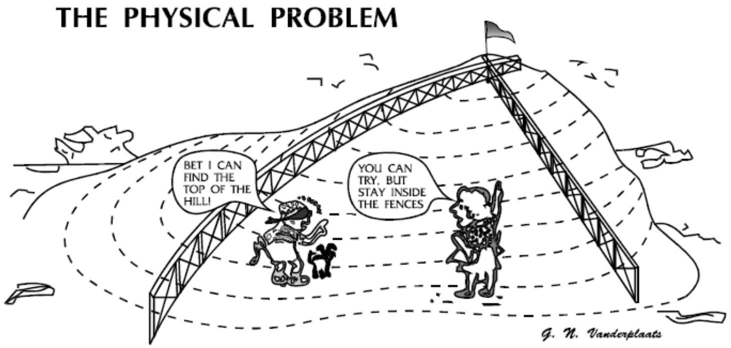
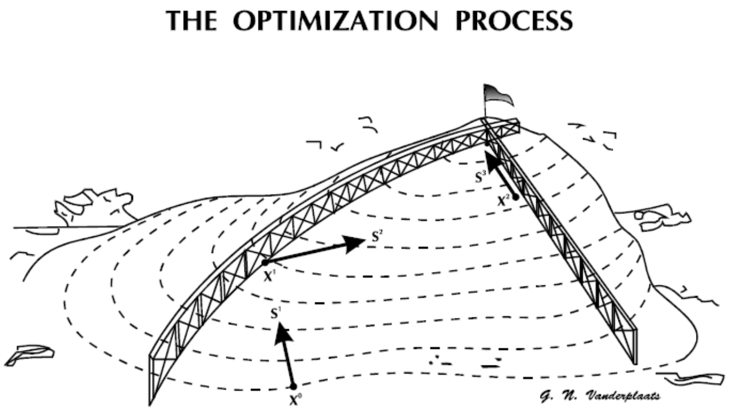
# Lecture Note-Numerical Analysis (9): Introduction to Optimization

## 1. Optimization Examples and Mathematical Description

- ○ There exists a target : Top of the hill → cost function (or objective function)
- ○ There exist various constraints: Inside the fence → constraints (→ inequality constraints)



○ Mathematical expression of the general optimization problem (nonlinear programming problem)

$$\boxed{\begin{aligned}&\min f(\mathbf{x})\\&\text{subject to}\\&\quad d_j(\mathbf{x}) \le a_j, \quad j = 1,2,\cdots,m\\&\quad e_j(\mathbf{x}) = b_j, \quad j = 1,2,\cdots,p\end{aligned}} \quad \text{or} \quad \boxed{\begin{aligned}&\max \{-f(\mathbf{x})\}\\&\text{subject to}\\&\quad d_j(\mathbf{x}) \le a_j, \quad j = 1,2,\cdots,m\\&\quad e_j(\mathbf{x}) = b_j, \quad j = 1,2,\cdots,p\end{aligned}}$$

where

| | | |
|---|---|---|
| $f(\mathbf{x})$ | : | cost function (or objective function) |
| $d_j(\mathbf{x})$ | : | inequality constraint |
| $e_j(\mathbf{x})$ | : | equality constraint |
| $\mathbf{x} = [x_1, \cdots, x_n]^T$ | : | design variables |
| $m$ | : | number of inequality constraints |
| $p$ | : | number of equality constraints |
| $n$ | : | number of design variables |

## 2. Category of Optimization Problem

- m=p=0: unconstrained nonlinear programming problem

$$\boxed{\min f(\mathbf{x})}$$

- m=0: equality constrained nonlinear programming problem

$$\boxed{\begin{array}{l}\min f(\mathbf{x}) \\ \text{s.t. } e_j(\mathbf{x}) = b_j, \quad j = 1,2,\cdots,p\end{array}}$$

- p=0: inequality constrained nonlinear programming problem

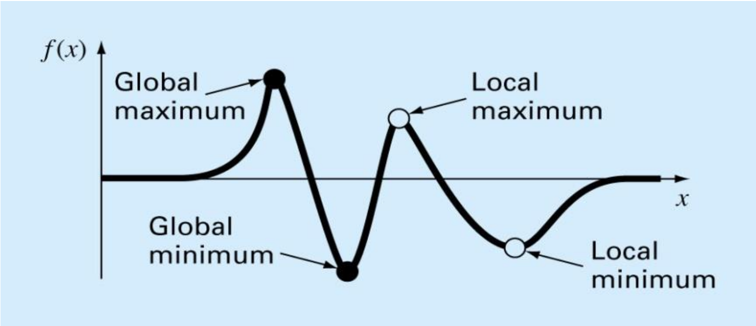
$$\boxed{\begin{array}{l}\min f(\mathbf{x}) \\ \text{s.t } d_j(\mathbf{x}) \le a_j, \quad j = 1,2,\cdots,m\end{array}}$$

- If $f(\mathbf{x}), d_j(\mathbf{x}), e_j(\mathbf{x})$ are linear: linear programming problem

$$\boxed{\begin{array}{l}\min \mathbf{a}^T\mathbf{x} \\ \text{subject to} \\ \quad \mathbf{g}_j^T\mathbf{x} \le b_j, \quad j = 1,2,\cdots,m \\ \quad \mathbf{h}_j^T\mathbf{x} = c_j, \quad j = 1,2,\cdots,p\end{array}} \rightarrow \text{or} \quad \boxed{\begin{array}{l}\min \mathbf{a}^T\mathbf{x} \\ \text{subject to} \\ \quad \mathbf{Gx} \le \mathbf{b}, \quad \mathbf{G} \in R^{m\times n} \\ \quad \mathbf{Hx} = \mathbf{c}, \quad \mathbf{H} \in R^{p\times n}\end{array}}$$

where $\mathbf{a} \in R^{n\times 1}$ and $\mathbf{g}_j \in R^{n\times 1}, \mathbf{h}_j \in R^{n\times 1}$ are vectors.

## 3. Local minimum/maximum and Global minimum/maximum




○ Local minimum/maximum condition for one dimensional problem

- local minimum condition ( $g(x)$ :gradient, $h(x)$ : Hessian)

1st order optimality condition: $g(x) = \dfrac{df(x)}{dx} = 0$

2nd order optimality condition: $h(x) = \dfrac{d^2 f(x)}{dx^2} > 0$

- local maximum condition ( $g(x)$ :gradient, $h(x)$ : Hessian)

1st order optimality condition: $g(x) = \dfrac{df(x)}{dx} = 0$

2nd order optimality condition: $h(x) = \dfrac{d^2 f(x)}{dx^2} < 0$

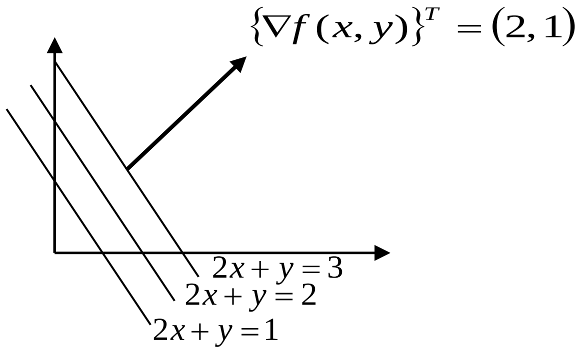
○ $\boxed{g(x) = \dfrac{df(x)}{dx} = 0}$ **→ <u>the 1st optimality condition becomes a nonlinear algebraic equation.</u>**

