Optimization Theory and Methods

Prof. Xiaoe Ruan

Email: wruanxe@xjtu.edu.cn

Tel:13279321898

2019， Fall

Chap.3 Unconstrained Optimization Methods

Global Convergence of Descent Algorithm

Steepest Descent Algorithm and Newton's Method

Quasi-Newton's Method

1. Global Convergence of Descent Algorithm

(1).Descent Algorithm

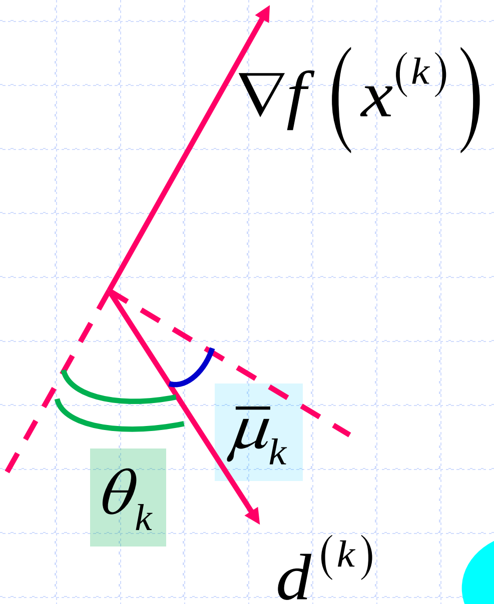
- **Construct sequence** $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$

s.t. $f\left(x^{(k+1)}\right) = f\left(x^{(k)} + \alpha_k d^{(k)}\right) < f\left(x^{(k)}\right).$

Because of

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = f\left(x^{(k)}\right) + \alpha_k \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)} + o\left(\left\|\alpha_k d^{(k)}\right\|\right).$$

Then $-\varepsilon < \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)} < 0$



Let $$\cos\theta_k = \frac{-\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}}{\left\|\nabla f\left(x^{(k)}\right)\right\| \left\|d^{(k)}\right\|}$$

Then $0 \le \theta_k \le \frac{\pi}{2} - \bar{\mu}_k, (\bar{\mu}_k > 0), k = 1, 2, \cdots$

Th.1. Let $\nabla f(x)$ be consistently continuous on level set $L\left(x^{(0)}\right)=\left\{x \mid f(x)\le f\left(x^{(0)}\right)\right\}$ and $0\le\theta_k\le\frac{\pi}{2}-\bar{\mu},(\bar{\mu}>0),k=1,2,\cdots$

$$f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\ge 0} f\left(x^{(k)}+\alpha d^{(k)}\right).$$

Iterative sequence $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)},k=0,1,\cdots.$

Then （1） There exists k, s.t. $\nabla f\left(x^{(k)}\right)=0$

or （2） $f\left(x^{(k)}\right)\to-\infty(k\to\infty)$

or （3） $\nabla f\left(x^{(k)}\right)\to 0(k\to\infty)$

Proof: If both （1） and （2） are not true.

Then $\nabla f\left(x^{(k)}\right)\ne 0(k=0,1,\cdots)$ and $-M\le f\left(x^{(k)}\right)$

Prove （3） following.

As $-M < \cdots < f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$, **then** $\lim\limits_{k\to\infty} f\left(x^{(k)}\right)$ **exists.**

That is $\left\| f\left(x^{(k)}\right) - f\left(x^{(k+1)}\right) \right\| \to 0 \left(k \to \infty\right)$ **(**)**

Assume that, by contradiction, （3） $\nabla f\left(x^{(k)}\right) \to 0$ **does not hold.**

Then there exists subsequence $\nabla f\left(x^{(k_i)}\right)$ **s.t.** $\left\|\nabla f\left(x^{(k_i)}\right)\right\| > \varepsilon$

Thus

$$\frac{-\nabla f\left(x^{(k_i)}\right)^{\mathrm{T}} d^{(k_i)}}{\left\|d^{(k_i)}\right\|} = \left\|\nabla f\left(x^{(k_i)}\right)\right\| \cos\theta_{k_i}$$

$$\geq \left\|\nabla f\left(x^{(k_i)}\right)\right\| \cos\left(\frac{\pi}{2} - \bar{\mu}\right) = \left\|\nabla f\left(x^{(k_i)}\right)\right\| \sin\bar{\mu} > \varepsilon \sin\bar{\mu} \triangleq \varepsilon_1$$

Then

$$f\left(x^{(k_i)} + \alpha d^{(k_i)}\right) = f\left(x^{(k_i)}\right) + \alpha \nabla f\left(x^{(k_i)} + \theta\alpha d^{(k_i)}\right)^{\mathrm{T}} d^{(k_i)}$$

$$= f\left(x^{(k_i)}\right) + \alpha \nabla f\left(x^{(k_i)}\right)^{\mathrm{T}} d^{(k_i)} + \alpha\left[\nabla f\left(x^{(k_i)} + \theta\alpha d^{(k_i)}\right) - \nabla f\left(x^{(k_i)}\right)\right]^{\mathrm{T}} d^{(k_i)}$$

