Optimization

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

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

What's New in MS&E 311?

- Provide mathematical proofs and in-depth theoretical analyses of optimization models/algorithms discussed in MS&E211
- Study additional conic-linear and nonlinear optimization models/problems comparing to MS&E310.
- Introduce/reintroduce effective optimization algorithms besides those discussed on the two courses.

Mathematical Optimization

The field of optimization is concerned with the study of maximization and minimization of mathematical functions. Very often the arguments of (i.e., variables or unknowns in) these functions are subject to side conditions or constraints. By virtue of its great utility in such diverse areas as applied science, engineering, economics, finance, medicine, and statistics, optimization holds an important place in the practical world and the scientific world. Indeed, as far back as the Eighteenth Century, the famous Swiss mathematician and physicist Leonhard Euler (1707-1783) proclaimed[a] that . . . nothing at all takes place in the Universe in which some rule of maximum or minimum does not appear.

---

[a]See Leonhardo Eulero, *Methodus Inviendi Lineas Curvas Maximi Minimive Proprietate Gaudentes*, Lausanne & Geneva, 1744, p. 245.

Mathematical Optimization/Programming (MP)

The class of mathematical optimization/programming problems considered in this course can all be expressed in the form

$$\begin{array}{lll} \text{(P)} & \text{minimize} & f(\mathbf{x}) \\ & \text{subject to} & \mathbf{x} \in \mathcal{X} \end{array}$$

where $\mathcal{X}$ usually specified by constraints:

$$\begin{array}{rcll} c_i(\mathbf{x}) & = & 0 & i \in \mathcal{E} \\ c_i(\mathbf{x}) & \leq & 0 & i \in \mathcal{I}. \end{array}$$

Global and Local Optimizers

A global minimizer for (P) is a vector $\mathbf{x}^*$ such that

$$\mathbf{x}^* \in \mathcal{X} \quad \text{and} \quad f(\mathbf{x}^*) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}.$$

Sometimes one has to settle for a local minimizer, that is, a vector $\bar{\mathbf{x}}$ such that

$$\bar{\mathbf{x}} \in \mathcal{X} \quad \text{and} \quad f(\bar{\mathbf{x}}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \cap N(\bar{x})$$

where $N(\bar{\mathbf{x}})$ is a neighborhood of $\bar{\mathbf{x}}$. Typically, $N(\bar{\mathbf{x}}) = B_\delta(\bar{\mathbf{x}})$, an open ball centered at $\bar{\mathbf{x}}$ having suitably small radius $\delta > 0$.

The value of the objective function f at a global minimizer or a local minimizer is also of interest. We call it the global minimum value or a local minimum value, respectively.

Important Terms

- decision variable/activity, data/parameter
- objective/goal/target
- constraint/limitation/requirement
- satisfied/violated
- feasible/allowable solutions
- optimal (feasible) solutions
- optimal value

Size and Complexity of Problems

- number of decision variables
- number of constraints
- bit number to store the problem input data
- problem difficulty or complexity
- algorithm complexity

Model Classifications

Optimization problems are generally divided into Unconstrained, Linear and Nonlinear Programming based upon the objective and constraints of the problem

- Unconstrained Optimization: Ω is the entire space R^n
- Linear Optimization: If both the objective and the constraint functions are linear
- Linearly Constrained Optimization: If the constraints are linear functions
- Quadratically Constrained Quadratic Optimization: If both the objective and the constraint functions are quadratic
- Nonlinear Optimization: If the constraints contain general nonlinear functions
- There are integer program, mixed-integer program etc.

Special Case: Linear Programming

$$
\begin{array}{rl}
\text{min(or max)imize} & c_1x_1 + c_2x_2 + ... + c_nx_n \\
\text{subject to} & a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \; \{\le, =, \ge\} \; b_1, \\
& a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \; \{\le, =, \ge\} \; b_2, \\
& \qquad ..., \\
& a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \; \{\le, =, \ge\} \; b_m, \\
& x_j \; \{\ge, \le\} \; u_j, \quad j = 1, ..., n,
\end{array}
$$

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

$$
\begin{aligned}
\text{min(or max)imize} \quad & \mathbf{c}^T \mathbf{x} \\
\text{subject to} \quad & A\mathbf{x} \,\{\leq, =, \geq\}\, \mathbf{b}, \\
& \mathbf{x} \,\{\geq, \leq\}\, \mathbf{0}.
\end{aligned}
$$

LP Terminology

- solution (decision, point): any specification of values for all decision variables, regardless of whether it is a desirable or even allowable choice
- feasible solution: a solution for which all the constraints are satisfied
- feasible region (constraint set, feasible set): the collection of all feasible solution
- interior, boundary, extreme point (corner) or basic feasible solution
- objective function contour (iso-profit, iso-cost line)
- optimal solution (optimum): a feasible solution that has the most favorable value of the objective function
- optimal (objective) value: the value of the objective function evaluated at an optimal solution
- active constraint (binding constraint), inactive constraint, redundant constraint