## 4. Examples of Unconstrained Optimization in two dimensions: Concept of the steepest descent.

○ **Meaning of the gradient of multi-variable functions**

**(a)** $f(x,y)=2x+y$

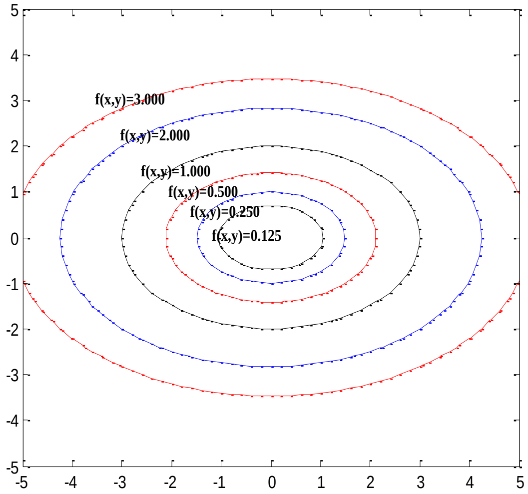
$$\nabla f(x,y)=\begin{pmatrix}\dfrac{\partial f}{\partial x}\\[2mm]\dfrac{\partial f}{\partial y}\end{pmatrix}=\begin{pmatrix}2\\1\end{pmatrix}$$




→ **The Gradient of function provides us with the information for the steepest ascending direction of function value**

**(b)** $f(x,y)=\dfrac{x^2}{9}+\dfrac{y^2}{4}$ : (x, y)= (0, 0) is the global minimum point with $f(x,y)=0$

```
n=50;
for i=1:n+1;
    x(i)=-5+(10/n)*(i-1);
    y(i)=x(i);
end;
a=1/9;b=1/4;
%
for i=1:n+1;
    for j=1:n+1;
        x1=x(i);
        y1=y(j);
%
        xx(i,j)=x1;
        yy(i,j)=y1;
        ff(i,j)=a*x1*x1+b*y1*y1;
    end;
end;
contour(xx,yy,ff,5)
```



- Gradient information at arbitrary points and the corresponding line

At $x=1,\ y=1$ $\rightarrow$ $\{\nabla f(1,1)\}^T = \left(\frac{2}{9},\frac{1}{2}\right)$, $y=\frac{9}{4}(x-1)+1=\frac{9}{4}x-\frac{5}{4}$

At $x=3,\ y=2$ $\rightarrow$ $\{\nabla f(3,2)\}^T = \left(\frac{2}{3},1\right)$, $y=\frac{3}{2}(x-\frac{2}{3})+1=\frac{3}{2}x$

- Analytical solution using the 1st order optimality condition

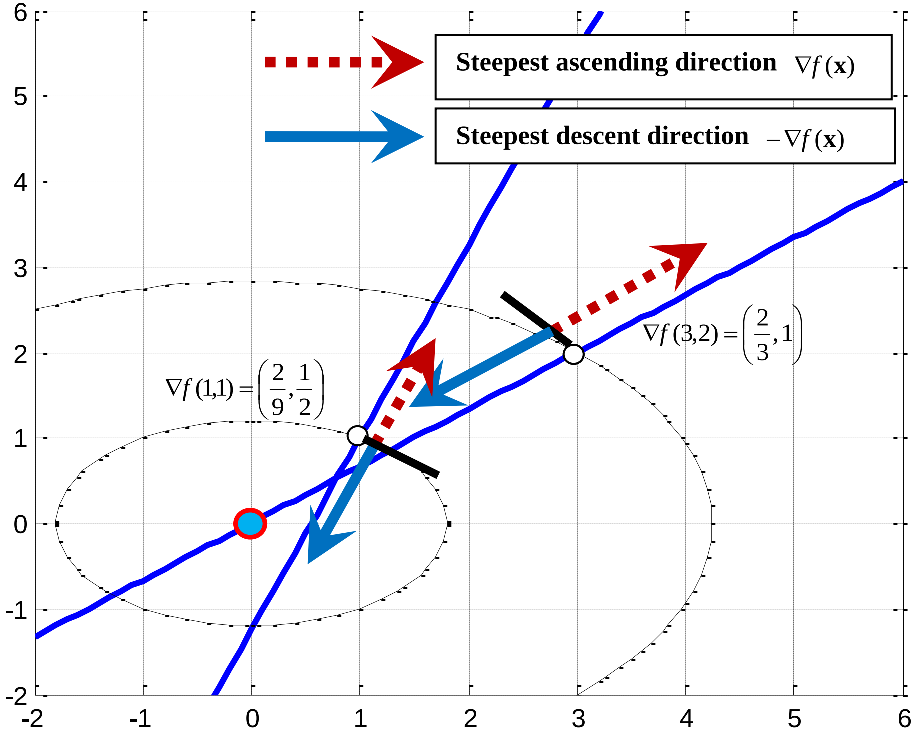
Gradient of the cost function: $f(x,y)=\frac{x^2}{9}+\frac{y^2}{4}$

$$\nabla f(x,y)=\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)=\left(\frac{2x}{9},\frac{y}{2}\right)$$

1st order optimality condition: $\nabla f(x,y)=0$

$$\left(\frac{2x}{9},\frac{y}{2}\right)=(0,0) \rightarrow \boxed{\begin{array}{l} x=0 \\ y=0 \end{array}}$$

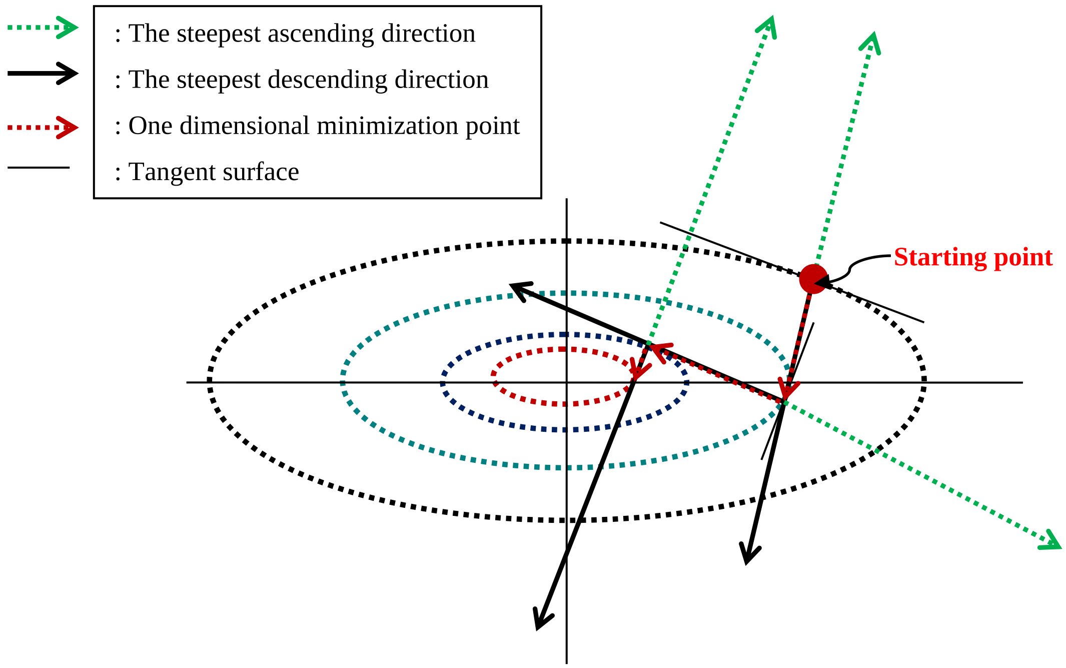
○ Gradient information at arbitrary points and the corresponding line




