Numerical Optimization

Constrained Optimization - Algorithms

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Barrier and Penalty Methods

Consider the problem:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} \in X \end{aligned}$$

where $X \in \mathbb{R}^n$.

Idea:

- Approximation by an unconstrained problem
- Solve a sequence of unconstrained optimization problems

Penalty Methods

Penalize for violating a constraint

Barrier Methods

Penalize for reaching the boundary of an inequality constraint

$$\begin{array}{rl} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

Define a function,

$$\psi(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in X \\ +\infty & \text{if } \boldsymbol{x} \notin X \end{cases}$$

Solve an *equivalent* unconstrained problem:

$$\min \ f(\boldsymbol{x}) + \psi(\boldsymbol{x})$$

- *Not a practical approach*
- Replace $\psi(\boldsymbol{x})$ by a sequence of continuous non-negative functions that approach $\psi(\boldsymbol{x})$

Penalty Methods

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} \in X \end{aligned}$$

- Let $\boldsymbol{x}^*$ be a local minimum
- Let $X = \{h_j(\boldsymbol{x}) \leq 0, \ j = 1, \ldots, l\}$
- Define
$$P(\boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{l} [\max(0, h_j(\boldsymbol{x})]^2$$
- Define $q(\boldsymbol{x}, c) = f(\boldsymbol{x}) + cP(\boldsymbol{x})$
- Define a sequence $\{c^k\}$ such that $c^k \geq 0$ and $c^{k+1} > c^k$ $\forall\, k$.
- Let $\boldsymbol{x}^k = \text{argmin}_{\boldsymbol{x}} \ q(\boldsymbol{x}, c^k)$
- Ideally, $\{\boldsymbol{x}^k\} \rightarrow \boldsymbol{x}^*$ as $\{c^k\} \rightarrow +\infty$

Nonlinear Program (**NLP**)

$$\begin{array}{rl} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \le 0, \ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, \ i = 1, \ldots, m \end{array}$$

Define

$$P(\boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{l} [\max(0, h_j(\boldsymbol{x})]^2 + \frac{1}{2} \sum_{i=1}^{m} e_i^2(\boldsymbol{x})$$

and

$$q(\boldsymbol{x}, c) = f(\boldsymbol{x}) + cP(\boldsymbol{x}).$$

- Assumption: f, h_j's and e_i's are *sufficiently smooth*

Lemma

If $\boldsymbol{x}^k = \text{argmin}_{\boldsymbol{x}} \; q(\boldsymbol{x}, c^k)$ *and* $c^{k+1} > c^k$, *then*

- $q(\boldsymbol{x}^k, c^k) \leq q(\boldsymbol{x}^{k+1}, c^{k+1})$
- $P(\boldsymbol{x}^k) \geq P(\boldsymbol{x}^{k+1})$
- $f(\boldsymbol{x}^k) \leq f(\boldsymbol{x}^{k+1})$.

Proof.

$$\begin{aligned} q(\boldsymbol{x}^{k+1}, c^{k+1}) &= f(\boldsymbol{x}^{k+1}) + c^{k+1}P(\boldsymbol{x}^{k+1}) \\ &\geq f(\boldsymbol{x}^{k+1}) + c^{k}P(\boldsymbol{x}^{k+1}) \\ &\geq f(\boldsymbol{x}^{k}) + c^{k}P(\boldsymbol{x}^{k}) \\ &= q(\boldsymbol{x}^k, c^k) \end{aligned}$$

Also, $f(\boldsymbol{x}^k) + c^kP(\boldsymbol{x}^k) \leq f(\boldsymbol{x}^{k+1}) + c^kP(\boldsymbol{x}^{k+1}) \quad \ldots (1)$

$f(\boldsymbol{x}^{k+1}) + c^{k+1}P(\boldsymbol{x}^{k+1}) \leq f(\boldsymbol{x}^k) + c^{k+1}P(\boldsymbol{x}^k). \quad \ldots (2)$

Adding (1) and (2), we get $P(\boldsymbol{x}^k) \geq P(\boldsymbol{x}^{k+1})$.

$f(\boldsymbol{x}^{k+1}) + c^kP(\boldsymbol{x}^{k+1}) \geq f(\boldsymbol{x}^k) + c^kP(\boldsymbol{x}^k) \; \Rightarrow \; f(\boldsymbol{x}^{k+1}) \geq f(\boldsymbol{x}^k)$ $\square$

Lemma

Let $\boldsymbol{x}^*$ *be a solution to the problem,*

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ s.t. & \boldsymbol{x} \in X. \end{array} \qquad \ldots (P1)$$

Then, for each k, $f(\boldsymbol{x}^k) \leq f(\boldsymbol{x}^*)$.

