

APM462 Course Notes

Kain Dineen

June 4, 2020

The following are lecture notes for APM462 (Nonlinear Optimization) offered in the Summer of 2020, taught by Jonathan Korman. These lecture notes were typed during lectures, and are not based off of any handwritten notes. These notes were created three weeks in to the course, and do not (as of now) include the material from the first two weeks.

Contents

1	Unconstrained Finite-Dimensional Optimization (May 19)	2
1.1	First Order Necessary Condition	2
1.2	Examples of using the FONC	3
1.3	Second Order Necessary Condition	4
1.4	Sylvester's Criterion	5
1.5	Examples of using the SONC	5
1.6	Completing the Square	6
2	Sufficient Condition for an Interior Local Minimizer (May 21)	7
2.1	A Sufficient Condition	7
2.2	Examples	7
3	Constrained Optimization (May 26)	9
3.1	Second Order Necessary Condition for a Local Minimizer	9
3.2	Optimization with Equality Constraints	9
4	Lagrange Multipliers (May 28)	11
4.1	First Order Necessary Condition for a Local Minimizer Under Equality Constraints .	11
4.2	The Box Example	11
4.3	Second Order Necessary Conditions for a Local Minimizer Under Equality Constraints	12
5	More on Optimization with Equality Constraints (June 2)	13
5.1	SONC and SOSC, Equality Constraints	13
5.2	Examples	15

6 Optimization under Inequality Constraints (June 4)	19
6.1 Kuhn-Tucker Conditions	19

1 Unconstrained Finite-Dimensional Optimization (May 19)

1.1 First Order Necessary Condition

Our main problem is

$$\min_{x \in \Omega} f(x) \quad f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R},$$

where Ω is one of the following three types:

- $\Omega = \mathbb{R}^n$.
- Ω open.
- Ω the closure of an open set.

We can consider minimization problems without any loss of generality, since any maximization problem can be converted to a minimization problem by taking the negative of the function in question: that is,

$$\max_{x \in \Omega} f(x) = \min_{x \in \Omega} -f(x).$$

Definition 1.1.1. Given $\Omega \subseteq \mathbb{R}^n$ and a point $x_0 \in \Omega$, we say that the vector $v \in \mathbb{R}^n$ is a feasible direction at x_0 if there is an $\bar{s} > 0$ such that $x_0 + sv \in \Omega$ for all $s \in [0, \bar{s}]$.

Theorem 1.1.1. (First order necessary condition for a local minimum, or FONC) Let $f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}$ be C^1 . If $x_0 \in \Omega$ is a local minimizer of f , then $\nabla f(x_0) \cdot v \geq 0$ for all feasible directions v at x_0 .

First we deduce a familiar case of the theorem - the one we know from second-year calculus.

Corollary 1.1.1. If $f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}$ is C^1 and x_0 is a local minimizer of f in the interior of Ω , then $\nabla f(x_0) = 0$.

Proof. If x_0 is an interior point of Ω , then all directions at x_0 are feasible. In particular, for any such v , we have $\nabla f(x_0) \cdot (v) \geq 0$ and $\nabla f(x_0) \cdot (-v) \geq 0$, which implies $\nabla f(x_0) = 0$ as all directions are feasible at x_0 . \square

Now we prove the theorem.

Proof. Reduce to a single-variable problem by defining $g(s) = f(x_0 + sv)$, where $s \geq 0$. Then 0 is a local minimizer of g . Taylor's theorem gives us

$$g(s) - g(0) = sg'(0) + o(s) = s\nabla f(x_0) \cdot v + o(s).$$

If $\nabla f(x_0) \cdot v < 0$, then for sufficiently small s the right side is negative. This implies that $g(s) < g(0)$ for those s , a contradiction. Therefore $\nabla f(x_0) \cdot v \geq 0$. \square

1.2 Examples of using the FONC

1. Consider the problem

$$\min_{x \in \Omega} f(x, y) = x^2 - xy + y^2 - 3y \quad \text{over } \Omega = \mathbb{R}^2.$$

By the corollary to the FONC, we want to find the points (x_0, y_0) where $\nabla f(x_0, y_0) = 0$. We have

$$\nabla f(x, y) = (2x - y, -x + 2y - 3),$$

so we want to solve

$$\begin{aligned} 2x - y &= 0 \\ -x + 2y &= 3, \end{aligned}$$

which has solution $(x_0, y_0) = (1, 2)$. Therefore $(1, 2)$ is the only *candidate* for a local minimizer. That is, if the function f has a local minimizer in \mathbb{R}^2 , then it must be $(1, 2)$.

