

APM462: Nonlinear Optimization

Tianyu Du

December 13, 2019

Contents

1 Preliminaries	2
1.1 Mean Value Theorems and Taylor Approximations.	2
1.2 Implicit Function Theorem	3
2 Convexity	3
2.1 Terminologies	3
2.2 Basic Properties of Convex Functions	3
2.3 Characteristics of C^1 Convex Functions	4
2.4 Minimum and Maximum of Convex Functions	5
3 Finite Dimensional Optimization	6
3.1 Unconstraint Optimization	6
3.2 Equality Constraints: Lagrangian Multiplier	10
3.2.1 Tangent Space to a (Hyper) Surface at a Point	10
3.3 Remark on the Connection Between Constrained and Unconstrained Optimizations	15
3.4 Inequality Constraints	16
4 Iterative Algorithms for Optimization	19
4.1 Newton's Method	19
4.2 Steepest/Gradient Descent	22
4.2.1 Steepest Descent: the Quadratic Case	24
4.3 Method of Conjugate Directions	27
4.4 Geometric Interpretations of Method of Conjugate Directions	31
5 Infinite Dimensional Optimization	32
5.1 Calculus of Variation	32
5.1.1 The Canonical Example	32
5.2 A Classical Example from Physics: Brachistochrone	34
5.3 General Class of Problems in Calculus of Variations	34
5.4 Euler-Lagrange Equations in \mathbb{R}^n	38
5.5 Equality Constraints: Isoperimetric Case	39
5.6 Equality Constraints: Holonomic Case	41
5.7 Geodesics on the Cylinder	42
6 Appendix	47

1 Preliminaries

1.1 Mean Value Theorems and Taylor Approximations.

Definition 1.1. Let $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, the **gradient** of f at $x \in S$, if exists, is a vector $\nabla f(x) \in \mathbb{R}^n$ characterized by the property

$$\lim_{v \rightarrow 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{\|v\|} = 0 \quad (1.1)$$

Theorem 1.1 (The First Order Mean Value Theorem). Let $f \in C^1(\mathbb{R}^n, \mathbb{R})$, then for any $x, v \in \mathbb{R}^n$, there exists some $\theta \in (0, 1)$ such that

$$f(x+v) = f(x) + \nabla f(x+\theta v) \cdot v \quad (1.2)$$

Proof. Let $x, v \in \mathbb{R}^n$, define $g(t) := f(x+tv) \in C^1(\mathbb{R}, \mathbb{R})$.

By the mean value theorem on \mathbb{R} , there exists $\theta \in (0, 1)$ such that $g(0+1) = g(0) + g'(\theta)(1-0)$, that is, $f(x+v) = f(x) + g'(\theta)$. Note that $g'(\theta) = \nabla f(x+\theta v) \cdot v$. ■

Proposition 1.1 (The First Order Taylor Approximation). Let $f \in C^1(\mathbb{R}^n, \mathbb{R})$, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + o(\|v\|) \quad (1.3)$$

that is

$$\lim_{\|v\| \rightarrow 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v}{\|v\|} = 0 \quad (1.4)$$

Proof. By the mean value theorem, $\exists \theta \in (0, 1)$ such that $f(x+v) - f(x) = \nabla f(x+\theta v) \cdot v$.

The limit becomes $\lim_{\|v\| \rightarrow 0} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{\|v\|} = \lim_{\|v\| \rightarrow 0; x+\theta v \rightarrow x} \frac{[\nabla f(x+\theta v) - \nabla f(x)] \cdot v}{\|v\|}$.

Since $f \in C^1$, $\lim_{x+\theta v \rightarrow x} \nabla f(x+\theta v) = \nabla f(x)$.

And $\frac{v}{\|v\|}$ is a unit vector, and every component of it is bounded, as the result, the limit of inner product vanishes instead of explodes. ■

Theorem 1.2 (The Second Order Mean Value Theorem). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^2 function, then for any $x, v \in \mathbb{R}^n$, there exists $\theta \in (0, 1)$ satisfying

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v^T \nabla^2 f(x+\theta v) v \quad (1.5)$$

Proposition 1.2 (The Second Order Taylor Approximation). Let $f : C^2(\mathbb{R}^n, \mathbb{R})$ function, and $x, v \in \mathbb{R}^n$, then

$$f(x+v) = f(x) + \nabla f(x) \cdot v + \frac{1}{2} v^T \nabla^2 f(x+\theta v) v + o(\|v\|^2) \quad (1.6)$$

that is

$$\lim_{\|v\| \rightarrow 0} \frac{f(x+v) - f(x) - \nabla f(x) \cdot v - \frac{1}{2} v^T \nabla^2 f(x) v}{\|v\|^2} = 0 \quad (1.7)$$

Proof. By the second mean value theorem, there exists $\theta \in (0, 1)$ such that the limit is equivalent to

$$\lim_{\|v\| \rightarrow 0} \frac{1}{2} \left(\frac{v}{\|v\|} \right)^T [\nabla^2 f(x+\theta v) - \nabla^2 f(x)] \frac{v}{\|v\|} \quad (1.8)$$

Since $f \in C^2$, the limit of $[H_f(x+\theta v) - H_f(x)]$ is in fact $\mathbf{0}_{n \times n}$. And every component of unit vector $\frac{v}{\|v\|}$ is bounded, the quadratic form converges to zero as an immediate result. ■

It is often noted that the gradient at a particular $x_0 \in \text{dom}(f) \subseteq \mathbb{R}^n$ gives the direction f increases most rapidly. Let $x_0 \in \text{dom}(f)$, and v be a unit vector representing a *feasible direction* of change. That is, there

exists $\delta > 0$ such that $x_0 + tv \in \text{dom}(f) \forall t \in [0, \delta]$. Then the rate of change of f along feasible direction v can be written as

$$\left. \frac{d}{dt} \right|_{t=0} f(x_0 + tv) = \nabla f(x_0) \cdot v = \|\nabla f(x_0)\| \|v\| \cos(\theta) \quad (1.9)$$

where $\theta = \angle(v, \nabla f(x_0))$. And the derivative is maximized when $\theta = 0$, that is, when v and ∇f point the same direction.

1.2 Implicit Function Theorem

Theorem 1.3 (Implicit Function Theorem). Let $f : C^1(\mathbb{R}^{n+1}, \mathbb{R})$, let $(a, b) \in \mathbb{R}^n \times \mathbb{R}$ such that $f(a, b) = 0$. If $\nabla f(a, b) \neq 0$, then $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : f(x, y) = 0\}$ is locally a graph of a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$.

Remark 1.1. $\nabla f(x_0) \perp$ level set of f near x_0 .

2 Convexity

2.1 Terminologies

Definition 2.1. Set $\Omega \subseteq \mathbb{R}^n$ is **convex** if and only if

$$\forall x_1, x_2 \in \Omega, \lambda \in [0, 1], \lambda x_1 + (1 - \lambda)x_2 \in \Omega \quad (2.1)$$

Definition 2.2. A function $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is **convex** if and only if Ω is convex, and

$$\forall x_1, x_2 \in \Omega, \lambda \in [0, 1], f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (2.2)$$

Definition 2.3. A function $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ is **strictly convex** if and only if Ω is convex and

$$\forall x_1, x_2 \in \Omega, \lambda \in (0, 1), f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (2.3)$$

2.2 Basic Properties of Convex Functions

Definition 2.4. A function $f : \Omega \rightarrow \mathbb{R}$ is **concave** if and only if $-f$ is **convex**.

Proposition 2.1. Properties of convex functions:

- (i) If f_1, f_2 are convex on Ω , so is $f_1 + f_2$;
- (ii) If f is convex on Ω , then for any $a > 0$, af is also convex on Ω ;
- (iii) Any **sub-level/lower contour set** of a convex function f

$$\mathcal{L}(c) := \{x \in \mathbb{R}^n : f(x) \leq c\} \quad (2.4)$$

is convex.

Proof of (iii). Let $c \in \mathbb{R}$, and $x_1, x_2 \in \mathcal{L}(c)$. Let $s \in [0, 1]$. Since $x_1, x_2 \in \mathcal{L}(c)$, and $f(\cdot)$ is convex, $f(sx_1 + (1 - s)x_2) \leq sf(x_1) + (1 - s)f(x_2) \leq sc + (1 - s)c = c$. Which implies $sx_1 + (1 - s)x_2 \in \mathcal{L}(c)$. ■

Example 2.1. ℓ_2 norm $f(x) : \mathbb{R}^n \rightarrow \mathbb{R} := \|x\|_2$ is convex.

Proof. Note that for any $u, v \in \mathbb{R}^n$, by triangle inequality, $\|u - (-v)\| \leq \|u - 0\| + \|0 - (-v)\| = \|u\| + \|v\|$. Consequently, let $u, v \in \mathbb{R}^n$ and $s \in [0, 1]$, then $\|su + (1-s)v\| \leq \|su\| + \|(1-s)v\| = s\|u\| + (1-s)\|v\|$. Therefore, $\|\cdot\|$ is convex. ■

Proposition 2.2. Any norm function $\|\cdot\|$ defined on a vector space $\mathcal{X}(\mathbb{R})$ is convex.

Proof. The proof follows the defining properties of norm,

$$\|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| \quad (2.5)$$

$$= \lambda\|x\| + (1-\lambda)\|y\| \quad (2.6)$$

■

2.3 Characteristics of C^1 Convex Functions

Theorem 2.1 (C^1 criteria for convexity). Let $f \in C^1$, then f is convex on a convex set Ω if and only if

$$\forall x, y \in \Omega, \quad f(y) \geq f(x) + \nabla f(x) \cdot (y - x) \quad (2.7)$$

that is, the linear approximation is never an overestimation of value of f .

Proof. (\implies) Suppose f is convex on a convex set Ω . Then $f(sy + (1-s)x) \leq sf(y) + (1-s)f(x)$ for every $x, y \in \Omega$ and $s \in [0, 1]$, which implies, for every $s \in (0, 1]$:

$$\frac{f(sy + (1-s)x) - f(x)}{s} \leq f(y) - f(x) \quad (2.8)$$

By taking the limit of $s \rightarrow 0$,

$$\lim_{s \rightarrow 0} \frac{f(x + s(y-x)) - f(x)}{s} \leq f(y) - f(x) \quad (2.9)$$

$$\implies \left. \frac{d}{ds} \right|_{s=0} f(x + s(y-x)) \leq f(y) - f(x) \quad (2.10)$$

$$\implies \nabla f(x) \cdot (y - x) \leq f(y) - f(x) \quad (2.11)$$

(\impliedby) Let $x_0, x_1 \in \Omega$, let $s \in [0, 1]$. Define $x^* := sx_0 + (1-s)x_1$, then

$$f(x_0) \geq f(x^*) + \nabla f(x^*) \cdot (x_0 - x^*) \quad (2.12)$$

$$\implies f(x_0) \geq f(x^*) + \nabla f(x^*) \cdot [(1-s)(x_0 - x_1)] \quad (2.13)$$

Similarly,

$$f(x_1) \geq f(x^*) + \nabla f(x^*) \cdot (x_1 - x^*) \quad (2.14)$$

$$\implies f(x_1) \geq f(x^*) + \nabla f(x^*) \cdot [s(x_1 - x_0)] \quad (2.15)$$

Therefore, $sf(x_0) + (1-s)f(x_1) \geq f(x^*)$. ■

Theorem 2.2 (C^2 criterion for convexity). $f \in C^2$ is a convex function on a convex set $\Omega \subseteq \mathbb{R}^n$ if and only if $\nabla^2 f(x) \succcurlyeq 0$ (i.e. positive semidefinite) for all $x \in \Omega$.

Corollary 2.1. When f is defined on \mathbb{R} , the C^2 criterion becomes $f''(x) \geq 0$.

Proof. (\Leftarrow) Suppose $\nabla^2 f(x) \succcurlyeq 0$ for every $x \in \Omega$, let $x, y \in \Omega$. By the second order MVT,

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + s(y - x))(y - x) \text{ for some } s \in [0, 1] \quad (2.16)$$

$$\implies f(y) \geq f(x) + \nabla f(x) \cdot (y - x) \quad (2.17)$$

So f is convex by the C^1 criterion of convexity.

(\implies) Let $v \in \mathbb{R}^n$. Suppose, for contradiction, that for some $x \in \Omega$, $\nabla^2 f(x) \not\succeq 0$.

If such $x \in \partial\Omega$, note that $v^T \nabla^2 f(\cdot) v$ is continuous because $f \in C^2$, then there exists $\varepsilon > 0$ such that $\forall x' \in V_\varepsilon(x) \cap \Omega^{int}$, $v^T \nabla^2 f(x') v < 0$.

Hence, one may assume with loss of generality that such $x \in \Omega^{int}$.

Because $x \in \Omega^{int}$, exists $\varepsilon' > 0$, such that $V_{\varepsilon'}(x) \subseteq \Omega^{int}$.

Define $\hat{v} := \frac{v}{\sqrt{\varepsilon'}}$, then for every $s \in [0, 1]$, $\hat{v}^T \nabla^2 f(x + s\hat{v}) \hat{v} < 0$.

Let $y = x + \hat{v}$, by the mean value theorem,

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f[x + s(y - x)](y - x) \quad (2.18)$$

for some $s \in [0, 1]$.

This implies $f(y) < f(x) + \nabla f(x) \cdot (y - x)$, which contradicts the C^1 criterion for convexity. \blacksquare

2.4 Minimum and Maximum of Convex Functions

Theorem 2.3. Let $\Omega \subseteq \mathbb{R}^n$ be a convex set, and $f : \Omega \rightarrow \mathbb{R}$ is a convex function. Let

$$\Gamma := \left\{ x \in \Omega : f(x) = \min_{x \in \Omega} f(x) \right\} \equiv \operatorname{argmin}_{x \in \Omega} f(x) \quad (2.19)$$

If $\Gamma \neq \emptyset$, then

(i) Γ is convex;

(ii) any local minimum of f is the global minimum.

Proof (i). Let $x, y \in \Gamma$, $s \in [0, 1]$, then $sx + (1 - s)y \in \Omega$ because Ω is convex. Since f is convex, $f(sx + (1 - s)y) \leq sf(x) + (1 - s)f(y) = \min_{x \in \Omega} f(x)$. The inequality must be equality since it would contradict the fact that $x, y \in \Gamma$. Therefore, $sx + (1 - s)y \in \Gamma$. \blacksquare

Proof (ii). Let $x \in \Omega$ be a local minimizer for f , but assume, for contradiction, it is not a global minimizer. That is, there exists some other y such that $f(y) < f(x)$. Since f is convex,

$$f(x + t(y - x)) = f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y) < f(x) \quad (2.20)$$

for every $t \in (0, 1]$. Therefore, for every $\varepsilon > 0$, there exists $t^* \in (0, 1]$ such that $x + t^*(y - x) \in V_\varepsilon(x)$ and $f(x + t^*(y - x)) < f(x)$, this contradicts the fact that x is a local minimum. \blacksquare

Theorem 2.4. Let $\Omega \subseteq \mathbb{R}^n$ be a convex and compact set, and $f : \Omega \rightarrow \mathbb{R}$ is a convex function. Then

$$\max_{x \in \Omega} f(x) = \max_{x \in \partial\Omega} f(x) \quad (2.21)$$

Proof. As we assumed, Ω is closed, therefore $\partial\Omega \subseteq \Omega$. Hence, $\max_{x \in \Omega} f \geq \max_{x \in \partial\Omega} f$.

Suppose, for contradiction, $\max_{x \in \Omega} f > \max_{x \in \partial\Omega} f$, then $x^* := \operatorname{argmax}_{x \in \Omega} f \in \Omega^{int}$.

Then we can construct a straight line through x^* and intersects $\partial\Omega$ at two points, $y_1, y_2 \in \partial\Omega$, such that

$x^* = sy_1 + (1-s)y_2$ for some $s \in (0, 1)$. Further, since f is convex, $\max_{x \in \Omega} f(x) = f(x^*) \leq sf(y_1) + (1-s)f(y_2) \leq s \max_{\partial\Omega} f + (1-s) \max_{\partial\Omega} f = \max_{\partial\Omega} f$, which leads to a contradiction. Therefore, $\max_{x \in \Omega} f = \max_{x \in \partial\Omega} f$. ■

Proposition 2.3. For $p, g > 1$ satisfying $\frac{1}{p} + \frac{1}{g} = 1$,

$$|ab| \leq \frac{1}{p}|a|^p + \frac{1}{g}|b|^g \quad (2.22)$$

Proof.

$$(-\log)|ab| = (-\log)|a| + (-\log)|b| \quad (2.23)$$

$$= \frac{1}{p}(-\log)|a|^p + \frac{1}{g}(-\log)|b|^g \quad (2.24)$$

$$(\because (-\log) \text{ is convex}) \geq (-\log) \left(\frac{1}{p}|a|^p + \frac{1}{g}|b|^g \right) \quad (2.25)$$

And since $(-\log)$ is monotonically decreasing,

$$|ab| \leq \frac{1}{p}|a|^p + \frac{1}{g}|b|^g \quad (2.26)$$

■

Corollary 2.2.

$$|ab| \leq \frac{|a|^2 + |b|^2}{2} \quad (2.27)$$

3 Finite Dimensional Optimization

3.1 Unconstraint Optimization

Theorem 3.1 (Extreme Value Theorem). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and $K \subseteq \mathbb{R}^n$ be a compact set, then the minimization problem $\min_{x \in K} f(x)$ has a solution.

