Constrainted problems

February 11, 2019

Contents

1 Introduction

Many variational problems ask for a constrained minimum of a variational functional. We will discuss several types of constraints.

(i) Isoperimetric problem (Section **??**). Minimize an integral functional if the values of the others integral functionals are given. The classic example is the Dido problem: maximize the area encircled by a rope of a given length.

$$\min_{y(x),T} \int_0^T y(x)\,dx \quad \text{subject to} \quad \int_0^T \sqrt{1+(y')^2}\,dx \quad u(0)=0, \quad u(T)=0 \tag{1}$$

where

The problem of minimization of the product, ratio, or superposition of the integrals can also be reduced to constrained variational problems, as described in Section **??**. An example is the problem of the principle eigenfrequency that is defined as a ratio between the total potential and kinetic energy of an oscillating body.

(ii) Problem with constraints imposed at each point of the trajectory (Section **??**). An example is a problem of geodesics: minimize the distance of the path between two points if the path everywhere belongs to a given surface $\phi(x,x,z) =$

0.

$$\min_{u(t)=\mathcal{U}} \int_0^1 ds(t)\,dt, \quad ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt \text{ subject to } \phi(x,x,z)=0 \tag{2}$$

where

$$\mathcal{U} = \{u = (u_1(t), u_2(t), u_3(t)) : \ u(0) = A, u(1) = B\}$$

(iii) Problems with differential constraints (Section **??**). The simplest variational problem can be rewriten this way:

$$\min_{u(x),v(x)} \int_0^T F(x,u,v)\,dx \quad \text{subject to } u' = v \tag{3}$$

More general differential constraints $L(u,u') = 0$ can be imposed as well. A example is the problem of minimization of the fuel consumption by a vehicle that moves in the given time T between two points A and B, if the rate of spend fuel f defines the speed of the vehicle u' through a differential equation of the motion that depends on f: $L(u,u',f) = 0$.

$$\min_{f>0} \int_0^T f(t)\,dt \quad \text{subject to } L(u,u',f)=0,\ u(0)=A,\ u(T)=B. \tag{4}$$

2 Lagrange multipliers: Vector problem

2.1 Lagrange multipliers method

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

$$J = \min_x f(x), \quad x \in R^n, \quad f \in C_2(R^n) \tag{5}$$

assuming that m constraints are given

$$g_i(x) = 0 \quad i = 1, \dots m, \quad m \le n, \tag{6}$$

The vector form of the constraints is

$$g(x) = 0$$

where g is an m-dimensional vector-function of an n-dimensional vector x.

To find the minimum, we add the constraints with the Lagrange multipliers $\lambda = (\lambda_1, \dots \lambda_m)$ and consider the problem for $J(x,\lambda)$

$$J(x,\lambda) = \min_x \left[f(x) + \sum_i^m \lambda_i g_i(x) \right]$$

The stationary conditions and the constraints form the system

$$\frac{dL}{dx_k} = \frac{df}{dx_k} + \sum_i^m \lambda_i \frac{dg_i}{dx_k} = 0, \quad k = 1, \dots n, \tag{7}$$

$$\frac{dL}{d\lambda_i} = g_i(x) = 0 \quad i = 1, \dots m \tag{8}$$

The vector form of these relations is

$$\frac{df}{dx} + W \cdot \lambda = 0, \quad g(x) = 0 \tag{9}$$

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

$$W = \frac{dg}{dx} \quad \text{or, by elements, } W_{ij} = \frac{dg_i}{dx_j}$$

The system (9) together with the constraints (6) forms a system of $n + p$ equations for $n + p$ unknowns: Components of the vectors x and λ.

Lagrange explanation Lagrange came to the method by considering an equilibrium of a particle on a surface. Assume that a particle moves minimizing its potential energy $V(x)$ that depends on the point $x = (x_1, x_2, x_3)$ in space. In the equilibrium point, energy $V(x)$ reaches a local minimum. The force f acting on the particle is $f = -\frac{dV}{dx}$. The particle reaches an equilibrium at a point x where $f(x) = 0$.

