Line search techniques for unconstrained optimization problems

- Algorithm:
  - Step 0 (Initialization/Inputs): Choose f, x^0 (initialization), $\varepsilon(> 0)$, β_1, β_2 (0<β_1<$\beta_2 < 1$) and $r \in (0,1)$. (β_1, β_2, and r are required only if we use inexact line search technique). Set $k := 0$.
  - Step 1 (Optimality/Stopping condition check): Compute $\nabla f(x^0)$. If $\|\nabla f(x^0)\| < \varepsilon$ then stop. Otherwise go to Step 2.
  - Step 2 (Compute descent direction) : Use different technique to compute suitable descent direction d^k.
  - Step 3 (Step length selection): Use exact/inexact (Armijo-Wolfe backtracking) line search technique to find suitable step length ($\alpha_k > 0$).
  - Step 4 (Update): Update x^{k+1} by $x^{k+1=x^k+\alpha_k}d^k$. Set $k := k + 1$ and go to Step 1.
- Output: An approximate stationary point.

- **Steepest Descent Method:**
  - In this method we use $\boldsymbol{d^k = -\nabla f(x^k)}$.
  - Clearly d^k is a descent direction since

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

- **Example:**
  - Suppose $f(x) = 4x_1^2 + x_2^2 - 2x_1x_2$.
  - Then $\nabla f(x) = \begin{pmatrix} 8x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{pmatrix}$. Clearly $\nabla f(x) = 0$ implies $x_1 = 0 = x_2$.
  - Now $\nabla^2 f(x) = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$.
  - At $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\nabla^2 f(x)$ is positive definite (verify).
  - Hence $x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is a local minima of f

Solve $\boldsymbol{min_{x\in R^2} f(x) = 4x_1^2 + x_2^2 - 2x_1x_2}$ ***using steepest descent method.***

- Choose initial approximation $x^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, $\beta_1 = 10^{-4}$, $\beta_2 = 0.9$, $r = 0.5$, and $\varepsilon = 10^{-3}$.
- Now $\nabla f(x^0) = \begin{pmatrix} 12 \\ 0 \end{pmatrix}$, clearly $||\nabla f(x^0)|| = 12 > \varepsilon$. So we calculate d^0 and proceed.
- $d^0 =- \nabla f(x^0) = \begin{pmatrix} -12 \\ 0 \end{pmatrix}$.
- Next select step length using exact line search technique:
  - $\alpha_0 = argmin_{\alpha>0}\, \phi(\alpha) = 4(2-12\alpha)^2 + 4 - 2(2-12\alpha)2$
  - $\phi'(\alpha) =- 96(2-12\alpha) + 48$
  - If $\alpha_0$is a minimizer then $\phi'(\alpha_0) = 0$. This implies$\alpha_0 = 0.125$

- So $x^1 = x^0 + \alpha_0 d^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + 0.125 \begin{pmatrix} -12 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 2 \end{pmatrix}$
- Clearly $f(x^1) = 3 < f(x^0) = 12$
- Now $\nabla f(x^0) = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$, clearly $\|\nabla f(x^0)\| = 3 > \varepsilon$. So we calculate d^1 and proceed.
- $d^1 =- \nabla f(x^1) = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$.
- Next select step length using exact line search technique:
  - $\alpha_1 = argmin_{\alpha>0}\, \phi(\alpha) = 1 + (2 - 3\alpha)^2 - 2 * 0.5(2 - 3\alpha)$
  - $\phi'(\alpha) =- 6(2 - 3\alpha) + 3$
  - If α_1 is a minimizer then $\phi'(\alpha_1) = 0$. This implies $\alpha_1 = 0.5$
- Then $x^2 = x^1 + \alpha_1 d^1 = \begin{pmatrix} 0.5 \\ 2 \end{pmatrix} + 0.5 \begin{pmatrix} 0 \\ -3 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$
- Clearly $f(x^2) = 0.75 < f(x^1)$ =0

- Now $\nabla f(x^2) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $d^2 = \begin{pmatrix} -3 \\ 0 \end{pmatrix}$.
- Observe that
  - $(d^0)^T d^1 = (-12,0)\begin{pmatrix} 0 \\ -3 \end{pmatrix} = 0$
  - $(d^1)^T d^2 = (0,-3)\begin{pmatrix} -3 \\ 0 \end{pmatrix} = 0$
- **In steepest descent method two consecutive descent directions are perpendicular to each other if step lengths are selected using exact line search technique.**

Solve $\boldsymbol{min_{x \in R^2} f(x) = (x_1 - 2)^2 + (2x_2 - x_1)^2}$ ***using steepest descent method (using Armijo-Wolfe inexact line search technique)***

- Choose initial approximation $x^0 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \beta_1 = 10^{-4}, \beta_2 = 0.9, r = 0.5$, and $\varepsilon = 10^{-3}$.
- We, $f(x^0) = 10$.
- Now $\nabla f(x^0) = \begin{pmatrix} -4 \\ 12 \end{pmatrix}$, clearly $||\nabla f(x^0)|| = 12.65 > \varepsilon$.

  So we calculate d^0 and proceed.
- $d^0 =- \nabla f(x^0) = \begin{pmatrix} 4 \\ -12 \end{pmatrix}$.

- Select step length using inexact line search technique:
  - For $\alpha = 1$, $f(x^0 + \alpha d^0) = 650 > f(x^0) + \alpha\beta_1 \nabla f(x^0)^T d^0$.
  - Update $\alpha = \alpha$r=0.5
  - For $\alpha = 0.5$, $f(x^0 + \alpha d^0) = 130 > f(x^0) + \alpha\beta_1 \nabla f(x^0)^T d^0$.
  - Update $\alpha = \alpha$r=0.25
  - For $\alpha = 0.25$, $f(x^0 + \alpha d^0) = 20 > f(x^0) + \alpha\beta_1 \nabla f(x^0)^T d^0$.
  - Update $\alpha = \alpha$r=0.125
  - For $\alpha = 0.125$, $f(x^0 + \alpha d^0) = 2.5 < f(x^0) + \alpha\beta_1 \nabla f(x^0)^T d^0$.
  - Also $\nabla f(x^0 + \alpha d^0)^T d^0 > \beta_2 \nabla f(x^0)^T d^0$ holds for $\alpha = 0.125$
  - So we choose $\alpha_0 = 0.125$

- Then $x^1 = x^0 + \alpha_0 d^0 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} + 1/8 \begin{pmatrix} 4 \\ -12 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 1.5 \end{pmatrix}$
- Clearly $f(x^1) = 2.5 < f(x^0)$ =10
- Final solution using stopping criteria $\|\nabla f(x^k)\| < \varepsilon$ is

$$x^{26} = \begin{pmatrix} 2.00022 \\ 1.00017 \end{pmatrix} \cong \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

- Major limitation of steepest descent method is: it converges linearly that is rate of convergence is 1.