AMA 505: Optimization Methods

Subject Lecturer: Ting Kei Pong

Lecture 9
Constrained Optimization
Penalty/Barrier methods

Problem settings

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & g_i(x) \le 0, \quad i = 1, \ldots, m. \end{array} \tag{1}$$

Here:

- f and g_i are all C^1 functions.
- For notational simplicity, we denote
$$I = \{1, \ldots, m\}.$$
- Present algorithmic ideas for solving the above.
- Assume that $\{x : \ g_i(x) \le 0 \ \forall i \in I\} \neq \emptyset$.
- Two main classes:
  - ⋆ Penalty method (exterior type);
  - ⋆ Barrier method (interior type).
- Can be generalized to include equality constraints.

Penalty method (exterior)

- Add a penalty if we get outside of the feasible set.
- If the penalty is sufficiently large, a minimizer is forced to be inside the feasible set.
- Initial idea: Define

$$P(x) := \begin{cases} 0 & \text{if } g_i(x) \leq 0 \ \ \forall i \in I, \\ \infty & \text{otherwise.} \end{cases}$$

Then we can consider the unconstrained problem:

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \ \ f(x) + P(x).$$

Except that... the objective is highly discontinuous!

Penalty method cont.

Definition. (Penalty functions)
A function $P : \mathbb{R}^n \to \mathbb{R}$ is a penalty function for the constraint set $\{x : \ g_i(x) \leq 0 \ \forall i \in I\}$ if

- $P(x) \geq 0$ for all $x \in \mathbb{R}^n$; and
- $P(x) = 0$ if and only if $g_i(x) \leq 0$ for all $i \in I$.

Examples:

- $P(x) = \sum_{i=1}^{m} \max\{g_i(x), 0\}$.
- $P(x) = \frac{1}{2} \sum_{i=1}^{m} \max\{g_i(x), 0\}^2$. Courant-Beltrami penalty function: C^1 function

Idea: Solve $\underset{x \in \mathbb{R}^n}{\text{Minimize}} \ \ f(x) + cP(x)$ for some very large $c > 0$? But...

- How large should c be?
- Can minimizers be found? (Existence? Just stationary points?)

Example 1

Example: Let $a < b$ be two real numbers. Consider the problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}}{\text{Minimize}} & f(x) \\ \text{subject to} & g_1(x) = x - b \le 0, \quad g_2(x) = a - x \le 0, \end{array}$$

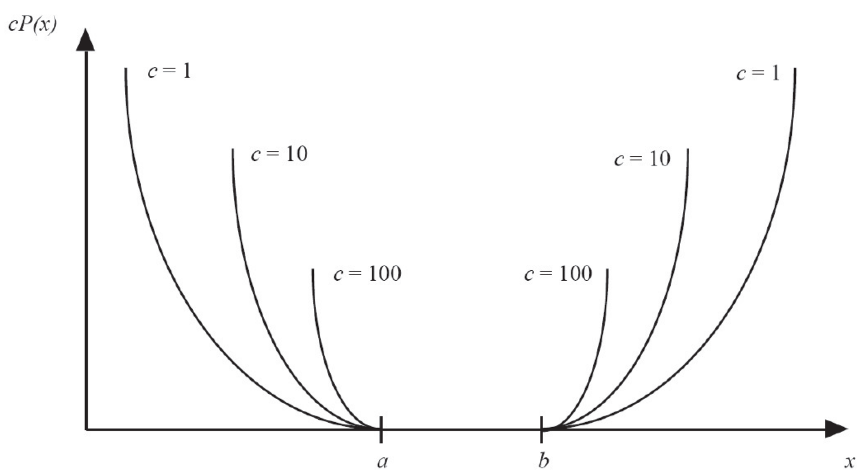
where $f \in C^1(\mathbb{R})$. The Courant-Beltrami penalty function becomes

$$\begin{aligned} P(x) &= \tfrac{1}{2}[\max\{g_1(x), 0\}^2 + \max\{g_2(x), 0\}^2] \\ &= \tfrac{1}{2}\begin{cases} (x-a)^2 & \text{if } x < a, \\ 0 & \text{if } a \le x \le b, \\ (x-b)^2 & \text{if } x > b. \end{cases} \end{aligned}$$

