CS 270 Combinatorial Algorithms and Data Structures, Spring 2015

Lecture 15: LP duality

We prove the strong duality theorem for linear programming and see some applications in combinatorial optimization and game theory.

Weak duality in linear programming

Let's start from linear programming. Consider the following linear program.

$$\max \ 2x_1 + 3x_2$$
$$4x_1 + 8x_2 \le 12$$
$$2x_1 + x_2 \le 3$$
$$3x_1 + 2x_2 \le 4$$
$$x_1, x_2 \ge 0$$

This is a maximization problem. To give a lower bound on the optimal value, we just need to give a feasible solution.

In this example, setting $x_1 = 1, x_2 = 1$ shows that the optimal value is at least 5.

Could you find a better solution? If not, how do we show that the optimal value is at most 5?

By looking at the first constraint, we notice that $2x_1 + 4x_2 \le 6$. Since $x_2 \ge 0$, this implies that $6 \ge 2x_1 + 3x_2$.

This shows that the optimal value is at most 6.

We can prove a better upper bound by a similar argument. Add the first two constraints and divide by three, we obtain that any feasible solution must satisfy $2x_1 + 3x_2 \le 5$. This shows the optimal value is 5.

Given a general linear program,

$$\begin{array}{ll} \max & \langle c, x \rangle \\ & \langle a_1, x \rangle \le b_1 \\ & \vdots \\ & \langle a_m, x \rangle \le b_m \\ & x \ge 0 \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \max & c^T x \\ & Ax \le b \\ & x \ge 0 \end{array} \quad \text{where } A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m,\ c, x \in \mathbb{R}^n.$$

We can also use the same strategy to give an upper bound on the optimal value.

Take some positive linear combination of the constraints. Say $\langle y_1 a_1 + y_2 a_2, x \rangle \le y_1 b_1 + y_2 b_2$, where y_1, y_2 positive is to guarantee the inequality holds in the right direction. If $y_1 a_1 + y_2 a_2$ "dominates" c, i.e. $y_1 a_1 + y_2 a_2 \ge c$, then we know that $\langle y_1 a_1 + y_2 a_2, x \rangle \ge \langle c, x \rangle$ since $x \ge 0$, and thus $y_1 b_1 + y_2 b_2$ is an upper bound on the optimal value.

How to find the best upper bound? This is itself a linear programming problem.

Associate a non-negative number y_i to each constraint above. The best (smallest) upper bound is:

$$\begin{aligned} &\min\ b^T y & & & &\min\ b^T y \\ &y_i a_i \geq c & &\longleftrightarrow & &y^T A \geq c \\ &y_i \geq 0 \quad \forall i & & & &y \geq 0 \end{aligned}$$

We call this pair of LP a primal-dual pair:

$$\text{(primal)} \quad \begin{aligned} &\max\ c^T x \\ &Ax \leq b \\ &x \geq 0 \end{aligned} \qquad\qquad \text{(dual)} \quad \begin{aligned} &\min\ b^T y \\ &y^T A \geq c \\ &y \geq 0 \end{aligned}$$

From our previous discussion, it is clear that the primal optimal value is upper bounded by the dual optimal value. This is called the weak duality theorem in linear programming.

<u>Weak duality theorem</u>. For each feasible solution x to the primal program and each feasible solution to the dual program, we have $c^T x \leq b^T y$.

<u>Complementary slackness conditions</u>

How to prove that a primal feasible solution x is an optimal solution?

It suffices if we could find a dual feasible solution y such that $c^T x = b^T y$, actually in this case we prove that both x is primal optimal and y is dual optimal.

When will $c^T x = b^T y$? Let's do the calculation for the weak duality theorem again.

We have $y^T b \geq y^T A x$ (taking a non-negative combination of the rows, by the primal constraints)

$\geq c^T x$ (the combination dominating the primal objective,

To have $y^T b = c^T x$, we must have both inequalities satisfied as equalities.

For the first inequality to hold as an equality, we should have $y_i > 0$ only if $\langle a_i, x \rangle = b_i$, i.e. we should use the i-th row only when the i-th constraint is tight.

For the second inequality to hold as an equality, we should have $\langle a^j, y \rangle = c_j$ where a^j denotes the j-th column of A whenever $x_j > 0$, i.e. if $x_j > 0$, then the coefficient in the combination should match exactly the coefficient in the objective function.

