15.093 Optimization Methods

Lecture 19: Line Searches
and Newton's Method

1 Last Lecture

Slide 1

- *Necessary Conditions for Optimality*
  (identifies candidates)
  x^* local min $\Rightarrow \nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ PSD
- *Sufficient Conditions for Optimality*
  $\nabla f(x^*) = 0$ $\nabla^2 f(x)$ psd $x \in B_\epsilon(x^*) \Rightarrow x^*$ loc. min.
- *Characterizations of Convexity*
  a. f convex $\Leftrightarrow f(y) \geq f(x) + \nabla f(x)'(y - x)$ $\forall x, y$
  b. f convex $\Leftrightarrow \nabla^2 f(x)$ PSD $\forall x$
- f convex then global min $\Leftrightarrow \nabla f(x^*) = 0$

2 Steepest descent

2.1 The algorithm

Slide 2

Step 0 Given $\boldsymbol{x}^0$, set $k := 0$.

Step 1 $\boldsymbol{d}^k := -\nabla f(\boldsymbol{x}^k)$. If $||\boldsymbol{d}^k|| \leq \epsilon$, then stop.

Step 2 Solve $\min_\lambda h(\lambda) := f(\boldsymbol{x}^k + \lambda \boldsymbol{d}^k)$ for the step-length λ^k, perhaps chosen by an exact or inexact line-search.

Step 3 Set $\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \lambda^k \boldsymbol{d}^k$, $k \leftarrow k + 1$.
Go to **Step 1**.

3 Outline

Slide 3

1. Bisection Method - Armijo's Rule
2. Motivation for Newton's method
3. Newton's method
4. Quadratic rate of convergence
5. Modification for global convergence

4 Choices of step sizes

Slide 4

- $Min_\lambda f(x^k + \lambda d^k)$
- Limited Minimization: $Min_{\lambda \in [0,s]} f(x^k + \lambda d^k)$
- Constant stepsize $\lambda^k = s$ constant

- Diminishing stepsize: $\lambda^k \to 0,\ \sum_k \lambda^k = \infty$
- Armijo Rule

4.1 Bisection Line- Search Algorithm

4.1.1 Convex functions

SLIDE 5

$$\bar{\lambda} := \arg\min_{\lambda} h(\lambda) := \arg\min_{\lambda} f(\bar{\boldsymbol{x}} + \lambda \bar{\boldsymbol{d}})$$

If $f(\boldsymbol{x})$ is convex, $h(\lambda)$ is convex.

Find $\bar{\lambda}$ for which $h'(\lambda) = 0$

$$h'(\lambda) = \nabla f(\bar{\boldsymbol{x}} + \lambda \bar{\boldsymbol{d}})' \, \bar{\boldsymbol{d}}$$

4.1.2 Algorithm

SLIDE 6

Step 0. Set $k = 0$. Set $\lambda_L := 0$ and $\lambda_U := \hat{\lambda}$.

Step k. Set $\tilde{\lambda} = \dfrac{\lambda_U + \lambda_L}{2}$ and compute $h'(\tilde{\lambda})$.
If $h'(\tilde{\lambda}) > 0$, re-set $\lambda_U := \tilde{\lambda}$. Set $k \leftarrow k + 1$.
If $h'(\tilde{\lambda}) < 0$, re-set $\lambda_L := \tilde{\lambda}$. Set $k \leftarrow k + 1$.
If $h'(\tilde{\lambda}) = 0$, stop.

4.1.3 Analysis

SLIDE 7

- At the kth iteration of the bisection algorithm, the current interval $[\lambda_L, \lambda_U]$ must contain a point $\bar{\lambda}$: $h'(\bar{\lambda}) = 0$
- At the k^{th} iteration of the bisection algorithm, the length of the current interval $[\lambda_L, \lambda_U]$ is

$$\text{length} = \left(\frac{1}{2}\right)^k (\hat{\lambda}).$$

- A value of λ such that $|\lambda - \bar{\lambda}| \leq \epsilon$ can be found in at most

$$\left\lceil \log_2 \left(\frac{\hat{\lambda}}{\epsilon}\right) \right\rceil$$

steps of the bisection algorithm.

