Globally convergent Modification of Newton's Method

10/12/2006

(5)

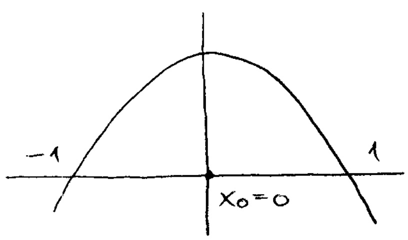
- Augment Newton with "globalization"
- To increase likelihood of convergence to some solution from an arbitrary initial guess.
- Note:
  - convergence is not (cannot) be guaranteed
  - also convergence is not necessarily be to any distinguished solution, e.g. a global minimizer
- Structure:
  - Begin with initial trial step (Newton step or something related to it)
  - Test for adequate progress
  - Modify if necessary to get a new trial step; return to test
- Criteria of which to base test
  - Suppose have $\{x_k\}$

    Want criteria that ideally imply $x_k \to x_*$ where $F(x_*)=0$

    Reasonable: require $\|F(x_{k+1})\| < \|F(x_k)\|$ for each k

    This isn't enough!

    Example: $f(x) = 1-x^2$



a) $x_k = x_{k-1} + \frac{1}{2^{k+1}}$

$\Rightarrow x_0 = 0$

$x_1 = 1/4$

$x_2 = 3/8$

$x_3 = 7/16$

$\vdots$

$x_k \longrightarrow 1/2$

and $f(\frac{1}{2}) \neq 0$

have $f(x_{k+1}) < f(x_k)$ but steps are too short

6) $x_k = -x_{k-1} + (-1)^{k-1}\, 2^{-k}$

$\Rightarrow x_0 = 0 \; ; \; x_1 = \frac{1}{2} \; ; \; x_2 = -\frac{3}{4} \; ; \; x_3 = \frac{7}{8} \; , \; \dots$

see $f(x_{k+1}) < f(x_k)$ again but $\lim_{k\to\infty} x_k$ doesn't exist

Problem: steps are too long

$x_* = 1$ and $x_{**} = -1$ are solutions and are both limit points of $\{x_k\}$

Def: x_* is limit point of $\{x_k\}$ if for any $\delta > 0$ and any k_0, there is $k > k_0$ such that

$$\| x_k - x_* \| < \delta$$

Note: x_* is a limit point of $\{x_k\}$ then there is a subsequence $\{x_{k_j}\}$ that converges to x_*

Def: $\{x_k\}$ converges to x_* $\left(\lim_{k\to\infty} x_k = x_* \right)$ if, for any $\delta > 0$ $\| x_k - x_* \| < \delta$ for <u>all</u> sufficiently large k

Will base criteria on:

actual reduction in $\|F\|$:

$$\text{ared}_k \equiv \| F(x_k) \| - \| F(x_{k+1}) \|$$

predicted reduction in $\|F\|$

$$\text{pred}_k \equiv \| F(x_k) \| - \| F(x_k) + F'(x_k)\, s_k \|$$

where $s_k = x_{k+1} - x_k$

(predicted by local linear model $F(x_k) + F'(x_k) \cdot s_k$)

relative predicted reduction

$$\text{relpred}_k = \frac{\text{pred}_k}{\| F(x_k) \|}$$

- will require each $pred_k \geq 0$
- will also require

$$ared_k \geq t \cdot pred_k \quad \left(\begin{array}{c}\text{sufficient decrease}\\ \text{condition}\end{array}\right)$$

for $t \in (0,1)$ independent of k

- will not need to require a minimum $pred_k$ for every step, will require a minimum cumulative reduction over all steps

Theorem: (Eisenstat - Walker 1994)

Suppose F is continuously differentiable and $\{x_k\}$ is such that for each k (and $s_k = x_{k+1} - x_k$)

$$ared_k \geq t \cdot pred_k \geq 0$$

for $t \in (0,1)$ independent of k.

