Constrained Optimization

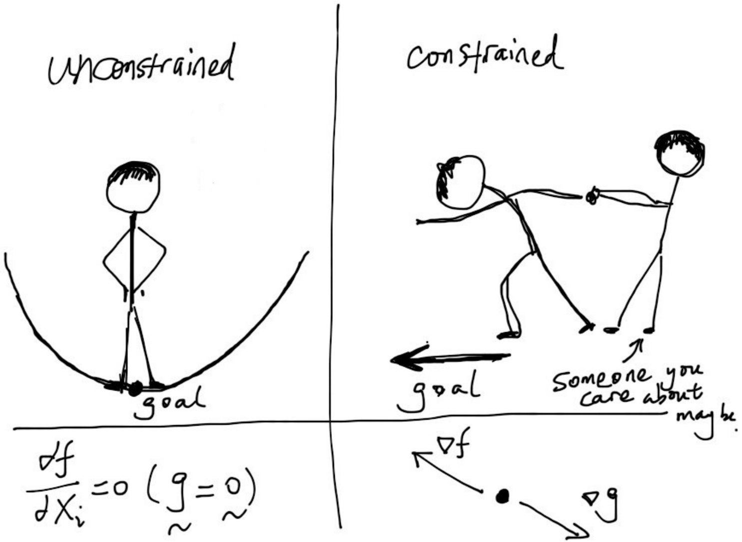
ME598/494 Lecture

Max Yi Ren

Department of Mechanical Engineering, Arizona State University

March 30, 2015

From unconstrained to constrained



Optimization with equality constraints (1/3)

A general optimization problem with only equality constraints is the following:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_j(\mathbf{x}) = 0, \quad j = 1, 2, ..., m. \end{aligned}$$

Let there be n variables and m equality constraints. In an ideal case, we can eliminate m variables by using the equalities and solve an unconstrained problem for the rest $n - m$ variables. However, such elimination may not be feasible in practice.

Consider taking a feasible perturbation from a feasible point $\mathbf{x}$, the perturbation $\partial\mathbf{x}$ needs to be such that the equality constraints are still satisfied. Mathematically, it requires the first-order approximations of the perturbations for constraints to be:

$$\partial h_j = \sum_{i=1}^{n} (\partial h_j / \partial x_i) \partial x_i = 0, \quad j = 1, 2, ...m. \tag{1}$$

Optimization with equality constraints (2/3)

Equation (1) contains a system of linear equations with $n - m$ degrees of freedom. Let us define *state variables* as:

$$s_i := x_i, \quad i = 1, ..., m,$$

and *decision variables* as:

$$d_i := x_i, \quad i = m + 1, ..., n.$$

The number of decision variables is equal to the number of degrees of freedom. Equation (1) can be rewritten as:

$$(\partial \mathbf{h}/\partial \mathbf{s})\partial \mathbf{s} = -(\partial \mathbf{h}/\partial \mathbf{d})\partial \mathbf{d}, \tag{2}$$

where the matrix $(\partial \mathbf{h}/\partial \mathbf{s})$ is

$$\begin{bmatrix} \partial h_1/\partial s_1 & \partial h_1/\partial s_2 & \cdots & \partial h_1/\partial s_m \\ \partial h_2/\partial s_1 & \partial h_2/\partial s_2 & \cdots & \partial h_2/\partial s_m \\ \vdots & \vdots & \ddots & \vdots \\ \partial h_m/\partial s_1 & \partial h_m/\partial s_2 & \cdots & \partial h_m/\partial s_m \end{bmatrix},$$

which is the Jacobian matrix with respect to the state variables. $\partial \mathbf{h}/\partial \mathbf{d}$ is then the Jacobian with respect to the decision variables.

Optimization with equality constraints (3/3)

From Equation (2), we can further have

$$\partial \mathbf{s} = -(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d})\partial \mathbf{d}. \qquad (3)$$

Equation (3) shows that for some perturbation for the decision variables, we can derive the corresponding perturbation for the state variables so that $\partial h(\mathbf{x}) = 0$ for first-order approximation. Notice that Equation (3) can only be derived when the Jacobian $\partial \mathbf{h}/\partial \mathbf{s}$ is invertible, i.e., the gradients of equality constraints must be linearly independent.