From consistent continuity of $\nabla f(x)$ **on level set** $L\left(x^{(0)}\right)$ **Yields,**

if $\alpha\left\|d^{(k_i)}\right\| \le \bar{\alpha}$ **then** $\left\|\nabla f\left(x^{(k_i)} + \theta\alpha d^{(k_i)}\right) - \nabla f\left(x^{(k_i)}\right)\right\| \le \frac{1}{2}\varepsilon_1.$

Let $\alpha = \frac{\bar{\alpha}}{\left\|d^{(k_i)}\right\|}$ **Then** $f\left(x^{(k_i)} + \frac{\bar{\alpha}}{\left\|d^{(k_i)}\right\|}d^{(k_i)}\right) - f\left(x^{(k_i)}\right) \le -\bar{\alpha}\varepsilon_1 + \frac{1}{2}\bar{\alpha}\varepsilon_1 = -\frac{1}{2}\bar{\alpha}\varepsilon_1.$

That is $$\left\|f\left(x^{(k_i)} + \frac{\bar{\alpha}}{\left\|d^{(k_i)}\right\|}d^{(k_i)}\right) - f\left(x^{(k_i)}\right)\right\| > \frac{1}{2}\bar{\alpha}\varepsilon_1.$$

As $$f\left(x^{(k_i)} + \alpha_{k_i}d^{(k_i)}\right) \le f\left(x^{(k_i)} + \frac{\bar{\alpha}}{\left\|d^{(k_i)}\right\|}d^{(k_i)}\right),$$

Therefore $$\left\|f\left(x^{(k_i)} + \alpha_{k_i}d^{(k_i)}\right) - f\left(x^{(k_i)}\right)\right\| > \frac{1}{2}\bar{\alpha}\varepsilon_1.$$

Which contradicts to (**).

(2).Convergence Rate

Lemma1 Let $\varphi(\alpha)$ be 2nd-order continuously differentiable on $[0,b]$. $\varphi'(0)<0$ and $\exists M>0$ s.t. $\varphi''(\alpha)\le M, \forall \alpha\in[0,b]$.

If $\alpha^*\in(0,b)$ is the unique minimizer of $\varphi(\alpha)$ then $\alpha^*\ge -\dfrac{\varphi'(0)}{M}$.

Proof: Let $\bar{\alpha}=-\dfrac{\varphi'(0)}{M}$. From $\varphi(\alpha)=\varphi(0)+\varphi'(0)\alpha+\dfrac{1}{2}\varphi''(\xi)\alpha^2$

yields
$$\varphi(\alpha^*)=\varphi(0)+\varphi'(0)\alpha^*+\frac{1}{2}\varphi''(\xi_1)\alpha^{*2} \qquad (1)$$

Further from
$$\varphi(\alpha)=\varphi(\alpha^*)+\frac{1}{2}\varphi''(\eta)(\alpha-\alpha^*)^2$$

makes
$$\varphi(0)=\varphi(\alpha^*)+\frac{1}{2}\varphi''(\xi_2)\alpha^{*2} \qquad (2)$$

(1)+(2) delivers
$$0<\alpha^*=-\frac{2\varphi'(0)}{\varphi''(\xi_1)+\varphi''(\xi_2)}\ge -\frac{\varphi'(0)}{M}=\bar{\alpha}$$

Lemma2 Let x^* be a minimizer of $f(x)$. If $f(x)$ is 2nd-order continuously differentiable in $N_\delta(x^*)$ and exist $\varepsilon > 0$, $0 < m < M$, s.t. $m\|u\|^2 \le u^{\mathrm{T}} \nabla^2 f(x) u \le M\|u\|^2$ for $\forall u \in R^n$ and $\|x - x^*\| < \varepsilon$

Then $\frac{1}{2} m\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2} M\|x - x^*\|^2$, $\|\nabla f(x)\| \ge m\|x - x^*\|$.

Proof: $$f(x) = f(x^*) + \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(\xi_1)(x - x^*)$$

Then $$\frac{1}{2} m\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2} M\|x - x^*\|^2,$$

From $$f(x^*) = f(x) + \nabla f(x)^{\mathrm{T}}(x^* - x) + \frac{1}{2}(x^* - x)^{\mathrm{T}} \nabla^2 f(\xi_2)(x^* - x)$$

makes $$\nabla f(x)^{\mathrm{T}}(x - x^*) = f(x) - f(x^*) + \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(\xi_2)(x - x^*)$$

Therefore $$\|\nabla f(x)\| \ge m\|x - x^*\|.$$

Th.2. **Let** $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right)$. $f\left(x^*\right)=\min_{x\in N\left(x^*\right)} f\left(x\right)$.

Iterative sequence $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}$ **satisfies** $\lim_{k\to\infty} x^{(k)}=x^*$.

