Optimum Design

Fall 2014

Myung-Il Roh

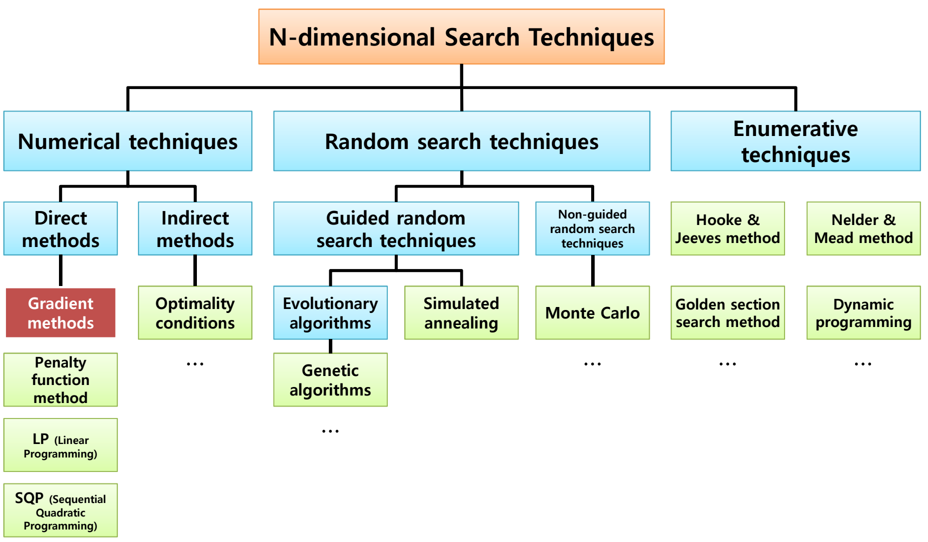
Department of Naval Architecture and Ocean Engineering
Seoul National University

Contents

Ch. 2 Unconstrained Optimization Method: Gradient Method

2.1 Steepest Descent Method
2.2 Conjugate Gradient Method
2.3 Newton's Method
2.4 Davidon-Fletcher-Powell (DFP) Method
2.5 Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method

Classes of Search Techniques

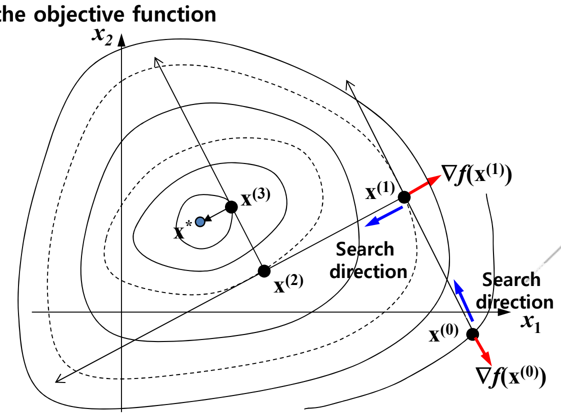


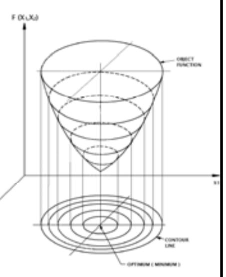
1. Steepest Descent Method (1/6)

- **Step 1: The search direction (d) is taken as the negative of the gradient of the objective function (f) at current iteration since the objective function decreases mostly rapidly toward that direction.**
  The direction of gradient vector of f, $\nabla f(\mathbf{x})$, is the direction of maximum increase of f at $\mathbf{x}$.

  Search direction $$\mathbf{d} = -\mathbf{c} \equiv -\nabla f(\mathbf{x})$$

- **Step 2: Iterate successively to find the optimum design point.**
  Ex) Minimize the objective function





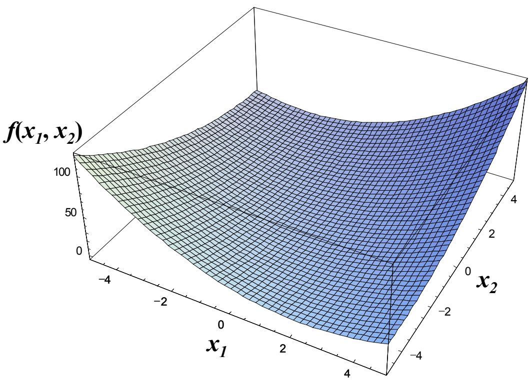
1. Steepest Descent Method (2/6): Example

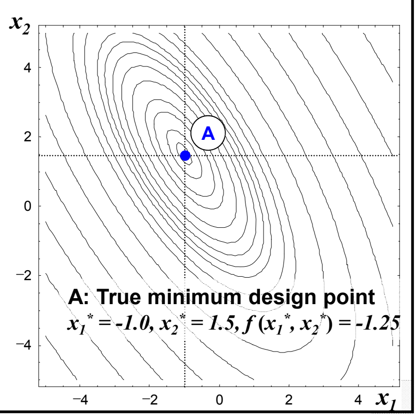
☑ **By using the steepest descent method, find the minimum design point for the following function of 2-variables.**

Given: Starting design point $\mathbf{x}^{(0)} = (0, 0)$, convergence tolerance $\varepsilon = 0.001$

Find: $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ ➡ Optimization problem with two unknown variables





1. Steepest Descent Method (3/6): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ **Starting design point** $\mathbf{x}^{(0)}$ **= (0, 0)**

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

■ **1st Iteration: Find $\mathbf{x}^{(1)}$**

$$\nabla f(\mathbf{x}^{(0)}) = \nabla f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha^{(0)} \nabla f(\mathbf{x}^{(0)})$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \alpha \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -\alpha \\ \alpha \end{pmatrix}$$

Replacing $\alpha^{(0)}$ to α for convenience

Substituting $\mathbf{x}^{(1)} = (-\alpha, \alpha)$ into the objective function

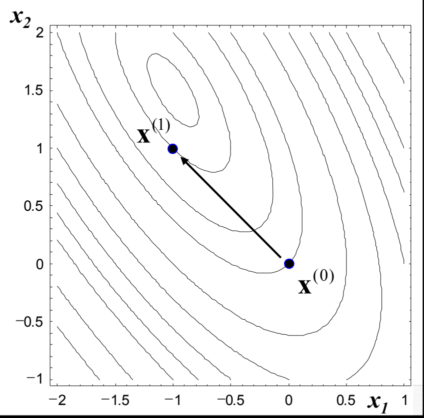
$$f(\mathbf{x}^{(1)}) = -\alpha - \alpha + 2\alpha^2 - 2\alpha^2 + \alpha^2$$

$$= \alpha^2 - 2\alpha$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\boxed{\frac{df(\mathbf{x}^{(1)})}{d\alpha}} = 2\alpha - 2 = 0 \rightarrow \alpha = 1.0 \qquad \therefore \mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

How can we differentiate f with respect to α?



1. Steepest Descent Method (4/6): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ **Starting design point** $\mathbf{x}^{(0)}$ **= (0, 0)**

■ **2nd Iteration: Find $\mathbf{x}^{(2)}$**

$$\nabla f(\mathbf{x}^{(1)}) = \nabla f\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - \alpha^{(1)} \nabla f(\mathbf{x}^{(1)})$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \alpha \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 + \alpha \\ 1 + \alpha \end{pmatrix}$$

Replacing $\alpha^{(1)}$ to α for convenience

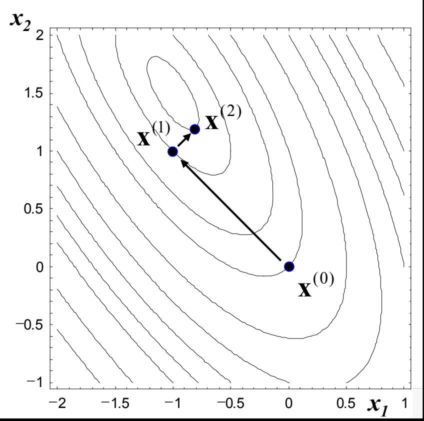
Substituting $\mathbf{x}^{(2)} = (-1 + \alpha, 1 + \alpha)$ into the objective function

$$f(\mathbf{x}^{(2)}) = 5\alpha^2 - 2\alpha - 1$$

To minimize $f(\mathbf{x}^{(2)})$,

$$\frac{df(\mathbf{x}^{(2)})}{d\alpha} = 10\alpha - 2 = 0 \rightarrow \alpha = 0.2$$

$$\therefore \mathbf{x}^{(2)} = \begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix}$$



