Note 3

1 Constrained minimum in vector problems

1.1 Lagrange Multipliers method

Reminding of the technique discussed in calculus, we first consider a finite-dimensional problem of constrained minimum. Namely, we want to find the condition of the minimum:

$$J = \min_x f(x), \quad x \in R^n, \quad f \in C_2(R^n) \tag{1}$$

assuming that m constraints are applied

$$g_i(x_1, \ldots x_n) = 0 \quad i = 1, \ldots m, \quad m \le n, \tag{2}$$

The vector form of the constraints is

$$\boldsymbol{g}(\boldsymbol{x}) = 0$$

where $\boldsymbol{g}$ is a m-dimensional vector-function of an n-dimensional vector $\boldsymbol{x}$.

To find the minimum, we add the constraints with the Lagrange multipliers $\boldsymbol{\mu} = (\mu_1, \ldots \mu_p)$ and end up with the problem

$$J = \min_{\boldsymbol{x}} \left[f(x) + \sum_i^m \mu_i g_i(x) \right]$$

The stationary conditions become:

$$\frac{\partial f}{\partial x_k} + \sum_i^m \mu_i \frac{\partial g_i}{\partial x_k} = 0, \quad k = 1, \ldots, n$$

or, in the vector form

$$\frac{\partial f}{\partial \boldsymbol{x}} + W \cdot \boldsymbol{\mu} = 0 \tag{3}$$

where the $m \times n$ Jacobian matrix W is

$$W = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \quad \text{or, by elements, } W_{nm} = \frac{\partial g_n}{\partial x_m}$$

The system (3) together with the constraints (2) forms a system of $n + p$ equations for $n + p$ unknowns: Components of the vectors $\boldsymbol{x}$ and $\boldsymbol{\mu}$.

Example Consider the problem

$$J = \min_x \sum_i A_i^2 x_i \quad \text{subject to} \quad \sum_i \frac{1}{x_i - k} = \frac{1}{c}.$$

Using Lagrange multiplier λ we rewrite it in the form:

$$J_a = \min_x \sum_i A_i^2 x_i + \lambda \left(\sum_i \frac{1}{x_i - k} - \frac{1}{c} \right).$$

From the condition $\frac{\partial J_a}{\partial \boldsymbol{x}} = 0$ we obtain

$$A_i^2 - \frac{\lambda}{(x_i - k)^2} = 0, \quad \text{or} \quad \frac{1}{x_i - k} = \frac{|A_i|}{\sqrt{\lambda}} \quad i = 1, \ldots, n.$$

We substitute these values into expression for the constraint and obtain an equation for λ

$$\frac{1}{c} = \sum_i \frac{1}{x_i - k} = \frac{1}{\sqrt{\lambda}} \sum_i |A_i|$$

Solving this equation, we find λ, the minimizer x_i

$$\sqrt{\lambda} = c \sum_i |A_i|, \quad x_i = k + \frac{\sqrt{\lambda}}{|A_i|},$$

and the value of the minimizing function J:

$$J = k \sum_i A_i^2 + c \left(\sum_i |A_i| \right)^2$$

Observe, the the minimum is a sum of squares of L_2 and L_1 norms of the vector $A = [A_1, \ldots, A_n]$.

How does it work? (Min-max approach) Consider again the finite-dimensional minimization problem

$$J = \min_{x_1, \ldots x_n} F(x_1, \ldots x_n) \tag{4}$$

subject to one constraint

$$g(x_1, \ldots x_n) = 0 \tag{5}$$

and assume that there exist solutions to (5) in the neighborhood of the minimal point.