Proof.

$$\begin{aligned} f(\boldsymbol{x}^k) &\leq f(\boldsymbol{x}^k) + c^k P(\boldsymbol{x}^k) \\ &\leq f(\boldsymbol{x}^*) + c^k P(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) \end{aligned}$$

□

Theorem

Any limit point of the sequence, $\{\boldsymbol{x}^k\}$ *generated by the penalty method is a solution to the problem (P1).*

Nonlinear Program (**NLP**)

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0, \; j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, \; i = 1, \ldots, m \end{aligned}$$

Penalty Function Method (to solve NLP)

---

(1) Input: $\{c^k\}_{k=0}^{\infty}, \epsilon$

(2) Set $k := 0$, initialize $\boldsymbol{x}^k$

(3) **while** $(q(\boldsymbol{x}^k, c^k) - f(\boldsymbol{x}^k)) > \epsilon$

  (a) $\boldsymbol{x}^{k+1} = \text{argmin}_{\boldsymbol{x}} \; q(\boldsymbol{x}, c^k)$

  (b) $k := k + 1$

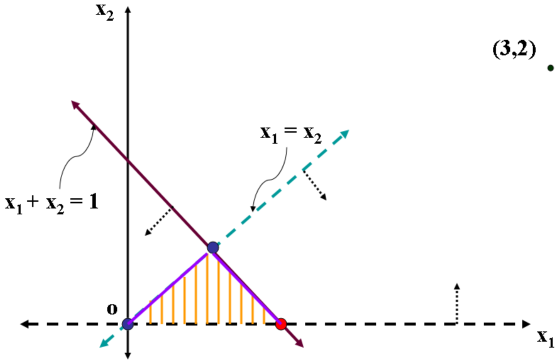
endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$

Example:

$$\begin{aligned} \min \quad & \tfrac{1}{2}[(x_1 - 3)^2 + (x_2 - 2)^2] \\ \text{s.t.} \quad & -x_1 + x_2 \le 0 \\ & x_1 + x_2 \le 1 \\ & -x_2 \le 0 \end{aligned}$$

$\boldsymbol{x}^* = (1, 0)^T$



$$\begin{array}{ll} \min & \frac{1}{2}[(x_1-3)^2+(x_2-2)^2] \\ \text{s.t.} & -x_1+x_2 \le 0 \\ & x_1+x_2 \le 1 \\ & -x_2 \le 0 \end{array}$$

$$q(\boldsymbol{x},c) = \frac{1}{2}[(x_1-3)^2+(x_2-2)^2] + \frac{c}{2}[(\max(0,-x_1+x_2))^2 + (\max(0,x_1+x_2-1))^2 + (\max(0,-x_2))^2]$$

- Let $\boldsymbol{x}^0 = (3,2)^T$ (Violates the constraint $x_1+x_2 \le 1$)
- At $\boldsymbol{x}^0$,

$$q(\boldsymbol{x},c) = \frac{1}{2}[(x_1-3)^2+(x_2-2)^2] + \frac{c}{2}[(x_1+x_2-1)^2].$$

- $\nabla_x\, q(\boldsymbol{x},c) = \boldsymbol{0} \Rightarrow \boldsymbol{x}^1(c) = \begin{pmatrix} \frac{2c+3}{2c+1} \\ \frac{2}{2c+1} \end{pmatrix} = \boldsymbol{x}^*(c)$

$$\begin{aligned} \min \quad & \tfrac{1}{2}[(x_1 - 3)^2 + (x_2 - 2)^2] \\ \text{s.t.} \quad & -x_1 + x_2 \leq 0 \\ & x_1 + x_2 \leq 1 \\ & -x_2 \leq 0 \end{aligned}$$

- At $\boldsymbol{x}^0 = (3, 2)^T$,

$$q(\boldsymbol{x}, c) = \frac{1}{2}[(x_1 - 3)^2 + (x_2 - 2)^2] + \frac{c}{2}[(x_1 + x_2 - 1)^2].$$

- $\nabla_x \, q(\boldsymbol{x}, c) = \boldsymbol{0} \;\Rightarrow\; \boldsymbol{x}^1(c) = \begin{pmatrix} \frac{2c+3}{2c+1} \\ \frac{2}{2c+1} \end{pmatrix} = \boldsymbol{x}^*(c)$
- Taking limit as $c \rightarrow \infty$, $\boldsymbol{x}^* = (1, 0)^T$