1. Steepest Descent Method (5/6): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ **Starting design point** $\mathbf{x}^{(0)} = (0, 0)$

- **3rd Iteration: Find $\mathbf{x}^{(3)}$**

$$\nabla f(\mathbf{x}^{(2)}) = \nabla f\begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix} = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix} = \begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix}$$

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} - \alpha^{(2)}\nabla f(\mathbf{x}^{(2)})$$

$$= \begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix} - \alpha\begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix} = \begin{pmatrix} -0.8-0.2\alpha \\ 1.2+0.2\alpha \end{pmatrix}$$

Replacing $\alpha^{(1)}$ to α for convenience

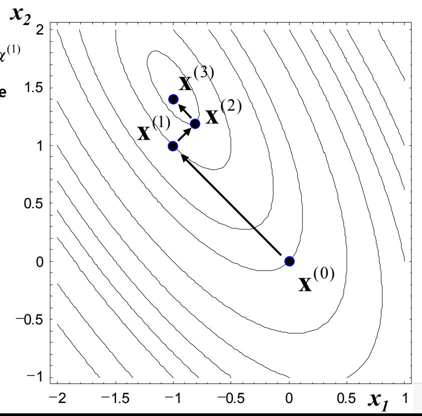
Substituting $\mathbf{x}^{(3)} = (-0.8-0.2\alpha, 1.2+0.2\alpha)$ **into the objective function**

$$f(\mathbf{x}^{(3)}) = 0.04\alpha^2 - 0.08\alpha - 1.2$$

To minimize $f(\mathbf{x}^{(3)})$**,**

$$\frac{df(\mathbf{x}^{(3)})}{d\alpha} = 0.08\alpha - 0.08 = 0 \;\rightarrow\; \alpha = 1.0$$

$$\therefore \mathbf{x}^{(3)} = \begin{pmatrix} -1 \\ 1.4 \end{pmatrix}$$



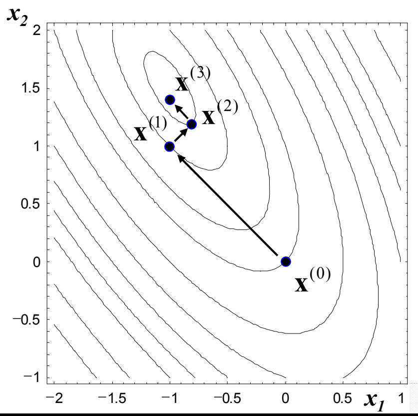
1. Steepest Descent Method (6/6): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ **Starting design point** $\mathbf{x}^{(0)} = (0, 0)$

- **4th Iteration: Find the minimum design point.**

  To obtain the minimum design point, we have to iterate.

  If $\left|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right| \le \varepsilon$**, then stop the iterative process because $\mathbf{x}^{(k+1)}$ can be assumed as the minimum design point.**



[Reference] Differentiation of Function of x with Respect to the Another Variable

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ Starting design point $\mathbf{x}^{(0)}$ = (0, 0)

$f(x_1, x_2) = f(\mathbf{x})$: f is the function of $\mathbf{x}$.

$\mathbf{x}^{(1)} = (-\alpha, \alpha)$: $\mathbf{x}^{(1)}$ is the function of α

➡ Substituting $\mathbf{x}^{(1)}$ into f, f is, then, a function of α and can be differentiated with respect to α.

Substituting $\mathbf{x}^{(1)} = (-\alpha, \alpha)$ into the objective function

$$f(\mathbf{x}^{(1)}) = -\alpha - \alpha + 2\alpha^2 - 2\alpha^2 + \alpha^2$$
$$= \alpha^2 - 2\alpha$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\frac{df(\mathbf{x}^{(1)})}{d\alpha} = 2\alpha - 2 = 0 \rightarrow \alpha = 1.0 \qquad \therefore \mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

How can we differentiate f with respect to α?

In the similar way, we can consider the followings:

To minimize $f(\mathbf{x}^* + \Delta\mathbf{x})$,

The second-order Taylor series expansion of $f(\mathbf{x}^* + \Delta\mathbf{x})$

$$f(\mathbf{x}^* + \Delta\mathbf{x}) = f(\mathbf{x}^*) + \mathbf{c}^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x}$$

$$f(\mathbf{x}^* + \Delta\mathbf{x}) - f(\mathbf{x}^*) = \mathbf{c}^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x}$$

In the above equation, we assume that x* is a constant and Δx is a variable.

$$f(\Delta\mathbf{x}) = \mathbf{c}^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x}$$

To minimize f,

$$\frac{df(\Delta\mathbf{x})}{d\Delta\mathbf{x}} = \mathbf{c} + \mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x} = 0$$

$$\Rightarrow \mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x} = -\mathbf{c}$$

$$\Rightarrow \boxed{\Delta\mathbf{x} = -\mathbf{H}(\mathbf{x}^*)^{-1}\mathbf{c}}$$ 'Newton's method'

2. Conjugate Gradient Method (1/5)

☑ This method requires only a simple modification to the steepest descent method and dramatically improves the convergence rate of the optimization process.

☑ The current steepest descent direction is modified by adding a scaled direction used in the previous iteration.

- Step 1: Estimate a starting design point as $\mathbf{x}^{(0)}$. Set the iteration counter $k = 0$. Also, specify a tolerance ε for stopping criterion. Calculate

$$\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} \equiv -\nabla f(\mathbf{x}^{(0)})$$

Check stopping criterion. If $\|\mathbf{c}^{(0)}\| < \varepsilon$, then stop. Otherwise, go to Step 4.

It is noted that Step 1 of the conjugate gradient method and steepest descent method is the same.

2. Conjugate Gradient Method (2/5)

- **Step 2: Compute the gradient of the objective function as $\mathbf{c}^{(k)} = \nabla f(\mathbf{x}^{(k)})$. If $\|\mathbf{c}^{(k)}\| < \varepsilon$, then stop; otherwise continue.**

- **Step 3: Calculate the new search direction as**

$$\boxed{\mathbf{d}^{(k)} = -\mathbf{c}^{(k)} + \beta_k \mathbf{d}^{(k-1)}} \rightarrow \text{Previous search direction}$$

$$\beta_k = \left(\|\mathbf{c}^{(k)}\| / \|\mathbf{c}^{(k-1)}\| \right)^2$$

The current search direction is calculated by adding a scaled direction used in the previous iteration.

- **Step 4: Compute a step size $\alpha = \alpha_k$ to minimize $f(\mathbf{x}^{(k)}+\alpha\mathbf{d}^{(k)})$.**

- **Step 5: Change the design point as follows, then set $k = k+1$ and go to Step 2.**

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$$

2. Conjugate Gradient Method (3/5): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ **Starting design point $\mathbf{x}^{(0)}$ = (0, 0)**

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

☑ **1st Iteration: Find $\mathbf{x}^{(1)}$**

$$\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = -\nabla f\left(\mathbf{x}^{(0)}\right) = -\nabla f\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$= -\begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix} = -\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0 \mathbf{d}^{(0)}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -\alpha \\ \alpha \end{pmatrix}$$

Replacing α_0 to α for convenience

Substituting $\mathbf{x}^{(1)} = (-\alpha, \alpha)$ into the objective function

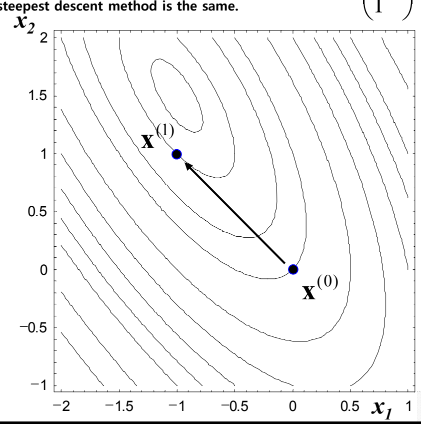
$$f(\mathbf{x}^{(1)}) = -\alpha - \alpha + 2\alpha^2 - 2\alpha^2 + \alpha^2$$
$$= \alpha^2 - 2\alpha$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\frac{df(\mathbf{x}^{(1)})}{d\alpha} = 2\alpha - 2 = 0 \rightarrow \alpha = 1.0$$