If $\sum_{k=1}^{\infty} relpred_k = \infty$, then $F(x_k) \to 0$

If also $\{x_k\} > 0$ has a limit point x_* such that $F'(x_*)$ is nonsingular, then $F(x_*) = 0$ and $x_k \to x_*$

Remarks: Theorem says exactly one of following must hold: (assuming $\sum_{k=1}^{\infty} relpred_k = \infty$)

1) $\|x_k\| \to \infty$ (no limit points)

2) $\exists$ limit points, and F' is singular at each

3) $x_k \to x_*$ with $F(x_*) = 0$, $F'(x_*)$ - nonsingular

The two "bad" cases can happen if F allows them to

example: $f(x) = e^{-x^2}$

for any $x_0 \neq 0$ Newton iterates $\{x_k\}$ satisfy

$ared_k \geq t \, pred_k \geq 0$ with $t = \frac{1}{4}$

for sufficiently large x_k, and

$x_k \to \pm \infty$

- Not hard to construct examples of $\{x_k\}$ with <u>multiple</u> limit points and F' singular at each (see Eisenstat – Walker '94)

- Also may not be possible to have $\sum \text{relpred}_k = \infty$

Example: $f(x) = 1 + x^2$

- If $\{x_k\}$ were such that $\sum \text{relpred}_k = \infty$ would have $f(x_k) \to 0$

<u>proof of theorem:</u>

easy part:

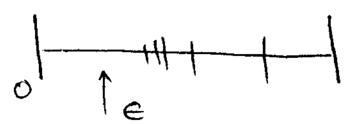
$$\mp \|F(x_1)\| \mp \|F(x_2)\| \mp \cdots \mp \|F(x_k)\|$$

$$\|F(x_0)\| \geq \|F(x_0)\| - \|F(x_{k+1})\|$$

$$= \sum_{j=0}^{k} \text{ared}_j \geq t \sum_{j=0}^{k} \text{pred}_j$$

$$= t \sum_{j=0}^{k} \text{relpred}_j \, \|F(x_j)\|$$

see: $\text{ared}_j \geq t\,\text{pred}_j \geq 0 \Rightarrow \|F(x_j)\|$ monotone decreasing



If $\|F(x_j)\| \not\to 0$ then $\exists\, \epsilon > 0$ such that

$$\|F(x_j)\| \geq \epsilon \quad \text{for all } j$$

Then

$$\|F(x_0)\| \geq t \cdot \epsilon \sum_{j=0}^{k} \text{relpred}_j \qquad \text{for all } k$$

$$\Rightarrow \sum_{j=0}^{\infty} \text{relpred}_j < \infty$$

so: $\sum \text{relpred}_j = \infty \Rightarrow F(x_j) \to \infty$

Hard part:

suppose $\sum \text{relpred}_k = \infty$ (so $F(x_k) \to 0$)

and x_* is a limit point of $\{x_k\}$ such that

$F'(x_*)$ is nonsingular.

Trivially $F(x_*)=0$ by continuity of F

To show: $x_k \to x_*$

By Handy Lemma (VI)

$$\| F(x) - \underbrace{F(x_*)}_{0} - F'(x_*)(x-x_*) \| \leq \omega\left(\|x-x_*\|\right) \|x-x_*\|$$

where

$$\omega(\delta) \equiv \max_{\|d\|<\delta} \| F'(x_*+d) - F'(x_*) \| \longrightarrow 0 \quad \text{as } \delta \to 0$$

Then

$$\omega\left(\|x-x_*\|\right)\|x-x_*\| \geqslant \| F'(x_*)(x-x_*) \| - \|F(x)\|$$

$$\Rightarrow \|F(x)\| \geqslant \|F'(x_*)(x-x_*)\| - \omega\left(\|x-x_*\|\right) \|x-x_*\|$$