Since $\mathbf{s}$ can be considered as functions of $\mathbf{d}$, the original constrained optimization problem can be treated as an unconstrained problem for minimizing

$$\min_{\mathbf{d}} \quad f(\mathbf{x}) := z(\mathbf{s}(\mathbf{d}), \mathbf{d}).$$

The gradient of this new objective function is

$$\partial z/\partial \mathbf{d} = (\partial f/\partial \mathbf{d}) + (\partial f/\partial \mathbf{s})(\partial \mathbf{s}/\partial \mathbf{d}).$$

Plug in Equation (3) to have

$$\partial z/\partial \mathbf{d} = (\partial f/\partial \mathbf{d}) - (\partial f/\partial \mathbf{s})(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d}). \qquad (4)$$

Lagrange multiplier

From Equation (4), a stationary point $\mathbf{x}_* = (\mathbf{s}_*, \mathbf{d}_*)^T$ will then satisfy

$$(\partial f/\partial \mathbf{d}) - (\partial f/\partial \mathbf{s})(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d}) = \mathbf{0}^T, \tag{5}$$

evaluated at $\mathbf{x}_*$. Equation (5) and $\mathbf{h} = \mathbf{0}$ together have n equalities and n variables. The stationary point $\mathbf{x}_*$ can be found when $\partial \mathbf{h}/\partial \mathbf{s}$ is invertible for some choice of $\mathbf{s}$.

Now introduce the *Lagrange multiplier* as

$$\boldsymbol{\lambda}^T := -(\partial f/\partial \mathbf{s})(\partial \mathbf{h}/\partial \mathbf{s})^{-1}. \tag{6}$$

From Equations (5) and (6), we can have

$$(\partial f/\partial \mathbf{d}) + \boldsymbol{\lambda}^T(\partial \mathbf{h}/\partial \mathbf{d}) = \mathbf{0}^T \quad \text{and} \quad (\partial f/\partial \mathbf{s}) + \boldsymbol{\lambda}^T(\partial \mathbf{h}/\partial \mathbf{s}) = \mathbf{0}^T.$$

Recall that $\mathbf{x} = (\mathbf{s}, \mathbf{d})^T$, then for a stationary point we have

$$(\partial f/\partial \mathbf{x}) + \boldsymbol{\lambda}^T(\partial \mathbf{h}/\partial \mathbf{x}) = \mathbf{0}. \tag{7}$$

Lagrangian function

Introduce the *Lagrangian* function for the original optimization problem with equality constraints:

$$L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}).$$

First-order necessary condition: $\mathbf{x}_*$ is a (constrained) stationary point if $\partial L/\partial \mathbf{x} = \mathbf{0}$ and $\partial L/\partial \boldsymbol{\lambda} = \mathbf{0}$.

This condition leads to Equation (7) and $\mathbf{h} = \mathbf{0}$, which in total has $m + n$ variables ($\mathbf{x}$ and $\boldsymbol{\lambda}$) and $m + n$ equalities. The stationary point solved using the Lagrangian function will be the same as that from the reduced gradient method in Equation (5).

Define the Hessian of the Lagrangian with respect to $\mathbf{x}$ as $L_{\mathbf{xx}}$.

Second-order sufficiency condition: If $\mathbf{x}_*$ together with some $\boldsymbol{\lambda}$ satisfies $\partial L/\partial \mathbf{x} = \mathbf{0}$ and $\mathbf{h} = \mathbf{0}$, and $\partial \mathbf{x}_*^T L_{\mathbf{xx}} \partial \mathbf{x}_* > 0$ for any $\partial \mathbf{x}_* \neq \mathbf{0}$ that satisfies $\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \partial \mathbf{x}_* = \mathbf{0}$, then $\mathbf{x}_*$ is a local (constrained) minimum.

Examples (1/4)

Exercise 5.2 For the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & (x_1 - 2)^2 + (x_2 - 2)^2 \\ \text{subject to} \quad & x_1^2 + x_2^2 - 1 = 0, \end{aligned}$$

find the optimal solution using constrained derivatives (reduced gradient) and Lagrange multipliers.

Exercise 5.3 For the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1^2 + x_2^2 - x_3^2 \\ \text{subject to} \quad & 5x_1^2 + 4x_2^2 + x_3^2 - 20 = 0, \\ & x_1 + x_2 - x_3 = 0, \end{aligned}$$

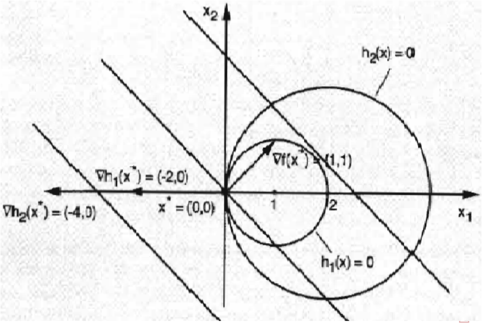
find the optimal solution using constrained derivatives and Lagrange multipliers.