Note: Step 1 of the conjugate gradient method and steepest descent method is the same.

$$\therefore \mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



2. Conjugate Gradient Method (4/5): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1 x_2 + x_2^2$

- **2nd Iteration: Find $\mathbf{x}^{(2)}$**

Compute the gradient of the objective function as

$$\mathbf{c}^{(1)} = \nabla f\left(\mathbf{x}^{(1)}\right)$$

$$= \nabla f \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

Calculate the new search direction as

$$\mathbf{d}^{(1)} = -\mathbf{c}^{(1)} + \beta_1 \mathbf{d}^{(0)} = -\mathbf{c}^{(1)} + \frac{\left\|\nabla f\left(\mathbf{x}^{(1)}\right)\right\|^2}{\left\|\nabla f\left(\mathbf{x}^{(0)}\right)\right\|^2} \mathbf{d}^{(0)}$$

$$= -\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \frac{2}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{d}^{(0)} = -\nabla f\left(\mathbf{x}^{(0)}\right) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{d}^{(k)} = -\mathbf{c}^{(k)} + \beta_k \mathbf{d}^{(k-1)}$$

$$\beta_k = \left(\left\|\mathbf{c}^{(k)}\right\| / \left\|\mathbf{c}^{(k-1)}\right\|\right)^2$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$$

2. Conjugate Gradient Method (5/5): Example

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1 x_2 + x_2^2$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \alpha_1 \mathbf{d}^{(1)}$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 + 2\alpha \end{pmatrix}$$ **Replacing α_1 to α for convenience**

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$$

$$\mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \mathbf{d}^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

Substituting $\mathbf{x}^{(2)} = (-1, 1 + 2\alpha)$ **into the objective function**

$$f(\mathbf{x}^{(2)}) = 4\alpha^2 - 2\alpha - 1$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\frac{df(\mathbf{x}^{(2)})}{d\alpha} = 8\alpha - 2 = 0 \;\rightarrow\; \alpha = 0.25$$

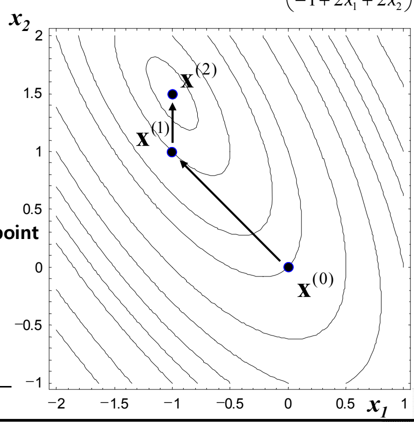
$$\therefore \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

→Minimum design point

Check stopping criterion.

$$\mathbf{c}^{(2)} = \nabla f\left(\mathbf{x}^{(2)}\right) = \nabla f \begin{pmatrix} -1 \\ 1.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

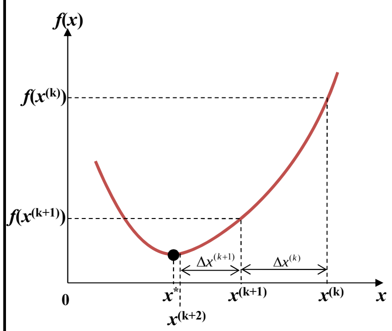
$$\left\|\mathbf{c}^{(2)}\right\| = 0 < \varepsilon$$ **→Stop!**



3. Newton's Method (1/9)

Given: $f(x)$

Find: x^* which minimizes $f(x)$



Assume that $f(x)$ has minimum at $x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$.

Consider the quadratic approximation of the function $f(x)$ at $x = x^{(k)}$ using the second-order Taylor expansion.

$$f(x^{(k)} + \Delta x^{(k)}) = f(x^{(k)}) + \frac{df(x^{(k)})}{dx}\Delta x^{(k)} + \frac{1}{2}\frac{d^2 f(x^{(k)})}{dx^2}(\Delta x^{(k)})^2 + \cancel{O((\Delta x^{(k)})^3)}$$

In this equation, $x^{(k)}$ is a constant and $\Delta x^{(k)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta x^{(k)}$.

$$f(x^{(k)} + \Delta x^{(k)}) = f(x^{(k)}) + \frac{df(x^{(k)})}{dx}\Delta x^{(k)} + \frac{1}{2}\frac{d^2 f(x^{(k)})}{dx^2}(\Delta x^{(k)})^2$$

Differentiate this equation with respect to $\Delta x^{(k)}$.

$$\frac{df(x^{(k)} + \Delta x^{(k)})}{d\Delta x^{(k)}} = \frac{df(x^{(k)})}{dx} + \frac{d^2 f(x^{(k)})}{dx^2}\Delta x^{(k)} = 0 \rightarrow$$ The necessary condition for minimization of this function

Calculate the small change $\Delta x^{(k)}$ in design.

$$\Delta x^{(k)} = \left(-\frac{df(x^{(k)})}{dx}\right) / \left(\frac{d^2 f(x^{(k)})}{dx^2}\right)$$

Is $|\Delta x^{(k)}| < \varepsilon$?

NO: $k = k + 1$

YES: Set $x^* = x^{(k+1)}$ and stop the iteration.

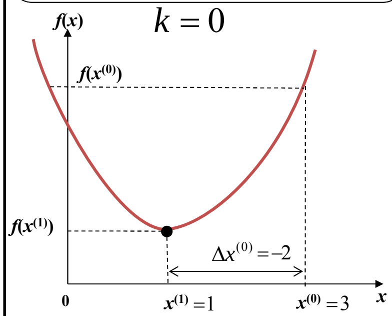
3. Newton's Method (2/9): Example

Given: $f(x) = x^2 - 2x + 2$

Starting design point $x^{(0)} = 3$

Find: x^* which minimizes $f(x)$

$k = 0$



Assume that $f(x)$ has minimum at $x^{(1)} = x^{(0)} + \Delta x^{(0)}$.

Consider the quadratic approximation of the function $f(x)$ at $x = x^{(0)}$ using the second-order Taylor expansion.

$$f(x^{(0)} + \Delta x^{(0)}) = f(x^{(0)}) + \frac{df(x^{(0)})}{dx}\Delta x^{(0)} + \frac{1}{2}\frac{d^2 f(x^{(0)})}{dx^2}(\Delta x^{(0)})^2$$

In this equation, $x^{(0)}$ is a constant and $\Delta x^{(0)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta x^{(0)}$.

$$f(x^{(0)} + \Delta x^{(0)}) = f(x^{(0)}) + \frac{df(x^{(0)})}{dx}\Delta x^{(0)} + \frac{1}{2}\frac{d^2 f(x^{(0)})}{dx^2}(\Delta x^{(0)})^2$$

Differentiate this equation with respect to $\Delta x^{(0)}$.

$$\frac{df(x^{(0)} + \Delta x^{(0)})}{d\Delta x^{(0)}} = \frac{df(x^{(0)})}{dx} + \frac{d^2 f(x^{(0)})}{dx^2}\Delta x^{(0)} = 0 \rightarrow$$ The necessary condition for minimization of this function

Calculate the small change $\Delta x^{(0)}$ in design.

$$\Delta x^{(0)} = \left(-\frac{df(x^{(0)})}{dx}\right) / \left(\frac{d^2 f(x^{(0)})}{dx^2}\right) = (-2x+2)_{x=3} / (2)_{x=3} = -2$$

Is $|\Delta x^{(0)}| < \varepsilon$?