It turns out that $(1, 2)$ is a global minimizer for f on $\Omega = \mathbb{R}^2$. By some work, we have

$$f(x, y) = \left(x - \frac{y}{2}\right)^2 + \frac{3}{4}(y - 2)^2 - 3.$$

In this form, it is obvious that a *global* minimizer occurs at the point where the squared terms are zero, if such a point exists. That point is $(1, 2)$.

2. Consider the problem

$$\min_{x \in \Omega} f(x, y) = x^2 - x + y + xy \quad \text{over } \Omega = \{(x, y) \in \mathbb{R}^2 : x, y \geq 0\}.$$

We have

$$\nabla f(x, y) = (2x + y - 1, x + 1).$$

To apply the FONC, we'll divide the feasible set Ω into four different regions. Suppose that (x_0, y_0) is a local minimizer of f on Ω .

- (i) (x_0, y_0) is an interior point:

By the corollary to the FONC, we must have $\nabla f(x_0, y_0) = 0$. Then $x_0 = -1$, which is not in the interior of Ω . This case fails.

- (ii) (x_0, y_0) on the positive x-axis:

Then we are considering $(x_0, 0)$. The feasible directions at $(x_0, 0)$ are those vectors $v \in \mathbb{R}^2$ with $v_2 \geq 0$. The FONC tells us that $\nabla f(x_0, 0) \cdot v \geq 0$ for all feasible directions v . We then have

$$(2x_0 - 1)v_1 + (x_0 + 1)v_2 \geq 0$$

for all v_1 and all $v_2 \geq 0$. In particular, this holds for $v_2 = 0$, so $(2x_0 - 1)v_1 \geq 0$ for all v_1 , implying $x_0 = 1/2$. Therefore $(1/2, 0)$ is a candidate for a local minimizer of f on Ω - this is the only candidate for a local minimizer of f on the positive x-axis.

(iii) (x_0, y_0) on the positive y -axis:

Then we are considering $(0, y_0)$. The feasible directions here are $v \in \mathbb{R}^2$ with $v_1 \geq 0$. Then we have

$$(y_0 - 1)v_1 + v_2 \geq 0$$

for any v_2 and $v_1 \geq 0$. This is a contradiction if we take $v_1 = 0$, so f has no local minimizers along the positive y -axis.

(iv) (x_0, y_0) is the origin:

Then we are considering $(0, 0)$. The feasible directions here are $v \in \mathbb{R}^2$ with $v_1, v_2 \geq 0$. Then we have

$$-v_1 + v_2 \geq 0$$

for all $v_1, v_2 \geq 0$, a contradiction. Therefore the origin is not a local minimizer of f .

We conclude that the only candidate for a local minimizer of f is $(1/2, 0)$. It turns out that this is actually a global minimizer of f on Ω . (This is to be seen.)

1.3 Second Order Necessary Condition

Theorem 1.3.1. (Second order necessary condition for a local minimum, or SONC) Let $f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}$ be C^2 . If $x_0 \in \Omega$ is a local minimizer of f , then for any feasible direction v at x_0 the following conditions hold:

(i) $\nabla f(x_0) \cdot v \geq 0$.

(ii) If $\nabla f(x_0) \cdot v = 0$, then $v^T \nabla^2 f(x_0) v \geq 0$.