#
$$\|V\| = \|A^{-1}AV\| \leq \|A^{-1}\| \, \|AV\|$$

$$\Rightarrow \|AV\| \geqslant \frac{\|V\|}{\|A^{-1}\|}$$

#

$$\geqslant \left[\frac{1}{K} - \omega\left(\|x-x_*\|\right)\right] \|x-x_*\| \quad , \quad K = \|F'(x_*)\|$$

Suppose $\delta > 0$ so small that $\omega(\delta) < \frac{1}{2K}$

Then $\|x-x_*\| \leq \delta \Rightarrow \|F(x)\| \geqslant \frac{1}{2K} \|x-x_*\|$

In particular, if $\|x_k - x_*\| \leq \delta$ then

$$\|F(x_k)\| \geqslant \frac{1}{2K} \|x_k - x_*\|$$

Since $F(x_k) \to 0$, only need to show

$\|x_k - x_*\| \leq \delta$ for all large k

Choose δ smaller if necessary so that

$$\|x - x_*\| \le \delta \Rightarrow \|F'(x)^{-1}\| \le 2k$$

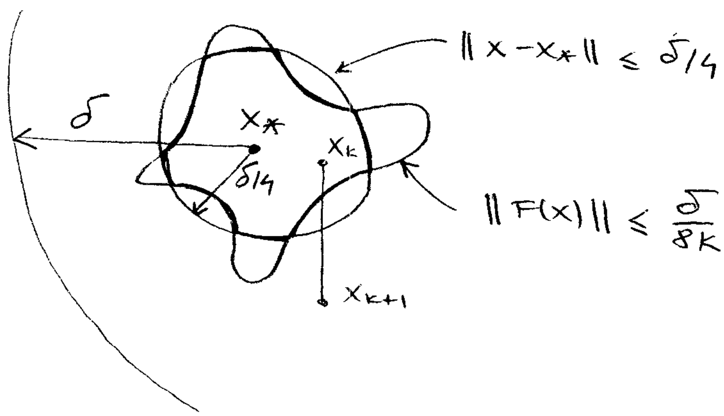
If x_k is such that $\|x_k - x_*\| \le \delta$ then

$$\|s_k\| = \|F'(x_k)^{-1}\{F'(x_k)s_k + F(x_k) - F(x_k)\}\|$$

$$\le \|F'(x_k)^{-1}\| \{ \|F(x_k)\| + \underbrace{\|F(x_k) + F'(x_k)s_k\|}_{\overset{\wedge\!\!|}{\|F(x_k)\|} \text{ since ared}_k \ge 0} \}$$

$$\le 2k \cdot 2\|F(x_k)\| = 4k\|F(x_k)\|$$

Set $S = \{x : \|x - x_*\| \le \frac{\delta}{4}$ and $\|F(x)\| \le \frac{\delta}{8k}\}$



Will show $x_k \in S$ for all sufficiently large k

Since x_* is a limit point $x_k \in S$ for some large k

Suffice to show: $x_k \in S \Rightarrow x_{k+1} \in S$

Then

$$\|s_k\| \le 4k\|F(x_k)\| \le 4k \frac{\delta}{8k} = \frac{\delta}{2}$$

$$\Rightarrow \|x_{k+1} - x_*\| \le \frac{\delta}{4} + \frac{\delta}{2} < \delta$$

But need to show $x_{k+1} \in S$

Have $\quad \|F(x_{k+1})\| \le \|F(x_k)\| = \dfrac{\delta}{8K}$

but also

$$\left.\begin{aligned} \|x_{k+1} - x_*\| &\le 2K\,\|F(x_{k+1})\| \\ &\le 2K\,\|F(x_k)\| \\ &\le 2K\,\frac{\delta}{8K} = \frac{\delta}{4} \end{aligned}\right\} \quad x_{k+1} \in S$$

$\sum_{k=0}^{\infty} \mathrm{relpred}_k = \infty$ may be difficult/impossible to verify

Won't need to do this – will construct algorithms that determine steps such that $\mathrm{ared}_k \ge t \cdot \mathrm{pred}_k \ge 0$ holds for each k and $\sum \mathrm{relpred}_k$ is automatically implicitly satisfied near points of interest.

