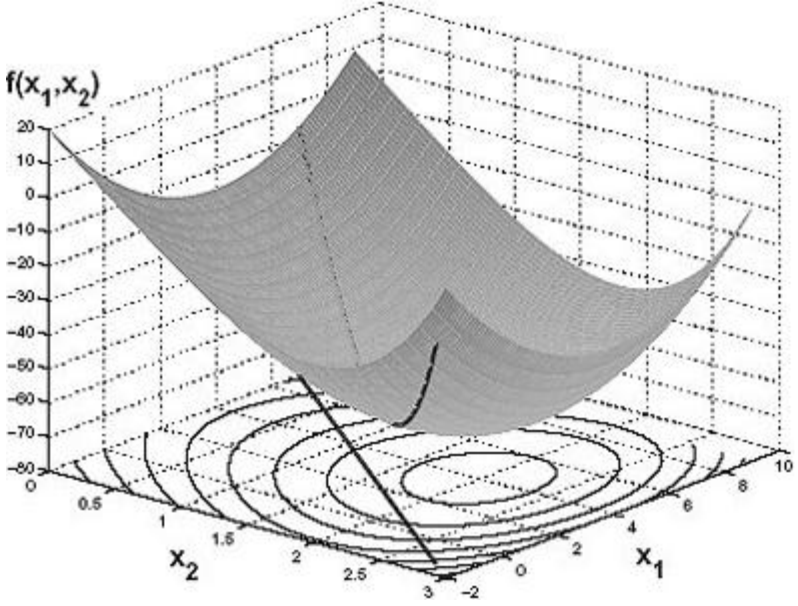
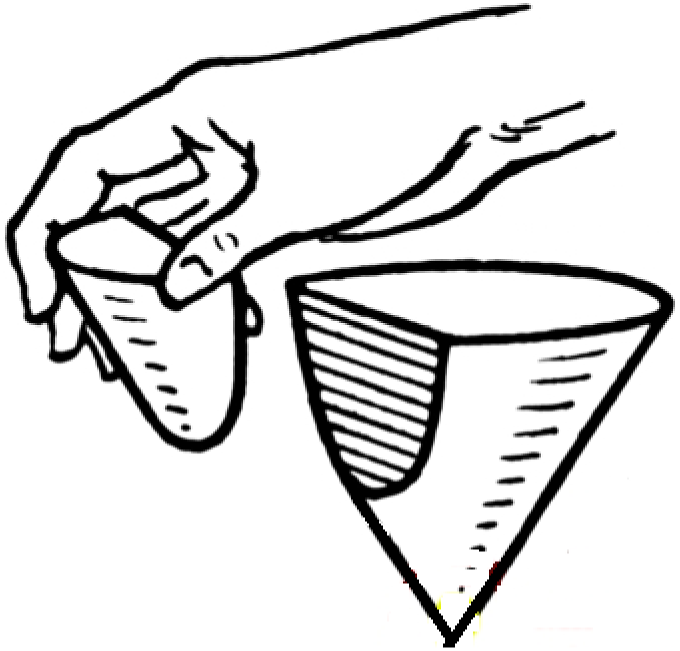
Multivariable problem