4.1.4 Example

SLIDE 8

$$\begin{aligned} h(\lambda) &= (x_1 + \lambda d_1) - 0.6(x_2 + \lambda d_2) + 4(x_3 + \lambda d_3) + \\ &\quad +0.25(x_4 + \lambda d_4) - \sum_{i=1}^{4} \log(x_i + \lambda d_i) - \\ &\quad - \log\left((5 - \sum_{i=1}^{4} x_i - \lambda \sum_{i=1}^{4} d_i\right) \end{aligned}$$

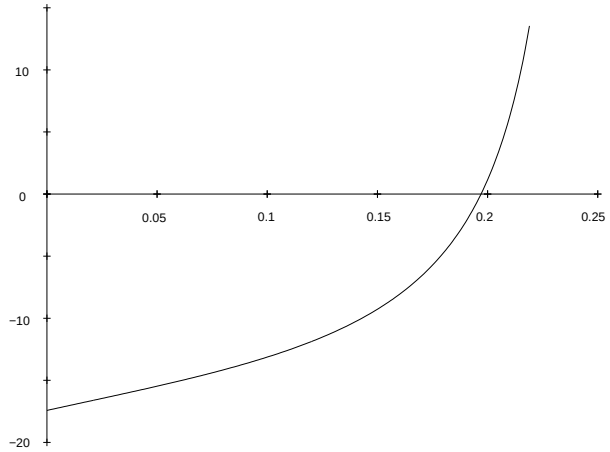
where $(x_1, x_2, x_3, x_4) = (1, 1, 1, 1)'$ and $(d_1, d_2, d_3, d_4) = (-1, 0.6, -4, -0.25)'$. SLIDE 9

$$\begin{aligned} h(\lambda) &= 4.65 - 17.4225\lambda - \log(1 - \lambda) - \log(1 + 0.6\lambda) - \\ &\quad - \log(1 - 4\lambda) - \log(1 - 0.25\lambda) - \log(1 + 4.65\lambda) \end{aligned}$$

and

$$\begin{aligned} h'(\lambda) &= -17.4225 + \frac{1}{1 - \lambda} - \frac{0.6}{1 + 0.6\lambda} + \frac{4}{1 - 4\lambda} + \\ &\quad + \frac{0.25}{1 - 0.25\lambda} - \frac{4.65}{1 + 4.65\lambda} \end{aligned}$$

SLIDE 10



SLIDE 11

k	λ_l^k	λ_u^k	$h'(\tilde{\lambda})$
1	0.0000000	1.0000000	NaN
2	0.0000000	0.5000000	NaN
3	0.0000000	0.2500000	−11.520429338348
4	0.1250000	0.2500000	−2.952901763683
5	0.1875000	0.2500000	13.286386294218
6	0.1875000	0.2187500	2.502969022220
7	0.1875000	0.2031250	−0.605144021505
8	0.1953125	0.2031250	0.831883373151
9	0.1953125	0.1992188	0.087369215988
10	0.1953125	0.1972656	−0.265032213496

SLIDE 12

k	λ_l^k	λ_u^k	$h'(\bar{\lambda})$
20	0.1970253	0.1970272	−0.000184301091
30	0.1970268	0.1970268	−0.000000146531
40	0.1970268	0.1970268	0.000000000189
41	0.1970268	0.1970268	0.000000000023
42	0.1970268	0.1970268	−0.000000000059
43	0.1970268	0.1970268	−0.000000000018
44	0.1970268	0.1970268	0.000000000003
45	0.1970268	0.1970268	−0.000000000008
46	0.1970268	0.1970268	−0.000000000002
47	0.1970268	0.1970268	0.000000000000

4.2 Armijo Rule

SLIDE 13

Bisection is accurate but may be expensive in practice
Need cheap method guaranteeing sufficient accuracy
Inexact line search method. Requires two parameters: $\epsilon \in (0,1)$, $\sigma > 1$.
$\bar{h}(\lambda) = h(0) + \lambda\epsilon h'(0)$
$\bar{\lambda}$ acceptable by Armijo's rule if:
• $h(\bar{\lambda}) \leq \bar{h}(\bar{\lambda})$
• $h(\sigma\bar{\lambda}) \geq \bar{h}(\sigma\bar{\lambda})$ (prevents the step size be small)
(i.e. $f(x^k + \bar{\lambda}d^k) \leq f(x^k) + \bar{\lambda}\epsilon\nabla f(x^k)'d^k$)

