上节课两道题.

① $\max\ -b^Tv-\lambda_1^Tu+\lambda_2^T l.$ $\qquad\qquad \max\ -v^Tb+l^T\underbrace{(A^Tv+c)^+}_{\lambda_2}-u^T\underbrace{(A^Tv+c)^-}_{\lambda_1}$

s.t. $A^Tv+\lambda_1-\lambda_2+c=0.$

$\lambda_1\ge 0,\ \lambda_2\ge 0.$

$\Leftrightarrow.$ $\max\ -v^Tb+l^T\lambda_2-u^T\lambda_1.$

s.t. $\left.\begin{array}{l}\lambda_2=(A^Tv+c)^+\\ \lambda_1=(A^Tv+c)^-\end{array}\right\}\longrightarrow\left|\begin{array}{l}\lambda_1\ge 0,\ \lambda_2\ge 0.\\ \lambda_2-\lambda_1=A^Tv+c\end{array}\right.$

②. 若$g(x,u,w)$为凸, 则$p(u,w)=\inf\limits_x g(x,u,w)$为凸.

$\forall\theta\in[0,1],\ \inf\limits_{x_1,x_2}\left(\theta g(x_1,u_1,w_1)+(1-\theta)g(x_2,u_2,w_2)\right)$

$\ge\inf\limits_{x_1,x_2}\ g(\theta x_1+(1-\theta)x_2,\ \theta u_1+(1-\theta)u_2,\ \theta w_1+(1-\theta)w_2).$

$\theta p(u_1,w_1)+(1-\theta)p(u_2,w_2)\ge p(\theta u_1+(1-\theta)u_2,\ \theta w_1+(1-\theta)w_2).$ 凸函数.

函数性质的分析 $f(x)$.

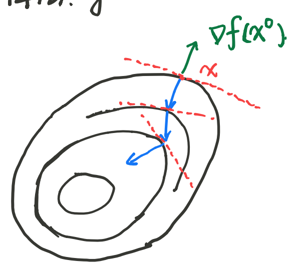
强凸性: $\exists m,\ \forall x\in\mathrm{dom}f,\ \nabla^2 f(x)\succeq mI.$

$\Leftrightarrow\ \forall x,y\in\mathrm{dom}f,\ \ f(y)\ge f(x)+\nabla^2 f(x)(y-x)+\frac{1}{2}m\|y-x\|_2^2.$

相应的, $\exists M,\ \forall x\in\mathrm{dom}f,\ \nabla^2 f(x)\preceq mI.$

$\Leftrightarrow\ \forall x,y\in\mathrm{dom}f,\ \ f(y)\le f(x)+\nabla^2 f(x)(y-x)+\frac{1}{2}m\|y-x\|_2^2.$

梯度下降法. gradient decent.



$d^k=-\nabla f(x^k)$

分析算法的收敛性. $\forall x\in\mathrm{dom}f,\ MI\succeq\nabla^2 f(x)\succeq mI.$

1) exact line search.

$$\tilde f(\alpha)=f(x^k+\alpha d^k)=f(\underbrace{x^k-\alpha\nabla f(x^k)}_{\Downarrow\, x^{k+1}}).$$

$$f(\underbrace{x^{k+1}}_{y})\le f(\underbrace{x^k}_{x})+\nabla f^T(x^k)(-\alpha\nabla f(x^k))+\frac{M}{2}\|-\alpha\nabla f(x^k)\|_2^2$$

$$\Leftrightarrow \quad \tilde{f}(\alpha) \le f(x^k) - \alpha\|\nabla f(x^k)\|_2^2 + \frac{M}{2}\alpha^2\|\nabla f(x^k)\|_2^2.$$

$$-\|\nabla f(x^k)\|_2^2 + M\alpha\|\nabla f(x^k)\|_2^2 = 0. \quad \Rightarrow \alpha = \frac{1}{M}$$

$$\min_{\alpha} \tilde{f}(\alpha) \le f(x^k) - \frac{1}{M}\|\nabla f(x^k)\|_2^2 + \frac{1}{2M}\|\nabla f(x^k)\|_2^2$$

$$\begin{cases} f(x^{k+1}) \le f(x^k) - \frac{1}{2M}\|\nabla f(x^k)\|_2^2. \\ p^* \ge f(x^k) - \frac{1}{2m}\|\nabla f(x^k)\|_2^2 \end{cases}$$

$$\frac{1}{2M}\|\nabla f(x^k)\|_2^2 + \underline{f(x^{k+1}) - p^*} \le \underline{f(x^k) - p^*} \qquad \times M$$

$$-\frac{1}{2m}\|\nabla f(x^k)\|_2^2 + \underline{f(x^k) - p^*} \le 0. \qquad \times m.$$

