# Applications of Optimality Condition and Duality Theory II

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

## Motivation Example: World Cup Information/Prediction Market

A market maker is organizing a combinatorial auction market to predict who would win the coming World-Cup game:

| Order/Bidder: | $\#1$ | $\#2$ | $\#3$ | $\#4$ | $\#5$ |
|---|---|---|---|---|---|
| Argentina | $1$ | $0$ | $1$ | $1$ | $0$ |
| Brazil | $1$ | $0$ | $0$ | $1$ | $1$ |
| Italy | $1$ | $0$ | $1$ | $1$ | $0$ |
| Germany | $0$ | $1$ | $0$ | $1$ | $1$ |
| France | $0$ | $0$ | $1$ | $0$ | $0$ |
| Bidding Price:$\pi$ | $0.75$ | $0.35$ | $0.4$ | $0.95$ | $0.75$ |
| Quantity limit:$\mathbf{q}$ | $10$ | $5$ | $10$ | $10$ | $5$ |
| Order fill:$\mathbf{x}$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

He or she would decide how many shares to sell to each bidder, knowing that one dollar/per share would be paid to the bidder whose selection includes the final winning team.

## Combinatorial Auction Market I: Abstraction

Given $m$ different states that are mutually exclusive and exactly one of them will be true at the maturity. A share contract of states is a paper agreement so that on maturity it is worth a notional $\$1$ if it includes the winning state and worth $\$0$ if is not on the winning state.

There are $n$ orders betting on one or a combination of states, with a bidding price and a quantity limit. Precisely, the $j$th order is given as $(\mathbf{a}_j \in R^m,\ \pi_j \in R_+,\ q_j \in R_+)$: $\mathbf{a}_j$ is the combination betting vector where each component is either $1$ or $0$

$$\mathbf{a}_j = (a_{1j};\ a_{2j}; ...;\ a_{mj})\ ,$$

where $1$ means the state is included and $0$ otherwise; $\pi_j$ is the bidding share price, and $q_j$ is the maximum number of shares the bidder like to buy.

## Combinatorial Auction Market II: Market Maker's Problem

Let $x_j$ be the number of shares awarded to the $j$th order. Then, the $j$th bidder will pay the amount

$$\pi_j \cdot x_j$$

and the total collected amount by the market maker is

$$\sum_{j=1}^{n} \pi_j \cdot x_j = \pi^T \mathbf{x}$$

If the $i$th state becomes the winning state, then the market maker needs to pay back

$$\left( \sum_{j=1}^{n} a_{ij} x_j \right) = \mathbf{a}_i \mathbf{x}.$$

The question is, how do the market maker decide $\mathbf{x} \in R^n$.

## Combinatorial Auction Market III: Worst-Case Profit Maximization

$$
\begin{array}{lrl}
\max & \pi^T \mathbf{x} - \max_i\{\mathbf{a}_i \mathbf{x}\} & \\
\text{s.t.} & \mathbf{x} & \le \mathbf{q}, \\
 & \mathbf{x} & \ge 0;
\end{array}
$$

or

$$
\begin{array}{lrl}
\max & \pi^T \mathbf{x} - z & \\
\text{s.t.} & A\mathbf{x} - \mathbf{e} \cdot z & \le \mathbf{0}, \\
 & \mathbf{x} & \le \mathbf{q}, \\
 & \mathbf{x} & \ge 0.
\end{array}
$$

$\pi^T \mathbf{x}$: the amount can be collected by the maker maker; $z$: the worst-case cost to the market maker; so that $\pi^T \mathbf{x} - z$ is the worst-case profit to the market maker.

## Combinatorial Auction Market IV: The Dual of the LP Problem

Let $A\mathbf{x} - \mathbf{e} \cdot z \leq \mathbf{0}$ be associated with dual/Lagrange variables $\mathbf{0} \leq \mathbf{p} \in R^m$ and $\mathbf{x} \leq \mathbf{q}$ be associated with dual/Lagrange variables $\mathbf{0} \leq \mathbf{s} \in R^n$. Then the dual would be:

$$
\begin{array}{rrl}
\min & \mathbf{q}^T\mathbf{s} & \\
\text{s.t.} & A^T\mathbf{p} + \mathbf{s} & \geq \pi, \\
& \mathbf{e}^T\mathbf{p} & = 1, \\
& (\mathbf{p}, \mathbf{s}) & \geq 0.
\end{array}
$$

In LP problems, $\mathbf{p}$ is interpreted as the state implicit/shadow prices (priced to each state); and $\mathbf{s}$ is interpreted as the order implicit/shadow prices (how much profit/per share the market maker can make from each bidder).