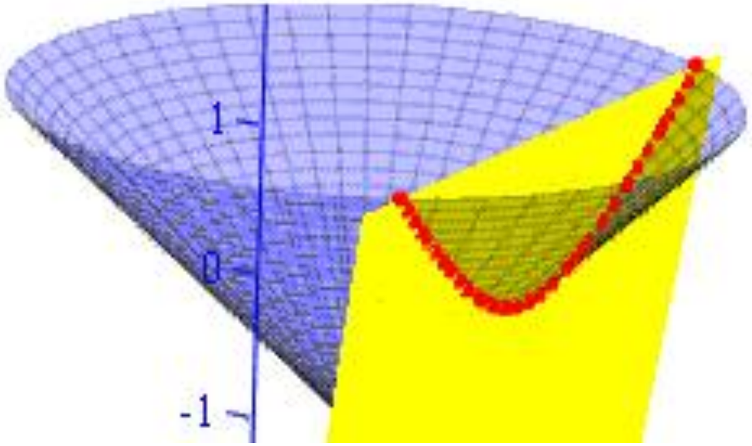
Prof. (Dr.) Rajib Kumar Bhattacharjya

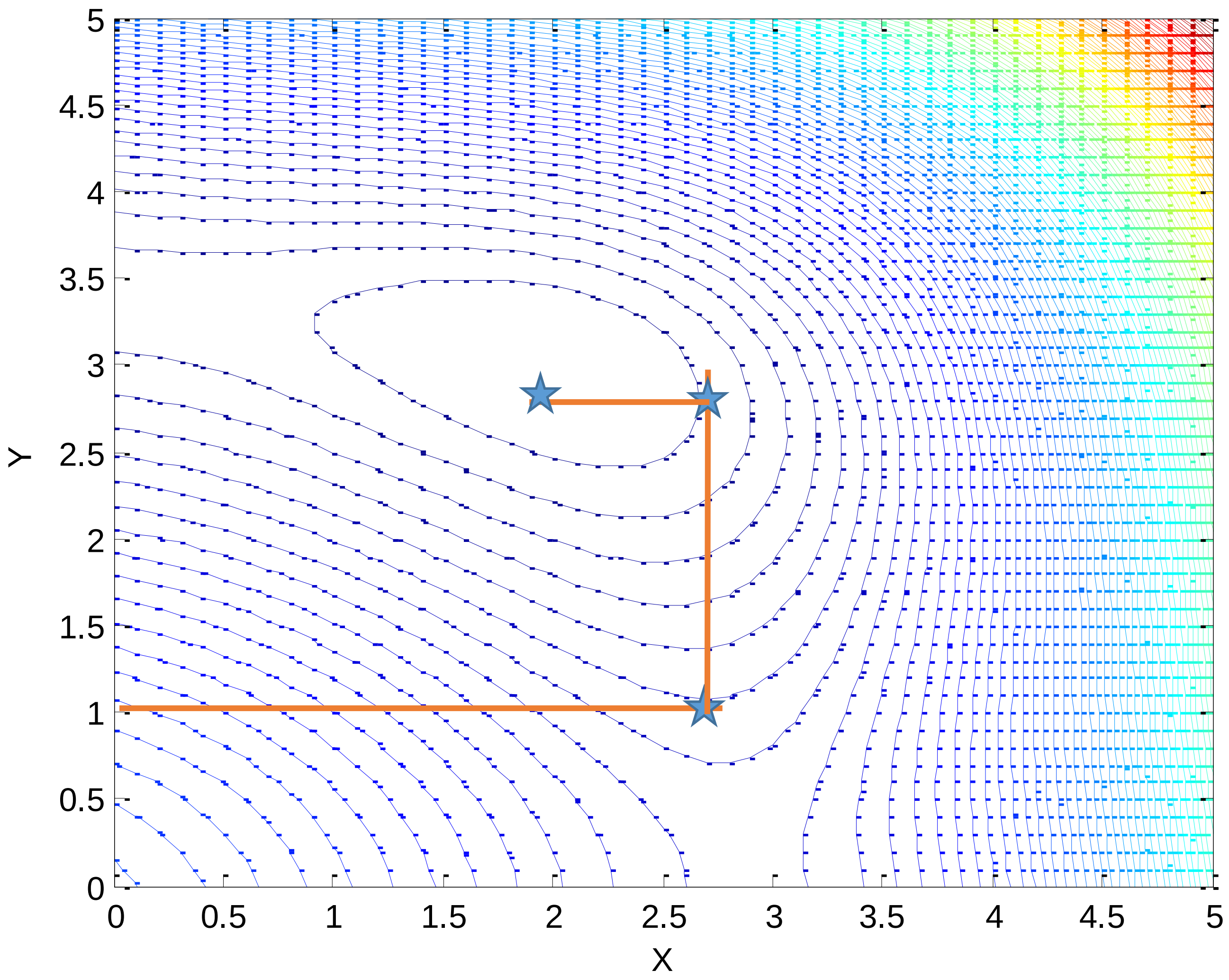
Professor, Department of Civil Engineering
Indian Institute of Technology Guwahati, India
Room No. 005, M Block
Email: rkbc@iitg.ernet.in, Ph. No 2428



