Study Notes for Optimization

Mike Mingcheng Wei

Contents

Chapter 1

Optimization Theorem

Definition 1 *Let $x \subseteq R^n$, ε-**neiborhood** of x is $N_\varepsilon(x) = \{y \in R^n | \, \|x - y\| \leq \varepsilon\}$.*

Definition 2 *Let $S \subseteq R^n$, if S contains an ε-neiborhood of each of its point. Then S is **open**.*

Theorem 3 *Let $S \subseteq R^n$, if for each $x \in S$, $\exists \ \varepsilon > 0$ such that $N_\varepsilon(x) \subseteq S$. Then S is **open***

Theorem 4 *Let $S \subseteq R^n$, if $S = Int(S)$. Then S is **open***

Definition 5 *Let $S \subseteq R^n$, if its complement $\bar{S} = R^n - S$ is open. Then S is **closed***

Theorem 6 *Let $S \subseteq R^n$, if $S = Closure(S)$. Then S is **closed***

Definition 7 *Let $S \subseteq R^n$, if $\exists \ \varepsilon > 0$, such that $N_\varepsilon(x) \subseteq S$. Then x is **interior point** of S. ($Int(S)$: set of all interior point of S)*

Definition 8 *Let $S \subseteq R^n$, if for each $\varepsilon > 0$, $N_\varepsilon(x)$ containts a point in S and a point not in S. Then x is **boundary point** of S. (∂S: set of all boundary point of S)*

Definition 9 *Let $S \subseteq R^n$, the **closure** of S is the union of S and boundary points of S. ($Closure(S) = S \cup \partial S$)*

Definition 10 *If $\exists m > 0$ such that $S \subseteq \{x \in R^n | \, \|x\| \leq m\}$. Then $S \subseteq R^n$ is **bounded***

Definition 11 *If it is closed and bounded. Then $S \subseteq R^n$ is **compact***

Theorem 12 ***Weierstrass Theorem**: A continuous function defined on a compact set $S \subseteq R^n$ attain a mimimum on S.*

1.1 General Condition

Generally, if we have objective function, which is differentiable at most points, then we can **use FOC to find the candidate point for local and global minimum**. Because the local and global minimum only occurse at the following cases: the boundary point, discontinuous points, continuous but undifferentiable points, and FOC points.

Theorem 13 *(Neccessary condition for interior solution optimization) Consider $f : C \to R$. If x^* is an interior local maximum/minimum of f and f is differentiable at x^*. Then*

$$\nabla f(x^*) = 0$$

Theorem 14 *(Transformation) Consider any function $f : C \to R$, where $C \in R^n$. Let $h : R \to R$ be any strictly decreasing/increaseing function. Then x^* minimize/maximize f on C iff x^* minimize/maximize $\hat{f} = h \circ f$ on C*

Theorem 15 [1] *(Sufficient condition for optimization) Let the mathematical programming problem is defined as follows*

$$\left\{ \begin{array}{c} \min_{x \in S} f(x) \\ S \text{ is convex set} \end{array} \right\}$$

Then:

$f(x)$	*convex*	*strictly convex*
$\bar{x}$ *local min*	$\bar{x}$ *global min*	$\bar{x}$ *unique global min*
$\nabla f(\bar{x}) = 0$	$\bar{x}$ *global min*	$\bar{x}$ *unique global min*

$f(x)$	*quasi-convex*	*strictly quasi-convex*	*strong quasi-convex*
$\bar{x}$ *local min*	NA	$\bar{x}$ *global min*	$\bar{x}$ *unique global min*
$\nabla f(\bar{x}) = 0$	NA	NA	NA

$f(x)$	*pseudo-convex*	*strictly pseudo-convex*
$\bar{x}$ *local min*	$\bar{x}$ *global min*	$\bar{x}$ *unique global min*
$\nabla f(\bar{x}) = 0$	$\bar{x}$ *global min*	$\bar{x}$ *unique global min*

This above theorem only state when the local minimum can be generalized to global minimum, but it does not say relationship between FOC and local minimum.