In fact, these are necessary and sufficient conditions for x, y to be optimal.

To summarize, we have the following optimality conditions.

Complementary slackness conditions. Let x be a primal feasible solution and y be a dual feasible solution.

Then x and y are optimal solutions if and only if they satisfy:

(primal conditions) if $x_j > 0$, then $\langle a^j, y \rangle = c_j$, where a^j is the j-th column of A.

(dual conditions) if $y_i > 0$, then $\langle a_i, x \rangle = b_i$, where a_i is the i-th row of A.

These complementary slackness conditions can guide us to search for optimal primal and dual solutions using a combinatorial algorithm, e.g. the Hungarian algorithm for weighted bipartite matching.

These are called primal-dual algorithms, using the primal-dual pair to guide us to search for optimal solutions without explicitly solving the LP (e.g. without the need to compute an optimal basic solution) and they lead to simple and efficient algorithms for combinatorial problems.

There are also "approximate complementary slackness conditions" to guide us to search for good approximate solutions. Some well-known examples are in designing approximation algorithms for facility location and network design problems. See the book by Williamson and Shmoys if you are interested.

Strong duality in linear programming

We have derived a necessary and sufficient condition for a pair of primal and dual solutions to be optimal.

But we have not proved that they must exist, and this is the content of the strong duality theorem, which says that the dual program always provides a tight upper bound for the primal program.

There are different ways to prove the strong duality theorem.

One is based on the simplex method, which is elementary and self-contained, but it is only specific to LP.

Instead, we will see an approach based on simple convex geometry, which can be applied to semidefinite programming as well (as we will cover SDP later in the course), but a slight disadvantage is that we need to assume some basic results in analysis and will not see all the details of the proof.

Separation theorem

To prove strong duality theorems, we need some basic results in convex analysis, showing that we can always find a separating hyperplane to separate a point not in a closed convex set from the convex set. (Recall that this is what we needed in the ellipsoid method.)

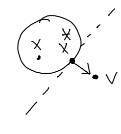
set. (Recall that this is what we needed in the ellipsoid method.)

Definition A set S is closed if every limit point of S is in S.

Definition A set S is convex if $x, z \in S$, then $\alpha x + (1-\alpha) z \in S$ for all $\alpha \in [0,1]$.

Theorem (separation theorem) Let $S \subseteq \mathbb{R}^n$ be a non-empty closed convex set and $v \notin S$.
Then there exists $y \in \mathbb{R}^n$ such that $\langle y, v \rangle > \langle y, x \rangle$ for all $x \in S$.

The proof plan is simple and intuitive. Given v, find the unique point $x^* \in S$ that is closest to v.
Then, argue that $\langle v - x^*, x - x^* \rangle \le 0$ for all $x \in S$, and this will give us the hyperplane with direction $v - x^*$.
Pictorically, the proof is summarized by the following picture:



Claim x^* is uniquely defined.

proof Let x be an arbitrary point in S. Consider $Z := \{ z \in S \mid \|z - v\|_2^2 \le \|x - v\|_2^2 \}$.
Then Z is bounded and closed, and thus compact.
A point closest to v is a minimizer of the continuous function $f(z) = \|z - v\|_2^2$ over compact set Z.
By Weierstrauss' theorem, the minimum value is attained by some point $x^* \in S$.
To see that x^* is uniquely defined, suppose by contradiction that $x_1 \neq x_2$ and both minimize the distance to v.
Consider $\bar{x} = \frac{1}{2}(x_1 + x_2)$. By convexity, $\bar{x} \in S$. Let $\mu = \|x_1 - v\|^2 = \|x_2 - v\|^2$ be the minimum value.
Check that $\|\bar{x} - v\|^2 = \frac{1}{2}\|x_1 - v\|^2 + \frac{1}{2}\|x_2 - v\|^2 - \frac{1}{4}\|x_1 - x_2\|^2 = \mu - \frac{1}{4}\|x_1 - x_2\|^2 < \mu$, a contradiction. □

Lemma x^* is the minimizer if and only if $\langle x - x^*, v - x^* \rangle \le 0$ for all $x \in S$.

proof Let $x \in S$. Consider $z = (1-\varepsilon)x^* + \varepsilon x$. By convexity, $z \in S$.