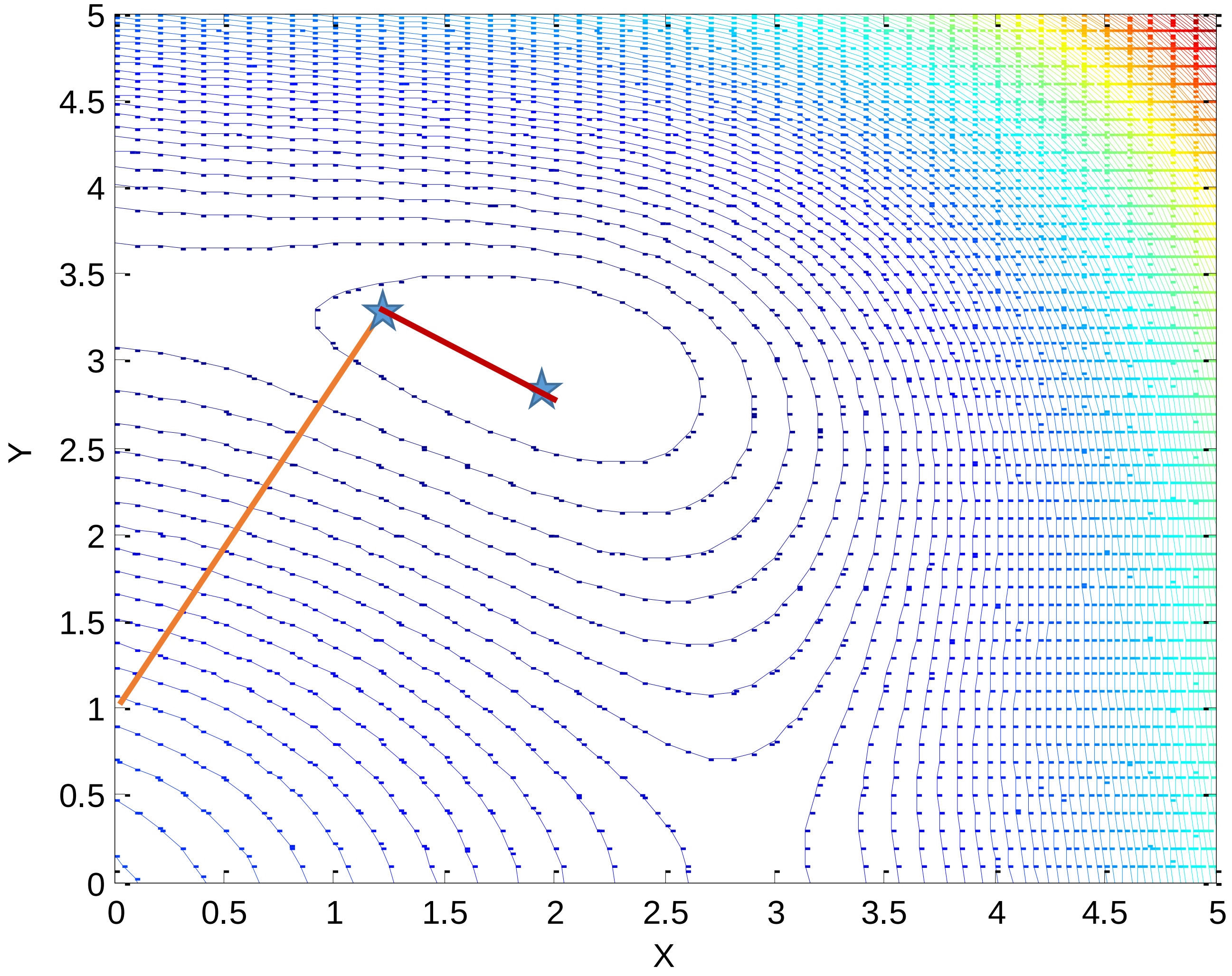
f(x1,x2)
x2
x1



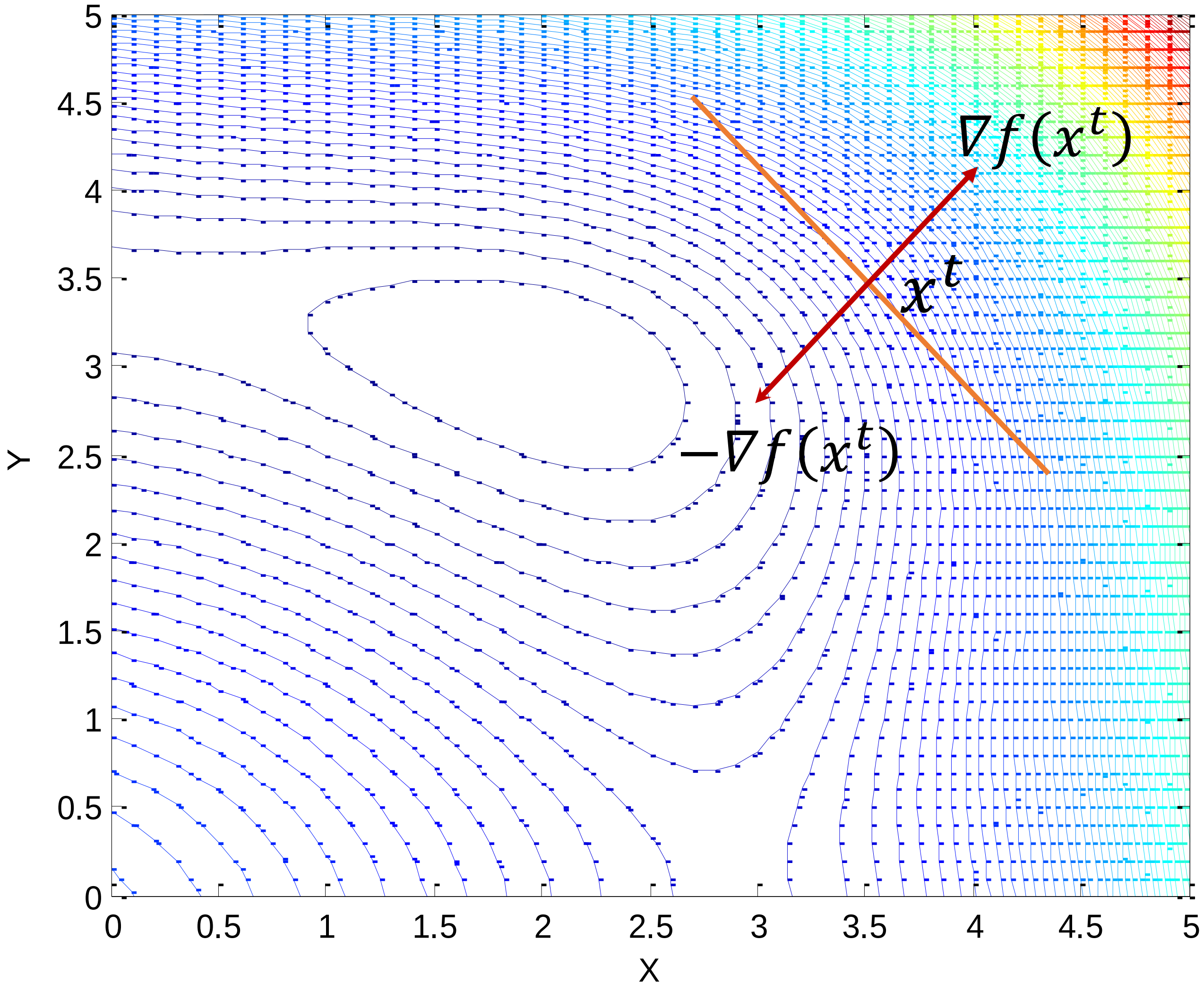




5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Y
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
X



Y
X
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5



Descent direction

A search direction d^t is a descent direction at point x^t if the condition $\nabla f(x^t).d^t < 0$ is satisfied in the vicinity of the point x^t.

$$f(x^{t+1}) = f(x^t + \alpha d^t)$$
$$= f(x^t) + \alpha \nabla^T f(x^t).d^t$$

The $f(x^{t+1}) < f(x^t)$

When $\alpha \nabla^T f(x^t).d^t < 0$

Or, $\nabla^T f(x^t).d^t < 0$

Newton's method

Taylor series $$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{n-1}(x) + \frac{h^n}{n!}f^n(x)$$