Remark 3.1. $f : \Omega \rightarrow \mathbb{R}$ is convex does not imply f is continuous.

Proposition 3.1. A convex function f defined on a convex open set is continuous.

Proof. Let $f : \Omega \rightarrow \mathbb{R}$ be a convex function, where $\Omega \subseteq \mathbb{R}^n$ is open. **TODO:** Is this true? ■

Proposition 3.2. A convex function f defined on an open interval in \mathbb{R} is continuous.

Proof. See homework 1. The proof involves squeeze theorem. ■

Proof of EVT.. Let $f : K \rightarrow \mathbb{R}$ be a continuous function defined on a compact set K .

WLOG, we only prove the existence of $\min f$, since the existence of \max can be easily proven by applying the exact same argument on $-f$.

That is, we claim the infimum of $f(K)$ is attained within K .

Because K is compact, the continuity of f implies $f(K)$ is compact.

By the completeness axiom of \mathbb{R} , $m := \inf_{x \in K} f(x)$ is well-defined. There exists a sequence $(x_i) \subseteq K$, such that $(f(x_i)) \rightarrow m$. Because K is compact, there exists a subsequence (x_{ik}) of (x_i) converges to some limit $x^* \in K$.

Since f is continuous, $(f(x_{ik})) \rightarrow f(x^*)$, which is a subsequence of the convergent sequence $(f(x_i))$, and they must converge to the same limit. Hence, $f(x^*) = m$, and the infimum is attained at $x^* \in K$. ■

Theorem 3.2 (Heine–Borel). Let $K \subseteq \mathbb{R}^n$, then the following are equivalent:

- (i) K is compact (every open cover of K has a finite sub-cover);
- (ii) K is closed and bounded;
- (iii) Every sequence in K has a convergent subsequence converges to a point in K .

Proposition 3.3. Let $\{h_i\}$ and $\{g_j\}$ be sets of continuous functions on \mathbb{R}^n , the the set of all points in \mathbb{R}^n that satisfy

$$\begin{cases} h_i(x) = 0 \quad \forall i \\ g_j(x) \leq 0 \quad \forall j \end{cases} \quad (3.1)$$

is a closed set. Moreover, if the qualified set is also bounded, then it is compact.

Proof. For every equality constraint h_i , it can be represented as the conjunction of two inequality constraint, namely $h_i^\alpha(x) := -h_i(x) \leq 0 \wedge h_i^\beta(x) := h_i(x) \leq 0$. Then the constraint collection is equivalent to

$$\begin{cases} h_i^\alpha(x) \leq 0 \quad \forall i \\ h_i^\beta(x) \leq 0 \quad \forall i \\ g_j(x) \leq 0 \quad \forall j \end{cases} \quad (3.2)$$

The subset of \mathbb{R}^n qualified by each individual constraint is closed by the property of continuous functions (i.e. a continuous function's pre-image of closed set is closed). And the intersection of arbitrarily many closed sets is closed. ■

Remark 3.2. Computer algorithms for solving minimization problems try to construct a sequence of (x_i) such that $f(x_i)$ decreases to $\min f$ rapidly.

The optimization problems investigated in this section can be formulated as

$$\min_{x \in \Omega} f(x) \quad (3.3)$$

where $\Omega \subseteq \mathbb{R}^n$. Typically, for simplicity, Ω are often \mathbb{R}^n , an open subset of \mathbb{R}^n , or the closure of some open subset of \mathbb{R}^n .

Everything above minimization discussed in this section is applicable to maximization as well using the proposition below.

Proposition 3.4. When $\Omega = \mathbb{R}^n$, the unconstrained minimization has the following properties

- (i) $\operatorname{argmax} f = \operatorname{argmin}(-f)$;
- (ii) $\max f = -\min(-f)$

Proof. Immediate by applying definitions of maximum and minimum. ■

Definition 3.1. A function $f : \Omega \rightarrow \mathbb{R}$ has **local minimum** at $x_0 \in \Omega$ if

$$\exists \varepsilon > 0 \text{ s.t. } \forall x \in V_\varepsilon(x_0) \cap \Omega, f(x_0) \leq f(x) \quad (3.4)$$

f attains **strictly local minimum** at x_0 if

$$\exists \varepsilon > 0 \text{ s.t. } \forall x \in V_\varepsilon(x_0) \cap \Omega \setminus \{x_0\} \quad f(x_0) < f(x) \quad (3.5)$$

f attains **global minimum** at x_0 if

$$\forall x \in \Omega \quad f(x_0) \leq f(x) \quad (3.6)$$

f attains **strict global minimum** at x_0 if

$$\forall x \in \Omega \setminus \{x_0\} \quad f(x_0) < f(x) \quad (3.7)$$

Note that strict global minimum is always unique.

Theorem 3.3 (Necessary Condition for Local Minimum). Let $C^1 \ni f : \Omega \rightarrow \mathbb{R}$, let $x_0 \in \Omega$ be a local minimum of f , then for every *feasible direction* v at x_0 ,

$$\nabla f(x_0) \cdot v \geq 0 \quad (3.8)$$

This theorem serves as the primary defining property of local minimum.

Definition 3.2. For $x_0 \in \Omega \subseteq \mathbb{R}^n$, $v \in \mathbb{R}^n$ is a **feasible direction** at x_0 if

$$\exists \bar{s} > 0 \text{ s.t. } \forall s \in [0, \bar{s}], x_0 + sv \in \Omega \quad (3.9)$$

Proof of Necessary Condition. Let $x_0 \in \Omega$ be a local minimum, and let v be a Define auxiliary function $g(s) := f(x_0 + sv)$. And since g attains minimum at $s = 0$, there exists some $\bar{s} > 0$ such that

$$g(s) - g(0) \geq 0 \quad \forall s \in [0, \bar{s}] \quad (3.10)$$

Therefore

$$g'(0) := \lim_{s \rightarrow 0} \frac{g(s) - g(0)}{s} \geq 0 \quad (3.11)$$

The alternative form of derivative can be derived using chain rule as

$$g'(0) = \nabla f(x_0 + sv) \cdot v \big|_{s=0} = \nabla f(x_0) \cdot v \quad (3.12)$$

By combining the two identities above, $\nabla f(x_0) \cdot v \geq 0$. ■

Alternative Proof of Necessary Condition (not that rigorous). The prove is almost immediate, if there exists a feasible direction v^* such that $\nabla f(x_0) \cdot v^* < 0$, for every $\varepsilon > 0$, one can construct $x' := x_0 + sv^*$ with sufficiently small s so that $x' \in V_\varepsilon(x_0) \cap \Omega$ and $f(x') < f(x_0)$. ■

Corollary 3.1. When Ω is open, then x_0 is a local minimum $\implies \nabla f(x_0) = 0$.

Proof. Since Ω is open, any sufficiently small $v \neq 0$ such that both v and $-v$ are feasible directions at x_0 , applying the necessary condition on both v and $-v$ provides the equality. ■

Theorem 3.4 (Second Order Necessary Condition for Local Minimum). Let $C^2 \ni f : \Omega \rightarrow \mathbb{R}$, let $x_0 \in \Omega$ be a local minimum of f , then for every non-zero feasible direction v at x_0 ,

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

$$(ii) \quad \nabla f(x_0) \cdot v = 0 \implies v^T \nabla^2 f(x_0) v \geq 0.$$

Proof. Let x_0 be a local minimum and v be a feasible direction at Ω , and $s \in (0, \bar{s}]$. The first statement is the immediate result of the first order necessary condition. Now suppose $\nabla f(x_0) = 0$, by the Taylor's theorem,

$$0 \leq f(x_0 + sv) - f(x_0) = s \nabla f(x_0) \cdot v + \frac{s^2}{2} v^T \nabla^2 f(x_0) v + o(s^2) \quad (3.13)$$

$$= \frac{s^2}{2} v^T \nabla^2 f(x_0) v + o(s^2) \quad (3.14)$$

Since $s^2 > 0$, divide both sides by s^2 and take limit,

$$\lim_{s \rightarrow 0} \frac{f(x_0 + sv) - f(x_0)}{s^2} = \lim_{s \rightarrow 0} \left\{ \frac{1}{2} v^T \nabla^2 f(x_0) v + \frac{o(s^2)}{s^2} \right\} \quad (3.15)$$

$$= \frac{1}{2} v^T \nabla^2 f(x_0) v + \lim_{s \rightarrow 0} \frac{o(s^2)}{s^2} \quad (3.16)$$

$$= \frac{1}{2} v^T \nabla^2 f(x_0) v \geq 0 \quad (3.17)$$

■

Example 3.1. $f(x, y) = x^2 - xy + y^2 - 3y : \Omega = \mathbb{R}^2 \rightarrow \mathbb{R}$. Then at $(x_0, y_0) = (1, 2)$,

$$\nabla f(x_0, y_0) = (2x_0 - y, -x_0 + 2y_0 - 3) = (0, 0) \quad (3.18)$$

$$\nabla^2 f(x_0, y_0) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \succ 0 \quad (3.19)$$

Definition 3.3. Let $A \in \mathbb{R}^{n \times n}$, A is

- (i) **Positive definite** (denoted as $A \succ 0$) if $x^T A x > 0 \forall x \neq 0$, if and only if all eigenvalues $\lambda_i > 0$;
- (ii) **Positive Semi-definite** (denoted as $A \succeq 0$) if $x^T A x \geq 0 \forall x \in \mathbb{R}^n$, if and only if all eigenvalues $\lambda_i \geq 0$.

Theorem 3.5 (Sylvester's Criterion). Let $A \in \mathbb{R}^{n \times n}$ be a Hermitian matrix (i.e. $A = \overline{A^T}$)¹, then

1. $A \succ 0 \iff$ all *leading principal minors* have positive determinants;
2. $A \succeq 0 \iff$ all leading principal minors have non-negative determinants.

Theorem 3.6 (Second Order Sufficient Condition for Interior Local Minima). Let $f : C^2(\Omega, \mathbb{R})$, for some $x_0 \in \Omega$, if

- (i) $\nabla f(x_0) = 0$,
- (ii) (and) $\nabla^2 f(x_0) \succ 0$.

then x_0 is a strictly local minimizer.

Lemma 3.1. Suppose $\nabla^2 f(x_0)$ is positive definite, then

$$\exists a > 0 \text{ s.t. } v^T \nabla^2 f(x_0) v \geq a \|v\|^2 \quad \forall v \quad (3.20)$$

That is, the quadratic form of a positive definite matrix is bounded away from zero.

¹ $\overline{A^T}$ denotes the complex conjugate of the transpose, a matrix with *real entries* is Hermitian if and only if it is symmetric.

Proof of the Lemma. Recall that a squared matrix Q is called **orthogonal** when every column and row of it is an orthogonal unit vector. So that for every orthogonal matrix Q , $Q^T Q = I$, which implies $Q^T = Q^{-1}$. Further, note that

$$\|Qv\|^2 = (Qv)^T(Qv) = v^T Q^T Q v = \|v\|^2 \quad (3.21)$$

$$\implies \|Qv\| = \|v\| \quad \forall v \in \mathbb{R}^n \quad (3.22)$$

Let $v \in \mathbb{R}^n$, consider the eigenvector decomposition of $\nabla^2 f(x_0)$, let w satisfy $v = Qw$:

$$Q^T \nabla^2 f(x_0) Q = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (3.23)$$

$$\implies v^T \nabla^2 f(x_0) v = (Qw)^T \nabla^2 f(x_0) (Qw) \quad (3.24)$$

$$= w^T Q^T \nabla^2 f(x_0) Q w \quad (3.25)$$

$$= w^T \text{diag}(\lambda_1, \dots, \lambda_n) w \quad (3.26)$$

$$= \lambda_1 w_1^2 + \dots + \lambda_n w_n^2 \quad (3.27)$$

Let $a := \min\{\lambda_1, \dots, \lambda_n\}$,

$$\dots \geq a\|w\|^2 = a\|Q^T v\|^2 = a\|v\|^2 \quad (3.28)$$

■

Proof of the Theorem. Let $x \in \Omega$, suppose $\nabla f(x_0) = 0$ and $\nabla^2 f(x_0) \succcurlyeq 0$. By the second order Taylor approximation,

$$f(x_0 + v) - f(x_0) = \nabla f(x_0)^T v + \frac{1}{2} v^T \nabla^2 f(x_0) v + o(\|v\|^2) \quad (3.29)$$

$$= \frac{1}{2} v^T \nabla^2 f(x_0) v + o(\|v\|^2) \quad (3.30)$$

$$\geq \frac{a}{2} \|v\|^2 + o(\|v\|^2) \text{ for some } a > 0 \quad (3.31)$$

$$= \|v\|^2 \left(\frac{a}{2} + \frac{o(\|v\|^2)}{\|v\|^2} \right) \quad (3.32)$$

$$> 0 \text{ for sufficiently small } v \quad (3.33)$$

Therefore, $f(x_0) < f(x) \quad \forall x \in V_\varepsilon(x_0)$. ■

3.2 Equality Constraints: Lagrangian Multiplier

3.2.1 Tangent Space to a (Hyper) Surface at a Point

Definition 3.4. A surface $\mathcal{M} \subseteq \mathbb{R}^n$ is defined as

$$\mathcal{M} := \{x \in \mathbb{R}^n : h_i(x) = 0 \quad \forall i\} \quad (3.34)$$

where h_i are all C^1 functions.

Definition 3.5. A **differentiable curve** on a surface \mathcal{M} is a C^1 function mapping from $(-\varepsilon, \varepsilon)$ to \mathcal{M} .

Remark: in previous calculus courses, differentiable curves are often referred to as parameterizations.

Let $x(s)$ be a differentiable curve on \mathcal{M} passes through $x_0 \in \mathcal{M}$, re-parameterize it so that $x(0) = x_0$.

Then vector

$$v := \left. \frac{d}{ds} \right|_{s=0} x(s) \quad (3.35)$$

touches \mathcal{M} *tangentially*.

Definition 3.6. Any vector v generated by some differentiable curve on \mathcal{M} and takes above form is a **tangent vector** on \mathcal{M} through x_0 .

Definition 3.7. The **tangent space** to \mathcal{M} at x_0 is defined to be the set of all tangent vectors:

$$T_{x_0}\mathcal{M} := \left\{ v \in \mathbb{R}^n : v := \left. \frac{d}{ds} \right|_{s=0} x(s) \text{ for some } x \in C^1((-\varepsilon, \varepsilon), \mathcal{M}) \text{ s.t. } x(0) = x_0 \right\} \quad (3.36)$$

Example 3.2. Define

$$\mathcal{M} := \{x \in \mathbb{R}^2 : \|x\|_2 = 1\} \quad (3.37)$$

By defining C^1 functions $g(x) := \|x\|_2^2 - 1$, \mathcal{M} is a surface. The tangent space of \mathcal{M} at x_0 is

$$T_{x_0}\mathcal{M} = \{v \in \mathbb{R}^n : \langle v, x_0 \rangle = 0\} \quad (3.38)$$

Definition 3.8. Let \mathcal{M} be a surface defined using C^1 functions, a point $x_0 \in \mathcal{M}$ is a **regular point** of the constraints if

$$\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\} \quad (3.39)$$

are linearly independent.

Remark: if there is only one constraint h , then x_0 is regular if and only if $\nabla h(x_0) \neq 0$.