SLIDE 14

We get a range of acceptable stepsizes.
Step 0: Set $k = 0$, $\lambda^0 = \bar{\lambda} > 0$
Step k: If $h(\lambda^k) \leq \bar{h}(\lambda^k)$, choose λ^k stop. If $h(\lambda^k) > \bar{h}(\lambda^k)$ let $\lambda^{k+1} = \frac{1}{\sigma}\lambda^k$
(for example, $\sigma = 2$)

5 Newton's method

5.1 History

SLIDE 15

Steepest Descent is simple but slow
Newton's method complex but fast
Origins not clear

Raphson became member of the Royal Society in 1691 for his book "Analysis Aequationum Universalis" with Newton method.
Raphson published it 50 years before Newton.

6 Newton's method

6.1 Motivation

SLIDE 16

Consider

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

- Taylor series expansion around $\bar{\boldsymbol{x}}$

$$f(\boldsymbol{x}) \approx g(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})'(x - \bar{x}) +$$

$$\frac{1}{2}(x - \bar{x})'\nabla^2 f(\bar{x})(x - \bar{x})$$

- Instead of $\min f(\boldsymbol{x})$, solve $\min g(\boldsymbol{x})$, i.e., $\nabla g(\boldsymbol{x}) = \mathbf{0}$
- $$\nabla f(\bar{\boldsymbol{x}}) + \nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) = \mathbf{0}$$

$$\Rightarrow \boldsymbol{x} - \bar{\boldsymbol{x}} = -(\nabla^2 f(\bar{\boldsymbol{x}}))^{-1}\nabla f(\bar{\boldsymbol{x}})$$

- The direction $\boldsymbol{d} = -(\nabla^2 f(\bar{\boldsymbol{x}}))^{-1}\nabla f(\bar{\boldsymbol{x}})$ is the Newton direction

6.2 The algorithm

SLIDE 17

Step 0 Given $\boldsymbol{x}^0$, set $k := 0$.

Step 1 $\boldsymbol{d}^k := -(\nabla^2 f(\boldsymbol{x}^k))^{-1}\nabla f(\boldsymbol{x}^k)$.
If $||\boldsymbol{d}^k|| \leq \epsilon$, then stop.

Step 2 Set $\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \boldsymbol{d}^k$, $k \leftarrow k + 1$.
Go to **Step 1**.

6.3 Comments

SLIDE 18

- The method assumes that $\nabla^2 f(\boldsymbol{x}^k)$ is nonsingular at each iteration
- There is no guaranteee that $f(\boldsymbol{x}^{k+1}) \leq f(\boldsymbol{x}^k)$
- We can augment the algorithm with a line-search:

$$\min_{\lambda} f(\boldsymbol{x}^k + \lambda \boldsymbol{d}^k)$$

6.4 Properties

SLIDE 19

Theorem If $\boldsymbol{H} = \nabla^2 f(\boldsymbol{x}^k)$ is PD, then $\boldsymbol{d}^k$ is a descent direction: $\nabla f(\boldsymbol{x}^k)'\boldsymbol{d}^k < 0$
Proof:

$$\nabla f(\boldsymbol{x}^k)'\boldsymbol{d}^k = -\nabla f(\boldsymbol{x}^k)'\boldsymbol{H}^{-1}\nabla f(\boldsymbol{x}^k) < 0$$

if $\boldsymbol{H}^{-1}$ is PD. But,

$$0 < \boldsymbol{v}'(\boldsymbol{H}^{-1})'\boldsymbol{H}\boldsymbol{H}^{-1}\boldsymbol{v} = \boldsymbol{v}'\boldsymbol{H}^{-1}\boldsymbol{v}$$

$$\Rightarrow \boldsymbol{H}^{-1} \text{ is PD}$$

6.5 Example 1

SLIDE 20

$$f(x) = 7x - \log x, \qquad x^* = \frac{1}{7} = 0.14857143$$

$$f'(x) = 7 - \frac{1}{x}, \qquad f''(x) = \frac{1}{x^2}$$

$$\Rightarrow d = -(f''(x))^{-1} f'(x) = -\left(\frac{1}{x^2}\right)^{-1}\left(7 - \frac{1}{x}\right) = x - 7x^2$$

$$x^{k+1} = x^k + \left(x^k - 7(x^k)^2\right) = 2x^k - 7(x^k)^2$$

SLIDE 21

k	x^k	x^k	x^k	x^k
0	1	0	0.01	0.1
1	-5	0	0.0193	0.13
2	-185	0	0.03599	0.1417
3	-239945	0	0.062917	0.14284777
4	-4E11	0	0.098124	0.142857142
5	-112E22	0	0.128849782	0.142857143
6		0	0.141483700	0.142857143
7		0	0.142843938	0.142857143
8		0	0.142857142	0.142857143