Backtracking

Idea: For each k, initial trial step is

$$s_k = s_k^N = -F'(x_k)^{-1} F(x_k)$$

If $\mathrm{ared}_k \ge t \cdot \mathrm{pred}_k \ge 0$, accept s_k

If not, shorten s_k $\quad (s_k = \lambda s_k^N, \; 0 < \lambda < 1)$ and repeat test

see: If $s = \lambda s^N$, $0 < \lambda < 1$ then

$$\begin{aligned} \|F + F's\| &= \|F + F'\lambda s^N\| \\ &= \|(1-\lambda)F + \lambda\,\underbrace{(F + F's^N)}_{0}\| \\ &= \|(1-\lambda)F\| = (1-\lambda)\|F\| \end{aligned}$$

$$\Rightarrow \quad \mathrm{pred} = \|F\| - \|F + F's\| = \lambda\|F\|$$

also

$$\begin{aligned}
\text{ared} &= \|F\| - \|F(x+s)\| \\
&= \|F\| - \|F(x+s) - F - F's + F + F's\| \\
&\geq \underbrace{\|F\| - \|F+F's\|}_{\substack{\text{pred} \\ \| \\ \lambda\|F\|}} - \|F(x+s) - F - F's\|
\end{aligned}$$

$$\|F(x+s) - F - F's\| \leq \omega(\|\lambda s^N\|)\,\|\lambda s^N\| \leq$$

$$\leq \omega(\|\lambda s^N\|)\,\lambda\,\|F'(x)^{-1}\|\,\|F(x)\|$$

pred

$$\begin{aligned}
&\geq \lambda\|F\| - \omega(\lambda\|s^N\|)\,\|F'(x)^{-1}\|\,\lambda\,\|F(x)\| \\
&= \left(1 - \omega(\lambda\|s^N\|)\,\|F'(x)^{-1}\|\right)\text{pred} \\
&\geq t\cdot\text{pred} \quad \text{, for sufficiently small } \lambda
\end{aligned}$$

So a sufficiently short step $s = \lambda s^N$ will satisfy

$$\text{ared} \geq t\cdot\text{pred} \geq 0$$

Newton's method with backtracking
(this is "safeguarded backtracking)

Given $t \in (0,1)$ and $0 < \theta_{min} < \theta_{max} < 1$

Until termination do

  solve $F'(x)\cdot s = -F(x)$ for s^N

  while ared $< t$ pred do

    choose $\theta \in [\theta_{min}, \theta_{max}]$

    update $s \leftarrow \theta s$

  update $x \leftarrow x + s$

Note: after p passes through while-loop with $\theta_1, \dots, \theta_p$ the θ-values have $s = \lambda_p s^N$

$\lambda_p = \theta_1 \cdots \theta_p$

Have seen

$$\text{ared} \geq \left(1 - \omega\left(\lambda_p \|s^N\|\right) \|F'(x)^{-1}\|\right) \text{pred}$$

since

$$\theta_{min}^p \leq \lambda_p \leq \theta_{max}^p$$

will have ared $\geq t$ pred for some finite p

<u>Theorem:</u> Suppose the algorithm produces $\{x_k\}$. If x_* is a limit point of $\{x_k\}$ such that $F'(x_*)$ is nonsingular, then $F(x_*) = 0$ and $x_k \to x_*$ and $s_k = x_{k+1} - x_k = -F'(x_k)^{-1} F(x_k)$ for all sufficiently large k

Note: - Don't need $\sum \text{relpred}_k = \infty$

- Full Newton steps are ultimately taken near x_* such that $F'(x_*)$ nonsingular