Notation 3.1. Define the T space on equality constraint as

$$T_{x_0} := \{x \in \mathbb{R}^n : \langle x_0, \nabla h_i(x_0) \rangle = 0 \ \forall i \in [k]\} \quad (3.40)$$

Example 3.3 (Counter example). Define

$$\mathcal{M} := \{(x, y) \in \mathbb{R}^2 : h(x, y) = xy = 0\} \quad (3.41)$$

Then it is easy to verify that $(0, 0)$ is not a regular point. And

$$T_{0,0} = \{(x, y) \in \mathbb{R}^2 : (x, y) \cdot (0, 0) = 0\} = \mathbb{R}^2 \quad (3.42)$$

$$\neq T_{0,0}\mathcal{M} = \{(x, y) \in \mathbb{R}^2 : x = 0 \vee y = 0\} \quad (3.43)$$

Theorem 3.7. Suppose x_0 is a *regular point* of $\mathcal{M} := \{h_i(x) = 0, i = 1, \dots, k\}$, then $T_{x_0} = T_{x_0}\mathcal{M}$.

Proof. Show $T_{x_0}\mathcal{M} \subseteq T_{x_0}$.

Suppose x_0 is a regular point of \mathcal{M} . Let $v \in T_{x_0}\mathcal{M}$, then there exists some differentiable curve $x(\cdot) : V_\varepsilon(0) \rightarrow \mathcal{M}$ such that $x(0) = x_0$, such that

$$v = \left. \frac{d}{ds} \right|_{s=0} x(s) \quad (3.44)$$

Note that $h_i(x(s)) = 0$ is constant for every $i \in [k]$, therefore

$$\left. \frac{d}{ds} \right|_{s=0} h_i(x(s)) \quad (3.45)$$

By the chain rule,

$$\nabla h_i(x_0) \cdot v = 0 \quad \forall i \quad (3.46)$$

Therefore $v \in T_{x_0}$.

Show $T_{x_0} \subseteq T_{x_0}\mathcal{M}$.

(i) x_0 is regular $\implies T_{x_0}\mathcal{M}$ is a vector space;

(ii) $T_{x_0} = \text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\}^\perp$.

Show $T_{x_0} \subseteq \text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\}^\perp$:

Let $v \in T_{x_0}$, then $v \perp \nabla h_i(x_0)$ for every i . Therefore v is orthogonal to every linear combination of $\nabla h_i(x_0)$, and therefore orthogonal to the span.

Show $\text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\}^\perp \subseteq T_{x_0}$:

Let v in the perp of the span, then v is orthogonal to every basis of the span, so $v \in T_{x_0}$. ■

Lemma 3.2. Let $f, h_1, \dots, h_k \in C^1$ defined on open subset $\Omega \subseteq \mathbb{R}^n$. Define $\mathcal{M} := \{x \in \mathbb{R}^n : h_i(x) = 0 \quad \forall i\}$. Suppose $x_0 \in \mathcal{M}$ is a local minimum of f on \mathcal{M} , then

$$\nabla f(x_0) \perp T_{x_0}\mathcal{M} \quad (3.47)$$

Proof. WLOG $\Omega = \mathbb{R}^n$, take $v \in T_{x_0}\mathcal{M}$. Then there exists some differentiable curve x on \mathcal{M} satisfying $v = x'(0)$. Because x_0 is a local minimum of f on Ω , $s = 0$ is a local minimum of $f(x(s))$, moreover, it is an interior minimum. By chain rule and the necessary condition of local minimum,

$$Df(x(0)) = \nabla f(x(0)) \cdot x'(0) = 0 \quad (3.48)$$

$$\implies \nabla f(x_0) \cdot v = 0 \quad (3.49)$$

Therefore $\nabla f(x_0) \perp T_{x_0}\mathcal{M}$. ■

Theorem 3.8 (Lagrange Multipliers: First Order Necessary Condition). Let $f, h_1, \dots, h_k \in C^1$ defined on open subset $\Omega \subseteq \mathbb{R}^n$. Let x_0 be a regular point of the constraint set $\mathcal{M} := \bigcap_{i=1}^k h_i^{-1}(0)$. Suppose x_0 is a local minimum of \mathcal{M} , then there exists $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ such that

$$\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) = 0 \quad (3.50)$$

Remark: if we define Lagrangian $\mathcal{L}(x, \lambda_i) := f(x) + \sum_{i=1}^k \lambda_i h_i(x)$, then the theorem says the local minimum is a critical point of \mathcal{L} .

Proof. Because x_0 is a regular point, then by previous lemma, $\nabla f(x_0) \perp T_{x_0}\mathcal{M}$. Moreover,

$$T_{x_0}\mathcal{M} = T_{x_0} = (\text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\})^\perp \quad (3.51)$$

Also, because x_0 is a local minimum,

$$\nabla f(x_0) \perp T_{x_0}\mathcal{M} \quad (3.52)$$

Therefore, $\nabla f(x_0) \in (T_{x_0}\mathcal{M})^\perp = (\text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\})^{\perp\perp} = \text{span}\{\nabla h_1(x_0), \dots, \nabla h_k(x_0)\}$, where the last equality holds in finite dimensional cases. Hence, it is obvious that we can write $\nabla f(x_0)$ as a linear combination of $\{\nabla h_i(x_0)\}$. ■

Theorem 3.9 (Second Order Necessary Condition). Let $f, h_i \in C^2$, if x_0 is a local minimum on previously defined surface \mathcal{M} , then there exists Lagrangian multipliers $\{\lambda_i\}$ such that

- (i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) = 0$ ($\nabla_x \mathcal{L} = 0$);
- (ii) And $\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \succcurlyeq 0$ on $T_{x_0}\mathcal{M}$ ($\nabla_x^2 \mathcal{L} \succcurlyeq 0$).

Remark: whenever x_0 is a local minimum, it must be a critical point of \mathcal{L} , and \mathcal{L} is positive semidefinite on the tangent space at x_0 .

Proof. The first result is exactly the same as the first order condition proven above.

To show the second result, let $x(s) \in \mathcal{M}$ be an arbitrary differentiable curve on \mathcal{M} such that $x(0) = x_0$. Then,

$$\frac{d}{ds} f(x(s)) = \nabla f(x(s)) \cdot x'(s) \quad (3.53)$$

$$\frac{d^2}{ds^2} f(x(s)) = x'(s)^T \nabla^2 f(x(s)) x'(s) + \nabla f(x(s)) x''(s) \quad (3.54)$$

By the second order Taylor theorem, for every s such that $x(s) \in \mathcal{M}$,

$$f(x(s)) - f(x_0) = s \nabla f(x_0) \cdot x'(0) + \frac{s^2}{2} [x'(0)^T \nabla^2 f(x_0) x'(0) + \nabla f(x_0) x''(0)] + o(s^2) \quad (3.55)$$

Note that by definition, $x'(0)$ is in the tangent space at x_0 . Also, we've shown previously that $\nabla f(x_0)$ is orthogonal to the tangent space at x_0 , therefore,

$$f(x(s)) - f(x_0) = \frac{s^2}{2} [x'(0)^T \nabla^2 f(x_0) x'(0) + \nabla f(x_0) x''(0)] + o(s^2) \quad (3.56)$$

Also, by the definition of \mathcal{M} , all constraints hold with equality:

$$f(x_0) = f(x_0) + \sum_{i=1}^k \lambda_i h_i(x_0) \quad (3.57)$$

where λ_i 's are from the first result. Hence,

$$f(x(s)) - f(x_0) = \frac{s^2}{2} \left[x'(0)^T \left(\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \right) x'(0) + \left(\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) \right) x''(0) \right] + o(s^2) \quad (3.58)$$

$$= \frac{s^2}{2} x'(0)^T \left(\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \right) x'(0) + o(s^2) \quad (3.59)$$

And above expression is greater or equal to zero because x_0 is a local minimum,

$$\frac{s^2}{2} x'(0)^T \left(\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \right) x'(0) + o(s^2) \geq 0 \quad (3.60)$$

$$\implies x'(0)^T \left(\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \right) x'(0) + \frac{o(s^2)}{s^2} \geq 0 \quad (3.61)$$

$$\xrightarrow{s \rightarrow 0} x'(0)^T \left(\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) \right) x'(0) \geq 0 \quad (3.62)$$

Where $x'(0)$ is a vector in the tangent space at x_0 by definition. Moreover, the curve $x(s)$ was chosen arbitrarily, so the argument works for every curve and therefore every tangent vector, and what's desired is shown. ■

Example 3.4.

$$\min f(x, y) = x^2 - y^2 \quad (3.63)$$

$$s.t. \ h(x, y) = y = 0 \quad (3.64)$$

First order condition suggests $(x_0, y_0) = (0, 0)$ Note that the tangent space at (x_0, y_0) is $\text{span}\{\nabla h_i\}^\perp$:

$$T_{x_0} \mathcal{M} = \{(u, 0) : u \in \mathbb{R}\} \quad (3.65)$$

and

$$\nabla_x^2 \mathcal{L} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \quad (3.66)$$

is obviously positive semidefinite (actually positive definition) on the tangent space.

Theorem 3.10 (Second Order Sufficient Conditions). Let $f, h_i \in C^2$ on open $\Omega \subseteq \mathbb{R}^n$, and $x_0 \in \mathcal{M}$ is a regular point, if there exists $\lambda_i \in \mathbb{R}$ such that

$$(i) \ \nabla_x \mathcal{L}(x_0, \lambda_i) = 0;$$

$$(ii) \ \nabla_x^2 \mathcal{L}(x_0, \lambda_i) \succ 0 \text{ on } T_{x_0} \mathcal{M},$$

then x_0 is a *strict* local minimum.

Proof. Recall that $\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)$ positive definite on $T_{x_0} \mathcal{M}$ implies there exists $a > 0$ (a is taken to be equal to the least eigenvalue of $\nabla_x^2 \mathcal{L}$) such that

$$v^T [\nabla^2 f(x_0) + \sum \lambda_i \nabla^2 h_i(x_0)] v \geq a \|v\|^2 \quad \forall v \in T_{x_0} \mathcal{M} \quad (3.67)$$

Let $x(s) \in \mathcal{M}$ be a curve such that $x(0) = x_0$ and $v = x'(0)$. WLOG, $\|x'(0)\| = 1$. By the second order

Taylor expansion,

$$f(x(s)) - f(x(0)) = s \left. \frac{d}{ds} \right|_{s=0} f(x(s)) + \frac{s^2}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} f(x(s)) + o(s^2) \quad (3.68)$$

$$= s \left. \frac{d}{ds} \right|_{s=0} \left[f(x(s)) + \sum \lambda_i h_i(x(s)) \right] + \frac{s^2}{2} \left. \frac{d^2}{ds^2} \right|_{s=0} \left[f(x(s)) + \sum \lambda_i h_i(x(s)) \right] + o(s^2) \quad (3.69)$$

$$= s \nabla_x \mathcal{L}(x_0, \lambda_i) \cdot x'(0) + \frac{s^2}{2} [x'(0)^T \nabla_x^2 \mathcal{L}(x_0, \lambda_i) x'(0) + \nabla_x \mathcal{L}(x_0, \lambda_i) x''(0)] + o(s^2) \quad (3.70)$$

$$= \frac{s^2}{2} x'(0)^T \nabla_x^2 \mathcal{L}(x_0, \lambda_i) x'(0) + o(s^2) \quad (3.71)$$

$$\geq \frac{s^2}{2} a \|x'(0)\|^2 + o(s^2) \quad \text{where } a > 0 \quad (3.72)$$

$$= s^2 \left(\frac{a}{2} + \frac{o(s^2)}{s^2} \right) \quad (3.73)$$

$$\stackrel{s \rightarrow 0}{>} 0 \quad (3.74)$$

Therefore, for sufficiently small s , $f(x(s)) - f(x(0)) > 0$. And this is true for every curve x on \mathcal{M} . So $x(0)$ is a strict local minimum. \blacksquare

3.3 Remark on the Connection Between Constrained and Unconstrained Optimizations

Example 3.5. Consider

$$\min f(x, y, z) \quad (3.75)$$

$$s.t. g(x, y, z) = z - h(x, y) = 0 \quad (3.76)$$

where \mathcal{M} is the graph of h . Using Lagrangian multiplier provides necessary condition: $\nabla f + \lambda \nabla g = 0$,

$$\begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} + \lambda \begin{pmatrix} -h_x \\ -h_y \\ 1 \end{pmatrix} = 0 \quad (3.77)$$

Convert the constrained optimization into an unconstrained optimization as

$$\min_{(x,y) \in \mathbb{R}^2} F(x, y) = f(x, y, h(x, y)) \quad (3.78)$$

The necessary condition for unconstrained optimization is

$$\nabla F(x, y) = \begin{pmatrix} f_x + f_z h_x \\ f_y + f_z h_y \end{pmatrix} \quad (3.79)$$

$$= \begin{pmatrix} f_x \\ f_y \end{pmatrix} - f_z \begin{pmatrix} -h_x \\ -h_y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (3.80)$$

Define $\lambda := -f_z$.

$$\nabla F(x, y) = \begin{pmatrix} f_x \\ f_y \\ f_z \end{pmatrix} + \lambda \begin{pmatrix} -h_x \\ -h_y \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (3.81)$$

3.4 Inequality Constraints

Definition 3.9. Let x_0 satisfy the set of constraints

$$(\dagger) \begin{cases} h_i(x) = 0 & i \in \{1, \dots, k\} \\ g_j(x) \leq 0 & j \in \{1, \dots, \ell\} \end{cases} \quad (3.82)$$

we say that the constraint g_i is **active** at x_0 if $g_i(x_0) = 0$, and is **inactive** at x_0 if $g_i(x_0) < 0$.

Definition 3.10. Split the collection of inequality constraints into active and inactive constraints, let $\Theta(x_0)$ denote the collection of active indices, that's:

$$g_j(x_0) = 0 \quad \forall j \in \Theta(x_0) \quad (3.83)$$

$$g_j(x_0) < 0 \quad \forall j \notin \Theta(x_0) \quad (3.84)$$

Then x_0 is said to be a **regular point** of the constraint if

$$\{\nabla h_i(x_0) \quad \forall i \in \{1, \dots, k\}; \underbrace{\nabla g_j(x_0) \quad \forall j \in \Theta(x_0)}_{\text{Active Constraints}}\} \quad (3.85)$$

is linearly independent.

Theorem 3.11 (The First Order Necessary Condition for Local Minimum: Kuhn-Tucker Conditions). Let Ω be an open subset of \mathbb{R}^n with constraints h_i and g_i to be C^1 on Ω . Suppose $x_0 \in \Omega$ is a regular point with respect to constraints, further suppose x_0 is a local minimum, then there exists some $\lambda_i \in \mathbb{R}$ and $\mu_j \in \mathbb{R}_+$ such that

- (i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) = 0$ (i.e. $\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$);
- (ii) $\mu_j g_j(x_0) = 0$ (*Complementary slackness*).

Remark 1: by complementary slackness, all μ_j corresponding to inactive inequality constraints are zero.

Remark 2: it is possible for an active constraint to have zero multiplier.

Proof. Let x_0 be a local minimum for f satisfying constraints, equivalently, it is a local minimum for equality constraints and active inequality constraints.

By the first order necessary condition for local minimum with equality constraints, there exists $\lambda_i, \mu_j \in \mathbb{R}$ such that

$$\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j \in \Theta(x_0)} \mu_j \nabla g_j(x_0) = 0 \quad (3.86)$$

Then by setting $\mu_j = 0$ for all $j \notin \Theta(x_0)$ one have

$$\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) = 0 \quad (3.87)$$

By construction, the complementary slackness is satisfied. At this stage, we have construct $\lambda_i \in \mathbb{R}$ and $\mu_j \in \mathbb{R}$ satisfying both conditions, we still need to argue that $\mu_j \geq 0$ for every j . ■

Theorem 3.12 (The Second Order Necessary Conditions). Let Ω be an open subset of \mathbb{R}^n , and $f, h_1, \dots, h_k, g_1, \dots, g_\ell \in C^2(\mathbb{R}^n, \mathbb{R})$. Let x_0 be a regular point of the constraints (†). Suppose x_0 is a local minimum of f subject to constraint (†), then there exists $\lambda_i \in \mathbb{R}$ and $\mu_j \geq 0$ such that

- (i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j=1}^\ell \mu_j \nabla g_j(x_0) = 0$;
- (ii) $\mu_j g_j(x_0) = 0$;
- (iii) $\nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) + \sum_{j=1}^\ell \mu_j \nabla^2 g_j(x_0)$ is positive semidefinite on the tangent space to activate constraints at x_0 .

Proof. (i) and (ii) are immediate result from the first order necessary condition.

Suppose x_0 is a local minimum for (†), then x_0 is a local minimum for active constraints at x_0 .

