lec47.

梯度下降法. $d^{k+1}=-\nabla f(x^k)$.

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

$k \sim \log(f(x^k)-p^*)$ 线性收敛.

最速下降法. steepest descent.

陡.

$x^k \quad \min_{x} f(x) \Rightarrow \min_{v} f(x^k+v).$

$$\min_v \{f(x^k)+\nabla f^T(x^k)v\}$$

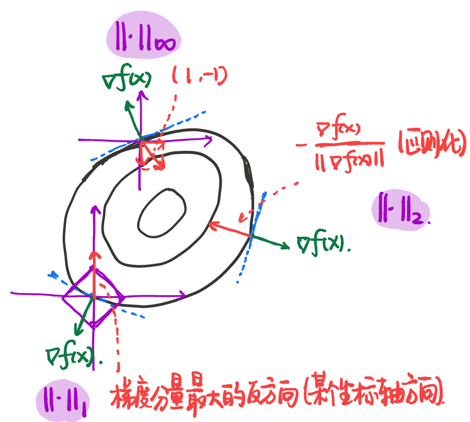
$$d^k=\arg\min_v \{f(x^k)+\nabla f^T(x^k)v \mid \|v\|=1\} \to \alpha^k$$

$$\begin{cases}\text{exact}\\ \text{Amijo}\end{cases}$$

$$d^k=\arg\min_v \{\nabla f^T(x^k)v \mid \|v\|_2=1\}.$$

$$d^k=\arg\min_v \{\nabla f^T(x^k)v \mid \|v\|=1\}.$$

$$\nabla f(x)=\begin{pmatrix}(\nabla f(x))_1\\ \vdots\\ o\\ \vdots\\ (\nabla f(x))_n\end{pmatrix} \quad \begin{pmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{pmatrix} \begin{pmatrix}0\\ \vdots\\ -1\\ \vdots\\ 0\end{pmatrix}$$



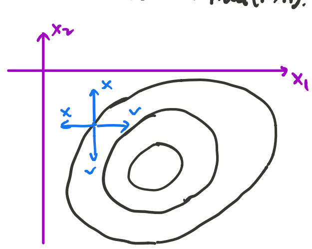
$$d^k=\arg\min_v \{\nabla f^T(x^k)v \mid \|v\|_\infty=1\}$$

$$\begin{pmatrix}+1\\ -1\\ \vdots\\ -1\\ +1\end{pmatrix} \quad \begin{matrix}(\nabla f^T(x^k))_i>0, (d^k)_i=-1\\ (\nabla f^T(x^k))_i<0, (d^k)_i=+1\end{matrix}$$

Gradient 与 Steepest Descent 的变种.

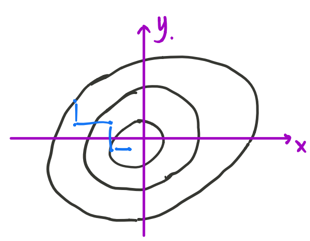
1). 坐标轮换法. Coordinate Descent

设 $x\in\mathbb{R}^n \quad d^k=e_{mod(k,n)}$.



Block Coordinate Descent.

$\min f(x,y)$.



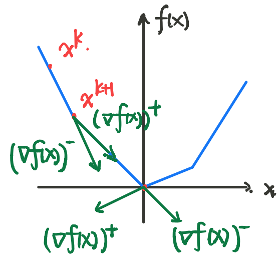
$$x^{k+1}=\arg\min_x f(x,y^k).$$

$$y^{k+1}=\arg\min_y f(x^{k+1},y).$$

$\vdots$

lec48.

2)若 $f(x)$ 在某些点不可微.



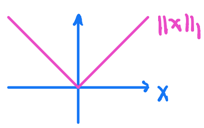
次梯度 subgradient.

$$\frac{\partial f(x)}{\partial x} = \begin{cases} \nabla f(x). \\ (\nabla f(x))^{+} \sim (\nabla f(x))^{-} \end{cases}$$

当 $(\nabla f(x))^{-} \sim (\nabla f(x))^{+}$ 包含0时, 取得最小值.

例: $\min f(x) = \frac{1}{2}\|Ax-b\|_2^2 + \lambda \|x\|_1$.

$$\frac{\partial f(x)}{\partial x} = A^T(Ax-b) + \lambda \frac{\partial \|x\|_1}{\partial x}$$



$$\begin{cases} \left(\frac{\partial \|x\|_1}{\partial x}\right)_i = 1 & x_i > 0. \\ \left(\frac{\partial \|x\|_1}{\partial x}\right)_i = -1 & x_i < 0 \\ -1 \le \left(\frac{\partial \|x\|_1}{\partial x}\right) \le 1 & x_i = 0 \end{cases}$$

牛顿法. Newton's method

x^k $\quad d^k = \arg\min_{v} \{f(x^k+v) \mid \|v\| = 1\}$. $\qquad \nabla^2 f(x) \succeq mI$.

$$= \arg\min_{v} \{f(x^k) + \underbrace{\nabla f^T(x^k) v + \tfrac{1}{2} v^T \nabla^2 f^T(x^k) v}\}$$ 关于v的二次函数, 有最小值.

$$\nabla f(x^k) + \nabla^2 f(x^k) v = 0.$$

$$v = -(\nabla^2 f(x^k))^{-1} f(x^k).$$ 牛顿方向.

当 $\nabla f^T(x^k) d^k = -\nabla f^T(x^k)(\nabla^2 f(x^k)) \nabla f(x^k)$.

牛顿法算法

Repeat

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

$\alpha^k = \arg\min_{0 \le \alpha \le \alpha_{max}} f(x^k + \alpha d^k)$.

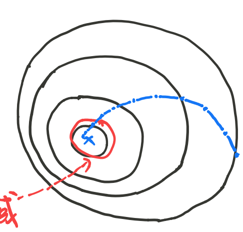
$x^{k+1} = x^k + \alpha^k d^k$.

Until convergence or $|\nabla f^T(x^k)(\nabla^2 f(x^k))^{-1} \nabla f(x^k)| \le \varepsilon$.

收敛性分析 $\exists \eta > 0$

①若 $\|\nabla f(x)\|_2 > \eta$ damped Newton phase.

②若 $\|\nabla f(x)\|_2 < \eta$ quadratically convergent phase.



在某邻域

收敛速度加快.

$$\frac{f(x^{k+1})-p^*}{f(x^k)-p^*} \sim u\,(<1) \text{ 线性.} \quad \frac{f(x^{k+1})-p^*}{f(x^k)-p^*} \sim u^2\,(<1) \text{ 二次.}$$

拟牛顿法. Quasi-Newton Method.

$$\nabla^2 f(x^k)\, d^k = -\nabla f(x^k).$$

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

(BFGS法)