ISE 426
Optimization models and applications

Lecture 22 — November 20, 2014

- Nonlinear Programming (NLP)
- Least squares example
- Quadratic Programming

Reading:

- Winston&Venkataramanan, Ch. 12 up to §12.4; §12.10

Nonlinear Programming (NLP)

Consider the **continuous** case: no integer variables.

$$\begin{array}{ll} \min & f(x) \\ & g_i(x) \le 0 \quad \forall i = 1, 2 \ldots, m \\ & h_j(x) = 0 \quad \forall j = 1, 2 \ldots, p \\ & x \in \mathbb{R}^n \end{array}$$

This is convex when:

- $f(x)$ is convex
- all $g_i(x)$ are convex
- all $h_i(x)$ are convex and their opposite $-h(x)$ is convex too

$\Leftrightarrow$ they are linear (aka affine):

$$h_i(x) = a_i^\top x - b_i$$

Unconstrained Optimization Problem

What is the main good news about convex optimization?

$$\min f(x) \quad x \in R^n$$

If f is differentiable, then optimality condition:

$$\nabla f(x) = 0$$

If f is convex, then this condition is sufficient.

Optimization Problem

$$\min f(x)$$
$$g_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$h_j(x) = 0, \quad j = 1, \ldots, p$$

Lagrangian function

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Dual function

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

(λ, ν) is called dual feasible if $\lambda \geq 0$ and $g(\lambda, \nu) > -\infty$.

Dual Problem

Lagrangian function

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Dual function

$$d(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

Want to find (λ, ν) such that $\operatorname{argmin}_x L(x, \lambda, \nu)$ is the solution to the original problem.
Lower bound: $d(\lambda, \nu) \leq f(x)$ for any feasible (λ, ν) and x.
Dual problem

$$\begin{aligned} \max \quad & d(\lambda, \nu) \\ & \lambda \geq 0 \end{aligned}$$

Weak and Strong Duality

Let f^* be solution of the primal problem and d^* - solution of the dual problem.

- Weak duality: $d^* \leq f^*$, always.
- Strong duality: If primal is convex then (usually) $d^* = f^*$.

Optimality conditions

Lagrangian function

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

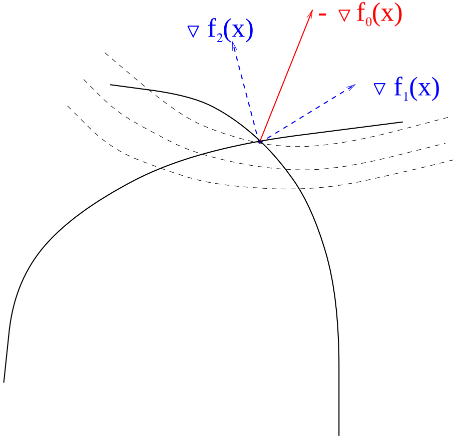
KKT (Karush-Kuhn-Tucker) optimality conditions: assume f_i and h_j are differentiable. If x^*, λ^*, ν^* are optimal with zero duality gap, then

$$g_i(x^*) \leq 0, \; h_i(x^*) = 0$$

$$\lambda_i \geq 0$$
$$\nabla f(x^*) + \sum_i \lambda_i^* \nabla g_i(x^*) + \sum_j \nu_j^* \nabla h_j(x)* = 0$$
$$\lambda_i^* g_i(x^*) = 0.$$

For convex problems: if x^*, λ^*, ν^* satisfy KKT conditions above, then they are optimal.

Optimalty conditions



(Mixed-Integer) Nonlinear Programming (**MINLP**)

In the general form,

$$\begin{aligned} \min \quad & f(x) \\ & g_i(x) \leq 0 \quad \forall i = 1, 2 \ldots, m \\ & h_j(x) = 0 \quad \forall j = 1, 2 \ldots, p \\ & x_k \in \mathbb{Z} \quad\quad \forall k \in K \subseteq \{1, 2 \ldots, n\} \end{aligned}$$

A "convex" MINLP problem is such that relaxing integrality of all variables yields a convex (continuous) problem.