If $f\left(x\right)$ **is 2nd-orde continuously differentiable in** $N_\delta\left(x^*\right)$

and there exist $\varepsilon>0,\quad 0<m<M,$ **s.t.**

$m\|u\|^2\leq u^{\mathrm{T}}\nabla^2 f\left(x\right)u\leq M\|u\|^2$ **for** $\forall u\in R^n$ **and** $\left\|x-x^*\right\|<\varepsilon$. **Then**

$\left\{f\left(x^{(k)}\right)\right\}$ **is Q-linearly convergent.** $\left\{x^{(k)}\right\}$ **is R-linearly convergent.**

Proof: Let $\left\{x^{(k)}\right\}\in N_\varepsilon\left(x^*\right),\quad \varphi(\alpha)=f\left(x^{(k)}+\alpha d^{(k)}\right)$.

Then there exists $0<\delta<<1$ **s.t.** $x^{(k)}+\left(\alpha_k+\delta\right)d^{(k)}\in N_\varepsilon\left(x^*\right)$.

where $\varphi(\alpha_k)=\min\varphi(\alpha),\quad \varphi'(0)=\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}<0,$

$$\varphi''(\alpha)=d^{(k)\mathrm{T}}\nabla^2 f\left(x^{(k)}+\alpha d^{(k)}\right)d^{(k)}<M\left\|d^{(k)}\right\|^2.$$

From Lemma1 conducts

$$\alpha_k \geq \frac{-\varphi'(0)}{M\left\|d^{(k)}\right\|^2} \geq \frac{\left\|\nabla f\left(x^{(k)}\right)\right\|\left\|d^{(k)}\right\| \sin\bar{\mu}}{M\left\|d^{(k)}\right\|^2} = \frac{\left\|\nabla f\left(x^{(k)}\right)\right\| \sin\bar{\mu}}{M\left\|d^{(k)}\right\|} \triangleq \bar{\alpha}_k$$

Let $\bar{x}^{(k)} = x^{(k)} + \bar{\alpha}_k d^{(k)}$ **Then** $\bar{x}^{(k)} \in N\left(x^*, \varepsilon\right)$ **Thus**

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) - f\left(x^{(k)}\right) \leq f\left(x^{(k)} + \bar{\alpha}_k d^{(k)}\right) - f\left(x^{(k)}\right)$$

$$= \bar{\alpha}_k \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)} + \frac{1}{2}\bar{\alpha}_k^2 d^{(k)\mathrm{T}} \nabla^2 f\left(\xi^{(k)}\right) d^{(k)}$$

$$\leq -\bar{\alpha}_k \left\|\nabla f\left(x^{(k)}\right)\right\|\left\|d^{(k)}\right\| \sin\bar{\mu} + \frac{1}{2}\bar{\alpha}_k^2 \left\|d^{(k)}\right\|^2 M$$

$$\leq -\frac{\left\|\nabla f\left(x^{(k)}\right)\right\|^2 \left\|d^{(k)}\right\| \sin^2\bar{\mu}}{M\left\|d^{(k)}\right\|} + \frac{1}{2}\frac{\left\|\nabla f\left(x^{(k)}\right)\right\|^2 \sin^2\bar{\mu}\left\|d^{(k)}\right\|^2 M}{M^2\left\|d^{(k)}\right\|^2}$$

$$= -\frac{1}{2}\frac{\left\|\nabla f\left(x^{(k)}\right)\right\|^2 \sin^2\bar{\mu}}{M} \le -\frac{1}{2}\frac{m^2\sin^2\bar{\mu}}{M}\left\|x^{(k)}-x^*\right\|^2$$

$$\le -\frac{m^2\sin^2\bar{\mu}}{M^2}\left[f\left(x^{(k)}\right)-f\left(x^*\right)\right]$$

Therefore $$f\left(x^{(k+1)}\right)-f\left(x^*\right) \le \left(1-\frac{m^2\sin^2\bar{\mu}}{M^2}\right)\left[f\left(x^{(k)}\right)-f\left(x^*\right)\right]$$

That means $\left\{f\left(x^{(k)}\right)\right\}$ **is Q-linearly convergent.** **And**

$$\frac{1}{2}m\left\|x^{(k+1)}-x^*\right\|^2 \le f\left(x^{(k+1)}\right)-f\left(x^*\right) \le \left(1-\frac{m^2\sin^2\bar{\mu}}{M^2}\right)^{k+1}\left[f\left(x^{(0)}\right)-f\left(x^*\right)\right]$$

Then

$$\sqrt[k+1]{\left\|x^{(k+1)}-x^*\right\|} \le \sqrt[2(k+1)]{\frac{2}{m}\left[f\left(x^{(0)}\right)-f\left(x^*\right)\right]}\sqrt{1-\frac{m^2\sin^2\bar{\mu}}{M^2}} \to \sqrt{1-\frac{m^2\sin^2\bar{\mu}}{M^2}} < 1$$

Therefore $\left\{x^{(k)}\right\}$ **is R-linearly convergent.**

Th.3. **Iterative sequence** $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$

$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min\limits_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right).$ **If** $f\left(x\right)$ **is consistently convex.**

i.e. $\exists \eta > 0,$ **s.t.** $\left[\nabla f\left(x\right) - \nabla f\left(y\right)\right]^{\mathrm{T}} \left(x - y\right) \geq \eta \left\| x - y \right\|^2, \forall x, y \in R^n.$