It is easy to see that the described constrained problem is equivalent to the unconstrained problem

$$J_* = \min_{x_1, \ldots x_n} \max_\lambda \left(F(x_1, \ldots x_n) + \lambda g(x_1, \ldots x_n) \right) \tag{6}$$

2

Indeed, the inner maximization gives

$$\max_{\lambda} \lambda g(x_1, \ldots x_n) = \begin{cases} \infty & \text{if } g \neq 0 \\ 0 & \text{if } g = 0 \end{cases}$$

because λ can be made arbitrary large or arbitrary small. This possibility forces us to choose such $\boldsymbol{x}$ that delivers equality in (5), otherwise the cost of the problem (6) would be infinite (recall that x "wants" to minimize J_*). By assumption, such x exists. At the other hand, the constrained problem (4)-(5) does not change its cost J if zero $g = 0$ is added to it. Thereby $J = J_*$ and the problem (4) and (5) is equivalent to (6).

If we interchange the sequence of the two extremal operations in (6), we would arrive at the *augmented* problem J_D

$$J_D(\boldsymbol{x}, \lambda) = \max_{\lambda} \min_{x_1, \ldots x_n} (F(x_1, \ldots x_n) + \lambda g(x_1, \ldots x_n)) \tag{7}$$

The interchange of max and min- operations preserves the problems cost if $F(x_1, \ldots x_n) + \lambda g(x_1, \ldots x_n)$ is a convex function of $x_1, \ldots x_n$; in this case $J = J_D$. In a general case, we arrive at an inequality $J \leq J_D$ (see the min-max theorem in Sectionintro)

The extended Lagrangian J_* depends on $n + 1$ variables. The stationary point corresponds to a solution to a system

$$\frac{\partial L}{\partial x_k} = \frac{\partial F}{\partial x_k} + \lambda \frac{\partial g}{\partial x_k} = 0, \quad k = 1, \ldots n, \tag{8}$$

$$\frac{\partial L}{\partial \lambda} = g = 0 \tag{9}$$

The procedure is easily generalized for several constrains. In this case, we add each constraint with its own Lagrange multiplier to the minimizing function and arrive at expression (3)

1.2 Exclusion of Lagrange multipliers and duality

We can exclude the multipliers $\boldsymbol{\mu}$ from the system (3) assuming that the constraints are independent, that is $\text{rank}(W) = m$. We project n-dimensional vector ∇F onto a $n - m$-dimensional subspace allowed by the constraints, and require that this projection is zero. The procedure is as follows.

1. Multiply (3) by W^T:

$$W^T \frac{\partial f}{\partial \boldsymbol{x}} + W^T W \cdot \boldsymbol{\mu} = 0, \tag{10}$$

Since the constraints are independent, $p \times p$ matrix $W^T W$ is nonsingular, $\det(W^T W) \neq 0$.

2. Find m-dimensional vector of multipliers $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = -(W^T W)^{-1} W^T \frac{\partial f}{\partial \boldsymbol{x}},$$

3. Substitute the obtained expression for $\boldsymbol{\mu}$ into (3) and obtain:

$$(I - W(W^T W)^{-1} W^T) \frac{\partial f}{\partial \boldsymbol{x}} = 0 \qquad (11)$$

Matrix $W(W^T W)^{-1} W^T$ is called the projector to the subspace W. Notice that the rank of the matrix $W(W^T W)^{-1} W^T$ is equal to p; it has p eigenvalues equal to one and $n - p$ eigenvalues equal to zero. Therefore the rank of $I - W(W^T W)^{-1} W^T$ is equal to $n - p$, and the system (11) produces $n - p$ independent optimality conditions. The remaining p conditions are given by the constraints (2): $g_i = 0$, $i = 1, \ldots p$. Together these two groups of relations produce n equations for n unknowns $x_1, \ldots, x_n$.

Below, we consider several special cases.

Degeneration: No constraints When there is no constraints, $W = 0$, the problem trivially reduces to the unconstrained on, and the necessary condition (11) becomes $\frac{\partial f}{\partial x} = 0$ holds.

Degeneration: n constraints Suppose that we assign n independent constraints. They themselves define vector $\boldsymbol{x}$ and no additional freedom to choose it is left. Let us see what happens with the formula (11) in this case. The rank of the matrix $W(W^T W)^{-1} W^T$ is equal to n, (W^{-1} exists) therefore this matrix-projector is equal to I:

$$W(W^T W)^{-1} W^T = I$$

and the equation (11) becomes a trivial identity. No new condition is produced by (11) in this case, as it should be. The set of admissible values of $\boldsymbol{x}$ shrinks to the point and it is completely defined by the n equations $\boldsymbol{g}(\boldsymbol{x}) = 0$.

