SySc 512, Session 8, Dyns of Optim

Gradient Descent Search:

Goal: Find minimizer $\vec{x}^*$ of object. fuct $f(\vec{x})$

1) Choose random start: $\vec{x}_0$
2) Compute gradient @ $\vec{x}_0$: $\nabla f(\vec{x})\big|_{\vec{x}=\vec{x}_0}$
3) Compute search direction:

$$\vec{S}_0 = -\nabla f(\vec{x}_0) / \|-\nabla f(x_0)\|$$

4) Minimize one-dimensional function:

$$F(t) = f(\vec{x}_0 + t\vec{S}_0) \Rightarrow t^*$$

– Not typically though final solution!

5) Choose next point:

$$\vec{x}_1 = \vec{x}_0 + t^*\vec{S}_0 \quad , \quad t^* \text{ minimizes } F(t)$$

6) Termination test:

$$\| \vec{x}_{k+1} - \vec{x}_k \| < \varepsilon$$

7) Iterate 2 → 6 until passes test 6.

Short cut: Quadratic Form

$$f(\vec{x}) = \tfrac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T\vec{x} + \cancel{c}^{\,0}$$

$\Rightarrow$

2) $$-\nabla f(\vec{x}) = \vec{b} - A\vec{x}$$

3) Residual: $\vec{r}_i = \vec{b} - A\vec{x}_i \quad , \quad \vec{x}_i \in \{\vec{x}_0, \vec{x}_1, \dots\}$

$$\vec{S}_i = \frac{-\nabla f(\vec{x}_i)}{\|-\nabla f(x_i)\|} = + \frac{\vec{r}_i}{\|\vec{r}_i\|}$$

4) $$\frac{d}{dt} F(t) = 0 = \nabla f(\vec{x}) \cdot \dot{\vec{x}} \Big|_{\vec{x}=\vec{x}_1} = \nabla f(\vec{x}_1) \cdot \frac{\vec{r}_0}{\|\vec{r}_0\|}$$

$$\Rightarrow -\nabla f(\vec{x}_1)\cdot \vec{r}_0 = 0 \Rightarrow \vec{r}_0 \perp (-\nabla f(\vec{x}))\Big|_{\vec{x}=\vec{x}_1}$$

$$\left.\begin{aligned} &\Rightarrow (\vec{r}_1)^T \cdot \vec{r}_0 = 0 \\ &(\vec{b} - A\vec{x}_1)^T \vec{r}_0 = 0 \\ &(b - A(\vec{x}_0 + t^*\vec{r}_0))^T \vec{r}_0 = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} &(b - A\vec{x}_0)^T\vec{r}_0 = t^*(A\vec{r}_0)^T\vec{r}_0 \\ &\vec{r}_0^{\,T}\vec{r}_0 = t^*\vec{r}_0^{\,T}(A\vec{r}_0) \\ &\Rightarrow \boxed{t^* = \frac{\vec{r}_0^{\,T}\vec{r}_0}{\vec{r}_0^{\,T} A \vec{r}_0}} \end{aligned}$$

SySc 512, Session 8, Dyn Opt (cont)

Thus, for quadratic form; Iterate:

A) Calculate residual: $\vec{r}_i = \vec{b} - A\vec{x}_i$

B) " minimizer: $t_i^* = |r_i|^2 / \vec{r}_i^T A \vec{r}_i$

C) " next $\vec{x}_{i+1} = \vec{x}_i + t_i^* \vec{r}_i$

Advantage: Only one matrix calculation ($A\vec{r}_i$) per iteration; because

$$\vec{b} - A x_{i+1} = \vec{b} - A x_i + t_i^* A \vec{r}_i$$

$$\Rightarrow \vec{r}_{i+1} = \vec{r}_i - t_i^* A \vec{r}_i$$