Then $f\left(x^{(k)}\right) - f\left(x^{(k+1)}\right) \geq \frac{1}{2} \eta \left\| \alpha_k d^{(k)} \right\|^2.$

Proof: **Let** $\varphi\left(\alpha\right) = f\left(x^{(k)} + \alpha d^{(k)}\right)$ **Then** $\varphi'\left(\alpha\right) = \nabla f\left(x^{(k)} + \alpha d^{(k)}\right)^{\mathrm{T}} d^{(k)}.$

$\varphi\left(0\right) - \varphi\left(\alpha_k\right) = -\int_0^{\alpha_k} \varphi'\left(\alpha\right) d\alpha.$ $\varphi'\left(\alpha_k\right) = \nabla f\left(x^{(k)} + \alpha_k d^{(k)}\right)^{\mathrm{T}} d^{(k)} = 0.$

Thus

$$f\left(x^{(k)}\right) - f\left(x^{(k)} + \alpha_k d^{(k)}\right) = -\int_0^{\alpha_k} \nabla f\left(x^{(k)} + \alpha d^{(k)}\right)^{\mathrm{T}} d^{(k)} d\alpha.$$

$$= \int_0^{\alpha_k} \frac{\left[\nabla f\left(x^{(k)} + \alpha_k d^{(k)}\right) - \nabla f\left(x^{(k)} + \alpha d^{(k)}\right)\right]^{\mathrm{T}} \left(\alpha_k - \alpha\right) d^{(k)}}{\left(\alpha_k - \alpha\right)} d\alpha.$$

$$\geq \int_0^{\alpha_k} \eta \left(\alpha_k - \alpha\right) \left\| d^{(k)} \right\|^2 d\alpha = \frac{1}{2} \eta \left\| \alpha_k d^{(k)} \right\|^2.$$

2. Steepest Descent Algorithm and Newton's Method

(1). Steepest Descent Algorithm

Let iterative sequence $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$

where $f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min\limits_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right).$

Remind

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = f\left(x^{(k)}\right) + \alpha_k \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)} + o\left(\left\|\alpha_k d^{(k)}\right\|\right)$$

Therefore **For the case when** $d^{(k)} = -\nabla f\left(x^{(k)}\right)$

the descent quantity of objective fcn is maximal.

Then, Steepest Descent Algorithm :

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right), k = 0,1,\cdots.$$

with $f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min\limits_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right).$

Lemma3 **Let** $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min\limits_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right)$. **If** $\exists M>0$,

s.t. $\left\|\nabla^2 f\left(x^{(k)}+\alpha d^{(k)}\right)\right\|\leq M, \forall\alpha>0, k=1,2,\cdots$. **Then**

$$f\left(x^{(k)}\right)-f\left(x^{(k)}+\alpha_k d^{(k)}\right)\geq\frac{1}{2M}\left\|\nabla f\left(x^{(k)}\right)\right\|^2\cos^2\theta_k, k=1,2,\cdots.$$

Proof: **Let** $\bar{\alpha}_k=-\dfrac{\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}}{M\left\|d^{(k)}\right\|^2}$ **Then** $\bar{\alpha}_k>0$ **and**

$$f\left(x^{(k)}+\bar{\alpha}_k d^{(k)}\right)-f\left(x^{(k)}\right)=\bar{\alpha}_k\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}+\frac{1}{2}\bar{\alpha}_k^2 d^{(k)\mathrm{T}}\nabla^2 f\left(\xi^{(k)}\right)d^{(k)}$$

$$\leq\bar{\alpha}_k\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}+\frac{1}{2}\bar{\alpha}_k^2\left\|d^{(k)}\right\|^2 M=-\frac{\left(\nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(k)}\right)^2}{2M\left\|d^{(k)}\right\|^2}=-\frac{1}{2M}\left\|\nabla f\left(x^{(k)}\right)\right\|^2\cos^2\theta_k.$$

In particular

$$f\left(x^{(k)}\right)-f\left(x^{(k)}+\alpha_k d^{(k)}\right)\geq f\left(x^{(k)}\right)-f\left(x^{(k)}+\bar{\alpha}_k d^{(k)}\right)\geq\frac{1}{2M}\left\|\nabla f\left(x^{(k)}\right)\right\|^2\cos^2\theta_k.$$

Lemma4 Iterative sequence $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$

with $f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right).$ Let $f(x)$ be

continuously differentiable in open set $D \subset R^n$ and $\bar{x} \in D$ be

an accumulation point of $\left\{x^{(k)}\right\}$ i.e. $\lim_{k_i \to \infty} x^{(k_i)} = \bar{x}.$

If $\exists M > 0$ s.t. $\left\|d^{(k_i)}\right\| \leq M$ and $\bar{d}$ is an accumulation

point of $\left\{d^{(k_i)}\right\}.$ Then $\nabla f(\bar{x})^{\mathrm{T}} \bar{d} = 0.$

Proof: If $\lim_{k_{i_j} \to \infty} d^{(k_{i_j})} = \bar{d} = 0,$ then $\nabla f(\bar{x})^{\mathrm{T}} \bar{d} = 0.$