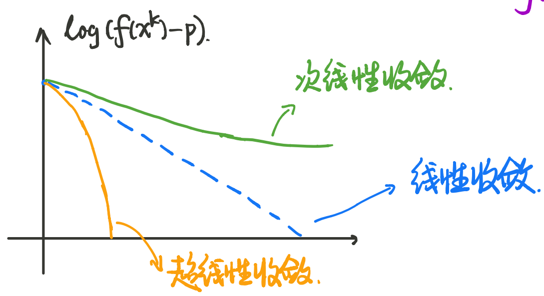
$$M(f(x^{k+1}) - p^*) + m(f(x^k) - p^*) \le M(f(x^k) - p^*).$$

$$\Leftrightarrow \underbrace{(f(x^{k+1}) - p^*)}_{\ge 0.} \le (1 - \frac{m}{M})\underbrace{(f(x^k) - p^*)}_{\ge 0.} \quad \text{越来越接近最优值.}$$

$$\left\| \frac{f(x^{k+1}) - p^*}{f(x^k) - p^*} \right\| \le \left\| 1 - \frac{m}{M} \right\|.$$

given ε, 何时 $\left\| \frac{f(x^{k+\tau}) - p^*}{f(x^k) - p^*} \right\| \le \varepsilon$ (经过τ步,十分接近最优解).

$$(1 - \frac{m}{M})^{\tau} = \varepsilon. \quad \Rightarrow \tau = \frac{\log \varepsilon}{\log(1 - \frac{m}{M})} \approx \frac{m}{M}.$$



$6 \to 5 \to 4 \to 3 \to \cdots$ 不是线性收敛.

$6 \to 6\times 0.9 \to 6\times 0.9^2 \to \cdots$ 线性收敛

2). Inexact line search (Armijo Rule).

当$0 \le \alpha \le \frac{1}{m}$时, 迭代必能停止.

$\tilde{f}(\alpha) = f(x^{k+1}) \le f(x^k) + \gamma\alpha \cdot \nabla f^T(x^k) d^k$ 时接受α.

当 $0\le\alpha\le\frac{1}{M}$ 时，必有 $(-\alpha+\frac{M\alpha^2}{2})\le-\frac{\alpha}{2}$

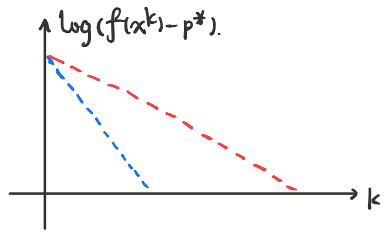
$$\tilde{f}(\alpha)=f(x^{k+1})\le f(x^k)-\alpha\|\nabla f(x^k)\|_2^2+\frac{M\alpha^2}{2}\|\nabla f(x^k)\|_2^2$$
$$\le f(x^k)-\frac{\alpha}{2}\|\nabla f(x^k)\|_2^2$$
$$\le f(x^k)-\gamma\alpha\|\nabla f(x^k)\|_2^2$$

$\alpha_{inexact}=\alpha_{max}$ 或 $\geqslant\frac{\beta}{M}$.

$$f(x^{k+1})=\tilde{f}(\alpha_{exact})\le f(x^k)-\frac{1}{2M}\|\nabla f(x^k)\|_2^2$$

$$f(x^{k+1})\le f(x)-\min\left\{\gamma\alpha_{max},\frac{\gamma\beta}{M}\right\}\|\nabla f(x)\|_2^2$$

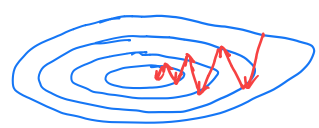
$$\frac{f(x^{k+1})-p^*}{f(x^k)-p^*}\le 1-\min\left\{2m\gamma\alpha_{max},\frac{2m\gamma\beta}{M}\right\}\to 1$$



例: $f(x)=\frac{1}{2}x^TPx$, $P\in S_+^n$

$\nabla^2f(x)=P$. $\underset{\lambda_{max}}{MI}\succeq P\succeq\underset{\lambda_{min}}{mI}$. $\frac{M}{m}$ 小，收敛快.

$P=\begin{pmatrix}100 & 0\\ 0 & 1\end{pmatrix}$



收敛慢. zig-zag.