Conjugate Gradient Method

- A set of nonzero vectors $\{d^1, d^2, \dots, d^n\}$ is said to be conjugate with respect to a matrix H if ${d^i}^T H d^j = 0$ for all $i \neq j$.
- In conjugate gradient method, the descent direction is defined as

$$d^k = -\nabla f\left(x^k\right) + \beta_k d^{k-1}$$

where

$$\beta_k = \frac{\left\|\nabla f\left(x^k\right)\right\|^2}{\|\nabla f(x^{k-1})\|^2}$$

- **If the objective function is quadratic with positive definite Hessian, then descent directions of conjugate gradient method (using exact line search technique) are conjugate with respect to Hessian.**

- Algorithm:
  - Step 0 (initialization): Choose objective function, initial approximation (x^0), $\varepsilon > 0$. Set $k := 0$
  - Step 1 (optimality check): If $\|\nabla f(x^k)\| < \varepsilon$, then stop. Otherwise go to Step 2.
  - Step 2 (descent direction calculation): Calculate $d^k = -\nabla f(x^k) + \beta_k d^{k-1}$

where $\beta_k = \begin{cases} 0, & if\ k = 0 \\ \dfrac{\|\nabla f(x^k)\|^2}{\|\nabla f(x^{k-1})\|^2} & if\ \ k > 0. \end{cases}$

  - Step 3 (step length selection): Select step length $\alpha_k > 0$ using exact line search technique. i.e. $\alpha_k = argmin_{\alpha>0}\, f(x^k + \alpha d^k)$.
  - Step 4 (update): Update $x^{k+1} = x^k + \alpha_k d^k$. Set $k := k + 1$ and go to Step 1.

- Consider $f(x) = 4x_1^2 + x_2^2 - 2x_1x_2$.
- Then $\nabla f(x) = \begin{pmatrix} 8x_1 - 2x_2 \\ -2x_1 + 2x_2 \end{pmatrix}$ and $\nabla^2 f(x) = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$. Clearly $\nabla^2 f(x)$ is positive definite.
- Choose $x^0 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. Then $\nabla f(x^0) = \begin{pmatrix} 10 \\ 2 \end{pmatrix}$ and $d^0 = -\nabla f(x^0) = \begin{pmatrix} -10 \\ -2 \end{pmatrix}$
- Next select $\alpha_0 = argmin_{\alpha>0}\, \varphi(\alpha) = f(x^0 + \alpha d^0)$
- For min of $\varphi(\alpha)$, $\varphi'(\alpha_0) = 0$. This implies ${d^0}^T \nabla f\,(x^0 + \alpha_0 d^0) = 0$.
- This implies $\begin{pmatrix} -10 & -2 \end{pmatrix} \begin{pmatrix} 8(2 - 10\alpha_0) - 2(3 - 2\alpha_0) \\ -2(2 - 10\alpha_0) + 2(3 - 2\alpha_0) \end{pmatrix} = 0$
- i.e. $\begin{pmatrix} -10 & -2 \end{pmatrix} \begin{pmatrix} 10 - 76\alpha_0 \\ 2 + 16\alpha_0 \end{pmatrix} = 0 \Rightarrow \alpha_0 = \frac{1}{7}$

- Now $x^1 = x^0 + \alpha_0 d^0 = \begin{pmatrix} 4/7 \\ 19/7 \end{pmatrix}$ and $\nabla f(x^1) = \begin{pmatrix} -0.8571 \\ 4.2857 \end{pmatrix}$
- Then $\beta_1 = 0.1837$ and $d^1 = -\nabla f(x^1) + \beta_1 d^0 = \begin{pmatrix} -0.9796 \\ -4.6531 \end{pmatrix}$.
- Now ${d^0}^T \nabla^2 f(x) d^1 = 0.000032 \approx 0$
- Similar to α_0, we select $\alpha_1 = 0.5833$.
- Hence $x^2 = \begin{pmatrix} -0.1843 \\ -0.0460 \end{pmatrix} * 10^{-6} \approx \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
- and $\|\nabla f(x^2)\| = 0.0000014 \approx 0.$
- One can observe that $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is a local minima of f. Since f is strictly convex for all x, $x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is the global minima of f.