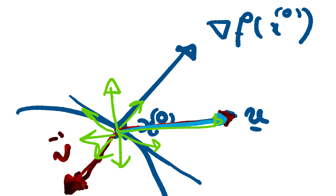
Lecture 11

Monday, 7 February 2022 2:00 PM

[Recall: $x^{(k+1)} = x^{(k)} - \boxed{F(x^{(k)})^{-1}} \nabla f(x^{(k)})$

Recall:

Directional derivative of f in the direction of $\underline{u}^{(\|\underline{u}\|=1)}$ at $\underline{x}^{(0)}$ is



$$\nabla_u f(\underline{x}^{(0)}) = Df(\underline{x}^{(0)})\,\underline{u}$$
$$= \nabla f(\underline{x}^{(0)})^T \underline{u} = \langle \nabla f(\underline{x}^{(0)}), \underline{u} \rangle .$$

Result: At a given point $\underline{x}^{(0)}$, $\underline{v} = -\nabla f(\underline{x}^{(0)})$ points in the direction of most rapid decrease for $f(\underline{x})$ and the rate of decrease of $f(\underline{x})$ at $\underline{x}^{(0)}$ in this direction is $-\|\nabla f(\underline{x}^{(0)})\|$.

[Direction of steepest descent]

Compute the direction of steepest descent

$$-\|\nabla f(\underline{x}^{(0)})\| = -\|\nabla f(\underline{x}^{(0)})\|\,\|\underline{u}\| \qquad (\because \|\underline{u}\| = 1)$$
$$\leq \langle \nabla f(\underline{x}^{(0)}), \underline{u} \rangle$$
$$= \nabla_u f(\underline{x}^{(0)})$$

[Cauchy-Schwarz
$\underbrace{|\langle a, b\rangle| \leq \|a\|\,\|b\|}$
$-\|a\|\,\|b\| \leq \langle a, b\rangle \leq \|a\|\,\|b\|$]

$$\nabla_u f(\underline{x}^{(0)}) \geq -\|\nabla f(\underline{x}^{(0)})\|$$

Minimum value is when we have an equality; that is when $\underline{u} = \dfrac{-\nabla f(\underline{x}^{(0)})}{\|\nabla f(\underline{x}^{(0)})\|}$.

[What we were looking for? Which direction $\underline{u}$ makes $\nabla_u f(x^{(0)})$ as small as possible? ↳ steepest descent direction]

Method of Steepest descent

Let $f(\underline{x})$ be a function with continuous partial derivatives on $\mathbb{R}^n$ and let $\underline{x}^{(0)} \in \mathbb{R}^n$. Steepest descent $\{\underline{x}^{(k)}\}$ with initial point $\underline{x}^{(0)}$ for minimizing $f(\underline{x})$ is given by

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k \nabla f(\underline{x}^{(k)}), \quad \alpha_k \geq 0$$

with α_k is the value of $\alpha \geq 0$ that minimizes the function

$$\phi_k(\alpha) = f(\underline{x}^{(k)} - \alpha \nabla f(\underline{x}^{(k)})), \quad \alpha \geq 0$$

Example: Find the first 3 terms of steepest descent of $f(x,y) = 4x^2 - 4xy + 2y^2$; $\underline{x}^{(0)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Soln

$$\nabla f(x,y) = \begin{pmatrix} 8x - 4y \\ -4x + 4y \end{pmatrix} \qquad (\nabla f)\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 8\times 2 - 4\times 3 \\ -4\times 2 + 4\times 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

$$x^{(1)} = x^{(0)} - \alpha \nabla f(\underline{x}^{(0)})$$

$$= \begin{pmatrix} 2 \\ 3 \end{pmatrix} - \alpha \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 - 4\alpha \\ 3 - 4\alpha \end{pmatrix}.$$

$$\phi_0(\alpha) = f(2-4\alpha, 3-4\alpha)$$

$$0 = \phi_0'(\alpha) = -\nabla f(2-4\alpha, 3-4\alpha) \cdot \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

