ISEN 629: Engineering Optimization

Lecture 5

Sergiy Butenko

Industrial and Systems Engineering
Texas A& M University

Fall 2007

Conjugate gradient algorithm

- To implement a conjugate direction algorithm, one needs to specify n Q-conjugate directions. This can be done using a procedure similar to the *Gram-Schmidt* orthogonalization.
- Conjugate gradient algorithm provides an alternative approach that allows to find Q-conjugate directions one by one as the steps of the algorithm progress.
- The directions used in this algorithm are related to the gradient. More specifically, the directions are chosen so that

$$\begin{aligned} d_0 &= -\nabla_0; \\ d_{k+1} &= -\nabla_{k+1} + \beta_k d_k, \ k = 0, \ldots, n-2. \end{aligned}$$

- Here β_k needs to be chosen so that $d_{k+1} Q d_k = 0$. We have

$$(-\nabla_{k+1} + \beta_k d_k)^T Q d_k = 0 \quad \Rightarrow \quad \beta_k = \frac{\nabla_{k+1}^T Q d_k}{d_k^T Q d_k}.$$

  It can be shown by induction that with such choice of β_k the directions $d_0, d_1, \ldots, d_{n-1}$ are Q-conjugate.

Conjugate gradient algorithm

Thus, the conjugate gradient algorithm proceeds as follows:

$d_0 = -\nabla_0$;
for $k = 0 : n-1$
 $\alpha_k = -\frac{\nabla_k^T d_k}{d_k^T Q d_k}$;
 $x_{k+1} = x_k + \alpha_k d_k$;
 $\beta_k = \frac{\nabla_{k+1}^T Q d_k}{d_k^T Q d_k}$;
 $d_{k+1} = -\nabla_{k+1} + \beta_k d_k$;
end

Non-quadratic problems

If the minimized function is not quadratic, one can use its quadratic approximation given by the Taylor's theorem:

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$$

- We could replace $f(x)$ with this quadratic approximation in x_k and apply the conjugate gradient algorithm to find the minimizer of the quadratic, which we would use as x_{k+1}.
- However, computing and evaluating the Hessian at each iteration are computationally expensive procedures that we would like to avoid.
- In the conjugate gradient algorithm above, two operations involve the matrix Q which would correspond to the Hessian. These operations are computing α_k and β_k.
- While α_k could be approximated using line search, we need to find a way to approximate β_k.
- Next we discuss several such methods.

Non-quadratic problems

The Hestenes-Stiefel formula

$$\beta_k = \frac{\nabla_{k+1}^T Q d_k}{d_k^T Q d_k} \approx \frac{\nabla_{k+1}^T(\nabla_{k+1} - \nabla_k)}{d_k^T(\nabla_{k+1} - \nabla_k)}.$$

The Polak-Ribiere formula

$$\beta_k \approx \frac{\nabla_{k+1}^T(\nabla_{k+1} - \nabla_k)}{d_k^T(\nabla_{k+1} - \nabla_k)} = \frac{\nabla_{k+1}^T(\nabla_{k+1} - \nabla_k)}{d_k^T \nabla_{k+1} - d_k^T \nabla_k} \approx \frac{\nabla_{k+1}^T(\nabla_{k+1} - \nabla_k)}{\nabla_k^T \nabla_k}.$$

The Fletcher-Reeves formula

$$\beta_k \approx \frac{\nabla_{k+1}^T(\nabla_{k+1} - \nabla_k)}{\nabla_k^T \nabla_k} = \frac{\nabla_{k+1}^T \nabla_{k+1} - \nabla_{k+1}^T \nabla_k}{\nabla_k^T \nabla_k} = \frac{\nabla_{k+1}^T \nabla_{k+1}}{\nabla_k^T \nabla_k}.$$

Rate of convergence

- In quadratic case, the method terminates in at most n iterations.
- In general case, we restart the method after each n iterations.
- In a neighborhood of a strict minimum the conjugate gradient schemes have a local n-step quadratic convergence:

$$\|x_{n+1} - x^*\| \leq C\|x_0 - x^*\|^2.$$

- The local convergence is slower than that of the quasi-Newton methods.
- However, the conjugate gradient schemes require less computational effort per iteration.