Linearly Constrained Programs in Standard Form

Linear Programming (LP)

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

Linearly Constrained Optimization Problem (LCOP)

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Conic Linear Programming (CLP)

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \in K. \end{array}$$

Quadratically Constrained Quadratic Programming (QCQP)

$$\begin{array}{ll} \text{minimize} & q_0(\mathbf{x}) \\ \text{subject to} & q_i(\mathbf{x}) \leq 0, \; \forall i = 1, ..., m \end{array}$$

where

$$q_i(\mathbf{x}) = \mathbf{x}^T Q_i \mathbf{x} + \mathbf{c}_i^T \mathbf{x}.$$

LP and LCOP Examles: Sparsest Data Fitting

We want to find a sparsest solution to fit exact data measurements, that is, to minimize the number of non-zero entries in $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$:

$$\begin{array}{rl} \text{minimize} & \|\mathbf{x}\|_0 = |\{j : x_j \neq 0\}| \\ \text{subject to} & A\mathbf{x} = \mathbf{b}. \end{array}$$

Sometimes this objective can be accomplished by

$$\begin{array}{rl} \text{minimize} & \|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j| \\ \text{subject to} & A\mathbf{x} = \mathbf{b}. \end{array}$$

Is this a linear program ?

Sparsest Data Fitting continued

It can be equivalently (?) represented by

$$\begin{array}{rl} \text{minimize} & \sum_{j=1}^n y_j \\ \text{subject to} & A\mathbf{x} = \mathbf{b},\ -\mathbf{y} \le \mathbf{x} \le \mathbf{y}; \end{array}$$

or

$$\begin{array}{rl} \text{minimize} & \sum_{j=1}^n (x_j' + x_j'') \\ \text{subject to} & A(\mathbf{x}' - \mathbf{x}'') = \mathbf{b},\ \mathbf{x}' \ge \mathbf{0},\ \mathbf{x}" \ge \mathbf{0}. \end{array}$$

Both are linear programs!

Sparsest Data Fitting continued

A better approximation of the objective can be accomplished by

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_p := \left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}; \end{array}$$

for some $0 < p < 1$. Or

$$\begin{array}{ll} \text{minimize} & \|\mathbf{x}\|_p^p = \sum_{j=1}^n |x_j|^p \\ \text{subject to} & A\mathbf{x} = \mathbf{b}. \end{array}$$

This is a linearly constrained optimization problem!

CLP Example: Facility Location

$\mathbf{c}_j$ is the location of client $j = 1, 2, ..., m$, and $\mathbf{y}$ is the location of a facility to be built.

$$\text{minimize} \quad \sum_j \|\mathbf{y} - \mathbf{c}_j\|_p.$$

Or equivalently (?)

$$\begin{array}{ll} \text{minimize} & \sum_j \delta_j \\ \text{subject to} & \mathbf{y} + \mathbf{x}_j = \mathbf{c}_j, \ \|\mathbf{x}_j\|_p \le \delta_j, \ \forall j. \end{array}$$

This is a p-order conic linear program for $p \ge 1$!

For simplicity, consider $m = 3$.