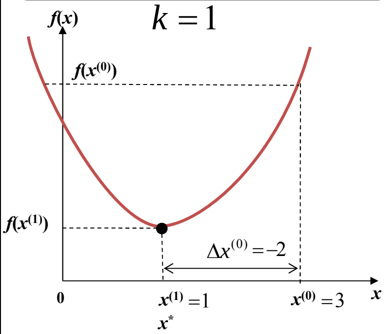
NO: $k = k + 1 = 0 + 1 = 1$

3. Newton's Method (3/9): Example

Given: $f(x) = x^2 - 2x + 2$

Starting design point $x^{(0)} = 3$

Find: x^* which minimizes $f(x)$



Is it possible to find the x* which minimizes a cubic function at once?

Assume that $f(x)$ has minimum at $x^{(1)} = x^{(0)} + \Delta x^{(0)}$.

Consider the quadratic approximation of the function $f(x)$ at $x = x^{(1)}$ using the second-order Taylor expansion.

$$f(x^{(1)} + \Delta x^{(1)}) = f(x^{(1)}) + \frac{df(x^{(1)})}{dx}\Delta x^{(1)} + \frac{1}{2}\frac{d^2 f(x^{(1)})}{dx^2}(\Delta x^{(1)})^2$$

In this equation, $x^{(1)}$ is a constant and $\Delta x^{(1)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta x^{(1)}$.

$$f(x^{(1)} + \Delta x^{(1)}) = f(x^{(1)}) + \frac{df(x^{(1)})}{dx}\Delta x^{(1)} + \frac{1}{2}\frac{d^2 f(x^{(1)})}{dx^2}(\Delta x^{(1)})^2$$

Differentiate this equation with respect to $\Delta x^{(1)}$.

$$\frac{df(x^{(1)} + \Delta x^{(1)})}{d\Delta x^{(1)}} = \frac{df(x^{(1)})}{dx} + \frac{d^2 f(x^{(1)})}{dx^2}\Delta x^{(1)} = 0 \rightarrow$$

The necessary condition for minimization of this function

Calculate the small change $\Delta x^{(1)}$ in design.

$$\Delta x^{(1)} = \left(-\frac{df(x^{(1)})}{dx}\right) / \left(\frac{d^2 f(x^{(1)})}{dx^2}\right)$$

$$= (-2x + 2)_{x=1} / (2)_{x=1} = 0$$

Is $|\Delta x^{(1)}| < \varepsilon$?

YES

Set $x^* = x^{(1)}$ and stop the iteration.

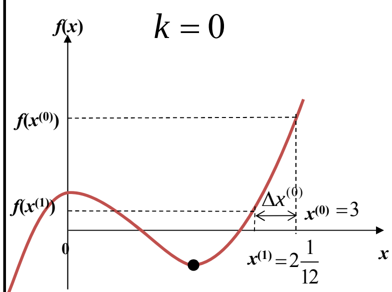
3. Newton's Method (4/9): Example

Is it possible to find the x* which minimizes a **cubic function** at once?

Given: $f(x) = x^3 - 3x^2 + 2x$

Starting design point $x^{(0)} = 3$

Find: x^* which minimizes $f(x)$



Assume that $f(x)$ has minimum at $x^{(1)} = x^{(0)} + \Delta x^{(0)}$.

Consider the quadratic approximation of the function $f(x)$ at $x = x^{(0)}$ using the second-order Taylor expansion.

$$f(x^{(0)} + \Delta x^{(0)}) = f(x^{(0)}) + \frac{df(x^{(0)})}{dx}\Delta x^{(0)} + \frac{1}{2}\frac{d^2 f(x^{(0)})}{dx^2}(\Delta x^{(0)})^2 + \cancel{O((\Delta x^{(0)})^3)}$$

In this equation, $x^{(0)}$ is a constant and $\Delta x^{(0)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta x^{(0)}$.

$$f(x^{(0)} + \Delta x^{(0)}) = f(x^{(0)}) + \frac{df(x^{(0)})}{dx}\Delta x^{(0)} + \frac{1}{2}\frac{d^2 f(x^{(0)})}{dx^2}(\Delta x^{(0)})^2$$

Differentiate this equation with respect to $\Delta x^{(0)}$.

$$\frac{df(x^{(0)} + \Delta x^{(0)})}{d\Delta x^{(0)}} = \frac{df(x^{(0)})}{dx} + \frac{d^2 f(x^{(0)})}{dx^2}\Delta x^{(0)} = 0 \rightarrow$$

The necessary condition for minimization of this function

Calculate the small change $\Delta x^{(0)}$ in design.

$$\Delta x^{(0)} = \left(-\frac{df(x^{(0)})}{dx}\right) / \left(\frac{d^2 f(x^{(0)})}{dx^2}\right)$$

$$= (-3x^2 + 6x - 2)_{x=3} / (6x - 6)_{x=3} = -\frac{11}{12}$$

Is $|\Delta x^{(0)}| < \varepsilon$?

NO

$k = k + 1$

$= 0 + 1 = 1$

3. Newton's Method (5/9): Example

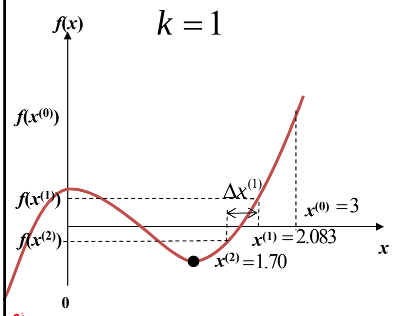
Is it possible to find the x* which minimizes a cubic function at once?

Given: $f(x) = x^3 - 3x^2 + 2x$

Starting design point $x^{(0)} = 3$

Find: x^* which minimizes $f(x)$

$k = 1$



Assume that $f(x)$ has minimum at $x^{(2)} = x^{(1)} + \Delta x^{(1)}$.

Consider the quadratic approximation of the function $f(x)$ at $x = x^{(1)}$ using the second-order Taylor expansion.

$$f(x^{(1)} + \Delta x^{(1)}) = f(x^{(1)}) + \frac{df(x^{(1)})}{dx}\Delta x^{(1)} + \frac{1}{2}\frac{d^2 f(x^{(1)})}{dx^2}(\Delta x^{(1)})^2 + \cancel{O((\Delta x^{(1)})^3)}$$

In this equation, $x^{(1)}$ is a constant and $\Delta x^{(1)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta x^{(1)}$.

$$f(x^{(1)} + \Delta x^{(1)}) = f(x^{(1)}) + \frac{df(x^{(1)})}{dx}\Delta x^{(1)} + \frac{1}{2}\frac{d^2 f(x^{(1)})}{dx^2}(\Delta x^{(1)})^2$$

Differentiate this equation with respect to $\Delta x^{(1)}$.

$$\frac{df(x^{(1)} + \Delta x^{(1)})}{d\Delta x^{(1)}} = \frac{df(x^{(1)})}{dx} + \frac{d^2 f(x^{(1)})}{dx^2}\Delta x^{(1)} = 0 \rightarrow$$ The necessary condition for minimization of this function

Calculate the small change $\Delta x^{(1)}$ in design.

$$\Delta x^{(1)} = \left(-\frac{df(x^{(1)})}{dx}\right) \Big/ \left(\frac{d^2 f(x^{(1)})}{dx^2}\right)$$
$$= \left(-3x^2 + 6x - 2\right)_{x=\frac{25}{12}} / \left(6x - 6\right)_{x=\frac{25}{12}} = -0.388$$

Is $|\Delta x^{(1)}| < \varepsilon$?

NO

$k = k + 1$
$= 1 + 1 = 2$

Why is it not possible to find the x* which minimizes a cubic function at once?

⇨ Since the second-order Taylor expansion is just an approximation for $f(x)$ at the point $x^{(0)}$ or $x^{(1)}$, $x^{(1)}$ or $x^{(2)}$ will probably not be the precise minimum design point of $f(x)$.

3. Newton's Method (6/9): Example of Function of Two Variables

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, Starting design point $\mathbf{x}^{(0)} = (0, 0)$

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}, \quad \mathbf{H}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

- **1st Iteration: Find $\mathbf{x}^{(1)}$**

Assume that $f(x)$ has minimum at $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)}$.