Therefore, $\nabla^2 \hat{\mathcal{L}} = \nabla^2 f(x_0) + \sum_{i=1}^k \lambda_i \nabla^2 h_i(x_0) + \sum_{j \in I(x_0)} \mu_j \nabla^2 g_j(x_0)$ is positive semidefinite on the tangent space to active constraints. Note that because $\mu_j = 0$ for inactive constraints, therefore $\nabla^2 \hat{\mathcal{L}} = \nabla^2 \mathcal{L}$ at x_0 , and both of them are positive semidefinite on the tangent space corresponding to active constraints. ■

Theorem 3.13 (The Second Order Sufficient Conditions). Let Ω be an open subset of \mathbb{R}^n , let $f, h_i, q_j \in C^2(\Omega)$. Consider minimizing $f(x)$ with the constraint

$$(\dagger) \begin{cases} h_i(x) = 0 & \forall i \\ g_j(x) \leq 0 & \forall j \\ x \in \Omega \end{cases} \quad (3.88)$$

Suppose there exists a feasible x_0 satisfying (†) and $\lambda_i \in \mathbb{R}$ and $\mu_j \in \mathbb{R}_+$ such that

- (i) $\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla h_i(x_0) + \sum_{j=1}^\ell \mu_j \nabla g_j(x_0) = 0$;
- (ii) $\mu_j g_j(x_0) = 0$ (*Complementary slackness*).

If the Hessian matrix for Lagrangian $\nabla_x^2 \mathcal{L}(x_0)$ is positive definite on \tilde{T}_{x_0} , the space of **strongly active** constraints at x_0 , then x_0 is a strict local minimum.

Definition 3.11. A constraint g_j is **strongly active** at x_0 if $g_j(x_0) = 0$ (so it is active) and $\mu_j > 0$.

Notation 3.2. For convenience, we can rearrange the collection of constraints such that, among the ℓ constrains in total, the first ℓ' constraints are *active* at x_0 and the first ℓ'' constraints are *strongly active*. Note that $\ell'' \leq \ell' \leq \ell$.

Define

$$\tilde{T}_{x_0} := \{v \cdot \nabla h_i(x_0) = 0 \ \forall i \wedge v \cdot \nabla g_j(x_0) \text{ for all } g_j \text{ active.}\} \quad (3.89)$$

$$\tilde{\tilde{T}}_{x_0} := \{v \cdot \nabla h_i(x_0) = 0 \ \forall i \wedge v \cdot \nabla g_j(x_0) \text{ for all } g_j \text{ strongly active.}\} \quad (3.90)$$

Clearly, $\tilde{\tilde{T}}_{x_0} \subseteq \tilde{T}_{x_0}$ because there are (weakly) more active constraints than strongly active constraints.

Proof of the Sufficient Condition. Suppose, for contradiction, x_0 is not a strict local minimum.

Claim 1: There exists 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 every i ;
(c) $\nabla g_j(x_0) \cdot v \leq 0$ for all $j \leq \ell'$ (active constraints).

Proof of Claim 1. Because x_0 is not a strictly local minimum, one can construct a sequence of feasible points $(x_k) \rightarrow x_0$ by setting $\varepsilon = \frac{1}{k}$ for every $k \in \mathbb{N}$ such that $f(x_k) \leq f(x_0)$.

Let $v_k := \frac{x_k - x_0}{\|x_k - x_0\|}$, $s_k := \|x_k - x_0\|$. Note that every v_k is in unit sphere, which is compact. Therefore, there exists a subsequence of (v_k) converges to some unit vector v .

$$0 \geq f(x_k) - f(x_0) = f(x_0 + s_k v_k) - f(x_0) \quad \forall k \in \mathbb{N} \quad (3.91)$$

The first order Taylor series suggests the following holds for every $k \in \mathbb{N}$:

$$0 \geq f(x_0 + s_k v_k) - f(x_0) \quad (3.92)$$

$$= s_k \nabla f(x_0) \cdot v_k + o(s_k) \quad (3.93)$$

$$0 = h_i(x_0 + s_k v_k) - h_i(x_0) = s_k \nabla h_i(x_0) \cdot v_k + o(s_k) \quad (3.94)$$

$$0 \geq g_j(x_0 + s_k v_k) - g_j(x_0) = s_k \nabla g_j(x_0) \cdot v_k + o(s_k) \quad \forall j \leq \ell' \quad (3.95)$$

Above inequalities are preserved by limit operation, therefore,

$$\nabla f(x_0) \cdot v_k + \frac{o(s_k)}{s_k} \rightarrow \nabla f(x_0) \cdot v \leq 0 \quad (3.96)$$

$$\nabla h_i(x_0) \cdot v_k + \frac{o(s_k)}{s_k} \rightarrow \nabla h_i(x_0) \cdot v = 0 \quad (3.97)$$

$$\nabla g_j(x_0) \cdot v_k + \frac{o(s_k)}{s_k} \rightarrow \nabla g_j(x_0) \cdot v \leq 0 \quad \forall j \leq \ell' \quad (3.98)$$

■

Claim 2: $\nabla g_j(x_0) \cdot v = 0$ for $j = 1, \dots, \ell''$.

Proof of Claim 2. Suppose not, there exists $j \in \{1, \dots, \ell''\}$ such that $\nabla g_j(x_0) \cdot v < 0$. Then by (i),

$$0 \geq \nabla f(x_0) \cdot v = - \sum_{i=1}^k \lambda_i \nabla h_i(x_0) \cdot v - \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) \cdot v \quad (3.99)$$

$$= - \sum_{j=1}^{\ell} \mu_j \nabla g_j(x_0) \cdot v > 0 \quad (3.100)$$

the last inequality is from the fact that $\mu_j \nabla g_j(x_0) \cdot v \leq 0$ for all active constraints and $\mu_j = 0$ for all inactive constraints.

■

(b) and claim 2 suggests $v \in \tilde{T}_{x_0}$.

By the second order Taylor approximation,

$$0 \geq f(x_k) - f(x_0) = s_k \nabla f(x_0) \cdot v_k + \frac{s_k^2}{2} v_k \cdot \nabla^2 f(x_0) \cdot v_k + o(s_k^2) \quad (3.101)$$

$$0 = h_i(x_k) - h_i(x_0) = s_k \nabla h_i(x_0) \cdot v_k + \frac{s_k^2}{2} v_k \cdot \nabla^2 h_i(x_0) \cdot v_k + o(s_k^2) \quad \forall i \quad (3.102)$$

$$0 \geq g_j(x_k) - g_j(x_0) = s_k \nabla g_j(x_0) \cdot v_k + \frac{s_k^2}{2} v_k \cdot \nabla^2 g_j(x_0) \cdot v_k + o(s_k^2) \quad \forall j \leq \ell' \quad (3.103)$$

Multiply the second equation by λ_i and third equation by μ_j , and use the fact that $\mu_j = 0$ for every $j > \ell'$. Also, given $\nabla \mathcal{L} = 0$ in (i):

$$0 \geq \frac{s_k^2}{2} v_k \cdot \nabla^2 \mathcal{L} \cdot v_k + o(s_k^2) \quad (3.104)$$

Divide by s_k^2 and take the limit $(v_k) \rightarrow v$:

$$v \cdot \nabla^2 \mathcal{L} \cdot v \leq 0 \quad (3.105)$$

which contradicts the assumption that $\nabla^2 \mathcal{L}$ is positive definite in \tilde{T}_{x_0} because we've shown that $v \in \tilde{T}_{x_0}$. ■

4 Iterative Algorithms for Optimization

4.1 Newton's Method

Example 4.1 (Motivation: a second order iterative algorithm). Let $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ where I is an open interval. Let $x_i \in I$ be a starting point, consider the second order linear approximation of f at x_0 :

$$g(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 \quad (4.1)$$

By construction, the second order Taylor polynomial, $g(x)$, is the best second order approximation to f at x_0 in the following sense:

$$g(x_0) = f(x_0) \quad (4.2)$$

$$g'(x_0) = f'(x_0) \quad (4.3)$$

$$g''(x_0) = f''(x_0) \quad (4.4)$$

The Newton's method aims to solve the critical point of $g(x)$ and define x_1 to be the critical point found:

$$g'(x_1) = f'(x_0) + f''(x_0)(x_1 - x_0) = 0 \quad (4.5)$$

$$\implies x_1 \leftarrow x_0 - \frac{f'(x_0)}{f''(x_0)} \quad (4.6)$$

Algorithm 4.1 (Newton's Method in \mathbb{R}). Given initial point $x_0 \in I$, while not terminated:

$$x_{n+1} \leftarrow x_n - \frac{f'(x_n)}{f''(x_n)} \quad (4.7)$$

Theorem 4.1. Let $f \in C^3$ on open interval $I \subseteq \mathbb{R}$. Suppose $x_* \in I$ satisfies $f'(x_*) = 0$ and $f''(x_*) \neq 0$, then the sequence of points (x_n) generated by Newton's method converges to x_* if x_0 is sufficiently close to x_* .

Example 4.2. Let $f(x) = x^2$, then $\frac{f'(x)}{f''(x)} = \frac{2x}{2} = x$. For any starting point x_0 , $x_1 = x_0 - \frac{2x_0}{2} = 0$. That is, the algorithm converges to the global minimum in one iteration.

Proof. Let $g(x) = f'(x)$ so that $x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$.

Because $f \in C^3$, then $g \in C^2$.

Note that by $g \in C^2$, $g' = f''$ is bounded away from zero near x_* .

And by continuity again, $g'' = f^{(3)}$ is bounded near the bounded region $V_\varepsilon(x_*)$.

That is, within small region near x_* , $V_\delta(x_*)$, there exists a sufficiently small $\alpha > 0$ such that

$$\begin{cases} |g'(x_1)| > \alpha \quad \forall x_1 \in V_\delta(x_*) \\ |g''(x_2)| < \frac{1}{\alpha} \quad \forall x_2 \in V_\delta(x_*) \end{cases} \quad (4.8)$$

Further, note that $g(x_*) = f'(x_*) = 0$.

WLOG, let $n \in \mathbb{N}$, suppose $x_n > x_*$:

$$x_{n+1} - x_* = x_n - \frac{g(x_n)}{g'(x_n)} - x_* \quad (4.9)$$

$$= x_n - x_* - \frac{g(x_n) - g(x_*)}{g'(x_n)} \quad (4.10)$$

$$= -\frac{g(x_n) - g(x_*) - g'(x_n)(x_n - x_*)}{g'(x_n)} \quad (4.11)$$

$$= -\frac{1}{2} \frac{g''(\xi)}{g'(x_n)} (x_n - x_*)^2 \quad \text{for some } \xi \in (x_*, x_n) \quad (4.12)$$

By taking the absolute values on both sides:

$$|x_{n+1} - x_*| = \frac{1}{2} \frac{|g''(\xi)|}{|g'(x_n)|} |x_n - x_*|^2 \quad (4.13)$$

$$< \frac{1}{2\alpha^2} |x_n - x_*|^2 \quad (4.14)$$

Let $\rho := \frac{1}{\alpha^2} |x_0 - x_*|^2$, choose x_0 sufficiently close to x_* such that $\rho < 1$.

Remark: we are showing the iterative algorithm induces a contraction map.

Then,

$$|x_1 - x_*| < \frac{1}{2\alpha^2} |x_0 - x_*|^2 \quad (4.15)$$

$$= \frac{1}{2\alpha^2} |x_0 - x_*| |x_0 - x_*| \quad (4.16)$$

$$= \rho |x_0 - x_*| \quad (4.17)$$

Inductively,

$$|x_2 - x_*| < \frac{1}{2\alpha^2} |x_1 - x_*|^2 \quad (4.18)$$

$$< \frac{1}{2\alpha^2} \rho^2 |x_0 - x_*|^2 \quad (4.19)$$

$$= \rho^3 |x_0 - x_*| \quad (4.20)$$

$$< \rho^2 |x_0 - x_*| \quad (4.21)$$

By induction,

$$|x_n - x_*| < \rho^2 |x_0 - x_*| \quad (4.22)$$

Therefore, as $n \rightarrow \infty$, $(x_n) \rightarrow x_*$. ■

Theorem 4.2 (2nd Order MVT).

$$g(x) = g(y) + g'(y)(x - y) + \frac{1}{2}g''(\xi)(x - y)^2 \quad \xi \in (x, y) \quad (4.23)$$

Algorithm 4.2 (Newton's Method in \mathbb{R}^n). Let $f : \Omega \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ where Ω is open, let initial point $x_0 \in \Omega$. Suppose $\nabla^2 f(x_n)$ is invertible for every generated n , and $\nabla f(x_*) = 0$ so that algorithm stops at minimum. The iterative algorithm is defined as following:

$$x_{n+1} \leftarrow x_n - [\nabla^2 f(x_n)]^{-1} \nabla f(x_n) \quad (4.24)$$

Theorem 4.3 (Generalization). Suppose $x_* \in \Omega$ and $f \in C^3(\Omega, \mathbb{R})$ such that $\nabla f(x_*) = 0$ and $\nabla^2 f(x_*)$ is invertible. **TODO: check this** Then if initial point x_0 is sufficiently closed to x_* , then Newton's method converges to x_* .

Proof. The basic idea is the same as the \mathbb{R} case: prove the iterative algorithm induces a contraction mapping. ■

Example 4.3 (Newton's Method Fails to Converge). Even if f has an unique global minimum x_* , and x_0 is arbitrarily close to the x_* , Newton's method could fail to converge.

Consider

$$f(x) = \frac{2}{3} |x|^{\frac{3}{2}} \quad (4.25)$$

Note that

$$f(x) = \begin{cases} \frac{2}{3} x^{\frac{3}{2}} & x \geq 0 \\ -\frac{2}{3} x^{\frac{3}{2}} & x < 0 \end{cases} \quad (4.26)$$

$$f'(x) = \begin{cases} x^{\frac{1}{2}} & x \geq 0 \\ -x^{\frac{1}{2}} & x < 0 \end{cases} \quad (4.27)$$

$$f''(x) = \begin{cases} \frac{1}{2} x^{-\frac{1}{2}} & x > 0 \\ -\frac{1}{2} x^{-\frac{1}{2}} & x < 0 \\ \text{DNE} & x = 0 \end{cases} \quad (4.28)$$

Therefore $f \notin C^2$.

Let $\delta > 0$ arbitrarily small, take initial point $x_0 \in V_\delta(0)$. WLOG, $x_0 = \varepsilon \in V_\delta(0)$ with $\varepsilon > 0$. The algorithm will oscillate between $\pm\varepsilon$ and never converge.

Remark 4.1. Newton's method does not necessarily converge to a global minimum, it may converge to local minimum or local maximum or even saddle point.

Example 4.4 (Newton's Method Converges to a Saddle Point). Consider $f(x) = x^3$, $x_{n+1} \rightarrow \frac{x_n}{2}$, which converges to 0 (a saddle point).

Example 4.5 (Newton's Method on Quadratic Function). Let Q be a symmetric $n \times n$ invertible matrix. Define quadratic form $f(x) := \frac{1}{2} x^T Q x : \mathbb{R}^n \rightarrow \mathbb{R}$. The optimal is $x = 0$.

Let $x_0 \in \mathbb{R}^n$, then $x_1 := x_0 - H_f(x_0)^{-1} \nabla f(x_0) = x_0 - Q^{-1} Q x_0 = 0$. Therefore, Newton's method converges in one iteration.

4.2 Steepest/Gradient Descent

Algorithm 4.3 (Steepest Descent). Let $f : \Omega \rightarrow \mathbb{R}$ where Ω is an open subset of \mathbb{R}^n . Let initial point $x_0 \in \Omega$.

To minimize f on Ω , iteratively update x follows at each step k :

$$x_{k+1} \leftarrow x_k - \alpha_k \nabla f(x_k) \quad (4.29)$$

where $\alpha_k = \operatorname{argmin}_{\alpha \geq 0} f(x_k - \alpha \nabla f(x_k))$.

Remark: There might be multiple minimizing α , in real world implementations, we take the least minimizer found.

Theorem 4.4 (Gradient Descending is Descending). At every step k , if $\nabla f(x_k) = 0$, the algorithm terminates. Otherwise,

$$f(x_{k+1}) < f(x_k) \quad (4.30)$$

Proof. Suppose $\nabla f(x_k) \neq 0$.