Now assume that the particle moves along a surface $g(x) = 0$. the Third law of Newton states that for every action force f, there is an equal and opposite reaction force. The reaction force exerted from the surface $R(x) = 0$, its magnitude λ is *a priori* unknown, and its direction is co-directed with the normal $\frac{dg}{dx}$ to the surface, $R = -\lambda \frac{dg}{dx}$. The equilibrium is reached at a point where the sum of the action and reaction forces is zero, $f + R = 0$, or

$$\frac{dV}{dx} + \lambda \frac{dg}{dx} = 0$$

In other words, force $f(x)$ in an equilibrium point x is directed along the normal to the surface $g(x)$.

2.2 Background of the method

Consider the finite-dimensional minimization problem

$$J = \min_{x_1, \dots x_n} F(x_1, \dots x_n) \tag{10}$$

subject to one constraint

$$g(x_1, \dots x_n) = 0 \tag{11}$$

and assume that solutions to (11) exist in a neighborhood of the minimal point. It is easy to see that the described constrained problem is equivalent to the unconstrained problem

$$J_* = \min_{x_1,\dots x_n} \max_{\lambda} \left(F(x_1,\dots x_n) + \lambda g(x_1,\dots x_n)\right) \tag{12}$$

Indeed, the operation of maximization gives

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

because λ can be made arbitrary large or arbitrary small. This possibility forces us to choose such x that delivers equality in (11), otherwise the cost of the problem (12) would be infinite (recall that we look for vector x that minimizes J_*). By assumption, such x exists. On the other hand, the constrained problem (10)-(11) does not change its cost J if the term $\lambda g(x) = 0$ is added to it. Therefore $J = J_*$ and the problem (10) - (11) is equivalent to (12).

Minimax and maximin Consider the minimax and maximin problems

$$I_{\min\max} = \min_u \max_v f(u,v) \tag{13}$$

$$I_{\max\min} = \max_v \min_u f(u,v) \tag{14}$$

that differ by the sequence of extremal operations. Let us show that

$$I_{\min\max} \geq I_{\max\min} \tag{15}$$

We call $\phi(v) = \min_u f(u,v)$ and $\psi(u) = \max_v f(u,v)$ The inequalities hold

$$\phi(v) \leq f(u,v) \leq \psi(u) \quad \forall u, v$$

for all values of u and v. In particular

$$\max_v \phi(v) \leq f(u,v) \leq \min_u \psi(u), \quad \forall u, v$$

which is equivalent to (15).

The interchange of max and min operations preserves the problem's cost if the function $f(u,v)$ has a stationary saddle point. This happens, if $f(u,v)$ is convex with respect to u and concave with respect to v, for all values of u and v. In this case.

$$I_{\min\max} = I_{\max\min}$$

Augmented problem If we interchange the sequence of the two extremal operations in (12), we would arrive at the *dual* problem J_D

$$J \geq J_D(x,\lambda) = \max_{\lambda} \min_{x} \left(F(x) + \lambda g(x)\right) \tag{16}$$

where x is a vector $x = (x_1, \dots x_n)$. The inequality $J \geq J_D$ is called the *weak duality relation*

The interchange of max and min- operations preserves the problem's cost if $f(x) + \lambda g(x)$ is a convex function of x because it is a linear function of λ. In this case $J = J_D$. This equality is called the *strong duality relation*

The procedure is easily generalized for the case of several constraints. In this case, we add each constraint with its own Lagrange multiplier to the minimizing functional and arrive at expression (9).

2.3 Examples

Example 2.1

$$J = \min_x \frac{1}{2} \sum_{i=1}^{n} x_i^2 \quad \text{subject to } a^T x = b$$

where $x, a \in R^n$, b is a scalar.

The extended Lagrangian is

$$I = \frac{1}{2} \sum_{i=1}^{n} x_i^2 + \lambda(a^T x - b)$$

where λ is a scalar Lagrange multiplier.

