9.1

Line Search Algorithms

Katta G. Murty, IOE 611 Lecture slides

Algorithms for one-dimensional opt. problems of form: min $f(\lambda)$ over $\lambda \geq 0$ (or $a \leq \lambda \leq b$ for some $a < b$).

Bracket: An interval in feasible region which contains the min.

A **2-point bracket:** is $\lambda_1 \leq \lambda \leq \lambda_2$ where $f'(\lambda_1) < 0$ and $f'(\lambda_2) > 0$.

A **3-point bracket** is $\lambda_1 < \lambda_2 < \lambda_3$ s. th. $f(\lambda_2) \leq \min\{f(\lambda_1), f(\lambda_3)\}$.

How to Select An Initial 3-pt. Bracket?

First consider min $f(\lambda)$ over $\lambda \geq 0$ where at 0 the function decreases as λ increases from 0 (otherwise 0 itself is a local min).

Select a step length Δ s. th. $f(\Delta) < f(0)$ (possible because of above facts). Define

$$\lambda_0 = 0$$
$$\lambda_r = \lambda_{r-1} + 2^{r-1}\Delta, \quad \text{for } r = 1, 2, \ldots$$

as long as $f(\lambda_r)$ keeps on decreasing, until a value k for r is found for first time s. th. $f(\lambda_{k+1}) > f(\lambda_k)$.

So we have $f(\lambda_k) < f(\lambda_{k-1})$ also. Among 4 points $\lambda_{k-1}, \lambda_k, \frac{1}{2}(\lambda_k + \lambda_{k+1}), \lambda_{k+1}$, drop either λ_{k-1} or λ_{k+1} whichever is farther from the point in pair $\{\lambda_k, \frac{1}{2}(\lambda_k + \lambda_{k+1})\}$ that yields smallest value for $f(\lambda)$; and remaining 3 pts. form an equi-distant 3-pt. bracket.

If problem is $\quad \min f(\lambda) \quad$ over $a \leq \lambda \leq b$ it is reasonable to assume that $f(\lambda)$ decreases as λ increases thro' a (otherwise a is a local min). Apply above procedure choosing Δ sufficiently small, keep defining λ_r as above until either a k defined as above is found, or until the upper bound b for λ is reached.

Methods Using 3-pt. Brackets

Golden Section Search: Works well when function is **unimodal** in initial bracket (i.e., it has unique local min there). Quite reasonable assumption in many applications.

Let λ_* be unknown min pt. Unimodality $\Rightarrow \lambda_1 < \lambda_2 < \lambda_*$ then $f(\lambda_1) > f(\lambda_2)$ and $\lambda_* < \lambda_1 < \lambda_2$ then $f(\lambda_2) > f(\lambda_1)$.

General Step: Let $[\alpha, \beta]$ be interval of current bracket. $f(\alpha), f(\beta)$ would already have been computed earlier.

$\tau = \frac{2}{1+\sqrt{5})} \simeq 0.618$ is the **Golden Ratio**. Let $\lambda_1 = \alpha + (1 - 0.618)(\beta - \alpha)$, $\lambda_2 = \alpha + 0.618(\beta - \alpha)$. If $f(\lambda_1), f(\lambda_2)$ not available, compute them. If

$f(\lambda_1) < f(\lambda_2)$ new bracket is $[\alpha, \lambda_1, \lambda_2]$

$f(\lambda_1) > f(\lambda_2)$ new bracket is $[\lambda_1, \lambda_2, \beta]$

$f(\lambda_1) = f(\lambda_2)$ new bracket interval is $[\lambda_1, \lambda_2]$

Repeat with new bracket, continue until bracket length becomes small.

Quadratic Fit Line Search Method

General Step: Let $(\lambda_1, \lambda_2, \lambda_3)$ be current 3-pt. bracket. Fit a Quad. approx. $Q(\lambda) = a\lambda^2 + b\lambda + c$ to $f(\lambda)$ using function values at $\lambda_1, \lambda_2, \lambda_3$. Because of bracket property, $Q(\lambda)$ will be convex. Its unique min is:

$$\hat{\lambda} = \frac{(\lambda_2^2 - \lambda_3^2)f(\lambda_1) + (\lambda_3^2 - \lambda_1^2)f(\lambda_2) + (\lambda_1^2 - \lambda_2^2)f(\lambda_3)}{2[(\lambda_2 - \lambda_3)f(\lambda_1) + (\lambda_3 - \lambda_1)f(\lambda_2) + (\lambda_1 - \lambda_2)f(\lambda_3)]}$$

If

$\hat{\lambda} > \lambda_2$ and $f(\hat{\lambda}) > f(\lambda_2)$ new bracket is $[\lambda_1, \lambda_2, \hat{\lambda}]$

$\hat{\lambda} > \lambda_2$ and $f(\hat{\lambda}) < f(\lambda_2)$ new bracket is $[\lambda_2, \hat{\lambda}, \lambda_3]$

$\hat{\lambda} > \lambda_2$ and $f(\hat{\lambda}) = f(\lambda_2)$ new bracket is either of the above.

If $\hat{\lambda} < \lambda_2$ similar to above.