Examples (2/4)

(A problem where Lagrangian multipliers cannot be found) For the problem

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1 + x_2 \\ \text{subject to} \quad & (x_1 - 1)^2 + x_2^2 - 1 = 0, \\ & (x_1 - 2)^2 + x_2^2 - 4 = 0, \end{aligned}$$

find the optimal solution and Lagrange multipliers. (Source: Fig. 3.1.2 D.P. Bertsekas, Nonlinear Programming)



Examples (3/4)

Important: In all development of theory hereafter, we assume that stationary points are *regular*. We will discuss in more details the regularity condition in the next section on KKT.

(A problem where Lagrangian multipliers are zeros) For the problem

$$\begin{aligned} \min_{x} \quad & x^2 \\ \text{subject to} \quad & x = 0, \end{aligned}$$

find the optimal solution and Lagrange multipliers.

Important: In all development of theory hereafter, we assume that all equality constraints are active.

Examples (4/4)

Example 5.6 Consider the problem with $x_i > 0$:

$$\begin{aligned} \min_{x_1, x_2, x_3} \quad & x_1^2 x_2 + x_2^2 x_3 + x_1 x_3^2 \\ \text{subject to} \quad & x_1^2 + x_2^2 + x_3^2 - 3 = 0. \end{aligned}$$

find the optimal solution using Lagrangian.

With inequality constraints

Let us now look at the constrained optimization problem with both equality and inequality constraints

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq 0, \qquad \mathbf{h}(\mathbf{x}) = 0. \end{aligned}$$

Denote $\hat{\mathbf{g}}$ as a set of inequality constraints that are active at a stationary point. Then following the discussion on the optimality conditions for problems with equality constraints, we have

$$(\partial f/\partial \mathbf{x}) + \boldsymbol{\lambda}^T(\partial \mathbf{h}/\partial \mathbf{x}) + \hat{\boldsymbol{\mu}}^T(\partial \hat{\mathbf{g}}/\partial \mathbf{x}) = \mathbf{0}^T, \tag{8}$$

where $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\mu}}$ are Lagrangian multipliers on $\mathbf{h}$ and $\hat{\mathbf{g}}$.

Nonnegative Lagrange multiplier

The Lagrange multipliers (at the local minimum) for inequality constraints $\boldsymbol{\mu}$ are nonnegative. This can be shown by examining the first-order perturbations for f, $\mathbf{g}$ and $\mathbf{h}$ at a local minimum for feasible nonzero perturbations $\partial\mathbf{x}$:

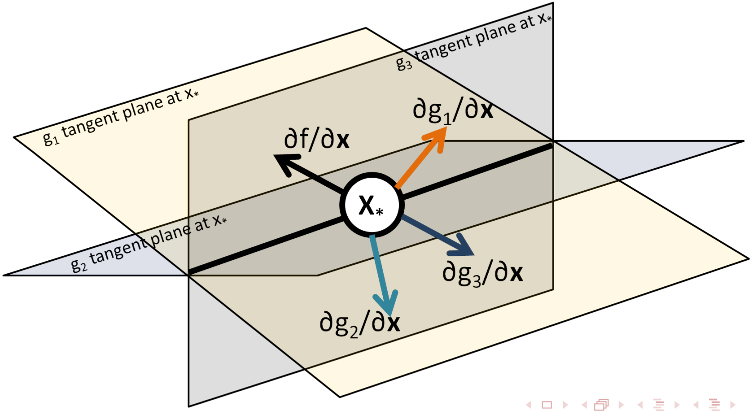
$$\frac{\partial f}{\partial \mathbf{x}}\partial\mathbf{x} \geq 0, \qquad \frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{x}}\partial\mathbf{x} \leq \mathbf{0}, \qquad \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\partial\mathbf{x} = \mathbf{0}. \tag{9}$$

Combining Equations (8) and (9) we get $\hat{\boldsymbol{\mu}}^T\partial\hat{\mathbf{g}} \leq 0$. Since $\partial\hat{\mathbf{g}} \leq \mathbf{0}$ for feasibility, we have $\hat{\boldsymbol{\mu}} \geq \mathbf{0}$.