One constraint Another special case occurs if only one constraint is imposed; in this case $p = 1$, the Lagrange multiplier $\boldsymbol{\mu}$ becomes a scalar, and the conditions (3) have the form:

$$\frac{\partial f}{\partial x_i} + \mu \frac{\partial g}{\partial x_i} = 0 \quad i = 1, \ldots n$$

Solving for μ and excluding it, we obtain $n - 1$ stationary conditions:

$$\frac{\partial f}{\partial x_1} \left(\frac{\partial g}{\partial x_1} \right)^{-1} = \ldots = \frac{\partial f}{\partial x_n} \left(\frac{\partial g}{\partial x_n} \right)^{-1} \qquad (12)$$

4

Let us find how does this condition follow from the system (11). This time, W is a $1 \times n$ matrix, or a vector,

$$W = \left[\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n} \right]$$

We have:

$$\text{rank } W(W^T W)^{-1} W^T = 1, \quad \text{rank}(I - W(W^T W)^{-1} W^T) = n - 1$$

Matrix $I - W(W^T W)^{-1} W^T$ has $n - 1$ eigenvalues equal to one and one zero eigenvalue that corresponds to the eigenvector W. At the other hand, optimality condition (11) states that the vector

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right]$$

lies in the null-space of the matrix $I - W(W^T W)^{-1} W^T$ that is vectors $\frac{\partial f}{\partial \boldsymbol{x}}$ and W are parallel. Equation (12) expresses parallelism of these two vectors.

Quadratic function Consider minimization of a quadratic function

$$F = \frac{1}{2} \boldsymbol{x}^T A \boldsymbol{x} + d^T x$$

subject to linear constraints

$$Bx = \beta$$

where $A > 0$ is a positive definite $n \times n$ matrix, B is a $n \times m$ matrix of constraints, d and β are the n- and m-dimensional vectors, respectively. Here, $W = B$. The optimality conditions consist of m constraints $Bx = \beta$ and $n - m$ linear equations

$$(I - B(B^T B)^{-1} B^T)(Ax + d) = 0$$

Duality Let us return to the constraint problem

$$J_D = \min_x \max_\mu (F(x) + \mu^T g(x))$$

with the stationarity conditions,

$$\nabla F + \mu^T W(x) = 0$$

Instead of excluding μ as is was done before, now we do the opposite: Exclude n-dimensional vector x from n stationarity conditions, solving them for x and thus expressing x through μ: $x = \phi(\mu)$. When this expression is substituted into original problem, the later becomes

$$J_D = \max_\mu \{ F(\phi(\mu)) + \mu^T g(\phi(\mu)) \};$$

it is called *dual* problem to the original minimization problem.

Dual form for quadratic problem   Consider again minimization of a quadratic. Let us find the dual form for it. We solve the stationarity conditions $Ax + d + B^T \mu$ for x, obtain

$$x = -A^{-1}(d + B^T \mu)$$

and substitute it into the extended problem:

$$J_D = \max_{\mu \in R_m} \left\{ \frac{1}{2}(d^T + \mu^T B)A^{-1}(d + B^T \mu) - \mu^T BA^{-1}(d + B^T \mu) - \mu^T \beta \right\}$$

Simplifying, we obtain

$$J_D = \max_{\mu \in R_m} \left\{ -\frac{1}{2}\mu^T BA^{-1}B^T \mu - \mu^T \beta + \frac{1}{2}d^T A^{-1} d \right\}$$

The dual problem is also a quadratic form over the m dimensional vector of Lagrange multipliers μ; observe that the right-hand-side term β in the constraints in the original problem moves to the sift term in the dual problem. The shift d in the original problem generates an additive term $\frac{1}{2}d^T A^{-1} d$ in the dual one.

