Gradient-Based Non-Linear Optimization – Lecture 04

Dr. Serhat Hosder

Associate Professor of Aerospace Engineering

Mechanical and Aerospace Engineering

290B Toomey Hall

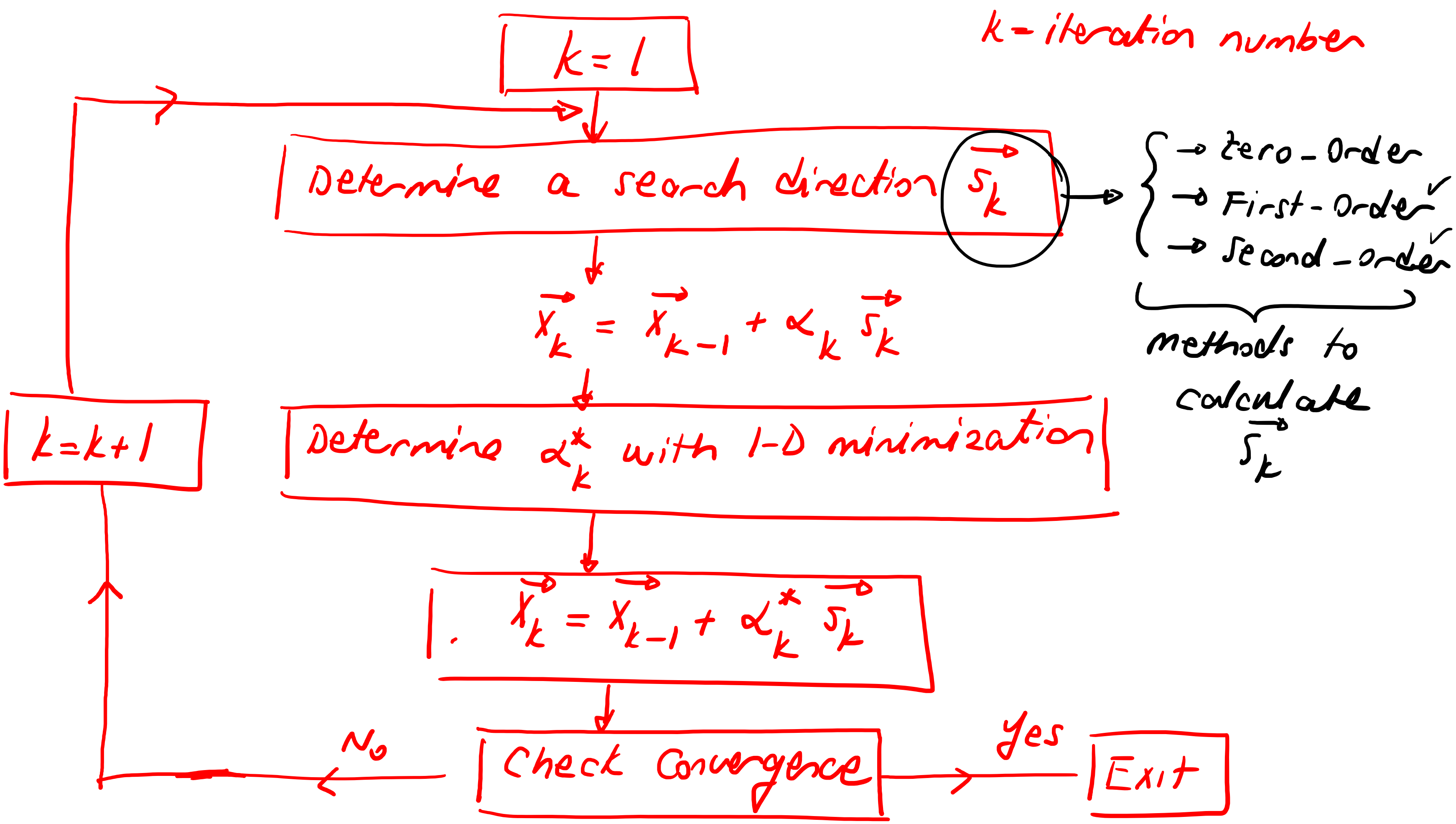
Missouri S&T

Rolla, MO 65409

Phone: 573-341-7239

E-mail: hosders@mst.edu

General Optimization Algorithm for Multi-Dimensional Unconstraint Optimization Problems :



MISSOURI S&T

First-Order Methods:

① Steepest Descent Method:

$\vec{X}$ = Design variable vector
$F(\vec{x})$ = objective function
$\vec{S}_k$ = search direction

$$\vec{X}_k = \vec{X}_{k-1} + \alpha_k^* \vec{S}_k \quad \text{where} \quad \vec{S}_k = -\frac{\nabla F(\vec{x}_{k-1})}{\|\nabla F(\vec{x}_{k-1})\|_2}$$

$\|\nabla F(\vec{x}_{k-1})\|_2$ → L_2 norm of $\nabla F(\vec{x}_{k-1})$

In general α_k^* is obtained with 1-D minimization.

If $F(\vec{x})$ is a quadratic function of "n" variables

Ex: $n=1 \rightarrow F(x) = a_0 x^2 + b_0 x + c_0$

Ex: $n=3 \rightarrow F(x_1,x_2,x_3) = a_0 x_1^2 + b_0 x_2^2 + c_0 x_3^2 + d_0 x_1 x_2 + e_0 x_1 x_3 + f_0 x_2 x_3 + g_0 x_1 + h_0 x_2 + i_0 x_3 + j_0$

$$\boxed{F(\vec{x}) = \frac{1}{2}(\vec{x})^T \cdot [Q] \cdot \vec{x} + (\vec{b})^T \cdot \vec{x} + c}$$ ⇐ general form

$[Q]$ = Hessian matrix
f_0 – a quadratic function

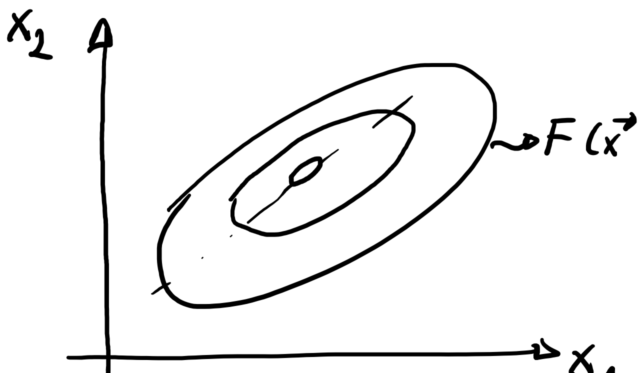
$$F(\vec{x}) = \frac{1}{2} \cdot \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix}^T \cdot \begin{bmatrix} 2a_0 & d_0 & e_0 \\ d_0 & 2b_0 & f_0 \\ e_0 & f_0 & 2c_0 \end{bmatrix} \cdot \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} + \begin{Bmatrix} g_0 \\ h_0 \\ i_0 \end{Bmatrix}^T \cdot \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} + j_0$$

The performance of the Steepest Descent Method will depend on the condition number of [Q] matrix.

$\underset{\longrightarrow \text{Hessian}}{[Q]\ \text{matrix}}$

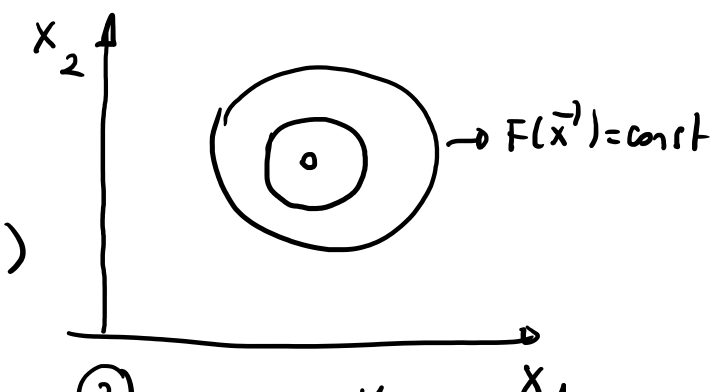
$$\text{Condition Number of } [Q] = \frac{\text{The largest eigenvalue of } [Q]}{\text{The smallest eigenvalue of } [Q]}$$

A large condition number will imply that the contours of the objective function $F(\vec{x})$ forms an elongated design space



① case with large condition number.

$F(\vec{x}) = F(x_1, x_2)$



② case with condition number ≈ 1.0

Example:

Example 4.2.1

The problem of determination of the maximum deflection and tip-rotation of a cantilever beam of length l shown in Figure (4.2.2) loaded at its tip is considered. Solution of this problem is formulated as a minimization of the total potential energy of the beam which is modelled using a single cubic beam finite element. For a two-noded beam element with two degrees of freedom at each node, the displacement field is assumed to be

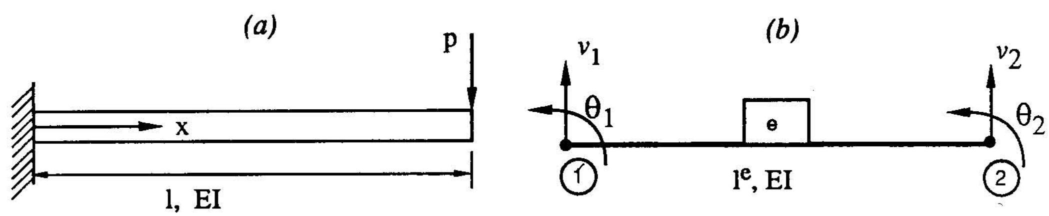


Figure 4.2.2 Tip loaded cantilever beam and its finite element model.