$$\|z - v\|^2 = \|(x^* - v) - \varepsilon(x^* - x)\|^2 = \|x^* - v\|^2 - 2\varepsilon\langle x^* - v, x^* - x\rangle + \varepsilon^2\|x^* - x\|^2.$$

Then x^* is the minimizer
$$\Leftrightarrow \quad -2\varepsilon\langle x^* - v, x^* - x\rangle + \varepsilon^2\|x^* - x\|^2 \ge 0 \quad \forall x\ \forall \varepsilon > 0$$
$$\Leftrightarrow \quad \langle x^* - v, x^* - x\rangle \le \frac{\varepsilon}{2}\|x^* - x\|^2 \quad \forall x\ \forall \varepsilon > 0$$
$$\Leftrightarrow \quad \langle x^* - v, x^* - x\rangle \le 0. \quad □$$

Proof of separation theorem Let $y = v - x^*$.

Then $\|v - x^*\|^2 > 0 \Rightarrow \langle v, v - x^* \rangle > \langle x^*, v - x^* \rangle \Rightarrow \langle v, y \rangle > \langle x^*, y \rangle$.

On the other hand, for $x \in S$, by the Lemma we have $\langle x - x^*, v - x^* \rangle \le 0 \Rightarrow \langle x, v - x^* \rangle \le \langle x^*, v - x^* \rangle$.

Therefore, $\langle v, y \rangle > \langle x, y \rangle$ for all $x \in S$. □

Farkas lemma and strong LP duality

Farkas lemma tells us exactly when a set of linear inequalities has no feasible solutions.

Theorem (Farkas lemma) The system $Ax=b$, $x \geq 0$ has no solution if and only if

$\exists y$ such that $y^T A \geq 0$ and $y^T b < 0$.

proof. Let $S = \{ Ax \mid x \geq 0 \}$. Then S is a closed convex set (see chapter 6 of Matousek-Gartner).

($\Rightarrow$) The system $Ax=b$, $x \geq 0$ has no solutions is equivalent to saying that $b \notin S$.

By the separation theorem, there exists y such that $\langle y, b \rangle < \langle y, z \rangle \quad \forall z \in S$,

which can be rewritten as $\langle y, b \rangle < \langle y, Ax \rangle \quad \forall x \geq 0$.

Since $0 \in S$, we have $y^T b < 0$.

Also, we must have $y^T A \geq 0$, as otherwise there exists $x \geq 0$ such that $\langle y, Ax \rangle = \langle y^T A, x \rangle = -\infty$,

contradicting that $\langle y, b \rangle < \langle y, Ax \rangle \quad \forall x \geq 0$.

($\Leftarrow$) If such a y exists, then the system must have no solution, as otherwise

$0 > y^T b = y^T A x \geq 0$, a contradiction. $\square$

Farkas lemma is often called a theorem of alternatives. Either the system has a solution or not, and in either case there is an easily verifiable condition for it. Now we are one step away from LP duality.

Theorem (strong LP-duality) If both the primal and dual are feasible and bounded, then they have the same objective value.

Proof of strong LP duality. Consider the primal linear program in equational form:

$$\text{(primal)} \quad \begin{array}{l} \max\ c^T x \\ Ax = b \\ x \geq 0 \end{array} \qquad\qquad \text{(dual)} \quad \begin{array}{l} \min\ b^T y \\ y^T A \geq c \end{array}$$

We would like to show that if the primal objective value is less than a value μ, then there is a dual feasible solution y with objective value less than μ, which would imply strong duality.

The primal objective value is less than μ is equivalent to saying that the following is infeasible:

$$\begin{array}{l} c^T x - s = \mu \\ Ax = b \\ x \geq 0,\ s \geq 0,\ s \in \mathbb{R} \end{array} \iff \begin{pmatrix} A & 0 \\ c^T & -1 \end{pmatrix} \begin{pmatrix} x \\ s \end{pmatrix} = \begin{pmatrix} b \\ \mu \end{pmatrix} \text{ has no solutions.} \quad x \geq 0,\ s \geq 0$$

By Farkas lemma, this is infeasible iff $\exists\, y \in \mathbb{R}^m$ (where $A \in \mathbb{R}^{m\times n}$) and $z \in \mathbb{R}$ such that

$$\begin{pmatrix} y^T & z \end{pmatrix} \begin{pmatrix} A & 0 \\ c^T & -1 \end{pmatrix} \geq 0 \qquad \begin{pmatrix} y^T & z \end{pmatrix} \begin{pmatrix} b \\ \mu \end{pmatrix} < 0.$$