Constrained minimization

Consider the problem

$$\begin{array}{ll} \min & f_0(x) \\ s.t. & f_i(x) \leq 0, \ i = 1, \dots, m, \end{array}$$

where $f_i(x)$ are smooth functions, $i = 0, \dots, m$.

We will briefly review two types of schemes of *sequential unconstrained minimization*, in which a solution to this problem is approximated by a sequence of solutions to some auxiliary unconstrained minimization problems:

- the penalty function methods;
- the barrier methods.

Penalty function methods

Definition

A continuous function $\Phi(x)$ is called a penalty function (or simply, a penalty) for a closed set Q if

- $\Phi(x) = 0$ for any $x \in Q$,
- $\Phi(x) > 0$ for any $x \notin Q$.

Examples: Denote by $(a)_+ = \max\{a, 0\}$. Then for

$$Q = \{x \in \mathbb{R}^n : f_i(x) \leq 0, i = 1, \dots, m\}$$

the following functions are penalties:

- Quadratic penalty: $\Phi(x) = \sum_{i=1}^{m} ((f_i(x))_+)^2$.
- Nonsmooth penalty: $\Phi(x) = \sum_{i=1}^{m} (f_i(x))_+$.

Penalty function methods

The penalty functions have the following property:

> *If $\Phi_1(x)$ is a penalty for Q_1 and $\Phi_2(x)$ is a penalty for Q_2, then $\Phi_1(x) + \Phi_2(x)$ is a penalty for intersection $Q_1 \cap Q_2$*

The general scheme of a penalty function methods is as follows.

Penalty function method

0. Choose $x_0 \in \mathbb{R}^n$. Choose a sequence of penalty coefficients: $0 < t_k < t_{k+1}$ and $t_k \to \infty$.
1. k^{th} iteration ($k \geq 0$): Find a point $x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \{f_0(x) + t_k \Phi(x)\}$ using x_k as a starting point.

Penalty function methods

Theorem

Let x^* be the global optimal solution to the considered constrained problem. If there exists a value $\bar{t} > 0$ such that the set

$$S = \{x \in \mathbb{R}^n : f_0(x) + \bar{t}\Phi(x) \leq f_0(x^*)\}$$

is bounded, then

$$\lim_{k\to\infty} f_0(x_k) = f_0(x^*), \quad \lim_{k\to\infty} \Phi(x_k) = 0.$$

Penalty function methods

Proof: Denote by $\Psi_k(x) = f_0(x) + t_k\Phi(x)$, $\Psi_k^* = \min_{x\in\mathbb{R}^n} \Psi_k(x)$.

(a) $\Psi_k^* \leq \Psi_k(x^*) = f_0(x^*)$;

(b) For any $x \in \mathbb{R}^n$ we have $\Psi_{k+1}(x) \geq \Psi_k(x) \;\Rightarrow\; \Psi_{k+1}^* \geq \Psi_k^*$.

(c) (a)+(b) $\Rightarrow$ there exists a limit $\lim_{k\to\infty} \Psi_k^* = \Psi^* \leq f_0(x^*)$.

(d) If $t_k > \bar{t}$ then $f_0(x_{k+1}) + \bar{t}\Phi(x_{k+1}) \leq f_0(x_{k+1}) + t_k\Phi(x_{k+1}) = \Psi_k^* \leq f_0(x^*)$.

(e) The sequence $\{x_k\}$ has a limit point (Bolzano-Weierstrass theorem: every bounded, infinite set of real numbers has a limit point.)

(f) Since $\lim_{k\to\infty} t_k = +\infty$, and $\lim_{k\to\infty} \Psi_k^* \leq f_0(x^*)$ (from (c)), we have $\Phi(x_*) = 0$ and $f_0(x_*) \leq f_0(x^*)$ for <u>any</u> limit point x_* of $\{x_k\}$.

