1 More on Optimization with Equality Constraints (June 2)

1.1 SONC and SOSC, Equality Constraints

Theorem 1.1. *(Second order necessary conditions for a local minimizer with equality constraints) Consider functions $f, h_1, \ldots, h_k$ which are C^2 on the open $\Omega \subseteq \mathbb{R}^n$. Suppose x_0 is a regular point of the constraints given by $h_1(x) = \cdots = h_k(x) = 0$, and that it is a local minimizer of f on $M = \bigcap h_i^{-1}(\{0\})$. Then*

1. *There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that*
$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) = 0.$$

2. *The Lagrangian*
$$L(x_0) = \nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)$$
is positive semi-definite on $T_{x_0}M$.

Proof. Let $x(s)$ be a smooth curve with $x(0) = 0$ in M. Recall that, by the product rule,

$$\begin{aligned} \frac{d}{ds} f(x(s)) &= \nabla f(x(s)) \cdot x'(s) \\ \frac{d^2}{ds^2} f(x(s)) &= x'(s) \cdot \nabla^2 f(x(s)) x'(s) + \nabla f(x(s)) \cdot x''(s). \end{aligned}$$

By the second order Taylor approximation, we have

$$0 \le f(x(s)) - f(x(0)) = s \left.\frac{d}{ds}\right|_{s=0} f(x(s)) + \frac{1}{2} s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} f(x(s)) + o(s^2).$$

This is, equivalently,

$$0 \le f(x(s)) - f(x(0)) = s \nabla f(x_0) \cdot \underbrace{x'(0)}_{\in T_{x_0}M} + \frac{1}{2} s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} f(x(s)) + o(s^2).$$

Since the gradient at a regular local minimizer is perpendicular to the tangent space there, the first-order term above vanishes. We have

$$0 \le \frac{1}{2} s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} f(x(s)) + o(s^2).$$

By the definition of M, we may write the above as

$$0 \le \frac{1}{2} s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} \left[f(x(s)) + \sum \lambda_i h_i(x(s)) \right] + o(s^2).$$

Or

$$0 \le \frac{1}{2}s^2 x'(0) \cdot \underbrace{\left(\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h(x_0)\right)}_{=L(x_0)} x'(0) + \frac{1}{2}s^2 \underbrace{\left(\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0)\right)}_{=0} \cdot x''(0) + o(s^2).$$

Divide by s^2:

$$0 \le \frac{1}{2}x'(0) \cdot L(x_0)x'(0) + \frac{o(s^2)}{s^2}.$$

By taking s small it follows that $0 \le \frac{1}{2}x'(0) \cdot L(x_0)x'(0)$. Since any tangent vector $v \in T_{x_0}M$ can be described as the tangent vector to a curve in M through x_0, it follows that $L(x_0)$ is positive semi-definite on $T_{x_0}M$. □

Theorem 1.2. *(Second order sufficient conditions for a local minimizer with equality constraints) Consider functions $f, h_1, \ldots, h_k$ which are C^2 on the open $\Omega \subseteq \mathbb{R}^n$. Suppose x_0 is a regular point of the constraints given by $h_1(x) = \cdots = h_k(x) = 0$. Let $M = \bigcap h_i^{-1}(\{0\})$. Suppose there exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that*

1. $$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) = 0$$

2. *The Lagrangian*
$$L(x_0) = \nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)$$
is positive definite on $T_{x_0}M$.

Then x_0 is a strict local minimizer of f on M.

Proof. Recall that if $L(x_0)$ is positive definite on $T_{x_0}M$, then there is an $a > 0$ such that $v \cdot L(x_0)v \ge a\|v\|^2$ for all $v \in T_{x_0}M$. (This is very easily proven by diagonalizing the matrix.) Let $x(s)$ be a smooth curve in M such that $x(0) = x_0$, and normalize the curve so that $\|x'(0)\| = 1$. We have which becomes

$$
\begin{aligned}
f(x(s)) - f(x(0)) &= s \left.\frac{d}{ds}\right|_{s=0} f(x(s)) + \frac{1}{2}s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} f(x(s)) + o(s^2) \\
&= s \left.\frac{d}{ds}\right|_{s=0} \Big[f(x(s)) + \sum \lambda_i h_i(x(s))\Big] + \frac{1}{2}s^2 \left.\frac{d^2}{ds^2}\right|_{s=0} \Big[f(x(s)) + \sum \lambda_i h_i(x(s))\Big] + o(s^2) \\
&= s \underbrace{[\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0)]}_{=0 \text{ by 1.}} \cdot x'(0) + \frac{1}{2}s^2 x'(0) \cdot L(x_0) x'(0) \\
&\qquad\qquad + \frac{1}{2}s^2 \underbrace{[\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0)]}_{=0 \text{ by 1.}} \cdot x''(0) + o(s^2) \\
&= \frac{1}{2}s^2 x'(0)^T L(x_0) x'(0) + o(s^2) \\
&\geq \frac{1}{2}s^2 a \|x'(0)\|^2 + o(s^2) \\
&= \frac{1}{2}s^2 a + o(s^2) \\
&= s^2 \left(\frac{1}{2}a + \frac{o(s^2)}{s^2}\right)
\end{aligned}
$$

For sufficiently small s, the above is positive, so $f(x(s)) > f(x_0)$ for all sufficiently small s. Since $x(s)$ was arbitrary, x_0 is a strict local minimizer of f on M. $\square$

1.2 Examples

1. Recall the box example: maximizing the volume of a box of sides $x, y, z \geq 0$ subject to a fixed surface area $A > 0$. We were really minimizing the negative of the volume. We got $(x_0, y_0, z_0) = (l, l, l)$, where $l = \sqrt{A/6}$. Our Lagrange multiplier was $\lambda = \frac{A}{8(x_0+y_0+z_0)} = \frac{A}{24l} > 0$. We had (after some calculation)

$$L(x_0, y_0, z_0) = (2\lambda - l) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