$$f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i)f'(x_i) + \frac{(x_{i+1} - x_i)^2}{2!}f''(x_i)$$

By setting derivative of the equation to zero for minimization of $f(x)$, we have

$$f'(x_{i+1}) = 0 + f'(x_i) + f''(x_i)(x_{i+1} - x_i) = 0$$

$$\boxed{x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}}$$

Quasi Newton's method

Sometime it is not possible to have the closed form expression of the function or it may be difficult to calculate the derivatives of the objective function. In such a scenarios, the derivative of the function can be calculated

$$f'(x_i) = \frac{f(x_i + \Delta x) - f(x_i - \Delta x)}{2\Delta x}$$

$$f''(x_i) = \frac{f(x_i + \Delta x) - 2f(x_i) + f(x_i - \Delta x)}{\Delta x^2}$$

$$\boxed{x_{i+1} = x_i - \frac{\Delta x[f(x_i + \Delta x) - f(x_i - \Delta x)]}{2[f(x_i + \Delta x) - 2f(x_i) + f(x_i - \Delta x)]}}$$

Convergence $|f'(x_i)| \leq \in$

Newton's method for multi-variable problem

Taylor series $f(X+h) = f(X) + h^T \nabla f(X) + \frac{1}{2!} h^T H h + \cdots$

$$f(X_{i+1}) = f(X_i) + \nabla f(X_i)^T (X_{i+1} - X_i) + \frac{1}{2!}(X_{i+1} - X_i)^T H (X_{i+1} - X_i) + \cdots$$

By setting partial derivative of the equation to zero for minimization of $f(X)$, we have

$$\nabla f = 0 + \nabla f(X_i) + H(X_{i+1} - X_i) = 0$$

$$\boxed{X_{i+1} = X_i - H^{-1} \nabla f}$$

Since higher order derivative terms have been neglected, the above equation can be iteratively used to find the value of the optimal solution