Proof. Fix a feasible direction v at x_0 . Then $f(x_0) \leq f(x_0 + sv)$ for sufficiently small s . By Taylor's theorem,

$$f(x_0 + sv) = f(x_0) + s \nabla f(x_0) \cdot v + \frac{1}{2} s^2 v^T \nabla^2 f(x_0) v + o(s^2),$$

so by the FONC,

$$f(x_0 + sv) - f(x_0) = \frac{1}{2} s^2 v^T \nabla^2 f(x_0) v + o(s^2).$$

If $v^T \nabla^2 f(x_0) v < 0$, then for sufficiently small s the right side is negative, implying that $f(x_0 + sv) < f(x_0)$ for such s , which contradicts local minimality of $f(x_0)$. Therefore $v^T \nabla^2 f(x_0) v \geq 0$. \square

Corollary 1.3.1. If $f : \mathbb{R}^n \supseteq \Omega \rightarrow \mathbb{R}$ is C^2 and x_0 is a local minimizer of f in the interior of Ω , then the following conditions hold:

(i) $\nabla f(x_0) = 0$.

(ii) $\nabla^2 f(x_0)$ is positive semidefinite.

1.4 Sylvester's Criterion

Here's a useful criterion for determining when a matrix is positive definite or positive semidefinite.

Definition 1.4.1. *A principal minor of a square matrix A is the determinant of a submatrix of A obtained by removing any k rows and the corresponding k columns, $k \geq 0$. A leading principal minor of A is the determinant of a submatrix obtained by removing the last k rows and k columns of A , $k \geq 0$.*

Theorem 1.4.1. *(Sylvester's criterion for positive definite self-adjoint matrices) If A is a self-adjoint matrix, then $A \succ 0$ if and only if all of the leading principal minors of A are positive.*

Theorem 1.4.2. *(Sylvester's criterion for positive semidefinite self-adjoint matrices) If A is a self-adjoint matrix, then $A \succeq 0$ if and only if all of the principal minors of A are non-negative.*

1.5 Examples of using the SONC

1. Consider the problem

$$\min_{x \in \Omega} f(x, y) = x^2 - xy + y^2 - 3y \quad \text{over } \Omega = \mathbb{R}^2.$$

Recall that $(1, 2)$ was the only candidate for a local minimizer of f on Ω . We now check that the SONC holds. Since $(1, 2)$ is an interior point of Ω , we must have $\nabla^2 f(1, 2) \succeq 0$. We have

$$\nabla^2 f(1, 2) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

All of the leading principal minors of $\nabla^2 f(1, 2)$ are positive, so $(1, 2)$ satisfies the SONC by Sylvester's criterion.

2. Consider the problem

$$\min_{x \in \Omega} f(x, y) = x^2 - x + y + xy \quad \text{over } \Omega = \{(x, y) \in \mathbb{R}^2 : x, y \geq 0\}.$$

Recall that $(1/2, 0)$ was the only candidate for a local minimizer of f . We have

$$\nabla^2 f(1/2, 0) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}.$$

To satisfy the SONC, we must have

$$v^T \nabla^2 f(1/2, 0) v \geq 0$$

for all feasible directions v at $(1/2, 0)$ such that $\nabla f(1/2, 0) \cdot v = 0$. We have

$$\nabla f(1/2, 0) = (0, 3/2),$$

so if $v = (v_1, 0)$, then v is a feasible direction at $(1/2, 0)$ with $\nabla f(1/2, 0) \cdot v = 0$. Then

$$v^T \nabla^2 f(1/2, 0) v = (v_1 \ 0) \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ 0 \end{pmatrix} = (v_1 \ 0) \begin{pmatrix} 2v_1 \\ v_1 \end{pmatrix} = 2v_1^2 \geq 0.$$

So the SONC is satisfied.

1.6 Completing the Square

Let A be a symmetric positive definite $n \times n$ matrix. Our problem is

$$\min_{x \in \Omega} f(x) = \frac{1}{2} x^T A x - b \cdot x \quad \text{over } \Omega = \mathbb{R}^n.$$

The FONC tells us that if x_0 is a local minimizer of f , then since x_0 is an interior point, $\nabla f(x_0) = 0$. We thus have $Ax_0 = b$, so since A is invertible (positive eigenvalues), $x_0 = A^{-1}b$. Therefore $x_0 = A^{-1}b$ is the *unique* candidate for a local minimizer of f on Ω .

