10.4 Steepest Descent techniques

Advantage: It Converges even for poor approx.
Disadvantage: Only linear.

Consider a nonlinear System

$$\begin{cases} f_1(x_1, x_2, .., x_n) = 0 \\ \vdots \\ f_n(x_1, x_2, .., x_n) = 0 \end{cases} \qquad (1)$$

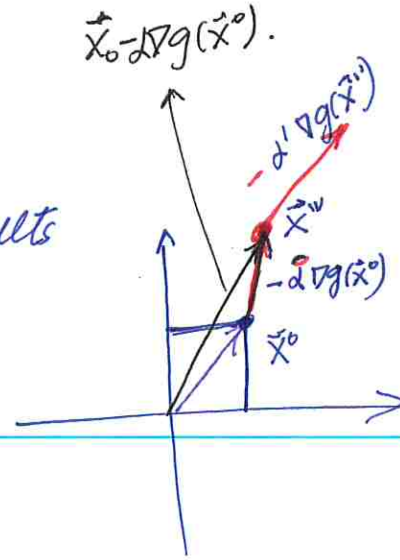
Finding the roots of (1) is equivalent to find the points where the function

$$g(x_1, x_2, .., x_n) = f_1^2(x_1, .., x_n) + \dots + f_n^2(x_1, .., x_n)$$

has the minimal value zero.

<u>Comment on why?</u>

<u>Steepest descent process</u>



1) Choose an initial guess: $\vec{x}^0$.
2) Determine a direction $\vec{u}$ from $\vec{x}^0$ that results in a decrease in the value of g.
3) $\vec{x}^1 = \vec{x}^0 + \alpha^0 \vec{u}$

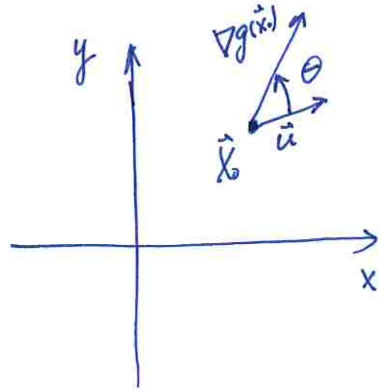
   where $g(\vec{x}^1) < g(\vec{x}^0)$.
4) Repeat steps 1–3. with $\vec{x}^0 = \vec{x}^1$.

Direction $\vec{u}$ from $\vec{X}^\circ$

Notice

Derivative of $g(\vec{x})$ in the direction $\vec{u}$.

$$D_{\vec{u}}\, g(\vec{x}^\circ) = \nabla g(\vec{x}^\circ) \cdot \vec{u} = |\nabla g(\vec{x}_0)|\,\overbrace{|\vec{u}|}^{=1} \cos\theta$$
$$= |\nabla g(\vec{x}_0)| \cos\theta$$



Then,

Maximum rate of change of g at $\vec{x}_0$ is in the direction of $\nabla g(\vec{x}_0)$ when $\theta = 0$

And $\boxed{\text{Max. rate of change} = |\nabla g(\vec{x}_0)|}$ and $\boxed{\text{the direction is } \nabla g(\vec{x}_0).}$

Similarly,

Min. rate of change of g at $\vec{x}_0$ $= -|\nabla g(\vec{x}_0)|$ when $\theta = \pi$

So $\boxed{\text{direction of min. rate of change is } -\nabla g(\vec{x}^\circ)}$

So $\vec{u} = -\nabla g(\vec{x}^\circ)$

and we consider the line: $\vec{x}^\circ + \alpha\vec{u} = \vec{x}^\circ - \alpha \nabla g(\vec{x}^\circ)$

our goal is to find $\alpha = \alpha^\circ$ such that

$$h(\alpha) = g\left(\vec{x}^\circ - \alpha \nabla g(\vec{x}^\circ)\right) \text{ is minimum. when } \alpha = \alpha_0$$