$$\frac{\partial (X^T A X)}{\partial X} = AX + A^T X$$

In this case

$$\frac{\partial (X^T A X)}{\partial X} = 2AX$$

$$\frac{\partial (AX)}{\partial X} = A^T$$

$$\frac{\partial (X^T A)}{\partial X} = A$$

$$\frac{\partial (A^T X)}{\partial X} = A$$

Q. Show that the Newton's method finds the minimum of a quadratic function in one iteration

A quadratic function can be written as

$$f(X) = \frac{1}{2}X^T AX + B^T X + \mathrm{C}$$

The minimum of the function is given by

$$\nabla f(X) = AX + B = 0$$

$$\mathrm{X} = -A^{-1}B$$

Now apply Newton's method. The iterative step gives

$$\mathrm{X}_{i+1} = \mathrm{X}_i - H^{-1}\nabla f(\mathrm{X}_i)$$

In this case $H = A$

$$\mathrm{X}_{i+1} = \mathrm{X}_i - A^{-1}(A\mathrm{X}_i + B)$$

$$\mathrm{X}_{i+1} = -A^{-1}B$$

$$\frac{\partial(X^T AX)}{\partial X} = AX + A^T X$$

In this case $A = A^T$

$$\frac{\partial(X^T AX)}{\partial X} = 2AX$$

$$\frac{\partial(AX)}{\partial X} = A^T$$

$$\frac{\partial(X^T A)}{\partial X} = A$$

$$\frac{\partial(A^T X)}{\partial X} = A$$

Marquardt Method

Steepest descent method $\boxed{X_{i+1} = X_i + \alpha[-\nabla f_i\ ]}$ Where $S_i = -\nabla f$

Newton's method $\boxed{X_{i+1} = X_i + \alpha\left[-{H_i}^{-1}\nabla f_i\right]}$ Where $S_i = -{H_i}^{-1}\nabla f_i$

Now we can combine these two method

$$\boxed{X_{i+1} = X_i + \alpha[-[H_i + \beta I]^{-1}\nabla f_i]}$$

For the large value of β, the effect of hessian matrix will be negligible and the method will be similar to steepest descent method. On the other hand when β is equal to zero, the method is similar to Newton's method.

Quasi Newton's method

Newton's method

$$X_{i+1} = X_i - [H_i]^{-1} \nabla f(X_i)$$

Considering $[A_i] = [H_i]$ and $[B_i] = [H_i]^{-1}$

$$X_{i+1} = X_i - [B_i] \nabla f(X_i)$$

We have

$$\nabla f(X) = \nabla f(X_0) + H(X - X_0)$$

Now consider two points X_i and X_{i+1}

$$\nabla f_{i+1} = \nabla f(X_0) + [A_i](X_{i+1} - X_0)$$

$$\nabla f_i = \nabla f(X_0) + [A_i](X_i - X_0)$$

Subtracting these two equations, we have

$$[A_i](X_{i+1} - X_i) = \nabla f_{i+1} - \nabla f_i$$

$$[A_i] d_i = g_i \qquad \Longrightarrow d_i = [B_i] g_i$$

Now $[B_i]$ needs to be updated in each iteration. It can be updated

$$[B_{i+1}] = [B_i] + [\Delta B_i]$$

Theoretically $[\Delta B_i]$ can have its rank as high as n. However, in practice we only use rank 1 or 2

$$[\Delta B_i] = czz^T \qquad [B_{i+1}] = [B_i] + czz^T$$

We have $d_i = [B_{i+1}]g_i \qquad \Rightarrow d_i = \big[[B_i] + czz^T\big]g_i = [B_i]g_i + cz(z^T g_i)$

Since $z^T g_i$ is a scalar, we have $\qquad cz = \dfrac{d_i - [B_i]g_i}{z^T g_i}$

The simple choice for z and c would be $\qquad c = \dfrac{1}{z^T g_i} \qquad z = d_i - [B_i]g_i$

Thus, the rank 1 update will be $\qquad [B_{i+1}] = [B_i] + \dfrac{zz^T}{z^T g_i}$

$$\boxed{\Rightarrow [B_{i+1}] = [B_i] + \frac{[d_i - [B_i]g_i][d_i - [B_i]g_i]^T}{[d_i - [B_i]g_i]^T g_i}}$$