The SONC then tells us that $\nabla^2 f(x_0) = A$, so that $\nabla^2 f(x_0) \succ 0$, implying that $x_0 = A^{-1}b$ is a candidate for a local minimizer of f on Ω .

In fact, the candidate x_0 is a global minimizer. Why? We will "complete the square". We can write

$$f(x) = \frac{1}{2} x^T A x - b \cdot x = \frac{1}{2} (x - x_0)^T A (x - x_0) - \frac{1}{2} x_0^T A x_0;$$

this relies on symmetry. (Long rearranging of terms.) In this form it is obvious that x_0 is a global minimizer of f over Ω .

2 Sufficient Condition for an Interior Local Minimizer (May 21)

2.1 A Sufficient Condition

Lemma 2.1.1. *If A is symmetric and positive-definite, then there is an $a > 0$ such that $v^T A v \geq a \|v\|^2$ for all v .*

Proof. There is an orthogonal matrix Q with $Q^T A Q = \text{diag}(\lambda_1, \dots, \lambda_n)$. If $v = Qw$,

$$\begin{aligned} v^T A v &= (Qw)^T A Qw \\ &= w^T (Q^T A Q) w \\ &= \lambda_1 w_1^2 + \dots + \lambda_n w_n^2 \\ &\geq \min\{\lambda_1, \dots, \lambda_n\} \|w\|^2 \\ &= \min\{\lambda_1, \dots, \lambda_n\} \|v\|^2 \quad \text{since } Q \text{ is orthogonal} \end{aligned}$$

Since A is positive-definite, every eigenvalue is positive and we are done. \square

Theorem 2.1.1. *(Second order sufficient conditions for interior local minimizers) Let f be C^2 on $\Omega \subseteq \mathbb{R}^n$, and let x_0 be an interior point of Ω such that $\nabla f(x_0) = 0$ and $\nabla^2 f(x_0) \succ 0$. Then x_0 is a strict local minimizer of f .*

Proof. The condition $\nabla^2 f(x_0) \succ 0$ implies there is an $a > 0$ such that $v^T \nabla^2 f(x_0) v \geq a \cdot \|v\|^2$ for all v . By Taylor's theorem we have

$$f(x_0 + v) - f(x_0) = \frac{1}{2} v^T \nabla^2 f(x_0) v + o(\|v\|^2) \geq \frac{1}{2} a \|v\|^2 + o(\|v\|^2) = \|v\|^2 \left(\frac{a}{2} + \frac{o(\|v\|^2)}{\|v\|^2} \right).$$

For sufficiently small v the right hand side is positive, so $f(x_0 + v) > f(x_0)$ for all such v . Therefore x_0 is a strict local minimizer of f on Ω . \square

2.2 Examples

- (i) Consider $f(x, y) = xy$. The gradient is $\nabla f(x, y) = (y, x)$ and the Hessian is

$$\nabla^2 f(x, y) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Suppose we want to minimize f on all of $\Omega = \mathbb{R}^2$. By the FONC, the only candidate for a local minimizer is $(0, 0)$. The Hessian's eigenvalues are ± 1 , so it is not positive definite. We conclude by the SONC that the origin is not a local minimizer of f .

- (ii) Consider the same function $f(x, y) = xy$ on $\Omega = \{(x, y) \in \mathbb{R}^2, x, y \geq 0\}$. We claim that every point of the boundary of Ω is a local minimizer of f .

Consider $(x, 0)$ with $x > 0$. The feasible directions here are v with $v_2 \geq 0$. The FONC tells us that $\nabla f(x, 0) \cdot v \geq 0$. This dot product is $xv_2 \geq 0$, so $(x, 0)$ satisfies the FONC. Therefore every point on the positive x-axis is a candidate for a local minimizer. As for the SONC, $\nabla f(x, 0) \cdot v = xv_2 = 0$ if and only if $v_2 = 0$. Then $v^T \nabla^2 f(x, 0) v = 0$. Of course, this tells us nothing; we need a sufficient condition that works for boundary points. That's for next lecture.