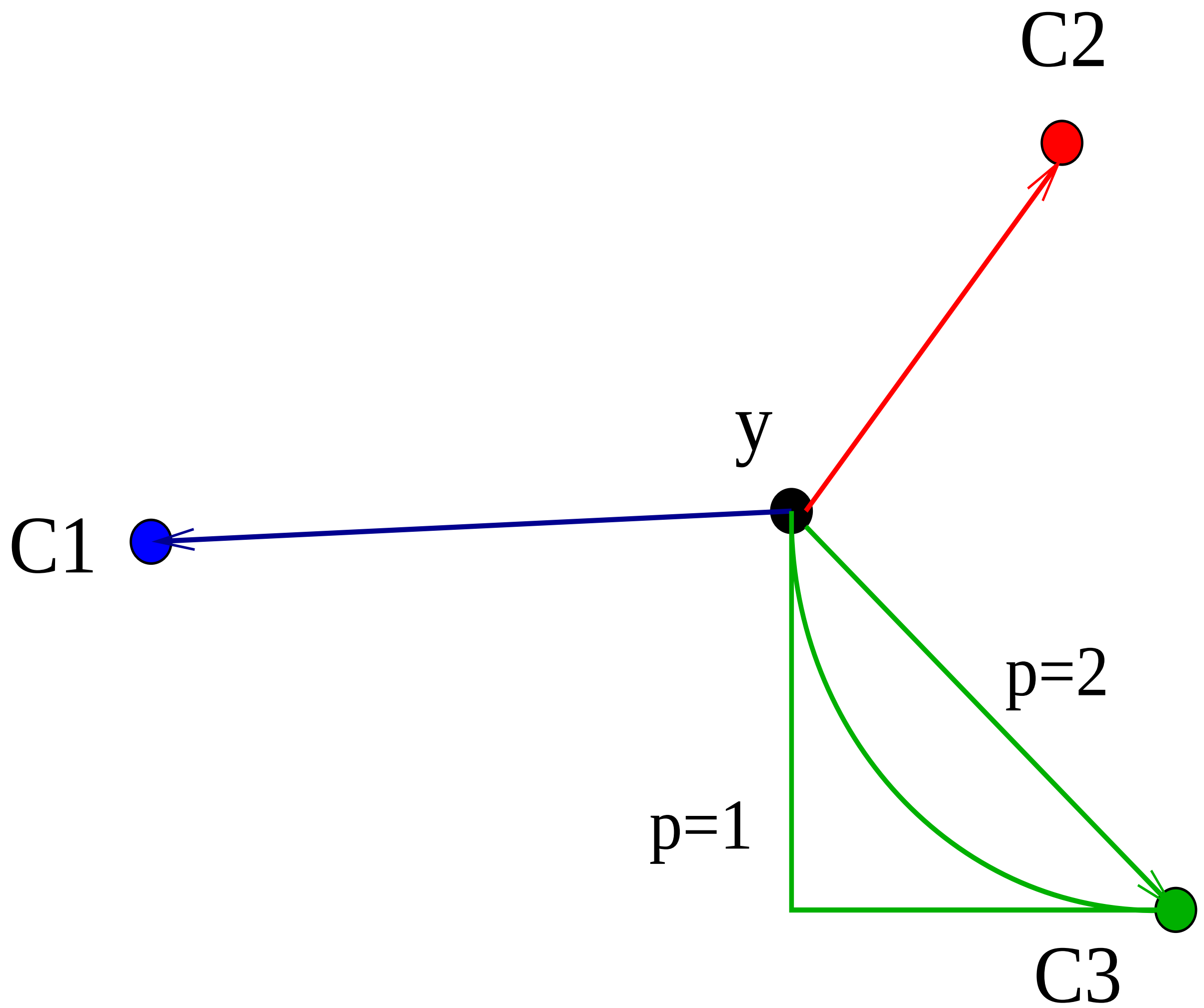

Figure 1: Facility Location at $\mathbf{y}$.

More LCOP and CLP Examples: Portfolio Management

For expected return vector $\mathbf{r}$ and co-variance matrix V of an investment portfolio, one management model is:

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

where $\mathbf{e}$ is the vector of all ones.

This is a convex quadratic program, a special case of LCOP.

QCQP Examples: Robust Portfolio Management

In applications, $\mathbf{r}$ and V may be estimated under various scenarios, say $\mathbf{r}_i$ and V_i for $i = 1, ..., m$. Then, we like

$$\begin{array}{rl} \text{minimize} & \max_i \mathbf{x}^T V_i \mathbf{x} \\ \text{subject to} & \min_i \mathbf{r}_i^T \mathbf{x} \geq \mu, \\ & \mathbf{e}^T \mathbf{x} = 1, \ \mathbf{x} \ \geq \ \mathbf{0}. \end{array}$$

$$\begin{array}{rl} \text{minimize} & \alpha \\ \text{subject to} & \mathbf{r}_i^T \mathbf{x} \geq \mu, \ \forall i \\ & \mathbf{x}^T V_i \mathbf{x} \leq \alpha, \ \forall i \\ & \mathbf{e}^T \mathbf{x} = 1, \ \mathbf{x} \ \geq \ \mathbf{0}. \end{array}$$

This is a quadratically constrained program.

Recall Supporting Vector Machine

Suppose we have two-class discrimination data. We assign the first class with 1 and the second with -1 for a binary varible. A powerful discrimination method is the Supporting Vector Machine (SVM).

Let the first class data points i be given by $\mathbf{a}_i \in R^d$, $i = 1, ..., n_1$ and the second class data points j be given by $\mathbf{b}_j \in R^d$, $j = 1, ..., n_2$.