If $\lim_{k_{i_j} \to \infty} d^{(k_{i_j})} = \bar{d} \neq 0,$ but $\lim_{k_{i_{j_l}} \to \infty} \alpha^{(k_{i_{j_l}})} = \bar{\alpha} = 0,$

From $\lim_{k_{i_{j_l}} \to \infty} \nabla f\left(x^{(k_{i_{j_l}})} + \alpha_{k_{i_{j_l}}} d^{(k_{i_{j_l}})}\right)^{\mathrm{T}} d^{(k_{i_{j_l}})} = 0.$ yields $\nabla f(\bar{x})^{\mathrm{T}} \bar{d} = 0.$

If $\lim\sup_{k_{i_j}\to\infty} \alpha^{\left(k_{i_j}\right)} = \bar{\alpha} > 0.$ **Then there exists** $\left\{\alpha^{\left(k_{i_{j_m}}\right)}\right\}$

s.t. $\lim_{k_{i_{j_m}}\to\infty} \alpha^{\left(k_{i_{j_m}}\right)} \geq \frac{\bar{\alpha}}{2} > 0.$

If $$\lim_{k_{i_{j_m}}\to\infty} \nabla f\left(x^{\left(k_{i_{j_m}}\right)} + \alpha_{k_{i_{j_m}}} d^{\left(k_{i_{j_m}}\right)}\right)^{\mathrm{T}} d^{\left(k_{i_{j_m}}\right)} = \nabla f\left(\bar{x}\right)^{\mathrm{T}} \bar{d} \neq 0$$

Then $$\exists\delta > 0,\ \tilde{\alpha} > 0.\ \textbf{s.t.}\ \nabla f\left(x^{\left(k_{i_{j_{m_n}}}\right)} + \tilde{\alpha} d^{\left(k_{i_{j_{m_n}}}\right)}\right)^{\mathrm{T}} d^{\left(k_{i_{j_{m_n}}}\right)} < -\frac{1}{2}\delta$$

Thus $$f\left(\bar{x}\right) - f\left(x^{(0)}\right) = \sum_{k_{i_{j_{m_n}}}} \left[f\left(x^{\left(k_{i_{j_{m_n}}}\right)}\right) - f\left(x^{\left(k_{i_{j_{m_{n-1}}}}\right)}\right) \right]$$

$$\leq \sum_{k_{i_{j_{m_n}}}} \alpha_{k_{i_{j_{m_{n-1}}}}} \nabla f\left(\xi^{\left(k_{i_{j_{m_{n-1}}}}\right)}\right)^{\mathrm{T}} d^{\left(k_{i_{j_{m_{n-1}}}}\right)} \leq \sum_{k_{i_{j_{m_n}}}} \frac{\bar{\alpha}}{2}\left(-\frac{\delta}{2}\right) \to -\infty$$

This contradicts to boundedness of $f\left(\bar{x}\right) - f\left(x^{(0)}\right)$

Therefore $\nabla f\left(\bar{x}\right)^{\mathrm{T}} \bar{d} = 0.$

Th.4. Let $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right). \quad d^{(k)} = -\nabla f\left(x^{(k)}\right), k = 0,1,\cdots.$$

If $\bar{x}$ is an accumulation point of $\left\{x^{(k)}\right\}$ Then $\nabla f\left(\bar{x}\right) = 0.$

Th.5. Let $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right). \quad d^{(k)} = -\nabla f\left(x^{(k)}\right), k = 0,1,\cdots.$$