MISSOURI S&T

$$v(\xi) = \begin{bmatrix}(1-3\xi^2+2\xi^3) & l(\xi-2\xi^2+\xi^3) & (3\xi^2-2\xi^3) & l(-\xi^2+\xi^3)\end{bmatrix} \begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \end{Bmatrix} , \tag{4.2.15}$$

where $\xi = x/l$. The corresponding potential energy of the beam model is given by

$$\Pi = \frac{EI}{2l^3} \int_0^l \left(\frac{\mathrm{d}^2 v}{\mathrm{d}\xi^2}\right)^2 \mathrm{d}\xi + p v_2 \ . \tag{4.2.16}$$

Because of the cantilever end condition at $\xi = 0$, the first two degrees of freedom in Eq. (4.2.15) are zero. Therefore, substituting Eq. (4.2.15) into Eq. (4.2.16) we obtain

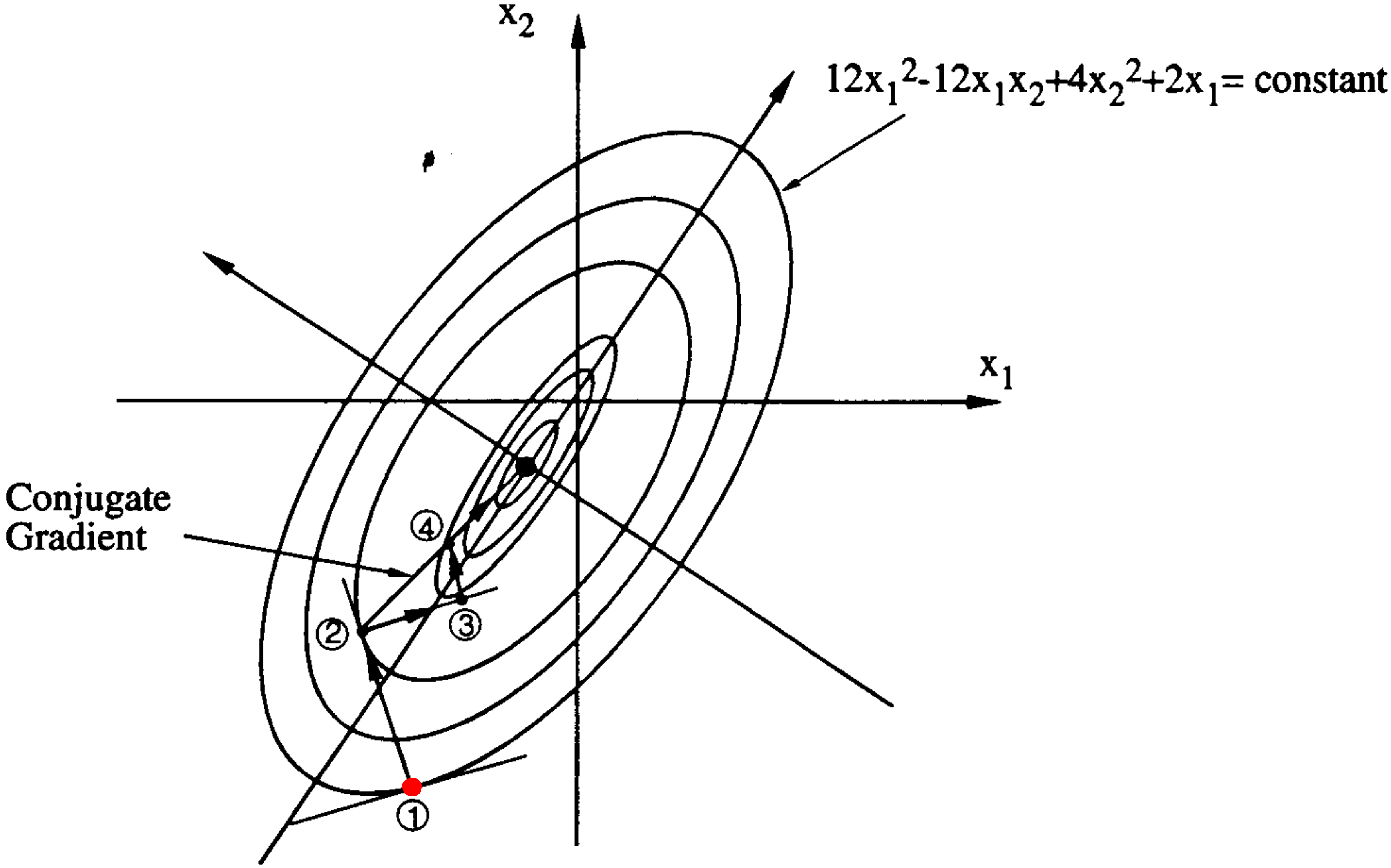
$$\Pi = \frac{EI}{2l^3}(12v_2^2 + 4\theta_2^2 l^2 - 12 v_2 \theta_2 l) + p v_2 \ . \tag{4.2.17}$$

Defining $f = 2\Pi l^3/EI$, $x_1 = v_2$, $x_2 = \theta_2 l$, and choosing $pl^3/EI = 1$, the problem of determining the tip deflection and rotation of the beam reduces to an unconstrained minimization of

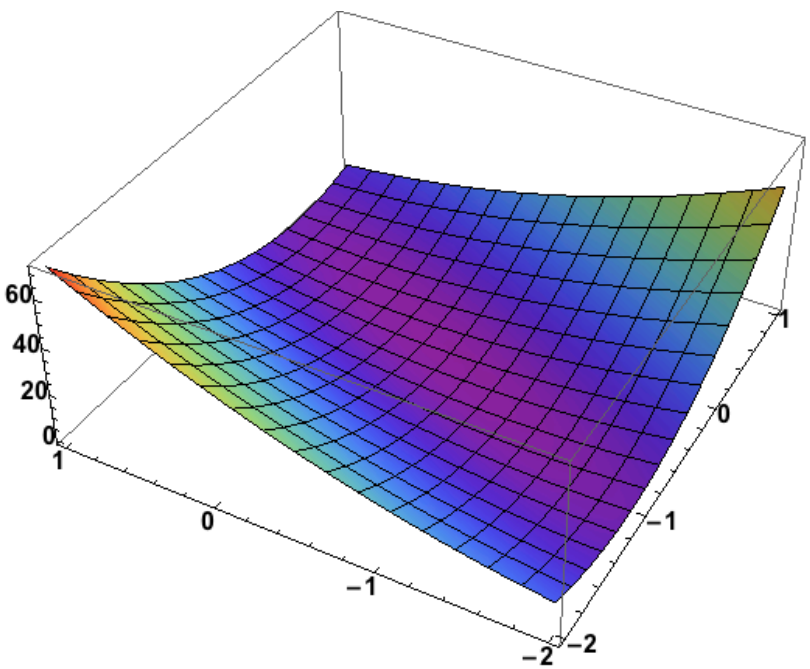
$$\boxed{f = 12x_1^2 + 4x_2^2 - 12x_1x_2 + 2x_1 \ .} \tag{4.2.18}$$

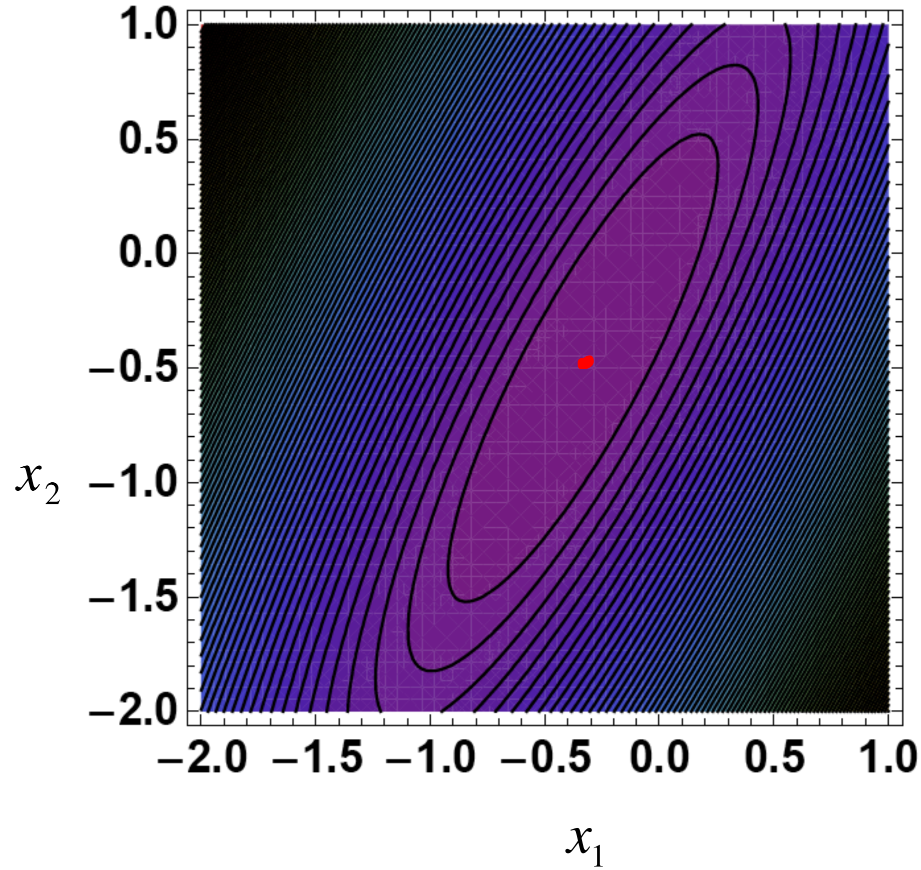
$f(\vec{x})$
$= f(x_1, x_2)$

Starting with an initial point of $\mathbf{x}_0^1 = (-1,\ -2)^T$ and $f(\mathbf{x}_0^1) = 2$ we will minimize f using Powell's conjugate directions method. The exact solution of this problem is at $\mathbf{x}^* = (-1/3,\ -1/2)^T$.

