Department of Systems Engineering
George Mason University

SYST611: Systems Methodology and Modeling #9

Kuo-Chu Chang
Fairfax, Virginia

SYST611: Outline

- Concepts and Introduction
- Resource Allocation and Optimization
- Unconstrained Optimization
- Constrained Optimization
- Numerical Techniques

Taxonomy of Optimization

- Optimization Techniques
  - Static
    - Unconstrained
    - Constrained
      - Nonlinear
      - Linear
  - Dynamic
    - Optimal Control
      - Continuous
      - Discrete
    - Dynamic Programming
      - Network (Discrete Event)
      - Discrete Time
      - Other

3

Introduction

- **Resource Allocation:** distribute resources to users so that the highest level of a specified objective can be achieved
  - Factory organization for maximum productivity
  - Spacecraft flight plan to minimize fuel consumption
  - Security and safety procedures to minimize risks and hazards
- **Optimization Problems:** to achieve a performance in the extreme within the resource available
  - Given budget resource, optimize performance (maximize reliability, minimize risk)
  - Given specification of performance, minimize the capital required

4

Optimization

- **Problem Formulation:** translate the desired system performance specifications and constraints into a mathematical form
  - Set up objective function (revenue, yield, cost, reliability, etc.)
  - Identify control variables
  - Specify constraints
- **Optimization Techniques:**
  - Unconstrained optimization
  - Constrained optimization
  - Path optimization (optimal control)

Unconstrained Optimization

Problem: Minimize $C(\mathbf{u})$, where $\mathbf{u} = [u_1, \dots, u_m]^T$

second-order Taylor series expansion

$$C(\mathbf{u}) = C(\mathbf{u_0}) + C_{\mathbf{u}}(\mathbf{u_0})(\mathbf{u} - \mathbf{u_0}) + \tfrac{1}{2}(\mathbf{u} - \mathbf{u_0})^T C_{\mathbf{uu}}(\mathbf{u_0})(\mathbf{u} - \mathbf{u_0}) + E(\mathbf{u} - \mathbf{u_0})$$

where

$$C_{\mathbf{u}}(\mathbf{u_0}) \equiv \nabla C(\mathbf{u_0}) = [C_{u_1}(\mathbf{u_0}), C_{u_2}(\mathbf{u_0}), \dots, C_{u_m}(\mathbf{u_0})]$$

$$= \left[\frac{\partial C}{\partial u_1}, \frac{\partial C}{\partial u_2}, \dots, \frac{\partial C}{\partial u_m}\right]_{\mathbf{u}=\mathbf{u_0}} \quad \text{is called } \textit{\textbf{gradient}}$$

$$C_{\mathbf{uu}}(\mathbf{u_0}) = \nabla^2 C(\mathbf{u_0}) = \begin{bmatrix} C_{u_1u_1} & \cdots & C_{u_1u_m} \\ \vdots & \ddots & \vdots \\ C_{u_mu_1} & \cdots & C_{u_mu_m} \end{bmatrix} = \left[\frac{\partial^2 C}{\partial u_i \partial u_j}\right]_{\mathbf{u}=\mathbf{u_0}} \quad \text{is called } \textit{\textbf{Hessian}}$$

Examples

$$\textbf{(I)}\ \mathbf{C}(\mathbf{x}_1, \mathbf{x}_2) = 2\mathbf{x}_1^2 + 2\mathbf{x}_1\mathbf{x}_2 + \mathbf{x}_2^2 - 10\mathbf{x}_1 - 10\mathbf{x}_2$$

$$\nabla \mathbf{C} = \begin{bmatrix} \frac{\partial \mathbf{C}}{\partial \mathbf{x}_1} & \frac{\partial \mathbf{C}}{\partial \mathbf{x}_2} \end{bmatrix} = \begin{bmatrix} 4\mathbf{x}_1 + 2\mathbf{x}_2 - 10 & 2\mathbf{x}_1 + 2\mathbf{x}_2 - 10 \end{bmatrix}$$