Consider the problem,

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & e(\boldsymbol{x}) = 0 \end{aligned}$$

- Let $(\boldsymbol{x}^*, \mu^*)$ be a KKT point $(\nabla f(\boldsymbol{x}^*) + \mu^* \nabla e(\boldsymbol{x}^*) = \boldsymbol{0})$
- Penalty Function: $q(\boldsymbol{x}, c) = f(\boldsymbol{x}) + cP(\boldsymbol{x})$
- As $c \to \infty$, $q(\boldsymbol{x}, c) = f(\boldsymbol{x})$

Consider the *perturbed* problem,

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & e(\boldsymbol{x}) = \theta \end{aligned}$$

and the penalty function,

$$\begin{aligned} \hat{q}(\boldsymbol{x}, c) &= f(\boldsymbol{x}) + c(e(\boldsymbol{x}) - \theta)^2 \\ &= f(\boldsymbol{x}) - 2c\theta e(\boldsymbol{x}) + ce(\boldsymbol{x})^2 \qquad \text{(ignoring constant term)} \\ &= \underbrace{f(\boldsymbol{x}) + \mu e(\boldsymbol{x})}_{\mathcal{L}(x,\mu)} + ce(\boldsymbol{x})^2 \\ &= \hat{\mathcal{L}}(\boldsymbol{x}, \mu, c) \qquad (\text{Augmented Lagrangian Function}) \end{aligned}$$

At $(\boldsymbol{x}^*, \mu^*)$, $\nabla_x \mathcal{L}(\boldsymbol{x}^*, \mu^*) = \nabla f(\boldsymbol{x}^*) + \mu^* \nabla e(\boldsymbol{x}^*) = \mathbf{0}$.

$$
\begin{aligned}
\therefore \nabla_x \hat{q}(\boldsymbol{x}^*, c) &= \nabla_x \hat{\mathcal{L}}(\boldsymbol{x}^*, \mu^*, c) \\
&= \nabla_x \mathcal{L}(\boldsymbol{x}^*, \mu^*) + 2ce(\boldsymbol{x}^*) \nabla e(\boldsymbol{x}^*) \\
&= \mathbf{0} \ \forall c
\end{aligned}
$$

Q. How to get an estimate of μ^*?
Let $\boldsymbol{x}_c^*$ be a minimizer of $\mathcal{L}(\boldsymbol{x}, \mu, c)$. Therefore,

$$
\begin{aligned}
\nabla_x \mathcal{L}(\boldsymbol{x}_c^*, \mu, c) &= \nabla f(\boldsymbol{x}_c^*) + \mu \nabla e(\boldsymbol{x}_c^*) + ce(\boldsymbol{x}_c^*) \nabla e(\boldsymbol{x}_c^*) = \mathbf{0} \\
\therefore \nabla f(\boldsymbol{x}_c^*) &= -\underbrace{(\mu + ce(\boldsymbol{x}_c^*))}_{\text{estimate of } \mu^*} \nabla e(\boldsymbol{x}_c^*)
\end{aligned}
$$

Program (EP)

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & e(\boldsymbol{x}) = 0 \end{aligned}$$

Augmented Lagrangian Method (to solve EP)

---

(1) Input: c, ϵ

(2) Set $k := 0$, initialize $\boldsymbol{x}^k, \mu^k$

(3) **while** $(\hat{\mathcal{L}}(\boldsymbol{x}^k, \mu^k, c) - f(\boldsymbol{x}^k)) > \epsilon$

  (a) $\boldsymbol{x}^{k+1} = \text{argmin}_{\boldsymbol{x}} \ \hat{\mathcal{L}}(\boldsymbol{x}^k, \mu^k, c)$

  (b) $\mu^{k+1} = \mu^k + ce(\boldsymbol{x}^k)$

  (c) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$

Nonlinear Program (**NLP**)

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{aligned}$$

- Easy to extend the Augmented Lagrangian Method to NLP
- Rewrite the inequality constraint, $h(\boldsymbol{x}) \leq 0$ as an equality constraint,

$$h(\boldsymbol{x}) + y^2 = 0$$

Barrier Methods

- Typically applicable to inequality constrained problems

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \end{array}$$

Let $X = \{\boldsymbol{x} : h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l\}$

- Some Barrier functions (defined on the *interior* of X)

$$B(\boldsymbol{x}) = -\sum_{j=1}^{l} \frac{1}{h_j(\boldsymbol{x})} \quad \text{or} \quad B(\boldsymbol{x}) = -\sum_{j=1}^{l} \log(-h_j(\boldsymbol{x}))$$

- Approximate problem using Barrier function (for $c > 0$)

$$\begin{array}{ll} \min & f(\boldsymbol{x}) + \frac{1}{c} B(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \text{ Interior of } X \end{array}$$

Cutting-Plane Methods

Primal Problem

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0, \; j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, \; i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{aligned}$$

Dual Problem

$$\begin{aligned} \max \quad & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

X is a compact set.