- Not really a convex problem, just an abuse of notation
- (integrality constraints make the problem nonconvex)

$\Rightarrow$ difficult, have to enumerate all local minima

- However, relaxing integrality gives a convex problem

$\Rightarrow$ it is relatively easy to find a lower bound

Example: solve the following problem by Branch&Bound:

$$\begin{aligned} \min \quad & (x - \tfrac{1}{3})^2 \\ & x \in \mathbb{Z} \end{aligned}$$

(Mixed-Integer) Nonlinear Programming (**MINLP**)

In the general form,

$$
\begin{array}{ll}
\min & f(x) \\
& g_i(x) \leq 0 \quad \forall i = 1, 2 \ldots, m \\
& h_j(x) = 0 \quad \forall j = 1, 2 \ldots, p \\
& x_k \in \mathbb{Z} \qquad \forall k \in K \subseteq \{1, 2 \ldots, n\}
\end{array}
$$

A "nonconvex" NLP can model any problems considered so far including integer variables. How?
In theory NLP=MINLP, but in practice MINLPs are the hardest to solve.

- The type of nonconvexity is unknown and hard to explore.

$\Rightarrow$ do not know where local minima are.

- Relaxing integrality does not give a convex problem.

Example: solve the following problem by Branch&Bound:

$$
\begin{array}{ll}
\min & -(\frac{1}{3}x^3 + \frac{1}{4}x^2 - \frac{1}{2}x) \\
& x \in \mathbb{Z},\ x \in [-10, 10]
\end{array}
$$

Equality constrained convex quadratic programming

$$
\begin{aligned}
\min \quad & x^\top Q x + c x \\
& a_1 x = b_1 \\
& a_2 x = b_2 \\
& \vdots \\
& a_m x = b_m
\end{aligned}
$$

where Q is a square psd matrix.

Positive (Semi)Definite Matrices

A square $n \times n$ matrix A is **Positive Definite** (PD) (denoted with $A \succ 0$) if, for any n-vector $x \neq 0$, the following holds:

$$x^\top A x \geq 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j =$$

$$\begin{array}{llll}
a_{11}x_1^2 & +a_{12}x_1x_2 & + \dots & +a_{1n}x_1x_n + \\
a_{21}x_2x_1 & +a_{22}x_2^2 & + \dots & +a_{2n}x_2x_n + \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1}x_nx_1 & +a_{n2}x_nx_1 & + \dots & +a_{nn}x_n^2
\end{array}$$

$$> 0$$

Positive (Semi)Definite Matrices

A symmetric matrix $A \in R^{n \times n}$ is **Positive Semidefinite** (PSD) (denote it with $A \succeq 0$) if all principal minors of A are nonnegative.

A symmetric matrix $A \in R^{n \times n}$ is **Positive Definite** (PD) (denote it with $A \succ 0$) if all leading principal minors of A are nonnegative.

A symmetric matrix $A \in R^{n \times n}$ is **Positive Semidefinite** (PSD) if and only if $A = BB^\top$ for some $B \in R^{n \times n}$.

A symmetric matrix $A \in R^{n \times n}$ is **Positive Semidefinite** (PSD) if for all $i = 1, \ldots, n$ $a_{ii} \geq \sum_{j=1, j \neq i}^{n} |a_{ij}|$.

A symmetric matrix $A \in R^{n \times n}$ is **Positive Semidefinite** (PSD) if all its eigenvalues are nonnegative.

Application: Least squares approximation

Problem: Perform regression analysis on a set of experimental observations to identify a trend.