Rank 2 update

$$[\Delta B_i] = c_1 z_1 {z_1}^T + c_2 z_2 {z_2}^T$$

$$[B_{i+1}] = [B_i] + c_1 z_1 {z_1}^T + c_2 z_2 {z_2}^T$$

$$d_i = [B_{i+1}] g_i \qquad d_i = \left[[B_i] + c_1 z_1 {z_1}^T + c_2 z_2 {z_2}^T\right] g_i$$

$$d_i = [B_i] g_i + c_1 z_1 ({z_1}^T g_i) + c_2 z_2 ({z_2}^T g_i)$$

$$d_i - [B_i] g_i = c_1 z_1 ({z_1}^T g_i) + c_2 z_2 ({z_2}^T g_i)$$

The following choices can be made $\quad c_1 = \dfrac{1}{{z_1}^T g_i} \qquad c_2 = -\dfrac{1}{{z_2}^T g_i}$

$$z_1 = d_i \qquad z_2 = [B_i] g_i$$

$$[B_{i+1}] = [B_i] + \frac{z_1 {z_1}^T}{{z_1}^T g_i} - \frac{z_2 {z_2}^T}{{z_2}^T g_i} = [B_i] + \frac{d_i {d_i}^T}{{d_i}^T g_i} - \frac{\left[[B_i] g_i\right]\left[[B_i] g_i\right]^T}{\left[[B_i] g_i\right]^T g_i}$$

Powell's conjugate direction method

Parallel subspace property

Given a quadratic function $f(X) = \frac{1}{2}X^T AX + B^T X + C$ of two variables and X^1 and X^2 are the two arbitrary but distinct points.