Dual Function: $z = \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{x \in X} f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x})$

Equivalent Dual problem

$$\begin{aligned} \max_{z, \mu, \lambda} \quad & z \\ \text{s.t.} \quad & z \leq f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x}), \;\; \boldsymbol{x} \in X \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

Linear Program with infinite constraints

Equivalent Dual problem

$$
\begin{aligned}
\max_{z,\mu,\lambda} \quad & z \\
\text{s.t.} \quad & z \leq f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x}), \;\; \boldsymbol{x} \in X \\
& \boldsymbol{\lambda} \geq \mathbf{0}
\end{aligned}
$$

Idea: Solve an approximate dual problem.
Suppose we know $\{\boldsymbol{x}^j\}_{j=0}^{k-1}$ such that

$$z \leq f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x}), \;\; \boldsymbol{x} \in \{\boldsymbol{x}^0, \ldots, \boldsymbol{x}^{k-1}\}$$

Approximate Dual Problem

$$
\begin{aligned}
\max_{z,\mu,\lambda} \quad & z \\
\text{s.t.} \quad & z \leq f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x}), \;\; \boldsymbol{x} \in \{\boldsymbol{x}^0, \ldots, \boldsymbol{x}^{k-1}\} \\
& \boldsymbol{\lambda} \geq \mathbf{0}
\end{aligned}
$$

Let $(z^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ be the optimal solution to this problem.

Approximate Dual Problem

$$
\begin{aligned}
\max_{z,\mu,\lambda} \quad & z \\
\text{s.t.} \quad & z \le f(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x}) + \boldsymbol{\mu}^T e(\boldsymbol{x}), \quad \boldsymbol{x} \in \{\boldsymbol{x}^0, \ldots, \boldsymbol{x}^{k-1}\} \\
& \boldsymbol{\lambda} \ge \boldsymbol{0}
\end{aligned}
$$

If $z^k \le f(\boldsymbol{x}) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}) \ \forall\, \boldsymbol{x} \in X$, then $(z^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ is the solution to the dual problem.

Q. How to check if $z^k \le f(\boldsymbol{x}) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}) \ \forall\, \boldsymbol{x} \in X$?

Consider the problem,

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{x} \in X
\end{aligned}
$$

and let $\boldsymbol{x}^k$ be an optimal solution to this problem.

$$\begin{array}{ll} \min & f(\boldsymbol{x}) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in X \end{array}$$

and let $\boldsymbol{x}^k$ be an optimal solution to this problem.

- If $z^k \leq f(\boldsymbol{x}^k) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}^k) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}^k)$, then $(\boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ is an optimal solution to the Lagrangian dual problem.
- If $z^k > f(\boldsymbol{x}^k) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}^k) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}^k)$, then add the constraint, $z \leq f(\boldsymbol{x}^k) + \boldsymbol{\lambda}^T h(\boldsymbol{x}^k) + \boldsymbol{\mu}^T e(\boldsymbol{x}^k)$ to the approximate dual problem.

Nonlinear Program (**NLP**)

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\
& e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \\
& \boldsymbol{x} \in X
\end{aligned}
$$

Summary of steps for Cutting-Plane Method:

- Initialize with a feasible point $\boldsymbol{x}^0$
- while stopping condition is not satisfied

$$
(z^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) = \begin{array}{l} \text{argmax}_{z,\lambda,\mu} \quad z \\ \text{s.t. } z \leq f(\boldsymbol{x}^j) + \boldsymbol{\lambda}^T h(\boldsymbol{x}^j) + \boldsymbol{\mu}^T e(\boldsymbol{x}^j),\ j = 0, \ldots, k-1 \\ \quad \boldsymbol{\lambda} \geq \boldsymbol{0} \end{array}
$$

$$
\boldsymbol{x}^k = \begin{array}{l} \text{argmin}_x \quad f(\boldsymbol{x}) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}) \\ \quad \boldsymbol{x} \in X \end{array}
$$

Stop if $z^k \leq f(\boldsymbol{x}^k) + \boldsymbol{\lambda}^{k^T} h(\boldsymbol{x}^k) + \boldsymbol{\mu}^{k^T} e(\boldsymbol{x}^k)$. Else, $k := k + 1$.