A

Method of Steepest Descent

Gradient descent is derived from 1st order Taylor approx:

$$f(x) \approx f(x^{(t)}) + \underbrace{\nabla f(x^{(t)})^T (x - x^{(t)})}$$

To min. $f(x)$, we want to make as small as possible.

For GD, we pick $x^{(t+1)} - x^{(t)} = -\mu \nabla f(x^{(t)})$
which gives

$$f(x^{(t+1)}) \approx f(x^{(t)}) - \mu \| \nabla f(x^{(t)}) \|_2^2$$

Another interpretation: we want to find direction p that minimizes $\nabla f(x^{(t)})^T p$.

$$\Rightarrow \left(\min_{\|p\| = \delta} \nabla f(x^{(t)})^T p \right) \quad \text{for some } \delta > 0.$$

we add $\|p\| = \delta$ to bound the problem and also because we are really only interested in the direction

This is equivalent to:

$$\min_p \frac{\delta \nabla f(x^{(t)})^T p}{\|p\|}$$

B

By Cauchy-Schwarz inequality:

$$-\|\nabla f(x^{(t)})\| \cdot \|p\| \leq \nabla f(x^{(t)})^T p \leq \|\nabla f(x^{(t)})\| \cdot \|p\|$$

$$\rightarrow \left\{ -\|\nabla f(x^{(t)})\| \leq \frac{\nabla f(x^{(t)})^T p}{\|p\|} \right\}$$

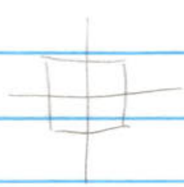
smallest possible value of

Assume we are using the <u>two-norm</u> and let $\delta = \|\nabla f(x^{(t)})\|$. Then the direction p that achieves the minimum value is:

$$\left\{ p^* = -\nabla f(x^{(t)}) \right\}$$ (ie. we get gradient descent)

Different norms give different methods of steepest descent

1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (sum of absolute values of x)



∞-norm $\|x\|_\infty = \max_{i=1 \dots n} |x_i|$ (max abs. value in x)

$$\mathrm{sign}(y) = \begin{cases} +1 & y>0 \\ 0 & y=0 \\ -1 & y<0 \end{cases}$$

c

For 1-norm

$$\min_{p} \frac{\nabla f(x^{(t)})^T p}{\|p\|_1} \Rightarrow p^* = -\mathrm{sign}(\nabla_j f(x^{(t)})) \begin{pmatrix} 0 \\ \vdots \\ \delta \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow \text{row } j \\ \\ \\ \end{array}$$

where $\delta = \|\nabla f(x^{(t)})\|_\infty = |\nabla_j f(x^{(t)})|$

j^{th} element of ∇f has the largest magnitude

This is aka "coordinate descent". We are moving along one axis at each iteration.

For ∞-norm

$$\min_{p} \frac{\nabla f(x^{(t)})^T p}{\|p\|_\infty} \Rightarrow p^* = -\delta \begin{pmatrix} \mathrm{sign}(\nabla_1 f(x^{(t)})) \\ \mathrm{sign}(\nabla_2 f(x^{(t)})) \\ \vdots \\ \mathrm{sign}(\nabla_n f(x^{(t)})) \end{pmatrix}$$

where $\delta = \|\nabla f(x^{(t)})\|_1$ and the sign of each ± 1 matches the sign of $\nabla f(x^{(t)})$.

(if $\nabla_i f(x^{(t)}) \neq 0 \ \forall i$, $p^* = -\delta \begin{pmatrix} \pm 1 \\ \vdots \\ \pm 1 \end{pmatrix}$)