Regularity

A *Regular* point $\mathbf{x}$ is such that the active inequality constraints and all equality constraints are linearly independent, i.e., $((\partial\hat{\mathbf{g}}/\partial\mathbf{x})^T, (\partial\mathbf{h}/\partial\mathbf{x}^T))$ should have independent columns.

Active constraints with zero multipliers are possible when $\mathbf{x}_*$ is not a regular point. This situation is usually referred to as *degeneracy*.



The Karush-Kuhn-Tucker (KKT) conditions

For the optimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq 0, \qquad \mathbf{h}(\mathbf{x}) = 0, \end{aligned}$$

its optimal solution $\mathbf{x}_*$ (*assumed to be regular*) must satisfy

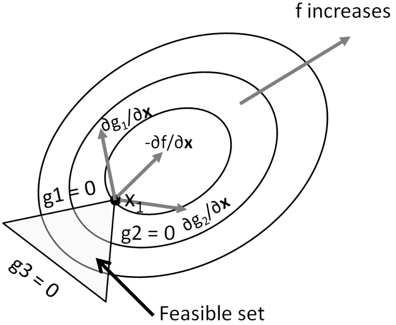
$$\begin{aligned} & \mathbf{g}(\mathbf{x}_*) \leq 0; \\ & \mathbf{h}(\mathbf{x}_*) = 0; \\ & (\partial f/\partial \mathbf{x}_*) + \boldsymbol{\lambda}^T(\partial \mathbf{h}/\partial \mathbf{x}_*) + \boldsymbol{\mu}^T(\partial \mathbf{g}/\partial \mathbf{x}_*) = \mathbf{0}^T, \\ & \text{where } \boldsymbol{\lambda} \neq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\mu}^T \mathbf{g} = 0. \end{aligned} \tag{10}$$

A point that satisfies the KKT conditions is called a *KKT point* and may not be a minimum since the conditions are not sufficient.

Second-order sufficiency conditions: If a KKT point $\mathbf{x}_*$ exists, such that the Hessian of the Lagrangian on feasible perturbations is positive-definite, i.e., $\partial\mathbf{x}^T L_{\mathbf{xx}} \partial\mathbf{x} > 0$ for any nonzero $\partial\mathbf{x}_*$ that satisfies $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\partial\mathbf{x} = \mathbf{0}$ and $\frac{\partial \hat{\mathbf{g}}}{\partial \mathbf{x}}\partial\mathbf{x} = \mathbf{0}$, then $\mathbf{x}_*$ is a local constrained minimum.

Geometry interpretation of KKT conditions

The KKT conditions (necessary) state that $-\partial f/\partial \mathbf{x}_*$ should belong to the cone spanned by the gradients of the active constraints at $\mathbf{x}_*$.



The second-order sufficiency conditions require both the objective functon and the feasible space be locally convex at the solution. Further, *if a KKT point exists for a convex function subject to a convex constraint set, then this point is a unique global minimizer.*

Example

Example 5.10: Solve the following problem using KKT conditions

$$\begin{aligned} \min_{x_1,x_2} \quad & 8x_1^2 - 8x_1x_2 + 3x_2^2 \\ \text{subject to} \quad & x_1 - 4x_2 + 3 \leq 0, \\ & -x_1 + 2x_2 \leq 0. \end{aligned}$$

Example with irregular solution: Solve the following problem

$$\begin{aligned} \min_{x_1,x_2} \quad & -x_1 \\ \text{subject to} \quad & x_2 - (1 - x_1)^3 \leq 0, \\ & -x_1 \leq 0, \\ & -x_2 \leq 0. \end{aligned}$$

Sensitivity analysis (1/2)

Consider the constrained problem with local minimum $\mathbf{x}_*$ and $\mathbf{h}(\mathbf{x}_*) = \mathbf{0}$ being the set of equality constraints and active inequality constraints. What will happen to the optimal objective value $f(\mathbf{x}_*)$ when we make a small perturbation $\partial\mathbf{h}$, e.g., slightly relax (restrain) the constraints?

