Tutorial 6

19 February 2022 11:04

Gradient Descent methods

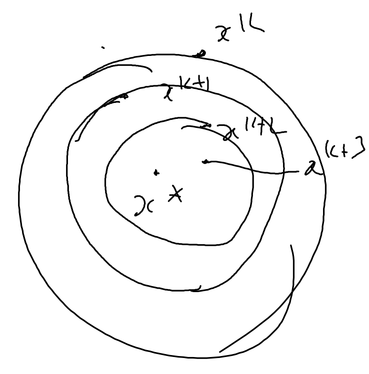
Steepest Gradient method

- fixed step size
- Exact line search
- Armijo's rule

- Newton's method
- Conjugate direction methods
  - Conjugate gradient method for quadratic problems
  - " " " non quadratic problems.

$$\|x_{k+1} - x^*\| \le |\bar{\lambda}(\nabla^2 f(x_k))| \, L \, \|x_k - x^*\|^2$$

$$\|x_{k+1} - x^*\| \le \left(\underbrace{|\bar{\lambda}(\nabla^2 f(x_k))| \, L \, \|x_k - x^*\|}_{<1}\right) \left(\|x^k - x^*\|\right)$$



suppose $|\bar{\lambda}(\nabla^2 f(x_k))| \, L < 1$

and $\|x_k - x^*\| < 1$

$$\Rightarrow |\bar{\lambda}(\nabla^2 f(x_k))| \, \|x_k - x^*\| < 1$$

(Actually this is enough)

$$\Rightarrow \|x_{k+1} - x^*\| < \|x_k - x^*\| < 1$$

$$\underbrace{|\bar{\lambda}(\nabla^2 f(x_k))|}_{\Lambda} \, \|x_k - x^*\| < 1$$

$$\Lambda \|x_k - x^*\| < 1$$

$$\|x_k - x^*\| < \frac{1}{\Lambda}$$

$$\|x_k - x^*\| \, |\bar{\lambda}(\nabla^2 f(x_k))| < 1$$

$\|x_k - x^*\| < 1/n$

To begin with $\|x_k - x^*\| \, |I(\nabla^2 f(x_k))| < 1$ may not be satisfied

- we modify Newton's method to address this:

$\boxed{0.000000\,00\,00783} \simeq 0 \qquad \frac{1}{0}$ is undefined

10

$$x_{k+1} = x_k - \alpha_k \underbrace{(\nabla^2 f(x_k) + \delta_k I)^{-1} \nabla f(x_k)}$$

x_{k+1}

$d_k = -(\nabla^2 f(x_k) + \delta_k I)^{-1} \nabla f(x_k) = d_k$

x_k

$\alpha_k \|(\nabla^2 f(x_k) + \delta_k)^{-1} \nabla f(x_k)\|$

$$x_{k+1} = x_k + \alpha_k d_k$$

$$x_{k+1} = x_k - (\nabla^2 f(x_k)^{-1}) \nabla f(x_k)$$

- two ways to see it

(a) $d_k = -\nabla f(x_k)$ (as in steepest descent)

$\alpha_k = \nabla^2 f(x_k)^{-1}$ (it is a matrix unlike gradient descent methods)

$d_k \|d_k\|$

x_k d_k

(b) $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

$\alpha_k = 1$

x_k $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

$\|\nabla^2 f(x_k)^{-1} \nabla f(x_k)\|$

Modified Newton's method

$f(x_{k+1}) \le f(x_k)$

Modified Newton's Method

If $\boxed{f(x_{k+1}) \le f(x_k)}$

$$d_k = \underbrace{-(\nabla^2 f(x_k))^{-1} \nabla f(x_k)}$$

$$\alpha_k = 1$$

Else

$$d_k = -(\nabla^2 f(x_k) + \delta_k I)^{-1} \nabla f(x_k)$$

α_k s.t. $f(x_{k+1}) \le f(x_k)$

Armijo's rule

choose α_k s.t.