- A set of $(n+1)$-dimensional points

$$
\begin{aligned}
&a_1 = (a_{11}, a_{12} \ldots, a_{1n}, b_1),\\
&a_2 = (a_{21}, a_{22} \ldots, a_{2n}, b_2),\\
&\vdots\\
&a_m = (a_{m1}, a_{m2} \ldots, a_{mn}, b_m),
\end{aligned}
$$

  are the result of the experiments
- i.e., each of them corresponds to a single observation and is a vector of numbers
- for instance, for each patient in a hospital, it is the vector of parameters: blood pressure, levels of cholesterol, etc.)

Application: Least squares approximation

- we think that these data are not random, but are connected to one another by a linear function
- i.e., there is a vector $(p_1, p_2 \dots, p_n)$ and a scalar q such that

$$\begin{array}{llllll}
b_1 = & p_1 a_{11} & + p_2 a_{12} & + \dots & + p_n a_{1n} & + q \\
b_2 = & p_1 a_{21} & + p_2 a_{22} & + \dots & + p_n a_{2n} & + q \\
\vdots & \vdots & \vdots & \ddots & \vdots & \\
b_m = & p_1 a_{m1} & + p_2 a_{m2} & + \dots & + p_n a_{mn} & + q
\end{array}$$

- …but we don't know $p_1, p_2 \dots, p_n$ or q.

Application: Least squares approximation

We may just solve the linear system in the previous slide, but:

- there may be errors in the observation
- to add some robustness to the observation, usually $m \gg n$

Thus, if there are errors or noise in the observation, we look for a vector p and scalar q that minimizes the the total error.

- Variables:
  - $p_i, i = 1, 2 \ldots, n$
  - q
- Objective:

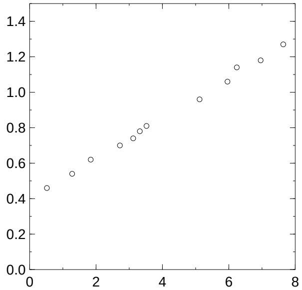
$$\sum_{i=1}^{m}(b_i - p_1 a_{i1} - p_2 a_{i2} \ldots - p_n a_{in} - q)^2 = \sum_{i=1}^{m}(b_i - \sum_{j=1}^{n}(p_j a_{ij} + q))^2$$

$\Rightarrow$ a continuous (unconstrained) NLP problem.

Example

a_1	a_2	a_3	a_4	a_5	a_6
0.52	1.28	1.84	2.72	3.12	3.32
0.46	0.54	0.62	0.70	0.74	0.78

a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
3.52	5.12	5.96	6.24	6.96	7.64
0.81	0.96	1.06	1.14	1.18	1.27



Are there p_1 and q satisfying:

$$
\begin{aligned}
0.46 &= 0.52p_1 + q\\
0.54 &= 1.28p_1 + q\\
&\vdots\\
1.27 &= 7.64p_1 + q
\end{aligned}
$$

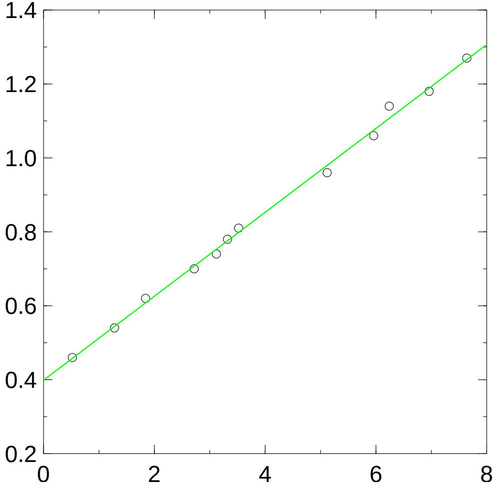
(I don't think so...)

At least minimize total error:

$$
\begin{array}{rl}
\min & (0.46 - 0.52p_1 - q)^2 + \\
+ & (0.54 - 1.28p_1 - q)^2 \\
+ & \vdots \\
+ & (1.27 - 7.64p_1 - q)^2
\end{array}
$$

Result

$p = 0.113495$, $q = 0.39875$.