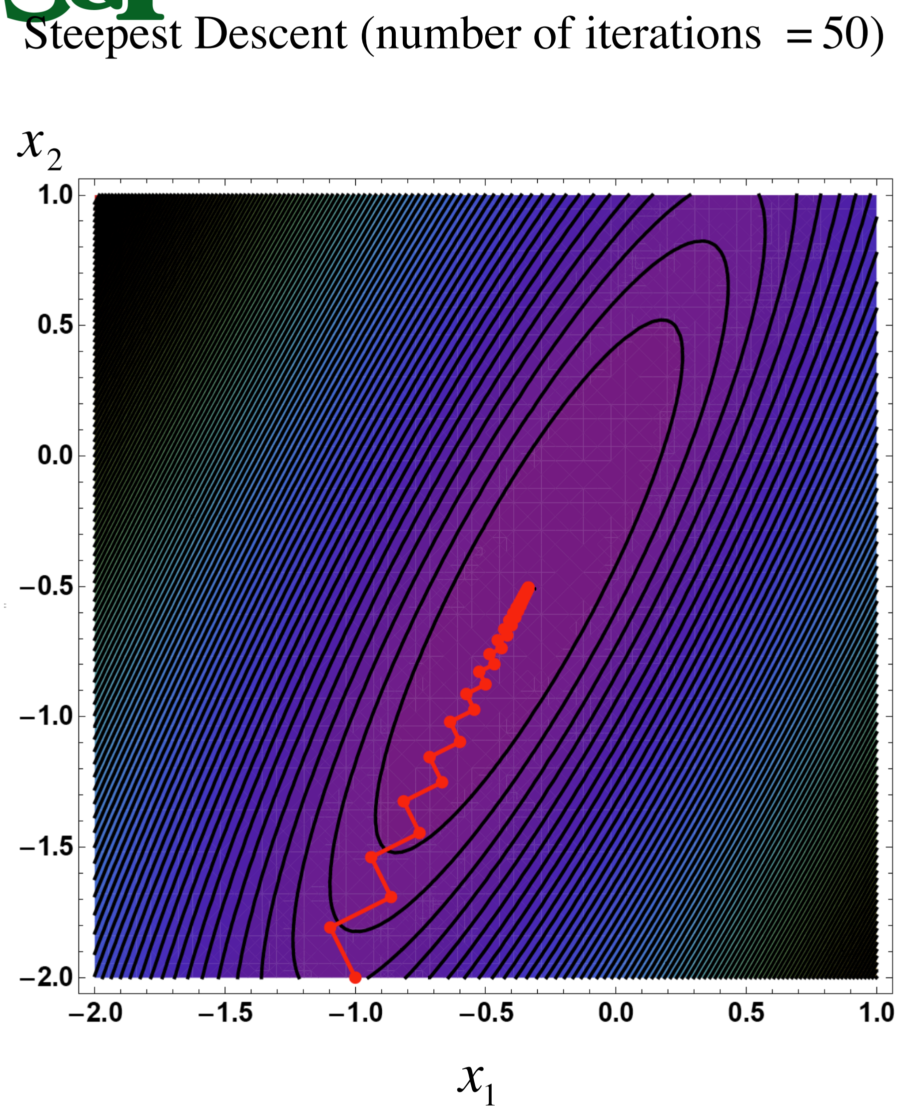

x_2
$12x_1^2-12x_1x_2+4x_2^2+2x_1=$ constant
x_1
Conjugate Gradient
4
2
3
1

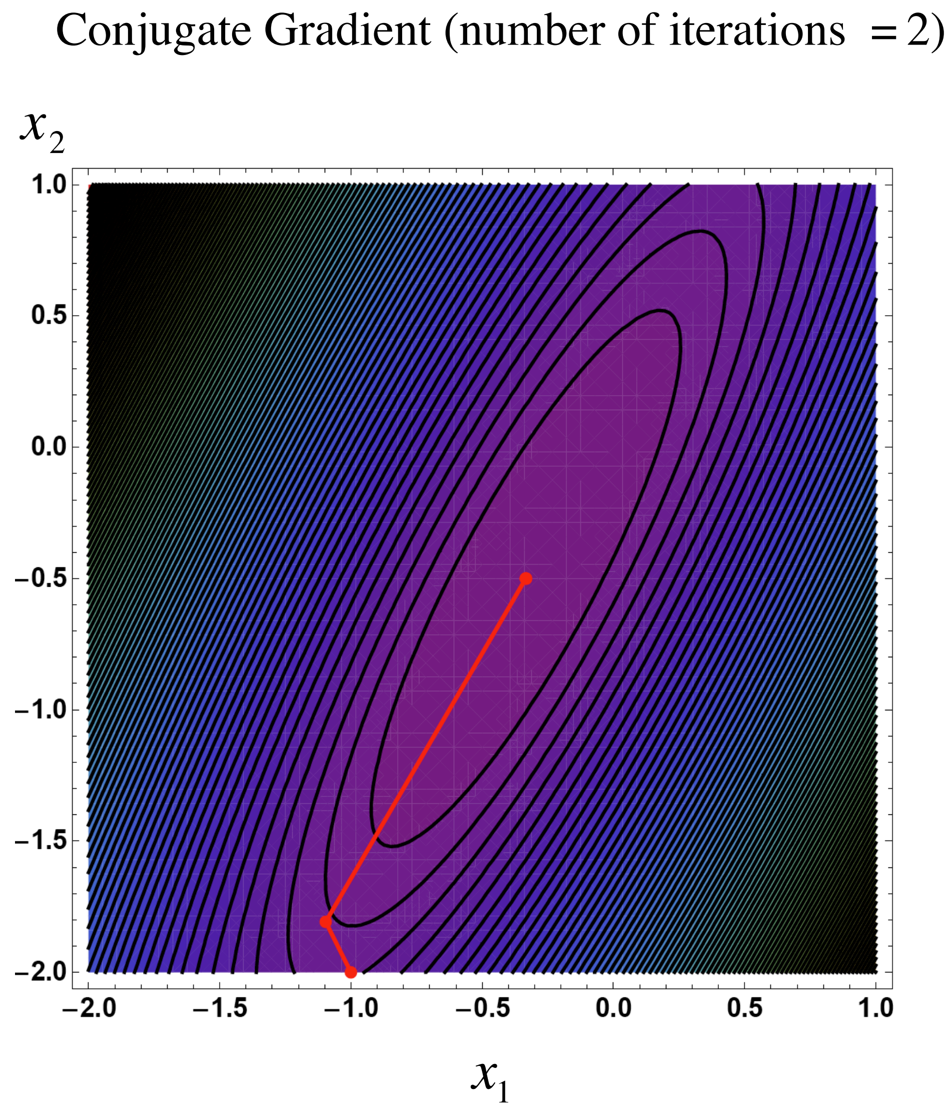
$$F(x_1,x_2) = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1$$





MISSOURI S&T


Steepest Descent (number of iterations = 50)
x_2
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
x_1


Conjugate Gradient (number of iterations = 2)
x_2
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
x_1

Fletcher-Reeves' Conjugate Gradient Method

1. At k = 0 start with the initial point $\vec{x}^0 = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}^0$

2. At k = 1, minimize $F(\vec{x})$ along the steepest descent direction

For $k=2$

$\vec{x}^2 = \vec{x}^1 + \alpha_2^* \vec{s}^2$

$\vec{s}^2 = -\nabla F(\vec{x}^1) + \beta_2 \vec{s}^1$

$$\vec{x}^1 = \vec{x}^0 + \alpha_1^* \vec{s}^1 \quad \text{where } \vec{s}^1 = -\nabla F(\vec{x}^0)$$

$\tilde{s}^1 = \frac{s^1}{\|s^1\|_2}$ can apply scaling to s^k

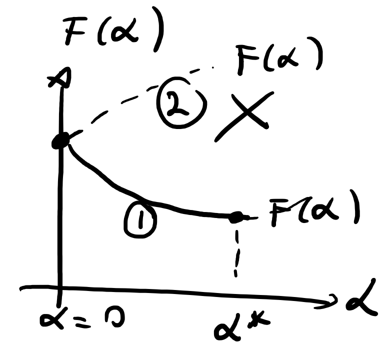
3. At k > 1, minimize $F(\vec{x}^k)$ with $\vec{x}^k = \vec{x}^{k-1} + \alpha_k^* \vec{s}^k$ where

$\beta_2 = \frac{\nabla F(\vec{x}^1)^T \cdot \nabla F(\vec{x}^1)}{\nabla F(\vec{x}^0)^T \cdot \nabla F(\vec{x}^0)}$

$$\boxed{s^k = -\nabla F(\vec{x}^{k-1}) + \beta_k \vec{s}^{k\text{-}1}} \quad \text{and} \quad \boxed{\tilde{s}^k = \frac{s^k}{\|s^k\|_2}}$$

$$\beta_k = \frac{\nabla F(\vec{x}^{k-1})^T \nabla F(\vec{x}^{k-1})}{\nabla F(\vec{x}^{k-2})^T \nabla F(\vec{x}^{k-2})}$$

4. The method should be restarted periodically due to possible error growth because of numerical imprecision



Calculate $\dfrac{dF(\alpha)}{d\alpha}$ for $\alpha = 0$ at each iteration.

It can be shown that $\dfrac{dF(\alpha)}{d\alpha}\Big|_{\alpha=0} = \nabla F(\vec{x}^{k-1})^T \vec{s}^{\,k}$

1 $\dfrac{dF}{d\alpha}\Big|_{\alpha=0} < 0$

If $\dfrac{dF(\alpha)}{d\alpha}\Big|_{\alpha=0} \geq 0 \rightarrow$ Indicator of numerical imprecision

2 $\dfrac{dF}{d\alpha}\Big|_{\alpha=0} \geq 0$

Then re-start the algorithm by setting $s^k = -\nabla F(\vec{x}^{k-1})$

5. Exit the iterations if the Convergence Criteria are satisfied:

$$\left\|\nabla F(\vec{x}^k)\right\|_2 < Tol_1 \text{ or } \left|F(\vec{x}^k) - F(\vec{x}^{k-1})\right| < Tol_2$$

Example for Conjugate-Gradient Method:

Example 4.2.3

We will show the effectiveness of this method on the cantilever beam problem for which we minimize

$$f = 12x_1^2 + 4x_2^2 - 12x_1x_2 + 2x_1 \, ,$$

starting with the initial design point $\mathbf{x}_0^T = (-1, -2)$. The initial move direction is calculated from the gradient

$$\nabla f(\mathbf{x}_0) = \begin{Bmatrix} 24x_1 - 12x_2 + 2 \\ 8x_2 - 12x_1 \end{Bmatrix}_{\mathbf{x}=\mathbf{x}_0} ,$$

$$\vec{\mathbf{s}}_{0\,1} = -\nabla f(\mathbf{x}_0) = \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} ,$$

$\vec{X}_k = \vec{X}_{k-1} + \alpha_k^* \vec{S}_k$

1st → $\vec{X}_1 = \vec{X}_0 + \alpha_1^* \vec{S}_1$ ✓

$\vec{X}_2 = \vec{X}_1 + \alpha_2^* \vec{S}_2$ ✓

and at the end of the first step we have

$$\boxed{\mathbf{x}_1 = \begin{Bmatrix} -1 \\ -2 \end{Bmatrix} + \alpha_1 \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} ,}$$

$$f(\alpha_1) = 12(-1 - 2\alpha_1)^2 + 4(-1 + 4\alpha_1)^2 - 12(-1 - 2\alpha_1)(-2 + 4\alpha_1) + 2(-1 - 2\alpha_1) \, .$$