1.3   Finite-dimensional variational problem

Consider the optimization problem for a finite-difference system of equations

$$J = \min_{y_1, \ldots, y_N} \sum_i^N f_i(y_i, z_i)$$

where $f_1, \ldots, f_N$ are given value of a function f, $y_1, \ldots, y_N$ is the N-dimensional vector of unknowns, and z_i $i = 2, \ldots, N$ are the finite differences of y_i:

$$z_i = \text{Diff}(y_i) \quad \text{where Diff}(y_i) = \frac{1}{\Delta}(y_i - y_{i-1}), \quad i = 1, \ldots, N \qquad (13)$$

Assume that the boundary values y_1 and y_n are given and take (13) as constraints. Using Lagrange multiplies $\mu_1, \ldots, \mu_N$ we pass to the augmented function

$$J_a = \min_{y_1, \ldots, y_N;\ z_1, \ldots, z_N} \sum_i^N \left[f_i(y_i, z_i) + \mu_i \left(z_i - \frac{1}{\Delta}(y_i - y_{i-1}) \right) \right]$$

The necessary conditions are:

$$\frac{\partial J_a}{\partial y_i} = \frac{\partial f_i}{\partial y_i} + \frac{1}{\Delta}(-\mu_i + \mu_{i+1}) = 0 \quad 2 = 1, \ldots, N - 1$$

and

$$\frac{\partial J_a}{\partial z_i} = \frac{\partial f_i}{\partial z_i} + \mu_i = 0 \quad i = 2, \ldots, N - 1$$

Excluding μ_i from the last equation and substituting their values into the previous one, we obtain the conditions:

$$\frac{\partial J_a}{\partial y_i} = \frac{\partial f_i}{\partial y_i} + \frac{1}{\Delta}\left(\frac{\partial f_i}{\partial z_i} - \frac{\partial f_{i+1}}{\partial z_{i+1}}\right) = 0 \quad i = 2,\ldots,N-1$$

or, recalling the definition of the Diff -operator,

$$\text{Diff}\left(\frac{\partial f_{i+1}}{\partial z_{i+1}}\right) - \frac{\partial f_i}{\partial y_i} = 0 \quad z_i = \text{Diff}(y_i) \tag{14}$$

One can see that the obtained necessary conditions have the form of the difference equation of second-order.

Formal passage to differential equation Diff-operator is an approximation of a derivative and the equation (14) is a finite-difference approximation of the Euler equation. When $N \to \infty$,

$$\text{Diff}(y_i) \to \frac{dy}{dx}$$

and we obtain the differential equation of the second order (the Euler equation):

$$\frac{d}{dx}\frac{\partial F}{\partial u'} - \frac{\partial F}{\partial u} = 0$$

for the unknown minimizer – function $y(x)$.

2 Inequality constraints

Nonnegative Lagrange multipliers Consider the problem with a constraint in the form of inequality:

$$\min_{x_1,\ldots x_n} F(x_1,\ldots x_n) \quad \text{subject to } g(x_1,\ldots x_n) \le 0 \tag{15}$$

In order to apply the Lagrangian multipliers technique, we reformulate the constraint:

$$g(x_1,\ldots x_n) + v^2 = 0$$

where v is a new auxiliary variable.

The augmented Lagrangian becomes

$$L_*(x,v,\lambda) = f(x) + \lambda g(x) + \lambda v^2$$

and the optimality conditions with respect to v are

$$\frac{\partial L_*}{\partial v} = 2\lambda v = 0 \tag{16}$$

$$\frac{\partial^2 L_*}{\partial v^2} = 2\lambda \ge 0 \tag{17}$$

The second condition requires the nonnegativity of the Lagrange multiplier and the first one states that the multiplier is zero, $\lambda = 0$, if the constraint is satisfied by a strong inequality, $g(x_0) > 0$.