If Y^1 is the solution of the problem

$$\text{Min } f(X^1 + \lambda d)$$

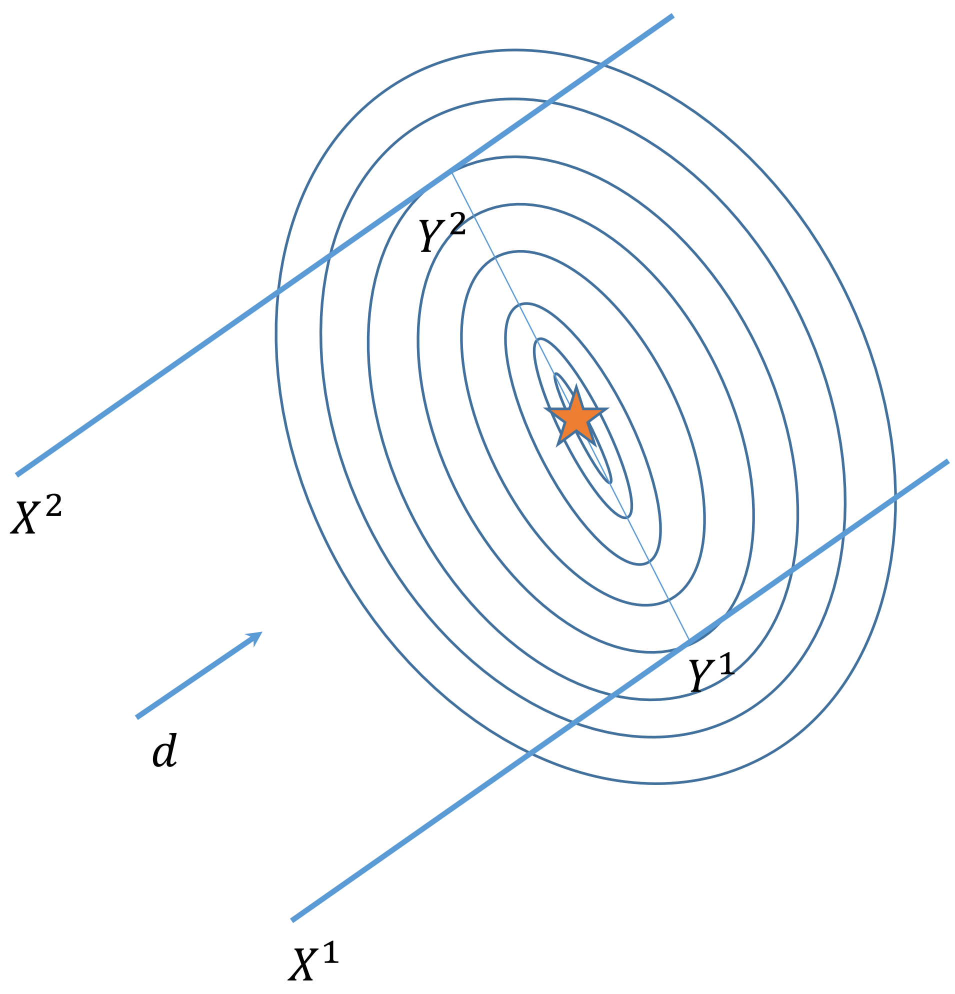
If Y^2 is the solution of the problem

$$\text{Min } f(X^2 + \lambda d)$$

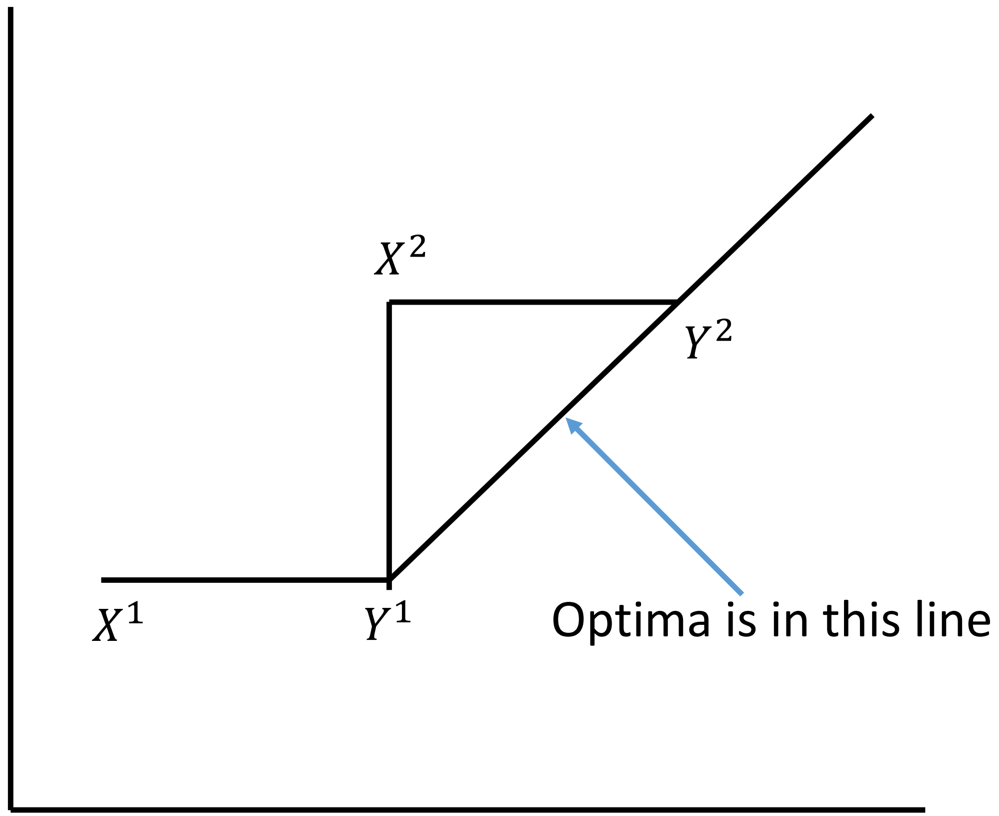
Then the direction $(Y^2 - Y^1)$ is conjugate to d, or other words, the quantity

$$(Y^2 - Y^1)^T Ad = 0$$

For quadratic function minimum lies on the direction $(Y^2 - Y^1)$



Y^2
X^2
Y^1
d
X^1



X^2
Y^2
X^1
Y^1
Optima is in this line