The value of α_1 for which the function f is a minimum is obtained from the condition $df/d\alpha_1 = 0$, or $\boxed{\alpha_1^* = 0.048077.}$ The new design point and the gradient at that point are

$$\mathbf{x}_1 = \begin{Bmatrix} -1.0961 \\ -1.8077 \end{Bmatrix} , \qquad \text{and} \quad \nabla f(\mathbf{x}_1) = \begin{Bmatrix} -2.6154 \\ -1.3077 \end{Bmatrix} .$$

✓

$\vec{s_2} = -\nabla f(\vec{x_1}) + \beta_2 \cdot \vec{s_1}$

Next, let $\vec{s_2} = -\nabla f(\mathbf{x}_1) + \beta_2 \vec{s_1}$ with β_1 from Eq. (4.2.37), or

$$\beta_2 = \frac{(-2.6154)^2 + (-1.3077)^2}{(-2)^2 + (4)^2} \quad = \quad 0.4275\,,$$

$$\beta_2 = \frac{\nabla f(x_1)^T . \nabla f(x_1)}{\nabla f(x_0)^T \nabla f(x_0)}$$

The new move direction is

$$\boxed{\vec{s_2} = -\begin{Bmatrix} -2.6154 \\ -1.3077 \end{Bmatrix} + 0.4275 \begin{Bmatrix} -2 \\ 4 \end{Bmatrix} = \begin{Bmatrix} 1.76036 \\ 3.0178 \end{Bmatrix}\,,}$$

For a quadratic function of "n" variables if $(\vec{s_i})^T . [Q] . \vec{s_j} = 0 \ (i \neq j)$ then $\vec{s_i}$ & $\vec{s_j}$ are Q-conjugate directions.

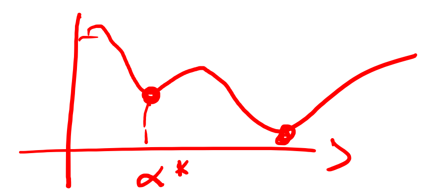
and

$$\mathbf{x}_2 = \begin{Bmatrix} -1.0961 \\ -1.8077 \end{Bmatrix} + \alpha_2 \begin{Bmatrix} 1.76036 \\ 3.0178 \end{Bmatrix}\,.$$

$\vec{x_2} = \vec{x_1} + \alpha^*_2 \, \vec{s_2}$

Again setting $\mathrm{d}f(\alpha_2)/\mathrm{d}(\alpha_2) = 0$ we obtain $\boxed{\alpha^*_2 = 0.4334,}$

$$\mathbf{x}_2 = \begin{Bmatrix} -0.3334 \\ -0.50 \end{Bmatrix}, \quad \text{and} \quad \nabla f(\mathbf{x}_2) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}\,.$$



Finally, since

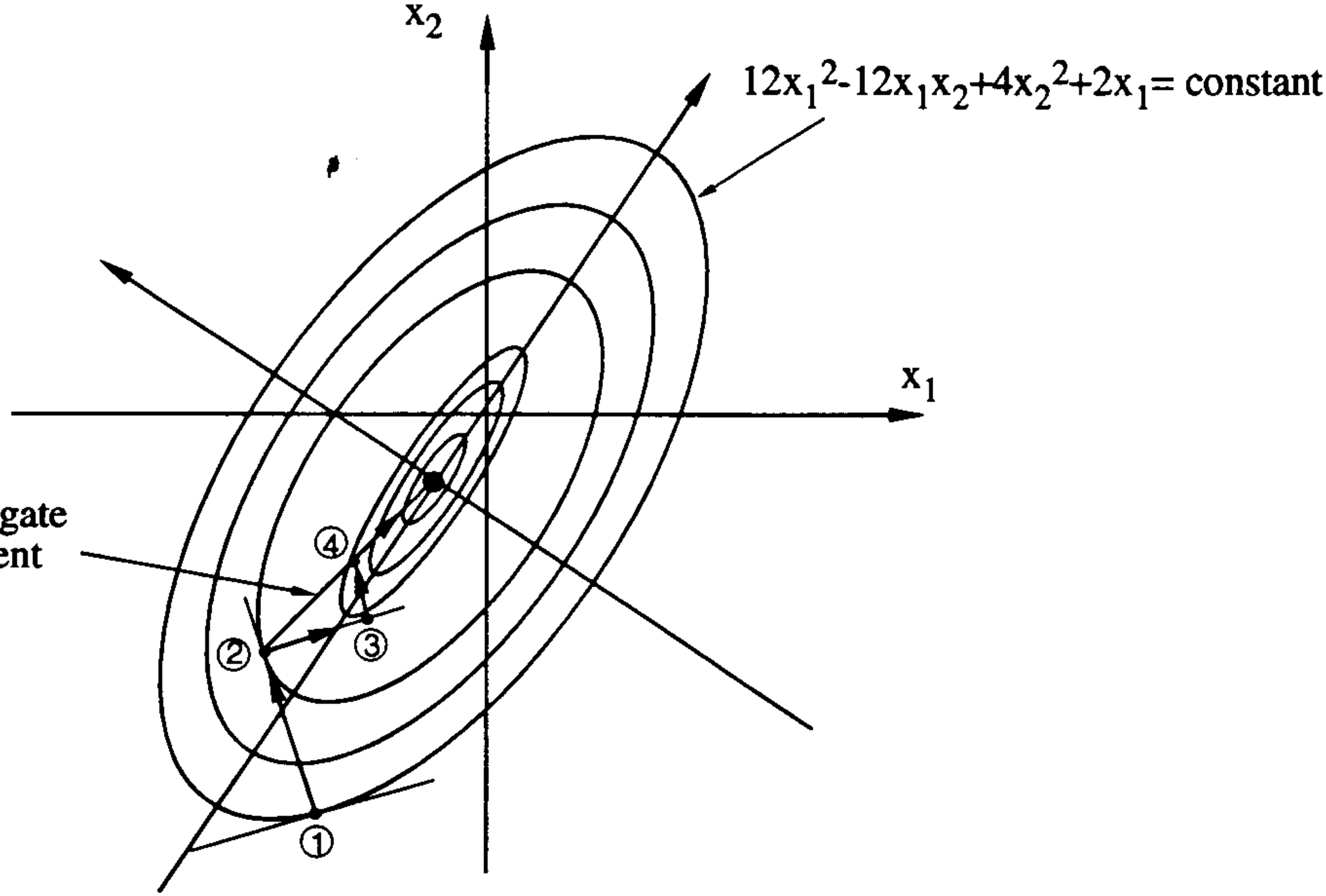
$$\underbrace{\begin{Bmatrix} -2 \\ 4 \end{Bmatrix}^T}_{\vec{s_1}} \overbrace{\begin{bmatrix} 24 & -12 \\ -12 & 8 \end{bmatrix}}^{[H]} \underbrace{\begin{Bmatrix} 1.76036 \\ 3.0178 \end{Bmatrix}}_{\vec{s_2}} \simeq 0\,.$$

we have verified the Q-conjugacy of the two directions $\mathbf{s}_0$ and $\mathbf{s}_1$. The progress of minimization using this method is illustrated in Figure (4.2.3). •••

$\vec{s_1}^T . [Q] . \vec{s_2} = 0$

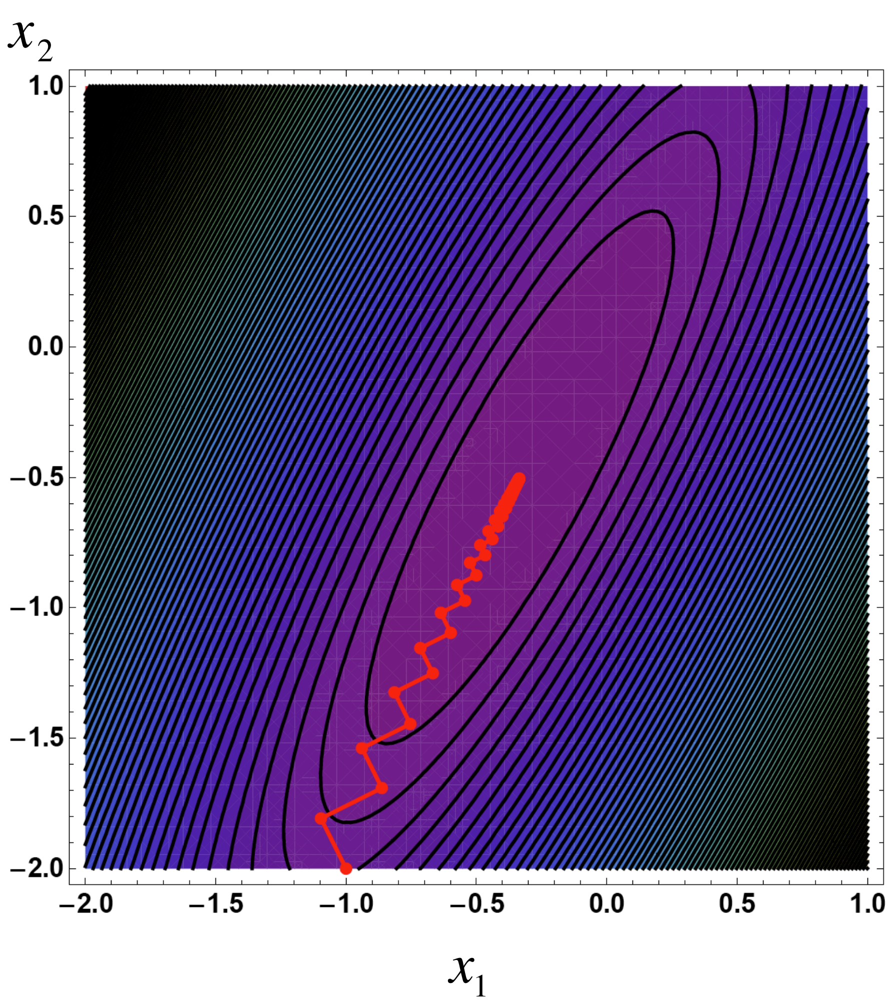
Hessian.