And the new approximation to the point where g reaches a minimum is

$$\vec{X}^{(1)} = \vec{X}^\circ - \alpha^\circ \nabla g(\vec{x}^\circ)$$

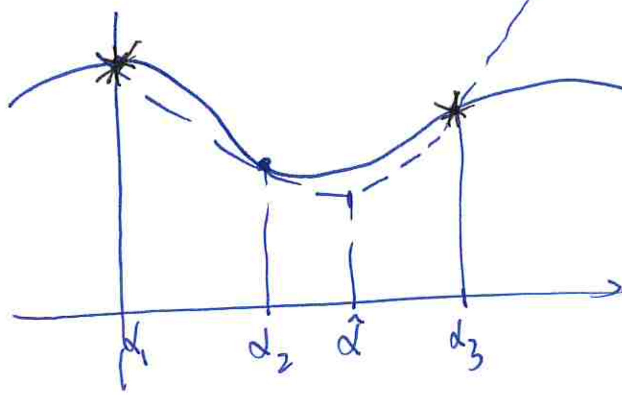
$h(\alpha)$ in general will be a strongly nonlinear function of α.

Now, the minimum value of $h(\alpha)$ is approx. considering a parabola $P_2(\alpha)$ interpol... $h(\alpha)$ at the points $(\alpha_1, h(\alpha_1)), \dots, (\alpha_3, h(\alpha_3))$

where $\alpha_1 = 0$, $\alpha_2 = \alpha_3/2$, α_3 (Usually $\alpha_3 = 1$)

and α_3 is chosen such that $h(\alpha_3) < h(\alpha_1)$.

$\hat{\alpha} \in (\alpha_1, \alpha_3]$.



<u>Example:</u> $\vec{F}(x,y) = \left(10(y-x^2),\ 1-x\right) = \vec{0}$ $\quad \begin{cases} y = x^2 \end{cases}$ $\quad \begin{matrix} x = 1 \\ y = 1 \end{matrix}$

$$\Rightarrow \quad g(\vec{x}) = 100(y-x^2)^2 + (1-x)^2 \qquad \boxed{\vec{x}^0 = (0.7, 0.7)}$$

$$g(0.7, 0.7) = 4.5.$$

$$\nabla g(\vec{x}) = \left(200(y-x^2)\cdot(-2x) - 2(1-x),\ 200(y-x^2)\right)$$

$$\nabla g(0.7, 0.7) \simeq (-59.4, 42)$$

$$g\left(\vec{x}^0 - \alpha \nabla g(\vec{x}^0)\right) = g\left((0.7,0.7) - \alpha(-59.4, 42)\right) = \underset{\text{along } \nabla g(\vec{x}^0) \text{ direction}}{\overset{\text{min.}}{g}}$$

is obtained at $\alpha^0 \simeq 1.6 \times 10^{-3}$.

$$\Rightarrow \quad \vec{x}^{(1)} = \vec{x}^0 - \alpha^0 \nabla g(\vec{x}^0) \simeq (0.796, 0.632)$$

$\vec{x}^{(k)}$ moves to the root very slow!

Constructing the parabola $P_2(\alpha)$ approximating $h(\alpha)$

It requires three interpolating points

$$(\alpha_1, h(\alpha_1)), \quad (\alpha_2, h(\alpha_2)), \quad (\alpha_3, h(\alpha_3)).$$

They are chosen as follows:

i) Choose $\alpha_1 = 0$ and compute

$$h(\alpha_1) = g(\vec{x}^0 - \overset{=0}{\alpha_1} \nabla g(\vec{x}^0)) = g(\vec{x}^0).$$

So, $(\alpha_1, h(\alpha_1)) = (0, g(\vec{x}^0))$

ii) Choose $\alpha_3 = 1$ and compare

$$g(\alpha_1) \text{ and } g(\alpha_3).$$

If $\quad g(\alpha_3) < g(\alpha_1)$

Choose $\alpha_2 = \alpha_3/2$

and define: $(\alpha_2, h(\alpha_2)) = (1/2, h(1/2))$

$$(\alpha_3, h(\alpha_3)) = (1, h(1))$$

iii) If $\quad g(\alpha_3) > g(\alpha_1)$, then $\alpha_3^2 = \alpha_3/2$

do this until $g(\alpha_3^k) < g(\alpha_1)$

If not possible after a maximum number of ^{k} divisions by 2 is reached accept α_1 as the minimum of $h(\alpha)$.