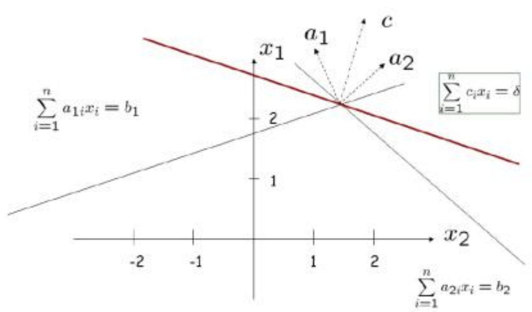
More compactly, $y^T A + z c^T \geq 0$, $-z \geq 0$, $b^T y + z\mu < 0$.

Consider two cases. Suppose $z = 0$. Then this implies that $y^T A \geq 0$ and $b^T y < 0$, which means that the primal linear program has no feasible solution at all.

Suppose $z \neq 0$. Then $z < 0$. We have $\left(\frac{y^T}{-z}\right) A \geq c^T$ and $\left(\frac{y^T}{-z}\right) b < \mu$.

Therefore, $-y^T/z$ is a dual feasible solution with objective value μ. $\square$

Geometric interpretation



If x is an optimal solution in the direction c, the objective function should be in the cone of the tight constraints, i.e. $\exists\, y \geq 0$ s.t. $y^T A = c$. Otherwise, we could find an improving solution in the direction c, and then x is not optimal.

Min-max theorems in combinatorial optimization

Min-max theorems are fundamental results in combinatorial optimization, as they give nice characterizations of the optimal value of the problem. Some examples include:

<u>Max-bipartite-matching min-vertex cover</u>: Given a bipartite graph, the maximum size of a matching is equal to the minimum size of a vertex cover. (Check that Hall's theorem is implied by it.)

<u>Max-flow min-cut</u>: Given a directed graph and two vertices s and t, the maximum value of an s-t flow is equal to the minimum value of an s-t cut.

We now see that these theorems can be proved systematically using LP duality.

<u>Bipartite matching</u>:

(primal) $\max \sum_e x_e$ (dual) $\min \sum_v y_v$

(primal) $\max \sum_e x_e$

$x(\delta(v)) \le 1 \quad \forall v \in V$

$x_e \ge 0 \quad \forall e \in E$

$\leftrightarrow$

(dual) $\min \sum_v y_v$

$y_u + y_v \ge 1 \quad \forall e = uv \in E$

$y_v \ge 0 \quad \forall v \in V.$

Note that any integral solution of the dual program corresponds to a vertex cover.

It can be shown that the dual LP has integral vertex solutions (exercise).

So, by the strong duality theorem, the max-bipartite-matching min-vertex-cover result follows.

Maximum flow

(primal) $\max x_{ts}$

$y_v \qquad \sum_{e \in \delta^{in}(v)} x_e - \sum_{e \in \delta^{out}(v)} x_e \le 0 \quad \forall v \in V$

$d_e \qquad x_e \le c_e \quad \forall e \in E$

$x_e \ge 0 \quad \forall e \in E$

$\leftrightarrow$

(dual) $\min \sum_e c_e d_e$

$d_{uv} + y_v - y_u \ge 0 \quad \forall uv \in E$

$y_s - y_t \ge 1$

$y_v \ge 0 \quad \forall v \in V$

$d_e \ge 0 \quad \forall e \in E$

To understand the dual program, first observe that $d_{uv} = y_u - y_v$ in any optimal solution, as $c_e \ge 0$.

Also, we can assume that $y_s = 1$ and $y_t = 0$, and $1 \ge y_v \ge 0$ for all $v \in V$.

Suppose $y_v \in \{0,1\}$ for all v, then it corresponds to the cut where all vertices with $y_v = 1$ are on the source side, while all vertices with $y_v = 0$ are on the sink side, and the objective function counts the total weight of the edges in this s-t cut.

Again, it can be proved that both the primal (HW1) and the dual (HW2?) are integral, and thus the max-flow min-cut theorem follows from the strong duality theorem.

Remark: There are different ways to show that a certain LP is integral. One well-known method is to show that the constraint matrix is "totally unimodular".