   Here, $2\lambda - l < 0$. We have

$$T_{(x_0,y_0,z_0)} M = \operatorname{span}(\nabla h(x_0, y_0, z_0))^{\perp} = \{(u, v, w) \in \mathbb{R}^3 : u + v + w = 0\},$$

since $\nabla h(x_0, y_0, z_0) = (4l, 4l, 4l)$. If $(u, v, w) \in T_{(x_0,y_0,z_0)}M$ is nonzero,

$$\begin{aligned}
\begin{pmatrix} u & v & w \end{pmatrix} (2\lambda - l) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} &= \begin{pmatrix} u & v & w \end{pmatrix} (2\lambda - l) \begin{pmatrix} v + w \\ u + w \\ u + v \end{pmatrix} \\
&= (2\lambda - l) \begin{pmatrix} u & v & w \end{pmatrix} \begin{pmatrix} -u \\ -v \\ -w \end{pmatrix} \\
&= -(2\lambda - l)(u^2 + v^2 + w^2) > 0,
\end{aligned}$$

so by the SOSC under equality constraints, our point (x_0, y_0, z_0) is a strict local maximizer of the volume. In fact, it is a strict global minimum (which is yet to be seen).

2. Consider the problem

$$\begin{aligned}
\text{minimize } & f(x, y) = x^2 - y^2 \\
\text{subject to } & h(x, y) = y = 0.
\end{aligned}$$

Then

$$\nabla f(x, y) + \lambda \nabla h(x, y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

implying that $\lambda = 0$ and that $(x, y) = (0, 0)$ is our candidate local minimizer. Since $\nabla h(x, y) \neq (0, 0)$, the candidate is a regular point. We have

$$L(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} + 0 \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix},$$

which is not positive semi-definite *everywhere*. What about on the tangent space $T_{(0,0)}(x\text{-axis}) = (x\text{-axis})$? Clearly it is positive definite on the x-axis, so by the SOSC that we just proved, $(0, 0)$ is a strict local minimizer of f on the x-axis. Thinking of level sets, this is intuitively true.

3. Consider the problem

$$\begin{aligned}
\text{minimize } & f(x, y) = (x - a)^2 + (y - b)^2 \\
\text{subject to } & h(x, y) = x^2 + y^2 - 1 = 0.
\end{aligned}$$

Let us assume that (a, b) satisfies $a^2 + b^2 > 1$. We have $\nabla h(x, y) = (2x, 2y)$, which is non-zero on S^1, implying that every point of S^1 is a regular point. Lagrange tells us that

$$\begin{pmatrix} 2(x - a) \\ 2(y - b) \end{pmatrix} + \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

as well as $x^2 + y^2 = 1$. This may be written

$$\begin{aligned}(1+\lambda)x &= a\\ (1+\lambda)y &= b\\ x^2 + y^2 &= 1\end{aligned}$$

By our assumption that $a^2 + b^2 > 1$, we have $\lambda \neq -1$. Therefore

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{1+\lambda}\begin{pmatrix} a \\ b \end{pmatrix},$$

which implies that

$$\frac{1}{1+\lambda} = \frac{1}{\sqrt{a^2+b^2}}$$

by the third equation. Therefore

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \frac{1}{\sqrt{a^2+b^2}}\begin{pmatrix} a \\ b \end{pmatrix}.$$

Thinking of level sets, this is intuitively true. The Lagrangian is

$$L(x_0, y_0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \lambda \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \underbrace{(1+\lambda)}_{>0}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

which, by the SOSC that we proved, proves that (x_0, y_0) is a strict local minimizer of f on S^1. In fact, this point is a global minimizer of f on S^1, which follows immediately by the fact that f necessarily takes on a global minimum on S^1 and that it only takes on the point (x_0, y_0).

4. For a special case, we will derive the Lagrange multipliers equation. Suppose we are working with C^1 functions f, h. Our problem is

$$\begin{aligned}&\text{minimize } f(x,y,z)\\ &\text{subject to } g(x,y,z) = z - h(x,y) = 0.\end{aligned}$$

That is, we are minimizing $f(x,y,z)$ on the graph Γ_h of h. The Lagrange equation tells us that

$$\nabla f(x,y,z) + \lambda g(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y,z) \\ \frac{\partial f}{\partial y}(x,y,z) \\ \frac{\partial f}{\partial z}(x,y,z) \end{pmatrix} + \lambda \begin{pmatrix} -\frac{\partial h}{\partial x}(x,y,z) \\ -\frac{\partial y}{\partial x}(x,y,z) \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We will derive the above formula by expressing it as an unconstrained minimization problem

$$\text{minimize }_{(x,y)\in\mathbb{R}^2} F(x,y)$$

for some function F. We will then find the first order necessary conditions for an unconstrained minimization, and then express it as the equation we would like to prove.

Define $F(x, y) = f(x, y, f(x, y))$. The constrained minimization problem is therefore equivalent to the unconstrained problem. By our theory of unconstrained minimization, $\nabla F(x_0, y_0) = (0, 0)$. That is,

$$\nabla F(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial h}{\partial x} \\ \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial h}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Rather,

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial h}{\partial x} = 0$$
$$\frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial h}{\partial y} = 0$$

Let $\lambda = -\frac{\partial f}{\partial z}$. The equation becomes

$$\frac{\partial f}{\partial x} - \lambda\frac{\partial h}{\partial x} = 0$$
$$\frac{\partial f}{\partial y} - \lambda\frac{\partial h}{\partial y} = 0$$
$$\frac{\partial f}{\partial z} + \lambda = 0$$

which is what we wanted.