Theorem 16 *(Avriel, 1976:146) If f is quasi-convex on a convex set and x is a strict local minimum, then x is a strict global minimum.*

Definition 17 *If f is a convex function and S is a convex set, then this mathematical programming is a* ***convex program***

Theorem 18 *Characterist of a global min: consider the convex program:*

$$\left\{ \begin{array}{c} \min_{x \in S} f(x) \\ S \text{ is convex set; } f(x) \text{ is convex function on } S \end{array} \right\}$$

then, we have the following properties:
a. $\bar{x}$ is a global minimal iff $\exists$ a subgradient ξ at $\bar{x}$ such that $\xi^t(x - \bar{x}) \geq 0$ for $\forall x \in S$;
b. if f is differentiable, then $\bar{x}$ is a global minimal iff $\nabla f(\bar{x})(x - \bar{x}) \geq 0$ for $\forall x \in S$;
c. if $\nabla f(\bar{x}) = 0$, then $\bar{x}$ is global minimal; (if, at $\bar{x}$, $\exists$ a subgradient $\xi = 0$, then $\bar{x}$ is global minimal);
d. if $\bar{x}$ is local minimal, and $\bar{x} \in int(S)$, then $\nabla f(\bar{x}) = 0$. (if $\bar{x}$ is local minimal, and $\bar{x} \in int(S)$, then $\exists$ a subgradient $\xi = 0$ at $\bar{x}$.)

Theorem 19 *Consider the convex program:*

$$\left\{ \begin{array}{c} \min_{x \in S} f(x) \\ S \text{ is convex set; } f(x) \text{ is convex function on } S \end{array} \right\}$$

If S is compact, then $\exists$ an extreme point optimal solution. (If f is strictly convex, then it is NOT neccessary the extrem point is unique.)

[1] Some textbook and notes use strongly quasi-convex as definition for strictly quasi-convex function

1.2 Geometric Optimality Condition

Definition 20 *For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ S \subseteq R^n, S \neq \phi \end{array} \right\}$$

Let $\bar{x} \in S$*, the* ***improving direction set at*** $\bar{x}$ *is defined as* $F = \{d \in R^n | f(\bar{x}+\lambda d) < f(\bar{x})$ *for* $\forall \lambda \in (0, \delta)$ *for some* $\delta > 0\}$*;*

Let $\bar{x} \in S$*, the* ***feasible direction set at*** $\bar{x}$ *is defined as* $D = \{d \in R^n | d \neq 0,\ \bar{x} + d\lambda \in S$ *for* $\forall \lambda \in (0, \delta)$ *for some* $\delta > 0\}$*;*

Claim 21 *If* f *is differentiable at* $\bar{x}$*, then improving direction set at* $\bar{x}$ *is defined as* $F_0 = \{d \in R^n | f(\bar{x})^t d < 0\}$*. (clearly,* $F_0 \subseteq F$*)*

Theorem 22 *For mathematical programming:*

$$\min_{x \in S} f(x)$$

Let $\bar{x} \in S$*, if* $\bar{x}$ *is local minimum, then* $F \cap D = \phi$*. (There is no feasible improvming direction)*

Theorem 23 *For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ f \text{ is Psudo-convex function; } S \text{ be convex set} \end{array} \right\}$$

Let $\bar{x} \in S$*, then* $F \cap D = \phi \underset{by\ continuity}{\Rightarrow} F_0 \cap D = \phi \Rightarrow \bar{x}$ *is local minimal*$\Rightarrow \bar{x}$ *is global minimal.*

Definition 24 *For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ S = \{x \in R^n | g_i(x) \leq 0,\ i = 1, ..., m\} \text{ and } g_i(x) \text{ is differentiable} \end{array} \right\}$$