# World Cup Information Market Result

| Order: | $\#1$ | $\#2$ | $\#3$ | $\#4$ | $\#5$ | State Price |
|---|---|---|---|---|---|---|
| Argentina | $1$ | $0$ | $1$ | $1$ | $0$ | $0.2$ |
| Brazil | $1$ | $0$ | $0$ | $1$ | $1$ | $0.35$ |
| Italy | $1$ | $0$ | $1$ | $1$ | $0$ | $0.2$ |
| Germany | $0$ | $1$ | $0$ | $1$ | $1$ | $0.25$ |
| France | $0$ | $0$ | $1$ | $0$ | $0$ | $0$ |
| Bidding Price:$\pi$ | $0.75$ | $0.35$ | $0.4$ | $0.95$ | $0.75$ | |
| Quantity limit:$\mathbf{q}$ | $10$ | $5$ | $10$ | $10$ | $5$ | |
| Order fill:$\mathbf{x}^*$ | $5$ | $5$ | $5$ | $0$ | $5$ | |

## Combinatorial Auction Market V: Optimality Condition Properties

| | |
|---|---|
| $x_j^* > 0$ | $\mathbf{a}_j^T\mathbf{p} + s_j = \pi_j$ and $s_j \geq 0$ so that $\mathbf{a}_j^T\mathbf{p} \leq \pi_j$ |
| $0 < x_j^* < q_j$ | $s_j = 0$ so that $\mathbf{a}_j^T\mathbf{p} = \pi_j$ |
| $x_j^* = q_j$ | $s_j > 0$ so that $\mathbf{a}_j^T\mathbf{p} < \pi_j$ |
| $x_j^* = 0$ | $\mathbf{a}_j^T\mathbf{p} + s_j > \pi_j$ and $s_j = 0$ so that $\mathbf{a}_j^T\mathbf{p} > \pi_j$ |

The market Fair: if a lower bid wins the auction, so does the higher bid on any same type of bids.

Question 1: Unique State Prices?

## Combinatorial Auction Market VI: Convex Programming Model

$$
\begin{array}{lrl}
\max & \pi^T \mathbf{x} - z + u(\mathbf{s}) & \\
\text{s.t.} & A\mathbf{x} - \mathbf{e} \cdot z + \mathbf{s} & = \mathbf{0}, \\
 & \mathbf{x} & \leq \mathbf{q}, \\
 & \mathbf{x},\ \mathbf{s} & \geq 0.
\end{array}
$$

$u(\mathbf{s})$: a value function for the market market on possible slack shares.

If $u(\cdot)$ is a strictly concave function, then the state price vector is unique.

Question 2: Online Auction?

## Online Combinatorial Auction Market

Suppose the $k$th order/bidder just arrived...

Approach 1 (SCPM):

$$\begin{array}{ll} \text{maximize}_{x_k,\mathbf{s}} & \pi_k x_k - z + u(\mathbf{s}) \\ \text{s.t.} & a_{ik}x_k - z + s_i = -\sum_{j=1}^{k-1} a_{ij}\bar{x}_j,\ \forall\, i = 1, 2, ..., m, \\ & 0 \le x_k \le q_k, \end{array}$$

where $\bar{x}_j,\ j = 1, 2, ..., k-1$ are the decisions made in previous steps.

Approach 2 (SLPM):

$$\begin{array}{ll} \text{maximize}_{x_1,\ldots,x_k} & \sum_{j=1}^{k} \pi_j x_j - z \\ \text{s.t.} & \sum_{j=1}^{k} a_{ij}x_j - z \le 0,\ \forall\, i = 1, 2, ..., m, \\ & 0 \le x_j \le q_j,\ \forall\, j = 1, ..., k. \end{array}$$

and use the state shadow prices of the LP to make the decisions for the next (few) bidders.

## Motivation Example: Sparse Portfolio Selection and Stock Tracking

Recall the modern portfolio selection problem:

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^T V \mathbf{x} \\ \text{subject to} & \mathbf{r}^T \mathbf{x} \geq \mu, \\ & \mathbf{e}^T \mathbf{x} = 1, \ \mathbf{x} \ \geq \ \mathbf{0}, \end{array}$$

where expect-value vector $\mathbf{r}$ and co-variance matrix $V$ are given, and $\mathbf{e}$ is the vector of all ones.

In shorting-allowed models, constraint $\mathbf{x} \ \geq \ \mathbf{0}$ is dropped. Also, frequently, the risk and expected return are aggregated into the objective function:

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^T V \mathbf{x} - w \mathbf{r}^T \mathbf{x} \\ \text{subject to} & \mathbf{e}^T \mathbf{x} = 1. \end{array}$$

But the final solution of the model are typically dense... One also like to track the market performance using fewer stocks than S&P500...