Or, you could just say that $f = 0$ on the boundary of Ω and is positive on the interior, so every point of the boundary of Ω is a local minimizer (not strict) of f .

3 Constrained Optimization (May 26)

Consider the following minimization problem:

$$\begin{aligned} & \text{minimize } f(x, y) = xy \\ & \text{subject to } x^2 + y^2 \leq 1 \end{aligned}$$

Let Ω be the feasible set. The feasible directions at a point $(x_0, y_0) \in \Omega$ are the $(v, w) \in \mathbb{R}^2$ such that $(v, w) \cdot (x_0, y_0) < 0$, or $vx_0 + wy_0 < 0$. By the FONC for a minimizer, $\nabla f(x_0, y_0) \cdot (v, w) \geq 0$, so $wx_0 + vy_0 \geq 0$. Note that a local minimum must occur on the boundary. (Why?) We have three cases, depending on the sign of $x_0 + y_0$.

- (i) $x_0 + y_0 < 0$: can't occur
- (ii) $x_0 + y_0 > 0$: can't occur
- (iii) $x_0 + y_0 = 0$: good!

(This part could not be finished as attention had to be diverted from the lecture.)

3.1 Second Order Necessary Condition for a Local Minimizer

Theorem 3.1.1. (Second order sufficient condition for a local minimizer) Let f be C^2 on $\Omega \subseteq \mathbb{R}^n$ and suppose $x_0 \in \Omega$ satisfies

- (i) $\nabla f(x_0) \cdot v \geq 0$ for all feasible directions v at x_0 ,
- (ii) if $\nabla f(x_0) \cdot v = 0$ for some such v , then $v^T \nabla^2 f(x_0) v > 0$.

Then x_0 is a local minimizer of f on Ω .

3.2 Optimization with Equality Constraints

Consider the minimization problem

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

Suppose (x_0, y_0) is a local minimizer. Two cases:

1. $\nabla f(x_0, y_0) \neq 0$: we claim that $\nabla f(x_0, y_0)$ is perpendicular to the tangent space to the unit circle $h^{-1}(\{0\})$ at (x_0, y_0) . If this is not the case, then we obtain a contradiction by looking at the level sets of f , to which ∇f is perpendicular. Therefore $\nabla f(x_0, y_0) = \lambda \nabla h(x_0, y_0)$ for some λ .
2. $\nabla f(x_0, y_0) = 0$: as in the previous case, $\lambda = 0$.

In either case, at a local minimizer, the gradient of the function to be minimized is parallel to the gradient of the constraints.

We now recall some elementary differential geometry.

Definition 3.2.1. For us, a surface is the set of common zeroes of a finite set of C^1 functions.

Definition 3.2.2. For us, a differentiable curve on the surface $M \subseteq \mathbb{R}^n$ is the image of a C^1 function $x : (a, b) \rightarrow M$.

Definition 3.2.3. Let $x(s)$ be a differentiable curve on M that passes through $x_0 \in M$ at time $x(0) = x_0$. The velocity vector $v = \frac{d}{ds}\big|_{s=0} x(s)$ of $x(s)$ at x_0 is, for us, said to be a tangent vector to the surface M at x_0 . The set of all tangent vectors to M at x_0 is called the tangent space to M at x_0 and is denoted by $T_{x_0}M$.

Definition 3.2.4. Let $M = \{x \in \mathbb{R}^n : h_1(x) = \cdots = h_k(x) = 0\}$ be a surface. If $\nabla h_1(x_0), \dots, \nabla h_k(x_0)$ are all linearly independent, then x_0 is said to be a regular point of M .