Consider the quadratic approximation of the function $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^{(0)}$ using the second-order Taylor expansion.

$$f(\mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)}) = f(\mathbf{x}^{(0)}) + \nabla f(\mathbf{x}^{(0)})^T \Delta\mathbf{x}^{(0)} + \frac{1}{2}(\Delta\mathbf{x}^{(0)})^T \mathbf{H}(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)}$$

How?

In this equation, $\mathbf{x}^{(0)}$ is a constant and $\Delta\mathbf{x}^{(0)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta\mathbf{x}^{(0)}$.

$$f(\mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)}) = f(\mathbf{x}^{(0)}) + \nabla f(\mathbf{x}^{(0)})^T \Delta\mathbf{x}^{(0)} + \frac{1}{2}(\Delta\mathbf{x}^{(0)})^T \mathbf{H}(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)}$$

3. Newton's Method (7/9): Example of Function of Two Variables

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, Starting design point $\mathbf{x}^{(0)}$ = (0, 0)

■ 1^{st} Iteration: Find $\mathbf{x}^{(1)}$

$$f(\mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)}) = f(\mathbf{x}^{(0)}) + \nabla f(\mathbf{x}^{(0)})^T \Delta\mathbf{x}^{(0)} + \frac{1}{2}(\Delta\mathbf{x}^{(0)})^T \mathbf{H}(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)}$$

How?

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

Differentiate this equation with respect to $\Delta\mathbf{x}^{(0)}$.

$$\frac{\partial f(\mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)})}{\partial(\Delta\mathbf{x}^{(0)})} = \nabla f(\mathbf{x}^{(0)}) + \mathbf{H}(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)} = 0 \longrightarrow$$

The necessary condition for minimization of function $f(x_1, x_2)$

Calculate the small change $\Delta\mathbf{x}^{(0)}$ in design.

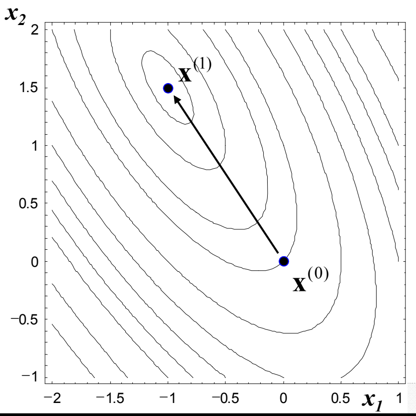
$$\mathbf{H}(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)} = -\nabla f(\mathbf{x}^{(0)})$$

$$\Delta\mathbf{x}^{(k)} = -[\mathbf{H}(\mathbf{x}^{(k)})]^{-1}\nabla f(\mathbf{x}^{(k)})$$

$$\downarrow \left(-\nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} -1 \\ 1 \end{pmatrix},\ \mathbf{H}\left(\mathbf{x}^{(0)}\right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \right)$$

$$\begin{pmatrix} \Delta\mathbf{x}_1^{(0)} \\ \Delta\mathbf{x}_2^{(0)} \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} \Delta\mathbf{x}_1^{(0)} \\ \Delta\mathbf{x}_2^{(0)} \end{pmatrix} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

$$\therefore x^{(1)} = x^{(0)} + \Delta\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1.5 \end{pmatrix} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$



3. Newton's Method (8/9): Example of Function of Two Variables

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, Starting design point $\mathbf{x}^{(0)}$ = (0, 0)

■ 2^{nd} Iteration: Find $\mathbf{x}^{(2)}$

$$x^{(1)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

In the same way as 1^{st} Iteration,

Assume that $f(x)$ has minimum at $\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \Delta\mathbf{x}^{(1)}$.

Consider the quadratic approximation of the function $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^{(1)}$ using the second-order Taylor expansion.

$$f(\mathbf{x}^{(1)} + \Delta\mathbf{x}^{(1)}) = f(\mathbf{x}^{(1)}) + \nabla f(\mathbf{x}^{(1)})^T \Delta\mathbf{x}^{(1)} + \frac{1}{2}(\Delta\mathbf{x}^{(1)})^T \mathbf{H}(\mathbf{x}^{(1)})\Delta\mathbf{x}^{(1)}$$

In this equation, $\mathbf{x}^{(1)}$ is a constant and $\Delta\mathbf{x}^{(1)}$ is a variable. So, the following equation is a quadratic function in terms of $\Delta\mathbf{x}^{(1)}$.

$$f(\mathbf{x}^{(1)} + \Delta\mathbf{x}^{(1)}) = f(\mathbf{x}^{(1)}) + \nabla f(\mathbf{x}^{(1)})^T \Delta\mathbf{x}^{(1)} + \frac{1}{2}(\Delta\mathbf{x}^{(1)})^T \mathbf{H}(\mathbf{x}^{(1)})\Delta\mathbf{x}^{(1)}$$

Differentiate this equation with respect to $\Delta\mathbf{x}^{(1)}$.

$$\frac{\partial f(\mathbf{x}^{(1)} + \Delta\mathbf{x}^{(1)})}{\partial(\Delta\mathbf{x}^{(1)})} = \nabla f(\mathbf{x}^{(1)}) + \mathbf{H}(\mathbf{x}^{(1)})\Delta\mathbf{x}^{(1)} = 0 \longrightarrow$$

The necessary condition for minimization of function $f(x_1, x_2)$

3. Newton's Method (9/9): Example of Function of Two Variables

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, Starting design point $\mathbf{x}^{(0)}$ = (0, 0)

$x^{(1)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$

$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} f_{x_1} \\ f_{x_2} \end{pmatrix} = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}$

■ 2nd Iteration: Find $\mathbf{x}^{(2)}$

Calculate the small change $\Delta\mathbf{x}^{(1)}$ in design.

$$\mathbf{H}(\mathbf{x}^{(1)})\Delta\mathbf{x}^{(1)} = -\nabla f(\mathbf{x}^{(1)})$$

$$\Delta\mathbf{x}^{(k)} = -[\mathbf{H}(\mathbf{x}^{(k)})]^{-1}\nabla f(\mathbf{x}^{(k)})$$

$$\downarrow \left(-\nabla f(\mathbf{x}^{(1)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ \mathbf{H}(\mathbf{x}^{(1)}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \right)$$

$$\begin{pmatrix} \Delta\mathbf{x}_1^{(1)} \\ \Delta\mathbf{x}_2^{(1)} \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \Delta\mathbf{x}_1^{(1)} \\ \Delta\mathbf{x}_2^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

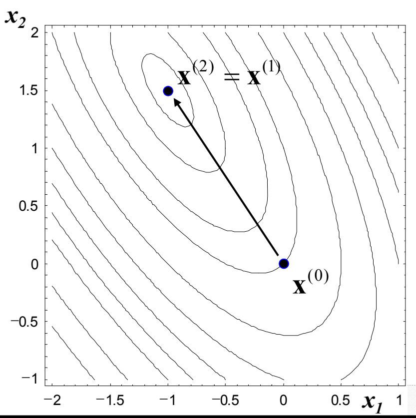
$$\therefore x^{(2)} = x^{(1)} + \Delta\mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

→Optimal design point

Check stopping criterion.

$$\left|\Delta\mathbf{x}^{(1)}\right| = 0 < \varepsilon$$

→Stop!



3. Modified Newton's Method (1/3)

☑ In this method, we treat $\Delta\mathbf{x}^{(k)} = -[\mathbf{H}(\mathbf{x}^{(k)})]^{-1}\nabla f(\mathbf{x}^{(k)})$ of the Newton's method as **the search direction** and use any of the one dimensional search methods to calculate the step size in the search direction.

■ Step 1: Estimate a starting design point $\mathbf{x}^{(0)}$.
Set iteration counter $k = 0$. Specify a tolerance ε for the stopping criterion.

■ Step 2: Calculate $c_i^{(k)} = \partial f(\mathbf{x}^{(k)}) / \partial x_i$ for $i = 1$ to n. If $\left\|\mathbf{c}^{(k)}\right\| < \varepsilon$, then stop the iterative process. Otherwise, continue.

