Sufficient Conditions and Necessary Conditions for Extreme Value Problems with Constraints

Chengshen Xu[1,]

[a]*Autohome Inc., 10th Floor Tower B, No. 3 Dan Ling Street Haidian District, Beijing, China*

Abstract

In this paper, we prove that the sufficient conditions for the extreme value problem with constraints requires that the projection of the gradient of the point in the final constraint surface is zero and that the second-order partial derivative matrix–Hessian matrix on the local linear subspace in the constraint surface is positive definite or negative definite. The necessary conditions require that the projection of the gradient of the point in the final constraint surface be zero and that the second-order partial derivative matrix–Hessian matrix on the local linear subspace in the constraint surface be semi-positive definite or semi-negative definite. Finally, we discuss the reconstruction of the micro-base vector on the local linear subspace in the constraint surface and the local coordinate system.

Keywords: Hessian matrix, Second-order Partial Derivative Matrix, Positive Definite, Semi-positive Definite, Negative Definite, Semi-negative Definite, Constraint, Local Linear Surface, Micro-base Vector, Local Coordinate System

1. Introduction

Extreme value problems

$$\begin{cases} \min\limits_{x_1,x_2,\cdots,x_n} f(x_1,x_2,\cdots,x_n) \\ g_1(x_1,x_2,\cdots,x_n)=0, \\ g_2(x_1,x_2,\cdots,x_n)=0, \\ \vdots, \\ g_m(x_1,x_2,\cdots,x_n)=0 \end{cases} \tag{1}$$

with constraints are usually solved by Lagrange method of multipliers [**?**]. Now let's study the content and geometric meaning of Lagrange method of multipliers in detail.

*Corresponding author: xcssgzs@126.com

We already know that the vector perpendicular to the level surface is the gradient vector, so the vector perpendicular to the constraint surface

$$g_i(x_1, x_2, \cdots, x_n) = 0 \tag{2}$$

is

$$\nabla g_i = (\frac{\partial g_i}{\partial x_1}, \frac{\partial g_i}{\partial x_2}, \cdots, \frac{\partial g_i}{\partial x_n})^T \tag{3}$$

Thus the final constraint surface S is the intersection of the m constraint surfaces and each linear combination of the gradient of each constraint surface

$$\sum_{i=1}^{m} \lambda_i \nabla g_i(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{m} \lambda_i (\frac{\partial g_i}{\partial x_1}, \frac{\partial g_i}{\partial x_2}, \cdots, \frac{\partial g_i}{\partial x_n})^T \tag{4}$$

is perpendicular to the final constraint plane S. Therefore, the necessary condition for the extreme value of the function $f(x_1, x_2, \cdots, x_n)$ is weakened to

$$\nabla f = \sum_{i=1}^{m} \lambda_i \nabla g_i \tag{5}$$

Let's write this expression in terms of its components combined with the constraints

$$\begin{cases} \frac{\partial f}{\partial x_1} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_1} = 0, \\ \frac{\partial f}{\partial x_2} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_2} = 0, \\ \qquad \vdots, \\ \frac{\partial f}{\partial x_n} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_n} = 0, \\ g_1(x_1, x_2, \cdots, x_n) = 0, \\ g_2(x_1, x_2, \cdots, x_n) = 0, \\ \qquad \vdots, \\ g_m(x_1, x_2, \cdots, x_n) = 0. \end{cases} \tag{6}$$

If we introduce the Lagrange function, the above equations can be reformulated as

$$\begin{cases} \frac{\partial L}{\partial x_1} = 0, \\ \frac{\partial L}{\partial x_2} = 0, \\ \quad \vdots, \\ \frac{\partial L}{\partial x_n} = 0, \\ \frac{\partial L}{\partial \lambda_1} = 0, \\ \frac{\partial L}{\partial \lambda_2} = 0, \\ \quad \vdots, \\ \frac{\partial L}{\partial \lambda_m} = 0. \end{cases} \tag{7}$$