If there exists $M > 0$ s.t. $\left\|\nabla^2 f\left(x\right)\right\| \le M.$

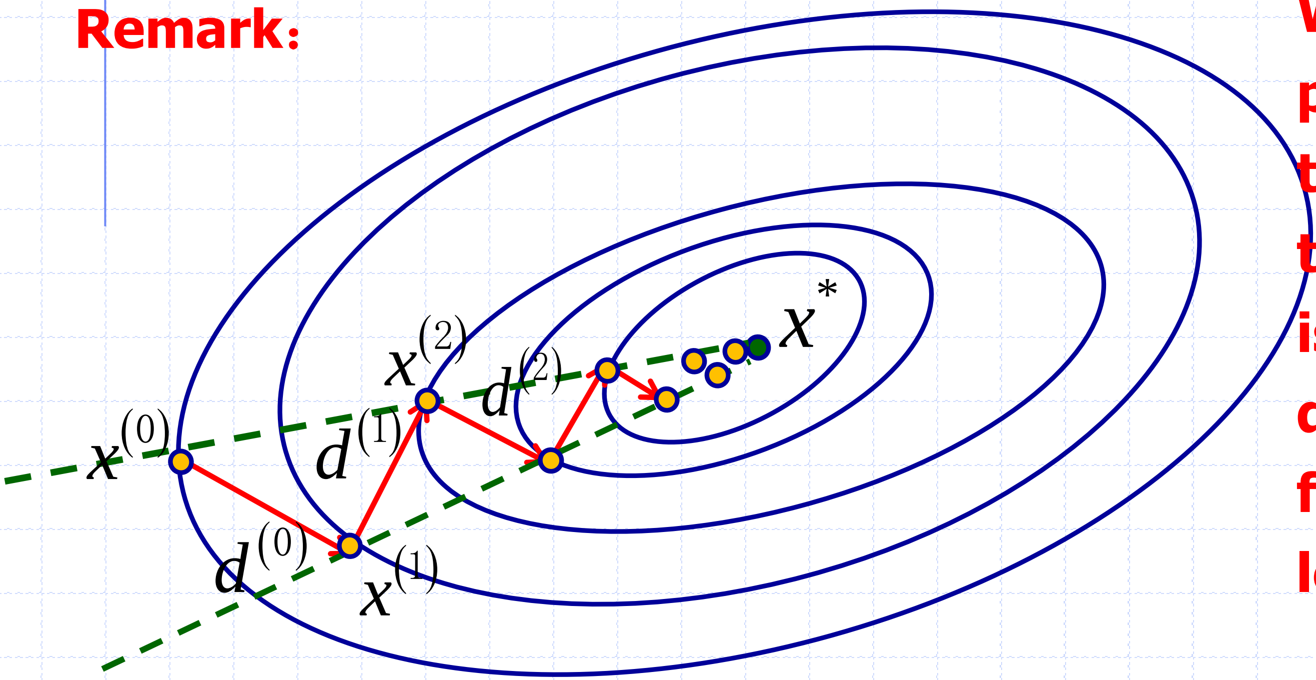
Then $\lim_{k \to \infty} \nabla f\left(x^{(k)}\right) = 0$ or $\lim_{k \to \infty} f\left(x^{(k)}\right) = -\infty$

From $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right)$ gives rise to

$$0=\left.\frac{\mathrm{d}f\left(x^{(k)}+\alpha d^{(k)}\right)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k}=\nabla f\left(x^{(k)}+\alpha_k d^{(k)}\right)^{\mathrm{T}} d^{(k)}=-d^{(k+1)\mathrm{T}}d^{(k)}.$$

This means that two adjacent directions are orthogonal.

Remark:



When the iterative point approaches to the minimizer, the convergent rate is not the fastest. It depends on the flatness of equi-level circle.

From defs. of convergence rate and condition number it conveys that more round implies faster the convergent rate.

Lemma5（Kantorovich-inequality） **Let** $A_{n\times n}$ **be RSPD and** $0<\lambda_1\le\lambda_2\le\cdots\le\lambda_n$ **be eigenvalues. Then for any** $0\ne x\in R^{(n)}$, **we have**

$$\frac{\left(x^{\mathrm{T}}x\right)^2}{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^{-1}x\right)}\ge\frac{4\lambda_1\lambda_n}{\left(\lambda_1+\lambda_n\right)^2}$$

Proof: **Let** $\Lambda=diag\left(\lambda_1,\lambda_2,\cdots,\lambda_n\right)$. **Then there exists orthogonal matrix** P, **s.t.** $A=P\Lambda P^{\mathrm{T}}$. **Let** $x=Py$. **Then**

$$\frac{\left(x^{\mathrm{T}}x\right)^2}{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^{-1}x\right)}=\frac{\left(y^{\mathrm{T}}y\right)^2}{\left(y^{\mathrm{T}}\Lambda y\right)\left(y^{\mathrm{T}}\Lambda^{-1}y\right)}=\frac{\left(\sum_{i=1}^{n}y_i^2\right)^2}{\left(\sum_{i=1}^{n}\lambda_i y_i^2\right)\left(\sum_{i=1}^{n}\frac{1}{\lambda_i}y_i^2\right)}$$

$$=\frac{1}{\left(\sum_{i=1}^{n}\lambda_i\xi_i\right)\left(\sum_{i=1}^{n}\frac{1}{\lambda_i}\xi_i\right)}$$

where $\xi_i=\dfrac{y_i^2}{\sum_{i=1}^{n}y_i^2}$

Denote $\lambda=\sum_{i=1}^{n}\lambda_i\xi_i$ **Then** $\lambda_1\le\lambda\le\lambda_n$ **As** $\lambda_1\le\lambda_2\le\cdots\le\lambda_n$

Thus $(\lambda_i-\lambda_1)(\lambda_i-\lambda_n)\le 0$ **i.e.** $\frac{1}{\lambda_i}\le\frac{\lambda_1+\lambda_n-\lambda_i}{\lambda_1\lambda_n}$

Then $$\left(\sum_{i=1}^{n}\lambda_i\xi_i\right)\left(\sum_{i=1}^{n}\frac{1}{\lambda_i}\xi_i\right)\le\lambda\left(\sum_{i=1}^{n}\frac{\lambda_1+\lambda_n-\lambda_i}{\lambda_1\lambda_n}\xi_i\right)$$

$$=\lambda\frac{\lambda_1+\lambda_n}{\lambda_1\lambda_n}\sum_{i=1}^{n}\xi_i-\frac{\lambda}{\lambda_1\lambda_n}\left(\sum_{i=1}^{n}\lambda_i\xi_i\right)=\frac{\lambda(\lambda_1+\lambda_n-\lambda)}{\lambda_1\lambda_n}$$