- <u>Gradient information gives us a local directional information to minimize f(x, y)</u>
- <u>However, the minimum point generally doesn't exit on the steepest descent line.</u>

**Since the gradient is local property, we should limit the maximum traveling distance along the steepest descent line up to the point where f(x, y) has the minimum value along that line.**

## → One dimension search (or One dimensional optimization)

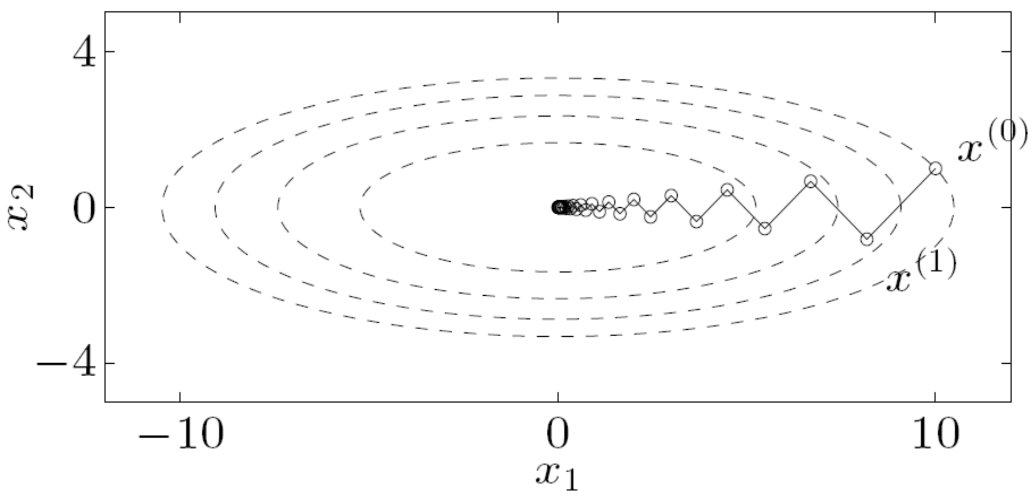


## 5. Steepest Descent Method

- Steepest descent direction: $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ → Search direction
- Iterative update using $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}$ → One dimensional search to define $\alpha$

find $\alpha$ as a solution of $\min_{\alpha} g(\alpha) = f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)})$

**[Example: the steepest descent method]**



## 6. One Dimensional Search / One Dimensional Optimization (Unconstrained case)

○ Problem statement

$$\boxed{\min f(x), \quad x \in R}$$

○ The 1st order optimality condition

$$g(x) = f'(x) = \frac{df(x)}{dx} = 0$$

(6-1) Solution using nonlinear algebraic equation solver: Newton-Raphason Method

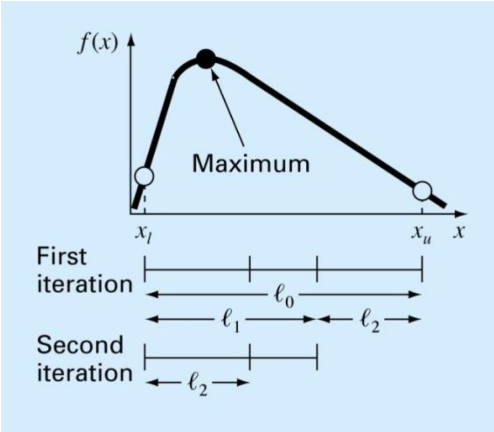
$$\boxed{x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}}$$

(6-2) Other Methods

**- Golden Section Search**

Maximum point exists in the interval $x \in [x_l, x_u]$ , if $f'(x_l)\, f'(x_u) \leq 0$

Crude Approximation of Maximum Value $x^{(k+1)} = \dfrac{x_l^{(k)} + x_u^{(k)}}{2}$




Golden Ratio: R

$$l_0 = l_1 + l_2$$
$$\frac{l_1}{l_0} = \frac{l_2}{l_1}$$

$$\frac{l_1}{l_1 + l_2} = \frac{l_2}{l_1} \qquad R = \frac{l_2}{l_1}$$
$$1 + R = \frac{1}{R} \qquad R^2 + R - 1 = 0$$
$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0.61803$$

**Golden Section Method**

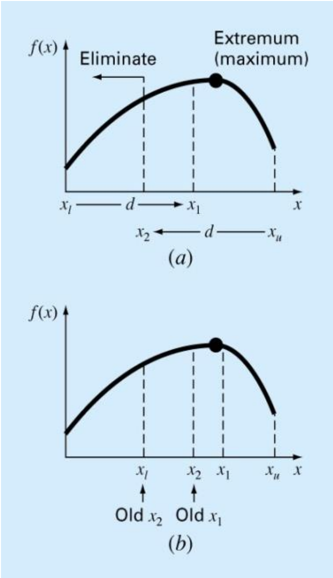
1. Function values at two points

$$d = R(x_u - x_l) = \frac{\sqrt{5}-1}{2}(x_u - x_l)$$
$$x_1 = x_l + d$$
$$x_2 = x_u - d$$
$$f(x_1)$$
$$f(x_2)$$

2. If $f(x_1) > f(x_2)$, $x_l \leftarrow x_2$

3. If $f(x_1) < f(x_2)$, $x_u \leftarrow x_1$

4. Repeat 1~3 until $|x_u - x_l| \le \varepsilon$



## - Interpolation Method (2nd order case)

Given three points of $(x_0, f_0), (x_1, f_1), (x_2, f_3)$

Local quadratic interpolation of f(x) using above three points

$$f(x) = ax^2 + bx + c$$

Local Minimum/maximum point: $f'(x) = 2ax + b = 0 \rightarrow x = -\dfrac{b}{2a}$

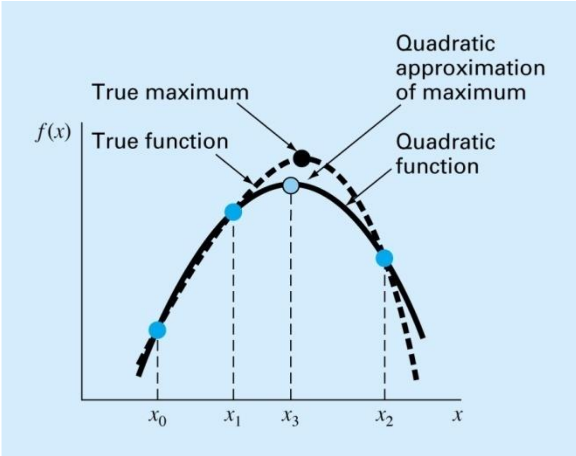
$$ax_0^2 + bx_0 + c = f_0$$
$$ax_1^2 + bx_1 + c = f_1$$
$$ax_2^2 + bx_2 + c = f_2$$

$$\rightarrow \begin{aligned} a(x_1^2 - x_0^2) + b(x_1 - x_0) = (f_1 - f_0) \\ a(x_2^2 - x_1^2) + b(x_2 - x_1) = (f_2 - f_1) \end{aligned}$$

