APM462 Course Notes

Kain Dineen

July 23, 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)

1.1 First Order Necessary Condition

Our main problem is

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

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 $\overline{s} > 0$ such that $x_0 + sv \in \Omega$ for all $s \in [0, \overline{s}]$.*

Theorem 1.1.1. *(First order necessary condition for a local minimum, or FONC) Let $f : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}$ 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 $f : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}$ 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. □

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 $\nabla f(x_0) \cdot v \geq 0$. □

1.2 Examples of using the FONC

1. Consider the problem

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

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

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

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 $f : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}$ 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,

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

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 $v^T \nabla^2 f(x_0) \geq 0$. □

Corollary 1.3.1. *If $f : \mathbb{R}^n \supseteq \Omega \to \mathbb{R}$ 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

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

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

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

Recall that $(1/2, 0)$ was the only candidate for a local minizer 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

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

So the SONC is satisfied.

1.6 Completing the Square

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

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

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 $v^T Av \geq a\|v\|^2$ for all v.*

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

$$\begin{aligned} v^T Av &= (Qw)^T AQw \\ &= w^T (Q^T AQ)w \\ &= \lambda_1 w_1^2 + \cdots + \lambda_n w_n^2 \\ &\geq \min\{\lambda_1, \ldots, \lambda_n\}\|w\|^2 \\ &= \min\{\lambda_1, \ldots, \lambda_n\}\|v\|^2 \qquad \text{since } Q \text{ is orthogonal} \end{aligned}$$

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

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. □

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 $x : (a, b) \to M$.*

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 $\nabla h_1(x_0), \ldots, \nabla h_k(x_0)$ 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 $f, h_1, \ldots, h_k$ 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 $f, h_1, \ldots, h_k$ 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 $h_1(x), \ldots, h_k(x) = 0$, which is also a regular point of these constraints. Then there are $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ ("Lagrange multipliers") such that*

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

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

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 $f, h_1, \ldots, h_k$ 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 $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that

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

(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 $M = h_1^{-1}(\{0\}) \cap \cdots \cap h_k^{-1}(\{0\})$.

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 $h_1(x) = \cdots = h_k(x) = 0$, and that it is a local minimizer of f on $M = \bigcap h_i^{-1}(\{0\})$. Then*

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

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

This is, equivalently,

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

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

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

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

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

Or

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

Divide by s^2:

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

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 $T_{x_0}M$. □

Theorem 5.1.2. *(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 $v \cdot L(x_0)v \geq a\|v\|^2$ 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

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

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. □

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

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

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,

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

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

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

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 $g_1(x_0) =, \ldots, g_{l'}(x_0) = 0$ are active, and $g_{l'+1}(x_0) \leq 0, \ldots, g_l(x_0) \leq 0$ are inactive. We say that x_0 is a regular point of these constraints if the vectors $\nabla h_1(x_0), \ldots, \nabla h_k(x_0), \nabla g_1(x_0), \ldots, \nabla g_{l'}(x_0)$ 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 $f, h_1, \ldots, h_k, g_1, \ldots, g_l$ 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 $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $\mu_1, \ldots, \mu_l \in \mathbb{R}^{\geq 0}$ 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

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

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

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

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

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

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

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

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 $\mu_j \geq 0$. □

7 More on Inequality Constraints (June 9)

As before, we are working on an open $\Omega \subseteq \mathbb{R}^n$, and we want to optimize f subject to $h_1, \ldots, h_k = 0$ and $g_1, \ldots, g_l \leq 0$. The smoothness of our functions varies.

7.1 Second Order Conditions

One might guess that the second order conditions under inequality constraints will be the same thing as before. However, the tangent space on which we evaluate the positive-definiteness of the Lagrangian is slightly different (in a very obvious way).

Theorem 7.1.1. *Suppose $f, h_1, \ldots, h_k, g_1, \ldots, g_k \in C^2(\Omega)$, where $\Omega \subseteq \mathbb{R}^n$. Suppose x_0 is a regular point of the constraints. If x_0 is a local minimizer of f subject to the constraints, then*

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

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

and $\mu_j g_j(x_0) = 0$ for each j.

(ii) The matrix

$$L(x_0) = \nabla^2 f(x_0) + \sum_i \lambda_i \nabla^2 h_i(x_0) + \sum_j \mu_j \nabla^2 g_j(x_0)$$

is positive semi-definite on the tangent space $T_{x_0}\tilde{M}$ to the active constraints at x_0. (Explicitly, $L(x_0)$ is positive semi-definite on the space

$$T_{x_0}\tilde{M} = \{v \in \mathbb{R}^n : \nabla h_i(x_0) \cdot v = 0 \text{ for all } i, \text{ and } \nabla g_j(x_0) \cdot v = 0 \text{ for all } 1 \leq j \leq l'\},$$

where the active g constraints are indexed precisely by $1, \ldots, l'$.)

Proof. x_0 is a local minimizer of f subject to the constraints, so it is also a local minimizer of f subject to only the active constraints. Since the Lagrange multiplies of the inactive constraints are zero, our theory of equality-constrained minimization finishes the problem. □

1. Consider, for example, the problem

$$\begin{aligned} \text{minimize } & f(x,y) := -x \\ \text{subject to } & g_1(x,y) := x^2 + y^2 \leq 1 \\ & g_2(x,y) := y + x - 1 \leq 0 \end{aligned}$$

The feasible set is the closed unit ball $\overline{B_1(0)}$ with an open semicircle removed from the top right. Geometrically, it is clear that the minimizer should be the point $(1, 0)$. It is not hard

to check that every feasible point is regular. Let's check that $(x_0, y_0) = (1, 0)$ satisfies the first order conditions. We look at

$$\nabla f(x_0, y_0) + \mu_1 \nabla g_1(x_0, y_0) + \mu_2 \nabla g_2(x_0, y_0) = (0, 0).$$

This becomes

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\begin{aligned} 2\mu_1 + \mu_2 &= 1 \\ \mu_2 &= 0. \end{aligned}$$

So $\mu_1 = 1/2$. Also, $g_1(1, 0) = 1^2 + 0^2 - 1 = 0$ and $g_(1, 0) = 0$ as well, so the complementary slackness conditions are satisfied. Therefore $(1, 0)$ satisfies the Kuhn-Tucker conditions, and so it is a candidate local minimizer. What about the second order conditions?

$$L(1, 0) = \nabla^2 f(x_0, y_0) + \mu_1 \nabla^2 g_1(x_0, y_0) + \mu_2 \nabla^2 g_2(x_0, y_0),$$

or

$$L(1, 0) = I.$$

Clearly the second order necessary conditions are satisfied, but let's check the tangent space anyway. We have $\nabla g_1(1, 0) = (2, 0)$ and $\nabla g_2(1, 0) = (1, 1)$; they are linearly independent, so the tangent space is a point. Therefore the second order necessary conditions are satisfied.

2. Consider the problem

$$\begin{aligned} \text{minimize } & f(x, y) := 2x^2 + 2xy + y^2 - 10x - 10y \\ \text{subject to } & g_1(x, y) = x^2 + y^2 - 5 \leq 0 \\ & g_2(x, y) := 3x + y - 6 \leq 0 \end{aligned}$$

The Kuhn-Tucker conditions are

$$\begin{pmatrix} 4x + 2y - 10 \\ 2x + 2y - 10 \end{pmatrix} + \mu_1 \begin{pmatrix} 2x \\ 2y \end{pmatrix} + \mu_2 \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

as well as $\mu_1, \mu_2 \geq 0$ and $\mu_1(x^2 + y^2 - 5) = 0$ and $\mu_2(3x + y - 6) = 0$. We consider four cases:

(i) Suppose g_1 is inactive and g_2 is active. Then $\mu_1 = 0$. The equations become

$$\begin{aligned} 4x + 2y - 10 + 3\mu_2 &= 0 \\ 2x + 2y - 10 + \mu_2 &= 0 \\ \mu_2(3x + y - 6) &= 0 \end{aligned}$$

Subtracting the second equation from the first gives $x = -\mu_2$. If $\mu_2 = 0$, then $x = 0$, which implies that $y = 6$ since the second constraint is active. We get the point $(0, 6)$; but this does not satisfy the constraints. Therefore $\mu_2 \neq 0$. (After some work one can conclude that no such point here satisfies the constraints.)