Therefore

$$\frac{\left(x^{\mathrm{T}}x\right)^2}{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^{-1}x\right)}=\frac{1}{\left(\sum_{i=1}^{n}\lambda_i\xi_i\right)\left(\sum_{i=1}^{n}\frac{1}{\lambda_i}\xi_i\right)}\ge\frac{\lambda_1\lambda_n}{\lambda(\lambda_1+\lambda_n-\lambda)}\ge\frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}$$

Th.6. **Unconstrained Optimization** $\min f(x)=\frac{1}{2}x^{\mathrm{T}}Ax$

where $A_{n\times n}$ **is RSPD.** **Let** $0<\lambda_1\le\lambda_2\le\cdots\le\lambda_n$ **be eigenvalues.**

Iterative sequence $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}, k=0,1,\cdots.$

with $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min\limits_{\alpha\ge 0} f\left(x^{(k)}+\alpha d^{(k)}\right),\quad d^{(k)}=-\nabla f\left(x^{(k)}\right).$

Then

$$f\left(x^{(k+1)}\right)\le\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 f\left(x^{(k)}\right),$$

$$\left\|x^{(k+1)}\right\|_2\le\sqrt{\frac{\lambda_n}{\lambda_1}}\left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)\left\|x^{(k)}\right\|_2.$$

That is, $\left\{f\left(x^{(k)}\right)\right\}$ **is Q-superlinearly/linearly convergent.**

Further $\left\{x^{(k)}\right\}$ **is R-superlinearly/linearly convergent.**

Proof: $d^{(k)} = -\nabla f\left(x^{(k)}\right) = -Ax^{(k)}$. **From** $f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min\limits_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right)$ **yields**

$$\left.\frac{\mathrm{d}f\left(x^{(k)} + \alpha d^{(k)}\right)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k} = \nabla f\left(x^{(k)} + \alpha_k d^{(k)}\right)^{\mathrm{T}} d^{(k)} = \left(Ax^{(k)} + \alpha_k Ad^{(k)}\right)^{\mathrm{T}} d^{(k)} = 0$$

Then $\alpha_k = \dfrac{d^{(k)\mathrm{T}} d^{(k)}}{d^{(k)\mathrm{T}} A d^{(k)}}$ **Due to** $x^{(k)\mathrm{T}} A x^{(k)} = d^{(k)\mathrm{T}} A^{-1} d^{(k)}$ **Thus**

$-d^{(k)}$

$$\begin{aligned}
f\left(x^{(k+1)}\right) &= \frac{1}{2}\left(x^{(k)} + \alpha_k d^{(k)}\right)^{\mathrm{T}} A\left(x^{(k)} + \alpha_k d^{(k)}\right) \\
&= \frac{1}{2}\left[x^{(k)\mathrm{T}} A x^{(k)} + 2\alpha_k d^{(k)\mathrm{T}} \boxed{Ax^{(k)}} + \alpha_k^2 d^{(k)\mathrm{T}} A d^{(k)}\right] \\
&= \frac{1}{2} x^{(k)\mathrm{T}} A x^{(k)} \left[1 - \frac{\left(d^{(k)\mathrm{T}} d^{(k)}\right)^2}{\left(d^{(k)\mathrm{T}} A d^{(k)}\right)\left(d^{(k)\mathrm{T}} A^{-1} d^{(k)}\right)}\right] \\
&\leq f\left(x^{(k)}\right)\left[1 - \frac{4\lambda_1\lambda_n}{\left(\lambda_1 + \lambda_n\right)^2}\right] = f\left(x^{(k)}\right)\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2
\end{aligned}$$

In addition $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ **Then for any** $x \in R^n$,

we have $\lambda_1 \|x\|_2^2 \le 2f(x) = x^{\mathrm{T}} A x \le \lambda_n \|x\|_2^2$ **Therefore**

$$\lambda_1 \left\| x^{(k+1)} \right\|_2^2 \le 2f\left(x^{(k+1)}\right) \le 2f\left(x^{(k)}\right)\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \le \lambda_n \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 \left\| x^{(k)} \right\|_2^2$$

i.e.

$$\left\| x^{(k+1)} \right\|_2 \le \sqrt{\frac{\lambda_n}{\lambda_1}} \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \left\| x^{(k)} \right\|_2$$

and

$$\left\| x^{(k+1)} \right\|_2^2 \le \frac{1}{\lambda_1}\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 2f\left(x^{(k)}\right) \le \cdots \le \frac{2f\left(x^{(0)}\right)}{\lambda_1}\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{2(k+1)}$$

This indicates $\left\{x^{(k)}\right\}$ **is R-superlinearly/linearly convergent.**

Define $cond(A) = \dfrac{\lambda_n}{\lambda_1}$ **as condition number or matrix** A

Then, the larger of $cond(A)$ **the slower of the convergent rate.**

On the contrary,

The smaller of $cond(A)$ **implies faster convergent rate.**

Th.7. Suppose $f(x)$ is 2nd-order continuously differentiable and

$$\nabla^2 f(x) = H(x) = \left[H_{ij}(x)\right]_{n\times n}$$ satisfies Lipschitz condition.