6.6 Example 2

SLIDE 22

$$f(x_1, x_2) = -\log(1 - x_1 - x_2) - \log x_1 - \log x_2$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} \dfrac{1}{1 - x_1 - x_2} - \dfrac{1}{x_1} \\ \dfrac{1}{1 - x_1 - x_2} - \dfrac{1}{x_2} \end{bmatrix}$$

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \left(\dfrac{1}{1 - x_1 - x_2}\right)^2 + \left(\dfrac{1}{x_1}\right)^2 & \left(\dfrac{1}{1 - x_1 - x_2}\right)^2 \\ \left(\dfrac{1}{1 - x_1 - x_2}\right)^2 & \left(\dfrac{1}{1 - x_1 - x_2}\right)^2 + \left(\dfrac{1}{x_1}\right)^2 \end{bmatrix}$$

SLIDE 23

$$(x_1^*, x_2^*) = \left(\frac{1}{3}, \frac{1}{3}\right), \qquad f(x_1^*, x_2^*) = 3.295836867$$

k	x_1^k	x_2^k	$\|\|\boldsymbol{x}-\boldsymbol{x}^*\|\|$
0	0.85	0.05	0.58925565
1	0.7170068	0.09659864	0.45083106
2	0.5129752	0.17647971	0.23848325
3	0.35247858	0.27324878	0.06306103
4	0.33844902	0.32623807	0.00874717
5	0.33333772	0.33325933	7.4133E-05
6	0.33333334	0.33333333	1.1953E-08
7	0.33333333	0.33333333	1.5701E-16

7 Quadratic convergence

SLIDE 24

- Recall

$$z = \lim_{n\to\infty} z_n, \qquad \limsup_{n\to\infty} \frac{|z_{n+1}-z|}{|z_n-z|^2} = \delta < \infty$$

- Example: $z_n = a^{2^n}$, $0 < a < 1$, $z = 0$

$$\frac{|z_{n+1}-z|}{|z_n-z|^2} = \frac{a^{2^{n+1}}}{a^{2^{n+1}}} = 1$$

7.1 Intuitive statement

SLIDE 25

Theorem Suppose $f(\boldsymbol{x}) \in C^3$ (thrice cont. diff/ble), $\boldsymbol{x}^*$: $\nabla f(x^*) = 0$ and $\nabla^2 f(\boldsymbol{x}^*)$ is nonsingular. If Newton's method is started sufficiently close to $\boldsymbol{x}^*$, the sequence of iterates converges quadratically to $\boldsymbol{x}^*$.

SLIDE 26

Lemma:
Suppose $f(\boldsymbol{x}) \in C^3$ on $\Re^n$ and x a given pt.
Then $\forall \epsilon > 0$, $\exists \beta > 0$: $||x-y|| \le \epsilon \Rightarrow$

$$||\nabla f(x) - \nabla f(y) - H(y)(x-y)|| \le \beta ||x-y||^2$$

7.2 Formal statement

SLIDE 27

Theorem
Suppose f twice cont. diff/ble. Let $\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$, $\boldsymbol{H}(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x})$, and $\boldsymbol{x}^* : \boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}$. Suppose $f(\boldsymbol{x})$ is twice differentiable and its Hessian satisfies:

- $||\boldsymbol{H}(\boldsymbol{x}^*)|| \ge h$
- $||\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{x}^*)|| \le L||\boldsymbol{x}-\boldsymbol{x}^*||$, $\forall \boldsymbol{x} : ||\boldsymbol{x}-\boldsymbol{x}^*|| \le \beta$