Least squares - convex quadratic problem

We solve

$$\begin{aligned} \min \quad & \tfrac{1}{2}\sum(a_i^\top x - b_i)^2 = \\ & \tfrac{1}{2}\|Ax - b\|^2 = \\ & \tfrac{1}{2}(Ax-b)^\top(Ax-b) = \\ & \tfrac{1}{2}x^\top A^\top Ax - b^\top Ax + \tfrac{1}{2}b^\top b \\ & \tfrac{1}{2}x^\top Qx + c^\top x + const = \\ & f(x) \end{aligned}$$

Q is positive semidefinite because $Q = A^\top A$.
Convex unconstrained problem $\Rightarrow$ every local minimum is a global minimum! To solve set derivative to zero:

$$\nabla f(x) = Qx + c = 0$$

$$x = -Q^{-1}c \quad \text{if } Q \succ 0 \ - \ \text{One solution!}$$

If Q is singular - two cases:

1. Unbounded solution
2. Infinite number of solutions

Another example of convex quadratic problem

You are considering purchasing a stock over the next n days. You can buy it on any number of the days. Since you represent a large investor and purchase large amount of the stock each time, your purchases affect the price of the stock. If you purchase the y_i amount of the stock on day i then the price on that day goes up by θy_i on that day, by next day the effect of your purchase is dampened a bit and the price raised by $\frac{\theta}{2} y_i$, then the next day it is $\frac{\theta}{4} y_i$ and so on. If the starting price on day i is p_i derive the optimal purchasing schedule over the next 10 days is you need to purchase the total of Y.

Formulation via a convex QP

- Variables y_i - the amount of stock purchased on day i.
- The price of the stock on day 1 is $\theta y_1 + p_1$.
- The price of the stock on day 2 is $\frac{\theta}{2} y_1 + \theta y_2 + p_2$.
- .....
- The price of the stock on day i is

$$\frac{\theta}{2^{i-1}} y_1 + \frac{\theta}{2^{i-2}} y_2 + \ldots + \theta y_i + p_i.$$

We need to solve

$$\begin{array}{rl} \min & \sum_{i=1}^{n} y_i \left(\sum_{j=1}^{i} \frac{\theta}{2^{i-j}} y_j + p_i \right) = \\ & \frac{1}{2} y^\top Q y + p^\top y \\ \text{s.t.} & \sum_{i=1}^{n} y_i = Y \end{array}$$

Formulation via a convex QP

$$\begin{array}{rl} \min & \frac{1}{2}y^\top Q y + p^\top y \\ \text{s.t.} & \sum_{i=1}^n y_i = Y \end{array}$$

$$Q = \begin{pmatrix} 2\theta & \frac{\theta}{2} & \cdots & \frac{\theta}{2^{n-1}} \\ \frac{\theta}{2} & 2\theta & \cdots & \frac{\theta}{2^{n-2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\theta}{2^{n-1}} & \frac{\theta}{2^{n-2}} & \cdots & 2\theta \end{pmatrix}$$

Q is positive definite because $Q_{ii} > \sum_{j \neq i} |Q_{ij}|$.
Convex quadratic problem $\Rightarrow$ every local minimum is a global minimum!
Only equality constraint - can solve by solving a system of linear equations:

$$\begin{bmatrix} Q & e \\ e^\top & 0 \end{bmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -p \\ Y \end{pmatrix}$$

If the matrix of the system is singular - two cases:

1. Unbounded solution

Conex Quadratic Programming (QP)

$$
\begin{aligned}
\min \quad & x^\top Q x + cx \\
& a_1 x \geq b_1 \\
& a_2 x \geq b_2 \\
& \vdots \\
& a_m x \geq b_m
\end{aligned}
$$

where Q is a square psd matrix.