Let $\bar{x} \in S$*, then the* ***index set of binding constraints at*** $\bar{x}$ *is defined as* $I = \{i | g_i(\bar{x}) = 0\}$*. And the feasible direction set at* $\bar{x}$ *is defined as* $G_0 = \{d \in R^n | \nabla g_i(\bar{x}) < 0$ *for* $\forall i \in I\}$*;*

Claim 25 $G_0 \subseteq \dot{D}$

Claim 26 *Let* g_i*,* $i = 1, ..., m$ *be strictly psudo-convex, then* $G_0 = \dot{D}$*.*

Claim 27 *If* $g_i(x)$ *is quasi-convex functions, then* S *is convex set.*

Theorem 28 *For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ S = \{x \in R^n | g_i(x) \leq 0, i = 1, ..., m\},\ g_i(x) \text{ is strictly Psudo-convex} \\ f \text{ is psudo-convex} \end{array} \right\}$$

Let $\bar{x} \in S$*, then* $F \cap D = \phi \underset{by\ continuity}{\Rightarrow} F_0 \cap G_0 = \phi \Rightarrow \bar{x}$ *is local minimal*$\Rightarrow \bar{x}$ *is global minimal*

1.3 Fritz-John Optimality Condition

Theorem 29 *(Fritz-John Optimality condition: Neccessary)For mathematical programming:*

$$\left\{\begin{array}{c}\min_{x\in S} f(x)\\ S=\{x\in R^n|g_i(x)\le 0, i=1,...,n\};\ f\ \textit{and}\ g_i(\bar{x}),\ i\in I\ \textit{is differentiable}\\ g_i(\bar{x}),\ i\notin I,\ \textit{continuous}\end{array}\right\}$$

Let $\bar{x}\in S$ and $I=\{i|g_i(\bar{x})=0\}$. If $\bar{x}$ is local minimum, then $\exists u_0$, u_i, $i\in I$ such that the following holds:

$$\left\{\begin{array}{c} u_0\nabla f(\bar{x})+\sum_{i\in I} u_i\nabla g_i(\bar{x})=0\\ u_0\ge 0, u_i\ge 0\ \textit{for}\ i\in I\\ \binom{u_0}{u_i, i\in I}\ne 0\end{array}\right\}$$

(PS: if $u_0\ne 0$, we then can refer to KKT condition.)(PS2: proof base on seperation theorem: Farkas' Lemma.)

Claim 30 *$F_0\cap G_0=\phi$ at $\bar{x}$ iff $\bar{x}$ is Fritz-John point. (Gorden's Theorem)*

Claim 31 *If $\nabla f(\bar{x})=0$ then $\bar{x}$ is a Fritz-John point.*

Claim 32 *If $\exists k\in I$ such that $\nabla g_k(\bar{x})=0$, then $\bar{x}$ is a Fritz-John point.*

Claim 33 *If exist a set of multiplier u_i, $i\in I$ such that $\sum_{i\in I} u_i\nabla g_i(\bar{x})=0$, and u_i are not all zero, then $\bar{x}$ is a Fritz-John point.*

Claim 34 *If $\bar{x}$ is local minimum, then $\bar{x}$ is a Fritz-John point.*

Claim 35 *If $\bar{x}$ is a Fritz-John point, then $\bar{x}$ is NOT neccessarily local minimum. (Can be inflection point)*

Theorem 36 *(Fritz-John Optimality condition: Sufficient)For mathematical programming:*

$$\left\{\begin{array}{c}\min_{x\in S} f(x)\\ S=\{x\in R^n|g_i(x)\le 0, i=1,...,n\};\ f\ \textit{and}\ g_i(\bar{x}),\ i\in I\ \textit{is differentiable}\\ g_i(\bar{x}),\ i\notin I,\ \textit{continuous}\end{array}\right\}$$

Let $\bar{x}\in S$ and $I=\{i|g_i(\bar{x})=0\}$. Let S' denotes relaxed feasible region for this problem in which the nonbinding constraints are dropped.