We can consider unconstrained optimization problems of the form

$$\underset{x\in\mathbb{R}}{\text{Minimize}} \; q_c(x) := f(x) + \frac{c}{2}\begin{cases} (x-a)^2 & \text{if } x < a, \\ 0 & \text{if } a \le x \le b, \\ (x-b)^2 & \text{if } x > b. \end{cases}$$

$cP(x)$



- Ideally, when $c \to \infty$, the solution point of the unconstrained penalty problem will converge to a solution of the original constrained problem.

Example 2

Example: Consider the problem

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{Minimize}} \quad & (x_1-6)^2+(x_2-7)^2 \\ \text{subject to} \quad & x_1+x_2-7\le 0. \end{aligned}$$

For $c>0$, define

$$q_c(x) := (x_1-6)^2+(x_2-7)^2+\frac{c}{2}(\max\{x_1+x_2-7,0\})^2. \quad \text{is convex}$$

Then q_c is convex check? and

$$\nabla q_c(x) = \begin{bmatrix} 2x_1-12+c\max\{x_1+x_2-7,0\} \\ 2x_2-14+c\max\{x_1+x_2-7,0\} \end{bmatrix}.$$

Then necessarily, q_c is minimized at a point with $x_1+x_2>7$ Why??:

$$x_1^*(c)=\frac{6+3c}{1+c} \quad x_2^*(c)=\frac{7+4c}{1+c}.$$

The limit as $c\to\infty$ is $x^*=(3,4)$. Check that this is the global minimizer of the constrained problem!

$$\begin{cases} x_1+x_2-7=0, \\ \nabla q(x)+\lambda\cdot\begin{pmatrix}1\\1\end{pmatrix}=0 \end{cases} \Rightarrow x_1=3,\ x_2=4,\ \lambda=6.$$

{ numerator
denominator

Penalty method: basic version

Penalty method for (1): basic version

Let $x^0\in\mathbb{R}^n$, $c>0$ and $\eta>1$. Set $c_1=c$. For $k=1,\ldots,$

- Find a minimizer x^k of
$$q_{c_k}(x) := f(x)+\frac{c_k}{2}\sum_{i=1}^{m}(\max\{g_i(x),0\})^2,$$
using x^{k-1} as the initial point for the iterative method.
- Update $c_{k+1}=\eta c_k$.

$$\begin{pmatrix} x^0 \\ q_{c_0} \\ x^1 \end{pmatrix} \rightarrow \begin{pmatrix} x^1 \\ q_{c_1} \\ x^2 \end{pmatrix}$$

Remark:

- As c increases, q_c becomes more ill-conditioned. The choice of x^{k-1} as a starting point for the iterative method helps alleviate the ill-conditioning.
- The above algorithm is only conceptual because finding global minimizers can be challenging if q_{c_k} is not convex. Global minimizers also may not exist! In general, only stationary points of q_{c_k} can be expected.

Penalty method: basic version cont.

Theorem 9.1: (Convergence of penalty method: basic version)
Consider (1) and suppose that $\inf f > -\infty$. Let $\{x^k\}$ be generated by the penalty method on Slide 7. Then any accumulation point x^* of $\{x^k\}$ is a globally optimal solution of (1).

Proof sketch: Assume that $\{x^{k_i}\}$ is a convergent subsequence with $\lim_{i\to\infty} x^{k_i} = x^*$.
Feasibility: Fix any feasible x. Then for each k_i, we have

$$\inf f + c_{k_i} P(x^{k_i}) \leq q_{c_{k_i}}(x^{k_i}) \leq q_{c_{k_i}}(x) = f(x),$$

where P is the Courant-Beltrami penalty function. Then

$$P(x^{k_i}) \leq \frac{f(x) - \inf f}{c_{k_i}}.$$

Hence, $P(x^*) = \lim_{i\to\infty} P(x^{k_i}) = 0$, showing that x^* is feasible.

Penalty method: basic version cont.