Application: Markowitz porfolio selection

- n stocks
- Decision variables x_j, $j = 1, \ldots, n$ represent number of shares of stock j to purchase.
- μ_j is mean of return of stock j (say after 1 year).
- σ_{jj} is variance of return of stock j (measures risk).
- σ_{ij} $(i \neq j)$ is covariance of return on one share of stock i and one of stock j.
- These parameters must all be estimated in practice.

Markowitz model

$$
\begin{aligned}
\min \quad & V(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{s.t.} \quad & \sum_{j=1}^{n} \mu_j x_j \geq L \\
& \sum_{j=1}^{n} P_j x_j \leq B \\
& x_j \geq 0 \quad \forall j = 1, \ldots, n
\end{aligned}
$$

- L is minimum acceptable expected return.
- P_j is price per share of stock j (today).
- B is budget.

Robust Markowitz model

$$
\begin{aligned}
\min \quad & V(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{s.t.} \quad & \sum_{j=1}^{n} \mu_j x_j \geq L \quad \mu_i \text{ is not known} \\
& \sum_{j=1}^{n} P_j x_j \leq B \\
& x_j \geq 0 \quad \forall j = 1, \ldots, n
\end{aligned}
$$

- L is minimum acceptable expected return.
- μ is not known, but has to be in some set M.
- How do we solve this?

Robust Markowitz model

$$
\begin{aligned}
\min \quad & V(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{s.t.} \quad & \min_{\mu \in M} \sum_{j=1}^{n} \mu_j x_j \geq L \\
& \sum_{j=1}^{n} P_j x_j \leq B \\
& x_j \geq 0 \quad \forall j = 1, \ldots, n
\end{aligned}
$$

- Optimize risk subject to minimum return requirement in the worst case.
- M can be a cubic set: $\{\mu : \mu_i^1 \leq \mu_i \leq \mu_i^2, \forall i = 1, \ldots, n\}$.
- M can be a ball: $\{\mu : \|\mu - \mu_0\| \leq R\}$.

MINLP Markowitz model

Say a portfolio can only have k stocks.

$$
\begin{aligned}
\min \quad & V(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \\
\text{s.t.} \quad & \sum_{j=1}^{n} \mu_j x_j \geq L \\
& \sum_{j=1}^{n} P_j x_j \leq B \\
& x_i \leq B y_i \\
& \sum_{i} y_i \leq k \\
& x_j \geq 0, \; y_i \in \{0,1\} \quad \forall j = 1, \ldots, n
\end{aligned}
$$

A convex MINLP commonly used in financial models.

Convex quadratically constrained quadratic problems (QCQP)

QCQPs are problems of the form

$$
\begin{array}{ll}
\min & x^\top Q^0 x + a_0 x \\
 & x^\top Q^1 x + a_1 x \le b_1 \\
 & x^\top Q^2 x + a_2 x \le b_2 \\
 & \vdots \\
 & x^\top Q^m x + a_m x \le b_m
\end{array}
$$

where Q_0, $Q_1 \ldots$, Q_m are square psd matrices.

Convex quadratically constrained quadratic problems (QCQP)

QCQPs are problems of the form

$$
\begin{array}{ll}
\min & \sum_{i=1}^n \sum_{j=1}^n q_{ij}^0 x_i x_j + \sum_{j=1}^n a_{0j} x_j \\
& \sum_{i=1}^n \sum_{j=1}^n q_{ij}^1 x_i x_j + \sum_{j=1}^n a_{1j} x_j \leq b_1 \\
& \sum_{i=1}^n \sum_{j=1}^n q_{ij}^2 x_i x_j + \sum_{j=1}^n a_{2j} x_j \leq b_2 \\
& \vdots \\
& \sum_{i=1}^n \sum_{j=1}^n q_{ij}^m x_i x_j + \sum_{j=1}^n a_{mj} x_j \leq b_m
\end{array}
$$

where Q_0, $Q_1 \ldots$, Q_m are square psd matrices.