$$\rightarrow \begin{aligned} a(x_1^2 - x_0^2)(f_2 - f_1) + b(x_1 - x_0)(f_2 - f_1) = (f_1 - f_0)(f_2 - f_1) \\ a(x_2^2 - x_1^2)(f_1 - f_0) + b(x_2 - x_1)(f_1 - f_0) = (f_1 - f_0)(f_2 - f_1) \end{aligned}$$

$$\rightarrow a\{(x_2^2 - x_1^2)(f_1 - f_0) - (x_1^2 - x_0^2)(f_2 - f_1)\} + b\{(x_2 - x_1)(f_1 - f_0) - (x_1 - x_0)(f_2 - f_1)\} = 0$$

$$\rightarrow \boxed{x_3 = -\frac{b}{2a} = \frac{1}{2}\frac{\{(x_2^2 - x_1^2)(f_1 - f_0) - (x_1^2 - x_0^2)(f_2 - f_1)\}}{\{(x_2 - x_1)(f_1 - f_0) - (x_1 - x_0)(f_2 - f_1)\}}}$$




If $x_0 \le x_3 < x_1$ , $\boxed{\begin{array}{l} x_0 \leftarrow x_0 \\ x_1 \leftarrow x_3 \\ x_2 \leftarrow x_1 \end{array}}$

If $x_1 \le x_3 \le x_2$ , $\boxed{\begin{array}{l} x_0 \leftarrow x_1 \\ x_1 \leftarrow x_3 \\ x_2 \leftarrow x_2 \end{array}}$

## Appendix A. Optimality Condition of the unconstrained multi-variable Nonlinear Programming Problems

Consider a function $f(\mathbf{x}), f \in R$ where $\mathbf{x}$ is a vector $\mathbf{x} = (x_1, x_2, \cdots, x_{n-1}, x_n)^T, \mathbf{x} \in R^n$. The gradient vector and the Hessian matrix of this function is given by the partial derivatives with respect to each of the independent variables. And higher derivatives of multi-variable functions are defined as in the single-variable case, but note that the number of gradient components increase by a factor of n for each differentiation. While the gradient of a function of n variables is an n-vector, the "second derivative" of an n-variable function is defined by $n^2$ partial derivatives of the n first partial derivatives with respect to the n variables as

$$\frac{\partial^2 f}{\partial x_i \partial x_j}, (i \neq j) \quad \text{and} \quad \frac{\partial^2 f}{\partial x_i^2}, (i = j)$$

If the partial derivatives $\frac{\partial f}{\partial x_i}, \frac{\partial f}{\partial x_j}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ are continuous and $f(\mathbf{x}), f \in R$ is single valued, then $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exists and $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$. Therefore the second order partial derivatives can be represented by a square symmetric matrix called the Hessian matrix, which is contains n(n + 1)=2 independent elements.

- Multi-variable function

$$f(\mathbf{x}) = f(x_1, x_2, \cdots, x_{n-1}, x_n), \quad f \in R, \mathbf{x} \in R^n$$

- Gradient vector and Hessian matrix

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \in R^{n \times 1}, \quad \mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \in R^{n \times n}$$

If $f(\mathbf{x}), f \in R$ is quadratic, the Hessian $\mathbf{H}$ of $f(\mathbf{x})$ is constant and can be expressed as

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

As in the single-variable case the optimality conditions can be derived from the Taylor-series expansion of $f(\mathbf{x})$ about $\mathbf{x}^*$,

$$f(\mathbf{x}^* + \varepsilon\mathbf{p}) \approx f(\mathbf{x}^*) + \varepsilon\mathbf{p}^T\mathbf{g} + \frac{1}{2}\varepsilon^2\mathbf{p}^T\mathbf{H}(\mathbf{x}^* + \varepsilon\theta\mathbf{p})\mathbf{p}$$

where $0 \le \theta \le 1$, $\varepsilon$ is a scalar, and $\mathbf{p}$ is an arbitrary n-vector.

For $\mathbf{x}^*$ to be a local minimum, then for any vector p there must be a finite such that $f(\mathbf{x}^* + \varepsilon\mathbf{p}) \ge f(\mathbf{x}^*)$ i.e. there is a neighborhood in which this condition holds. If this condition is satisfied, then

$$f(\mathbf{x}^* + \varepsilon\mathbf{p}) - f(\mathbf{x}^*) \ge 0$$

and the first and second order terms in the Taylor-series expansion must be greater of equal to zero. As in the single variable case, and for the same reason, we start by considering the first order terms. Since $\mathbf{p}$ is an arbitrary vector and $\varepsilon$ can be positive or negative, every component of the gradient vector $\mathbf{g}(\mathbf{x})$ must be zero. Now we have to consider the second order term, $\frac{1}{2}\varepsilon^2\mathbf{p}^T\mathbf{H}(\mathbf{x}^* + \varepsilon\mathbf{p})\mathbf{p}$. For this term to be non-negative, $\mathbf{H}(\mathbf{x}^* + \varepsilon\mathbf{p})$ has to be positive semi-definite, and by continuity, the Hessian at the optimum, $\mathbf{H}(\mathbf{x}^*)$ must also be positive semi-definite.

- Therefore we can get the following minimum conditions (or minimum optimality condition)

Necessary condition

$\|g(\mathbf{x}^*)\| = 0$ and $\mathbf{H}(\mathbf{x}^*)$ is positive semi-definite

Sufficient condition

$\|g(\mathbf{x}^*)\| = 0$ and $\mathbf{H}(\mathbf{x}^*)$ is positive definite

**[Example] Critical points of a function**

Consider the function: $f(\mathbf{x}) = 1.5x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + 0.5x_1^4$

calculate gradient vector:

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 3x_1 - 2x_2 + 6x_1^2 + 2x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix}$$

Find stationary points

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 3x_1 - 2x_2 + 6x_1^2 + 2x_1^3 \\ 2x_2 - 2x_1 \end{pmatrix} = 0$$

$$\rightarrow \begin{cases} 0 = 3x_1 - 2x_2 + 6x_1^2 + 2x_1^3 \\ x_1 = x_2 \end{cases} \rightarrow \begin{cases} 0 = x_1 + 6x_1^2 + 2x_1^3 \\ x_1 = x_2 \end{cases} \rightarrow \begin{cases} x_1 = x_2 = 0 \\ x_1 = x_2 = -3 \pm \sqrt{7} \end{cases}$$

Therefore, the stationary points are

$(x_1, x_2) \rightarrow$ $(0,0)$, $(-3+\sqrt{7}, -3+\sqrt{7})$, $(-3-\sqrt{7}, -3-\sqrt{7})$

Calculate Hessian matrix at each stationary point and the corresponding eigenvalues :

$$\mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 3+12x_1+8x_1^2 & -2 \\ -2 & 2 \end{pmatrix}$$

$$\mathbf{H}(0,0) = \begin{pmatrix} 3 & -2 \\ -2 & 2 \end{pmatrix} \rightarrow \begin{vmatrix} 3-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = 0 \rightarrow \lambda^2 - 5\lambda + 2 = 0 \rightarrow \lambda = \frac{5 \pm \sqrt{17}}{2} > 0$$