2. The Sufficient Conditions and Necessary Conditions for Taking an Extreme Value Without Constraints

Now let's think about what's sufficient or necessary to take the extremum when the first derivative i.e. the gradient is zero. We expand the function to the second order

$$\Delta f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \Delta x_i + \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k} \Delta x_j \Delta x_k \tag{8}$$

whose vector form is

$$\Delta f(x_1, x_2, \cdots, x_n) = \nabla f \cdot \Delta \boldsymbol{x} + \Delta \boldsymbol{x}^T H \Delta \boldsymbol{x} \tag{9}$$

where

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}. \tag{10}$$

When

$$\nabla f = \mathbf{0} \tag{11}$$

the necessary condition that

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \tag{12}$$

is an extreme value is that H is either semi-positive definite (minimum) or semi-negative definite (maximum). And of course, this is a necessary condition, not a sufficient condition, because even if it's a semi-positive definite or a semi-negative definite, once some eigenvalue is zero, that is, the determinant is zero, we still have to look at higher derivatives of $f(x_1, x_2, \cdots, x_n)$.

When

$$\nabla f = \mathbf{0} \tag{13}$$

the sufficient condition that

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \tag{14}$$

is an extreme value is that H is either positive definite (minimum) or negative definite (maximum). And of course, this is a sufficient condition, not a necessary condition, because when

$$|H| = 0 \tag{15}$$

it may also be an extreme value.

3. The Sufficient Conditions and Necessary Conditions for Taking an Extreme Value With Constraints

For constrained problems, when the gradient is zero, the necessary conditions and sufficient conditions for the function to take an extreme value are not as strong as unconstrained problems for the second derivative matrix. Since the argument can only be active in the final constraint surface, the necessary condition is that H is semi-positive definite or semi-negative definite in the linear subspace of the final constraint surface, and the sufficient condition is that H is positive definite or negative definite.

If the m constraints are linearly independent at a specific point, then the dimension of the constraint surface at this point is $n-m$. If it is linearly dependent, the final dimension will be larger than $n-m$. There is no harm to suppose that it is linearly independent because the process of argument is the same if dependent. We select a set of base vectors within the constraint surface

$$\hat{\boldsymbol{e}}'_1, \hat{\boldsymbol{e}}'_2, \cdots, \hat{\boldsymbol{e}}'_{n-m} \tag{16}$$

The vector group can be extended to a group of base vectors in the whole space

$$\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \cdots, \hat{\boldsymbol{e}}_n \tag{17}$$

Suppose that the transition matrix from the original base vectors to the new base vectors is S, i.e

$$\begin{pmatrix} \hat{\boldsymbol{e}}'_1 & \hat{\boldsymbol{e}}'_2 & \cdots & \hat{\boldsymbol{e}}'_n \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{e}}_1 & \hat{\boldsymbol{e}}_2 & \cdots & \hat{\boldsymbol{e}}_n \end{pmatrix} S, \tag{18}$$

so its matrix form is

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} \Delta x'_1 \\ \Delta x'_2 \\ \vdots \\ \Delta x'_n \end{pmatrix} \tag{19}$$

Therefore, we have

$$\Delta\boldsymbol{x}^T H \Delta\boldsymbol{x} = \Delta\boldsymbol{x}'^T S^T H S \Delta\boldsymbol{x}' \tag{20}$$

Then the second derivative matrix under the new base vectors is

$$H' = S^T H S = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \tag{21}$$

where A is a symmetric matrix of order $(n-m)\times(n-m)$, B is a matrix of order $(n-m)\times m$, and C is a matrix of order $m\times m$. Then the necessary condition for taking the extreme value only requires A to be semi-positive definite or semi-negative definite rather than require H to be semi-positive definite or semi-negative definite. And the sufficient condition for taking the extreme value only requires A to be positive definite or negative definite rather than require H to be positive definite or negative definite.