(Reminder: If $\boldsymbol{A}$ is a square matrix:

$$||\boldsymbol{A}|| = \max_{\boldsymbol{x}:||\boldsymbol{x}||=1} ||\boldsymbol{A}\boldsymbol{x}|| = \max\{|\lambda_1|, \ldots, |\lambda_n|\})$$

Suppose SLIDE 28

$$||\boldsymbol{x} - \boldsymbol{x}^*|| \leq \min\left(\beta, \frac{2h}{3L}\right) = \gamma.$$

Let $\bar{\boldsymbol{x}}$ be the next iterate in the Newton method. Then,

$$||\bar{\boldsymbol{x}} - \boldsymbol{x}^*|| \leq ||\boldsymbol{x} - \boldsymbol{x}^*||^2 \left(\frac{2h}{3L}\right) \quad \text{and}$$

$$||\bar{\boldsymbol{x}} - \boldsymbol{x}^*|| < ||\boldsymbol{x} - \boldsymbol{x}^*|| < \gamma.$$

7.3 Proof

SLIDE 29

$\bar{\boldsymbol{x}} - \boldsymbol{x}^*$

$$\begin{aligned}
&= \boldsymbol{x} - \boldsymbol{H}(\boldsymbol{x})^{-1}\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{x}^* \\
&= \boldsymbol{x} - \boldsymbol{x}^* + \boldsymbol{H}(\boldsymbol{x})^{-1}\Big(\boldsymbol{g}(\boldsymbol{x}^*) - \boldsymbol{g}(\boldsymbol{x})\Big) \\
&= \boldsymbol{x} - \boldsymbol{x}^* + \boldsymbol{H}(\boldsymbol{x})^{-1} \\
&\quad \int_0^1 \boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x^*} - \boldsymbol{x}))(\boldsymbol{x^*} - \boldsymbol{x})dt \quad \text{(Lemma 1)} \\
&= \boldsymbol{H}(\boldsymbol{x})^{-1} \int_0^1 \left(\boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x^*} - \boldsymbol{x})) - \boldsymbol{H}(\boldsymbol{x})\right)(\boldsymbol{x^*} - \boldsymbol{x})dt
\end{aligned}$$

SLIDE 30

$||\bar{\boldsymbol{x}} - \boldsymbol{x}^*||$

$$\begin{aligned}
&\leq ||\boldsymbol{H}(\boldsymbol{x})^{-1}|| \\
&\quad \int_0^1 \left|\left|\boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x^*} - \boldsymbol{x})) - \boldsymbol{H}(\boldsymbol{x})\right|\right| \cdot ||\boldsymbol{x} - \boldsymbol{x^*}||dt \\
&\leq ||\boldsymbol{H}(\boldsymbol{x})^{-1}|| \cdot ||\boldsymbol{x} - \boldsymbol{x^*}|| \int_0^1 L||\boldsymbol{x} - \boldsymbol{x^*}||\ t\ dt \\
&= \frac{1}{2}L||\boldsymbol{H}(\boldsymbol{x})^{-1}|| \cdot ||\boldsymbol{x} - \boldsymbol{x^*}||^2
\end{aligned}$$

SLIDE 31

Now

$$\begin{aligned}
h &\leq ||\boldsymbol{H}(\boldsymbol{x}^*)|| \\
&= ||\boldsymbol{H}(\boldsymbol{x}) + \boldsymbol{H}(\boldsymbol{x}^*) - \boldsymbol{H}(\boldsymbol{x})|| \\
&\leq ||\boldsymbol{H}(\boldsymbol{x})|| + ||\boldsymbol{H}(\boldsymbol{x}^*) - \boldsymbol{H}(\boldsymbol{x})|| \\
&\leq ||\boldsymbol{H}(\boldsymbol{x})|| + L||\boldsymbol{x}^* - \boldsymbol{x}|| \\
&\Rightarrow ||\boldsymbol{H}(\boldsymbol{x})|| \geq h - L||\boldsymbol{x} - \boldsymbol{x}^*||
\end{aligned}$$