(g) $\Phi(x_*) = 0 \;\Rightarrow\; x_* \in Q$, so x_* is a global minimizer of the problem. $\square$

Penalty function methods

Practical questions to address:

- What kind of penalty function should we use?
- How to choose the sequence of penalty coefficients $\{t_k\}$?
- How to solve the auxiliary unconstrained problems?

These questions are very difficult to answer in the framework of general nonlinear optimization theory.

Barrier methods

Definition
Let Q be a closed set with a nonempty interior. A continuous function $F(x)$ is called a barrier function (or simply, a barrier) for Q if $F(x) \to \infty$ when x approaches the boundary of Q.

Examples: Let $Q = \{x \in \mathbb{R}^n : f_i(x) \leq 0, i = 1, \ldots, m\}$ satisfy the Slater condition: there exists $\bar{x}$ such that $f_i(\bar{x}) < 0, i = 1, \ldots, m$. Then the following functions are barriers for Q:

- Power-function barrier: $F(x) = \sum_{i=1}^{m} \frac{1}{(-f_i(x))^p}, p \geq 1$.
- Logarithmic barrier: $F(x) = -\sum_{i=1}^{m} \ln(-f_i(x))$.
- Exponential barrier: $F(x) = -\sum_{i=1}^{m} \exp\left(\frac{1}{-f_i(x)}\right)$.

Barrier methods

The barrier functions have the following property:

If $F_1(x)$ is a barrier for Q_1 and $F_2(x)$ is a barrier for Q_2, then $F_1(x) + F_2(x)$ is a barrier for intersection $Q_1 \cap Q_2$

The general scheme of a barrier function methods is as follows.

Barrier function method

0. Choose $x_0 \in \operatorname{int} Q$. Choose a sequence of penalty coefficients: $0 < t_k < t_{k+1}$ and $t_k \to \infty$.
1. k^{th} iteration ($k \geq 0$):
   Find a point $x_{k+1} = \arg\min_{x \in Q}\{f_0(x) + \frac{1}{t_k} F(x)\}$ using x_k as a starting point.

Barrier methods

Denote by

$$\Psi_k(x) = f_0(x) + \frac{1}{t_k} F(x), \; \Psi_k^* = \min_{x \in Q} \Psi_k(x).$$

Theorem
Let barrier $F(x)$ be bounded from below on Q. Then

$$\lim_{k \to \infty} \Psi_k^* = f^*,$$

where f^* is the optimal value of the considered problem.

Barrier methods

Proof: Let $F(x) \geq F^*$ for all $x \in Q$. For any $\bar{x} \in \operatorname{int} Q$ we have

$$\limsup_{k \to \infty} \Psi_k^* \leq \lim_{k \to \infty} \left[f_0(\bar{x}) + \frac{1}{t_k} F(\bar{x}) \right] = f_0(\bar{x}),$$

thus $\boxed{\limsup_{k \to \infty} \Psi_k^* \leq f^*}$ (since the opposite would yield $\limsup_{k \to \infty} \Psi_k^* > f_0(\bar{x})$ for some $\bar{x} \in \operatorname{int}(Q)$).

On the other hand,

$$\Psi_k^* = \min_{x \in Q} \left\{ f_0(x) + \frac{1}{t_k} F(x) \right\} \geq \min_{x \in Q} \left\{ f_0(x) + \frac{1}{t_k} F^* \right\} = f^* + \frac{1}{t_k} F^*,$$

so $\boxed{\liminf_{k \to \infty} \Psi_k^* \geq f^*}$.

Thus, $\lim_{k \to \infty} \Psi_k^* = f^*$. $\square$

Barrier methods

Again, many practical questions need to be addressed in order to use this method:

- How to find a starting point x_0?
- How to choose the barrier function?
- How to update the penalty coefficients?
- How to solve the auxiliary problems?

The above questions are very difficult to answer in general, but can be answered precisely in the framework of <u>convex</u> optimization.