(ii) Suppose g_1 is active and g_2 is inactive. Then

$$\begin{aligned} 4x + 2y - 10 + 2\mu_1 x &= 0 \\ 2x + 2y - 10 + 2\mu_1 y &= 0 \\ \mu_1(x^2 + y^2 - 5) &= 0 \\ \mu_1 &\geq 0 \end{aligned}$$

The solution is $(1, 2)$ and $\mu_1 = 1$. It is not hard to see that this point is regular. Therefore the point $(1, 2)$ is a candidate. The Lagrangian is, after some work,

$$L(1, 2) = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}.$$

This matrix is clearly positive definite, so we conclude that the second order necessary (and, as we'll see later, sufficient) conditions are satisfied.

(iii) Suppose g_1 and g_2 are active.

(iv) And so on. (This problem was not completed during lecture.)

7.2 Second Order Sufficient Conditions

Theorem 7.2.1. *Suppose $f, h_1, \ldots, h_k, g_1, \ldots, g_k \in C^2(\Omega)$, where $\Omega \subseteq \mathbb{R}^n$ is open. Suppose that x_0 is feasible. If*

1. *There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $\mu_1, \ldots, \mu_l \geq 0$ such that*

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

2. *$\mu_j g_j(x_0) = 0$ for each j.*

3. *The matrix*

$$L(x_0) = \nabla^2 f(x_0) + \sum_i \lambda_i \nabla^2 h_i(x_0) + \sum_j \mu_j \nabla^2 g_j(x_0)$$

is positive definite on the tangent space to the "strongly active constraints" at x_0. That is, it is positive definite on the space

$$\tilde{T}_{x_0} = \{v \in \mathbb{R}^n : \nabla h_i(x_0) = 0 \text{ for all } i, \text{ and } \nabla g_j(x_0) = 0 \text{ for all } 1 \leq k \leq l''\},$$

where $\{1, \ldots, l''\}$ is the set of all indices of active constraints whose Lagrange multipliers are positive.

Then x_0 is a strict local minimizer of f subject to the usual constraints.

Proof. Will be given on Thursday. (Copypaste it here?) □

Let's consider some more examples.

1. Here's an example. Given (a, b) with $a, b > 0$ and $a^2 + b^2 > 1$. Consider the minimization problem:

$$\begin{aligned} &\text{minimize } f(x, y) := (x - a)^2 + (y - b)^2 \\ &\text{subject to } g_1(x, y) := x^2 + y^2 - 1 \leq 0 \end{aligned}$$

Our intuition says that the minimizer should be $\left(\frac{a}{\sqrt{a^2+b^2}}, \frac{b}{\sqrt{a^2+b^2}}\right)$. We have $\nabla g(x, y) = (2x, 2y)$, so clearly all feasible points are regular. The Kuhn-Tucker conditions are

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

and $\mu g(x, y) = 0$. That is,

$$\begin{aligned} (1 + \mu)x &= a \\ (1 + \mu)y &= b \\ \mu(x^2 + y^2 - 1) = 0, \mu &\geq 0 \end{aligned}$$

Suppose $\mu = 0$. Then $x = a$ and $y = b$; since we assumed $a^2 + b^2 > 1$, we would have that (x, y) is not feasible. Therefore $\mu \neq 0$, and so $x^2 + y^2 = 1$ by the third equation. Squaring the first two equations and adding them gives

$$(1 + \mu)^2(x^2 + y^2) = a^2 + b^2,$$

implying that $\mu = -1 + \sqrt{a^2 + b^2}$ - we took the positive root because $\mu > 0$. This is actually positive, since $a^2 + b^2 > 1$. Those first equations again give us

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

as expected. What do the second order conditions tell us? The Lagrangian is

$$L(x_0, y_0) = 2I + 2\mu I = 2(1 + \mu)I = 2\sqrt{a^2 + b^2}I,$$

which is everywhere positive definite. Therefore the second-order sufficient conditions are satisfied. For practice, however, let's compute the tangent space to the "strongly active constraints". The only constraint is g; since g is active and its Lagrange multiplier μ is positive, the constraint g is strongly active at (x_0, y_0). Therefore the tangent space we are interested in is the tangent space to S^1 at (x_0, y_0): that space is $\{v \in \mathbb{R}^2 : av_1 + bv_2 = 0\}$.

2. Consider the problem

$$\begin{aligned} &\text{minimize } f(x,y) := x^3 + y^2 \\ &\text{subject to } g(x,y) := (x+1)^2 + y^2 - 1 \leq 0. \end{aligned}$$

We have $\nabla g(x,y) = (2(x+1), 2y)$, which makes it clear that every feasible point is regular. The Kuhn-Tucker conditions are

$$\begin{pmatrix} 3x^2 \\ 2y \end{pmatrix} + \mu \begin{pmatrix} 2(x+1) \\ 2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

with $\mu((x+1)^2 + y^2 - 1) = 0$ and $\mu \geq 0$.

Consider $(x_0, y_0) = (0,0)$. The Kuhn-Tucker conditions imply $\mu = 0$. In particular, g is active at $(0,0)$, but not strongly active there. The tangent space to the active constraint at $(0,0)$ is the y-axis. The Lagrangian at $(0,0)$ is

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

which is clearly positive definite on this tangent space. However, we cannot conclude anything, since the constraint g is not strongly active. In fact, it is clear that $(0,0)$ is not a local minimizer: for $x < 0$ sufficiently close to 0, $f(x,0)$ is negative, yet it is 0 at $(0,0)$.

8 Proof of the Second Order Sufficient Conditions (June 11)

8.1 Second Order Sufficient Conditions

Theorem 8.1.1. *(Second order sufficient conditions for a minimizer under inequality constraints) Suppose $\Omega \subseteq \mathbb{R}^n$ is open and $f, h_1, \ldots, h_k, g_1, \ldots, g_l \in C^2(\Omega)$. Consider the minimization problem*

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

Suppose x_0 is a feasible point of the constraints. If the following three conditions are satisfied:

1. *There exist $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $\mu_1, \ldots, \mu_l \geq 0$ such that*

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

2. *$\mu_j g_j(x_0) = 0$ for each j.*

3. *The matrix*

$$L(x_0) = \nabla^2 f(x_0) + \sum_i \lambda_i \nabla^2 h_i(x_0) + \sum_j \mu_j \nabla^2 g_j(x_0)$$

is positive definite on the tangent space to the "strongly active constraints" at x_0. That is, it is positive definite on the space

$$\tilde{\tilde{T}}_{x_0} = \{v \in \mathbb{R}^n : \nabla h_i(x_0) = 0 \text{ for all } i, \text{ and } \nabla g_j(x_0) = 0 \text{ for all } 1 \leq k \leq l''\},$$

where $\{1, \ldots, l''\}$ is the set of all indices of active constraints whose Lagrange multipliers are positive.

then x_0 is a strict local minimizer of f.

Proof. Suppose x_0 is not a strict local minimizer of f. We claim that there then exists a unit vector $v \in \mathbb{R}^n$ such that

(a) $\nabla f(x_0) \cdot v \leq 0$.

(b) $\nabla h_i(x_0) \cdot v = 0$ for each $i = 1, \ldots, k$.

(c) $\nabla g_j(x_0) \cdot v \leq 0$ for all the active constraints (hereafter labelled by $j = 1, \ldots, l'$).

Intuitively, (a) says that f is non-increasing in the direction of $v \neq 0$, and (b) and (c) say that v is a feasible direction. Let us prove the claim.