a). If there exists a $N_\varepsilon(\bar{x})$ such that $f(x)$ is psudo-convex over $N_\varepsilon(\bar{x})\cap S'$ and $g_i(\bar{x})$, $i\in I$, are strictly psudo-convex over $N_\varepsilon(\bar{x})\cap S'$, then $\bar{x}$ is a local minimal.

b). If $f(x)$ is psudo-convex at $\bar{x}$ and $g_i(\bar{x})$, $i\in I$, are both strictly psudo-convex and quasi-convex at $\bar{x}$, then $\bar{x}$ is global optimal solution.

1.4 KKT Optimality Condition

Theorem 37 *KKT condition = F-J condition + $u_0>0$*

Theorem 38 *(KKT condition for Inequality Constraint: neccessary)For mathematical programming:*

$$\left\{\begin{array}{c}\min_{x\in S} f(x)\\ S=\{x\in R^n|g_i(x)\le 0, i=1,...,n\};\ f\ \textit{and}\ g_i(\bar{x}),\ i\in I\ \textit{is differentiable}\\ g_i(\bar{x}),\ i\notin I,\ \textit{continuous}\end{array}\right\}$$

Let $\bar{x}\in S$ and $I=\{i|g_i(\bar{x})=0\}$. If $\bar{x}$ is local minimum and one Constraint Qualification, CQ, holds, then $\exists u_i$, $i\in I$, such that the following holds:

$$\nabla f(\bar{x})+\sum_{i\in I} u_i\nabla g_i(\bar{x})=0,\ u_i\ge 0$$

Claim 39 *Constraint Qualification:*
a). $\nabla g_i(\bar{x})$, $i \in I$, are linearly independent;
b). $G_0 \neq 0$;
c). $g_i(\bar{x})$ is linear; (Abaidie's CQ)
d). Exit a strict interrial point
e). ...

Claim 40 *If $\bar{x}$ is local minimum, then $\bar{x}$ is NOT neccessarily a KKT point. (For example, if $\nabla g_i(\bar{x})$, $i \in I$, are linearly dependent, then $\sum_{i \in I} u_i \nabla g_i(\bar{x}) = 0$ for some u_i) (Neccessary condition needs some CQs)*

Claim 41 *Suppose we have a convex program, if $\bar{x}$ is a global minimum, then $\bar{x}$ is NOT neccessarily a KKT point. (For example, only one fesible point and $\nabla g_i(\bar{x})$ are dependent.) (Neccessary condition needs some CQs)*

Claim 42 *If $\bar{x}$ is a KKT point, then $\bar{x}$ is NOT neccessarily a local minimum. (For example, $\bar{x}$ is saddle point and no binding constraints.)*

Claim 43 *Suppose we have a convex program, if $\bar{x}$ is a KKT point, then $\bar{x}$ is a global minimum.*

Theorem 44 *Let $G_0^{'} = \{d \in R^n | d \neq 0, \nabla g_i(\bar{x})^t d \leq 0 \text{ for } \forall i \in I\}$, then $F_o \cap G_0^{'} = \phi$ at $\bar{x}$ iff $\bar{x}$ is KKT point. (By Farkas' Lemma)*

Theorem 45 *(KKT condition for Inequality Constraint: sufficient)For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ S = \{x \in R^n | g_i(x) \leq 0, i = 1, ..., n\};\ f \text{ and } g_i(\bar{x}),\ i \in I \text{ is differentiable} \\ g_i(\bar{x}),\ i \notin I, \text{ continuous} \end{array} \right\}$$

Let $\bar{x} \in S$ is a KKT point and $I = \{i | g_i(\bar{x}) = 0\}$. Let $S^{'}$ denotes relaxed feasible region for this problem in which the nonbinding constraints are dropped.

a). If there exists a $N_\varepsilon(\bar{x})$ such that $f(x)$ is psudo-convex over $N_\varepsilon(\bar{x}) \cap S^{'}$ and $g_i(\bar{x})$, $i \in I$, are quasi-convex over $N_\varepsilon(\bar{x}) \cap S^{'}$, then $\bar{x}$ is a local minimal.

b).If $f(x)$ is psudo-convex at $\bar{x}$ and $g_i(\bar{x})$, $i \in I$, are quasi-convex at $\bar{x}$, then $\bar{x}$ is global optimal solution.