→ Positive definite

$$\mathbf{H}(-3\pm\sqrt{7}, -3\pm\sqrt{7}) = \begin{pmatrix} -1-12\left(-3\pm\sqrt{7}\right) & -2 \\ -2 & 2 \end{pmatrix} = \begin{pmatrix} 35\mp 12\sqrt{7} & -2 \\ -2 & 2 \end{pmatrix}$$

$$\begin{vmatrix} 35\mp 12\sqrt{7}-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = 0 \rightarrow \lambda^2 - \left(37\mp 12\sqrt{7}\right)\lambda + 66 \mp 24\sqrt{7} = 0$$

$\lambda = 68.7, \quad 0.036$ for $x_1 = x_2 = -3-\sqrt{7}$ → positive definite

$\lambda = 2.62 \pm 11.07j$ for $x_1 = x_2 = -3+\sqrt{7}$ → indefinite

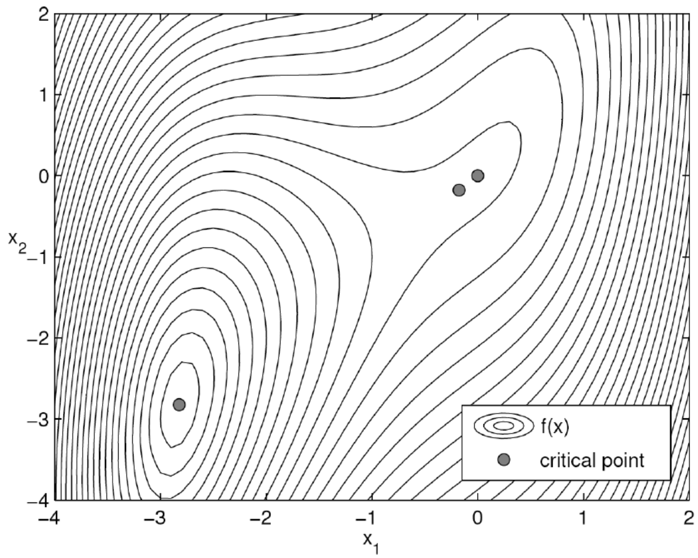



Figure 3.1: Critical points of $f(x) = 1.5x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + 0.5x_1^4$

Therefore, the stationary points are classified as

$(x_1, x_2) = (0,0)$ : local minimum point

$(x_1, x_2) = (-3+\sqrt{7}, -3+\sqrt{7})$ : saddle point

$(x_1, x_2) = (-3-\sqrt{7}, -3-\sqrt{7})$ : global minimum point

## Appendix B. Lagrange multiplier method for the constrained multi-variable Nonlinear Programming Problems

### (B-1) Problem statement (Example)

$$\min\ f(x) = (x_1 + x_2)^2 + (x_2 + x_3)^2$$

$$s.t.$$

$$x_1 + 2x_2 + 3x_3 - 1 = 0$$

**Start with** $x_0 = (-4,1,1)$, $f(x_0) = 13$

**[Answer]**

$$x^* = (0.5, -0.5, 0.5)$$

$$f(x^*) = 0$$

### (B-2) Solution by direct substitution

$$\begin{aligned} f(x) &= (x_1 + x_2)^2 + (x_2 + x_3)^2 \\ &= (-2x_2 - 3x_3 + 1 + x_2)^2 + (x_2 + x_3)^2 \\ &= (1 - x_2 - 3x_3)^2 + (x_2 + x_3)^2 \end{aligned}$$

$\rightarrow$

$$\frac{\partial f(x)}{\partial x_2} = -2(1 - x_2 - 3x_3) + 2(x_2 + x_3) = 0$$

$$\frac{\partial f(x)}{\partial x_3} = -6(1 - x_2 - 3x_3) + 2(x_2 + x_3) = 0$$

**Therefore,**

$$x_2 = -x_3$$

$$1 - x_2 - 3x_3 = 1 - 2x_3 = 0$$

$$x_1 + 2x_2 + 3x_3 - 1 = 0$$

$\rightarrow$

$$x_1 = 2 - 3/2 = 1/2$$

$$x_2 = -1/2$$

$$x_3 = 1/2$$

**(B-3) Solution Using Lagrange multiplier**

**Building Lagrangian function using the Lagrange multiplier**

$$L(x) = f(x) - \mu g(x)$$

$$= (x_1 + x_2)^2 + (x_2 + x_3)^2 - \mu(x_1 + 2x_2 + 3x_3 - 1)$$

**Treat the Lagrange multipliers as design variables to get a first-order optimality condition**

**(4 unknowns** $x_1, x_2, x_3, \mu$ **with 4 equations)**

$$\frac{\partial L(x)}{\partial x_1} = 2(x_1 + x_2) - \mu = 0$$

$$\frac{\partial L(x)}{\partial x_2} = 2(x_1 + x_2) + 2(x_2 + x_3) - 2\mu = 0$$

$$\frac{\partial L(x)}{\partial x_3} = 2(x_2 + x_3) - 3\mu = 0$$

$$\frac{\partial L(x)}{\partial \mu} = -(x_1 + 2x_2 + 3x_3 - 1) = 0$$

**Hessian Matrix**

$$\rightarrow \begin{pmatrix} 2 & 2 & 0 & 1 \\ 2 & 4 & 2 & 2 \\ 0 & 2 & 2 & 3 \\ 1 & 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ -\mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \textbf{or} \quad \begin{matrix} x_1 + x_2 = \mu/2 \\ x_2 + x_3 = \mu/2 \\ -2\mu = 0 \\ 2x_3 - 1 = 0 \end{matrix} \rightarrow \begin{matrix} x_1 = 1/2 \\ x_2 = -1/2 \\ x_2 = 1/2 \\ \mu = 0 \end{matrix}$$

**Transpose of Constraint gradient**

## Appendix C : Formal Optimality Conditions for the NLP with Nonlinear Equality Constraints

The optimality conditions for nonlinearly constrained problems are important because they form the basis for algorithms for solving such problems.

Suppose we have the following optimization problem with equality constraints,

$$\boxed{\begin{aligned}&\min\ f(\mathbf{x})\\&s.t.\quad h_j(\mathbf{x})=0\quad j=1,\cdots,m\\&(f\in R,\mathbf{x}\in R^n,\ h_j\in R)\end{aligned}}$$

To solve this problem, we could solve for m components of $\mathbf{x}$ by using the equality constraints to express them in terms of the other components. The result would be an unconstrained problem with *n-m* variables. However, this procedure is only feasible for simple explicit functions.