Proof of Theorem 9.1 sketch cont.: Assume that $\{x^{k_i}\}$ is a convergent subsequence with $\lim_{i\to\infty} x^{k_i} = x^*$.

Optimality: Fix any feasible x. Then for each k_i, we have

$$f(x^{k_i}) \leq f(x^{k_i}) + c_{k_i} P(x^{k_i}) = q_{c_{k_i}}(x^{k_i}) \leq q_{c_{k_i}}(x) = f(x),$$

where P is the Courant-Beltrami penalty function. Then $f(x^*) \leq f(x)$.

Since this is true for any feasible x and x^* is feasible, we conclude that x^* solves (1).

Role of CQ (constraint qualification)

Remarks:

- A more practical penalty method reads as follows:

> **Penalty method for (1): practical version**
> Let $x^0 \in \mathbb{R}^n$, $c > 0$ and $\eta > 1$. Set $c_1 = c$. For $k = 1, \ldots,$
> - ⋆ Find an x^k satisfying $\nabla q_{c_k}(x^k) \approx 0$, using x^{k-1} as the initial point for the iterative method.
> - ⋆ Update $c_{k+1} = \eta c_k$.

- Note that only an approximate stationary point is required in each step.
- However, since x^k is not minimizing q_{c_k}, even though $c_k \to \infty$, we cannot guarantee that accumulation pts of $\{x^k\}$ are feasible.
- If an accumulation point x^* is feasible and if the MFCQ holds at x^*, one can still show that x^* is a stationary point of (1).

Barrier method

(interior)

Recall that

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i \in I. \end{aligned} \tag{2}$$

Here:

- f and g_i are all C^1 functions.
- For barrier methods, we assume in addition that
$$S^0 := \{x : \ g_i(x) < 0 \ \ \forall i \in I\} \neq \emptyset.$$
- For simplicity, we focus on the case that all f and g_i are convex.
- In contrast to Penalty methods that are exterior methods, Barrier methods are interior methods: every iterate stays within the feasible region.

Barrier method cont.

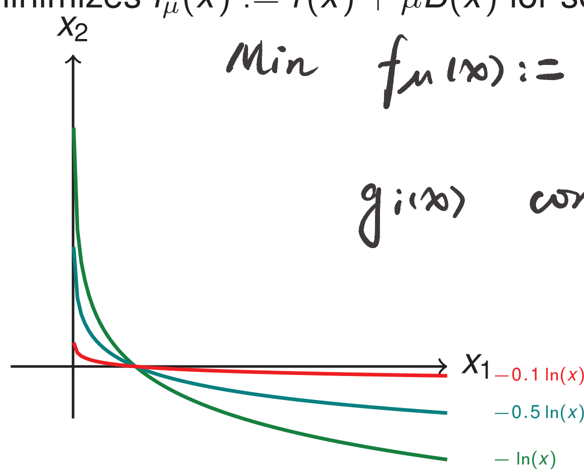
- One standard way is to make use of the log-barrier function:

$$B(x) := -\sum_{i=1}^{m} \ln[-g_i(x)]. \qquad -g_i(x) \geq 0$$

Then one minimizes $f_\mu(x) := f(x) + \mu B(x)$ for some $\mu > 0$.

Min $f_\mu(x) := f(x) + \mu B(x)$.

$g_i(x)$ convex $\Rightarrow$ $-\ln(-g_i(x))$ convex



slater point $g_i(x^0) < 0$

Barrier method: basic version

Barrier method for (2): basic version

Let $x^0 \in S^0$, $\mu > 0$ and $\eta > 1$. Set $\mu_1 = \mu$. For $k = 1, \ldots,$

- Find a minimizer x^k of
$$f_{\mu_k}(x) := f(x) - \mu_k \sum_{i=1}^{m} \ln[-g_i(x)],$$
using x^{k-1} as the initial point for the iterative method.
- Update $\mu_{k+1} = \mu_k/\eta$.

Remarks:

- Notice that μ_k is being decreased instead of being increased. Thus, at each x, we have

$$\lim_{k\to\infty} \mu_k B(x) = \begin{cases} 0 & \text{if } g_i(x) < 0 \ \ \forall i \in I, \\ \infty & \text{otherwise.} \end{cases}$$

Barrier method: basic version cont.