$$0 = -4 \begin{pmatrix} 8(2-4\alpha) - 4(3-4\alpha) \\ -4(2-4\alpha) + 4(3-4\alpha) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 2(2-4\alpha) - (3-4\alpha) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= -16 \begin{pmatrix} 2(2-4\alpha) - (3-4\alpha) \\ -2+4\alpha + 3-4\alpha \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$0 = -16 \left[4 - 8\alpha - 3 + 4\alpha - 2 + \cancel{4\alpha} + 3 - \cancel{4\alpha} \right]$$

$$0 = -16 \left[2 - 4\alpha \right] \Rightarrow \boxed{\alpha_0 = \tfrac{1}{2}} \qquad \phi''(\tfrac{1}{2}) \to$$

$$x^{(1)} = x^{(0)} - \frac{1}{2} \nabla f(x^{(0)})$$

$$= \begin{pmatrix} 2 - 4 \times 1/2 \\ 3 - 4 \times 1/2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

$$x^{(2)} = x^{(1)} - \alpha \nabla f(x^{(1)})$$

Check:

$$\varphi_1(\alpha) = f \begin{pmatrix} 4\alpha \\ 1-4\alpha \end{pmatrix}$$

$$\varphi_1'(\alpha) = 0 \Rightarrow \boxed{\alpha_1 = \frac{1}{10}}$$ Check HW. $x^{(2)} = \begin{pmatrix} 2/5 \\ 3/5 \end{pmatrix}$

$$x^{(3)} = \begin{pmatrix} 0 \\ 1/5 \end{pmatrix}.$$

$$f(x,y) = 4x^2 - 4xy + 2y^2$$

$$= 2\left[x^2 + x^2 - 2xy + y^2 \right]$$

$$= 2\left[x^2 + \underbrace{(x-y)^2} \right] \geq 0$$

$\boxed{\text{min is } 0}$ at $(0,0)$.

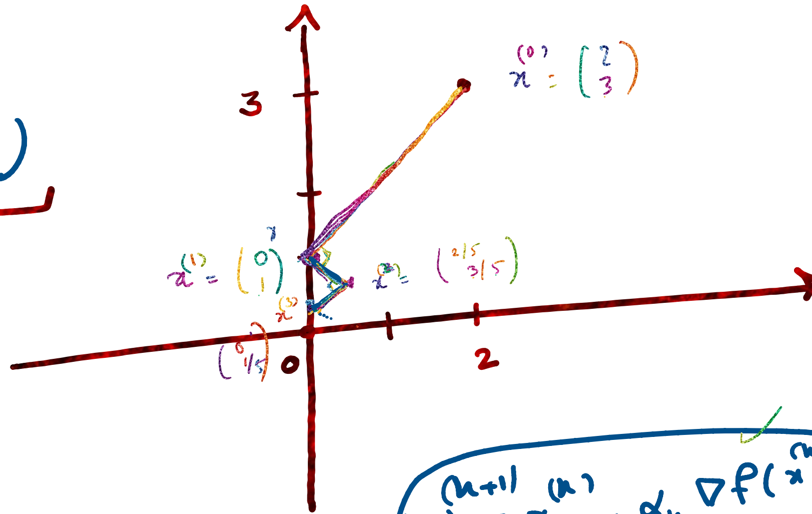
$x^{(0)}$ $x^{(1)}$ $x^{(2)}$ $x^{(3)}$ $x^{(0)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

Plot $x^{(0)}, x^{(1)}, x^{(2)}, x^{(3)}$

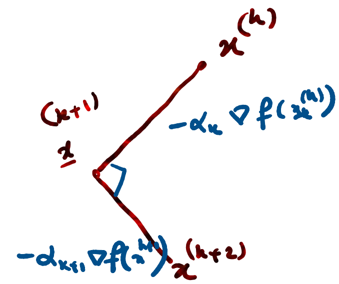
$$(\underline{x}^{(k+1)} - \underline{x}^{(k)}) \perp (\underline{x}^{(k+2)} - \underline{x}^{(k+1)})$$

Observe that the directions are orthogonal !