$$\nabla^2 \mathbf{C} = \begin{bmatrix} \frac{\partial^2 \mathbf{C}}{\partial^2 \mathbf{x}_1} & \frac{\partial^2 \mathbf{C}}{\partial \mathbf{x}_1 \partial \mathbf{x}_2} \\ \frac{\partial^2 \mathbf{C}}{\partial \mathbf{x}_1 \partial \mathbf{x}_2} & \frac{\partial^2 \mathbf{C}}{\partial^2 \mathbf{x}_2} \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

$$\textbf{(II)}\ \mathbf{C}(\mathbf{x}_1, \mathbf{x}_2) = 2\mathbf{x}_1^2 - 6\mathbf{x}_1 + \mathbf{x}_2^2 - 2\mathbf{x}_1\mathbf{x}_2$$

$$\nabla \mathbf{C} = \begin{bmatrix} 4\mathbf{x}_1 - 6 - 2\mathbf{x}_2 & 2\mathbf{x}_2 - 2\mathbf{x}_1 \end{bmatrix}$$

$$\nabla^2 \mathbf{C} = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

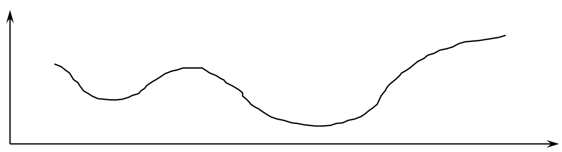
Local and Global Minimum

Local Minimum:

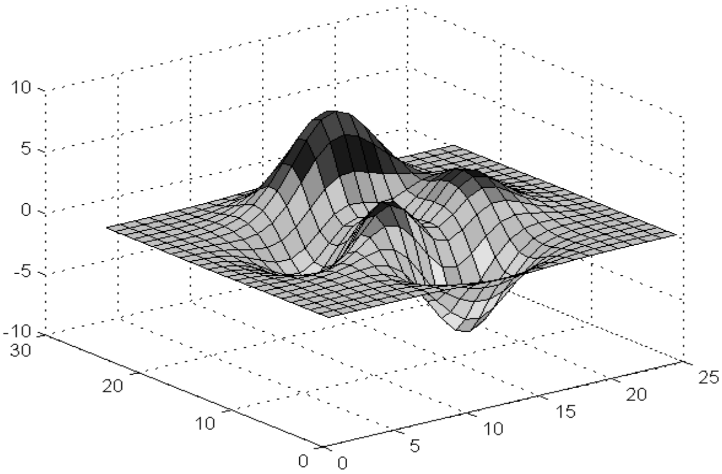
A point u_0 is said to be a relative or local minimum if there exists an $\varepsilon > 0$, such that $C(u) \geq C(u_0)$ for all u, such that $|u - u_0| < \varepsilon$ and $u \in \Omega$. This definition says that if we can draw an arbitrarily small box around our point u_0 and no other values of C(u) within that small box have a lower value than $C(u_0)$, then we have a local minimum.

Global Minimum:

A point u^* is said to be a global minimum if $C(u) \geq C(u^*)$ for all $u \in \Omega$.



Example



Necessary and Sufficient Conditions

Local Minimum:

Necessary conditions:

first - order $\nabla C(\mathbf{x_0}) = \mathbf{0}$ ($\mathbf{x_0}$:critical point)

second - order $C_{\mathbf{uu}}(\mathbf{x_0}) \equiv Q$ is positive semidefinite $(\mathrm{x}^T Q\mathrm{x} \geq 0, \forall \mathrm{x} \neq 0)$

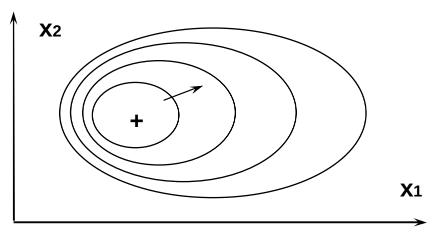
Sufficient conditions:

first - order $\nabla C(\mathbf{x_0}) = \mathbf{0}$

second - order $C_{\mathbf{uu}}(\mathbf{x_0}) \equiv Q$ is positive definite $(\mathrm{x}^T Q\mathrm{x} > 0, \forall \mathrm{x} \neq 0)$