That is $\exists \beta > 0,$ such that

$$\left|H_{ij}(x) - H_{ij}(y)\right| \le \beta \|x-y\|, \quad \forall x, y \in R^n, i, j = 1,\cdots,n.$$

Let $$d^{(k)} = -\nabla f\left(x^{(k)}\right), k = 0,1,\cdots.$$

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha\ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right).$$

Iterative sequence $$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, k = 0,1,\cdots.$$

Then $\left\{ f\left(x^{(k)}\right)\right\}$ is Q-linearly convergent.

(2) Newton's Method

Quadratic approximation

$$f(x)=\boxed{f\left(x^{(k)}\right)+\nabla f\left(x^{(k)}\right)^{\mathrm{T}}\left(x-x^{(k)}\right)+\frac{1}{2}\left(x-x^{(k)}\right)^{\mathrm{T}}\nabla^2 f\left(x^{(k)}\right)\left(x-x^{(k)}\right)}$$

$$+o\left(\left\|x-x^{(k)}\right\|^2\right), \qquad x\in N_1\left(x^{(k)}\right), x^{(k)}\in N_2\left(x^*\right)$$

Then $f(x)\approx q_k(x)$ **If** $\nabla^2 f\left(x^{(k)}\right)$ **is positive definite,** $q_k(x)$

From $\nabla q_k(x)=\nabla f\left(x^{(k)}\right)+\nabla^2 f\left(x^{(k)}\right)\left(x-x^{(k)}\right)=0$ **makes**

$$x-x^{(k)}=-\left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1}\nabla f\left(x^{(k)}\right)$$

Thus, construct iterative sequence

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

$d^{(k)}$ **satisfying** $\nabla^2 f\left(x^{(k)}\right)d^{(k)}=-\nabla f\left(x^{(k)}\right)$

named as Newton's Formula.

Th.8. Suppose $f\left(x\right)$ is 2nd-order continuously differentiable and

$$f\left(x^*\right)=min\ f\left(x\right), x\in N\left(x^*\right),\quad \nabla f\left(x^*\right)=0,$$

$$\nabla^2 f\left(x^*\right)=H\left(x^*\right)=\left[H_{ij}\left(x^*\right)\right]_{n\times n}$$ is positive definite.

$\nabla^2 f\left(x\right)=H\left(x\right)$ satisfies Lipschitz condition, i.e. $\exists\beta>0,$

s.t. $\left|H_{ij}\left(x\right)-H_{ij}\left(y\right)\right|\le\beta\left\|x-y\right\|,\quad \forall x,y\in R^n, i,j=1,\cdots,n.$

Then, for the case when $\left\|x^{(0)}-x^*\right\|<<1,$

Sequence $x^{(k+1)}=x^{(k)}-\left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1}\nabla f\left(x^{(k)}\right)$ is well-defined

and $\lim\limits_{k\to\infty} x^{(k)}=x^*,\ \lim\limits_{k\to\infty}\dfrac{\left\|x^{(k+1)}-x^*\right\|}{\left\|x^{(k)}-x^*\right\|^2}=\eta,\ \lim\limits_{k\to\infty}\sqrt[k+1]{\left\|x^{(k+1)}-x^*\right\|}<1.$

That is $\left\{x^{(k)}\right\}$ is R-linearly convergent.

(3) Damped Newton's Method

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right) \qquad k = 0,1,\cdots$$

Th.9 Suppose that $f(x) \in C^2$ and for any $x^{(0)} \in R^n$,

there exists $m > 0,$ s.t. $m\|u\|^2 \le u^{\mathrm{T}} H(x) u, \ \forall x \in L\left(x^{(0)}\right).$

that is, $f(x)$ is consistently convex.

Then

(1) For the case when $\left\{x^{(k)}\right\} = \left\{x^{(0)}, x^{(1)}, \cdots, x^{(k_0)}\right\}$

$$f\left(x^{(k_0)}\right) = min\ f(x), x \in R^n$$

(2) For the case when $\lim\limits_{k \to \infty} x^{(k)} = x^*,$

$$f\left(x^*\right) = min\ f(x), x \in R^n$$

(4) Damped Newton's with protection

$$x^{(k+1)} = x^{(k)} + d^{(k)},$$ **where** $d^{(k)}$ **is chosen as:**

(1) $\nabla f\left(x^{(k)}\right) \neq 0$ **and** $\nabla^2 f\left(x^{(k)}\right)$ **is positive definite.**

Then $d^{(k)} = -\left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right)$

(2) $\nabla^2 f\left(x^{(k)}\right)$ **not positive definite but** $\nabla f\left(x^{(k)}\right)^{\mathrm{T}} \nabla^2 f\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right) < 0.$

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

(3) $\nabla^2 f\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right) = 0$ **or** $\nabla^2 f\left(x^{(k)}\right)$ **singular.**

Then $d^{(k)} = -\nabla f\left(x^{(k)}\right)$

Remark:

At the beginning iterations we adopt Steepest Descent Method and nearing the minimizer we adopt Newton's Method.

THANK YOU FOR ATTENDING