Since x_0 is not a strict local minimizer, there exists a sequence x_k of feasible points unequal to x_0 converging to x_0 such that $f(x_k) \leq f(x_0)$. Then $f(x_k) - f(x_0) \leq 0$ for each k. Let $v_k = \frac{x_k - x_0}{\|x_k - x_0\|}$, and let $s_k = \|x_k - x_0\|$. Then $x_k = x_0 + s_k v_k$, with which we may rewrite the inequality as $f(s_k v_k + x_0) - f(x_0) \leq 0$. Since each $v_k \in S^1$, we may assume that the sequence v_k is convergent and that it converges to some $v \in S^1$. We claim that this vector v has the three desired properties.

By Taylor's theorem we have

$$0 \geq f(s_k v_k + x_0) - f(x_0) = s_k \nabla f(x_0) \cdot v_k + o(s_k) \tag{A}$$
$$0 = h_i(s_k v_k + x_0) - h_i(x_0) = s_k \nabla h_i(x_0) \cdot v_k + o(s_k) \tag{B}$$
$$0 \geq g_k(s_k v_k + x_0) - g_j(x_0) = s_k \nabla g_j(x_0) \cdot v_k + o(s_k) \tag{C}$$

(The last equation is ≤ 0 because $g_j(x_0) = 0$.) Divide everything by s_k and take the limit as $k \to \infty$. Then

$$0 \geq \nabla f(x_0) \cdot v \tag{a}$$
$$0 = \nabla h_i(x_0) \cdot v \tag{b}$$
$$0 \geq \nabla g_j(x_0) \cdot v, \tag{c}$$

which proves the earlier claim.

We now claim that equality actually holds in (c). Suppose for the sake of contradiction that there is some $1 \leq k \leq l'$ such that $\nabla g_j(x_0) \cdot v < 0$ for some j for which g_j is strongly active at x_0. By the first condition of the theorem,

$$0 \geq \underbrace{\nabla f(x_0) \cdot v}_{\geq 0 \text{ by (a)}} = -\underbrace{\sum \lambda_i \nabla h_i(x_0) \cdot v}_{= 0 \text{ by (b)}} \underbrace{- \sum \mu_j \nabla g_j(x_0) \cdot v}_{\geq 0 \text{ by (c)}},$$

and so the right hand side is strictly greater than zero, because we only considered strongly active constraints. This is a contradiction, so we conclude that $\nabla g_j(x_0) = 0$ for all j such that g_j is strongly active at x_0. Therefore $v \in \tilde{\tilde{T}}_{x_0}$.

Again, by Taylor's theorem

$$0 \geq f(s_k v_k + x_0) - f(x_0) = s_k \nabla f(x_0) \cdot v_k + \frac{1}{2} s_k^2 v_k^T \nabla^2 f(x_k) \cdot v_k + o(s_k^2)$$
$$0 = h_i(s_k v_k + x_0) - h_i(x_0) = s_k \nabla h_i(x_0) \cdot v_k + \frac{1}{2} s_k^2 v_k^T \nabla^2 h_i(x_k) \cdot v_k + o(s_k^2)$$
$$0 \geq g_k(s_k v_k + x_0) - g_j(x_0) = s_k \nabla g_j(x_0) \cdot v_k + \frac{1}{2} s_k^2 v_k^T \nabla^2 g_j(x_k) \cdot v_k + o(s_k^2)$$

Multiply the second line by λ_i and the third by μ_j and add everything up to get

$$0 \geq s_k \underbrace{\left[\nabla f(x_0) + \sum \lambda_i \nabla h_i(x_0) + \sum \mu_j \nabla g_j(x_0)\right]}_{= 0 \text{ by condition 1}} v_k + \frac{s_k^2}{2} v_k^T \underbrace{\left[\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0) + \sum \mu_j \nabla^2 g_j(x_0)\right]}_{= L(x_0)} v_k + o(s_k^2$$

Divide everything by s_k^2 to get

$$0 \geq \frac{1}{2} v_k^T \left[\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0) + \sum \mu_j \nabla^2 g_j(x_0) \right] \cdot v_k + \frac{o(s_k^2)}{s_k^2}$$

Taking the limit $k \to \infty$ gives

$$0 \leq v^T L(x_0) \cdot v,$$

which violates condition 3 of the theorem. We have a contradiction, so we conclude that x_0 must be a strict local minimizer. □

8.2 A Quick Example

Consider the example from last class:

$$\begin{aligned} \text{minimize } & f(x,y) = -x \\ \text{subject to } & g_1(x,y) = x^2 + y^2 - 1 \leq 0 \\ & g_2(x,y) = y + x - 1 \leq 0 \end{aligned}$$

We found that $(1,0)$ was a good candidate: that it satisfied the necessary conditions. Recall that $\mu_1 = 1/2$, $g_1(1,0) = 0$ and $\mu_2 = 0$, $g_2(1,0) = 0$. Therefore the first constraint is strongly active. The Lagrangian is the identity matrix, so the second order sufficient conditions are satisfied. Therefore $(1,0)$ is a strict local minimizer of f.

9 Newton's Method and Steepest Descent (July 7)

9.1 Motivation for Newton's Method

Consider a twice-differentiable function $f : I \to \mathbb{R}$ defined on an interval $I \subseteq \mathbb{R}$. We would like to find the minima of f. We shall do so by considering quadratic approximations of f.

Let us start at a point $x_0 \in I$. Consider

$$q(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2,$$

the (best) quadratic approximation to f at x_0. Note that $q(x_0) = f(x_0)$, $q'(x_0) = f'(x_0)$ and $q''(x_0) = f''(x_0)$. We will now find the local minimizer x_1 for the quadratic q. That is, we would like to find x_1 such that

$$0 = q'(x_1) = f'(x_0) + f''(x_0)(x_1 - x_0),$$

implying that, so long as $f''(x_0) \neq 0$,

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

The idea of Newton's method is to iterate this procedure. (Consider the Newton's method for finding roots of functions; this is the same as finding the root of the derivative of the function.)

9.2 Newton's Method in One Dimension

Precisely, we pick a starting point $x_0 \in I$. Then we recursively define

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}.$$

We hope that the sequence x_n converges to a minimizer of f. For the sake of the rest of the lecture, let $g = f'$. With this notation we may write Newton's method as

$$x_0 \in I$$
$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}.$$

Theorem 9.2.1. *(Convergence of Newton's Method) Let $g \in C^2(I)$ (i.e. $f \in C^3(I)$). Suppose there is an $x_* \in I$ satisfies $g(x_*) = 0$ and $g'(x_*) \neq 0$. If x_0 is sufficiently close to x_*, then the sequence x_n generated by Newton's method converges to x_*.*

Proof. Since $g'(x_0) \neq 0$, there is, by continuity of g', an $\alpha > 0$ such that

1. $|g'(x_1)| > \alpha$ for all x_1 in a neighbourhood of x_0, and

2. $|g''(x_2)| < \frac{1}{\alpha}$ for all x_2 in the neighbourhood of x_0.

The proof of the first claim is a simple continuity argument. The proof of the second claim follows from continuity of g'' and the extreme value theorem applied to this neighbourhood('s closure). (That is, we can choose an α to bound $|g'|$ from below, and then shrink it possibly to ensure $1/\alpha$ bounds $|g''|$ from above.)