## Sparse-Regression I: Cardinality Constrained Regression

Consider the problem:

$$\text{Minimize}_x \quad f_p(\mathbf{x}) := \|A\mathbf{x} - \mathbf{b}\|_2^2, \quad \text{s.t. } \|\mathbf{x}\|_0 \le k,$$

where data $A \in R^{m\times n}, \mathbf{b} \in R^m$, $k$ is a given limit on the number non-zeros in the final solution, and

$$\|\mathbf{x}\|_0 := |\{j : \; x_j \neq 0\}|.$$

Often, the problem can be reformulated as a regularized problem:

$$\text{Minimize}_x \quad f_0(\mathbf{x}) := \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_0. \tag{1}$$

In general, for a $0 \le p < 1$.

$$\|\mathbf{x}\|_p = (\sum_{1\le j\le n} |x_j|^p)^{1/p}$$

with a $0 < p < 1$ is called quasi-norm of vector $\mathbf{x}$.

## Sparse-Regression II: Quasi-Norm Regularized Regression

Since $\|\mathbf{x}\|_0$ is not a continuous function, often it is approximated by (typically with $p = 1/2$):

$$\text{Minimize}_x \quad f_0(\mathbf{x}) := \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_p^p. \tag{2}$$

and it worked in many applications such as Sparse image reconstruction, Sparse signal recovering, Compressed sensing, etc.

Although non-convex, the classical KKT optimality conditions can be applied in analyses.

## Theory of $L_p$ Regularized Regression I

**Theorem 1** *(The first order bound) Let $\mathbf{x}^*$ be any first-order KKT solution of (2) and*

$$\ell_j = \left(\frac{\lambda p}{2\|\mathbf{a}_j\|\sqrt{f_p(\mathbf{x}^*)}}\right)^{\frac{1}{1-p}},$$

*where $\mathbf{a}_j$ is the $j$th column of $A$. Then, the following property holds:*

$$\text{for each } j, \qquad x_j^* \in (-\ell_j, \ell_j) \quad \Rightarrow \quad x_j^* = 0.$$

*Moreover, the number of nonzero entries in $\mathbf{x}^*$ is bounded by*

$$\|\mathbf{x}^*\|_0 \le \min\left(m, \frac{f_p(\mathbf{x}^*)}{\lambda \ell^p}\right);$$

*where $\ell = \min\{\ell_j\}$.*

## Sketch of Proof

Let $\mathbf{x}^*$ be a first-order KKT solution. Then it remains an KKT solution after eliminating those variables whose values are zeros. That is, for the nonzero-value variables, they must still satisfy the first-order KKT conditions:

$$2\mathbf{a}_j^T(A\mathbf{x}^* - \mathbf{b}) + \lambda p(|x_j^*|^{p-1} \cdot \mathrm{sign}(x_j^*)) = 0.$$

Thus,

$$|x_j^*|^{1-p} \geq \frac{\lambda p}{2\|\mathbf{a}_j\|\|A\mathbf{x}^* - \mathbf{b}\|} \geq \frac{\lambda p}{2\|\mathbf{a}_j\|\sqrt{f_p(\mathbf{x}^*)}}.$$

Now we show the second part of the theorem. Again,

$$\lambda\|\mathbf{x}^*\|_p^p \leq \|A\mathbf{x}^* - \mathbf{b}\| + \lambda\|\mathbf{x}^*\|_p^p = f_p(\mathbf{x}^*).$$

From the first part of this theorem, any nonzero entry of $\mathbf{x}^*$ is bounded from below by $\ell$ so that we have the desired result.

## Theory of $L_p$ Regularized Regression II

**Theorem 2** *(The second order bound) Let* $\mathbf{x}^*$ *be any second-order KKT solution of (2), and* $\kappa_j = \left(\frac{\lambda p(1-p)}{2\|\mathbf{a}_j\|^2}\right)^{\frac{1}{2-p}}$, $j \in \mathcal{N}$. *Then the following property holds:*

$$\text{for each } j, \qquad x_j^* \in (-\kappa_j, \kappa_j) \quad \Rightarrow \quad x_j^* = 0.$$

Again, we remove zero-value variables from $\mathbf{x}^*$ and the remain variables must still satisfy the second-order KKT condition for (2):

$$\nabla^2 f_p(\mathbf{x}) = 2A^TA - \lambda p(1-p)\text{Diag}(|x_j^*|^{p-2}) \succeq \mathbf{0}.$$

Then all diagonal entries of the Hessian must be nonnegative, which gives the proof.