Note that for the first minimizing α_k found:

$$f(x_{k+1}) = f(x_k - \alpha_k \nabla f(x_k)) \quad (4.31)$$

$$\leq f(x_k - \alpha \nabla f(x_k)) \quad \forall 0 \leq \alpha \leq \alpha_k \quad (4.32)$$

Recall that

$$\left. \frac{d}{ds} \right|_{s=0} f(x_k - s \nabla f(x_k)) = -\nabla f(x_k) \cdot \nabla f(x_k) = -\|\nabla f(x_k)\|_2^2 < 0 \quad (4.33)$$

Therefore,

$$f(x_{k+1}) \leq f(x_k - \alpha \nabla f(x_k)) < f(x_k) \text{ for small } \alpha \quad (4.34)$$

■

Theorem 4.5 (Gradient Descending Induces Perpendicular Steps). The consecutive steps induced by gradient descending are perpendicular. That is

$$(x_{k+2} - x_{k+1}) \cdot (x_{k+1} - x_k) = 0 \quad (4.35)$$

Proof. Note that

$$(x_{k+2} - x_{k+1}) \cdot (x_{k+1} - x_k) = (-\alpha_{k+1} \nabla f(x_{k+1})) \cdot (-\alpha_k \nabla f(x_k)) \quad (4.36)$$

$$= \alpha_k \alpha_{k+1} \nabla f(x_k) \cdot \nabla f(x_{k+1}) \quad (4.37)$$

If $\alpha_k = 0$, done.

If $\alpha_k > 0$,

$$f(x_{k+1}) = f(x_k) - \alpha_k \nabla f(x_k) \quad (4.38)$$

$$= \min_{\alpha \geq 0} \{f(x_k - \alpha \nabla f(x_k))\} \quad (4.39)$$

$$= \min_{\alpha > 0} \{f(x_k - \alpha \nabla f(x_k))\} \quad (4.40)$$

$$\implies \left. \frac{\partial}{\partial \alpha} \right|_{\alpha=\alpha_k} f(x_k - \alpha \nabla f(x_k)) = 0 \quad (4.41)$$

$$\implies -\nabla f(x_k - \alpha_k \nabla f(x_k)) \cdot \nabla f(x_k) = 0 \quad (4.42)$$

$$\implies -\nabla f(x_{k+1}) \cdot \nabla f(x_k) = 0 \quad (4.43)$$

■

Theorem 4.6 (Sufficient Condition for Gradient Descent to Converge). Let $f \in C^1$ on open $\Omega \subseteq \mathbb{R}^n$.

Let $\{x_k\}$ be the sequence generated by gradient descent: $x_{k+1} \leftarrow x_k - \alpha_k \nabla f(x_k)$.

If (x_k) is bounded in Ω , that is, there exists a compact set $K \subseteq \Omega$ such that $(x_k) \subseteq K$, then every convergent subsequence of (x_k) converges to a critical point $x_* \in \Omega$ of f .

Proof. **TODO:** Need to fix this proof. Let $x_k \in K$ compact.

Then there exists subsequence $x_{k_i} \rightarrow x_* \in K$.

Show: $\nabla f(x_*) = 0$.

Note that $f(x_k) \geq f(x_{k+1})$ for every $k \in \mathbb{N}$, therefore $f(x_{k_i}) \searrow f(x_*)$. Therefore, $f(x_k) \searrow f(x_*)$. **TODO:**

Show this. Suppose, for contradiction, $\nabla f(x_*) \neq 0$.

By continuity of ∇f , $(\nabla f(x_{k_i})) \rightarrow \nabla f(x_*)$.

Let $y_{k_i} := x_{k_i} - \alpha_{k_i} \nabla f(x_{k_i}) = x_{k_{i+1}}$.

Note that y_{k_i} has a convergent subsequence converging to y_* .

WLOG, $(y_{k_i}) \rightarrow y_*$.

Observe

$$\alpha_{k_i} = \frac{|y_{k_i} - x_{k_i}|}{\|\nabla f(x_{k_i})\|} \quad (4.44)$$

$$\implies \lim_{k_i \rightarrow \infty} \alpha_{k_i} = \frac{|y_* - x_*|}{\|\nabla f(x_*)\|} =: \alpha_* \quad (4.45)$$

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

Now $f(y_{k_i}) = f(x_{k_{i+1}}) = \min_{\alpha \geq 0} f(x_{k_i} - \alpha \nabla f(x_{k_i}))$, which implies

$$f(y_{k_i}) \leq f(x_{k_i} - \alpha \nabla f(x_{k_i})) \quad \forall \alpha \geq 0 \quad (4.46)$$

$$\forall \alpha \geq 0 \quad \lim_{i \rightarrow \infty} f(y_{k_i}) = f(y_*) \leq \lim_{i \rightarrow \infty} f(x_{k_i} - \alpha \nabla f(x_{k_i})) = f(x_* - \alpha \nabla f(x_*)) \quad (4.47)$$

$$\implies f(y_*) \leq \min_{\alpha \geq 0} f(x_* - \alpha \nabla f(x_*)) < f(x_*) \quad (4.48)$$

Further note that

$$f(y_*) = \lim_{i \rightarrow \infty} f(y_{k_i}) = \lim_{i \rightarrow \infty} f(x_{k_{i+1}}) = f(x_*) \quad (4.49)$$

Contradiction. ■

4.2.1 Steepest Descent: the Quadratic Case

Example 4.6. Let f follow the general quadratic form

$$f(x) = \frac{1}{2}x^T Qx - b^T x \quad (4.50)$$

with $b, x \in \mathbb{R}^n$ and Q is positive definite.

Let $0 < \lambda = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = \Lambda$ be eigenvalues of Q .

Proposition 4.1. Gradient descent on strictly convex(concave) quadratic functions is guaranteed to converge to the unique global minimum(maximum).

Essentially, we are going to define an auxiliary function g , preserving the optimizing behaviour in a sense that

$$\operatorname{argmin} g(x) = \operatorname{argmin} f(x) \quad (4.51)$$

and then show the convergence property on $f(x)$ indirectly by showing converging property on $g(x)$.

Lemma 4.1. Recall that positive definite Q implies the existence of unique minimizer x_* . The minimizer satisfies the first order necessary condition.

$$Qx_* - b = 0 \quad (4.52)$$

$$\iff x_* = Q^{-1}b \quad (4.53)$$

Where the second equation came from the invertibility of positive definite matrices. One can rewrite auxiliary function capturing the optimization behaviours f as following

$$g(x) := \frac{1}{2}(x - x_*)^T Q(x - x_*) \quad (4.54)$$

$$= \underbrace{\frac{1}{2}x^T Qx - \overbrace{x^T Qx_*}^{x^T b}}_{f(x)} + \frac{1}{2}x_*^T Qx_* \quad (4.55)$$

Because Q is positive definite:

$$g(x) \geq 0 \quad \forall x \in \mathbb{R}^n \quad (4.56)$$

$$g(x) = 0 \iff x = x_* \quad (4.57)$$

$$\nabla f(x) = \nabla g(x) = Qx - b \quad (4.58)$$

Where the last equation came from the fact that $f(x)$ and $g(x)$ differ by a constant. Therefore, using the method of steepest descent,

$$x_{k+1} = x_k - \alpha_k \nabla g(x_k) \quad (4.59)$$

where $\alpha_k \in \operatorname{argmin}_{\alpha \geq 0} f(x_k - \alpha \nabla g(x_k))$.

The necessary condition for minimizations suggests α_k must satisfy

$$0 = \frac{d}{d\alpha} \bigg|_{\alpha=\alpha_k} f(x_k - \alpha \nabla g(x_k)) = \nabla f(x_k - \alpha_k \nabla g(x_k)) \cdot (-\nabla g(x_k)) \quad (4.60)$$

$$= -[Q(x_k - \alpha_k \nabla g(x_k)) - b] \cdot \nabla g(x_k) \quad (4.61)$$

$$= -[Qx_k - \alpha_k Q \nabla g(x_k) - b] \cdot \nabla g(x_k) \quad (4.62)$$

$$= -[\nabla g(x_k) - \alpha_k Q \nabla g(x_k)] \cdot \nabla g(x_k) \quad (4.63)$$

$$= -\|\nabla g(x_k)\|_2^2 + \alpha_k \nabla g(x_k)^T Q \nabla g(x_k) \quad (4.64)$$

$$\implies \alpha_k = \frac{\|\nabla g(x_k)\|_2^2}{\nabla g(x_k)^T Q \nabla g(x_k)} \quad (4.65)$$

Assumption 4.1. TODO: *Need to check if this assumption is required.*

Assume Q is symmetric.

Lemma 4.2. The iterative updating from gradient descent on $g(x)$ is a contraction mapping. That is,

$$g(x_{k+1}) = \underbrace{\left(1 - \frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-1} \nabla g(x_k)]}\right)}_{\in [-1, 1]} g(x_k) \quad (4.66)$$

Proof.

$$g(x_{k+1}) \equiv g(x_k - \alpha_k \nabla g(x_k)) \quad (4.67)$$

$$\equiv \frac{1}{2} [x_k - \alpha_k \nabla g(x_k) - x_*]^T Q [x_k - \alpha_k \nabla g(x_k) - x_*] \quad (4.68)$$

$$= \frac{1}{2} [x_k - x_* - \alpha_k \nabla g(x_k)]^T Q [x_k - x_* - \alpha_k \nabla g(x_k)] \quad (4.69)$$

$$= \underbrace{\frac{1}{2} (x_k - x_*)^T Q (x_k - x_*)}_{g(x_k)} - \alpha_k \nabla g(x_k)^T Q (x_k - x_*) + \frac{1}{2} \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k) \quad (4.70)$$

$$\implies g(x_k) - g(x_{k+1}) = -\frac{1}{2} \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k) + \alpha_k \nabla g(x_k)^T Q \underbrace{(x_k - x_*)}_{=: y_k} \quad (4.71)$$

$$\implies \frac{g(x_k) - g(x_{k+1})}{g(x_k)} = \frac{-\frac{1}{2} \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k) + \alpha_k \nabla g(x_k)^T Q y_k}{\frac{1}{2} y_k^T Q y_k} \quad (4.72)$$

$$= \frac{2\alpha_k \nabla g(x_k)^T Q y_k - \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k)}{y_k^T Q y_k} \quad (4.73)$$

Note that the first order condition implies $Qx_* = b$.

Therefore, $\nabla g(x_k) = Qx_k - b = Qx_k - Qx_* = Qy_k$, which implies $y_k = Q^{-1} \nabla g(x_k)$.

$$\frac{2\alpha_k \nabla g(x_k)^T Q y_k - \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k)}{y_k^T Q y_k} = \frac{2\alpha_k \nabla g(x_k)^T Q Q^{-1} \nabla g(x_k) - \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k)}{\nabla g(x_k)^T Q^{-T} Q Q^{-1} \nabla g(x_k)} \quad (4.74)$$

$$= \frac{2\alpha_k \|\nabla g(x_k)\|_2^2 - \alpha_k^2 \nabla g(x_k)^T Q \nabla g(x_k)}{\nabla g(x_k)^T Q^{-T} \nabla g(x_k)} \quad (4.75)$$

Plug in the α_k computed before:

$$\dots = \frac{2 \frac{\|\nabla g(x_k)\|_2^2}{\nabla g(x_k)^T Q \nabla g(x_k)} \|\nabla g(x_k)\|_2^2 - \frac{\|\nabla g(x_k)\|_2^4}{(\nabla g(x_k)^T Q \nabla g(x_k))^2} \nabla g(x_k)^T Q \nabla g(x_k)}{\nabla g(x_k)^T Q^{-T} \nabla g(x_k)} \quad (4.76)$$

$$= \frac{2 \frac{\|\nabla g(x_k)\|_2^4}{\nabla g(x_k)^T Q \nabla g(x_k)} - \frac{\|\nabla g(x_k)\|_2^4}{\nabla g(x_k)^T Q \nabla g(x_k)}}{\nabla g(x_k)^T Q^{-T} \nabla g(x_k)} \quad (4.77)$$

$$= \frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-T} \nabla g(x_k)]} \quad (4.78)$$

$$= \frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-1} \nabla g(x_k)]} \quad \because Q \in \mathbb{S}^n \quad (4.79)$$

$$\implies g(x_k) - g(x_{k+1}) = \left\{ \frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-1} \nabla g(x_k)]} \right\} g(x_k) \quad (4.80)$$

$$\implies g(x_{k+1}) = \left\{ 1 - \left[\frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-1} \nabla g(x_k)]} \right] \right\} g(x_k) \quad (4.81)$$

Lemma 4.3 (Kantorovich Inequality). Let Q be a $n \times n$ positive definite symmetric matrix with eigenvalues $0 < \lambda = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = \Lambda$. Then, for any $v \in \mathbb{R}^n$:

$$\frac{\|v\|_2^4}{(v^T Q v)(v^T Q^{-1} v)} \geq \frac{4\lambda\Lambda}{(\lambda + \Lambda)^2} \quad (4.82)$$

Proof of Kantorovich Inequality. **TODO:** Prove this lemma ■

Therefore,

$$g(x_{k+1}) = \left\{ 1 - \left[\frac{\|\nabla g(x_k)\|_2^4}{[\nabla g(x_k)^T Q \nabla g(x_k)][\nabla g(x_k)^T Q^{-1} \nabla g(x_k)]} \right] \right\} g(x_k) \quad (4.83)$$

$$\leq \left\{ 1 - \frac{4\lambda\Lambda}{(\lambda + \Lambda)^2} \right\} g(x_k) \quad (4.84)$$

$$= \frac{(\lambda - \Lambda)^2}{(\lambda + \Lambda)^2} g(x_k) \quad (4.85)$$

■

Theorem 4.7. For any initial point $x_0 \in \mathbb{R}^n$, gradient descent converges to the unique minimum point x_* of the quadratic $f(x) = x^T Q x - b^T x$.

Proof. Define $q(x) := \frac{1}{2}(x - x_*)^T Q(x)(x - x_*)$.

Note that $q(x)$ and $f(x)$ differ by a constant, therefore $\operatorname{argmin} q(x) = \operatorname{argmin} f(x)$.

Moreover, we've shown:

$$q(x_{k+1}) \leq \underbrace{\left(\frac{\Lambda - \lambda}{\Lambda + \lambda} \right)^2}_{\in [0,1)} q(x_k) \quad (4.86)$$

It is easy to notice that

$$q(x_k) \leq r q(x_{k-1}) \quad (4.87)$$

$$\implies q(x_k) \leq r^k q(x_0) \quad (4.88)$$

$$\implies q(x_k) \in \{x \in \mathbb{R}^k : q(x) \leq r^k q(x_0)\} =: \mathcal{L}_k \quad (4.89)$$

Note that the sub-level set \mathcal{L}_k is strictly decreasing (i.e. $\mathcal{L}_{k+1} \subsetneq \mathcal{L}_k$).

Further, note that x_* is the only point satisfying the inequality at the limit:

$$q(x_*) = 0 = \lim_{k \rightarrow \infty} q(x_0) \quad (4.90)$$

Therefore, $\lim_{k \rightarrow \infty} \mathcal{L} = \{0\}$, and $(x_k) \rightarrow x_*$. ■

Remark 4.2. Note that

$$r = \left(\frac{\frac{\Lambda}{\lambda} - 1}{\frac{\Lambda}{\lambda} + 1} \right)^2 \in [0, 1) \quad (4.91)$$

$$= \left(\frac{C - 1}{C + 1} \right)^2 \quad (4.92)$$

where $C = \frac{\Lambda}{\lambda}$ is the **condition number** of Q .

Clearly, when $\lambda = \Lambda$, $r = 0$ and gradient descent converges to the unique global minimum after only one epoch.

While $C \gg 1$, $r \approx 1$ and the worst case of gradient descent converges slowly.

4.3 Method of Conjugate Directions

Motivation Method of conjugate directions is designed for quadratic functions with form $\frac{1}{2}x^T Qx - b^T x$. For other functional forms, one can approximate the function using quadratic form firstly and then apply method of conjugate directions.

Definition 4.1. Let $Q \in \mathbb{S}^n$, $d, d' \in \mathbb{R}^n$ are **Q -orthogonal** or **Q -conjugate** if

$$d^T Q d' = 0 \quad (4.93)$$

Remark: note that $d^T Q d' = d'^T Q d$ because Q is symmetric.

Definition 4.2. Finite set $D = (d_0, \dots, d_k) \subseteq \mathbb{R}^n$ is a **Q -orthogonal** set if

$$\forall i \neq j, d_i^T Q d_j = 0 \quad (4.94)$$

That is, D is *orthogonal in pairs*.

Example 4.7. When $Q = I_n$, the notional of Q -orthogonal becomes the conventional notion of orthogonality.

Proposition 4.2. Let d, d' be two eigenvectors of Q with different eigenvalues λ, λ' , then d, d' are Q -orthogonal.

Proof. Let $Qv = \lambda v$ and $Qw = \lambda' w$ where $\lambda \neq \lambda'$.