Use the partition $\partial\mathbf{x} = (\partial\mathbf{d}, \partial\mathbf{s})^T$. We have

$$\partial\mathbf{h} = (\partial\mathbf{h}/\partial\mathbf{d})\partial\mathbf{d} + (\partial\mathbf{h}/\partial\mathbf{s})\partial\mathbf{s}.$$

Assuming $\mathbf{x}_*$ is regular thus $(\partial\mathbf{h}/\partial\mathbf{s})^{-1}$ exists, we further have

$$\partial\mathbf{s} = \left(\frac{\partial\mathbf{h}}{\partial\mathbf{s}}\right)^{-1}\partial\mathbf{h} - \left(\frac{\partial\mathbf{h}}{\partial\mathbf{s}}\right)^{-1}\left(\frac{\partial\mathbf{h}}{\partial\mathbf{d}}\right)\partial\mathbf{d}. \tag{11}$$

Recall that the perturbation of the objective function is

$$\partial f = \left(\frac{\partial f}{\partial\mathbf{d}}\right)\partial\mathbf{d} + \left(\frac{\partial f}{\partial\mathbf{s}}\right)\partial\mathbf{s}. \tag{12}$$

Sensitivity analysis (2/2)

Use Equation (11) in Equation (12) to have

$$\partial f = \left(\frac{\partial f}{\partial \mathbf{s}}\right)\left(\frac{\partial \mathbf{h}}{\partial \mathbf{s}}\right)^{-1}\partial \mathbf{h} + \left(\frac{\partial z}{\partial \mathbf{d}}\right)\partial \mathbf{d}. \tag{13}$$

Notice that the reduced gradient $(\partial z/\partial \mathbf{d})$ is zero at $\mathbf{x}_*$. Therefore

$$\begin{aligned} \partial f(\mathbf{x}_*) &= \left(\frac{\partial f}{\partial \mathbf{s}}\right)\left(\frac{\partial \mathbf{h}}{\partial \mathbf{s}}\right)^{-1}\partial \mathbf{h} \\ &= -\boldsymbol{\lambda}^T\partial \mathbf{h}. \end{aligned} \tag{14}$$

To conclude, for a unit perturbation in active (equality and inequality) constraints $\partial \mathbf{h}$, the optimal objective value will be changed by $-\boldsymbol{\lambda}$. Note that the analysis here is based on first-order approximation and is only valid for small changes in constraints.

Generalized reduced gradient (1/2)

We discussed the optimality conditions for constrained problems. Generalized reduced gradient (GRG) is an iterative algorithm to find solutions for $(\partial z/\partial \mathbf{d}) = \mathbf{0}^T$.

Similar to the gradient descent method for unconstrained problems, we update the decision variables by

$$\mathbf{d}_{k+1} = \mathbf{d}_k - \alpha(\partial z/\partial \mathbf{d})_k^T.$$

The corresponding state variables can be found by

$$\begin{aligned}\mathbf{s}'_{k+1} &= \mathbf{s}_k - (\partial \mathbf{h}/\partial \mathbf{s})_k^{-1}(\partial \mathbf{h}/\partial \mathbf{d})_k \partial \mathbf{d}_k \\ &= \mathbf{s}_k + \alpha_k(\partial \mathbf{h}/\partial \mathbf{s})_k^{-1}(\partial \mathbf{h}/\partial \mathbf{d})_k(\partial z/\partial \mathbf{d})_k^T.\end{aligned}$$

Generalized reduced gradient (2/2)

Note that the above calculation is based on the linearization of the constraints and it will not satisfy the constraints exactly unless they are all linear. However, a solution to the nonlinear system

$$\mathbf{h}(\mathbf{d}_{k+1}, \mathbf{s}_{k+1}) = \mathbf{0},$$

given $\mathbf{d}_{k+1}$ can be found iteratively using $\mathbf{s}'_{k+1}$ as an initial guess and the following iteration

$$[\mathbf{s}_{k+1}]_{j+1} = [\mathbf{s}_{k+1} - (\partial \mathbf{h}/\partial \mathbf{s})^{-1}_{k+1} \mathbf{h}(\mathbf{d}_{k+1}, \mathbf{s}_{k+1})]_j.$$

The iteration on the decision variables may also be performed based on Newton's method:

$$\mathbf{d}_{k+1} = \mathbf{d}_k - \alpha(\partial^2 z/\partial \mathbf{d}^2)^{-1}_k (\partial z/\partial \mathbf{d})^T_k.$$

The state variables can also be adjusted by the quadratic approximation

$$\mathbf{s}_{k+1} = \mathbf{s}_k + (\partial \mathbf{s}/\partial \mathbf{d})_k \partial \mathbf{d}_k + (1/2)\partial \mathbf{d}^T_k (\partial^2 \mathbf{s}/\partial \mathbf{d}^2)_k \partial \mathbf{d}_k.$$

The GRG algorithm can be used with the presence of inequality constraints when accompanied by an active set algorithm. This will be discussed in Chapter 7.