Is this an accident ?



$$\boxed{\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k \nabla f(x^{(k)})}$$



$$\left(\underline{x}^{(k+1)} - \underline{x}^{(k)}\right) \cdot \left(\underbrace{\underline{x}^{(k+2)} - \underline{x}^{(k+1)}}\right)$$

$$= \alpha_{k+1} \alpha_k \nabla f(x^{(k)}) \nabla f(x^{(k+1)}) = 0$$

$\underline{x}^{k+1} = \underline{x}^{(k)} - \alpha_k \nabla f(\underline{x}^{(k)})$ where α_k is a minimizer of

$$\phi_k(\alpha) = f(\underbrace{x^{(k)} - \alpha \nabla f(x^{(k)})}) \qquad \alpha \geq 0.$$

α_k satisfies

$$0 = \phi_k'(\alpha_k) = \nabla f(x^{(k+1)}) \cdot \nabla f(x^{(k)})$$

We proved : The method of steepest descent moves in perpendicular steps. If $\{\underline{x}^{(k)}\}$ is a steepest descent sequence for the function $f(\underline{x})$, then for each $k \in \mathbb{N}$, the vector joining $\underline{x}^{(k)}$ to $\underline{x}^{(k+1)}$ is orthogonal to the vector joining $\underline{x}^{(k+1)}$ to $\underline{x}^{(k+2)}$.

Drawbacks : Inherently slow ; same direction

Drawbacks: Inherently slow; same direction. Computationally inefficient.

Direction of negative gradient is the best when we start, but later on ceases to be the best; zigzag path as it pursues the search in the same directions.

Theorem: If $\{x^{(k)}\}$ is the steepest descent sequence for $f(\underline{x})$ and if $\nabla f(\underline{x}^{(k)}) \neq 0$ for some k, then

$$f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)}).$$

+ point over Newton's method.

Pf:
$$\underline{x}^{(k+1)} = x^{(k)} - \alpha_k \nabla f(\underline{x}^{(k)})$$

⊛ – α_k is the minimizer of $\phi_k(\alpha) = f(\underbrace{x^{(k)} - \alpha \nabla f(x^{(k)})})$

$\boxed{f(\underline{x}^{(k)}) = \phi_k(0)}$

$$f(\underline{x}^{(k+1)}) = \phi_k(\alpha_k) \leq \phi_k(\alpha)$$

for some α or all α $(\alpha \geq 0)$.

as ⊛ holds.

Akash

$$f(\underline{x}^{(k+1)}) = f(\underline{x}^{(k)}) + \nabla f(\underline{x}^{k})(x^{(k+1)} - \cancel{x^{(k)}}) + O(x^{(k+1)} - x^{(k)})$$

$$\Downarrow$$

$$\cancel{x^{(k)}} - \alpha_k \nabla f(x^{k})$$

$$f(x^{(k+1)}) = f(x^{(k)}) + \alpha_k \|\nabla f(x^{(k)})\|^2 + \underbrace{O(x^{(k+1)} - x^{(k)})}_{\text{Small}}$$

$$f(x^{(k+1)}) - f(x^{(k)}) = -\alpha_k \|\nabla f(x^{(k)})\|^2 < 0.$$

⊖

$$\phi_k'(\alpha) = -\nabla f(x^{(k)} - \alpha \nabla f(\underline{x}^{(k)})) \cdot \nabla f(\underline{x}^{(k)})$$

$$\phi_k(\alpha) = f(x^{(k)} - \alpha \nabla f(x^{(k)}))$$

$$\phi_k'(\alpha) = -\nabla f\left(x^{(k)} - \alpha \nabla f(x^{(k)})\right) \cdot \nabla f(x^{(k)})$$

$$\phi_k'(0) = -\|\nabla f(x^{(k)})\|^2 < 0$$

ϕ_k is a decreasing fn.

$\exists\ \bar{\alpha} > 0$ s.t. $\quad \phi_k(0) > \phi_k(\alpha) \quad 0 \leq \alpha \leq \bar{\alpha}$.

$$\Rightarrow f(x^{(k+1)}) = \phi_k(\alpha_k) \leq \varphi_k(\bar{\alpha}) < \varphi_k(0) = f(x^{(k)}).$$

Descent methods.

<u>Remark:</u>

$\min\limits_{x \in \mathbb{R}^n} f(x)$ $\quad f$ is $\begin{cases} \text{coercive} \\ \text{convex} \\ \frac{1}{2} x^T Q x - b^T x + c. \end{cases}$ | positive outcome of steepest descent.