Note that

$$\langle v, Qw \rangle = \langle v, \lambda' w \rangle \quad (4.95)$$

$$= \lambda' \langle v, w \rangle \quad (4.96)$$

because inner product is bilinear.

Similarly,

$$\langle v, Qw \rangle = \langle Q^T v, w \rangle \quad (4.97)$$

$$= \langle Qv, w \rangle \quad (4.98)$$

$$= \langle \lambda v, w \rangle \quad (4.99)$$

$$= \lambda \langle v, w \rangle \quad (4.100)$$

$$\implies (\lambda' - \lambda) \langle v, w \rangle = 0 \quad (4.101)$$

where $\lambda' - \lambda \neq 0$.

Therefore, $\langle v, w \rangle = 0$.

Further,

$$v^T Qw = \langle v, Qw \rangle \quad (4.102)$$

$$= \langle v, \lambda w \rangle \quad (4.103)$$

$$= \lambda \langle v, w \rangle = 0 \quad (4.104)$$

So v, w are Q -orthogonal. ■

Proposition 4.3. Let $Q \in \mathbb{S}^n$, then there exists a set of orthogonal eigen-basis of Q .

Corollary 4.1. The orthogonal eigen-basis of Q is also Q -orthogonal.

Proof.

$$\forall i \neq j, \quad d_i^T Qd_j = d_i^T \lambda_j d_j = \lambda_j \langle d_i, d_j \rangle = 0 \quad (4.105)$$

■

Proposition 4.4. Let $Q \in \mathbb{S}_+^n$, let $d_0, \dots, d_k \neq 0$ be a set of Q -orthogonal vectors with $k \leq n - 1$, then d_0, \dots, d_k are linearly independent.

Proof. Suppose

$$\alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_k d_k = 0 \quad (4.106)$$

For every $i \in \{0, 1, \dots, k\}$, multiply $d_i^T Q$ on both sides of the equation:

$$\underbrace{\alpha_0 d_i^T Q d_0 + \alpha_1 d_i^T Q d_1 + \dots + \alpha_i d_i^T Q d_i}_{=0} + \underbrace{\alpha_{i+1} d_i^T Q d_{i+1} + \dots + \alpha_k d_i^T Q d_k}_{=0} = 0 \quad (4.107)$$

Further, $d_i^T Q d_i > 0$ because Q is positive definite. Hence, $\alpha_i = 0$ for every i , and d_0, \dots, d_k are linearly independent. ■

Lemma 4.4 (Theorems Covered so Far). Recall that

- (i) d_i, d_j are Q -orthogonal if $d_i^T Q d_j = 0$;
- (ii) Eigen-vectors with different eigenvalues are Q -orthogonal;
- (iii) Q symmetric \implies there exists an orthogonal basis \implies the set of basis is Q -orthogonal as well;
- (iv) Q -orthogonal vectors are linearly independent.

Example 4.8 (Special Case: Method of Conjugate Direction on Quadratic Functions). Let $Q \in \mathbb{S}_{++}^n$ and minimizing the quadratic function

$$\min f(x) = \frac{1}{2}x^T Qx - b^T \quad (4.108)$$

Recall that the unique global minimum is $x^* = Q^{-1}b$.

Let d_0, d_1, \dots, d_{n-1} be non-zero Q -orthogonal vectors.

Note that they are linearly independent by the previous theorem.

Therefore, they form a basis of \mathbb{R}^n .

The global minimum can be represented as

$$x^* = \sum_{j=1}^{n-1} \alpha_j d_j \quad \alpha_j \in \mathbb{R} \quad (4.109)$$

For every j , the following holds

$$d_j^T Qx^* = \alpha_j d_j^T Qd_j \quad (4.110)$$

$$\implies \alpha_j = \frac{d_j^T Qx^*}{d_j^T Qd_j} \quad (4.111)$$

Algorithm 4.4 (Method of Conjugate Directions). Let $Q \in \mathbb{S}_{++}^n$ and $\{d_j\}_{j=0}^{n-1}$ be a set of non-zero Q -orthogonal vectors, note that they form a basis of \mathbb{R}^n .

Given initial point $x_0 \in \mathbb{R}^n$, the method of conjugate direction generates a sequence of points $\{x_k\}_{k=0}^n$ as the following:

$$x_{k+1} \leftarrow x_k + \alpha_k d_k \quad (4.112)$$

$$\alpha_k := -\frac{\langle g_k, d_k \rangle}{d_k^T Qd_k} \quad g_k := \nabla f(x_k) \quad (4.113)$$

Theorem 4.8. Given the method of conjugate, the sequence of points generated eventually reaches the global minimum. That is, $x_n = x^*$.

Proof. Let $x^*, x_0 \in \mathbb{R}^n$, consider

$$x^* - x_0 = \sum_{j=0}^{n-1} \beta_j d_j \quad (4.114)$$

$$\iff x^* = x_0 + \sum_{j=0}^{n-1} \beta_j d_j \quad (4.115)$$

$$d_j^T Q(x^* - x_0) = \beta_j d_j^T Qd_j \quad (4.116)$$

$$\implies \beta_j = \frac{d_j^T Q(x^* - x_0)}{d_j^T Qd_j} \quad (4.117)$$

Note that the algorithm generates the sequence as following:

$$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j d_j \quad (4.118)$$

$$\implies (x_k - x_0) = \sum_{j=0}^{k-1} \alpha_j d_j \quad (4.119)$$

$$\implies d_k^T Q(x_k - x_0) = \sum_{j=0}^{k-1} \alpha_j d_k^T Q d_j = 0 \quad (4.120)$$

Therefore,

$$\beta_k = \frac{d_k^T Q(x^* - x_0)}{d_k^T Q d_k} \quad (4.121)$$

$$= \frac{d_k^T Q(x^* - x_0) - d_k^T Q(x_k - x_0)}{d_k^T Q d_k} \quad (4.122)$$

$$= \frac{d_k^T Q(x^* - x_k)}{d_k^T Q d_k} \quad (4.123)$$

$$= \frac{d_k^T (Qx^* - Qx_k)}{d_k^T Q d_k} \quad (4.124)$$

The first order necessary condition suggests $Qx_* = b$,

$$\beta_k = \frac{d_k^T (Qx^* - Qx_k)}{d_k^T Q d_k} \quad (4.125)$$

$$= \frac{d_k^T (b - Qx_k)}{d_k^T Q d_k} \quad (4.126)$$

$$= - \frac{d_k^T (Qx_k - b)}{d_k^T Q d_k} \quad (4.127)$$

$$= - \frac{d_k^T \nabla f(x_k)}{d_k^T Q d_k} = \alpha_k \quad (4.128)$$

Consequently,

$$x^* = x_0 + \sum_{j=0}^{n-1} \beta_j d_j \quad (4.129)$$

$$= x_0 + \sum_{j=0}^{n-1} \alpha_j d_j \quad (4.130)$$

$$= x_n \quad (4.131)$$

■

4.4 Geometric Interpretations of Method of Conjugate Directions

Theorem 4.9. Let $f \in C^1(\Omega, \mathbb{R})$, where Ω is a convex subset of \mathbb{R}^n , then x_0 is a local minimum of f on Ω if and only if

$$\nabla f(x_0) \cdot (y - x_0) \geq 0 \quad \forall y \in \Omega \quad (4.132)$$

Corollary 4.2. Now consider the special case in which Ω is an affine hyperplane, that is,

$$\Omega = \{x \in \mathbb{R}^n : cx + b = 0\} \quad (4.133)$$

where $\dim(\Omega)$ is $n - 1$.

Note that for every $y \in \Omega$, $\nabla f(x_0) \cdot (y - x_0) \geq 0$. For any feasible direction a at point x_0 , by the definition of hyperplane, $-a$ is a feasible direction as well.

Consequently, $a \cdot \nabla f(x_0)$ for every feasible direction. That is, $\nabla f(x_0) \perp \Omega$.

Notation 4.1. Let d_0, d_1, \dots, d_{n-1} be a set of non-zero Q -orthogonal vectors in \mathbb{R}^n . For every $k \in \{0, 1, \dots, n\}$, let

$$\mathcal{B}_k = \text{span}\{d_0, \dots, d_{k-1}\} \quad (4.134)$$

In particular, $\mathcal{B}_0 = \{0\}$ and $\mathcal{B}_n = \mathbb{R}^n$.

Theorem 4.10. The sequence $\{x_k\}$ generated from $x_0 \in \mathbb{R}^n$ by conjugate direction method has the property that x_k minimizes $f(x) = \frac{1}{2}x^T Qx - b^T x$ on the affine hyperplane $x_0 + \mathcal{B}_k$.

That is,

$$x_k \in \underset{x \in x_0 + \mathcal{B}_k}{\text{argmin}} f(x) \quad (4.135)$$

Proof. Recall that x_k is the the minimizer of f on the affine hyperplane if and only if $\nabla f(x_k) \perp x_0 + \mathcal{B}_k$.

It is enough to show that $\nabla f(x_k) \perp \mathcal{B}_k$. *Base Case:* $g_0 := \nabla f(x_0) \perp \{0\}$ trivially.

Inductive Step: Assume $g_k \perp \mathcal{B}_k$, show $g_{k+1} \perp \mathcal{B}_{k+1}$:

$$g_k \perp \mathcal{B}_k \quad (4.136)$$

$$\implies g_k \perp \text{span}\{d_0, \dots, d_{k-1}\} \quad (4.137)$$

$$\langle g_{k+1}, d_k \rangle = \langle g_k + \alpha_k Qd_k, d_k \rangle \quad (4.138)$$

$$= \langle g_k, d_k \rangle + \alpha_k \langle Qd_k, d_k \rangle \quad (4.139)$$

$$= \langle g_k, d_k \rangle - \frac{\langle g_k, d_k \rangle}{d_k^T Q d_k} d_k^T Q d_k = 0 \quad (4.140)$$

$$\implies g_{k+1} \perp d_k \quad (4.141)$$

$$\langle g_{k+1}, d_i \rangle = \langle g_k + \alpha_k Qd_k, d_i \rangle \quad 0 \leq i < k \quad (4.142)$$

$$= \langle g_k, d_i \rangle + \langle \alpha_k Qd_k, d_i \rangle \quad (4.143)$$

$$= 0 + 0 = 0 \quad (4.144)$$

$$\implies g_{k+1} \perp \mathcal{B}_k \quad (4.145)$$

■

Corollary 4.3. x_n (the final output of the method of conjugate gradient) minimizes $f(x)$ on $x_0 + \mathcal{B}_n = \mathbb{R}^n$.

Corollary 4.4. Let $q(\cdot)$ be a quadratic function, then

$$0 \leq q(x_k) = \min_{x \in x_0 + \mathcal{B}_k} q(x) \leq q(x_k) = \min_{x \in x_0 + \mathcal{B}_{k-1}} q(x) \quad (4.146)$$

Proof. The result is immediate by noting that $x_0 + \mathcal{B}_{k-1} \subseteq x_0 + \mathcal{B}_k$. ■

5 Infinite Dimensional Optimization

5.1 Calculus of Variation

5.1.1 The Canonical Example

Problem Setup Let

$$\mathcal{A} := \{u : [0, 1] \rightarrow \mathbb{R} : u \in C^1 \wedge u(0) = u(1) = 1\} \quad (5.1)$$

And let $F[\cdot] : \mathcal{A} \rightarrow \mathbb{R}$ denote the **functional** defined as

$$F[u(\cdot)] := \frac{1}{2} \int_0^1 u^2(x) + u'^2(x) \, dx \quad (5.2)$$

Consider the minimization problem over the *space of functions*:

$$\min F[u] \text{ s.t. } u \in \mathcal{A} \quad (5.3)$$

Optimizing via Disturbance (Test Functions) The general idea of solving such infinite dimensional problem would be reducing it to a collection of one dimensional problems.

Let $v(\cdot) \in C^1([0, 1], \mathbb{R})$ such that $v(0) = v(1) = 0$.

Suppose $u^*(\cdot)$ is the minimizer of the problem stated above.

Consider

$$u^*(\cdot) + sv(\cdot) \quad s \in \mathbb{R} \quad (5.4)$$

Notice that by construction, $u^*(\cdot) + sv(\cdot) \in \mathcal{A}$ for every $s \in \mathbb{R}$.

Define the auxiliary function

$$f(s) := F[u^* + sv] \in \mathbb{R} \quad (5.5)$$

The attained when $s = 0$, the first order necessary condition of minimization suggests:

$$f'(0) = \left. \frac{d}{ds} \right|_{s=0} F[u_*(\cdot) + sv(\cdot)] \quad (5.6)$$

$$= \left. \frac{d}{ds} \right|_{s=0} \frac{1}{2} \int_0^1 [u_*(x) + sv(x)]^2 + [u'_*(x) + sv'(x)]^2 dx \quad (5.7)$$

$$= \left. \frac{d}{ds} \right|_{s=0} \left\{ \frac{1}{2} \int_0^1 [u_*(x)^2 + u'_*(x)^2] dx + \int_0^1 s[u_*(x)v(x) + u'_*(x)v'(x)] dx + \frac{s^2}{2} \int_0^1 v^2(x) + v'^2(x) dx \right\} \quad (5.8)$$

$$= \left. \frac{d}{ds} \right|_{s=0} \int_0^1 s[u_*(x)v(x) + u'_*(x)v'(x)] dx \quad (5.9)$$

$$= \int_0^1 \left. \frac{d}{ds} \right|_{s=0} s[u_*(x)v(x) + u'_*(x)v'(x)] dx \quad (5.10)$$

$$= \int_0^1 u_*(x)v(x) + u'_*(x)v'(x) dx = 0 \quad (\dagger') \quad (5.11)$$

Where (\dagger') is the primitive form of the first order necessary condition.

Furthermore,

$$\int_0^1 u_*(x)v(x) + u'_*(x)v'(x) dx = \int_0^1 u_*(x)v(x) dx + \int_0^1 u'_*(x)v'(x) dx \quad (5.12)$$

$$= \int_0^1 u_*(x)v(x) dx + u_*(x)v(x)|_0^1 + \int_0^1 u''_*(x)v(x) dx \quad (5.13)$$

$$= \int_0^1 u_*(x)v(x) dx + \int_0^1 u''_*(x)v(x) dx \quad (5.14)$$

$$= \int_0^1 [u_*(x) + u''_*(x)]v(x) dx \quad (5.15)$$

The first order necessary condition is amount to

$$\int_0^1 [u_*(x) + u''_*(x)]v(x) dx = 0 \quad (5.16)$$

for every function $v \in C^1([0, 1], \mathbb{R})$ such that $v(0) = v(1) = 0$.

Lemma 5.1 (Du Bois-Reymond Lemma). Provided that $u_*(\cdot) \in \mathcal{A}$ is C^1 , altogether with the fact that it is the minimizer, $u_*(\cdot)$ is automatically C^2 .

Notation 5.1. We call all $v \in C^1([0, 1], \mathbb{R})$ such that $v(0) = v(1) = 0$ **test functions**.

Lemma 5.2 (The Fundamental Lemma of Calculus of Variation). Suppose g is continuous on interval $[a, b]$ such that

$$\int_a^b g(x)v(x) dx = 0 \quad \forall \text{ test function } v(\cdot) \quad (5.17)$$

Then $g(x) \equiv 0$ on $[a, b]$.

Proof. Suppose, for contradiction, $g(\eta) \neq 0$ for some $\eta \in [0, 1]$.

WLOG, assume $g(\eta) > 0$.

Note that if η is at either endpoint, by continuity of g , one may choose another $\eta' \in V_\varepsilon(\eta)$ such that $\eta' \in (a, b)$.

Therefore, we may assume $\eta \in (a, b)$ without loss of generality.

Then there exists an open interval $(c, d) \subsetneq (a, b)$ such that $g(x) > 0 \forall x \in (c, d)$.

Define

$$v(x) := \begin{cases} (x-c)^2(x-d)^2 & \text{if } x \in (c, d) \\ 0 & \text{otherwise} \end{cases} \quad (5.18)$$

Clearly, $v(\cdot) \in C^1([0, 1], \mathbb{R})$ by noticing

$$v'(x) = \begin{cases} 2(x-c)(x-d)^2 + 2(x-c)^2(x-d) & \text{if } x \in (c, d) \\ 0 & \text{otherwise} \end{cases} \quad (5.19)$$

Obviously,

$$\int_a^b g(x)v(x) dx = \int_c^d g(x)v(x) dx > 0 \quad (5.20)$$

$\Rightarrow \Leftarrow$

Proposition 5.1.

$$\int_0^1 (u_*(x) + u''_*(x))v(x) dx = 0 \forall v \implies u_*(\cdot) + u''_*(\cdot) \equiv 0 \quad (5.21)$$

Proof. By Du Bois-Reymond Lemma, $u_*(\cdot) + u''_*(\cdot)$ is continuous. Because $v(\cdot)$ is a test function and by the fundamental lemma of calculus of variation, $u_*(\cdot) + u''_*(\cdot) \equiv 0$. ■

5.2 A Classical Example from Physics: Brachistochrone

Brachistochrone (Galileo, 1638) Find the curve connecting two points A and B on which a point mass moves without friction under the influence of gravity moves from A to B in the least time possible.