$$f(x^k + \alpha_k d_k) \le f(x_k) - \sigma \alpha_k \|d_k\|^2$$

($d_k = -\nabla f(x_k)$ in gradient descent)

Exact line search

$$\alpha_k = \arg\min_{\alpha} f(x^k - \alpha (\nabla^2 f(x_k) + \delta_k I)^{-1} \nabla f(x_k))$$

Q.2 $f : \mathbb{R}^n \to \mathbb{R}$

$$f(x) = \frac{x^T A x}{2} - b^T x + c \qquad (A \text{ is symmetric positive definite})$$

$$\nabla f(x_k) = A x - b$$

Steepest descent method

$$x_{k+1} = x_k - \alpha (A x_k - b)$$

Exact line search

$$\alpha_k = \arg\min_{\alpha} f(x_k - \alpha A x_k + \alpha b)$$

$$= \arg\min_{\alpha} f(\underbrace{(I - \alpha A) x_k + \alpha b})$$

Define $g(\alpha) = f((I - \alpha A) x_k + \alpha b)$ $\qquad g : \mathbb{R} \to \mathbb{R}$

Is g convex? (Exercise)

Is g convex? (Exercise)

$g'(\alpha) = \nabla f\left(\left((I-\alpha A)x_k + \alpha b\right)^T (b - Ax_k)\right)$

$= \left(A(I-\alpha A)x_k + \alpha A b - b\right)^T (b - Ax_k)$

$= \left(A(I-\alpha A)x_k - (I-\alpha A)b\right)^T (b - Ax_k)$

$A(I-\alpha A) = (I-\alpha A)A$

$= \left((I-\alpha A)(Ax_k - b)\right)^T (b - Ax_k)$

$= -(Ax_k - b)^T (I-\alpha A)(Ax_k - b)$

$g''(\alpha)$

α_k is obtained via setting $g'(\alpha) = 0$

(b) $b - Ax_0 = \begin{bmatrix} 10 \\ 5 \end{bmatrix} - \underbrace{\begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}}_{=A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$

Exact line search

$\alpha_0 = \dfrac{(8 \;\; 1)\begin{pmatrix} 8 \\ 1 \end{pmatrix}}{[8 \;\; 1]\begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}\begin{bmatrix} 8 \\ 1 \end{bmatrix}}$

$\boxed{x_1 = x_0 + \alpha_0 d_0}$ $\quad (d_0 = b_0 - Ax_0)$

* $(Ax_k - b)^T (I - \alpha A)(Ax_k - b) = 0$

$\Rightarrow (Ax_k - b)^T \left((Ax_k - b) - \alpha A(Ax_k - b)\right) = 0$

$\Rightarrow (Ax_k - b)^T (Ax_k - b) = \alpha (Ax_k - b)^T A (Ax_k - b)$

$\Rightarrow \alpha_k = \dfrac{\|Ax_k - b\|^2}{(Ax_k - b)^T A (Ax_k - b)}$

$n=1$

$$f(x) = \frac{1}{2}ax^2 - bx + c$$

$$\alpha_k = \frac{(ax_k - b)^2}{(ax_k - b)\, a(ax_k - b)} = \frac{1}{a}$$

$$f'(x) = ax - b$$

$$x_{k+1} = x_k - \alpha(ax_k - b)$$

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}}\ \frac{1}{2}a\left(x_k - \alpha(ax_k - b)\right)^2 - b\left(x_k - \alpha(ax_k - b)\right) + c$$

define $g(\alpha) = f\left(x_k - \alpha(ax_k - b)\right)$

$$g'(\alpha) = 0$$

$$-f'\left(x_k - \alpha(ax_k - b)\right) \cdot (ax_k - b) = 0$$

$$-\left(a\left(x_k - \alpha(ax_k - b)\right) - b\right)\underbrace{(ax_k - b)}_{\neq 0} = 0$$

$$ax_k - b = \alpha a(ax_k - b)$$

$$\Rightarrow \alpha = 1/a$$