4. The Method of Constructing Micro Base Vectors in the Constrained Subspace

In order to find the above matrix A, we need to establish a group of micro base vectors in the local linear subspace of the constraint surface. Taking the partial derivative of the constraints, we can obtain

$$\begin{cases} \sum\limits_{i=1}^{n} \frac{\partial g_1}{\partial x_i} \Delta x_i = 0, \\ \sum\limits_{i=1}^{n} \frac{\partial g_2}{\partial x_i} \Delta x_i = 0, \\ \quad \vdots, \\ \sum\limits_{i=1}^{n} \frac{\partial g_m}{\partial x_i} \Delta x_i = 0. \end{cases} \tag{22}$$

whose matrix form is

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{23}$$

Since the m constraints may be linearly dependent, the column rank l of the matrix may satisfy

$$l < m \tag{24}$$

By solving this system of linear equations, we can find $n - l$ free variables, supposing the $n - l$ independent free variables are

$$x_{i_{l+1}}, x_{i_{l+2}}, \cdots, x_{i_n} \tag{25}$$

where $i_1, i_2, \cdots, i_n$ is a rearrangement of $1, 2, \cdots, n$. We select

$$\begin{pmatrix} \hat{\boldsymbol{e}}_{i_1} & \hat{\boldsymbol{e}}_{i_2} & \cdots & \hat{\boldsymbol{e}}_{i_n} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{e}}_1 & \hat{\boldsymbol{e}}_2 & \cdots & \hat{\boldsymbol{e}}_n \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix} \tag{26}$$

as the new base vectors, then the transformation relation between the old coordinate and the new coordinate is

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix} \begin{pmatrix} \Delta x_{i_1} \\ \Delta x_{i_2} \\ \vdots \\ \Delta x_{i_n} \end{pmatrix} \tag{27}$$

where

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix} \tag{28}$$

is the permutation matrix (In each row, only one element is 1, and the rest are 0; the same is true for each column). So if we solve this equation we finally get

$$\begin{cases} \Delta x_{i_1} = \sum\limits_{j=l+1}^{n} k_{1j} \Delta x_{i_j}, \\ \Delta x_{i_2} = \sum\limits_{j=l+1}^{n} k_{2j} \Delta x_{i_j}, \\ \qquad \vdots, \\ \Delta x_{i_l} = \sum\limits_{j=l+1}^{n} k_{lj} \Delta x_{i_j}. \end{cases} \tag{29}$$

It can be seen that $\Delta x_{i_1}, \Delta x_{i_2}, \cdots, \Delta x_{i_l}$ are not an independent variables and are uniquely determined by $\Delta x_{i_{l+1}}, \Delta x_{i_{l+2}}, \cdots, \Delta x_{i_n}$. Therefore we can select the vectors

$$\begin{cases} (k_{1,l+1}, k_{2,l+1}, \cdots, k_{l,l+1}, 1, 0, 0, \cdots, 0), \\ (k_{1,l+2}, k_{2,l+2}, \cdots, k_{l,l+2}, 0, 1, 0, \cdots, 0), \\ \qquad \vdots, \\ (k_{1,n}, \quad k_{2,n}, \quad \cdots, \quad k_{l,n}, \quad 0, 0, 0, \cdots, 1). \end{cases} \tag{30}$$

as new base vectors,

$$\begin{cases} \hat{\boldsymbol{e}}'_1 \\ \hat{\boldsymbol{e}}'_2, \\ \ \vdots, \\ \hat{\boldsymbol{e}}'_{n-l}. \end{cases} \tag{31}$$

So we have found the basis vectors that satisfies the above condition in last section. We extend it further to a set of base vectors, namely

$$\begin{cases} (k_{1,l+1}, k_{2,l+1}, \cdots, k_{l,l+1}, 1, 0, 0, \cdots, 0), \\ (k_{1,l+2}, k_{2,l+2}, \cdots, k_{l,l+2}, 0, 1, 0, \cdots, 0), \\ \qquad \vdots, \\ (k_{1,n}, \quad k_{2,n}, \quad \cdots, \quad k_{l,n}, \quad 0, 0, 0, \cdots, 1), \\ (k_{1,1}, \quad k_{2,1}, \quad \cdots, \quad k_{l,1}, \quad 0, 0, 0, \cdots, 0), \\ (k_{1,2}, \quad k_{2,2}, \quad \cdots, \quad k_{l,2}, \quad 0, 0, 0, \cdots, 0), \\ \qquad \vdots, \\ (k_{1,l}, \quad k_{2,l}, \quad \cdots, \quad k_{l,l}, \quad 0, 0, 0, \cdots, 0). \end{cases} \tag{32}$$

i.e.