iv) If α_3^k is such that

$$g(\alpha_3^k) < g(\alpha_1) = g(0).$$

then the three interpolating points are:

$$(\alpha_1, g(\alpha_1)) = (0, g(\vec{x}^0))$$

$$(\alpha_2\ g(\alpha_2)) = \left(\alpha_3^k/2, g\left(\vec{x}^0 - \frac{\alpha_3}{2}\nabla g(\vec{x}^0)\right)\right)$$

$$(\alpha_3, g(\alpha_3)) = \left(\alpha_3^k, g\left(\vec{x}^0 - \alpha_3^k \nabla g(\vec{x}^0)\right)\right)$$

v) Construction of $P_2(\alpha)$ using forward divided-diff.

$$P_2(\alpha) = a_0 + a_1(\alpha - \overset{=0}{\alpha_1}) + a_2(\alpha - \overset{=0}{\alpha_1})(\alpha - \alpha_2)$$

$\alpha_1 = 0$ $\quad \boxed{g(\alpha_1)}$

$\qquad \boxed{h_1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1}}$

α_2 $\quad g(\alpha_2)$ $\qquad \boxed{\frac{h_2 - h_1}{\alpha_3 - \alpha_1} = h_3}$

$\qquad h_2 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2}$

α_3 $\quad g(\alpha_3)$

Then,
$$\boxed{P_2(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2).}$$

vi) Critical point of $P_2(\alpha)$ is when

$$P_2'(\hat{\alpha}) = h_1 - h_3\alpha_2 + 2h_3\hat{\alpha} = 0$$

$$\Rightarrow \quad \hat{\alpha} = \frac{h_3\alpha_2 - h_1}{2h_3} = \frac{1}{2}\left(\alpha_2 - \frac{h_1}{h_3}\right).$$

VII) If $P_2(\hat{\alpha}) < P_2(\alpha_3)$

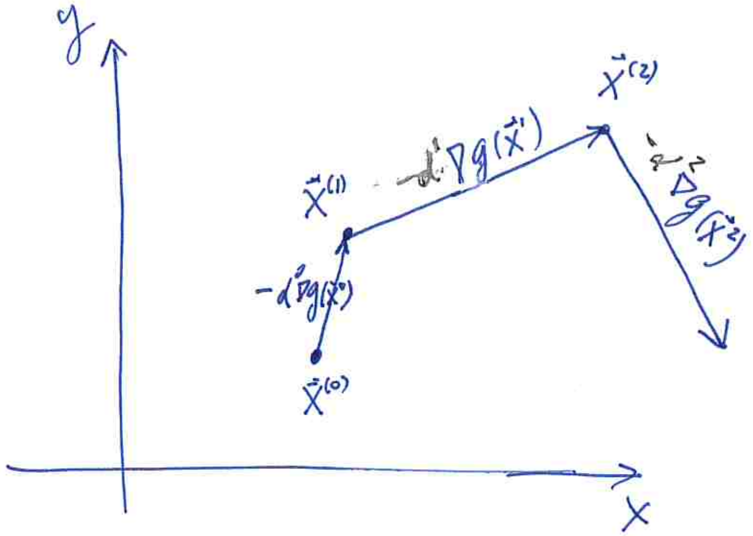
accept $\hat{\alpha}$ as the point
Where $h(\alpha)$ reaches a minimum

and define

$$\vec{X}^1 = \vec{X}^0 - \hat{\alpha}\,\nabla g(\vec{X}^0)$$

VIII) Repeat this process for $\vec{X}^k$, $k = 2, \ldots k_{max}$.

SD_4



y
$\vec{x}^{(2)}$
$-\alpha^1 \nabla g(\vec{x}^1)$
$\vec{x}^{(1)}$
$-\alpha^2 \nabla g(\vec{x}^2)$
$-\alpha^0 \nabla g(\vec{x}^0)$
$\vec{x}^{(0)}$
x