■ Step 3: Calculate the Hessian matrix $\mathbf{H}^{(k)}$ at current design point $\mathbf{x}^{(k)}$.

$$\mathbf{H}(\mathbf{x}^{(k)}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right], \quad i = 1, \cdots, n;\ j = 1, \cdots, n$$

3. Modified Newton's Method (2/3)

- Step 4: Calculate the search direction as follows:

$$\boxed{\mathbf{d}^{(k)} = \Delta \mathbf{x}^{(k)} = -\mathbf{H}^{-1}\mathbf{c}^{(k)}}$$

When $f(\mathbf{x}^* + \Delta\mathbf{x}) = f(\mathbf{x}^*) + \mathbf{c}^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x}$,

the necessary condition for minimization of this function is as follows:

$df(\Delta\mathbf{x})/d\Delta\mathbf{x} = \mathbf{c} + \mathbf{H}(\mathbf{x}^*)\Delta\mathbf{x} = 0$

$\Rightarrow \mathbf{H}(\mathbf{x}^*)\,\Delta\mathbf{x} = -\mathbf{c} \Rightarrow \ \Delta\mathbf{x} = -\mathbf{H}(\mathbf{x}^*)^{-1}\mathbf{c}$

- Step 5: Update the design point as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)}$, where α is calculated to minimize $f(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)})$. Any one dimensional search method may be used to calculate α.

- Step 6: Set $k = k+1$ and go to Step 2.

3. Modified Newton's Method (3/3)
- Disadvantages of the Newton's Method

The Newton's method is not very useful in practice, due to following features of the method:

1. It requires the storing of the $n \times n$ matrix $\mathbf{H}(\mathbf{x}^{(k)})$.
2. It becomes very difficult and sometimes, impossible to compute the elements of the matrix $\mathbf{H}(\mathbf{x}^{(k)})$.
3. It requires the inversion of the matrix $\mathbf{H}(\mathbf{x}^{(k)})$ at each iteration.
4. It requires the evaluation of the quantity $\mathbf{H}(\mathbf{x}^{(k)})^{-1}\nabla f(\mathbf{x}^{(k)})$ at each iteration.

4. Davidon-Fletcher-Powell (DFP) Method (1/6)

☑ This method builds an approximation for the inverse of the Hessian matrix of $f(\mathbf{x})$ using only the first derivatives.

- Step 1: Estimate a starting design point $\mathbf{x}^{(0)}$.
  Choose a symmetric positive definite $n\times n$ matrix $\mathbf{A}^{(0)}$ as an approximation for the inverse of the Hessian matrix of the objective function. In the absence of more information, $\mathbf{A}^{(0)} = \mathbf{I}$ may be chosen. Also, specify a tolerance ε for the stopping criterion. Set $k = 0$ and compute the gradient vector as $\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} \equiv -\nabla f(\mathbf{x}^{(0)})$.

- Step 2: Calculate the norm of the gradient vector as $\|\mathbf{c}^{(k)}\|$.
  If $\|\mathbf{c}^{(k)}\| < \varepsilon$, then stop the iterative process. Otherwise, continue.
  It is noted that Step 1 and 2 of this method and the steepest descent method are the same.

4. Davidon-Fletcher-Powell (DFP) Method (2/6)

- Step 3: Calculate the search direction as follows:

$$\boxed{\mathbf{d}^{(k)} = -\mathbf{A}^{(k)}\mathbf{c}^{(k)}}$$

Newton's method

$$\Delta\mathbf{x}^{(k)} = -[\mathbf{H}(\mathbf{x}^{(k)})]^{-1}\nabla f(\mathbf{x}^{(k)})$$
$$\therefore \mathbf{d}^{(k)} = -(\mathbf{H}^{(k)})^{-1}\mathbf{c}^{(k)}$$

  Here, the matrix $\mathbf{A}$ is used as an estimate for the inverse of the Hessian matrix $\mathbf{H}^{-1}$ of the objective function.

- Step 4: Compute optimum step size $\alpha = \alpha_k$ to minimize $f(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)})$.

- Step 5: Update the design point as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k\mathbf{d}^{(k)}$.

4. Davidon-Fletcher-Powell (DFP) Method (3/6)

- **Step 6: Update the matrix $\mathbf{A}^{(k)}$ - approximation for the inverse of the Hessian matrix of the objective function - as follows:**

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \mathbf{B}^{(k)} + \mathbf{C}^{(k)} \quad ; \; n \times n \text{ matrix}$$

where, the correction matrices $\mathbf{B}^{(k)}$ and $\mathbf{C}^{(k)}$ are calculated as below.

$$\mathbf{B}^{(k)} = \frac{\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}} \quad ; \; n \times n \text{ matrix} \qquad \mathbf{C}^{(k)} = \frac{-\mathbf{z}^{(k)}\mathbf{z}^{(k)^T}}{\mathbf{y}^{(k)^T}\mathbf{z}^{(k)}} \quad ; \; n \times n \text{ matrix}$$

$$\mathbf{s}^{(k)} = \alpha_k \mathbf{d}^{(k)} \quad : \; n \times 1 \text{ matrix}$$

$$\mathbf{y}^{(k)} = \mathbf{c}^{(k+1)} - \mathbf{c}^{(k)} \quad : \; n \times 1 \text{ matrix}$$

$$\mathbf{c}^{(k+1)} = \nabla f(\mathbf{x}^{(k+1)}) \quad : \; n \times 1 \text{ matrix}$$

$$\mathbf{z}^{(k)} = \mathbf{A}^{(k)}\mathbf{y}^{(k)} \quad : \; [n \times n][n \times 1] = [n \times 1] \text{ matrix}$$

$\mathbf{d}^{(k)}$: search direction

$\alpha^{(k)}$: optimum step size

- **Step 7: Set $k = k+1$ and go to Step 2.**

4. Davidon-Fletcher-Powell (DFP) Method (4/6): Example

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point $\mathbf{x}^{(0)}$ = (0, 0)**

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

- **1st Iteration: Find $\mathbf{x}^{(1)}$**

$$\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; \mathbf{A}^{(0)} = \mathbf{I}$$

$$\mathbf{c}^{(0)} = \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 1 + 4\cdot 0 + 2\cdot 0 \\ -1 + 2\cdot 0 + 2\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Check stopping criterion.

$$\left\|\mathbf{c}^{(0)}\right\| = \sqrt{1^2 + (-1)^2} = \sqrt{2} > \varepsilon$$

$$\mathbf{d}^{(0)} = -\mathbf{A}^{(0)}\mathbf{c}^{(0)} = -\mathbf{I}\,\mathbf{c}^{(0)} = -\mathbf{c}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0 \mathbf{d}^{(0)}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -\alpha \\ \alpha \end{pmatrix}$$

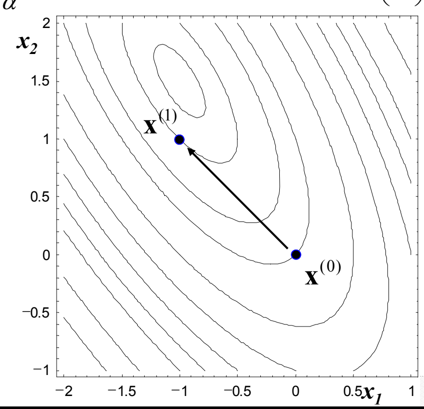
Replacing α_0 to α for convenience

Substitute $\mathbf{x}^{(1)} = (-\alpha, \alpha)$ into the objective function

$$f(\mathbf{x}^{(1)}) = \alpha^2 - 2\alpha$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\frac{df(\mathbf{x}^{(1)})}{d\alpha} = 2\alpha - 2 = 0 \;\rightarrow\; \alpha = 1.0 \qquad \therefore \mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



4. Davidon-Fletcher-Powell (DFP) Method (5/6): Example

$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}$

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point $\mathbf{x}^{(0)}$ = (0, 0)**