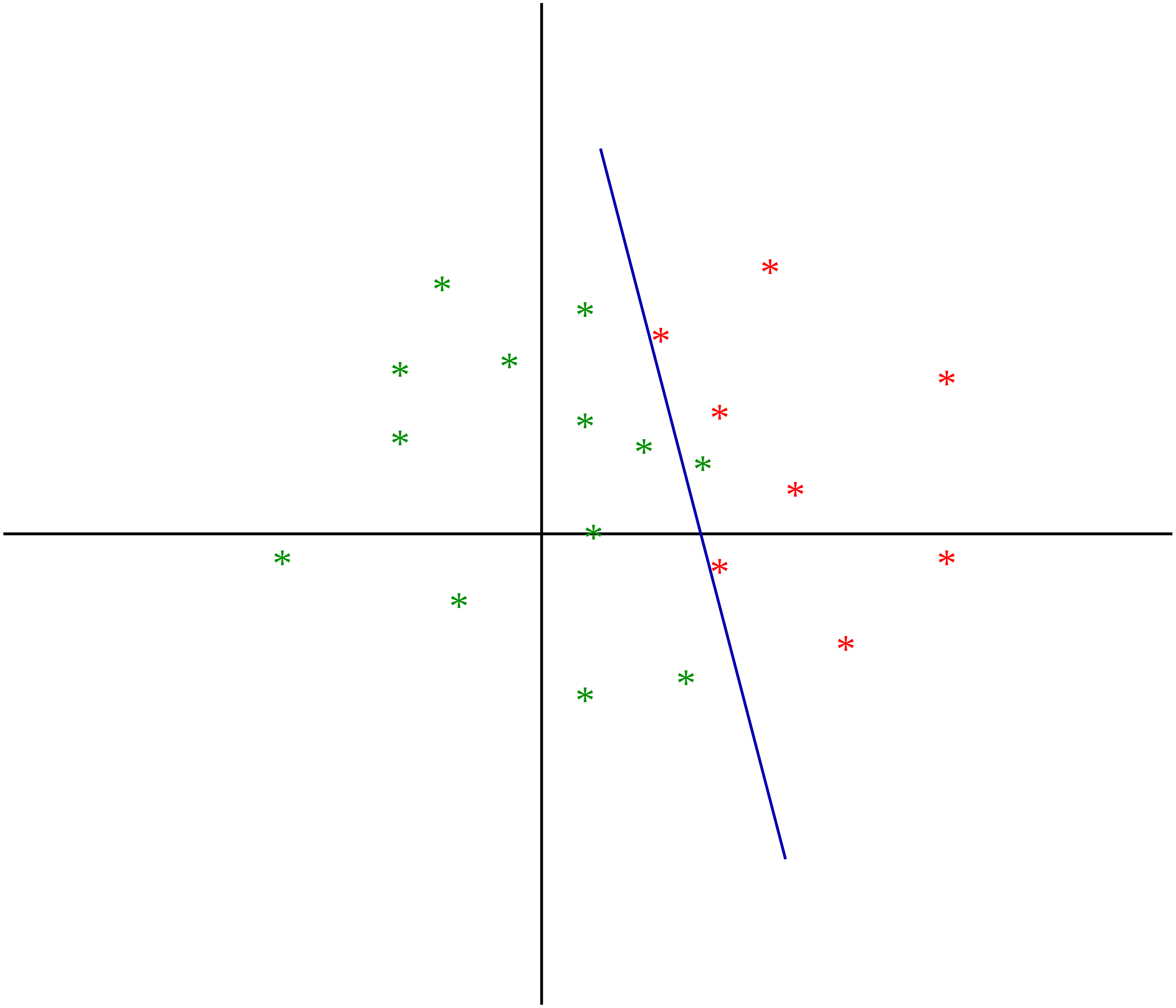

Figure 2: Linear Support Vector Machine

Supporting Vector Machine: ellipsoidal separation?

$$\begin{array}{ll} \text{minimize} & \text{trace}(\Omega) + \|\omega\|_2 \\ \text{subject to} & \mathbf{a}_i^T \Omega \mathbf{a}_i + \mathbf{a}_i^T \omega + \beta \geq 1, \ \forall i, \\ & \mathbf{b}_j^T \Omega \mathbf{b}_j + \mathbf{b}_j^T \omega + \beta \leq -1, \ \forall j, \\ & \Omega \succeq \mathbf{0}. \end{array}$$

This type of problems is semidefinite programming.

$$\begin{array}{ll} \text{minimize} & \text{trace}(\Omega) + \alpha + \mu \left(\sum_i e_i' + \sum_j e_j'' \right) \\ \text{subject to} & \mathbf{a}_i^T \Omega \mathbf{a}_i + \mathbf{a}_i^T \omega + \beta \geq 1 - e_i', \ \forall i, \\ & \mathbf{b}_j^T \Omega \mathbf{b}_j + \mathbf{b}_j^T \omega + \beta \leq -1 + e_j'', \ \forall j, \\ & e_i', \ e_j'' \geq 0, \ \forall i, j, \\ & \Omega \succeq \mathbf{0}, \ \|\omega\|_2 \leq \alpha. \end{array}$$

This problems is a mixed linear, SOCP and sDP program.

Figure 3: Quadratic Support Vector Machine