Theorem 3.2.1. At a regular point $x_0 \in M$, the tangent space $T_{x_0}M$ is given by

$$T_{x_0}M = \{y \in \mathbb{R}^n : \nabla \mathbf{h}(x_0)y = 0\}.$$

Proof. It's in the book. Use the implicit function theorem. □

Lemma 3.2.1. Let f, h_1, \dots, h_k be C^1 functions on the open set $\Omega \subseteq \mathbb{R}^n$. Let $x_0 \in M = \{x \in \Omega : h_1(x) = \cdots = h_k(x) = 0\}$. Suppose x_0 is a local minimizer of f subject to the constraints $h_i(x) = 0$. Then $\nabla f(x_0)$ is perpendicular to $T_{x_0}M$.

Proof. Without loss of generality, suppose $\Omega = \mathbb{R}^n$. Let $v \in T_{x_0}M$. Then $v = \frac{d}{ds}\big|_{s=0} x(s)$ for some differentiable curve $x(s)$ in M with $x(0) = x_0$. Since x_0 is a local minimizer of f , 0 is a local minimizer of $f \circ x$, so $\nabla f(x_0) \cdot x'(0) = \nabla f(x_0) \cdot v = 0$. □

4 Lagrange Multipliers (May 28)

4.1 First Order Necessary Condition for a Local Minimizer Under Equality Constraints

Here is the first order necessary condition for a local minimizer under equality constraints.

Theorem 4.1.1. (*Lagrange multipliers*) Let f, h_1, \dots, h_k be C^1 functions on some open $\Omega \subseteq \mathbb{R}^n$. Suppose x_0 is a local minimizer of f subject to the constraints $h_1(x), \dots, h_k(x) = 0$, which is also a regular point of these constraints. Then there are $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ ("Lagrange multipliers") such that

$$\nabla f(x_0) + \lambda_1 \nabla h_1(x_0) + \dots + \lambda_k \nabla h_k(x_0) = 0.$$

Proof. Since x_0 is regular, $T_{x_0}M = \text{span}(\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\})^\perp$. By a lemma from last class, $\nabla f(x_0) \in (T_{x_0}M)^\perp$. Therefore $\nabla f(x_0) \in \text{span}(\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\})$, since we are dealing with a finite dimensional vector space. We are done. \square

4.2 The Box Example

Given a fixed area $A > 0$, how do we construct a box of maximum volume with surface area A ? Suppose the volume is $V(x, y, z) = xyz$ and the area is $A(x, y, z) = 2(xy + xz + yz)$. Our problem is stated as a maximization problem, so we have to convert it to a minimization problem. Let $f = -V$. We are therefore dealing with the problem

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

But we don't know how to deal with inequality constraints right now, so we have to make some changes. Note that if any one of x, y, z is zero, then the volume is zero. Therefore the problem we want to consider is really the problem

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

Now, if $\Omega = \{(x, y, z) \in \mathbb{R}^3 : x, y, z > 0\}$, then the above minimization problem may be solved using the first order necessary condition we gave above, for the set Ω is open.

Suppose (x_0, y_0, z_0) is a local minimizer of f subject to the constraint $h(x, y, z) = 0$. This point is regular because we are only considering points whose coordinates are all positive. Then there is a $\lambda \in \mathbb{R}$ such that $\nabla f(x_0, y_0, z_0) + \lambda \nabla h(x_0, y_0, z_0) = 0$. Therefore

$$(-y_0 z_0, -x_0 z_0, -x_0 y_0) + \lambda(2y_0 + 2z_0, 2x_0 + 2z_0, 2x_0 + 2y_0) = (0, 0, 0).$$

Equivalently,

$$\begin{aligned} 2\lambda(y_0 + z_0) &= y_0 z_0 \\ 2\lambda(x_0 + z_0) &= x_0 z_0 \\ 2\lambda(x_0 + y_0) &= x_0 y_0 \end{aligned}$$

Add all of these equations together:

$$2\lambda(2x_0 + 2y_0 + 2z_0) = x_0 z_0 + x_0 y_0 + y_0 z_0 = \frac{A}{2} > 0$$

implying that $\lambda > 0$. The first two equations tell us that

$$\begin{aligned} 2\lambda x_0(y_0 + z_0) &= x_0 y_0 z_0 \\ 2\lambda y_0(x_0 + z_0) &= x_0 y_0 z_0. \end{aligned}$$

Subtracting these two equations gives $2\lambda(x_0 z_0 - y_0 z_0) = 0$. Cancelling the z_0 's gives $2\lambda(x_0 - y_0) = 0$, and since $\lambda > 0$, we have $x_0 = y_0$. Since we could have done the same thing with the other pairs of equations, we get $x_0 = y_0 = z_0$.