5.3 General Class of Problems in Calculus of Variations

Constraint

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R} : u \in C^1, u(a) = A, u(b) = B\} \quad (5.22)$$

Objective function

$$F[u(\cdot)] := \int_a^b L(x, u(x), u'(x)) dx \quad (5.23)$$

where $L(x, z, p) : [a, b] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$.

Example 5.1.

$$L(x, z, p) = \frac{1}{2}(z^2 + p^2) \quad (5.24)$$

$$F[u(\cdot)] = \int_0^1 \frac{u(x)^2 + u'(x)^2}{2} dx \quad (5.25)$$

Definition 5.1. Given $u(\cdot) \in \mathcal{A}$, suppose \exists function $g(\cdot)$ on $[a, b]$ such that

$$\left. \frac{d}{ds} \right|_{s=0} F[u(\cdot) + sv(\cdot)] = \int_a^b g(x)v(x) dx \quad \forall \text{ test functions } v(\cdot) \quad (5.26)$$

Then $g(\cdot)$ is called the **variational derivative** of F at $u(\cdot)$, often denoted as $\frac{\delta F}{\delta u}(u)(\cdot)$.

Remark 5.1 (Analogy to the finite dimensional case). In finite dimensional case:

$$\left. \frac{d}{ds} \right|_{s=0} f(u + sv) = \nabla f(u) \cdot v = \sum_{i=1}^n \nabla_i f(u) v_i \quad \forall v \in \mathbb{R}^n \quad (5.27)$$

And in the infinite dimensional case:

$$\left. \frac{d}{ds} \right|_{s=0} F[u(\cdot) + sv(\cdot)] = \int_a^b \frac{\delta F}{\delta u}(u)(x)v(x) dx \quad \forall v \in \mathcal{T} \quad (5.28)$$

Lemma 5.3 (Necessary condition on the variational derivative). Let

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R} : u \in C^1, u(a) = A, u(b) = B\} \quad (5.29)$$

If $u_*(\cdot) \in \mathcal{A}$ minimizes F over \mathcal{A} , and if $\frac{\delta F}{\delta u}(u_*)(\cdot)$ exists and is continuous, then it must satisfy

$$\frac{\delta F}{\delta u}(u_*)(\cdot) \equiv 0 \quad (5.30)$$

Proof. Note that for every test function v , $u_*(\cdot) + sv(\cdot) \in \mathcal{A}$.

Suppose u_* minimizes F over \mathcal{A} , equivalently,

$$f(0) = F[u_*] \leq F[u_* + sv] = f(s) \text{ for every test function } v \quad (5.31)$$

$$\iff f(0) \leq f(s) \quad \forall s \in \mathbb{R} \quad (5.32)$$

$$\implies 0 = f'(0) = \left. \frac{d}{ds} \right|_{s=0} f(s) = \left. \frac{d}{ds} \right|_{s=0} F[u_* + sv] \quad (5.33)$$

$$= \int_a^b \frac{\delta F}{\delta u}(u)(x)v(x) dx \quad (5.34)$$

$$\implies \frac{\delta F}{\delta u}(u)(\cdot) \equiv 0 \quad (5.35)$$

■

Theorem 5.1 (Euler-Lagrange). Let

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R} : u \in C^1, u(a) = A, u(b) = B\} \quad (5.36)$$

Let $L \in C^2$ such that

$$F[u(\cdot)] = \int_a^b L(x, u(x), u'(x)) dx \quad (5.37)$$

Then, if $u(\cdot) \in C^1$, then $\frac{\delta F}{\delta u}(u)(\cdot)$ exists and is continuous.

Moreover,

$$\frac{\delta F}{\delta u}(u)(\cdot)(x) = -\frac{d}{dx}[L_p(x, u(x), u'(x))] + L_z(x, u(x), u'(x)) \quad (5.38)$$

This equation is often referred to as the **Euler-Lagrange equation**.

Proof. Let v be a test function.

$$\left. \frac{d}{ds} \right|_{s=0} F[u(\cdot) + sv(\cdot)] = \left. \frac{d}{ds} \right|_{s=0} \int_a^b L(x, u(x) + sv(x), u'(x) + sv'(x)) dx \quad (5.39)$$

$$= \int_a^b \left. \frac{d}{ds} \right|_{s=0} L(x, z, p) dx \quad (5.40)$$

$$= \int_a^b L_z(\cdot)v(x) + L_p(\cdot)v'(x) dx \quad (5.41)$$

$$= \int_a^b L_z(\cdot)v(x) dx + \int_a^b L_p(\cdot)v'(x) dx \quad (5.42)$$

$$= \int_a^b L_z(\cdot)v(x) dx + L_p(\cdot)v(x) \Big|_a^b - \int_a^b \frac{d}{dx} L_p(\cdot)v(x) dx \quad (5.43)$$

$$= \int_a^b \left[-\frac{d}{dx} L_p(\cdot) + L_p(\cdot) \right] v(x) dx \quad \forall \text{ test functions } v(\cdot) \quad (5.44)$$

The result follows definition of variational derivative.

Further, because $L(\cdot) \in C^2$, $-\frac{d}{dx} L_p(\cdot)$ is continuous. Moreover, $u(\cdot)$ and $u'(\cdot)$ are continuous, so is the composite function. Hence, the variational derivative is continuous. ■

Example 5.2. In the model example,

$$L(x, z, p) = \frac{1}{2}(z^2 + p^2) \quad (5.45)$$

$$L_p(x, z, p) = p \implies L_p(x, u(x), u'(x)) = u'(x) \quad (5.46)$$

$$L_z(x, z, p) = z \implies L_z(x, u(x), u'(x)) = u(x) \quad (5.47)$$

$$\frac{\delta F}{\delta u}(u)(\cdot) = -\frac{d}{dx}[u'(x)] + u(x) = -u''(x) + u(x) \quad (5.48)$$

By above theorem, if u_* is the minimizer, it must be

$$-u''(x) + u(x) \equiv 0 \quad (5.49)$$

Example 5.3 (Min Arclength). Suppose we are trying to minimize

$$F[u] = \int_a^b \sqrt{1 + u'(x)^2} dx = \text{arclength of } u(\cdot) \quad (5.50)$$

where

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R} : u \in C^1, u(a) = A, u(b) = B\} \quad (5.51)$$

and

$$L(x, z, p) = \sqrt{1 + p^2} \quad (5.52)$$

$$L_z = 0 \quad (5.53)$$

$$L_p = \frac{p}{\sqrt{1 + p^2}} \quad (5.54)$$

$$\implies \frac{\delta F}{\delta u}(u)(\cdot) = -\frac{d}{dx} \frac{u'(x)}{\sqrt{1 + u'(x)^2}} \equiv 0 \quad (5.55)$$

$$\implies \frac{u'(x)}{\sqrt{1 + u'(x)^2}} = C \quad (5.56)$$

$$\implies u'(x)^2 = C(1 + u'(x)^2) \quad (5.57)$$

$$\implies u'(x)^2 = C \quad (5.58)$$

$$\implies u'(x) = \alpha \quad (5.59)$$

$$\implies u(x) = \alpha x + \beta \quad (5.60)$$

Example 5.4 (Surface Area of Revolution). Suppose $u(\cdot) \in C^1$ on $[a, b]$, the surface area of rotating the curve u connecting a and b can be computed as

$$F[u(\cdot)] = \int_a^b 2\pi u(x) \sqrt{1 + u'(x)^2} dx \quad (5.61)$$

For simplicity, assume $u > 0$.

In this example, the space of feasible functions is

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R} : u \in C^1, u(a) = A, u(b) = B, u > 0\} \quad (5.62)$$

If $u(\cdot)$ solves the minimization problem, it must be the case that

$$\frac{\delta F}{\delta u}(u)(\cdot) \equiv 0 \quad (\dagger) \quad (5.63)$$

Notice

$$L(x, z, p) = 2\pi z \sqrt{1 + p^2} \quad (5.64)$$

$$L_z(x, z, p) = 2\pi \sqrt{1 + p^2} \quad (5.65)$$

$$L_p(x, z, p) = 2\pi z \frac{p}{\sqrt{1 + p^2}} \quad (5.66)$$

Claim: the family of $u(\cdot) = \beta \cosh\left(\frac{x-\alpha}{\beta}\right)$ solves the necessary condition (\dagger) .

Instance 1 When $a = 0, b = 1, A = B = 1$, plugging in the initial condition gives

$$\begin{cases} \beta \cosh\left(\frac{0-\alpha}{\beta}\right) &= 1 \\ \beta \cosh\left(\frac{1-\alpha}{\beta}\right) &= 1 \end{cases} \quad (5.67)$$

solving above system of equations provides the solution.

Instance 2 When $a = 0, b = 1, A = 1, B = 0$, plugging in these initial conditions gives

$$\begin{cases} \beta \cosh\left(\frac{0-\alpha}{\beta}\right) &= 1 \\ \beta \cosh\left(\frac{1-\alpha}{\beta}\right) &= 0 \end{cases} \quad (5.68)$$

because $\cosh > 0$, the second equation suggests $\beta = 0$, but in this case the first equation would never hold. Therefore, there is no solution to this calculus of variation.

In face, the surface area is minimized by

$$u(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.69)$$

5.4 Euler-Lagrange Equations in \mathbb{R}^n

Setup

$$F[u(\cdot)] = \int_a^b L(x, u(x), u'(x)) \, dx \quad (5.70)$$

$$u : [a, b] \rightarrow \mathbb{R}^n \quad (5.71)$$

$$L(x, z, p) : [a, b] \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \quad (5.72)$$

$$\mathcal{A} := \{u : [a, b] \rightarrow \mathbb{R}^n : u \in C^1, u(a) = \mathbf{A}, u(b) = \mathbf{B}\} \quad (5.73)$$

Theorem 5.2 (Euler-Lagrange Equations in Vector Forms).

$$-\frac{d}{dx} \nabla_p L(x, z, p) + \nabla_z L(x, z, p) = \mathbf{0} \in \mathbb{R}^n \quad (\dagger) \quad (5.74)$$

Example 5.5 (Classical Lagrangian Mechanics).

$$V(x) : \mathbb{R}^n \rightarrow \mathbb{R} \text{ potential energy} \quad (5.75)$$

$$\frac{1}{2} m ||v||_2^2 \text{ kinetic energy} \quad (5.76)$$

$$L(t, x, v) := \frac{1}{2} m ||v||_2^2 - V(x) \text{ difference between KE and PE} \quad (5.77)$$

Consider a path $x(t)$ in \mathbb{R}^n , define objective function as

$$F[x(\cdot)] = \int_a^b L(t, x(t), x'(t)) \, dt \quad (5.78)$$

$$= \int_a^b \frac{1}{2} m ||\dot{x}(t)||_2^2 - V(x(t)) \, dt \quad (5.79)$$

The Euler-Lagrange equation in vector form implies

$$-\frac{d}{dt} \nabla_{(3)} L(t, x(t), \dot{x}(t)) + \nabla_{(2)} L(t, x(t), \dot{x}(t)) = 0 \quad (5.80)$$

$$\implies -\frac{d}{dt} m \dot{x}(t) - \nabla V(x(t)) = 0 \quad (5.81)$$

$$\implies m \ddot{x}(t) = \nabla V(x(t)) \quad (\dagger\dagger) \quad (5.82)$$

Remark 5.2. $(\dagger\dagger)$ is often referred to as *Newton's second law*: object moves along the path on which the total conversion between kinetic and potential energies is minimized.

Example 5.6 (3-Dimensional Pendulum). Suppose the pendulum is moving on a path such that the total conversion between kinetic and potential energies is minimized, that is

$$\min \int_a^b L(t) dt = \int_a^b \frac{1}{2} m (\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2) - mgz(t) dt \quad (5.83)$$

with the restriction that $\|\mathbf{x}(t)\| = \ell$, where ℓ is the radius of the sphere.

The restriction can be embodied by framing the problem using spherical coordinates:

$$x := \ell \cos \varphi \sin \theta \quad (5.84)$$

$$y := \ell \sin \varphi \sin \theta \quad (5.85)$$

$$z := -\ell \cos \theta \quad (5.86)$$

where the path of motion can be characterized using $(\theta(t), \varphi(t))$.

The objective function is therefore

$$L \left(t, \begin{pmatrix} \theta(t) \\ \varphi(t) \end{pmatrix}, \begin{pmatrix} \dot{\theta}(t) \\ \dot{\varphi}(t) \end{pmatrix} \right) = \frac{1}{2} m \ell^2 (\dot{\theta}^2 + \dot{\varphi}^2 \sin^2(\theta)) + mg\ell \cos \theta \quad (5.87)$$

So the Euler-Lagrange equation can be written as

$$-\frac{d}{dt} \nabla_{(3)} L \left(t, \begin{pmatrix} \theta(t) \\ \varphi(t) \end{pmatrix}, \begin{pmatrix} \dot{\theta}(t) \\ \dot{\varphi}(t) \end{pmatrix} \right) + \nabla_{(2)} L \left(t, \begin{pmatrix} \theta(t) \\ \varphi(t) \end{pmatrix}, \begin{pmatrix} \dot{\theta}(t) \\ \dot{\varphi}(t) \end{pmatrix} \right) = \mathbf{0} \quad (5.88)$$

5.5 Equality Constraints: Isoperimetric Case

Recall: Finite Dimensional Case Given $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$, consider

$$\begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ g(x) = \text{constant} \end{cases} \quad (5.89)$$

Suppose x_* is a regular point ($\nabla g(x_*) \neq 0$) is a minimizer, then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_*) + \lambda \nabla g(x_*) = 0 \quad (\text{Lagrange Multiplier}) \quad (5.90)$$

Remark 5.3. x_* is a critical point minimizing $f + \lambda g$ as well. Effectively, the method of Lagrange multiplier converts the constrained optimization to an unconstrained optimization problem.