If $\hat{\lambda} = \lambda_2$, quad. fit failed to produce a new pt. In this case:

if $\lambda_3 - \lambda_1 \leq \epsilon =$ positive tolerance, stop with λ_2 as best pt.

$$\lambda_3 - \lambda_1 > \epsilon, \text{ define } \hat{\lambda} = \begin{cases} \lambda_2 + \epsilon/2 & \text{if } \lambda_2 - \lambda_1 < \lambda_3 - \lambda_2 \\ \lambda_2 - \epsilon/2 & \text{if } \lambda_2 - \lambda_1 > \lambda_3 - \lambda_2 \end{cases}$$

Compute $f(\hat{\lambda})$ and go to the above step again.

Terminate method when either of $\max\{f(\lambda_1), f(\lambda_3)\} - f(\lambda_2)$ or $\lambda_3 - \lambda_1$ or $|f'(\lambda_2)|$ becomes smaller than a tolerance.

Methods Using 2-pt. Brackets

The method of Bisection: Let $[a, b]$ be current bracket. Compute $f'((a+b)/2)$. If

$$f'((a+b)/2)\begin{cases} = 0 & (a+b)/2 \text{ is a stationary pt., terminate} \\ > 0 & \text{continue with } [a, (a+b)/2] \\ < 0 & \text{continue with } [(a+b)/2, b] \end{cases}$$

Not preferred, because method does not use function values.

Cubic Interpolation Method: Let $[\lambda_1, \lambda_2]$ be current bracket. Fit a cubic approx. to $f(\lambda)$ using values of $f(\lambda_1)$, $f(\lambda_2)$, $f'(\lambda_1)$, $f'(\lambda_2)$. Because of bracket conds., min of this cubic func. λ_* is inside bracket. It is:

$$\lambda_* = \lambda_1 + (\lambda_2 - \lambda_1)\left(1 - \frac{f'(\lambda_2) + \nu - \eta}{f'(\lambda_2) - f'(\lambda_1) + 2\nu}\right)$$

where $\eta = \frac{3(f(\lambda_1) - f(\lambda_2))}{\lambda_2 - \lambda_1} + f'(\lambda_1) + f'(\lambda_2)$ and $\nu = (\eta^2 - f'(\lambda_1)f'(\lambda_2))^{1/2}$.

If $|f'(\lambda_*)|$ is small accept λ_* as approx. to min. Otherwise if $f'(\lambda_*) > 0 \quad (< 0) \quad$ continue with $[\lambda_1, \lambda_*] \quad ([\lambda_*, \lambda_2])$.

Line Search Method Based on Piecewise Linear Approximation

Let $[a, b]$ be initial 2-pt. bracket for $f(\lambda)$.

A **piecewise linear** (or **polyhedral) approximation** for $f(\lambda)$ in the bracket $[a, b]$ is the pointwise supremum function of the linearizations of $f(\lambda)$ at a & b, $P(\lambda) = \max\{f(a)+f'(a)(\lambda-a), f(b) + f'(b)(\lambda - b)\}$.

From bracket conds., the min of $P(\lambda)$ occurs in $[a, b]$ at the point where the two linearizations are equal, it is $a + d^P$ where

$$d^P = \frac{f(a) - f(b) - f'(b)(a - b)}{f'(b) - f'(a)}$$

The quadratic Taylor series approximation of $f(\lambda)$ at a is $Q(\lambda) = f(a) + f'(a)(\lambda - a) + \frac{1}{2}(\lambda - a)^2 f''(a)$. The min of $Q(\lambda)$ occurs at $a + d^Q$ where

$$d^Q = \begin{cases} f'(a)/f''(a) & \text{if } f''(a) > 0 \\ \pm\infty & \text{if } f''(a) \leq 0 \end{cases}$$

One simple strategy is to take the new point to be $\lambda_1 = a + \min\{d^P, d^Q\}$ if $d^Q > 0$, or $a + d^P$ otherwise; and take the next bracket to be the 2-pt. bracket among $[a, \lambda_1]$, $[\lambda_1, b]$ and continue.

But C. Lemarechal & R. Mifflin make several modifications to guarantee convergence to a stationary pt. even when $f(\lambda)$ is nonconvex. Also see Murty LCLNP.

Newton's Method for line search

2nd order method for: min $f(\lambda)$ over entire real line.

It is application of Newton-Raphson method to solve the eq. $f'(\lambda) = 0$.

Starting with initial point λ_0, generates iterates by

$$\lambda_{r+1} = \lambda_r - \frac{f'(\lambda_r)}{f''(\lambda_r)}$$

assuming all $f''(\lambda_r) > 0$. Not suitable if 2nd derivative ≤ 0 at a point encountered.

Suitable if a near opt. sol. known, then function is locally convex in the nbhd. of opt.

Secant Method: A modified Newton's method. Method initiated with a 2-pt. bracket, and replaces $f''(\lambda_r)$ by the finite difference approx. $\frac{f'(\lambda_r)-f'(\lambda_{r-1})}{\lambda_r-\lambda_{r-1}}$ in Newton formula.