Lagrange devised a method to solve this problem.
At a stationary point, the total differential of the objective function has to be equal to zero, i.e.,

$$df=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\cdots+\frac{\partial f}{\partial x_n}dx_n=0$$

For a feasible point, the total differential of the constraints must also be zero, and so

$$h_j+dh_j=h_j+\frac{\partial h_j}{\partial x_1}dx_1+\frac{\partial h_j}{\partial x_2}dx_2+\cdots+\frac{\partial h_j}{\partial x_n}dx_n=0,\quad j=1,\cdots,m$$

Lagrange suggested that one could multiply each constraint variation by a scalar $\lambda_j$ (called the Lagrange multiplier) and subtract it to from the objective function rather than eliminating some of $dx_k$ s as

$$\begin{aligned}df-\sum_{j=1}^{m}\lambda_j dh_j&=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2+\cdots+\frac{\partial f}{\partial x_n}dx_n-\sum_{j=1}^{m}\lambda_j\left(\frac{\partial h_j}{\partial x_1}dx_1+\frac{\partial h_j}{\partial x_2}dx_2+\cdots+\frac{\partial h_j}{\partial x_n}dx_n\right)\\&=\left(\frac{\partial f}{\partial x_1}-\sum_{j=1}^{m}\lambda_j\frac{\partial h_j}{\partial x_1}\right)dx_1+\left(\frac{\partial f}{\partial x_2}-\sum_{j=1}^{m}\lambda_j\frac{\partial h_j}{\partial x_2}\right)dx_2+\cdots+\left(\frac{\partial f}{\partial x_n}-\sum_{j=1}^{m}\lambda_j\frac{\partial h_j}{\partial x_n}\right)dx_n\end{aligned}$$

Note that the components of variation vector $dx_k$ are independent and arbitrary, since we have already accounted for the constraint. Thus, for this equation to be satisfied, we need a vector $\lambda_j$ such that the expression inside the parenthesis vanishes, i.e.,

$$\left( \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial h_j}{\partial x_i} \right) = 0, \quad i = 1, 2, \cdots, n$$

Therefore, we have (n+m) equations to solve (n+m) variables as

$$\left( \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda_j \frac{\partial h_j}{\partial x_i} \right) = 0, \quad i = 1, 2, \cdots, n$$

$$h_j(\mathbf{x}) = 0 \quad j = 1, \cdots, m$$

with $\mathbf{x} = (x_1, x_2, \cdots, x_n)$

$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)$

If we define the Lagrangian function with constant $\lambda_j$ s as

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{j=1}^{m} \lambda_j h_j(\mathbf{x})$$

If $\mathbf{x}^*$ is a stationary point of this function, using the necessary conditions for unconstrained problems, we obtain

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}} f(\mathbf{x}) - \sum_{j=1}^{m} \lambda_j \nabla_{\mathbf{x}} h_j(\mathbf{x}) = 0$$ (the constraint normal is the same direction as $\nabla_{\mathbf{x}} f(\mathbf{x})$ )

$$\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}) = -\mathbf{h}(\mathbf{x}) = 0 \text{ or } h_j(\mathbf{x}) = 0$$

These first order conditions are known as the **Karush-Kuhn-Tucker (KKT) conditions** and are necessary conditions for the optimum of a constrained problem.

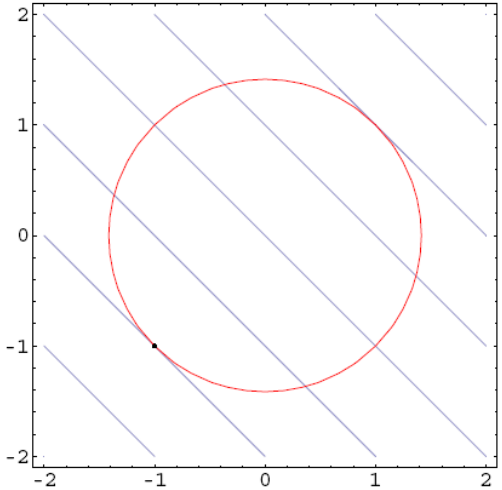
Note that the Lagrangian function is defined such that minimizing it with respect to the design variables and the Lagrange multipliers, we obtain the constraints of the original function. We have transformed a constrained optimization problem of n variables and m constraints into an unconstrained problem of n +m variables.

[Example]

$$\boxed{\begin{aligned} &\min \ f(\mathbf{x}) = x_1 + x_2 \\ &s.t. \quad x_1^2 + x_2^2 - 2 = 0 \end{aligned}}$$

with the graphic solution.

By inspection we can see that the feasible region for this problem is a circle of radius $\sqrt{2}$. The solution $\mathbf{x}$ is obviously $(-1,-1)^T$. From any other point in the circle it is easy to find a way to move in the feasible region (the boundary of the circle) while decreasing f.




**Solution**

Lagrangian $\quad L(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2)$

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1} = 0 \rightarrow 1 - 2\lambda x_1 = 0$$

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial x_1} = 0 \rightarrow 1 - 2\lambda x_2 = 0 \qquad \rightarrow \quad \boxed{\begin{aligned} &x_1 = x_2 = \pm 1 \\ &\lambda = \frac{1}{2x_1} = \pm\frac{1}{2} \end{aligned}}$$

$$\frac{\partial L(x_1,x_2,\lambda)}{\partial \lambda} = 0 \rightarrow -(x_1^2 + x_2^2 - 2) = 0$$

$\boxed{x_1 = x_2 = 1}$ is the maximum point and $\boxed{x_1 = x_2 = -1}$ is the minimum point. We can distinguish these two situations by checking for positive definiteness of the Hessian of the Lagrangian.

**Appendix D : NLP with inequality constraints: Introduction to the active set strategy**

## (D-1) Definition of Active set

Let's define the index set $I(\mathbf{x})$ for active inequality constraints as

$$I(\mathbf{x}) = \{j : g_j(\mathbf{x}) \geq 0\}$$

The index set for equality constraints, which is always active, can be expresses as

$$E(\mathbf{x}) = \{j : h_j(\mathbf{x}) = 0,\ j = 1, \cdots, l\}$$

Then we can define the active set $A(\mathbf{x})$ as

$$A(\mathbf{x}) = I(\mathbf{x}) \bigcup E(\mathbf{x})$$

***(Note) You should be careful on the local property of active set, which means the set $A(\mathbf{x})$ is function of $\mathbf{x}$. Therefore, its components can be changed during iterative numerical solution process***

## (D-2) Active set Method

If we consider the active constraints, then the NLP equation can be rewritten since inactive constraints have no effect on the optimal solution

$$\boxed{\begin{aligned}&\min\ f(\mathbf{x})\\&s.t.\quad g_j(\mathbf{x})=0\quad j\in I(\mathbf{x})\\&\qquad\ \ h_j(\mathbf{x})=0\quad j\in A(\mathbf{x})=\{1,\cdots,l\}\end{aligned}}$$

The same as the previous section, the Lagrangian can be defined as

$$\begin{aligned}L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu}) &= f(\mathbf{x})-\sum_{j=1}^{m}\lambda_j h_j(\mathbf{x})-\sum_{j\in I(\mathbf{x})}\mu_j g_j(\mathbf{x})\\&= f(\mathbf{x})-\boldsymbol{\lambda}^T\mathbf{h}(\mathbf{x})-\boldsymbol{\mu}^T\mathbf{g}(\mathbf{x})\end{aligned}$$

where

$$\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$$

$$\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_l)^T$$

$$\boldsymbol{\mu}=(\mu_j,\quad j\in I(\mathbf{x}))^T$$

$$\mathbf{h}=(h_1,h_2,\cdots,h_l)^T$$

$$\mathbf{g}=(g_j,\quad j\in I(\mathbf{x}))^T$$

The corresponding KKT condition can be defined as

$$\nabla_{\mathbf{x}}L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s})=\nabla_{\mathbf{x}}f(\mathbf{x})-\sum_{j=1}^{m}\lambda_j\nabla_{\mathbf{x}}h_j(\mathbf{x})-\sum_{j\in I(\mathbf{x})}\mu_j\nabla_{\mathbf{x}}g_j(\mathbf{x})=0$$