Global Minimum:

compare all local minimums



Example

$$\min\ \mathbf{C}(\mathbf{x}_1,\mathbf{x}_2) = 2\mathbf{x}_1^2 - 6\mathbf{x}_1 + \mathbf{x}_2^2 - 2\mathbf{x}_1\mathbf{x}_2$$

$$\nabla\mathbf{C} = [4\mathbf{x}_1 - 6 - 2\mathbf{x}_2,\quad 2\mathbf{x}_2 - 2\mathbf{x}_1] = [0\quad 0]$$

$$\Rightarrow \mathbf{x}_1 = \mathbf{x}_2 = 3 \quad (\textbf{critical point})$$

$$\mathbf{Q} = \nabla^2\mathbf{C} = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix},\ \forall \mathbf{x}_1,\mathbf{x}_2 \neq 0$$

$$\Rightarrow [\mathbf{x}_1\quad \mathbf{x}_2]\mathbf{Q}\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix} = (2\mathbf{x}_1 - \mathbf{x}_2)^2 + \mathbf{x}_2^2 > 0$$

$$\Rightarrow \mathbf{Q}\ \textbf{is positive definite}$$

$$\Rightarrow \mathbf{x}_1 = \mathbf{x}_2 = 3\ \textbf{is a local minimum}$$

Constrained Optimization

Equality Constraints

Minimize $f(\mathbf{x})$

subject to $\mathbf{h(x)} = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{bmatrix} = \mathbf{0}$

Inequality Constraints

Minimize $f(\mathbf{x})$

subject to $\mathbf{h(x)} = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{bmatrix} = \mathbf{0}$ and $\mathbf{g(x)} = \begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_p(\mathbf{x}) \end{bmatrix} \leq \mathbf{0}$

Lagrange Multipliers

Equality Constraints :

Minimize $f(\mathbf{x})$

subject to $\mathbf{h(x) = 0}$

Lagrangian : $l(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda \mathbf{h}(\mathbf{x})$

Necessary conditions for local minimum

first - order : $\nabla_{\mathbf{x}} l(\mathbf{x}, \lambda) = \nabla_{\mathbf{x}} f(\mathbf{x}) + \lambda \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}) = \mathbf{0}$

$\nabla_{\lambda} l(\mathbf{x}, \lambda) = \mathbf{h}(\mathbf{x}) = \mathbf{0}$

second - order :

$\nabla^2_{\mathbf{x}} l(\mathbf{x}, \lambda)\big|_{\mathbf{M}} \equiv \mathbf{L}(\mathbf{x})\big|_{\mathbf{M}}$ is positive semidefinite

Sufficient conditions for local minimum

first - order : same as above

second - order : $\nabla^2_{\mathbf{x}} l(\mathbf{x}, \lambda)\big|_{\mathbf{M}} \equiv \mathbf{L}(\mathbf{x})\big|_{\mathbf{M}}$ is positive definite

where $\mathrm{M} = \{\mathbf{y}: \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}\}$

Example

Minimize $-(x_1 x_2 + x_2 x_3 + x_3 x_1)$

subject to $x_1 + x_2 + x_3 = 3$

Lagrangian: $l(\mathbf{x}, \lambda) = -(x_1 x_2 + x_2 x_3 + x_3 x_1) + \lambda(x_1 + x_2 + x_3 - 3)$

conditions

first - order: $\nabla_{\mathbf{x}} l(\mathbf{x}, \lambda) = \nabla_{\lambda} l(\mathbf{x}, \lambda) = \mathbf{0} \Rightarrow x_1 = x_2 = x_3 = 1,\ \lambda = 2$

second - order:

$$\mathbf{L}(\mathbf{x}) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix},\ \mathrm{M} = \{\mathbf{y}: [\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x})]\mathbf{y} = \mathbf{0}\} = \{\mathbf{y}: y_1 + y_2 + y_3 = 0\}$$

$\Rightarrow \mathbf{y}^{\mathbf{T}} \mathbf{L} \mathbf{y} = y_1(-y_2 - y_3) + y_2(-y_1 - y_3) + y_3(-y_1 - y_2) = y_1^2 + y_2^2 + y_3^2 > 0$