Since $g(x_*) = 0$, the formula of Newton's method now implies

$$x_{n+1} - x_* = x_n - x_* - \frac{g(x_n) - g(x_*)}{g'(x_n)} = -\frac{g(x_n) - g(x_*) - g'(x_n)(x_n - x_*)}{g'(x_n)}. \qquad (*)$$

By the second order mean value theorem, there exists a ξ sufficiently close to x_* such that

$$g(x_*) = g(x_n) + g'(x_n)(x_* - x_n) + \frac{1}{2}g''(\xi)(x_* - x_n)^2.$$

Then (*) becomes

$$x_{n+1} - x_* = \frac{1}{2}\frac{g''(\xi)}{g'(x_n)}(x_n - x_*)^2.$$

The bounds on g' and g'' we found at the start of the proof imply that

$$|x_{n+1} - x_*| < \frac{1}{2\alpha^2}|x_n - x_*|^2. \qquad (**)$$

Let ρ be the constant $\rho = \frac{1}{\alpha^2}|x_0 - x_*|$. Choose x_0 close enough to x_* so that $\rho < 1$. Then (**) implies

$$|x_1 - x_*| < \frac{1}{2\alpha^2}|x_0 - x_*||x_0 - x_*| = \rho|x_0 - x_*| < |x_0 - x_*|.$$

Similarly, (**) gives

$$|x_2 - x_*| < \frac{1}{2\alpha^2}|x_1 - x_*|^2 < \frac{1}{2\alpha^2}\rho^2|x_0 - x_*|^2 < \rho^2|x_0 - x_*|.$$

Continuing in the same way we obtain

$$|x_n - x_*| < \rho^n|x_0 - x_*|,$$

implying that Newton's method converges in our neighbourhood. □

9.3 Newton's Method in Higher Dimensions

Consider a function $f : \Omega \to \mathbb{R}$ defined on an open set $\Omega \subseteq \mathbb{R}^n$. We choose a starting point $x_0 \in \Omega$, and recursively define

$$x_{n+1} = x_n - \nabla^2 f(x_n)^{-1}\nabla f(x_n).$$

For a general f, the algorithm requires that $\nabla^2 f(x_n)$ is invertible. The algorithm stops if $\nabla f(x_n) = 0$ at some point (that is, the sequence given by Newton's method becomes constant if $\nabla f(x_n) = 0$ for some x_n.) Our main result is

Theorem 9.3.1. *(Convergence of Newton's Method) Suppose $f \in C^3(\Omega)$. Suppose also that there is an $x_* \in \Omega$ such that $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*)$ is invertible. Then the sequence x_n defined by*

$$x_{n+1} = x_n - \nabla^2 f(x_n)^{-1} \nabla f(x_n)$$

converges for all x_0 sufficiently close to x_*.

The goal of Newton's method was to find a minimizer of f, but it is possible for it to fail, for it only searches for *critical points*, not necessarily extrema.

9.4 Things That May Go Wrong

It is possible for Newton's method to fail to converge even when f has a unique global minimizer x_* and the initial point x_0 can be taken arbitrarily close to x_*. Consider

$$f(x) = \frac{2}{3}|x|^{3/2} = \begin{cases} \frac{2}{3}x^{3/2} & x \geq 0 \\ \frac{2}{3}(-x)^{3/2} & x \leq 0 \end{cases}.$$

This function is differentiable, and its derivative is

$$f'(x) = \begin{cases} x^{1/2} & x \geq 0 \\ -(-x)^{1/2} & x \leq 0 \end{cases}$$

and its second derivative is

$$f''(x) = \begin{cases} \frac{1}{2}x^{-1/2} & x > 0 \\ \frac{1}{2}(-x)^{-1/2} & x < 0 \\ \text{N/A} & x = 0 \end{cases},$$

so $f \notin C^3$ (it is not even C^2). Let $x_0 = \varepsilon$. Then

$$x_1 = \varepsilon - \frac{f'(\varepsilon)}{f''(\varepsilon)} = \varepsilon - \frac{\varepsilon^{1/2}}{\frac{1}{2}\varepsilon^{-1/2}} = \varepsilon - 2\varepsilon = -\varepsilon,$$

and

$$x_2 = -\varepsilon - \frac{f'(-\varepsilon)}{f''(-\varepsilon)} = -\varepsilon - \frac{-\varepsilon^{1/2}}{\frac{1}{2}\varepsilon^{-1/2}} = -\varepsilon + 2\varepsilon = \varepsilon.$$

So Newton's method gives an alternating sequence $\varepsilon, -\varepsilon, \varepsilon, -\varepsilon, \ldots$. This definitely does not converge. This does not contradict the theorem of convergence because the function in question does not satisfy the conditions of the theorem.

Now we consider an example in which the function in question converges, just not to a minimizer. Consider $f(x) = x^3$, which has derivatives $f'(x) = 3x^2$ and $f''(x) = 6x$. Starting at x_0, we have

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} = x_n - \frac{3x_n^2}{6x_n} = x_n - \frac{1}{2}x_n = \frac{1}{2}x_n.$$

So Newton's method definitely converges to the critical point 0, no matter the choice of $x_0 \in \mathbb{R}$. However, the function f in question does not have a global minimizer, so, while Newton's method converges, it does not converge to an extrema of any sorts.

9.5 Motivation for Steepest Descent

Consider a C^1 function $f : \Omega \to \mathbb{R}$ defined on an open set $\Omega \subseteq \mathbb{R}^n$. The idea is: at every point in the "landscape" of f (the graph of f in $\mathbb{R}^{n+1}$), make a step "downwards" in the steepest direction. (If you're on a mountain and want to descend to the bottom as fast as possible, how do you do so? You, at your current position, take a step down in the steepest direction, and repeat until you're done.)

Since the gradient $\nabla f(x_0)$ represents the direction of greatest increase of f at x_0, the vector $-\nabla f(x_0)$ represents the direction of steepest decrease at x_0. We would therefore like to move in the direction of the negative gradient. We will do so, with the condition that we move until we have a minimizer in the direction of the negative gradient (at which point we will stop moving and repeat).

9.6 Steepest Descent

Here is the steepest descent algorithm:

$$\begin{aligned} x_0 &\in \Omega \\ x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \end{aligned}$$

where $\alpha_k \geq 0$ satisfies

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k)).$$

We call α_k the *optimal step*, since it is chosen so that x_{k+1} is the minimum of f sufficiently close to x_k. We also call x_{k+1} the *minimum point on the half-line* $x_k - \alpha \nabla f(x_k), \alpha \geq 0$. We now describe some properties of the method of steepest descent.

Theorem 9.6.1. *The steepest descent algorithm is actually descending; $f(x_{k+1}) < f(x_k)$ so long as $\nabla f(x_k) \neq 0$.*

Proof. We have

$$f(x_{k+1}) = f(x_k - \alpha_k \nabla f(x_k)) \leq f(x_k - s \nabla f(x_k))$$

for all $s \in [0, \alpha_k]$. Also,

$$\left.\frac{d}{ds}\right|_{s=0} f(x_k - s \nabla f(x_k)) = \nabla f(x_k) \cdot (-\nabla f(x_k)) = -\|\nabla f(x_k)\|^2 < 0.$$

Then for sufficiently small $s \geq 0$,

$$f(x_k - s \nabla f(x_k)) < f(x_k),$$

proving the claim. □

Theorem 9.6.2. *The steepest descent algorithm moves in perpendicular steps; for all k, we have $(x_{k+2} - x_{k+1}) \cdot (x_{k+1} - x_k) = 0$.*

Proof. We have

$$(x_{k+2} - x_{k+1}) \cdot (x_{k+1} - x_k) = \alpha_{k+1}\alpha_k \nabla f(x_{k+1}) \cdot \nabla f(x_k).$$

Recall that $\alpha_k \geq 0$. If $\alpha_k = 0$, then the whole expression is zero and we're done. Consider the possibility that $\alpha_k > 0$. Then

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{s>0} f(x_k - s\nabla f(x_k)),$$

implying that α_k is a minimizer of the function on the right in the above. Then

$$0 = \frac{d}{ds}\bigg|_{s=\alpha_k} f(x_k - s\nabla f(x_k)) = \nabla f(x_k - \alpha_k \nabla f(x_k)) \cdot (-\nabla f(x_k)) = -\nabla f(x_{k+1}) \cdot \nabla f(x_k),$$

proving the claim. □

The fact that the steepest descent algorithm moves in perpendicular steps implies that the method may converge very slowly. Consider the example of a quadratic function $f(x) = x^T Q x$ in $\mathbb{R}^2$ for Q positive definite, and its elliptical level sets.