The stationary conditions with respect to x

$$\begin{array}{ll} \nabla f = 0 & \text{if } g \leq 0 \\ \nabla f + \lambda \nabla g = 0 & \text{if } g = 0 \end{array}$$

state that either the minimum correspond to an inactive constraint ($g > 0$) and coincide with the minimum in the corresponding unconstrained problem, or the constraint is active ($g = 0$) and the gradients of f and g are parallel and directed in opposite directions:

$$\frac{\nabla f(x_b) \cdot \nabla g(x_b)}{|\nabla f(x_b)| \, |\nabla g(x_b)|} = -1, \quad x_b : \ g(x_b) = 0$$

In other terms, the projection of $\nabla f(x_b)$ on the subspace orthogonal to $\nabla g(x_b)$ is zero, and the projection of $\nabla f(x)$ on the direction of $\nabla g(x_b)$ is nonpositive.

The necessary conditions can be expressed by a single formula using the notion of infinitesimal variation of x or a differential. Let x_0 be an optimal point, x_{trial} – an admissible (consistent with the constraint) point in an infinitesimal neighborhood of x_0, and $\delta x = x_{\text{trial}} - x_0$. Then the optimality condition becomes

$$\nabla f(x_0) \cdot \delta x \geq 0 \quad \forall \delta x \tag{18}$$

Indeed, in the interior point x_0 ($g(x_0) > 0$) the vector δx is arbitrary, and the condition (18) becomes $\nabla f(x_0) = 0$. In a boundary point x_0 ($g(x_0) = 0$), the admissible points are satisfy the inequality $\nabla g(x_0) \cdot \delta x \leq 0$, the condition (18) follows from (17).

It is easy to see that the described constrained problem is equivalent to the unconstrained problem

$$L_* = \min_{x_1, \ldots x_n} \max_{\lambda > 0} \left(F(x_1, \ldots x_n) + \lambda g(x_1, \ldots x_n) \right) \tag{19}$$

that differs from (7) by the requirement $\lambda > 0$.

Several constraints: Kuhn-Tucker conditions Several inequality constraints are treated similarly. Assume the constraints in the form

$$g_1(x) \leq 0, \ldots, g_m(x) \leq 0.$$

The stationarity condition can be expressed through nonnegative Lagrange multipliers

$$\nabla f + \sum_{i=1}^{m} \lambda_i \nabla g_i = 0, \tag{20}$$

where
$$\lambda_i \geq 0, \quad \lambda_i g_i = 0, \quad i = 1, \ldots, m. \tag{21}$$

The minimal point corresponds either to an inner point of the permissible set (all constraints are inactive, $g_i(x_0) < 0$), in which case all Lagrange multipliers λ_i are zero, or to a boundary point where $p \leq m$ constraints are active. Assume for definiteness that the first p constraints are active, that is

$$g_1(x_0) = 0, \quad \ldots, \quad g_p(x_0) = 0. \tag{22}$$

The conditions (21) show that the multiplier λ_i is zero if the ith constrain is inactive, $g_i(x) > 0$. Only active constraints enter the sum in (23), and it becomes

$$\nabla f + \sum_{i=1}^{p} \lambda_i \nabla g_i = 0, \quad \lambda_i > 0, \quad i = 1, \ldots, p. \tag{23}$$

The term $\sum_{i=1}^{p} \lambda_i \nabla g_i(x_0)$ is a cone with the vertex at x_0 stretched on the rays $\nabla g_i(x_0) > 0$, $i = 1, \ldots, p$. The condition (23) requires that the negative of $\nabla f(x_0)$ belongs to that cone.

Alternatively, the optimality condition can be expressed through the admissible vector δx,

$$\nabla f(x_0) \cdot \delta x \geq 0 \tag{24}$$

Assume again that the first p constraints are active, as in (**??**)

$$g_1(x_0) =, \ldots, = g_p(x_0) = 0.$$

In this case, the minimum is given by (24) and the admissible directions of δx satisfy the system of linear inequalities

$$\delta x \cdot \nabla g_i \geq 0, \quad i = 1, \ldots, p. \tag{25}$$

These conditions are called Kuhn-Tucker conditions, see []