Minimax theorem in game theory

There are also various uses of the strong duality theorem in game theory.

The most fundamental one is the following minimax theorem in two player zero-sum game.

Two player zero-sum game can be described by a matrix, where each row corresponds to a strategy of the row player, and each column corresponds to a strategy of the column player.

If the row player chooses strategy i and the column player chooses strategy j, then the payoff for the players is the (i,j)-th entry of the matrix.

The row player's goal is to maximize the payoff while the column player's goal is to minimize it.

	P	S	R
P	0	-1	1
S	1	0	-1
R	-1	1	0

e.g. this is the payoff matrix of the paper-scissor-rock game

A Nash equilibrium is a pair of strategies of the players, such that even if a player knows the strategy of the other player, he/she can not gain by changing his/her own strategy.

It should be clear that there is no pure strategy solution in the paper-scissor-rock game, but the mixed strategy $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is an equilibrium solution for both players.

We now show that any two player zero-sum game has a mixed strategy equilibrium solution.

Let $A \in \mathbb{R}^{m \times n}$ be the payoff matrix.

Let $x = (x_1, \dots, x_m)$ be the probability distribution of the row strategies, s.t. $\sum_{i=1}^m x_i = 1$, $x_i \geq 0 \ \forall i$.

We denote it by $x \in \Delta^m$. Similarly, the column mixed strategy is denoted by $y \in \Delta^n$.

If the row player plays x and the column player plays y, then the payoff is simply $x^T A y$.

The minimax theorem says that if both players play optimally, then it doesn't matter who announce his/her (mixed) strategy first, and thus these form an equilibrium solution.

<u>Theorem</u> (minimax theorem) $\displaystyle \max_{x \in \Delta^m} \min_{y \in \Delta^n} x^T A y = \min_{y \in \Delta^n} \max_{x \in \Delta^m} x^T A y$.

(On LHS, the row player announces first, while on RHS, the column player announces first.)

<u>Proof</u> First, observe that once the row player fixes a strategy x, then the column player can compute $x^T A = (z_1, z_2, \dots, z_n)$ where z_i is the expected payoff if the column player plays the (pure) strategy i, and playing the strategy i with minimum z_i is a best response.

So, we can simplify $\displaystyle \max_{x \in \Delta^m} \min_{y \in \Delta^n} x^T A y$ as $\displaystyle \max_{x \in \Delta^m} \min_i (x^T A)_i$, which can be written as the following LP:

$$
\begin{aligned}
\max \quad & t \\
& \sum_{i=1}^m x_i a_{ij} \geq t && \forall\, 1 \leq j \leq n \\
& \sum_{i=1}^m x_i = 1 \\
& x_i \geq 0 && \forall\, 1 \leq i \leq m
\end{aligned}
$$

Similarly, $\min \max x^T A y$ can be simplified as $\min \max (Ay)_i$, and can be written as

$$x_i \geq 0 \qquad \forall\ 1 \leq i \leq m$$

Similarly, $\min_{y \in \Delta^m} \max_{x \in \Delta^m} x^T A y$ can be simplified as $\min_{y \in \Delta^m} \max_j (Ay)_j$, and can be written as the following LP:

$$\min \quad r$$
$$\sum_{j=1}^{n} a_{ij} y_j \leq r \qquad \forall\ 1 \leq i \leq m$$
$$\sum_{j=1}^{n} y_j = 1$$
$$y_j \geq 0 \qquad \forall\ 1 \leq j \leq n$$

Now, just verify that these two programs are a primal-dual pair.

Hence, the minimax theorem follows from the strong duality theorem for LP. □

Yao's minimax principle

Yao observed that the minimax principle can be used to prove lower bounds for randomized algorithms.

The worst case running time of a randomized algorithm is its running time on the worst input.

Suppose we want to prove a lower bound on the running time of any randomized algorithm to solve a problem.

Notice that a randomized algorithm is just a distribution of deterministic algorithm, i.e. when the random string is fixed, it behaves deterministically.

Think of it as a two player zero-sum game where there is an adversary who likes to play the worst distribution of inputs to maximize the running time, while the randomized algorithm player wants to play the best distribution of algorithms to minimize the running time.

Suppose the number of inputs is finite and the number of deterministic algorithm is finite.