$$\begin{pmatrix} \hat{\boldsymbol{e}}_1' & \hat{\boldsymbol{e}}_2' & \cdots & \hat{\boldsymbol{e}}_n' \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{e}}_{i_1} & \hat{\boldsymbol{e}}_{i_2} & \cdots & \hat{\boldsymbol{e}}_{i_n} \end{pmatrix} \tag{33}$$

$$\begin{pmatrix} k_{1,l+1} & k_{1,l+2} & \cdots & k_{1,n} & k_{1,1} & k_{1,2} & \cdots & k_{1,l} \\ k_{2,l+1} & k_{2,l+2} & \cdots & k_{2,n} & k_{2,1} & k_{2,2} & \cdots & k_{2,l} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ k_{l,l+1} & k_{l,l+2} & \cdots & k_{l,n} & k_{l,1} & k_{l,2} & \cdots & k_{l,l} \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{34}$$

Then the transformation relation of coordinates is

$$\begin{pmatrix} \Delta x_{i_1} \\ \Delta x_{i_2} \\ \vdots \\ \Delta x_{i_n} \end{pmatrix} = \begin{pmatrix} k_{1,l+1} & k_{1,l+2} & \cdots & k_{1,n} & k_{1,1} & k_{1,2} & \cdots & k_{1,l} \\ k_{2,l+1} & k_{2,l+2} & \cdots & k_{2,n} & k_{2,1} & k_{2,2} & \cdots & k_{2,l} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ k_{l,l+1} & k_{l,l+2} & \cdots & k_{l,n} & k_{l,1} & k_{l,2} & \cdots & k_{l,l} \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \Delta x_1' \\ \Delta x_2' \\ \vdots \\ \Delta x_n' \end{pmatrix} \tag{35}$$

Making

$$K = \begin{pmatrix} k_{1,l+1} & k_{1,l+2} & \cdots & k_{1,n} & k_{1,1} & k_{1,2} & \cdots & k_{1,l} \\ k_{2,l+1} & k_{2,l+2} & \cdots & k_{2,n} & k_{2,1} & k_{2,2} & \cdots & k_{2,l} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ k_{l,l+1} & k_{l,l+2} & \cdots & k_{l,n} & k_{l,1} & k_{l,2} & \cdots & k_{l,l} \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{36}$$

we have

$$\Delta \boldsymbol{x} = T \begin{pmatrix} \Delta x_{i_1} \\ \Delta x_{i_2} \\ \vdots \\ \Delta x_{i_n} \end{pmatrix} = TK \begin{pmatrix} \Delta x_1' \\ \Delta x_2' \\ \vdots \\ \Delta x_n' \end{pmatrix} = TK\Delta \boldsymbol{x}' \tag{37}$$

It is obvious from the last section that the matrix TK is the matrix S we are looking for, in other words, we do the coordinate transformation

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = S^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{38}$$

then Hessian matrix of the function $f(x'_1, x'_2, \cdots, x'_n)$ in the new coordinate system is the H' we need in the last section.

Note that the satisfactory $\Delta \boldsymbol{x}'$ is not unique, because the first $n-l$ rows of K as linearly independent linear combinations can combine an equivalent linearly independent group to replace the first $n-l$ rows as a new $\Delta \boldsymbol{x}'$. On the other hand, the degree of freedom of the last l rows of K is very high when the vector group is extended. We don't even need all of the entries in the $n-l$ columns to be 0, as long as the final matrix is linearly independent.

5. Conclusions

In this paper, we obtain the sufficient conditions and necessary conditions for the extreme value problem with constraints.

The sufficient conditions for taking an extreme value are that the projection of the gradient of the point in the constraint surface is zero and the second-order partial derivative matrix on the local linear subspace in the constraint surface be positive or negative definite. The necessary conditions are that the projection of the gradient of the point in the constraint surface is zero and that the second-order partial derivative matrix on the local linear subspace in the constraint surface be semi-positive or semi-negative definite.

Finally, we discuss the construction of the local coordinate system and the micro base vector on the local linear subspace in the constraint surface to find the second-order partial derivative matrix on the local linear subspace in the constraint surface. This construction has a high degree of freedom.