## Theory of $L_p$ Regularized Regression III

- The first-order theorem indicates that the lower the objective value, the sparser the solution cardinality bound. Also, for $\lambda$ sufficiently large but finite, the number of nonzero entries in any KKT solution reduces to $0$.
- The result of the second-order theorem depends only on $\lambda$ and $p$. In practice, one would typically choose $p = 1/2$.
- The two theorems establish relations between model parameters $p,\ \lambda$ and the desired degree of sparsity of the solution. In particular, it gives a guidance on how to choose the combination of $\lambda$ and $p$.
- Later, we would show that a second-order KKT solution of (2) would be relatively easy to compute, either in theory or practice.

## The $L_p$ Regularized Sparse Portfolio Selection

$$\begin{array}{ll} \text{minimize} & \mathbf{x}^T V \mathbf{x} - w\mathbf{r}^T\mathbf{x} + \lambda\|\mathbf{x}\|_p \\ \text{subject to} & \mathbf{e}^T\mathbf{x} = 1. \end{array}$$

**Theorem 3** *(The second order theorem) Let $\mathbf{x}^* \in R^K$ be any second-order KKT solution (after removing zero-value entries) and $V^*$ be the corresponding covariance sub-matrix. Furthermore, let*

$$\kappa_j = V^*_{jj} - \frac{2}{K}(V^*\mathbf{e})_j + \frac{1}{K^2}(\mathbf{e}^T V^*\mathbf{e}),\ j \in P^*,$$

*which are the diagonal entries of matrix $\left(1 - \frac{1}{K}\mathbf{e}\mathbf{e}^T\right) V^* \left(1 - \frac{1}{K}\mathbf{e}\mathbf{e}^T\right)$. Then:*

- $(K-1)K^{3/2} \le \frac{4}{\lambda}\sum_j \kappa_j$.
- *If there is $\kappa_j = 0$, then $K = 1$ and $x^*_j = 1$; otherwise,*

$$x^*_j \ge \left(\frac{\lambda(1-\frac{1}{K})^2}{4\kappa_j}\right)^{2/3}.$$

## Proof

In the proof, we only consider variables $j \in P^*$. The second-order condition requires that the Hessian of the Lagrangian function

$$V^* - \frac{\lambda}{4}\text{Diag}\left[(x_j^*)^{-3/2}\right]$$

must be positive semidefinite in the null space of $\mathbf{e} \in R^K$. Or, the projected Hessian matrix

$$\left(I - \frac{1}{K}\mathbf{e}\mathbf{e}^T\right)\left(V^* - \frac{\lambda}{4}\text{Diag}\left[(x_j^*)^{-3/2}\right]\right)\left(I - \frac{1}{K}ee^T\right) \succeq \mathbf{0},$$

must be positive semidefinite.

Thus, the $j$th diagonal entry of the projected Hessian matrix

$$\kappa_j - \frac{\lambda}{4}\left((x_j^*)^{-3/2}\left(1 - \frac{2}{K}\right) + \frac{\sum_k (x_k^*)^{-3/2}}{K^2}\right) \geq 0, \tag{3}$$

and the trace of projected Hessian matrix

$$\sum_k \kappa_k - \frac{\lambda}{4}\frac{K-1}{K}\sum_k (x_k^*)^{-3/2} \geq 0.$$

The quantity $\sum_k (x_k^*)^{-3/2}$, with $\sum_k x_k^* = 1,\ x_k^* \geq 0$ achieves its minimum at $x_k^* = 1/K$ for all $k$ with the minimum value $K \cdot K^{3/2}$. Thus,

$$\frac{\lambda}{4}(K-1)K^{3/2} \leq \sum_k \kappa_k,$$

or

$$(K-1)K^{3/2} \leq \frac{4\sum_k \kappa_k}{\lambda},$$

which complete the proof of the first item.

Again, from (3) we have

$$\frac{\lambda}{4}\left((x_j^*)^{-3/2}\left(1-\frac{2}{K}\right)+\frac{\sum_k (x_k^*)^{-3/2}}{K^2}\right) \leq \kappa_j.$$

Or

$$\frac{\lambda}{4}\left((x_j^*)^{-3/2}\left(1-\frac{1}{K}\right)^2+\frac{\sum_{k,k\neq j} (x_k^*)^{-3/2}}{K^2}\right) \leq \kappa_j,$$

which implies

$$\frac{\lambda}{4}(x_j^*)^{-3/2}\left(1-\frac{1}{K}\right)^2 \leq \kappa_j.$$

Hence, if any $\kappa_j = 0$, we must have $K = 1$ and $x_j^*$ is the only non-zero entry in $\mathbf{x}^*$ so that $x_j^* = 1$. Otherwise, we have the desired second statement in the Theorem.