Physically, this tells us that in order to maximize the volume of a rectangular solid of fixed area, we must make a cube. Note that we haven't actually solved the maximization problem; we've only figured out what form its solutions must take.

4.3 Second Order Necessary Conditions for a Local Minimizer Under Equality Constraints

Theorem 4.3.1. *Let f, h_1, \dots, h_k be C^2 on some open set $\Omega \subseteq \mathbb{R}^n$. Suppose x_0 is a regular point which is a local minimizer of f subject to the constraints. Then*

(i) *There are $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ such that*

$$\nabla f(x_0) + \lambda_1 \nabla h_1(x_0) + \dots + \lambda_k \nabla h_k(x_0) = 0.$$

(ii) *The "Lagrangian"*

$$L(x_0) = \nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)$$

is positive semi-definite on the tangent space $T_{x_0}M$, where $M = h_1^{-1}(\{0\}) \cap \dots \cap h_k^{-1}(\{0\})$.

5 More on Optimization with Equality Constraints (June 2)

5.1 SONC and SOSC, Equality Constraints

Theorem 5.1.1. *(Second order necessary conditions for a local minimizer with equality constraints)* Consider functions f, h_1, \dots, 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 $h_1(x) = \dots = h_k(x) = 0$, and that it is a local minimizer of f on $M = \bigcap h_i^{-1}(\{0\})$. Then

1. There exist $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ 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

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

This is, equivalently,

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

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

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

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

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

Or

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

Divide by s^2 :

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

By taking s small it follows that $0 \leq \frac{1}{2}x'(0) \cdot L(x_0)x'(0)$. 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 $T_{x_0}M$. \square

Theorem 5.1.2. *(Second order sufficient conditions for a local minimizer with equality constraints) Consider functions f, h_1, \dots, 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 $h_1(x) = \dots = h_k(x) = 0$. Let $M = \bigcap h_i^{-1}(\{0\})$. Suppose there exist $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ 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 $v \cdot L(x_0)v \geq a\|v\|^2$ 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

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

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

5.2 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

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

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,

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

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

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

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,

$$\begin{aligned} \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z} \frac{\partial h}{\partial x} &= 0 \\ \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z} \frac{\partial h}{\partial y} &= 0 \end{aligned}$$

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

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

which is what we wanted.

6 Optimization under Inequality Constraints (June 4)

6.1 Kuhn-Tucker Conditions

Our problem is of the form

$$\begin{aligned} & \text{minimize } f(x) \\ & \text{subject to } h_1(x) = \cdots = h_k(x) = 0 \\ & \quad g_1(x), \dots, g_l(x) \leq 0. \end{aligned}$$

Definition 6.1.1. Let x_0 satisfy the above constraints. We call the inequality constraint $g_i(x) \leq 0$ active at x_0 if $g_i(x_0) = 0$. Otherwise, it is inactive at x_0 .

Since we are only studying local properties of functions, we will only be concerned with active constraints.

Definition 6.1.2. Suppose there is an index $l' \leq l$ such that $g_1(x_0) = \dots, g_{l'}(x_0) = 0$ are active, and $g_{l'+1}(x_0) \leq 0, \dots, g_l(x_0) \leq 0$ are inactive. We say that x_0 is a regular point of these constraints if the vectors $\nabla h_1(x_0), \dots, \nabla h_k(x_0), \nabla g_1(x_0), \dots, \nabla g_{l'}(x_0)$ are linearly independent.