10 More on Steepest Descent (July 9)

10.1 Convergence of Steepest Descent

Theorem 10.1.1. *Suppose f is a C^1 function on an open set $\Omega \subseteq \mathbb{R}^n$. Let $x_0 \in \Omega$, and let $\{x_k\}_{k=0}^{\infty}$ be the sequence generated by the method of steepest descent. If there is a compact $K \subseteq \Omega$ containing all x_k, then every convergent subsequence of $\{x_k\}_{k=0}^{\infty}$ in K will converge to a critical point x_* of f.*

Proof. Choose a convergent subsequence $\{x_{k_i}\}$ converging to a point $x_* \in K$. Note that $\{f(x_{k_i})\}$ decreases and converges to $f(x_*)$. Since $\{f(x_k)\}$ is a decreasing sequence, it also converges to $f(x_*)$.

Suppose for the sake of contradiction that $\nabla f(x_*) \neq 0$. Since f is C^1, $\nabla f(x_{k_i})$ converges to $\nabla f(x_*)$. Define $y_{k_i} = x_{k_i} - \alpha_{k_i} \nabla f(x_{k_i})$ (i.e. $y_{k_i} = x_{k_1+1}$). We may therefore assume without loss of generality that y_{k_i} converges to some $y_* \in K$. Since $\nabla f(x_*) \neq 0$, we may write

$$\alpha_{k_i} = \frac{|y_{k_i} - x_{k_i}|}{|\nabla f(x_{k_i})|}.$$

Taking the limit as $i \to \infty$, we have

$$\alpha_* := \lim_{i \to \infty} \alpha_{k_i} = \frac{|y_* - x_*|}{|\nabla f(x_*)|}$$

Taking the same limit in the definition of y_{k_i} we have

$$y_* = x_* - \alpha_* \nabla f(x_*).$$

Note that

$$f(y_{k_i}) = f(x_{k_i+1}) = \min_{\alpha \geq 0} f(x_{k_i} - \alpha \nabla f(x_{k_i})).$$

Thus $f(y_{k_i}) \leq f(x_{k_i} - \alpha \nabla f(x_{k_i}))$ for all $\alpha \geq 0$. For any fixed $\alpha \geq 0$, taking the limit $i \to \infty$ gives us

$$f(y_*) \leq f(x_* - \alpha \nabla f(x_*)),$$

implying

$$f(y_*) \leq \min_{\alpha \geq 0} f(x_* - \alpha \nabla f(x_*)) < f(x_*),$$

since the function f decreases in the direction of $-\nabla f(x_*) \neq 0$.

We can also argue the following: $f(x_{k_i+1}) \to f(x_*)$. But since $x_{k_i+1} = y_{k_i}$, we have $f(y_{k_i}) \to f(y_*)$, implying $f(x_*) = f(y_*)$, a contradiction. □

10.2 Steepest Descent in the Quadratic Case

Consider a function f of the form $f(x) = \frac{1}{2} x^T Q x - b^T x$ for $b, x \in \mathbb{R}^n$ and Q an $n \times n$ symmetric positive definite matrix. Let $\lambda = \lambda_1 \leq \cdots \leq \lambda_n = \Lambda$ be the eigenvalues of Q. (Note that they are all

strictly positive.) Note that $\nabla^2 f(x) = Q$ for any x, so f is strictly convex. There therefore exists a unique global minimizer x_* of f in $\mathbb{R}^n$ such that $Qx_* = b$.

Let

$$q(x) = \frac{1}{2}(x - x_*)^T Q(x - x_*) = f(x) + \frac{1}{2}x_*^T Q x_*.$$

So q and f differ by a constant. Therefore it suffices to find the minimizer of q, rather than f. Note that $q(x) \geq 0$ for all x, since Q is positive definite. So we shall study the minimizer x_* of q.

Note that $\nabla f(x) = \nabla q(x) = Qx - b$; let $g(x) = Qx - b$. The method of steepest descent may therefore be written as

$$x_{k+1} = x_k - \alpha_k g(x_k).$$

We would like a formula for the optimal step α_k. Recall that α_k is defined to be the minimizer of the function $f(x_k - \alpha g(x_k))$ over $\alpha \geq 0$. Thus

$$0 = \frac{d}{d\alpha}\bigg|_{\alpha=\alpha_k} f(x_k - \alpha g(x_k)) = \nabla f(x_k - \alpha_k g(x_k)) \cdot (-g(x_k)).$$

This simplifies to

$$0 = (Q(x_k - \alpha_k g(x_k)) - b) \cdot (-g(x_k)) = -(\underbrace{Qx_k - b}_{=g(x_k)} - \alpha_k Q g(x_k)) \cdot g(x_k)$$

giving

$$0 = -|g(x_k)|^2 + \alpha_k g(x_k)^T Q g(x_k).$$

Therefore

$$\alpha_k = \frac{|g(x_k)|^2}{g(x_k)^T Q g(x_k)}. \tag{*}$$

Theorem 10.2.1.

$$q(x_{k+1}) = \left(1 - \frac{|g(x_k)|^4}{(g(x_k)^T Q g(x_k))(g(x_k)^T Q^{-1} g(x_k))}\right) q(x_k)$$

Proof.

$$\begin{aligned}
q(x_{k+1}) &= q(x_k - \alpha_k g(x_k)) \\
&= \frac{1}{2}(x_k - \alpha_k g(x_k) - x_*)^T Q(x_k - \alpha_k g(x_k) - x_*) \\
&= \frac{1}{2}(x_k - x_* - \alpha_k g(x_k))^T Q(x_k - x_* - \alpha_k g(x_k)) \\
&= \frac{1}{2}(x_k - x_*)^T Q(x_k - x_*) - \alpha_k g(x_k)^T Q(x_k - x_*) + \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k) \\
&= q(x_k) - \alpha_k g(x_k)^T Q(x_k - x_*) + \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k),
\end{aligned}$$

implying

$$q(x_k) - q(x_{k+1}) = \alpha_k g(x_k)^T Q(x_k - x_*) - \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k).$$

Dividing by $q(x_k)$ gives

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{\alpha_k g(x_k)^T Q(x_k - x_*) - \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k)}{\frac{1}{2}(x_k - x_*)^T Q(x_k - x_*)}.$$

Let $g_k = g(x_k)$ and $y_k - x_k - x_*$. Then

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{\alpha_k g_k^T Q y_k - \frac{1}{2}\alpha_k^2 g_k^T Q g_k}{\frac{1}{2} y_k^T Q y_k}.$$

Note that $g_k = Qx_k - b = Q(x - x_*) = Qy_k$, so $y_k = Q^{-1} g_k$. The above formula therefore simplifies to

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{2\alpha_k |g_k|^2 - \alpha_k^2 g_k^T Q g_k}{g_k^T Q^{-1} g_k}.$$

Now recall the formula

$$\alpha_k = \frac{|g_k|^2}{g_k^T Q g_k}. \qquad (*)$$

This implies that

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{2\frac{|g_k|^4}{g_k^T Q g_k} - \frac{|g_k|^4}{g_k^T Q g_k}}{g_k^T Q^{-1} g_k} = \frac{|g_k|^4}{(g_k^T Q g_k)(g_K^T Q^{-1} g_k)},$$

proving the theorem. □

11 Steepest Descent Convergence, Conjugate Directions (July 14)

11.1 Recap

Consider $f(x) = \frac{1}{2}x^TQx - b^Tx$, where Q is positive definite symmetric, and has eigenvalues $\lambda = \lambda_1 \leq \cdots \leq \lambda_n = \Lambda$. Since Q is positive definite, there is a unique minimizer x_* such that $Qx_* = b$. Let $g(x) = \nabla f(x) = Qx - b$. We may as well minimize $q(x) = \frac{1}{2}(x - x_*)^TQ(x - x_*) = f(x) + \text{const}$. Moreover, q is always positive except at $x = x_*$, so q is nicer to work with. Note that $\nabla q(x) = \nabla f(x) = g(x) = Qx - b$. Denote by g_k the point $g(x_k) = Qx_k - b$. Then, if x_k is generated by steepest descent, we derived the expression

$$q(x_{k+1}) = \left(1 - \frac{|g_k|^4}{(g_k^TQg_k)(g_k^TQ^{-1}g_k)}\right) q(x_k).$$

We may use this to study the rate of convergence of gradient descent.