Interchange the extremal operations, compute derivative with respect of x and set it to zero:

$$\frac{dI}{dx_i} = x_i + \lambda a_i = 0 \quad \text{or} \quad x = -\lambda a$$

This conditions shows that vectors x and a are proportional to each other: $x = \lambda a$. Now we find λ from the constraint:

$$a^T x - b = \lambda |a|^2 - b = 0 \quad \Rightarrow \quad \lambda = \frac{b}{|a|^2}$$

and find optimal value x_0 of x

$$x_0 = -\lambda a = -\frac{b}{|a|^2} a$$

The cost J of the problem is

$$J = \frac{1}{2} x_0^T x_0 = \frac{1}{2} \frac{b^2}{|a|^2}$$

Remark 2.1 Notice that the derivative of the scalar product $a^T x$ with respect of a vector x is

$$\frac{d(a^T x)}{dx} = a$$

Similarly, the derivative of a quadratic form is computed as

$$\frac{d(x^T A x)}{dx} = 2Ax$$

Here, matrix A is assumed to be symmetric: $A = A^T$.

The next example shows the eigenvalue problem.

Example 2.2

$$J = \min_x x^T A x \quad \text{subject to } x^T x = 1$$

where $x, \in R^n$, A is a symmetric $n \times n$ matrix.

The extended Lagrangian is

$$L = x^T A x + \lambda(x^T x - 1)$$

The optimality condition

$$\frac{dL}{dx} = 2\,(Ax - \lambda x) = 2(A - \lambda I)x = 0$$

shows that λ is an eigenvalue of A and x is the corresponding normalized eigenvector. The cost J is $J = \lambda$. Clearly, the minimum corresponds to the smallest eigenvalue $J = \lambda_{\min}$.

Example 2.3 Consider the problem

$$J = \min_x \sum_{i=1}^n A_i^2 x_i \quad \text{subject to } \sum_{i=1}^n \frac{1}{x_i - k} = \frac{1}{c}, \quad x_i > k, c > 0, k > 0.$$

where $A = (A_1, \ldots, A_n)$ is a given vector, c and k are given constants.

Using Lagrange multiplier λ we rewrite it in the form:

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

The condition $\frac{dJ_a}{dx} = 0$ shows that

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

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

$$\frac{1}{c} = \sum_{i=1}^n \frac{1}{x_i - k} = \frac{1}{\sqrt{\lambda}} \sum_{i=1}^n |A_i|$$

Solving this equation, we find λ and the minimizer x_i

$$\sqrt{\lambda} = c \sum_{i=1}^n |A_i|, \quad x_i = k + \frac{\sqrt{\lambda}}{|A_i|},$$

as well as the optimal value of J:

$$J = k \sum_{i=1}^n A_i^2 + c \left(\sum_{i=1}^n |A_i| \right)^2$$

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

2.4 Exclusion of Lagrange multipliers

We can exclude the multipliers λ from the system (9) assuming that the constraints are independent, that is that $\text{rank}(W) = m$ (Recall that $W = \frac{dg}{dx}$). We project the n-dimensional vector $\frac{d}{dx}F$ onto an $(n-m)$-dimensional subspace allowed by the constraints, and require that this projection is zero. The procedure is as follows.

1. Multiply (9) by W^T:
$$W^T \frac{df}{dx} + W^T W \cdot \lambda = 0, \tag{17}$$
Since the constraints are independent, $m \times m$ matrix $W^T W$ is nonsingular, $\det(W^T W) \neq 0$.

2. Find m-dimensional vector of multipliers λ:
$$\lambda = -(W^T W)^{-1} W^T \frac{df}{dx},$$

3. Substitute the obtained expression for λ into (9) and obtain:
$$(I - P)\frac{df}{dx} = 0 \quad P = W(W^T W)^{-1} W^T \tag{18}$$

Projection Matrices P and $I - P$ are called the projector to the subspace W and projector to the orthogonal subspace, respectively.

Definition A symmetric $n \times n$ matrix $P = W(W^T W)^{-1} W^T$ is called a projector, if its eigenvalues are equal to one or zero.

One can check that P is a projector by checking equality $P^2 = P$:
$$P^2 = [W(W^T W)^{-1} W^T][W(W^T W)^{-1} W] = W(W^T W)^{-1} W^T = P$$
(we see this immediately by opening parenthesis and canceling $W^T W(W^T W)^{-1} = I$.