■ **2nd Iteration: Find $\mathbf{x}^{(2)}$**

Update the matrix $\mathbf{A}^{(1)}$ - approximation for the inverse of the Hessian matrix of the objective function - as follows:

$$\mathbf{A}^{(1)} = \mathbf{A}^{(0)} + \mathbf{B}^{(0)} + \mathbf{C}^{(0)}$$

$$\mathbf{B}^{(0)} = \frac{\mathbf{s}^{(0)}\mathbf{s}^{(0)^T}}{\mathbf{s}^{(0)^T}\mathbf{y}^{(0)}}$$

$$\mathbf{s}^{(0)} = \alpha\mathbf{d}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{c}^{(0)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{c}^{(1)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{y}^{(0)} = \mathbf{c}^{(1)} - \mathbf{c}^{(0)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$\mathbf{s}^{(0)}\mathbf{s}^{(0)^T} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\mathbf{s}^{(0)^T}\mathbf{y}^{(0)} = 2$$

$$= \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix}$$

$$\mathbf{C}^{(0)} = \frac{-\mathbf{z}^{(0)}\mathbf{z}^{(0)T}}{\mathbf{y}^{(0)^T}\mathbf{z}^{(0)}}$$

$$\mathbf{A}^{(0)} = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{z}^{(0)} = \mathbf{A}^{(0)}\mathbf{y}^{(0)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$\mathbf{y}^{(0)^T}\mathbf{z}^{(0)} = 4$$

$$\mathbf{z}^{(0)}\mathbf{z}^{(0)^T} = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{A}^{(1)} = \mathbf{A}^{(0)} + \mathbf{B}^{(0)} + \mathbf{C}^{(0)}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 1.5 \end{pmatrix}$$

33

4. Davidon-Fletcher-Powell (DFP) Method (6/6): Example

$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}$

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point $\mathbf{x}^{(0)}$ = (0, 0)**

■ **2nd Iteration: Find $\mathbf{x}^{(2)}$**

Check stopping criterion.

$$\left\|\mathbf{c}^{(1)}\right\| = \sqrt{2} > \varepsilon$$

$$\mathbf{d}^{(1)} = -\mathbf{A}^{(1)}\mathbf{c}^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \alpha_1\mathbf{d}^{(1)}$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \alpha\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1+\alpha \end{pmatrix}$$

Replacing α_1 to α for convenience

Substitute $\mathbf{x}^{(2)} = (-1, 1+\alpha)$ **into the objective function**

$$f(\mathbf{x}^{(2)}) = \alpha^2 - \alpha - 1$$

To minimize $f(\mathbf{x}^{(2)})$,

$$\frac{df(\mathbf{x}^{(2)})}{d\alpha} = 2\alpha - 1 = 0 \;\rightarrow\; \alpha = 0.5$$

$$\therefore \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

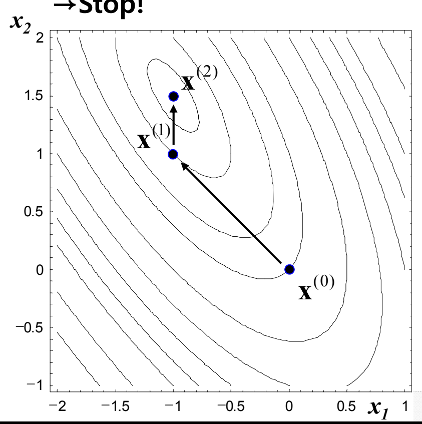
→Optimal design point

$$\mathbf{c}^{(2)} = \nabla f(\mathbf{x}^{(2)}) = \begin{pmatrix} 1+4\cdot(-1)+2\cdot1.5 \\ -1+2\cdot(-1)+2\cdot1.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Check stopping criterion.

$$\left\|\mathbf{c}^{(2)}\right\| = 0 \le \varepsilon$$

→Stop!



5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (1/6)

☑ This method updates the Hessian matrix rather than its inverse at every iteration.

- Step 1: Estimate a starting design point $\mathbf{x}^{(0)}$.
  Choose a symmetric positive definite $n \times n$ matrix $\tilde{\mathbf{H}}^{(0)}$ as an approximation for the Hessian matrix of the objective function.
  In the absence of more information, let $\tilde{\mathbf{H}}^{(0)} = \mathbf{I}$. Specify a tolerance ε for the stopping criterion. Set $k = 0$, and compute the gradient vector as $\mathbf{c}^{(0)} = \nabla f(\mathbf{x}^{(0)})$.

- Step 2: Calculate the norm of the gradient vector as $\left\|\mathbf{c}^{(k)}\right\|$.
  If $\left\|\mathbf{c}^{(k)}\right\| < \varepsilon$, then stop the iterative process. Otherwise, continue.
  It is noted that Step 1 and 2 of this method and the steepest descent method are the same.

5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (2/6)

- Step 3: Solve the linear system of the following equation to obtain the search direction.

$$\boxed{\mathbf{d}^{(k)} = -(\tilde{\mathbf{H}}^{(k)})^{-1}\mathbf{c}^{(k)}}$$

Newton's method

$\Delta\mathbf{x}^{(k)} = -[\mathbf{H}(\mathbf{x}^{(k)})]^{-1}\nabla f(\mathbf{x}^{(k)})$

$\therefore \mathbf{d}^{(k)} = -(\mathbf{H}^{(k)})^{-1}\mathbf{c}^{(k)}$

  This equation looks like $\mathbf{d}^{(k)} = -(\mathbf{H}^{(k)})^{-1}\mathbf{c}^{(k)}$ of the Newton's method, but $\tilde{\mathbf{H}}^{(k)}$ is an approximated Hessian matrix $\mathbf{H}^{(k)}$, comprised of the first order derivatives.

- Step 4: Compute optimum step size $\alpha = \alpha_k$ to minimize $f(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)})$.

- Step 5: Update the design point as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k\mathbf{d}^{(k)}$.

5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (3/6)

- **Step 6: Update the matrix $\tilde{\mathbf{H}}^{(k)}$ - approximation for the Hessian matrix of the objective function - as follows:**

$$\tilde{\mathbf{H}}^{(k+1)} = \tilde{\mathbf{H}}^{(k)} + \mathbf{D}^{(k)} + \mathbf{E}^{(k)} \quad : \quad n \times n \text{ matrix}$$

where, the correction matrices $\mathbf{D}^{(k)}$ and $\mathbf{E}^{(k)}$ are given as below.

$$\mathbf{D}^{(k)} = \frac{\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}}{\mathbf{y}^{(k)^T}\mathbf{s}^{(k)}}; \qquad \mathbf{E}^{(k)} = \frac{\mathbf{c}^{(k)}\mathbf{c}^{(k)^T}}{\mathbf{c}^{(k)^T}\mathbf{d}^{(k)}};$$

$$\mathbf{s}^{(k)} = \alpha_k \mathbf{d}^{(k)} \quad \text{: change in design}$$

$$\mathbf{y}^{(k)} = \mathbf{c}^{(k+1)} - \mathbf{c}^{(k)} \quad \text{: change in gradient}$$

$$\mathbf{c}^{(k+1)} = \nabla f(\mathbf{x}^{(k+1)})$$

$\mathbf{d}^{(k)}$: search direction

$\alpha^{(k)}$: optimum step size

- **Step 7: Set $k = k+1$ and go to Step 2.**

5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (4/6): Example

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point $\mathbf{x}^{(0)}$ = (0, 0)**

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

- **1st Iteration: Find $\mathbf{x}^{(1)}$**

$$\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \tilde{\mathbf{H}}^{(0)} = \mathbf{I}$$

$$\mathbf{c}^{(0)} = \nabla f(\mathbf{x}^{(0)}) = \begin{pmatrix} 1 + 4\cdot 0 + 2\cdot 0 \\ -1 + 2\cdot 0 + 2\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Check stopping criterion.

$$\left\|\mathbf{c}^{(0)}\right\| = \sqrt{1^2 + (-1)^2} = \sqrt{2} > \varepsilon$$

$$\mathbf{d}^{(0)} = -(\tilde{\mathbf{H}}^{(0)})^{-1}\mathbf{c}^{(0)} = -\mathbf{I}\,\mathbf{c}^{(0)} = -\mathbf{c}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0 \mathbf{d}^{(0)}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -\alpha \\ \alpha \end{pmatrix}$$