$\vec{s_1}$ and $\vec{s_2}$ are conjugate directions wrt $[Q]$

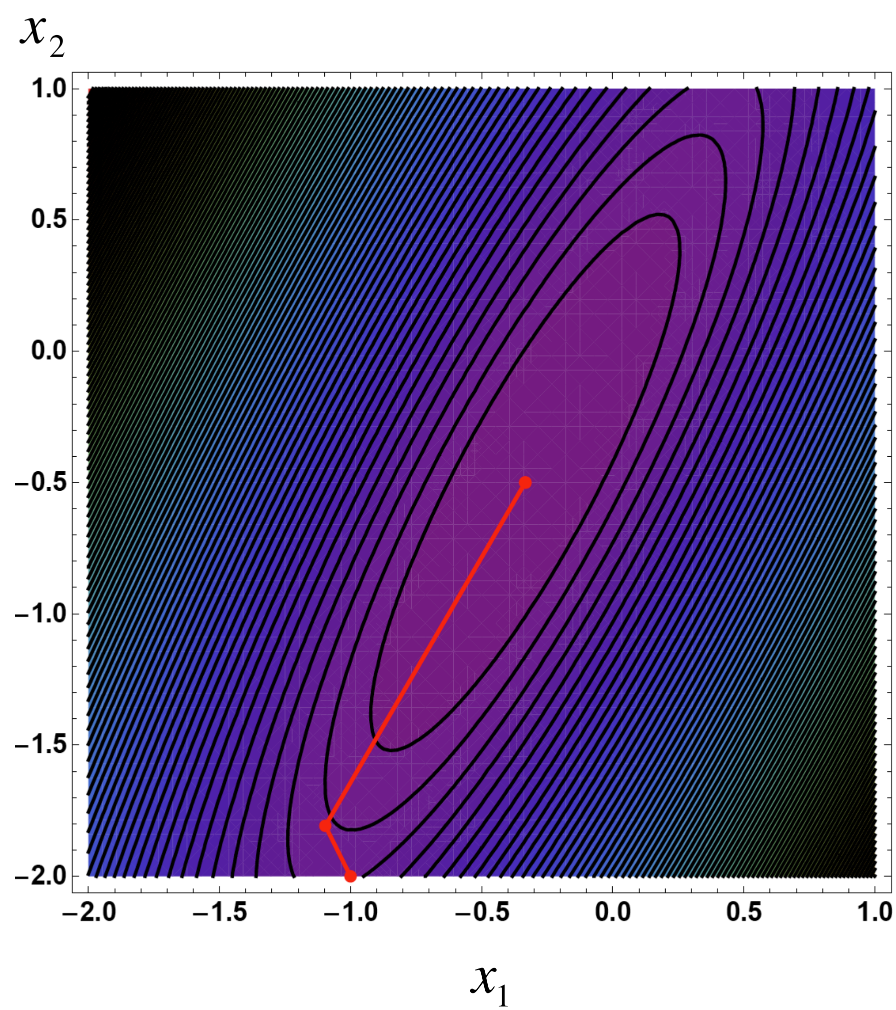


x2
x1
12x1 2-12x1x2+4x2 2+2x1= constant
Conjugate Gradient
1
2
3
4

Steepest Descent (number of iterations $=50$)



Conjugate Gradient (number of iterations $=2$)



MISSOURI S&T

Second-Order Method Basics:

$$\Delta\vec{x}_k = \vec{x}_k - \vec{x}_{k-1}$$

write the Taylor series expansion for $F(\vec{x}_k)$ at $\vec{x}_{k-1}$ point.

$$F(\vec{x}_k) = F(\vec{x}_{k-1}) + (\Delta\vec{x}_k)^T \cdot \nabla F(\vec{x}_{k-1}) + \frac{1}{2!}(\Delta\vec{x}_k)^T \cdot [H(\vec{x}_{k-1})]\,\Delta\vec{x}_k + \cdots$$

$$\boxed{\vec{x}_k - \vec{x}_{k-1} = \alpha^*_k \vec{s}_k}$$

Take the gradient of both sides

$$-H^{-1}(\vec{x}_{k-1}) \cdot \nabla F(\vec{x}_{k-1}) = \Delta\vec{x}_k$$

$$\nabla F(\vec{x}_k) = \nabla F(\vec{x}_{k-1}) + [H(\vec{x}_{k-1})] \cdot \Delta\vec{x}_k$$

Assume $\vec{x}_k \simeq x^*$ (optimum point) then $\nabla F(\vec{x}_k) = 0$

$$0 = \nabla F(\vec{x}_{k-1}) + [H(\vec{x}_{k-1})] \cdot \Delta\vec{x}_k \Rightarrow \boxed{-\nabla F(\vec{x}_{k-1}) = [H(\vec{x}_{k-1})] \cdot \Delta\vec{x}_k}$$

Example if $F(\vec{x}) = F(x_1, x_2)$ 2-D problem

$$-\begin{Bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \end{Bmatrix}_{\vec{x}_{k-1}} = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} \end{bmatrix}_{\vec{x}_{k-1}} \begin{Bmatrix} \Delta x_1 \\ \Delta x_2 \end{Bmatrix}_k$$

$$(\Delta x_1)_k = (x_1)_k - (x_1)_{k-1}$$

$$(\Delta x_2)_k = (x_2)_k - (x_2)_{k-1}$$

$$\vec{x}_k = \vec{x}_{k-1} + \Delta\vec{x}_k$$

Variable Metric (Quasi-Newton) Methods

Basic concept : Create an array which approximates the inverse of the Hessian Matrix as the optimization progresses

For these methods, search direction is defined as :

$$\vec{s}^{\,k} = -\hat{H}^{k}\nabla F(\vec{x}^{\,k-1})$$

where $\hat{H}^{k}$ approaches to the inverse of the Hessian Matrix for quadratic functions as the optimization progresses

Then using this search direction, minimize $F(\vec{x}^{\,k})$ with

$$\vec{x}^{\,k} = \vec{x}^{\,k-1} + \alpha_k^{*}\vec{s}^{\,k}$$

At k = 1, $\hat{H}^{1} = I$ (Identitiy Matrix) then $\vec{s}^{\,k} = -\nabla F(\vec{x}^{\,0})$

At the end of the k$^{\text{th}}$ iteration define $\hat{H}^{k+1} = \hat{H}^k + D^k$

where symmetric update matrix D^k is given by

$$D^k = \frac{\sigma + \theta\tau}{\sigma^2}\vec{p}\vec{p}^T + \frac{\theta - 1}{\tau}\hat{H}^k\vec{y}\left(\hat{H}^k\vec{y}\right)^T - \frac{\theta}{\sigma}\left[\hat{H}^k\vec{y}\vec{p}^T + \vec{p}\left(\hat{H}^k\vec{y}\right)^T\right]$$

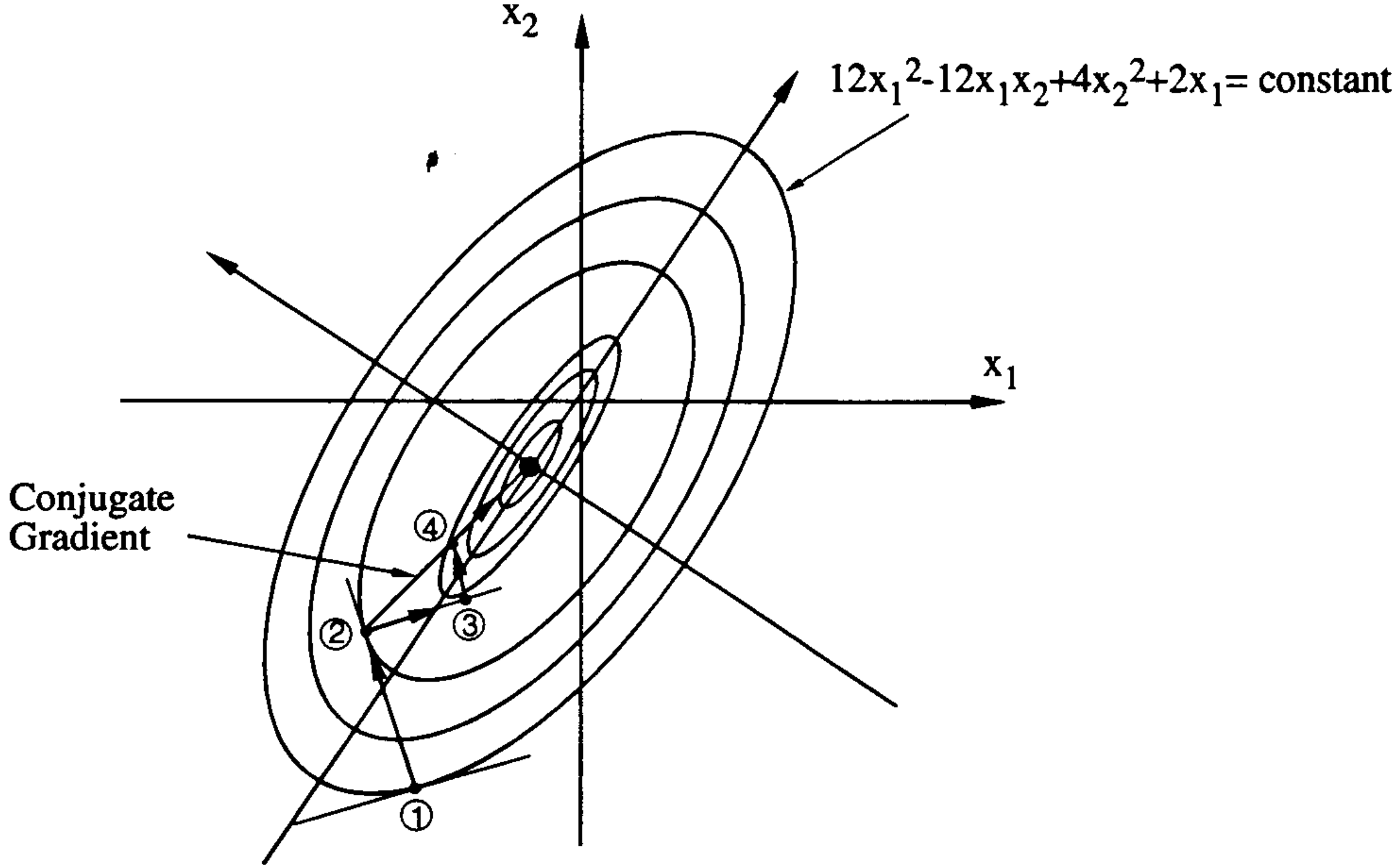
Where the change vectors are defined as:

$\vec{p} = \vec{x}^k - \vec{x}^{k-1}$ and $\vec{y} = \nabla F(\vec{x}^k) - \nabla F(\vec{x}^{k\text{-}1})$

and the scalers are defined as:

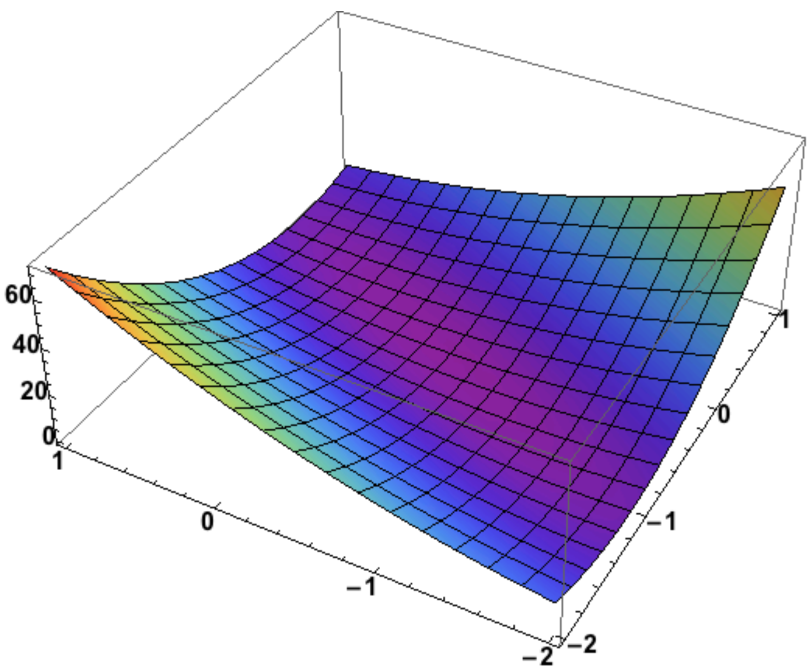
$\sigma = \vec{p}^T\vec{y}$ and $\tau = \vec{y}^T\hat{H}^k\vec{y}$

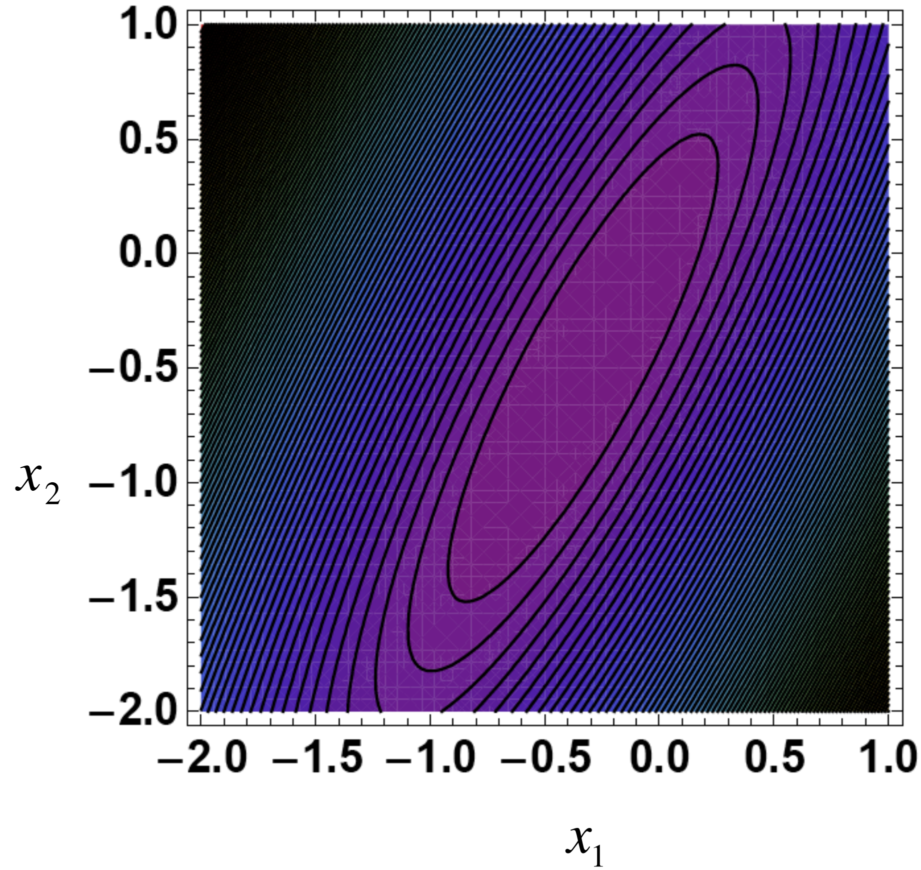
1. For $\theta = 0 \rightarrow$ DFP (Davidon - Fletcher - Powell) Method
2. For $\theta = 1 \rightarrow$ BFGS (Braydon - Fletcher - Goldfarb - Shanno) Method



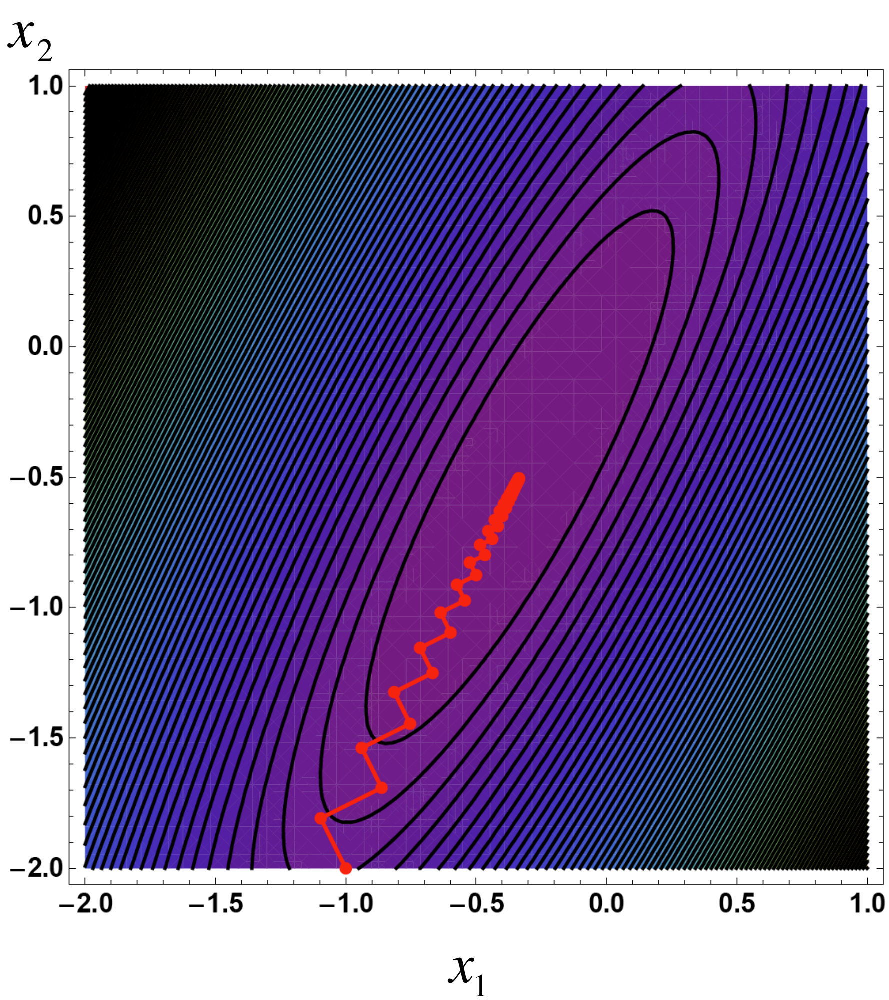
x_2
$12x_1^2-12x_1x_2+4x_2^2+2x_1=$ constant
x_1
Conjugate Gradient
1
2
3
4