Theorem 46 *(KKT neccessary condition for Inequality & Equality Constraint)For mathematical programming:*

$$\left\{ \begin{array}{c} \min\limits_{x \in S} f(x) \\ g_i(x) \leq 0, i = 1, ..., n \\ h_j(x) = 0, j = 1, ..., l \\ f\ ,\ h_j(x), \text{ and } g_i(\bar{x}),\ i \in I, \text{ is differentiable; } g_i(\bar{x}),\ i \notin I, \text{ continuous} \end{array} \right\}$$

Let $\bar{x} \in S$ and $I = \{i | g_i(\bar{x}) = 0\}$. If $\bar{x}$ is local minimum and one Constraint Qualification, CQ, holds, then $\exists u_i$, $i \in I$, such that the following holds:

$$\nabla f(\bar{x}) + \sum_{i \in I} u_i \nabla g_i(\bar{x}) + \sum\nolimits_{j=1}^{l} \bar{v}_j \nabla h_i(\bar{x}) = 0,\ u_i \geq 0$$

Claim 47 *Constraint Qualification:*
a). $\nabla g_i(\bar{x})$, $i \in I$, are linearly independent;
b). ...

Theorem 48 *(KKT condition for speacial function form: sufficient) Let the mathematical programming is defined as follows:*

$$\left\{\begin{array}{c} Min\ f(x) \\ g_i(x) \le 0,\ i = 1, ..., m \\ h_j(x) = 0,\ j = 1, ..., l \\ f\ convex;\ g_i(x)\ convex;\ h_j(x)\ linear;\ all\ function\ differentiable \end{array}\right\}$$

If the following KKT condition holds for $\bar{x}$:

$$\left\{\begin{array}{c} Primal\ Feasibility\ condition:\ {}^{g_i(\bar{x})\le 0,\ i=1,...,m}_{h_j(\bar{x})=0,\ j=1,...,l} \\ Dual\ Feasibility\ condition:{}^{\nabla f(\bar{x})+\sum_{i=1}^{m}\bar{u}_i\nabla g_i(\bar{x})+\sum_{j=1}^{l}\bar{v}_j\nabla h_i(\bar{x})=0}_{\bar{u}_i\ge 0, i=1,...,m} \\ Complementary\ Slackness\ condition:\ \bar{u}_i g_i(\bar{x}) = 0, i = 1, ..., m \end{array}\right\}$$

Then, $\bar{x}$ is global minimal point.

1.5 Duality Theory & Saddle Point Optimality Theorem

Let the original problem defined as follows:

$$(P):\ \left\{\begin{array}{c} Min\ f(x) \\ g(x) = \begin{pmatrix} g_1(x) \\ ... \\ g_n(x) \end{pmatrix} \le 0 \\ x \in S \subseteq R^m \end{array}\right\}$$

Define the duality problem as follows:

$$(D):\ \left\{\begin{array}{c} Max\ \theta(u) \\ u \ge 0 \\ \text{where } \theta(u) = \underset{x\in S}{Min}\{f(x) + u^t g(x)\} \end{array}\right\}$$

Theorem 49 *(Bazaraa, Sherali, and Shetty:Theorem6.3.1)The dual function, $\theta(u) = \underset{x\in S}{Min}\{f(x) + u^t g(x)\}$, is concave function.*