Let $A \in \mathbb{R}^{m \times n}$ be the payoff matrix, where each row corresponds to an input, and each column corresponds to a deterministic algorithm. and the (i,j)-th input is the running time of algorithm j on input i.

Let $\vec{i} \in \Delta^m$ be a probability distribution of the inputs, and $\vec{r} \in \Delta^n$ be a probability distribution of the deterministic algorithms.

The complexity that we are interested in analyzing is $\min_{\vec{r} \in \Delta^n} \max_{\vec{i} \in \Delta^m} \vec{i} A \vec{r}$, the running time of the best randomized algorithm.

By the minimax theorem, it is equal to $\max_{\vec{i} \in \Delta^m} \min_{\vec{r} \in \Delta^n} \vec{i} A \vec{r}$.

So, to prove lower bound on the complexity of the best randomized algorithm, it is enough to come up with a distribution of the inputs and prove that any deterministic algorithm would take a long time. It is usually easier to reason about deterministic algorithms, and this approach is widely used.

Note that any distribution would give a lower bound (actually just the weak duality theorem), but the minimax theorem says that one could prove the optimal lower bound this way.

See chapter 2 of "randomized algorithms" for an example of proving lower bound this way.

Tolls for multicommodity flow

We sketch one more application with a game theory flavor.

Suppose there is a traffic network or a communication network.

There are n users in the network, where each user i wants to send d_i units of information from vertex s_i to vertex t_i.

If we let each user to choose their own paths to send the information, they may all send along the shortest paths and this may cause high congestions on some edges, and a bad utilization of the network.

Instead of directly controlling their behaviors, what you could do is to give prices on edges, and charge the users on the edges they used, and the hope is to avoid over congestion and lead to a better social welfare.

The problem can be formalized as follows.

Let P_i be the set of all possible s_i-t_i paths. For each path $p \in P_i$, we have a variable f_p^i.

Let c_e be the congestion of edge e that we like to enforce.

Let l_p^c be the latency of the path p, given the current congestion pattern.

Our objective is to minimize the total latency while enforcing the congestion:

$$\min \sum_{i=1}^{n} \sum_{p \in P_i} l_p^c \cdot f_p^i$$

$$\sum_{i=1}^{n} \sum_{p \in P_i : e \in p} f_p^i \le c_e \qquad \forall e \qquad \text{(congestion constraints)}$$

$$\sum_{p \in P_i} f_p^i = d_i \qquad \forall i \qquad \text{(information constraints)}$$

$$f_p^i \ge 0 \qquad \forall p \; \forall i$$

Consider the dual program.

$$\max \sum_i d_i z_i + \sum_e c_e t_e$$

$$z_i - \sum_{e : e \in p} t_e \le l_p^c \qquad \forall i \; \forall p \in P_i$$

$$t_e \ge 0 \qquad \forall e \in E$$

Now, if we set the price on each edge e to be t_e, and a user needs to pay $\sum_{e \in p} t_e$ dollars for sending one unit of information along path p.

Then, we can direct the users to follow the optimal primal solution that minimizes congestion and the total latency.

The reason is that the paths with $f_p^i > 0$ would have $z_i = l_p^c + \sum_{e: e \in p} t_e$ by the complementary slackness condition, while all other paths p would have $z_i \le l_p^c + \sum_{e: e \in p} t_e$.

So, the paths with $f_p^i > 0$ are the paths that the selfish users would have no incentive to switch, and so this is an equilibrium solution (and under some conditions this is the unique equilibrium solution).

To conclude, by setting prices on edges, the network administrator can set the global optimal solution to be the (unique) equilibrium solution, so that the selfish users will be directed to this solution.

<u>References and pointers</u>

See chapter 6 of "understanding and using linear programming" for more information about LP duality.

See chapter 7 of "the design of approximation algorithms" for primal-dual approximation algorithms.

See "combinatorial optimization: polyhedra and efficiency" for integrality proofs of LP relaxations.

See chapter 2 of "randomized algorithms" for a lower bound proof using Yao's minimax principle.

See the paper "tolls for heterogenous selfish users in multicommodity networks ..." by Fleischer, Jain, Mahdian for more information about the last part.

There are some very interesting algorithmic results using the primal-dual method to design "online" algorithms. The survey "the design of competitive online algorithms via a primal-dual approach" by Buchbinder and Naor, and the paper "randomized primal-dual analysis of RANKING for online bipartite matching" are highly recommended.