Remarks cont.:

- In principle, one can still apply descent methods though the function $B(x)$ is not defined on the whole $\mathbb{R}^n$. This is because descent methods make sure f_μ value decreases; in particular, f_μ will remain finite, keeping x^k feasible.
- Unlike penalty function q_c, the function $-\ln(\cdot)$ is an analytic function on $\mathbb{R}_{++}$. Indeed, when f and $g \in C^2(\mathbb{R}^n)$, one typically uses Newton's method (or its variants) to minimize f_μ.
- The above algorithm is only conceptual. In practice, μ has to be decreased judiciously to avoid getting too close to the boundary. Moreover, minimizers of f_μ may not exist.

Example

Example: Consider the problem:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^2}{\text{Minimize}} & \frac{1}{2}(x_1 + x_2 - 6)^2 \\ \text{subject to} & x \geq 0. \end{array}$$

For $\mu > 0$, define

$$f_\mu(x) := \frac{1}{2}(x_1 + x_2 - 6)^2 - \mu \ln(x_1) - \mu \ln(x_2).$$

Then

$$\nabla f_\mu(x) = \begin{bmatrix} x_1 + x_2 - 6 - \frac{\mu}{x_1} \\ x_1 + x_2 - 6 - \frac{\mu}{x_2} \end{bmatrix}.$$

Setting the gradient to zero, we get $x_1 = x_2$ and hence

$$2x_1^2 - 6x_1 - \mu = 0.$$

Since $x_1 > 0$, we have $x_1^*(\mu) = \frac{1}{2}(3 + \sqrt{9 + 2\mu}) = x_2^*(\mu)$. Thus,

$$\lim_{\mu \downarrow 0} x^*(\mu) = (3, 3).$$

This is clearly the minimizer. And look at how we approach it.

No exam on IPM!

A glimpse into IPM

Consider the linear programming problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & c^T x \\ \text{subject to} & Ax = b,\ x \geq 0. \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

- Recall that when $b \in \mathbb{R}^m_+$ and A has full row rank, one can apply simplex method to the above problem.
- However, the worst case complexity of simplex method is exponential.
- Surprisingly, a polynomial-time algorithm for LP can be derived based on barrier method, with careful update of μ, assuming in addition that the generalized Slater's condition holds for the above LP.

A glimpse into IPM cont.

Consider the linear programming problem (MINIMIZATION)

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & c^T x \\ \text{subject to} & Ax = b,\ x \geq 0. \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, and suppose that the generalized Slater's condition holds.

For each $\mu > 0$, consider the function

$$\ell_\mu(x) = c^T x - \mu \sum_{i=1}^{n} \ln(x_i).$$

$\exists y$, s.t. $A^T y < c$

Assuming additionally that the dual problem satisfies the Slater's condition, one can show that there is always a minimizer for the barrier problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \ell_\mu(x) \\ \text{subject to} & Ax = b. \end{array}$$

$e_1 = \frac{x_1}{\|x_1\|}$

$e_2 = \frac{x_2}{\|x_2\|} - (x_1, e_1)\cdot e_1$

complementary slackness ? $Ax = b$, so no

A glimpse into IPM cont.

KKT conditions: With $x_i > 0$ and $s_i > 0$ for $i = 1, \ldots, n$,

$$\begin{aligned} A^T y + s &= c, \\ Ax &= b, \\ x_i s_i &= \mu, \ \forall i = 1, \ldots, n. \end{aligned}$$

Idea: Use Newton's method to solve the above system. Apply some backtracking techniques to obtain (x_μ, y_μ, s_μ). Use this as an initial point to re-solve the system for a (judiciously chosen) smaller μ.

Remark: $F(x_\mu, y_\mu, s_\mu)$

- Backtracking scheme is used to make sure $x_\mu > 0$ and $s_\mu > 0$: wide / narrow neighborhood.
- μ has to be judiciously chosen to guarantee sufficient decrease of some measure.
- This class of method was generalized to a large class of conic optimization problems and standard solvers are available.