$\Rightarrow$ solution is at least a local minimum

Inequality Constraints

Inequality Constraints:

Minimize $f(\mathbf{x})$

subject to $\mathbf{h(x) = 0}, \ \ \mathbf{g(x) \le 0}$

Lagrangian: $l(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda \mathbf{h(x)} + \mu \mathbf{g(x)}$

Necessary conditions:

First-order (***Kuhn-Tucker*** Conditions): if $\mathbf{x_0}$ is a local minimum, then $\exists$ a vector λ and a vector $\mu \ge \mathbf{0}$ such that

$$\nabla_{\mathbf{x}} f(\mathbf{x_0}) + \lambda \nabla_{\mathbf{x}} \mathbf{h(x_0)} + \mu \nabla_{\mathbf{x}} \mathbf{g(x_0)} = \mathbf{0}$$
$$\mu \mathbf{g(x_0)} = \mathbf{0}$$

second-order:

$\nabla^2_{\mathbf{x}} l(\mathbf{x}, \lambda, \mu) \equiv \mathbf{L(x_0)}$ is positive semidefinite on the tangent subspace of the active constraints at $\mathbf{x_0}$

Sufficient conditions for local minimum

first-order: same as above

second-order: $\nabla^2_{\mathbf{x}} l(\mathbf{x}, \lambda, \mu) \equiv \mathbf{L(x_0)}$ is positive definite

Example

Minimize $2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2$

subject to $x_1^2 + x_2^2 \le 5, \ 3x_1 + x_2 \le 6$

first-order: $4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0$

$$2x_1 + 2x_2 - 10 + 2\mu_1 x_2 + \mu_2 = 0$$
$$\mu_1 \ge 0, \ \mu_2 \ge 0$$
$$\mu_1(x_1^2 + x_2^2 - 5) = 0$$
$$\mu_2(3x_1 + x_2 - 6) = 0$$

assuming first constraint is active

$\Rightarrow x_1 = 1, x_2 = 2, \mu_1 = 1, \mu_2 = 0$

Lagrangian Relaxation

Separable Cost Function and Constraints :

$$\text{Minimize } f(\mathbf{x}) = \sum_i f(x_i)$$

$$\text{subject to } \mathbf{h}(\mathbf{x}) = \sum_i h(x_i) < 0$$

Lagrangian : $l(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda \mathbf{h}(\mathbf{x})$

$$\min_{\mathbf{x}} l(\mathbf{x}, \lambda) = \min_{\mathbf{x}} \sum_i \left[f(x_i) + \lambda h(x_i) \right]$$

$$= \sum_i \min_{x_i} \left[f(x_i) + \lambda h(x_i) \right]$$

(1) The decomposed optimization problems become trivial
(2) Search over the λ space instead of the x space, it is much simpler

Numerical Optimization Techniques

- **Motivation: analytical solution may not be feasible**
  - **Cost function may not be differentiable**
  - **Evaluating the expression for the derivative may be difficult**
- **Algorithm Properties:**
  - ***Iterative* in nature**
  - ***Descending* in sequence**
  - ***Globally convergent* is desirable but often not achievable**
- **Numerical Techniques:**
  - **Search methods**
  - **Gradient methods**

Search Methods

Constructing shrinking bounding intervals

Fibonacci, Golden section

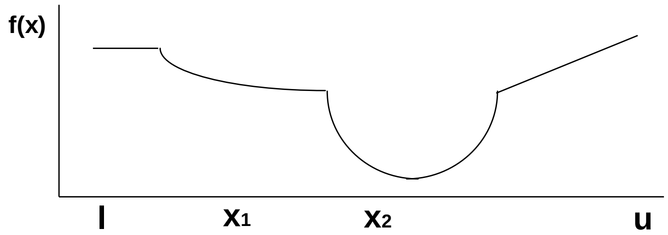
Often improved by curve - fitting line search

Basic property of unimodal function:

(1) if $f(x_1) > f(x_2) \Rightarrow x^* \in [x_1, \mathrm{u}]$

(2) if $f(x_1) < f(x_2) \Rightarrow x^* \in [\mathrm{l}, x_2]$

(3) if $f(x_1) = f(x_2) \Rightarrow x^* \in [x_1, x_2]$



Gradient Methods

- **Steepest Descent: oldest and most widely known method for minimizing multi-variable functions**
  - **Simple and not difficult to analyze**
  - **Many modifications on the basic technique exist for better convergence properties**
- **The Method:**

To minimize $f(\mathbf{x})$

$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{g}_k$ where $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$

and α_k is a nonnegative scalar minimizing $f(\mathbf{x}_k - \alpha_k \mathbf{g}_k)$

The Quadratic Case

Consider $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathbf{T}} Q\mathbf{x} - \mathbf{x}^{\mathbf{T}}\mathbf{b}$

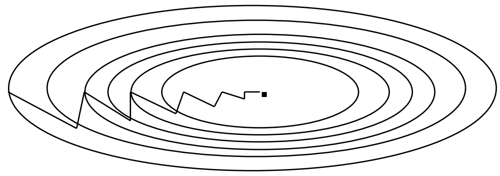
$\Rightarrow \mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = Q\mathbf{x} - \mathbf{b}$

but $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k - \alpha_k \mathbf{g}_k)$

$= \frac{1}{2}(\mathbf{x}_k - \alpha_k \mathbf{g}_k)^T Q(\mathbf{x}_k - \alpha_k \mathbf{g}_k) - (\mathbf{x}_k - \alpha_k \mathbf{g}_k)^T \mathbf{b}$

which is minimized at $\alpha_k = \dfrac{\mathbf{g}_k{}^{\mathbf{T}}\mathbf{g}_k}{\mathbf{g}_k{}^{\mathbf{T}} Q\mathbf{g}_k}$

$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - \left(\dfrac{\mathbf{g}_k{}^{\mathbf{T}}\mathbf{g}_k}{\mathbf{g}_k{}^{\mathbf{T}} Q\mathbf{g}_k}\right)\mathbf{g}_k$ where $\mathbf{g}_{\mathbf{k}} = Q\mathbf{x}_{\mathbf{k}} - \mathbf{b}$



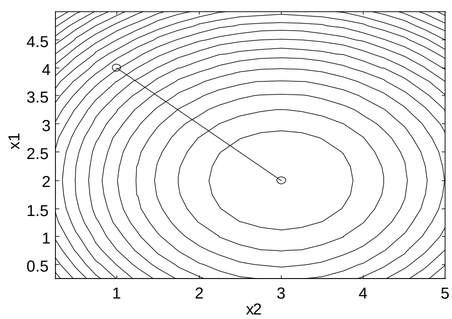
Example I

Ex: $f(x_1, x_2) = x_1^2 - 4x_1 + x_2^2 - 6x_2 + 30;\quad x_1(0) = 4, x_2(0) = 1$

$\nabla f(\mathbf{x(0)}) = \begin{bmatrix} 2x_1 - 4 & 2x_2 - 6 \end{bmatrix}_{\mathbf{x(0)}} = \begin{bmatrix} 4 & -4 \end{bmatrix},\ x_1(1) = x_1(0) - 4\alpha, x_2(1) = x_2(0) + 4\alpha$

$f(\mathbf{x(1)}) = (4-4\alpha)^2 - 4(4-4\alpha) + (1+4\alpha)^2 - 6(1+4\alpha) + 30 = 32\alpha^2 - 32\alpha + 25 \equiv h(\alpha)$

$\min h(\alpha) \Rightarrow \alpha = 0.5 \Rightarrow x_1(1) = 2, x_2(1) = 3$



Example II

$Ex: f(x_1, x_2) = 2x_1^2 - 6x_1 + x_2^2 - 2x_1x_2; x_1(0) = 0, x_2(0) = 0$

$\nabla f(\mathbf{x(0)}) = [4x_1 - 6 - 2x_2 \quad 2x_2 - 2x_1]_{\mathbf{x(0)}} = [-6 \quad 0], x_1(1) = x_1(0) + 6\boldsymbol{\alpha}, x_2(1) = x_2(0) + 0\boldsymbol{\alpha}$