If P is a projector on a subspace, $(I - P)$ is a projector on the orthogonal subspace, since the eigenvalue of P that are equal to one become zero eigenvalues of $I - P$, and vise versa. The number m of linearly independent constraints equals to the rank of P: it has m eigenvalues equal to one and $n-m$ eigenvalues equal to zero. Therefore the rank of $I-P$ is equal to $n-m$, and the system (18) produces $n-m$ independent optimality conditions. The remaining m conditions are given by the constraints (6): $g_i = 0,\ i = 1, \dots m$. Together, these two groups of relations produce n equations for n unknowns $x_1, \dots, x_n$.

Below, we consider several special cases.

Degeneration: No constraints When there are no constraints, $W = 0$, the problem trivially reduces to the unconstrained one, and the necessary condition (18) becomes $\frac{df}{dx} = 0$ holds.

Degeneration: n constraints Suppose that we assign n independent constraints. They themselves define vector x and no additional freedom is left. Let us see what happens with the formula (18) in this case. The rank of the matrix $W(W^TW)^{-1}W^T$ is equal to n, (W^{-1} exists) therefore this matrix-projector is equal to I:

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

The equation (18) becomes a trivial identity. No new condition is produced by (18) in this case, as it should be. The set of admissible values of x shrinks to one point and it is completely defined by the n equations $g(x) = 0$.

One constraint Another special case occurs if only one constraint is imposed; in this case $m = 1$, the Lagrange multiplier λ becomes a scalar, and the conditions (9) have the form:

$$\frac{df}{dx_i} + \lambda \frac{dg}{dx_i} = 0 \quad i = 1, \dots n$$

which show that the vectors $\frac{df}{dx_i}$ and $\frac{dg}{dx_i}$ are parallel, $\frac{df}{dx_i} || \frac{dg}{dx_i}$

Example 2.4 (Quadratic function) Consider minimization of a quadratic function

$$f(x) = \frac{1}{2}x^TAx + a^Tx$$

subject to linear constraints

$$Bx = \beta$$

where $A > 0$ is a symmetric positive definite $n \times n$ matrix, B is a $m \times n$ matrix of constraints, a 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

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

2.5 Duality

Let us return to the constrained problem

$$J = \min_x \max_\lambda L(x, \lambda), \quad L(x, \lambda) = (f(x) + \lambda^T g(x))$$

Interchanging the max and min operations, we find the dual problem

$$J_D = \max_\lambda \min_x (f(x) + \lambda^T g(x))$$

The cost J_D of the dual problem is not larger than J,

$$J_D \leq J \tag{19}$$

The stationarity conditions for the dual problem are:

$$\frac{df}{dx} + \lambda^T W(x) = 0, \quad W = \frac{dg}{dx}.$$

Instead of excluding λ that has been done above, now we do the opposite: Exclude n-dimensional vector x from n stationarity conditions, solving them for x and thus expressing x through λ: $x = \phi(\lambda)$. When this expression is substituted into the original problem, it becomes

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

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

The dual problem asks for maximization of the cost, therefore any admissible vector λ provides the *lower bound* for J_D and therefore, as $J_D \leq J$, for J. Recall, that any admissible vector x provides the upper bound for the original minimization problem. Therefore, the pair of the admissible vectors x and λ give the two-side bounds for the cost function.

Dual form for quadratic problem Consider again minimization of a quadratic function in example 2.4

$$f(x) = \frac{1}{2}x^T A x + a^T x$$

subject to linear constraints

$$Bx = \beta$$

Let us find the dual form for it. The Lagrangian is

$$L = \frac{1}{2}x^T A x + a^T x + \lambda^T (Bx - \beta).$$

The stationarity conditions are:

$$\frac{dL}{dx} = Ax + d + B^T \lambda = 0$$

We solve them for x, obtaining

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

and substitute it into the Lagrangian:

$$\begin{aligned} L_D(\lambda) &= \frac{1}{2}(d^T + \lambda^T B)A^{-1}(d + B^T\lambda) - \lambda^T B A^{-1}(d + B^T\lambda) \\ &- \lambda^T \beta - d^T A^{-1}(d + B^T \lambda) \end{aligned}$$

Simplifying, we obtain the dual Lagrangian

$$L_D(\lambda) = -\lambda^T \beta - \frac{1}{2}(B^T\lambda + d)^T A^{-1}(B^T \lambda + d)$$

and the dual problem

$$J_D = \max_{\lambda \in R^m} L_D(\lambda)$$

is also a quadratic form over the m dimensional vector of Lagrange multipliers λ.