Possibilities: Exactly one holds

(1) $\|x_k\| \longrightarrow \infty$

(2) $\{x_k\}$ has limit points and F' is singular at each

(3) $x_k \longrightarrow x_*$ such that $F(x_*) = 0$, $F'(x_*)$ nonsingular and convergence is that of Newton's method (superlinear or quadratic)

proof: Have $s_k = \lambda_k s_k^N$ for $0 < \lambda_k < 1$

and $\text{pred}_k = \lambda_k \|F_k\|$

$$\Rightarrow \text{relpred}_k = \frac{\text{pred}}{\|F_k\|} = \lambda_k$$

To show: $\sum_{k=1}^{\infty} \lambda_k = \infty$

Let $\delta > 0$ be sufficiently small that if $\|x - x_*\| \leq \delta$ then

i) $\|F'(x)^{-1}\| \leq 2k$ where $k = \|F'(x_*)^{-1}\|$

ii) $w(2\delta) \equiv \max\limits_{\|d\| \leq \delta} \|F'(x+d) - F'(x)\| \leq \dfrac{1-t}{2k}$

(ok since F' is uniformly continuous on δ-balls about x_*)

suppose $\|x_k - x_*\| < \delta$ and $\|x_{k+s} - x_*\| < \delta$,

$s = \lambda s_k^N$, $0 < \lambda < 1$

$$\begin{aligned}
\text{ared} &= \|F(x_k)\| - \|F(x_k+s)\| \\
&\geq \|F_k\| - \|F_k + F_k' s\| - \underbrace{\|F(x_k+s) - F_k - F_k' s\|}_{\leq\ w(\lambda\|s_k^N\|)\|F_k'^{-1}\| \underbrace{\lambda\|F_k\|}_{\text{pred}}} \\
&\geq \left(1 - \|F_k'^{-1}\|\, w(\|s\|)\right) \text{pred} \\
&\geq \left(1 - 2k\, w(2\delta)\right) \text{pred} \\
&\geq \left(1 - 2k \frac{1-t}{2k}\right) \text{pred} = t\ \text{pred}
\end{aligned}$$

Also,

$$\begin{aligned}
\|x_k - x_*\| \leq \delta \Rightarrow \|s_k^N\| &\leq \|-F'(x_k)^{-1} F(x_k)\| \\
&\leq 2k\,\|F(x_k)\| \\
&\leq 2kM
\end{aligned}$$

where $\|F(x)\| \leq M$ whenever $\|x - x_*\| \leq \delta$

Now if $\|x_k - x_*\| \leq \frac{\delta}{2}$ and $s = \lambda s_k^N$ with

$0 \leq \lambda \leq \lambda_* \equiv \min\left\{1, \frac{\delta}{4KM}\right\}$ then

$$\|s\| \leq \lambda_* \|s_k^N\| \leq \lambda_* 2KM \leq \frac{\delta}{2}$$

$$\Rightarrow \|x_k + s - x_*\| \leq \delta \quad \text{and} \quad \text{ared} \geq t\,\text{pred}$$

So if $\|x_k - x_*\| < \frac{\delta}{2}$, backtracking terminates with $s_k = \lambda_k s_k^N$ after at most p steps where $\theta_{max}^p \leq \lambda_*$

Then have $\theta_{min}^p \leq \lambda_k$ and so

$$\sum_k \text{relpred}_k = \sum_k \lambda_k \geq \sum_{k \in I} \lambda_k$$

$$I = \left\{k : \|x_k - x_*\| \leq \frac{\delta}{2}\right\}$$

$$\geq \sum_{k \in I} \theta_{min}^p = \infty$$

since $\|x_k - x_*\| \leq \frac{\delta}{2}$ for infinitely many k

Results follows from theorem

For k sufficiently large, both x_k and $x_k + s^N$ are within δ of x_*. Then ared $\geq$ t pred and $x_{k+1} = x_k + s_k^N$