$$F(x_1, x_2) = 12x_1^2 - 12x_1x_2 + 4x_2^2 + 2x_1$$

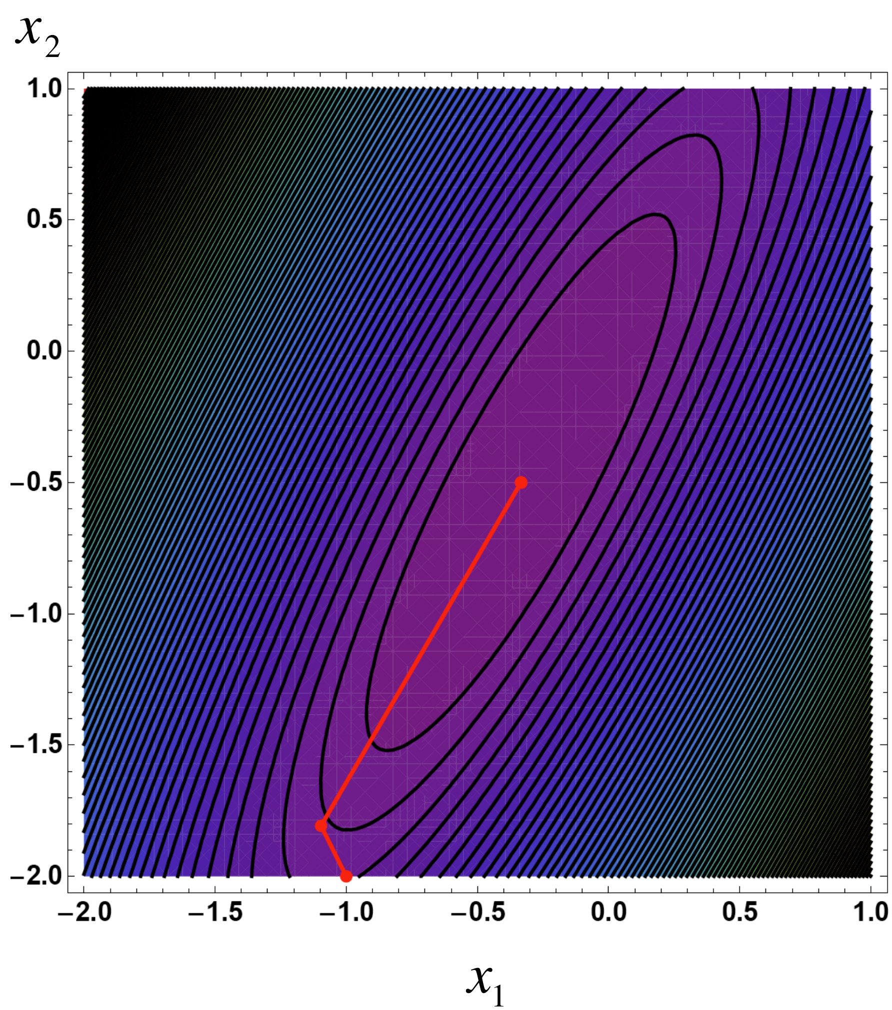




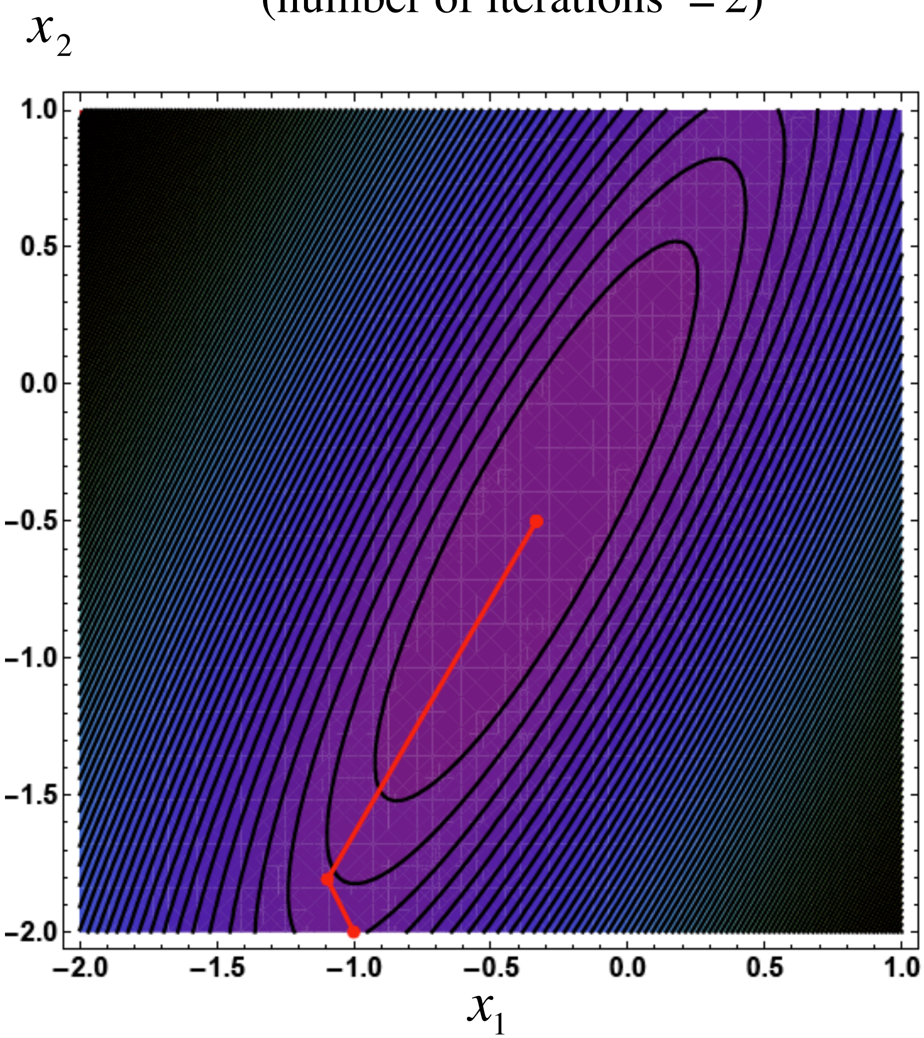
Steepest Descent (number of iterations $=50$)



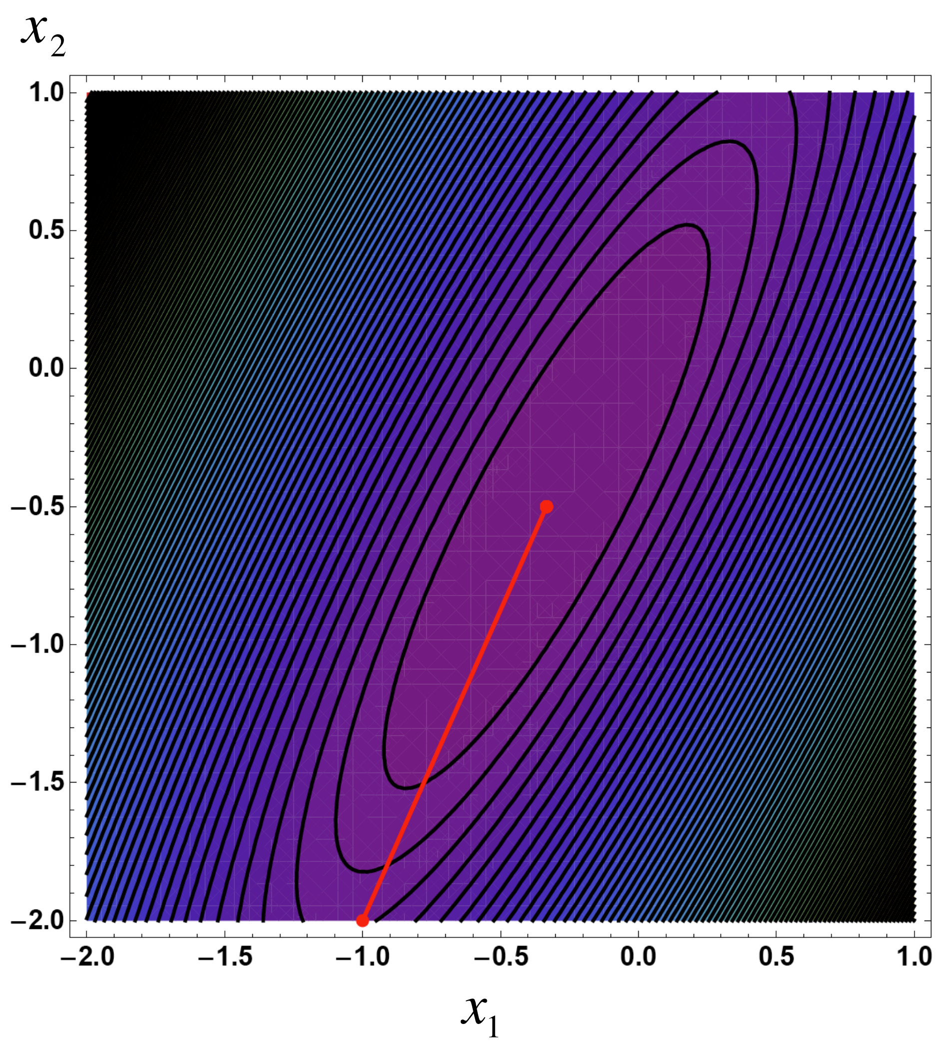
Conjugate Gradient (number of iterations $=2$)



Quasi - Newton (BFGS)
(number of iterations = 2)



Newton (number of iterations = 1)



Example:

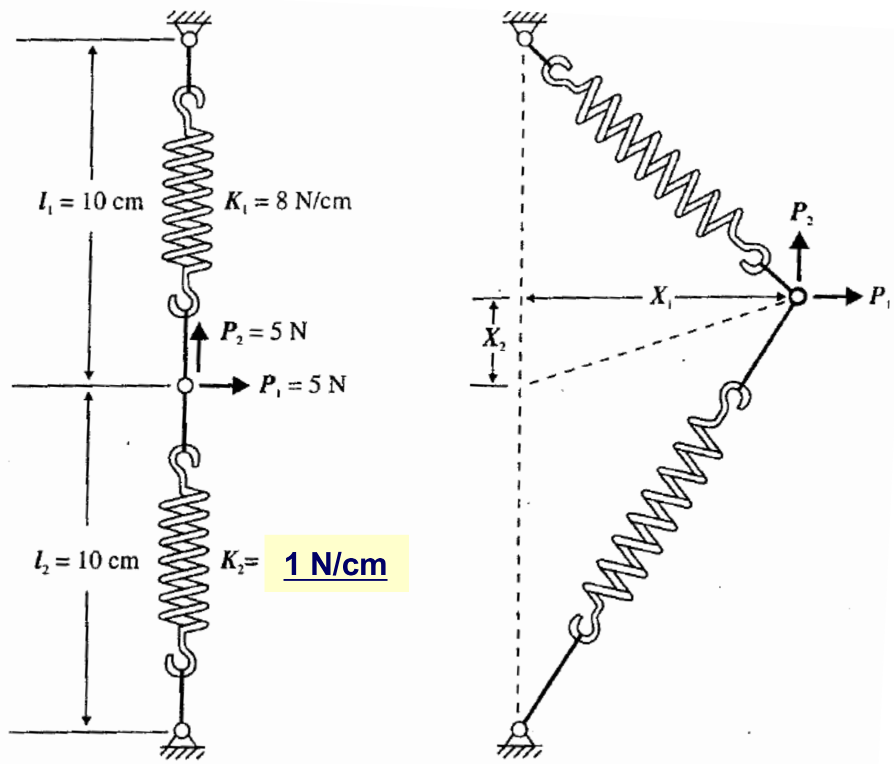
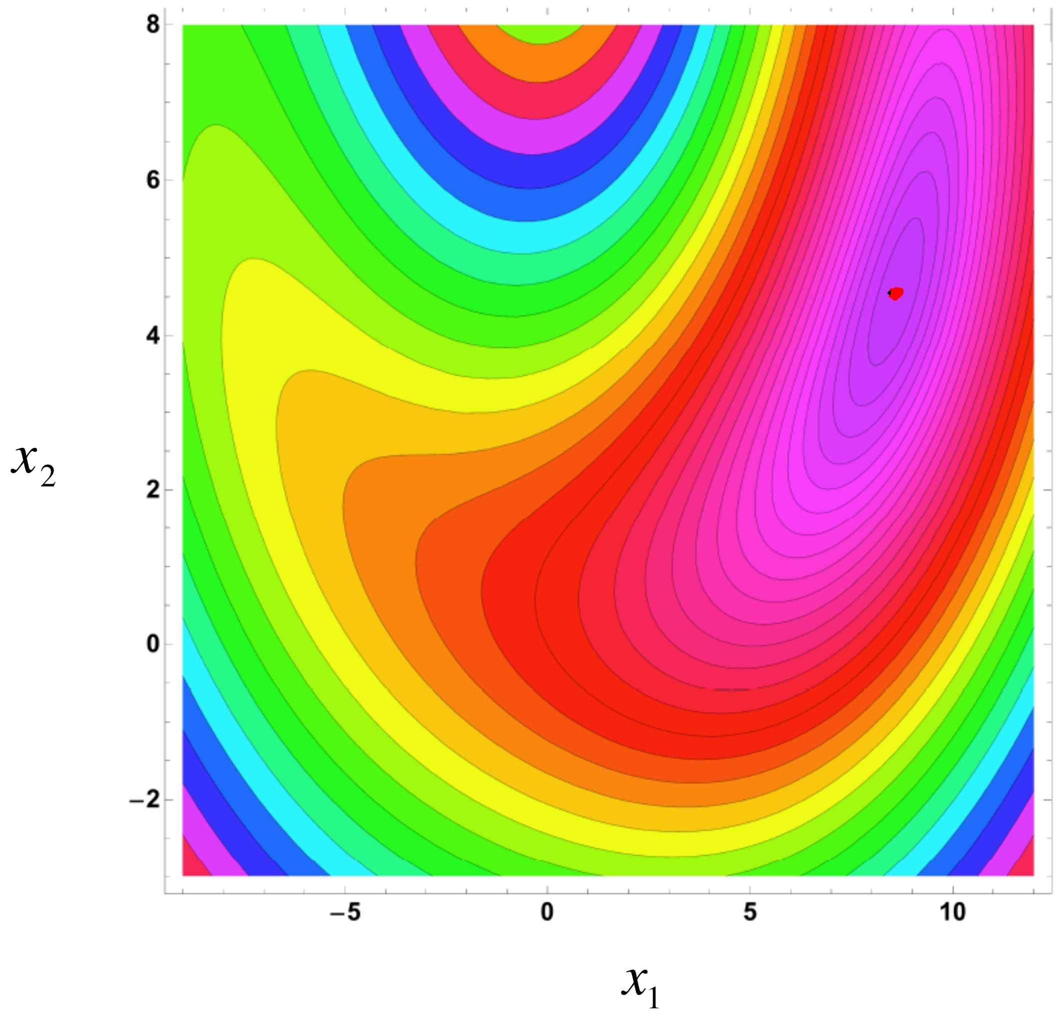