2.6 Inequality constraints

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

$$\min_{x_1,\dots x_n} F(x_1,\dots x_n) \quad \text{subject to } g(x_1,\dots x_n) \le 0 \tag{20}$$

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

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

where v is a new auxiliary variable.

The extended Lagrangian becomes

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

and the optimality conditions with respect to v are

$$\frac{\partial L_*}{\partial v} = 2\lambda v = 0 \tag{21}$$

$$\frac{\partial^2 L_*}{\partial v^2} = 2\lambda \ge 0 \tag{22}$$

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 as a strong inequality, $g(x_0) > 0$.

The stationary conditions with respect to x

$$\begin{array}{ll} \frac{d}{dx} f = 0 & \text{if } g \le 0 \\ \frac{d}{dx} f + \lambda \frac{d}{dx} g = 0 & \text{if } g = 0 \end{array}$$

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 ($g(x_b) = 0$) and the gradients of f and g are parallel and directed in opposite directions:

$$\frac{\frac{d}{dx} f(x_0) \cdot \frac{d}{dx} g(x_b)}{|\frac{d}{dx} f(x_b)|\,|\frac{d}{dx} g(x_b)|} = -1, \quad x_b : \; g(x_b) = 0$$

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

$$\frac{d}{dx} f(x_0) \cdot \delta x \le 0 \quad \forall \delta x \tag{23}$$

Indeed, in the interior point x_0 ($g(x_0) < 0$) the vector δx is arbitrary, and the condition (23) becomes $\frac{d}{dx} f(x_0) = 0$. In a boundary point x_0 ($g(x_0) = 0$), the admissible points satisfy the inequality $\frac{d}{dx} g(x_0) \cdot \delta x \leq 0$, the condition (23) follows from (22).

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

$$L_* = \min_{x_1, \dots x_n} \max_{\lambda \geq 0} \left(F(x_1, \dots x_n) + \lambda g(x_1, \dots x_n) \right) \tag{24}$$

that differs from (16) by the requirement $\lambda \geq 0$.

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

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

The stationarity condition can be expressed through nonnegative Lagrange multipliers

$$\frac{d}{dx} f(x) + \sum_{i=1}^{m} \lambda_i \frac{d}{dx} g_i(x) = 0, \tag{25}$$

where

$$\lambda_i \geq 0, \quad \lambda_i g_i(x) = 0, \quad i = 1, \dots, m. \tag{26}$$

The minimal point corresponds either to an interior point of the admissible 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

$$g_1(x_0) = 0, \quad \dots, \quad g_p(x_0) = 0. \tag{27}$$

The conditions (26) show that the multiplier λ_i is zero if the ith constraint is inactive, $g_i(x) > 0$. Only active constraints enter the sum in (28), and it becomes

$$\frac{d}{dx} f(x) + \sum_{i=1}^{p} \lambda_i \frac{d}{dx} g_i(x) = 0, \quad \lambda_i > 0, \quad i = 1, \dots, p. \tag{28}$$

In the space R^m of the components of Lagrange multiplier vector, the term $\sum_{i=1}^{p} \lambda_i \frac{d}{dx} g_i(x_0)$ is a cone with the vertex at x_0 stretched on the rays $\frac{d}{dx} g_i(x_0) > 0$, $i = 1, \dots, p$. The condition (28) requires that the negative of $\frac{d}{dx} f(x_0)$ belongs to that cone.

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

$$\frac{d}{dx} f(x_0) \cdot \delta x \geq 0 \tag{29}$$

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

$$g_1(x_0) = \dots = g_p(x_0) = 0.$$

In this case, the minimum is given by (29) and the admissible directions of δx satisfy the system of linear inequalities

$$\delta x \cdot \frac{d}{dx} g_i \geq 0, \quad i = 1, \ldots, p. \tag{30}$$

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