$$\mathbf{h}=(h_1,h_2,\cdots,h_l)^T=\mathbf{0}$$

$$\mathbf{g}=(g_j,\quad j\in I(\mathbf{x}))^T=\mathbf{0}$$

which says (if we renumber the index for active inequality constraints, $\mathbf{g} = (g_j, \quad j \in I(\mathbf{x}))^T = \mathbf{0}$

$$\begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix} - \begin{pmatrix} \lambda_1 \dfrac{\partial h_1}{\partial x_1} + \lambda_2 \dfrac{\partial h_2}{\partial x_1} + \cdots + \lambda_l \dfrac{\partial h_l}{\partial x_1} \\ \lambda_1 \dfrac{\partial h_1}{\partial x_2} + \lambda_2 \dfrac{\partial h_2}{\partial x_2} + \cdots + \lambda_l \dfrac{\partial h_l}{\partial x_2} \\ \vdots \\ \lambda_1 \dfrac{\partial h_1}{\partial x_n} + \lambda_2 \dfrac{\partial h_2}{\partial x_n} + \cdots + \lambda_l \dfrac{\partial h_l}{\partial x_n} \end{pmatrix} - \begin{pmatrix} \mu_1 \dfrac{\partial g_1}{\partial x_1} + \mu_2 \dfrac{\partial g_2}{\partial x_1} + \cdots + \mu_{I(x)} \dfrac{\partial g_{I(x)}}{\partial x_1} \\ \mu_1 \dfrac{\partial g_1}{\partial x_2} + \mu_2 \dfrac{\partial g_2}{\partial x_2} + \cdots + \mu_{I(x)} \dfrac{\partial g_{I(x)}}{\partial x_2} \\ \vdots \\ \mu_1 \dfrac{\partial g_1}{\partial x_n} + \mu_2 \dfrac{\partial g_2}{\partial x_n} + \cdots + \mu_{I(x)} \dfrac{\partial g_{I(x)}}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_l \end{pmatrix} = \mathbf{0}, \quad \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{I(\mathbf{x})} \end{pmatrix} = \mathbf{0}$$

or

$$\left.\begin{aligned}&\nabla_{\mathbf{x}} f(\mathbf{x}) - \{\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x})\}\boldsymbol{\lambda} - \{\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x})\}\boldsymbol{\mu} = 0\\ &\mathbf{h}(\mathbf{x}) = \mathbf{0}\\ &\mathbf{g}(\mathbf{x}) = \mathbf{0}\end{aligned}\right. \rightarrow \boxed{\begin{aligned}&\nabla_{\mathbf{x}} f(\mathbf{x}) - \mathbf{A}\boldsymbol{\lambda} - \mathbf{G}\boldsymbol{\mu} = 0\\ &\mathbf{h}(\mathbf{x}) = \mathbf{0}\\ &\mathbf{g}(\mathbf{x}) = \mathbf{0}\end{aligned}}$$

where the gradient of constraint functions are defined as

$$\mathbf{A} = \nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1}, & \frac{\partial h_2}{\partial x_1}, & \cdots, & \frac{\partial h_l}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2}, & \frac{\partial h_2}{\partial x_2}, & \cdots, & \frac{\partial h_l}{\partial x_2} \\ & \vdots & & \\ \frac{\partial h_1}{\partial x_n}, & \frac{\partial h_2}{\partial x_n}, & \cdots, & \frac{\partial h_l}{\partial x_n} \end{pmatrix}, \quad \mathbf{G} = \nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}, & \frac{\partial g_2}{\partial x_1}, & \cdots, & \frac{\partial g_{I(x)}}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2}, & \frac{\partial g_2}{\partial x_2}, & \cdots, & \frac{\partial g_{I(x)}}{\partial x_2} \\ & \vdots & & \\ \frac{\partial g_1}{\partial x_n}, & \frac{\partial g_2}{\partial x_n}, & \cdots, & \frac{\partial g_{I(x)}}{\partial x_n} \end{pmatrix}$$

## (D-3) NLP with Nonlinear Equality and Inequality Constraints

### Problem Statement

$$\boxed{\begin{aligned}&\min f(\mathbf{x})\\ &s.t.\quad g_j(\mathbf{x}) \le 0 \quad j=1,\cdots,m\\ &\qquad\;\; h_j(\mathbf{x}) = 0 \quad j=1,\cdots,l\\ &(f\in R,\ \mathbf{x}\in R^n,\ g_j\in R,\ h_j\in R)\end{aligned}}$$

We can modify the above problem by introducing slack variables as

$$\boxed{\begin{aligned}&\min f(\mathbf{x})\\ &s.t.\quad g_j(\mathbf{x}) + s_j^2 = 0 \quad j=1,\cdots,m\\ &\qquad\;\; h_j(\mathbf{x}) = 0 \quad j=1,\cdots,l\end{aligned}}$$

The same as the previous section, the Lagrangian can be defined as

$$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s}) = f(\mathbf{x}) - \sum_{j=1}^{m}\lambda_j h_j(\mathbf{x}) - \sum_{j=1}^{l}\mu_j\left\{g_j(\mathbf{x}) + s_j^2\right\}$$

The corresponding stationary condition can be defined as

$$\nabla_{\mathbf{x}} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s}) = \nabla_{\mathbf{x}} f(\mathbf{x}) - \sum_{j=1}^{m}\lambda_j \nabla_{\mathbf{x}} h_j(\mathbf{x}) - \sum_{j=1}^{l}\mu_j \nabla_{\mathbf{x}} g_j(\mathbf{x}) = 0 \qquad (1)$$

$$\nabla_{\boldsymbol{\lambda}} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s}) = -\mathbf{h}(\mathbf{x}) = 0 \ \text{ or } h_j(\mathbf{x}) = 0,\ \ j = 1,2,\cdots,m \qquad (2)$$

$$\nabla_{\boldsymbol{\mu}} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s}) = -\mathbf{g}(\mathbf{x}) = 0 \ \text{ or } g_j(\mathbf{x}) = 0,\ \ j = 1,2,\cdots,l \qquad (3)$$

$$\nabla_{s_j} L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\mu},\mathbf{s}) = -2\mu_j s_j = 0,\ \ j = 1,2,\cdots,l \qquad (4)$$

Now we have *n* + *m* + *2l* equations and two main possibilities for each inequality constraint

(1) $s_j \neq 0$: the j-th inequality constraint is inactive , and $\mu_j$ = 0.

(2) $s_j = 0$: the j-th inequality constraint is active , and $\mu_j$ must then be non-positive, otherwise from the first equations, the gradient of objective and gradient of constraint point in the same direction.

Again the equations (1)~(4) are as the **Karush-Kuhn-Tucker (KKT) conditions** and are necessary conditions for the optimum of a constrained problem. The point where the KKT conditions satisfied is call a KKT point.