Inexact Line Search Procedures

Exact line searches expensive in subroutines for solving higher dimensional min. problems. Inexact line searches with sufficient degree of descent do guarantee convergence of overall algo.

Consider: min $f(\lambda) = \theta(\bar{x} + \lambda y)$ over $\lambda \geq 0$. $\bar{x}$ is current point, y is a descent direction at $\bar{x}$.

Conditions for Global Convergence

1. Sufficient Rate of Descent Condition: If $\bar{\lambda}$ is step length choosen, new pt. is $\bar{x} + \bar{\lambda} y$. Cond. is that average rate of descent from $\bar{x}$ to $\bar{x} + \bar{\lambda} y$ be $\geq$ specified fraction of initial rate of decrease in $theta(x)$ at $\bar{x}$ in direction y. Select $\alpha \in (0, 1)$ and require $\bar{\lambda}$ satisfy:

$$f(\bar{\lambda}) \leq f(0) + \bar{\lambda} \alpha f'(0)$$

2. Step Length not too Small: Cond. 1 will always be satisfied for any $\alpha < 1$ by sufficiently small steps. This requires that step lengths are not too small. Several equivalent ways to ensure this. One is to require that $\bar{\lambda}$ satisfy:

$$(2.1) \quad f'(\bar{\lambda}) \geq \beta f'(0)$$

for some selected $\beta \in (\alpha, 1)$. Sats step must be long enough that rate of change in f at $\bar{\lambda}$ is $\geq$ specified fraction of its magnitude at 0.

Theorem: If $\theta(x)$ is bounded below on R^n, $\nabla\theta(\bar{x})y < 0$, $0 < \alpha < \beta < 1$, there exists $0 < \lambda_1 < \lambda_2$ s. th. for any $\lambda \in [\lambda_1, \lambda_2]$, $\bar{x} + \lambda y$ satisfies both conds. 1 & (2.1).

Theorem: $\theta(x)$ is cont. diff., and there exists $\delta \geq 0$ s. th. $||\nabla\theta(z) - \nabla\theta(x)|| \leq \delta||z - x|| \ \forall x, z \in R^n$. Starting with x^0 suppose the sequence $\{x^r\}$ generated by choosing y^r satisfying $\nabla\theta(x^r)y^r < 0$, and the iteration $x^{r+1} = x^r + \lambda_r y^r$ where λ_r satisfies conds. 1 & (2.1). Then one of following holds.

(a) Either $\nabla\theta(x^k) = 0$ for some k, or

(b) $\lim_{r\to\infty} f(x^r) = -\infty$, or

(c) $\lim_{r\to\infty} \frac{\nabla\theta(x^r)y^r}{||y^r||} = 0$.

From (c) we know that unless the angle between $\nabla\theta(x^r)$ & y^r converges to 90^0 as $r \to \infty$, either $\nabla\theta(x^r) \to 0$, or $f(x^r) \to -\infty$, or both.

If we can guarantee that angle between $\nabla\theta(x^r)$ & y^r is bounded away from 90^0 (this can be guaranteed by proper choice of y^r, for example in Quasi-Newton methods $y^r = -H_r\nabla\theta(x^r)$ where H_r is PD, this property will hold if cond. numbers of $\{H_r\}$ are uniformly bounded above) then either $\nabla\theta(x^r) \to 0$, or $f(x^r) \to -\infty$, or both.

Other equivalent conds. to ensure step lengths not too small are:

$$(2.2) \qquad f(\bar{\lambda}) \geq f(0) + \gamma\bar{\lambda}f'(0)$$

for some selected $\gamma \in (0, \alpha)$.

A simple procedure to satisfy these conds. uses $0 < \epsilon_1 < 1 < \epsilon_2$ ($\epsilon_1 = 0.2, \epsilon_2 = 2$ commonly used). ϵ plays role of α in above.

Cond. 1 is equivalent to requiring step length $\bar{\lambda}$ satisfy $f(\bar{\lambda}) \leq \ell(\bar{\lambda}) = f(0) + \bar{\lambda}\epsilon_1 f'(0)$.

Cond. 2 is enforced by requiring $f(\epsilon_2\bar{\lambda}) > \ell(\epsilon_2\bar{\lambda}) = f(0) + \epsilon_2\bar{\lambda}\epsilon_1 f'(0)$.

Step 1: See if length of 1 satisfies both conds. 1 and 2. If so, select $\bar{\lambda} = 1$ as step length and terminate procedure. If step length of 1 violates cond. 1 [cond. 2] go to Step 2 [Step 3].

Step 2: Take $\bar{\lambda} = \frac{1}{\epsilon_2^t}$ where t is the smallest integer > 1 for which $f(\frac{1}{\epsilon_2^t}) \leq \ell(\frac{1}{\epsilon_2^t})$, and terminate procedure.

Step 3: Take $\bar{\lambda} = \epsilon_2^t$ where t is largest integer ≥ 0 for which $f(\epsilon_2^t) \leq \ell(\epsilon_2^t)$, and terminate procedure.

Other ways of implementing procedure uses quadratic fits to determine new step lengths when current step length (step length $= 1$ initially) violates one of the conds.