Figure 3-1 Equilibrium of a spring-force system.

$$F(x_1,x_2)=\frac{1}{2}K_1\left\{\sqrt{x_1^2+(l_1-x_2)^2}-l_1\right\}^2+\frac{1}{2}K_2\left\{\sqrt{x_1^2+(l_2+x_2)^2}-l_2\right\}^2-P_1x_1-P_2x_2$$

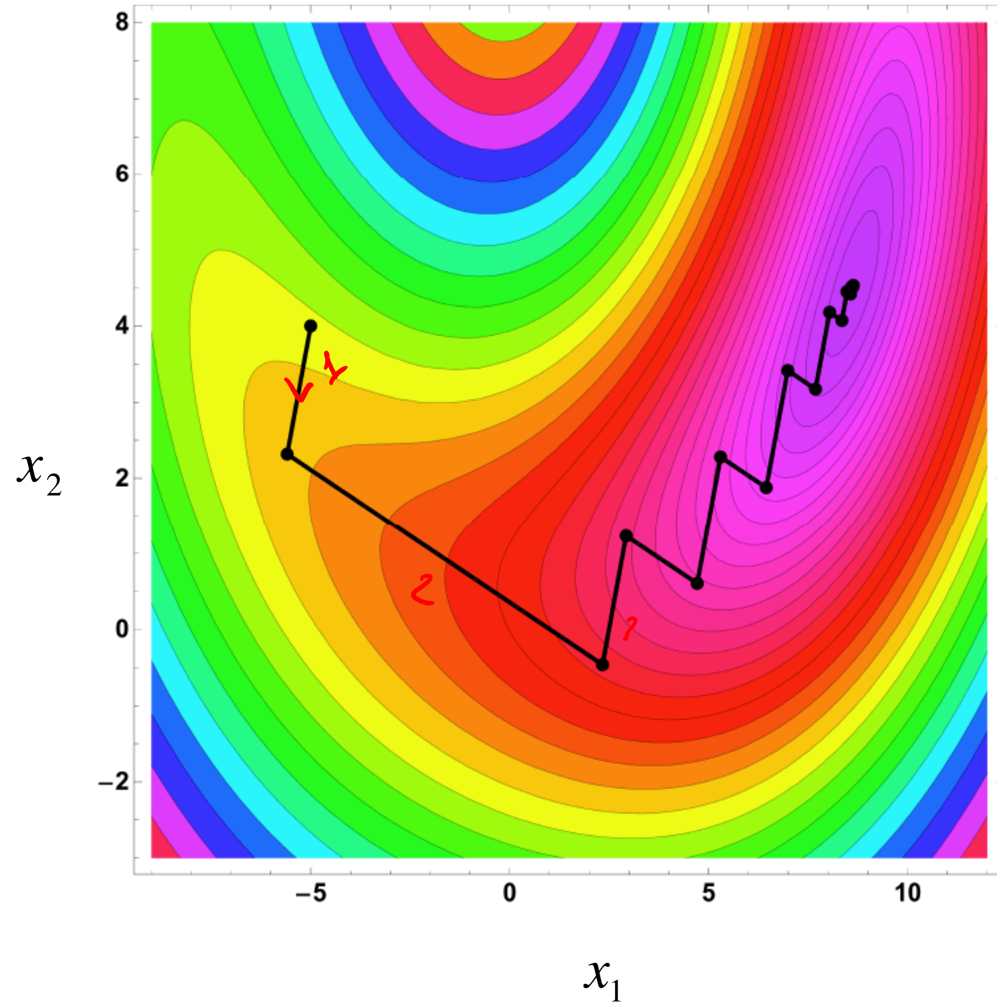


At the optimum

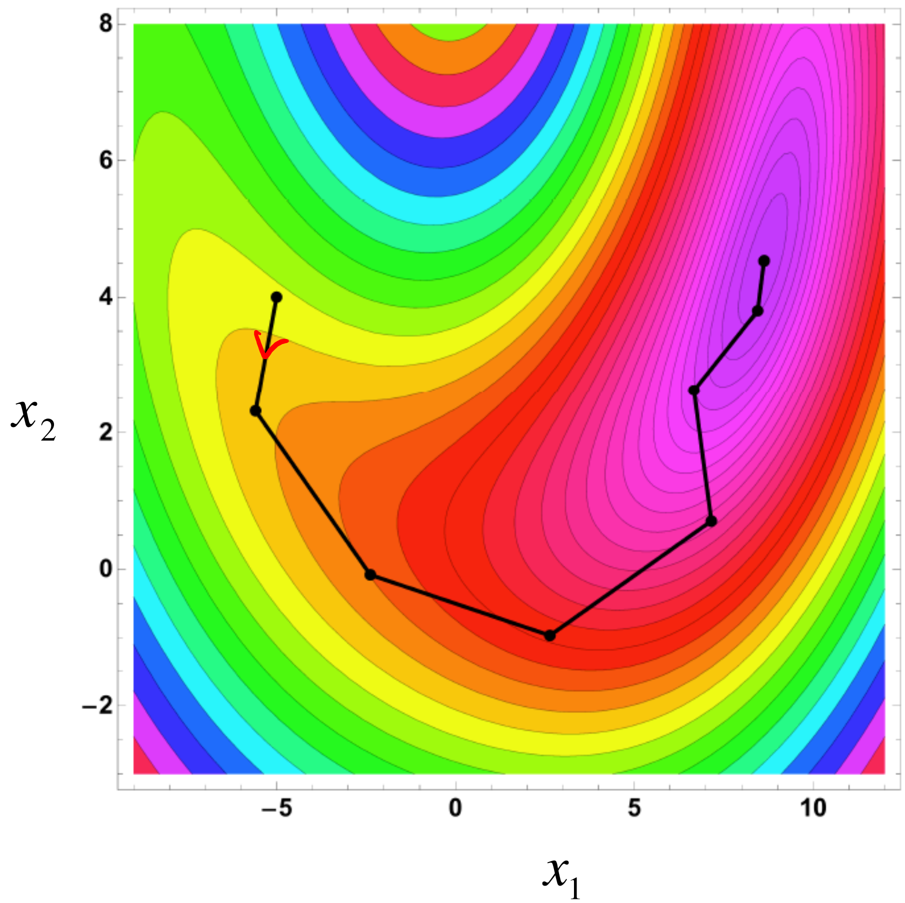
$x_1 = 8.631$ cm $x_2 = 4.533$ cm

$F(x_1,x_2) = -41.81$ N - cm

Steepest Descent (number of iterations $= 27$)



Quasi - Newton (BFGS)
(number of iterations $=12$)



Solution of Non-Linear set of equations as an unconstraint non-linear optimization problem (with optimization algorithms)

solve $\vec{f} = 0$ where $\vec{f} = \begin{Bmatrix} f_1(x_1, x_2, \dots x_n) \\ f_2(x_1, x_2, \dots x_n) \\ \vdots \\ f_n(x_1, x_2, \dots x_n) \end{Bmatrix}$ ✓

Find $\underline{\underline{\vec{x}^*}} \Rightarrow \vec{f}(\vec{x}^*) = 0$

↳ solution vector

$$\boxed{F(\vec{x}) = f_1^2 + f_2^2 + \dots + f_n^2} \quad \text{(Objective function)}$$

Minimize $F(\vec{x})$ to obtain $\vec{x} = \vec{x}^*$ (minimum point is the solution for $\vec{f} = 0$)

At $\vec{x}^k$ $\nabla F(\vec{x}^*) = 0$ (Necessary condition)

$$\nabla F(\vec{x}) = \left.\begin{Bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{Bmatrix}\right|_{\vec{x}^*} \qquad \left.\frac{\partial F}{\partial x_i}\right|_{\vec{x}^*} = 2 \overbrace{f_1}^{=0} \left.\frac{\partial f_1}{\partial x_i}\right|_{\vec{x}^*} + 2 \overbrace{f_2}^{=0} \left.\frac{\partial f_2}{\partial x_i}\right|_{\vec{x}^*} + \dots + 2 \overbrace{f_n}^{=0} \left.\frac{\partial f_n}{\partial x_i}\right|_{\vec{x}^*}$$

$i = 1, \dots n$