The last condition $\mu_j s_j = 0$ is known **as a complementarity condition (or complementarity slackness**) and implies that the Lagrange multiplier can be strictly negative only when the constraint is active. Therefore, the equation (3) can be applicable only for the active inequality constraints.

$$\nabla_{\boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{s}) = -\mathbf{g}(\mathbf{x}) = 0 \text{ or } g_j(\mathbf{x}) = 0, \ \ j = 1, 2, \cdots, l \qquad (3)$$

only for active inequality constraints

## Appendix E: Example NLP Test Problems

---

**Problem 1**

$$\min\ f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$s.t.$$

$$-1.5 \le x_2$$

**Start with** $x_0 = (-2,1)$, $f(x_0) = 909$

**[Answer]**

$$x^* = (1,1)$$

$$f(x^*) = 0$$

---

**Problem 4**

$$\min\ f(x) = \frac{1}{3}(x_1 + 1)^2 + x_2$$

$$s.t.$$

$$1 \le x_1$$

$$0 \le x_2$$

**Start with** $x_0 = (1.125, 0.125)$, $f(x_0) = 3.323568$

**[Answer]**

$$x^* = (1,0)$$

$$f(x^*) = \frac{8}{3}$$

**Problem 13**

$$\min f(x) = (x_1 - 2)^2 + x_2^2$$

$$s.t.$$

$$(1 - x_1)^3 - x_2 \geq 0$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

**Start with** $x_0 = (-2, -2)$, $f(x_0) = 20$

**[Answer]**

$$x^* = (1, 0)$$

$$f(x^*) = 1$$

**Problem 14**

$$\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$$

$$s.t.$$

$$-0.25x_1^2 - x_2^2 + 1 \geq 0$$

$$x_1 - 2x_2 + 1 = 0$$

**Start with** $x_0 = (2, 2)$, $f(x_0) = 1$

**[Answer]**

$$x^* = (0.5(\sqrt{7} - 1), 0.5(\sqrt{7} + 1))$$

$$f(x^*) = 9 - 2.875\sqrt{7}$$

**Problem 15**

$$\min\ f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$s.t.$$

$$x_1 x_2 - 1 \geq 0$$

$$x_1 + x_2^2 \geq 0$$

$$x_1 \leq 0.5$$

**Start with** $x_0 = (-2, 1)$, $f(x_0) = 909$

**[Answer]**

$$x^* = (0.5, 2)$$

$$f(x^*) = 306.5$$

---

**Problem 20**

$$\min\ f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$s.t.$$

$$x_1 + x_2^2 \geq 0$$

$$x_1^2 + x_2 \geq 0$$

$$x_1^2 + x_2^2 - 1 \geq 0$$

$$-0.5 \leq x_1 \leq 0.5$$

**Start with** $x_0 = (-2, 1)$, $f(x_0) = 909$

**[Answer]**

$$x^* = (0.5, 0.5\sqrt{3})$$

$$f(x^*) = 81.5 - 25\sqrt{3}$$

**Problem 22**

$$\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$$

$$s.t.$$

$$-x_1 - x_2 + 2 \geq 0$$

$$x_1^2 + x_2 \geq 0$$

**Start with** $x_0 = (2,2)$, $f(x_0) = 1$

**[Answer]**

$$x^* = (1,1)$$

$$f(x^*) = 1$$

**Problem 26**

$$\min f(x) = (x_1 - x_2)^2 + (x_2 - x_3)^2$$

$$s.t.$$

$$(1 + x_2^2)x_1 + x_3^4 - 3 = 0$$

**Start with** $x_0 = (2.6,2,2)$, $f(x_0) = 21.16$

**[Answer]**

$$x^* = (1,1,1),(a,a,a)$$

$$f(x^*) = 0$$

**Problem 27**

$$\min\ f(x) = 0.01(x_1 - 1)^2 + (x_2 - x_1^2)^2$$

$$s.t.$$

$$x_1 + x_3^2 + 1 = 0$$

**Start with** $x_0 = (2,2,2)$, $f(x_0) = 4.01$

**[Answer]**

$$x^* = (-1,1,0)$$

$$f(x^*) = 0.04$$

---

**Problem 28**

$$\min\ f(x) = (x_1 + x_2)^2 + (x_2 + x_3)^2$$

$$s.t.$$

$$x_1 + 2x_2 + 3x_3 - 1 = 0$$

**Start with** $x_0 = (-4,1,1)$, $f(x_0) = 13$

**[Answer]**

$$x^* = (0.5,-0.5,0.5)$$

$$f(x^*) = 0$$

---

**Problem 29**

$$\min\ f(x) = -x_1 x_2 x_3$$

$$s.t.$$

$$-x_1^2 - 2x_2^2 - 4x_3^2 + 48 \geq 0$$

**Start with** $x_0 = (1,1,1)$, $f(x_0) = -1$

**[Answer]**

$$x^* = (a,b,c),(a,-b,-c),(-a,b,-c) \leftarrow a = 4, b = 2\sqrt{2}, c = 2$$

$$f(x^*) = -16\sqrt{2}$$

**Problem 32**

$$\min\ f(x) = (x_1 + 3x_2 + x_3)^2 + 4(x_1 - x_2)^2$$

$$s.t.$$

$$6x_2 + 4x_3 - x_1^2 - 3 \geq 0$$

$$1 - x_1 - x_2 - x_3 = 0$$

$$0 \leq x_1$$

$$0 \leq x_2$$

$$0 \leq x_3$$

**Start with** $x_0 = (0.1, 0.7, 0.2)$, $f(x_0) = 7.2$

**[Answer]**

$$x^* = (0,0,1)$$

$$f(x^*) = 1$$

**Problem 32**

$$\min\ f(x) = -x_1$$

$$s.t.$$

$$x_2 - x_1^3 - x_3^2 = 0$$

$$x_1^2 - x_2 - x_4^2 = 0$$

**Start with** $x_0 = (2,2,2,2)$, $f(x_0) = -2$

**[Answer]**

$$x^* = (1,1,0,0)$$

$$f(x^*) = -1$$

---

**Problem 40**

$$\min\ f(x) = -x_1 x_2 x_3 x_4$$

$$s.t.$$

$$x_1^3 + x_2^2 - 1 = 0$$

$$x_1^2 x_4 - x_3 = 0$$

$$x_4^2 - x_2 = 0$$

**Start with** $x_0 = (0.8,0.8,0.8,0.8)$, $f(x_0) = -0.4096$

**[Answer]**

$$x^* = (2^a, 2^{2b}, (-1)^i 2^c, (-1)^i 2^b) \leftarrow i = 1,2 \quad a = -1/3, b = -1/4, c = -11/12$$

$$f(x^*) = -0.25$$

**Problem 42**

$$\min\ f(x) = (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 3)^2 + (x_4 - 4)^2$$

$s.t.$

$$x_1 - 2 = 0$$

$$x_3^2 + x_4^2 - 2 = 0$$

**Start with** $x_0 = (1,1,1,1)$, $f(x_0) = 14$

**[Answer]**

$$x^* = (2,2,0.6\sqrt{2},0.8\sqrt{2})$$

$$f(x^*) = 28 - 10\sqrt{2}$$

**Problem 43**

$$\min\ f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

$s.t.$

$$8 - x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 \geq 0$$

$$10 - x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 \geq 0$$

$$5 - 2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 \geq 0$$

**Start with** $x_0 = (0,0,0,0)$, $f(x_0) = 0$

**[Answer]**

$$x^* = (0,1,2,-1)$$

$$f(x^*) = -44$$

**Problem 47**

$$\min \; f(x) = (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^2 + (x_4 - x_5)^2$$

$s.t.$

$$x_1 + x_2^2 + x_3^3 - 3 = 0$$

$$x_2 - x_3^2 + x_4 - 1 = 0$$

$$x_1 x_5 - 1 = 0$$

**Start with** $x_0 = (2, \sqrt{2}, -1, 2 - \sqrt{2}, 0.5)$

$f(x_0) = 12.4954368$

**[Answer]**

$$x^* = (1,1,1,1,1)$$

$$f(x^*) = 0$$