Think about how this would enter into the individual problems solved by each user to get $\bar{x}_j$).

[4]Methods of this form are called *dual decomposition* algorithms. Note that while the final flows will be feasible in the limit, some care needs to be taken to ensure that the approximate solution is feasible.

References

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[LL99] Steven Low and David Lapsley. “Optimization flow control. I. Basic algorithm and convergence”. In: *IEEE/ACM Transactions on networking* 7.6 (1999), pp. 861–874.

[Low03] Steven Low. “A duality model of TCP and queue management algorithms”. In: *IEEE/ACM Transactions On Networking* 11.4 (2003), pp. 525–536.