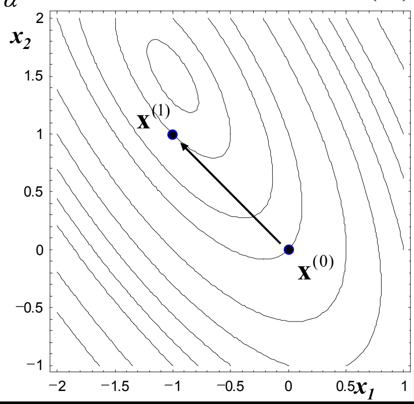
Replacing α_0 to α for convenience

Substitute $\mathbf{x}^{(1)} = (-\alpha, \alpha)$ into the objective function

$$f(\mathbf{x}^{(1)}) = \alpha^2 - 2\alpha$$

To minimize $f(\mathbf{x}^{(1)})$,

$$\frac{df(\mathbf{x}^{(1)})}{d\alpha} = 2\alpha - 2 = 0 \;\rightarrow\; \alpha = 1.0 \quad \therefore \mathbf{x}^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$



5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (5/6): Example

$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1+4x_1+2x_2 \\ 2x_1+2x_2 \end{pmatrix}$

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point** $\mathbf{x}^{(0)}$ **= (0, 0)**

■ **2nd Iteration: Find $\mathbf{x}^{(2)}$**

Update the matrix $\tilde{\mathbf{H}}^{(0)}$ - approximation for the Hessian matrix of the objective function - as follows:

$$\tilde{\mathbf{H}}^{(1)} = \tilde{\mathbf{H}}^{(0)} + \mathbf{D}^{(0)} + \mathbf{E}^{(0)}$$

$$\mathbf{D}^{(0)} = \frac{\mathbf{y}^{(0)}\mathbf{y}^{(0)^T}}{\mathbf{y}^{(0)^T}\mathbf{s}^{(0)}}$$

$$\mathbf{s}^{(0)} = \alpha\mathbf{d}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{c}^{(0)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{c}^{(1)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{y}^{(0)} = \mathbf{c}^{(1)} - \mathbf{c}^{(0)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$\mathbf{y}^{(0)}\mathbf{y}^{(0)^T} = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{y}^{(0)^T}\mathbf{s}^{(0)} = 2$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{E}^{(0)} = \frac{-\mathbf{c}^{(0)}\mathbf{c}^{(0)T}}{\mathbf{c}^{(0)^T}\mathbf{d}^{(0)}}$$

$$\mathbf{c}^{(0)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{d}^{(0)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{c}^{(0)}\mathbf{c}^{(0)^T} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\mathbf{c}^{(0)^T}\mathbf{d}^{(0)} = -2$$

$$= \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$$

$$\tilde{\mathbf{H}}^{(1)} = \tilde{\mathbf{H}}^{(0)} + \mathbf{D}^{(0)} + \mathbf{E}^{(0)}$$
$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}$$
$$= \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

39

5. Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method (6/6): Example

$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = \begin{pmatrix} 1+4x_1+2x_2 \\ 2x_1+2x_2 \end{pmatrix}$

Minimize $f(\mathbf{x}) = f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$, **Starting design point** $\mathbf{x}^{(0)}$ **= (0, 0)**

■ **2nd Iteration: Find $\mathbf{x}^{(2)}$**

Check stopping criterion.

$$\left\|\mathbf{c}^{(1)}\right\| = \sqrt{2} > \varepsilon$$

$$\tilde{\mathbf{H}}^{(1)}\mathbf{d}^{(1)} = -\mathbf{c}^{(1)} \qquad \mathbf{d}^{(1)} = -(\tilde{\mathbf{H}}^{(1)})^{-1}\mathbf{c}^{(1)}$$

$$\tilde{\mathbf{H}}^{(1)} = \begin{pmatrix} 2.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \quad \mathbf{d}^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \mathbf{c}^{(1)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \alpha_1\mathbf{d}^{(1)}$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \alpha\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1+2\alpha \end{pmatrix}$$ **Replacing α_1 to α for convenience**

Substitute $\mathbf{x}^{(2)} = (-1, 1+2\alpha)$ into the objective function

$$f(\mathbf{x}^{(2)}) = 4\alpha^2 - 2\alpha - 1$$

To minimize $f(\mathbf{x}^{(2)})$,

$$\frac{df(\mathbf{x}^{(2)})}{d\alpha} = 8\alpha - 2 = 0 \;\rightarrow\; \alpha = 0.25$$

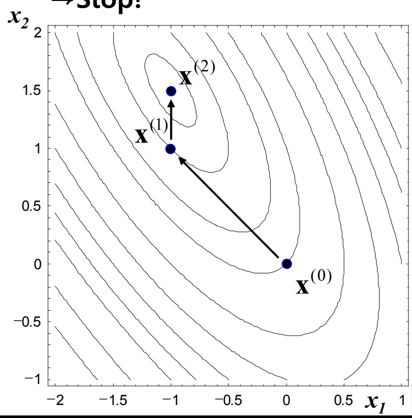
$$\therefore \mathbf{x}^{(2)} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$ **→Optimal design point**

$$\mathbf{c}^{(2)} = \nabla f(\mathbf{x}^{(2)}) = \begin{pmatrix} 1+4\cdot(-1)+2\cdot 1.5 \\ -1+2\cdot(-1)+2\cdot 1.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Check stopping criterion.

$$\left\|\mathbf{c}^{(2)}\right\| = 0 \le \varepsilon$$

→Stop!



Reference Slides

[Review] Taylor Series Expansion for the Function of Two Variables

Taylor series expansion for the function of two variables $f(x_1, x_2)$ at (x_1^*, x_2^*)

$$f(x_1,x_2) = f(x_1^*,x_2^*) + \frac{\partial f}{\partial x_1}(x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x_2 - x_2^*)$$

$$+\frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2}(x_1 - x_1^*)^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 - x_1^*)(x_2 - x_2^*) + \frac{\partial^2 f}{\partial x_2^2}(x_2 - x_2^*)^2\right) + R$$

Each term can be represented as follows:

$$\frac{\partial f}{\partial x_1}(x_1 - x_1^*) + \frac{\partial f}{\partial x_2}(x_2 - x_2^*) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}^T \begin{bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{bmatrix} = \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*)$$

$$\frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2}(x_1 - x_1^*)^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 - x_1^*)(x_2 - x_2^*) + \frac{\partial^2 f}{\partial x_2^2}(x_2 - x_2^*)^2\right) = \frac{1}{2}\left[\frac{\partial^2 f}{\partial x_1^2}(x_1 - x_1^*) + \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_2 - x_2^*) \quad \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 - x_1^*) + \frac{\partial^2 f}{\partial x_2^2}(x_2 - x_2^*)\right]\begin{bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} x_1 - x_1^* & x_2 - x_2^* \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \begin{bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \end{bmatrix}$$

$$= \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$$

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + R$$

Element of the 2x2 Matrix

$$\left(\ \mathbf{x} = (x_1, x_2)^T,\ \mathbf{x}^* = (x_1^*, x_2^*)^T,\ \mathbf{H} \in M_{2\times 2}\right)$$

[Summary] Optimality Conditions for Function of Several Variables

- The Taylor series expansion of $f(\mathbf{x})$, which is the function of n variables, gives

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2}\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + R$$

- From this equation, the change in the function at $\mathbf{x}^*$, i.e., $\Delta f(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^*)$, is given as

$$\Delta f = \nabla f(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2}\mathbf{d}^T \mathbf{H}(\mathbf{x}^*)\mathbf{d} + R$$

- If we assume a local minimum is at $\mathbf{x}^*$, then Δf must be positive.

  1) the first-order necessary condition:
  If $\nabla f(\mathbf{x}^*)^T = \mathbf{0}$, i.e., $\frac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0$, $(i = 1, 2, \cdots n)$, $\mathbf{x}^*$ is a stationary point (minimum, maximum, or inflection point).

  2) the sufficient condition:
  If $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} > 0$, then the stationary point ($\nabla f(\mathbf{x}^*)^T = \mathbf{0} \Rightarrow \nabla f(\mathbf{x}^*) = \mathbf{0}$) is a local minimum.
  To be $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} > 0$, $\mathbf{H}(\mathbf{x}^*)$ must be positive definite.