11.2 Rate of Convergence of Steepest Descent

If $v = g_k$, then the term in the brackets may be written

$$1 - \frac{|v|^4}{(v^TQv)(v^TQ^{-1}v)}.$$

Kantorovich's inequality says that if Q is an $n \times n$ positive definite symmetric matrix with eigenvalues $\lambda = \lambda_1 \leq \cdots \leq \lambda_n = \Lambda$, then

$$\frac{|v|^4}{(v^TQv)(v^TQ^{-1}v)} \geq \frac{4\lambda\Lambda}{(\lambda + \Lambda)^2} \qquad \text{for all } v \in \mathbb{R}^n.$$

Thus

$$q(x_{k+1}) = \left(1 - \frac{|v|^4}{(v^TQv)(v^TQ^{-1}v)}\right) q(x_k) \leq \left(1 - \frac{4\lambda\Lambda}{(\lambda + \Lambda)^2}\right) q(x_k),$$

which simplifies to, after some work,

$$q(x_{k+1}) \leq \underbrace{\frac{(\lambda - \Lambda)^2}{(\lambda + \Lambda)^2}}_{r} q(x_k).$$

Then $0 \leq r < 1$. We shall call the constant r the *rate of convergence*. We state some properties of steepest descent in the quadratic case. The only thing we have to prove in the following theorem is that steepest descent converges.

Theorem 11.2.1. *(Steepest descent, quadratic case) For $x_0 \in \mathbb{R}^n$, the method of steepest descent starting at x_0 converges to the unique minimizer x_* of the function f, and we have $q(x_{k+1}) \leq rq(x_k)$.*

Proof. We know that $q(x_{k+1}) \leq r^k q(x_0)$. Since $0 \leq r < 1$, when $k \to \infty$, $r^k \to 0$. Note that

$$x_k \in \{x \in \mathbb{R}^n : q(x) \leq r^k q(x_0)\}.$$

This set is a sublevel set of q. The sublevel sets of q look like concentric filled-in ellipses centred at x_*, and as $k \to \infty$, they seem to "shrink" into x_*. Therefore steepest descent converges in the quadratic case. □

Note that

$$r = \frac{(\Lambda - \lambda)^2}{(\Lambda - \lambda)^2} = \frac{(\Lambda/\lambda - 1)^2}{(\Lambda/\lambda - 1)^2},$$

so r depends only on the ratio Λ/λ. This number is called the *condition number of* Q. (The condition number may be defined as $\|Q\|\|Q^{-1}\|$ in the operator norm on matrices; it is not hard to see that these numbers agree in our case.)

If the condition number $\Lambda/\lambda \gg 1$ (large), then convergence is very slow. If $\Lambda/\lambda = 1$, then $r = 0$, and so convergence is achieved in one step.

11.3 Method of Conjugate Directions

We will develop a new method for finding the minimizers of quadratic functions $\frac{1}{2}x^T Q x - b^T x$.

Definition 11.3.1. *Let Q be symmetric. We say that d, d' are Q-conjugate or Q-orthogonal if $d^T Q d' = 0$. A finite set $d_0, \ldots, d_k$ of vectors is called Q-orthogonal if $d_i^T Q d_j = 0$ for all $i \geq j$.*

For example, if $Q = I$, then Q-orthogonality is equivalent to regular orthogonality. For another example, if Q has more than one distinct eigenvalue, let d and d' be eigenvectors corresponding to distinct eigenvalues. Then $d^T Q d' = \lambda' d^T d' = 0$, since the distinct eigenspaces of a symmetric matrix are orthogonal subspaces.

Recall that any symmetric matrix Q may the orthogonally diagonalized; there exists an orthonormal basis $d_0, \ldots, d_{n-1}$ of eigenvectors of Q. These eigenvectors are also Q-orthogonal. Hence to any symmetric matrix is a basis of orthonormal vectors that are also orthogonal with respect to the matrix, as just defined.

Theorem 11.3.1. *If Q is symmetric and positive definite, then any set of non-zero Q-orthogonal vectors $\{d_i\}$ is linearly independent.*

Proof. If $\sum \alpha_i d_i = 0$, then left-multiplying by $d_j^T Q$ gives $\alpha_j d_j^T Q d_j = 0$. Positive definiteness implies $\alpha_j = 0$. □

Let Q be an $n \times n$ symmetric positive definite matrix. Recall that $f(x) = \frac{1}{2}x^T Q x - b^T x$ has the unique global minimizer $x_* = Q^{-1}b$. Let $d_0, \ldots, d_{n-1}$ be non-zero Q-orthogonal vectors. Then $d_0, \ldots, d_{n-1}$ form a basis of $\mathbb{R}^n$. Thus there are scalars $\alpha_0, \ldots, \alpha_{n-1}$ such that $x_* = \sum \alpha_i d_i$. We would like a formula for the α_i's.

Multiplying both sides of the sum $x_* = \sum \alpha_i d_i$ by $d_j^T Q$ implies that $d_j^T Q x_* = \alpha_j d_j^T Q d_j$, implying that

$$\alpha_j = \frac{d_j^T b}{d_j^T Q d_j}.$$

Therefore

$$x_* = \sum_{i=1}^{n-1} \frac{d_i^T b}{d_i^T Q d_i} d_i.$$

This implies that we can actually solve for x_* by computing the $d_0, \ldots, d_{n-1}$ and the coefficients above. Computationally, computing inner products is very easy. The disadvantage is that we do not know how to find the vectors $d_0, \ldots, d_{n-1}$.

Theorem 11.3.2. *(Method of Conjugate Directions) Let $d_0, \ldots, d_{n-1}$ be a set of non-zero Q-orthogonal vectors. For a starting point $x_0 \in \mathbb{R}^n$, consider the sequence $\{x_l\}$ defined by*

$$x_{k+1} = x_k + \alpha_k d_k,$$

where

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k} \qquad \text{where } g_k = Q x_k - b.$$

The sequence $\{x_k\}$ converges to the minimizer x_* it at most n steps; $x_n = x_*$.

Proof. Write $x_* - x_0 = \alpha_0' d_0 + \cdots + \alpha_{n-1}' d_{n-1}$. Multiply both sides by $d_i^T Q$ to get

$$d_i^T Q(x_* - x_0) = \alpha_i d_i^T Q d_i,$$

giving us the expression

$$\alpha_i' = \frac{d_i^T Q(x_* - x_0)}{d_i^T Q d_i}. \tag{*}$$

Note that

$$\begin{aligned} x_1 &= x_0 + \alpha_0 d_0 \\ x_2 &= x_0 + \alpha_0 d_0 + \alpha_1 d_1 \\ &\vdots \\ x_k &= x_0 + \alpha_0 d_0 + \cdots + \alpha_{k-1} d_{k-1}, \end{aligned}$$

implying that

$$x_k - x_0 = \alpha_0 d_0 + \cdots + \alpha_{k-1} d_{k-1}.$$

Multiplying both sides by $d_k^T Q$ gives $d_k^T Q(x_k - x_0) = 0$. By (*) we have

$$\alpha_k' = \frac{d_k^T Q(x_* - x_0) - d_k^T Q(x_k - x_0)}{d_k^T Q d_k} = \frac{d_k^T Q(x_* - x_k)}{d_k^T Q d_k} = -\frac{(Q x_k - Q x_*)^T d_k}{d_k^T Q d_k}$$

simplifying to

$$\alpha_k' = -\frac{g_k^T d_k}{d_k^T Q d_k} = \alpha_k.$$

So

$$x_* = x_0 + \alpha_0 d_0 + \cdots + \alpha_{n-1} d_{n-1} = x_n.$$

So after n steps, we reach the minimizer. □

(There may be an error in the above calculations. The professor will send a note on this.)

11.4 Geometric Interpretation of Conjugate Directions

Let $d_0, \ldots, d_{n-1}$ be a set of non-zero Q-orthogonal vectors in $\mathbb{R}^n$. Let B_k be the span of the first k of these vectors. Note that B_k has dimension k and contains $B_1, \ldots, B_{k-1}$, so $B_1, \ldots, B_n$ is a sequence of expanding subspaces of $\mathbb{R}^n$. Let us agree that $B_0 = \{0\}$.