Theorem 6.1.1. (First order necessary conditions for minimizers under inequality constraints) Let $\Omega \subseteq \mathbb{R}^n$ be open and consider C^1 functions $f, h_1, \dots, h_k, g_1, \dots, g_l$ on Ω . Suppose x_0 is a local minimizer of f subject to the constraints, and that x_0 is regular as defined above. Then

(i) There exist $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ and $\mu_1, \dots, \mu_l \in \mathbb{R}^{\geq 0}$ such that

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

(ii) ("Complementary slackness conditions") For all j , $\mu_j g_j(x_0) = 0$, or equivalently, $\sum \mu_j g_j(x_0) = 0$.

These conditions are also known as the *Kuhn-Tucker conditions*.

Suppose the active constraints at x_0 are the first l' constraints. Since each $\mu_j \geq 0$, condition (ii) is equivalent to saying that if $j \geq l' + 1$, then $\mu_j = 0$.

Proof. If x_0 is a local minimizer of f subject to the constraints, then it is certainly a local minimizer of f subject to only the active constraints. That is, x_0 is also a local minimizer of f subject to the equality constraints

$$h_1(x) = \cdots = h_k(x) = g_1(x) = \cdots = g_{l'}(x) = 0.$$

□

We know how to work with this! Let M be the surface defined by these equality constraints. By the Lagrange multipliers theorem,

$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \sum \mu_j \nabla g_j(x_0) = 0,$$

for some $\lambda_i \in \mathbb{R}, \mu_j \in \mathbb{R}$. (Note that we have not yet shown that the μ_j 's are non-negative.)

Note that $g_1(x_0) = \dots = g_{l'}(x_0) = 0$. Therefore $\mu_{l'+1} = \dots = \mu_l = 0$, so it follows that

$$\mu_1 g_1(x_0) = 0, \dots, \mu_{l'} g_{l'}(x_0) = 0,$$

which implies that $\mu_j g_j(x_0) = 0$ for all j . We have proven condition (ii).

We must now verify the non-negativity of the μ_j 's. Suppose for the sake of contradiction that some $\mu_j < 0$; WLOG assume $j = 1$. Let

$$\tilde{M} = \{x \in \Omega : h_i(x) = 0, g_i(x) = 0, i \neq 1\}.$$

Since x_0 is a regular point of M , x_0 is a regular point of \tilde{M} . Therefore

$$T_{x_0} \tilde{M} = \text{span}(\{\nabla h_1(x_0), \dots, \nabla h_k(x_0), \nabla g_2(x_0), \dots, \nabla g_l(x_0)\})^\perp.$$

The vector $\nabla g_1(x_0)$ does not lie in this span, so there is a $v \in T_{x_0} \tilde{M}$ such that $\nabla g_1(x_0) \cdot v < 0$. That is, g_1 is strictly decreasing in the direction of v , or in more precise language, $g_1(x_0 + sv) < g_1(x_0)$ for all sufficiently small s , as we have

$$\left. \frac{d}{ds} \right|_{s=0} g_1(x_0 + sv) = \nabla g_1(x_0) \cdot v < 0.$$

Therefore v is a feasible direction for $g_1(x) \leq 0$ at x_0 , and also, v is tangential to the other constraints. Since x_0 is a regular point of \tilde{M} , we may find a curve $x(s)$ on \tilde{M} such that $x(0) = x_0$ and $x'(0) = v$. Also, $s = 0$ is a local minimizer of $f \circ x$, so

$$\left. \frac{d}{ds} \right|_{s=0} f(x(s)) = \nabla f(x_0) \cdot v \geq 0.$$

On the other hand,

$$\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \mu_1 \nabla g_1(x_0) + \sum_{j=2}^{l'} \mu_j \nabla g_j(x_0) = 0.$$

Taking the dot product of the above equation by v kills the two sums above and gives

$$\nabla f(x_0) \cdot v + \mu_1 \nabla g_1(x_0) \cdot v = 0,$$

implying $\nabla f(x_0) \cdot v < 0$, a contradiction. So every $\mu_j \geq 0$.