Infinite Dimensional Case Consider the objective and constraint functions

$$F[u(\cdot)] = \int_a^b L^F(x, u(x), u'(x)) dx \quad (5.91)$$

$$G[u(\cdot)] = \int_a^b L^G(x, u(x), u'(x)) dx \quad (5.92)$$

and the optimization problem

$$\begin{cases} \min_{u(\cdot) \in \mathcal{A}} F[u(\cdot)] \\ G[u(\cdot)] = \text{constant} \end{cases} \quad (5.93)$$

Theorem 5.3. Suppose $u_*(\cdot)$ is a regular point, that is, the variational derivative $\frac{\delta G}{\delta u}(u_*) \neq 0$. Further, if $u_*(\cdot)$ is a minimizer of above constrained optimization problem, then $\exists \lambda \in \mathbb{R}$ such that:

$$\frac{\delta F}{\delta u}[u_*] + \lambda \frac{\delta G}{\delta u}[u_*] \equiv 0 \quad (5.94)$$

Example 5.7.

$$\mathcal{A} := \{u : [-a, a] \rightarrow \mathbb{R}, u \in C^1, u(-a) = u(a) = 0\} \quad (5.95)$$

$$L^G(x, z, p) = \sqrt{1 + p^2} \quad (5.96)$$

$$G[u(\cdot)] = \int_a^b \sqrt{1 + u'(x)^2} dx = \ell > 0 \quad (5.97)$$

$$L^F(x, z, p) = z \quad (5.98)$$

$$L^{-F}(x, z, p) = -z \quad (5.99)$$

$$F[u(\cdot)] = \int_a^b u(x) dx \quad (5.100)$$

$$\begin{cases} \min_{u(\cdot) \in \mathcal{A}} (-F)[u(\cdot)] \\ G[u(\cdot)] = \ell \end{cases} \quad (5.101)$$

Let $u_*(\cdot)$ be a minimizer, then Euler-Lagrange equation suggests

$$\frac{\delta(-F)}{\delta u} = -\frac{d}{dx} L_p^{-F} + L_z^{-F} \equiv 0 \quad (5.102)$$

$$\frac{\delta G}{\delta u} = -\frac{d}{dx} L_p^G + L_z^G \equiv 0 \quad (5.103)$$

By the necessary condition of minimization,

$$\frac{\delta(-F)}{\delta u} + \lambda \frac{\delta G}{\delta u} \equiv 0 \quad (5.104)$$

$$\implies -1 + \lambda \left(-\frac{d}{dx} \frac{p}{\sqrt{1+p^2}} + 0 \right) \equiv 0 \quad (5.105)$$

$$\implies \lambda \frac{d}{dx} \frac{p}{\sqrt{1+p^2}} \equiv -1 \quad (5.106)$$

$$\implies \lambda \frac{p}{\sqrt{1+p^2}} \equiv -x + C_1 \quad (5.107)$$

$$\implies \lambda^2 \frac{u'_*(x)^2}{1 + u'_*(x)^2} \equiv (C_1 - x)^2 \quad (\dagger) \quad (5.108)$$

Claim: Note that any solution $u_*(\cdot)$ to (\dagger) satisfies

$$(x - C_1)^2 + (u_*(x) - C_2)^2 = \lambda^2 \quad (5.109)$$

Check:

$$\frac{d}{dx} [2(x - C_1) + 2(u_*(x) - C_2)u'_*(x)] = 0 \quad (5.110)$$

which implies

$$u'_*(x) = -\frac{x - C_1}{u_*(x) - C_2} \quad (5.111)$$

$$\implies u'_*(x)^2 = \frac{(x - C_1)^2}{(u_*(x) - C_2)^2} \quad (\S) \quad (5.112)$$

Also,

$$(u'_*(x)^2)(u_*(x) - C_2)^2 = (x - C_1)^2 + (u_*(x) - C_2)^2 = \lambda^2 \quad (5.113)$$

$$\implies (u_*(x) - C_2)^2 = \frac{\lambda^2}{1 + u'_*(x)^2} \quad (\S\S) \quad (5.114)$$

Combine (\S) and $(\S\S)$,

$$\frac{\lambda^2}{1 + u'_*(x)^2} u'_*(x)^2 = (x - C_1)^2 \quad (5.115)$$

5.6 Equality Constraints: Holonomic Case

Setup (3-Dim Special Case) Minimize

$$F[x(\cdot), y(\cdot), z(\cdot)] = \int_a^b L(t, x(t), y(t), z(t), \dot{x}(t), \dot{y}(t), \dot{z}(t)) dt \quad (5.116)$$

with constraint

$$H(x(t), y(t), z(t)) \equiv 0 \quad (5.117)$$

Theorem 5.4 (Euler-Lagrange Equations). Let $\mathbf{x}_*(t) := \begin{pmatrix} x_*(t) \\ y_*(t) \\ z_*(t) \end{pmatrix}$ be the minimizer subject to constraint, then

$$\begin{pmatrix} \frac{\delta F}{\delta x} [x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \\ \frac{\delta F}{\delta y} [x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \\ \frac{\delta F}{\delta z} [x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \end{pmatrix} + \lambda(t) \begin{pmatrix} H_x[x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \\ H_y[x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \\ H_z[x_*(\cdot), y_*(\cdot), z_*(\cdot)](t) \end{pmatrix} = 0 \quad \forall t \in \mathbb{R} \quad (5.118)$$

where $\lambda(t)$ is a function here.

Example 5.8 (3-Dim Pendulum).

$$\min_{\mathbf{x}(\cdot) \in \mathcal{A}} F[x, y, z] = \int_a^b \frac{1}{2} m (\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2) - mgz(t) dt \quad (5.119)$$

$$s.t. \ H = \frac{1}{2} (x^2 + y^2 + z^2 - \ell^2) = 0 \quad (5.120)$$

In this case,

$$L(t, z_1, z_2, z_3, p_1, p_2, p_3) = \frac{1}{2}m(p_1^2 + p_2^2 + p_3^2) - mgz_3 \quad (5.121)$$

$$\begin{cases} L_{p_1} = mp_1 & L_{p_2} = mp_2 & L_{p_3} = mp_3 \\ L_{z_1} = 0 & L_{z_2} = 0 & L_{z_3} = -mg \end{cases} \quad (5.122)$$

Because

$$\frac{\delta F}{\delta x_i(t)} = -\frac{d}{dt}L_{p_i} + L_{z_i} \quad (5.123)$$

The Euler-Lagrange equation in this case is

$$\begin{pmatrix} -m\ddot{x}(t) \\ -m\ddot{y}(t) \\ -m\ddot{z}(t) - mg \end{pmatrix} + \lambda(t) \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} \equiv 0 \quad (\dagger\dagger) \quad (5.124)$$

Combing $(\dagger\dagger)$ and the constraint

$$\begin{cases} m\ddot{x}(t) = x(t) \\ m\ddot{y}(t) = y(t) \\ m\ddot{z}(t) = z(t) \\ x^2 + y^2 + z^2 = \ell^2 \end{cases} \quad (5.125)$$

Solving above system of equations provides the minimizer function to the proposed problem.

5.7 Geodesics on the Cylinder

Problem Find the string connecting two points on a cylinder of radius r such that the arc-length is minimized:

$$\min F[x(t), y(t), z(t)] := \int_a^b \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2} dt \quad (5.126)$$

$$s.t. \ H := x^2 + y^2 - r^2 = 0 \quad (5.127)$$

The Euler-Lagrange equations in this case are

$$\begin{cases} -\frac{d}{dt}L_{\dot{x}} + L_x = -2\lambda(t)x(t) \\ -\frac{d}{dt}L_{\dot{y}} + L_y = -2\lambda(t)y(t) \\ -\frac{d}{dt}L_{\dot{z}} + L_z = -2\lambda(t)z(t) \end{cases} \quad (5.128)$$

$$\text{where } L := \sqrt{p_1^2 + p_2^2 + p_3^2} \quad (5.129)$$

The Euler-Lagrange equations can be reduced to

$$\frac{d}{dt} \begin{pmatrix} x' / \sqrt{x'^2 + y'^2 + z'^2} \\ y' / \sqrt{x'^2 + y'^2 + z'^2} \\ z' / \sqrt{x'^2 + y'^2 + z'^2} \end{pmatrix} = \lambda(t) \begin{pmatrix} 2x \\ 2y \\ 0 \end{pmatrix} \quad (\dagger) \quad (5.130)$$

Definition 5.2. A curve is called **geodesic** if it satisfies the Euler-Lagrange necessary condition (†).

Definition 5.3. Given two points on the cylinder, being geodesic is not a sufficient condition to show one curve is actually the distance minimizing curve connecting these two points.

Example 5.9. Spirals are geodesics.

Proof. A spiral can be parametrized using time t as

$$\begin{cases} x(t) &= r \cos(t) \\ y(t) &= r \sin(t) \\ z(t) &= r\alpha \end{cases} \implies \begin{cases} x'(t) &= -r \sin(t) \\ y'(t) &= r \cos(t) \\ z'(t) &= \alpha \end{cases} \quad (5.131)$$

Therefore,

$$\sqrt{x'^2 + y'^2 + z'^2} = \sqrt{r^2 + \alpha^2} = \text{constant} \quad (5.132)$$

The Euler-Lagrange equations become

$$\frac{1}{\text{constant}} \frac{d}{dt} \begin{pmatrix} -r \sin(t) \\ r \cos(t) \\ \alpha \end{pmatrix} = 2\lambda(t) \begin{pmatrix} r \cos(t) \\ r \sin(t) \\ 0 \end{pmatrix} \quad (5.133)$$

$$\implies \frac{1}{\text{constant}} \begin{pmatrix} -r \cos(t) \\ -r \sin(t) \\ \alpha \end{pmatrix} = 2\lambda(t) \begin{pmatrix} r \cos(t) \\ r \sin(t) \\ 0 \end{pmatrix} \quad (5.134)$$

$$\implies \lambda(t) = -\frac{1}{2 \text{ constant}} \quad (5.135)$$

$$= -\frac{1}{2\sqrt{r^2 + \alpha^2}} \quad (5.136)$$

■

Example 5.10. Straight line is also geodesic.

Proof. Consider the straight line parameterized as

$$\begin{cases} x(t) &= \alpha \\ y(t) &= \beta \\ z(t) &= \gamma t \end{cases} \implies \begin{cases} x'(t) &= 0 \\ y'(t) &= 0 \\ z'(t) &= \gamma \end{cases} \quad (5.137)$$

In this case, $\sqrt{x'^2 + y'^2 + z'^2} = \gamma$.

The Euler-Lagrange equations are

$$\frac{d}{dt} \frac{1}{\gamma} \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix} = 2\lambda(t) \begin{pmatrix} \alpha \\ \beta \\ 0 \end{pmatrix} \quad (5.138)$$

$$\implies \lambda(t) = 0 \quad (5.139)$$

■

Example 5.11 (Another Example of Isoperimetric Constraints). Consider

$$F[u(\cdot)] = \int_0^a u(x) \sqrt{1 + u'(x)^2} dx \quad (5.140)$$

$$L^F = z \sqrt{1 + p^2} \quad (5.141)$$

$$u \in \mathcal{A} := \{u : [0, a] \rightarrow \mathbb{R} : u \in C^1, u(x) \geq 0, u(0) = b, u(a) = 0\} \quad (5.142)$$

$$G[u(\cdot)] = \int_0^a u(X) dx = \frac{ab}{2} \quad (5.143)$$

$$L^G = z - \frac{b}{2} \quad (5.144)$$

Part 1: Derive the Euler-Lagrange equations.

$$\frac{\delta F}{\delta u} = -\frac{d}{dx} \frac{zp}{\sqrt{1+p^2}} + \sqrt{1+p^2} \quad (5.145)$$

$$\frac{\delta G}{\delta u} = 1 \quad (5.146)$$

The Euler-Lagrange equations are

$$\frac{d}{dx} \frac{zp}{\sqrt{1+p^2}} - \sqrt{1+p^2} = \lambda \quad (5.147)$$

Part 2: Show that the following function satisfies the necessary condition.

$$u(x) = b \left(1 - \frac{x}{a}\right) \quad (5.148)$$

$$u'(x) = -\frac{b}{a} \quad (5.149)$$

Therefore,

$$\frac{d}{dx} \frac{zp}{\sqrt{1+p^2}} - \sqrt{1+p^2} = \frac{p^2}{\sqrt{1+p^2}} - \sqrt{1+p^2} \quad (5.150)$$

$$= \frac{\frac{b^2}{a^2}}{\sqrt{1+\frac{b^2}{a^2}}} - \sqrt{1+\frac{b^2}{a^2}} \quad (5.151)$$

$$= \frac{\frac{b^2}{a}}{\sqrt{a^2+b^2}} - \frac{1}{a} \sqrt{a^2+b^2} \quad (5.152)$$

$$= \frac{b^2}{a\sqrt{a^2+b^2}} - \frac{a^2+b^2}{a\sqrt{a^2+b^2}} \quad (5.153)$$

$$= \frac{-a}{\sqrt{a^2+b^2}} = \lambda \quad (5.154)$$

Obviously, $u(\cdot) \in C^1$ and $u(0) = b, u(a) = 0$.

Also,

$$G[u(\cdot)] = \int_0^a b\left(1 - \frac{x}{a}\right) dx \quad (5.155)$$

$$= \int_0^a b - \frac{bx}{a} dx \quad (5.156)$$

$$= bx - \frac{bx^2}{2a} \Big|_0^a \quad (5.157)$$

$$= ab - \frac{ba^2}{2a} \quad (5.158)$$

$$= \frac{ab}{2} \quad (5.159)$$

Example 5.12 (The Brachistochrone Problem). Parametrize the problem using $c : [0, T] \rightarrow \mathbb{R}^2$ to describe the curve such that $c(0) = A = (0, a)$ and $c(T) = B = (b, 0)$.

In particular, for each path of motion $x(t)$, the parametric representation can be constructed as

$$c(t) = (x(t), y(t)) = (x(t), u(x(t))) \quad (5.160)$$

$$\implies v(t) = \frac{d}{dt}c(t) = \dot{x}(t) \begin{pmatrix} 1 \\ \dot{u}(x(t)) \end{pmatrix} \quad (5.161)$$

The kinetic energy at time t is

$$T(t) = \frac{1}{2}m\dot{x}(t)^2[1 + \dot{u}(x(t))^2] \quad (5.162)$$

The potential energy at time t is

$$V(t) = mgu(x(t)) \quad (5.163)$$

By the conservation of energy, the total energy $E(t) := T(t) + V(t)$ is a constant function.

Re-write the total energy function

$$E(t) = \frac{1}{2}m\dot{x}(t)^2[1 + \dot{u}(x(t))^2] + mgu(x(t)) = mga \quad \forall t \in [0, T] \quad (5.164)$$

$$\implies \dot{x}(t) = \sqrt{\frac{2g(a - u(x(t)))}{1 + \dot{u}(x(t))^2}} \quad (5.165)$$

The total time for the point mass to travel from A to B is T :

$$T = \int_0^T 1 \, dt = \int_0^T \frac{1}{\dot{x}(t)} \dot{x}(t) \, dt \quad (5.166)$$

$$= \int_0^T \sqrt{\frac{1 + \dot{u}(x(t))^2}{2g(a - u(x(t)))}} \dot{x}(t) \, dt \quad (5.167)$$

$$= \int_{x(t_0)}^{x(t_1)} \sqrt{\frac{1 + \dot{u}(x(t))^2}{2g(a - u(x(t)))}} \, dx \quad (5.168)$$

$$= \int_0^b \sqrt{\frac{1 + \dot{u}(x(t))^2}{2g(a - u(x(t)))}} \, dx \quad (5.169)$$

The minimization problem becomes

$$\min F[u(\cdot)] = \int_0^b \sqrt{\frac{1 + \dot{u}(x(t))^2}{2g(a - u(x(t)))}} dx \quad (5.170)$$

$$s.t. u(\cdot) \in \mathcal{A} = \{u : [0, b] \rightarrow \mathbb{R} : u \in C^1, u(0) = a, u(b) = 0\} \quad (5.171)$$

The Euler-Lagrange equation is **TODO:** *verify this point*

$$(1 + u'(x)^2)(a - u(x))c^2 = 1 \text{ where } c = \text{constant} \quad (5.172)$$

Claim: the solution to above differentiable equation takes the form

$$u(x(t)) = a - k(1 - \cos t) \quad (5.173)$$

To verify the validation of the proposed solution

$$\frac{d}{dt} u'(x(t)) x'(t) = -k \sin t \quad (5.174)$$

$$\implies u'(x(t)) = -\frac{k \sin t}{x'(t)} \quad (5.175)$$

$$\implies c^2 \left(1 + \frac{k^2 \sin^2 t}{x'(t)^2}\right) k(1 - \cos t) = 1 \quad (5.176)$$

In particular, choose $c^2 k = \frac{1}{2}$:

$$\left(1 + \frac{k^2 \sin^2 t}{x'(t)^2}\right) (1 - \cos t) = 2 \quad (5.177)$$

$$\implies k^2 \sin^2 t (1 - \cos t) = x'(t)^2 (1 + \cos t) \quad (5.178)$$

$$\implies k^2 \sin^2 t (1 - \cos t)^2 = x'(t)^2 (1 + \cos t) (1 - \cos t) \quad (5.179)$$

$$= x'(t)^2 (1 - \cos^2 t) = x'(t) \sin^2 t \quad (5.180)$$

$$\implies k^2 (1 - \cos t)^2 = x'(t)^2 \quad (5.181)$$

$$\implies k(1 - \cos t) = x'(t) \quad (5.182)$$

$$\implies kt - k \sin t = x(t) + C \quad (5.183)$$

$$\implies c(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} x(t) \\ u(x(t)) \end{pmatrix} \quad (5.184)$$

$$= \begin{pmatrix} kt - k \sin t \\ a - k(1 - \cos t) \end{pmatrix} \quad (5.185)$$

With initial conditions:

$$c(0) = (0, a) = A \quad (5.186)$$

$$c(T) = \begin{pmatrix} k(T - \sin T) \\ a - k(1 - \cos T) \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \quad (5.187)$$

6 Appendix

Some Basic ODEs

$$u'(x) = 0 \implies u(x) = c \tag{6.1}$$

$$u'(x) = u(x) \implies u(x) = Ae^x \tag{6.2}$$

$$u''(x) = 0 \implies u(x) = Ax + B \tag{6.3}$$

$$u''(x) = u(x) \implies u(x) = Ae^x + Be^{-x} \tag{6.4}$$

$$u''(x) = -u(x) \implies u(x) = A \sin x + B \cos x \tag{6.5}$$