Fix $x_0 \in \mathbb{R}^n$ and consider the affine subspaces $x_0 + B_k$ each with "origin" x_0. We now have a sequence of expanding affine subspaces of $\mathbb{R}^n$.

Theorem 11.4.1. *The sequence $\{x_k\}$ generated from x_0 by the method of conjugate directions has the property that x_k is the minimizer of $f(x) = \frac{1}{2}x^T Q x - b^T x$ on the affine subspace $x_0 + B_k$.*

12 More on Conjugate Directions (July 16)

12.1 Geometric Interpretation

Let $d_0, \ldots, d_{n-1}$ be a set of non-zero Q-orthogonal vectors in $\mathbb{R}^n$, where Q is symmetric and positive definite. Note that these vectors are linearly independent by a result from last lecture. Let B_k denote the subspace spanned by the first k vectors. We have an increasing sequence

$$B_0 \subsetneq B_1 \subsetneq \cdots \subsetneq B_n,$$

and $\dim(B_k) = k$.

Theorem 12.1.1. *The sequence $\{x_k\}_{k=0}^{\infty}$ generated from x_0 by the method of conjugate directions has the property that x_k minimizes $f(x) = \frac{1}{2}x^TQx - b^Tx$ on the affine subspace $x_0 + B_k$.*

Recall the function $q(x) = \frac{1}{2}(x - x_*)^TQ(x - x_*)$, which differs from $f(x)$ by a constant. They have the same minimizers.

If $Q = I$, then $q(x) = \frac{1}{2}|x - x_*|^2$. Then $x_k \in x_0 + B_k$ is the closest point in $x_0 + B_k$ to x_*, by the theorem.

Before proving the theorem, recall the following result about convex functions.

Lemma 12.1.1. *Let f be a C^1 convex function defined on a convex domain $\Omega \subseteq \mathbb{R}^n$. Suppose there is an $x_* \in \Omega$ such that $\nabla f(x_*) \cdot (y - x_*) \geq 0$ for all $y \in \Omega$. Then x_* is a global minimizer of f on Ω. The converse is obviously true.*

Geometrically, this means that if we move in any feasible direction from the point x_*, the function is increasing. Hence x_* is a local minimizer; convexity implies it is global. With this result in mind, we prove the theorem.

Proof. The affine subspace $\Omega = x_0 + B_k$ is convex. **(This proof could not be finished as attention had to be diverted from the lecture.)** □

Corollary 12.1.1. *x_n minimizes $f(x)$ on $\mathbb{R}^n$. That is, $x_n = x_*$; the method of conjugate directions for this function f terminates in at most n steps.*

When $Q = I$, then $q(x)$ is half the distance squared from x to x_*. What if $Q \neq I$. q is still a metric on $\mathbb{R}^n$. Thus x_k is the point "closest" to x_* on the affine subspace $x_0 + B_k$. **(These notes are incomplete.)**

13 Conjugate Gradients, Introduction to The Calculus of Variations (July 21)

13.1 Conjugate Gradient Method

Assume all of the conditions of the previous class.

We will describe a new optimization algorithm that is a type of conjugate direction method. Start at $x_0 \in \mathbb{R}^n$. Choose $d_0 = -g_0 = -\nabla f(x_0) = b - Qx_0$. Recursively define $d_{k+1} = -g_{k+1} + \beta_k d_k$, where $g_{k+1} = Qx_{k+1} - b$ and

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$$

and

$$x_{k+1} = x_k + \alpha_k d_k,$$

where

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}.$$

Given an initial point x_0, take $d_0 = -g_0 = b - Qx_0$. By definition, $x_1 = x_0 + \alpha_0 d_0$; we need to find α_0. This is

$$\alpha_0 = -\frac{g_0^T d_0}{g_0^T Q g_0}.$$

Then $x_2 = x_1 + \alpha_1 d_1$. By definition, $\alpha_1 = -\frac{g_1^T d_1}{d_1^T Q d_1}$, where $d_1 = -g_1 + \beta_0 d_0$, where $\beta_0 = \frac{g_1^T Q d_0}{d_0^T Q d_0}$.

Some remarks:

1. Like the other conjugate direction methods, this method converges to the minimizer x_* in n steps.
2. We have a procedure to find the direction vectors d_k.
3. This method makes good *uniform* progress towards the solution at every step.

13.2 Bounds on Convergence

As before, consider $q(x) = \frac{1}{2}(x - x_*)^T Q (x - x_*) = f(x) + \text{const}$. It's better to look at q rather than f, since q behaves like a distance function relative to x_*. (More on this in HW7.)

Theorem 13.2.1.

$$q(x_{k+1}) \leq \left(\max_{\substack{\lambda \\ \textit{eigval of } Q}} (1 + \lambda P_k(\lambda))^2 \right) q(x_k),$$

where P_k is any polynomial of degree k.

Proof. In the textbook; will not be proven in class. □

For example, suppose Q has $m \leq n$ distinct eigenvalues. Choose a polynomial P_{m-1} such that $1 + \lambda P_{m-1}(\lambda)$ has its m zeroes at the m eigenvalues of Q. With such a polynomial, we would get $q(x_m) \leq 0$, implying that $q(x_m) = 0$; the conjugate gradient method terminates at the mth step, i.e. $x_m = x_*$.

13.3 Introducing The Calculus of Variations

Consider the problem

$$\begin{aligned} &\text{minimize } F[u] \\ &\qquad u \in \mathcal{A}, \end{aligned}$$

where $\mathcal{A}$ is a set of functions. Here, F is a function of functions, often called a *functional*. This is the general unconstrained calculus of variations problem.

For example, consider

$$\mathcal{A} = \{u \in C^1([0,1], \mathbb{R}) : u(0) = u(1) = 1\}.$$

Define $F : \mathcal{A} \to \mathbb{R}$ by

$$F[u(\cdot)] \coloneqq \frac{1}{2} \int_0^1 (u(x)^2 + u'(x)^2)\, dx.$$

To solve the minimization problem

$$\begin{aligned} &\text{minimize } F[u] \\ &\qquad u \in \mathcal{A} \end{aligned}$$

is to find a $u^* \in \mathcal{A}$ such that $F[u^*] \leq F[u]$ for all $u \in \mathcal{A}$. To do so, we will

1. We will derive first order necessary conditions for a minimizer.
2. We will find a function satisfying these conditions.
3. We will check that this function is indeed a minimizer. (This is not always possible.)

(Consider the obvious parallels with finite dimensional optimization.)

Fix $v \in C^1([0,1], \mathbb{R})$ with $v(0) = v(1) = 0$. Suppose u^* is a minimizer of F on $\mathcal{A}$. Clearly $u^* + sv \in \mathcal{A}$ for all $s \in \mathbb{R}$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(s) \coloneqq F[u^* + sv]$. Then $f(s) \geq f(0)$ for all s, since u^* is a minimizer of F. Then 0 is a minimizer of f, implying $f'(0) = 0$. How do we actually compute $f'(0)$? Since

$$\begin{aligned} f(s) &= \frac{1}{2} \int_0^1 (u^*(x) + sv(x))^2 + (u^{*'}(x) + sv'(x))^2\, dx \\ &= \frac{1}{2} \int_0^1 (u^*(x)^2 + u^{*'}(x)^2)\, dx + s \int_0^1 (u^*(x)v(x) + u^{*'}(x)v'(x))\, dx + \frac{s^2}{2} \int_0^1 (v(x)^2 + v'(x)^2)\, dx, \end{aligned}$$

implying that

$$f'(s) = \int_0^1 (u^*(x)v(x) + u^{*'}(x)v'(x))\,dx + s\int_0^1 (v(x)^2 + v'(x)^2)\,dx,$$

or

$$0 = \int_0^1 (u^*(x)v(x) + u^{*'}(x)v'(x))\,dx \text{ for all } v \in C^1([0,1],\mathbb{R}) \text{ such that } v(0) = v(1) = 0. \qquad (*)$$

The above equality holds for all $v \in C^1([0,1],\mathbb{R})$ such that $v(0) = v(1) = 0$. This is a *primitive* form of the first order necessary conditions.