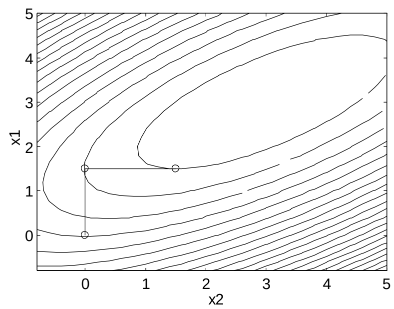
$f(\mathbf{x(1)}) = 2(6\boldsymbol{\alpha})^2 - 6(6\boldsymbol{\alpha}) + (0)^2 - 2(6\boldsymbol{\alpha})(0) = 72\boldsymbol{\alpha}^2 - 36\boldsymbol{\alpha} \equiv h(\boldsymbol{\alpha})$

$\min h(\boldsymbol{\alpha}) \Rightarrow \boldsymbol{\alpha} = 0.25 \Rightarrow x_1(1) = 1.5, x_2(1) = 0$

$\nabla f(\mathbf{x(1)}) = [4x_1 - 6 - 2x_2 \quad 2x_2 - 2x_1]_{\mathbf{x(1)}} = [0 \quad -3], x_1(2) = x_1(1) + 0\boldsymbol{\alpha}, x_2(2) = x_2(1) + 3\boldsymbol{\alpha}$

$f(\mathbf{x(1)}) = 2(1.5)^2 - 6(1.5) + (3\boldsymbol{\alpha})^2 - 2(1.5)(3\boldsymbol{\alpha}) = 9\boldsymbol{\alpha}^2 - 9\boldsymbol{\alpha} - 4.5 \equiv h(\boldsymbol{\alpha})$

$\min h(\boldsymbol{\alpha}) \Rightarrow \boldsymbol{\alpha} = 0.5 \Rightarrow x_1(1) = 1.5, x_2(1) = 1.5$



Newton's Method

Approximate the function locally by a quadratic function

$$f(\mathbf{x}) \cong f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) + \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_k)' \mathbf{F}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$$

the right-hand side is minimized at

$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{F}(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k) \Leftarrow$ Newton's Method

For real quadratic functions $f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}' Q\mathbf{x} - \mathbf{x}'\mathbf{b}$

$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - Q^{-1}(Q\mathbf{x}_k - \mathbf{b}) = Q^{-1}\mathbf{b}$

reach solution in **one iteration**

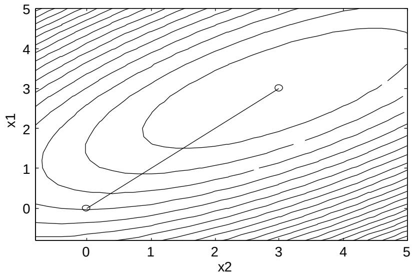
Example

$$Ex: f(x_1, x_2) = 2x_1^2 - 6x_1 + x_2^2 - 2x_1x_2; x_1(0) = 0, x_2(0) = 0$$

$$\nabla f(\mathbf{x(0)}) = \left[4x_1 - 6 - 2x_2 \quad 2x_2 - 2x_1\right]_{\mathbf{x(0)}} = \left[-6 \quad 0\right], \nabla^2 f(\mathbf{x(0)}) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

$$\left[\nabla^2 f(\mathbf{x(0)})\right]^{-1} \nabla f(\mathbf{x(0)})^T = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} -6 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$$

$$x_1(1) = x_1(0) - (-3) = 3, x_2(1) = x_2(0) - (-3) = 3$$



References

1. Bradley W. Dickinson ***Systems- Analysis, Design, and Computation***, **Chap. 5, Prentice Hall, 1991.**

2. Boardman, J. ***Systems Engineering, An introduction***, **Prentice Hall, 1990**

3. David G. Luenberger ***Introduction to Linear and Nonlinear Programming,*** **MA: Addison-Wesley, 1973.**

4. **Optimization software guide: http://www-fp.mcs.anl.gov/otc/Guide/SoftwareGuide/index.html**