SLIDE 32

$$\Rightarrow ||\boldsymbol{H}(\boldsymbol{x})^{-1}|| \leq \frac{1}{||\boldsymbol{H}(\boldsymbol{x})||} \leq \frac{1}{h - L||\boldsymbol{x} - \boldsymbol{x}^*||}$$

$$
\begin{aligned}
\Rightarrow ||\bar{\boldsymbol{x}} - \boldsymbol{x}^*|| &\leq ||\boldsymbol{x} - \boldsymbol{x}^*||^2 \frac{L}{2(h - L||\boldsymbol{x} - \boldsymbol{x}^*||)} \\
&\leq ||\boldsymbol{x} - \boldsymbol{x}^*||^2 \frac{L}{2(h - \frac{2}{3}h)} \\
&= \frac{3L}{2h} ||\boldsymbol{x} - \boldsymbol{x}^*||^2
\end{aligned}
$$

7.4 Lemma 1

Slide 33

- Fix $\boldsymbol{w}$. Let $\phi(t) = \boldsymbol{g}(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))'\boldsymbol{w}$
- $\phi(0) = \boldsymbol{g}(\boldsymbol{x})'\boldsymbol{w}, \quad \phi(1) = \boldsymbol{g}(\boldsymbol{x}^*)'\boldsymbol{w}$
- $\phi'(t) = \boldsymbol{w}'\boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))(\boldsymbol{x}^* - \boldsymbol{x})$
- $\phi(1) = \phi(0) + \int_0^1 \phi'(t)dt \Rightarrow$

Slide 34

$$
\forall \boldsymbol{w}: \quad \boldsymbol{w}'(\boldsymbol{g}(\boldsymbol{x}^*) - \boldsymbol{g}(\boldsymbol{x})) =
$$

$$
\boldsymbol{w}' \int_0^1 \boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))(\boldsymbol{x}^* - \boldsymbol{x})dt
$$

$$
\Rightarrow \boldsymbol{g}(\boldsymbol{x}^*) - \boldsymbol{g}(\boldsymbol{x}) = \int_0^1 \boldsymbol{H}(\boldsymbol{x} + t(\boldsymbol{x}^* - \boldsymbol{x}))(\boldsymbol{x}^* - \boldsymbol{x})dt
$$

7.5 Critical comments

Slide 35

- The iterates from Newton's method are equally attracted to local min and local max.
- We do not know $\beta,\ h,\ L$ in general.
- Note, however, that they are only used in the proof, not in the algorithm.
- We do not assume convexity, only that $H(\boldsymbol{x}^*)$ is nonsingular and not badly behaved near $\boldsymbol{x}^*$.

Slide 36

7.6 Properties of Convergence

Proposition:

Let $r^k = ||x^k - x^*||$ and $C = \frac{3L}{2h}$. If $\|x^0 - x^*|| < \gamma$ then

$$
r^k \leq \frac{1}{C}(Cr^0)^{2^k}
$$

Slide 37

Proposition:

If $||x^0 - x^*|| > \epsilon > 0 \Rightarrow ||x^k - x^*|| < \epsilon$

$$\forall k \geq \left\lceil \frac{\log(\frac{\log(\frac{1}{C\epsilon})}{\log(\frac{1}{Cr^0})}}{\log 2} \right\rceil$$

8 Solving systems of equations

Slide 38

$$\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$$

$$\boldsymbol{g}(\boldsymbol{x}^{t+1}) \approx \boldsymbol{g}(\boldsymbol{x}^t) + \nabla \boldsymbol{g}(\boldsymbol{x}^t)(\boldsymbol{x}^{t+1} - \boldsymbol{x}^t) \Rightarrow$$

$$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - (\nabla \boldsymbol{g}(\boldsymbol{x}^t))^{-1} \boldsymbol{g}(\boldsymbol{x}^t)$$

Application in optimization: $\boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$

9 Modifications for global convergence

Slide 39

- Perform line search
- When Hessian is singular or near singular, use:

$$\boldsymbol{d}^k = -\left(\nabla^2 f(\boldsymbol{x}^k) + \boldsymbol{D}^k\right)^{-1} \nabla f(\boldsymbol{x}^k)$$

- $\boldsymbol{D}^k$ is a diagonal matrix:

$$\nabla^2 f(\boldsymbol{x}^k) + \boldsymbol{D}^k \text{ is PD}$$

10 Summary

Slide 40

1. Line search methods:
   Bisection Method.
   Armijo's Rule.
2. The Newton's method
3. Quadratic rate of convergence
4. Modification for global convergence

15.093J / 6.255J Optimization Methods

Fall 2009