tests for local extrema in Lagrange multiplier method

Canonical name	TestsForLocalExtremaInLagrangeMultiplierMethod
Date of creation	2013-03-22 15:28:52
Last modified on	2013-03-22 15:28:52
Owner	stevecheng (10074)
Last modified by	stevecheng (10074)
Numerical id	6
Author	stevecheng (10074)
Entry type	Result
Classification	msc 26B12
Classification	msc 49-00
Classification	msc 49K35
Related topic	HessianForm
Related topic	RelationsBetweenHessianMatrixAndLocalExtrema
Defines	projected Hessian
Defines	reduced Hessian

Let U be open in $\mathbb{R}^n$, and $f\colon U \to \mathbb{R}$, $g\colon U \to \mathbb{R}^m$ be twice continuously differentiable functions. Assume that $p \in U$ is a stationary point for f on $M = g^{-1}(\{0\})$, and $\mathrm{D}\, g$ has full rank everywhere[1] on M. Then we know that p is the solution to the Lagrange multiplier system

$$\mathrm{D}\, f(p) = \lambda \cdot \mathrm{D}\, g(p)\,, \tag{1}$$

for a Lagrange multiplier vector $\lambda = (\lambda_1, \dots, \lambda_m)$.

Our aim is to develop an analogue of the second derivative test for the stationary point p.

The most straightforward way to proceed is to consider a coordinate chart $\alpha\colon V \to M$ for the manifold M, and consider the Hessian of the function $f \circ \alpha\colon V \to \mathbb{R}^n$ at $0 = \alpha^{-1}(p)$. This Hessian is in fact just the Hessian form of $f\colon M \to \mathbb{R}$ expressed in the coordinates of the chart α. But the whole point of using Lagrange multipliers is to avoid calculating coordinate charts directly, so we find an equivalent expression for $\mathrm{D}^2(f \circ \alpha)(0)$ in terms of $\mathrm{D}^2 f(p)$ without mentioning derivatives of α.

To do this, we differentiate $f \circ \alpha$ twice using the chain rule and product rule[2]. To reduce clutter, from now on we use the prime notation for derivatives rather than D.

$$\begin{aligned}
(f \circ \alpha)''(0) &= ((f' \circ \alpha) \cdot \alpha')'(0) \\
&= \big((f'' \circ \alpha) \cdot \alpha' \cdot \alpha' + (f' \circ \alpha) \cdot \alpha''\big)(0) \\
&= f''(\alpha(0)) \cdot \alpha'(0) \cdot \alpha'(0) + f'(\alpha(0)) \cdot \alpha''(0) \\
&= f''(p) \cdot \alpha'(0) \cdot \alpha'(0) + f'(p) \cdot \alpha''(0)\,.
\end{aligned} \tag{2}$$

If we interpret $(f \circ \alpha)''(0)$ as a bilinear mapping of vectors $u, v \in \mathbb{R}^{n-m}$, then formula (**??**) really means

$$(f \circ \alpha)''(0) \cdot (u, v) = f''(p) \cdot \big(\alpha'(0) \cdot u, \alpha'(0) \cdot v\big) + f'(p) \cdot \big(\alpha''(0) \cdot (u, v)\big)\,. \tag{3}$$

To obtain the quadratic form, we set $v = u$; also we abbreviate the vector $\alpha'(0) \cdot u$ by h, which belongs to the tangent space $\mathrm{T}_p M$ of M at p. So,

$$(f \circ \alpha)''(0) \cdot u^2 = f''(p) \cdot h^2 + f'(p) \cdot \big(\alpha''(0) \cdot u^2\big)\,. \tag{4}$$

[1] Actually, only $\mathrm{D}\, g(p)$ needs to have full rank, and the arguments presented here continue to hold in that case, although M would not necessarily be a manifold then.

[2] Note that the "product" operation involved (second equality of (**??**)) is the operation of *composition of two linear mappings*. Think hard about this if you are not sure; it took me several tries to get this formula right, since multi-variable iterated derivatives have a complicated structure.

Naïvely, we might think that $(f \circ \alpha)''(0)$ is simply $f''(p)$ restricted to the tangent space $\mathrm{T}_p M$. This happens to be the first term in (**??**), but there is also an additional contribution by the second term involving $\alpha''(0)$; intuitively, $\alpha''(0)$ is the curvature of the surface (manifold) M, "changing the geometry" of the graph of f.

But the second term of (**??**) still involves α. To eliminate it, we differentiate the equation $g \circ \alpha = 0$ twice.

$$0 = (g \circ \alpha)''(0) = g''(p) \cdot \alpha'(0) \cdot \alpha'(0) + g'(p) \cdot \alpha''(0)\,. \tag{5}$$

(It is derived the same way as (**??**) but with f replaced by g.) Now we can substitute (**??**) and (**??**) in (**??**) to eliminate the term $f'(p) \cdot \alpha''(0)$:

$$\begin{aligned} (f \circ \alpha)''(0) &= f''(p) \cdot \alpha'(0) \cdot \alpha'(0) + \lambda \cdot g'(p) \cdot \alpha''(0) \\ &= f''(p) \cdot \alpha'(0) \cdot \alpha'(0) - \lambda \cdot g''(p) \cdot \alpha'(0) \cdot \alpha'(0)\,, \end{aligned} \tag{6}$$

or expressed as a quadratic form,

$$(f \circ \alpha)''(0) \cdot u^2 = f''(p) \cdot h^2 - \lambda \cdot g''(p) \cdot h^2\,. \tag{7}$$

Thus, to understand the nature of the stationary point p, we can study the modified Hessian:

$$f''(p) - \lambda \cdot g''(p)\,, \qquad \text{restricted to } \mathrm{T}_p M. \tag{8}$$

For example, if this bilinear form is positive definite, then p is a local minimum, and if it is negative definite, then p is a local maximum, and so on. All the tests that apply to the usual Hessian in $\mathbb{R}^n$ apply to the modified Hessian (**??**).

In coordinates of $\mathbb{R}^n$, the modified Hessian (**??**) takes the form

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left. \frac{\partial^2 f}{\partial x^i \partial x^j} \right|_p - \sum_{k=1}^{m} \lambda_k \left. \frac{\partial^2 g^k}{\partial x^i \partial x^j} \right|_p \right) h^i h^j\,. \tag{9}$$

We emphasize that the vector h can be restricted to lie in the tangent space $\mathrm{T}_p M$, when studying the stationary point p of f restricted to M.

In matrix form (**??**) can be written

$$B = \left[\frac{\partial^2 f}{\partial x^i \partial x^j} - \sum_{k=1}^{m} \lambda_k \frac{\partial^2 g^k}{\partial x^i \partial x^j} \right]_{ij}\,. \tag{10}$$

But again, the test vector h need only lie on $\mathrm{T}_p M$, so if we want to apply positive/negative definiteness tests for matrices, they should instead be applied to the *projected* or *reduced* Hessian:

$$Z^{\mathrm{t}} B Z \tag{11}$$

where the columns of the $n \times (n-m)$ matrix Z form a *basis* for $\mathrm{T}_p M = \ker g'(p) \subset \mathbb{R}^n$.