Let us call the functions v described in (*) the *test functions on* $[0,1]$. We would like to write (*) in a more useful way. Let us make the simplifying assumption that u^* is C^2. Integration by parts gives

$$\int_0^1 u^{*'}(x)v'(x)\,dx = \cancel{u^{*'}(x)v(x)\Big|_0^1} - \int_0^1 u^{*''}(x)v(x)\,dx = \int_0^1 u^{*''}(x)v(x)\,dx.$$

By substituting this into (*) we obtain

$$\int_0^1 (u^*(x)v(x) - u^{*''}(x)v(x))\,dx = 0.$$

Factor the common v out to get

$$\int_0^1 (u^*(x) - u^{*''}(x))v(x)\,dx = 0 \text{ for all test functions } v \text{ on } [0,1].$$

So we have a continuous function $u^*(x) - u^{*''}(x)$ that is zero whenever "integrated against test functions". We claim that any function satisfying this condition must be zero. This result or its variations is called the *fundamental lemma of the calculus of variations*. We shall show that $u^* = u^{*''}$ on $[0,1]$; this gives us the first order necessary conditions we wanted in the first place.

Theorem 13.3.1. *(Fundamental lemma of the calculus of variations) Suppose* $g \in C^0([a,b])$. *If*

$$\int_a^b g(x)v(x)\,dx = 0$$

for all test functions v *on* $[a,b]$, *then* $g \equiv 0$ *on* $[a,b]$.

So the first order necessary condition we derived are that $u^* = u^{*''}$ on $[0,1]$, as well as $u^*(0) = u^*(1) = 1$. By MAT267, $u^*(x) = c_1e^x + c_2e^{-x}$ for some constants $c_1, c_2 \in \mathbb{R}$. Some work gives $c_1 = \frac{1}{e+1}$ and $c_2 = \frac{e}{e+1}$. Therefore

$$u^*(x) = \frac{1}{e+1}e^x + \frac{e}{e+1}e^{-x}$$

is the only C^1 minimizer candidate.

We must finally check that u^* is indeed a minimizer. Some work shows that $F[u^* + sv] \geq F[u^*]$ for all $s \in \mathbb{R}$ and all test functions v on $[0,1]$. Choose $u \in \mathcal{A}$. Let $v = u - u^*$; this is a test function on $[0,1]$, so $F[u] \geq F[u^*]$, showing that u^* is, in fact, the global minimizer.

14 The Brachistochrone Problem (July 23)

14.1 Fundamental Lemma of the Calculus of Variations

Recall that v is said to be a *test function on* $[a,b]$ if it is C^1 and $v(a) = v(b) = 0$.

Theorem 14.1.1. *(Fundamental Lemma of the Calculus of Variations) If g is a continuous function on $[a,b]$ with the property that*

$$\int_a^b g(x)v(x)\,dx = 0$$

for all test functions v on $[a,b]$, then $g \equiv 0$.

Proof. Suppose $g \not\equiv 0$. Then there is an $x_0 \in (a,b)$ such that $g(x_0) \neq 0$. (We can ensure that x_0 is in the interior of the interval because of continuity.) Assume without loss of generality that $g(x_0) > 0$. There exists an open neighbourhood (c,d) of x_0 inside (a,b) on which g is positive. Let v be a C^1 function on $[a,b]$ such that $v > 0$ on (c,d) and $v = 0$ otherwise. Then v is a test function on $[a,b]$, so by the hypotheses,

$$0 = \int_a^b g(x)v(x)\,dx = \int_c^d g(x)v(x)\,dx > 0,$$

a contradiction. □

The test function v we chose in the proof of the preceding theorem could be, for example,

$$v(x) = \begin{cases} (x-c)^2(x-d)^2 & x \in [c,d] \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$v(x) = \begin{cases} 2(x-c)(x-d)^2 + 2(x-c)^2(x-d) & x \in (c,d) \\ 0 & \text{otherwise} \end{cases},$$

which is easily seen to be continuous. Therefore v is a test function on $[a,b]$ which is positive only on (c,d).

14.2 The Brachistochrone Problem

The brachistrochrone problem is the problem from which the calculus of variations was born. In approximately 1638, Galileo Galilei was studying the problem of a ball rolling along a slope from a point A to a point B. Galileo experimented with multiple kinds of slopes, such as a straight line from A to B, or some non-straight curve from A to B, and so on. He measured the time it takes for the ball to roll. He first noticed that the straight line from A to B did *not* minimize the time. He posed the following problem:

Find the curve connecting A and B on which a point mass moves without friction under the influence of gravity in the least time possible.

Around 1696, Johan Bernoulli posted this problem somewhere as a challenge to the mathematicians of the world.

Let us pose the problem more mathematically. Let $c : [0, T] \to \mathbb{R}^2$ describe a curve (the graph of a function) that starts at A at time $t = 0$ and ends at B at time $t = T$. So if $c(t) = (x(t), y(t))$ satisfies $c(0) = A$ and $c(T) = B$. Assuming $y = u(x)$, we have $c(t) = (x(t), u(x(t)))$. Assume $A = (0, a)$ and $B = (b, 0)$.

Now what is the velocity of the point mass along this curve?

$$v(t) = \frac{d}{dt}c(t) = \begin{pmatrix} x'(t) \\ u'(x(t))x'(t) \end{pmatrix} = x'(t) \begin{pmatrix} 1 \\ u'(x(t)) \end{pmatrix}.$$

The kinetic energy of the point mass is $K(t) = \frac{1}{2}m|v|^2 = \frac{m}{2}x'(t)^2(1 + u'(x(t)))^2$, and the potential energy is $V(t) = mgy = mgu(x(t))$. The total energy is $E = K + P$. There is no friction, so energy is conserved, hence the total energy at any time is equal to the total energy at time $t = 0$: $E(t) = E(0)$ for all t. Written out, this means

$$\frac{m}{2}x'(t)^2(1 + u'(x(t)))^2 + mgu(x(t)) = mga.$$

Some algebra shows that this is equal to

$$\frac{1}{2}x'(t)^2 = \frac{g(a - u(x(t)))}{1 + u'(x(t))^2}.$$

Multiplying by 2 and taking square roots gives

$$x'(t) = \sqrt{\frac{2g(a - u(x(t)))}{1 + u'(x(t))^2}},$$

a differential equation in x! What is the total time it takes the point mass to go from A to B along c? We have

$$T = \int_0^T 1\,dt = \int_0^T \sqrt{\frac{1 + u'(x(t))^2}{2g(a - u(x(t)))}}x'(t)\,dt. \qquad (*)$$

How does this give us anything we want? It appears that T is on both sides, so that this reveals nothing about T.

Let f be some function, and consider the integral

$$\int_{t_0}^{t_1} f(x(t))x'(t)\,dt = \int_{t_0}^{t_1} F'(x(t))x'(t)\,dt = F(x(t_1)) - F(x(t_0)) = \int_{x(t_0)}^{x(t_1)} f(x)\,dx, \qquad (**)$$

where F is an antiderivative of f.

Now, (**) applied to (*) gives

$$T = \int_{x(0)}^{x(1)} \sqrt{\frac{1+u'(x)^2}{2g(a-u(x))}}\,dx = \int_0^b \sqrt{\frac{1+u'(x)^2}{2g(a-u(x))}}\,dx.$$

With this, we may pose Galileo's original problem as a minimization problem: the *brachistochrone problem in the calculus of variations.*

$$\begin{aligned} \text{minimize } F[u] &\coloneqq \int_0^b \sqrt{\frac{1+u'(x)^2}{2g(a-u(x))}}\,dx \\ u \in \mathcal{A} &\coloneqq \{u \in C^1([0,b],\mathbb{R}) : u(0) = a, u(b) = 0\}. \end{aligned}$$

This is a problem in the calculus of variations. Next time, we will find first order conditions for a minimizer and attempt to find a function u_* satisfying these conditions.