Theorem 50 *(Bazaraa, Sherali, and Shetty:Theorem6.3.3) If S is compact, f and g be continuous, the optimal solutions to the dual function is singleton. Then dual function is differentiable at u with $\nabla_u \theta(u) = g(x^*(u))$. ($x^*(u)$ is the optimal solution function for the dual function.)*

Theorem 51 *(Weak Duality Theorem) If $\bar{x}$ is feasible to (P) and $\bar{u}$ is feasible to (D), then $\theta(\bar{u}) \le f(\bar{x})$. ($\theta(\bar{u})$ is lower bound for original problem; $f(\bar{x})$ is upper bound for the dual problem)*

Claim 52 *If $\bar{x}$ is feasible to (P) and $\bar{u}$ is feasible to (D) and $\theta(\bar{u}) = f(\bar{x})$, then $\bar{x}$ solves (P) and $\bar{u}$ solves (D).*

Claim 53 *If (D) objective is unbounded (goes to ∞), then (P) is infeasible.*

Claim 54 *If (P) objective is unbounded (goes to $-\infty$), then $\theta(u) = -\infty$ for $\forall u \ge 0$.*

Theorem 55 *(Strong Duality Theorem) If $f(x)$ and $g(x)$ are convex functions and S is convex set, and $\exists \bar{x} \in S$ such that $g(x) < 0$, then there is no duality gap. ($\exists \bar{x} \in S$ such that $g(x) < 0$ sever as Slater's CQ)*

Theorem 56 *(Saddle Point Optimality Theorem: SPOT) Define the Lagraingian Function $L(x, u) = f(x) + u^t g(x)$. If there exist $\bar{x} \in S$ and $\bar{u} \ge 0$ such that $L(\bar{x}, u) \le L(\bar{x}, \bar{u}) \le L(x, \bar{u})$ for $\forall x \in S$ and $\forall u \ge 0$, then $\bar{x}$ solves (P) and $\bar{u}$ solves (D).*

Claim 57 *(Relationship with KKT point): If $f(x)$ and $g(x)$ are convex functions, then if $\bar{x}$ is KKT point, then $(\bar{x}, \bar{u})$ satisfies the SPOT.*

Claim 58 *(Relationship with KKT point): If $\bar{x} \in Int(S)$, then if $(\bar{x}, \bar{u})$ satisfies the SPOT, then $\bar{x}$ is KKT point.*

Chapter 2

Value Function and Solution Function Properties Under Optimization

2.1 Value Function properties

Theorem 59 *Non-negative weighted maximum:* $f = \max\{w_1 f_1, ..., w_n f_n\}$ *where* $f_1, ..., f_n$ *are convex;* $w_1, ..., w_n$ *are non-negative. Then* f *is convex.*

Theorem 60 *Non-negative weighted maximum:* $f = \max\{w_1 f_1, ..., w_n f_n\}$ *where* $f_1, ..., f_n$ *are quasi-convex;* $w_1, ..., w_n$ *are non-negative. Then* f *is quasi-convex.*

Theorem 61 *Non-negative weighted maximum:* $f = \max\{w_1 f_1, ..., w_n f_n\}$ *where* $f_1, ..., f_n$ *are supermodular;* $w_1, ..., w_n$ *are non-negative. Then* f *is supermodular.*

Theorem 62 *If* Y *is a nenempty set and* $f(\cdot, y)$ *is a quasi-convex function on a convex set* X *for every* $y \in Y$*. Then* $g(x) = \sup_{y \in Y} f(x, y)$ *is a quasi-convex function on* X*.*

Theorem 63 *(Preservation under minimization) Let* $f(x, y) : X \times Y(x) \to R$*. If* $Y(x)$ *is a nonempty set for every* $x \in X$*,* X *is convex set, and* $(X, Y(x))$ *is convex set,* $f(x, y)$ *is quasi-convex function on* $(X, Y(x))$*,* $g(x) > -\infty$ *for* $\forall x \in X$*. Then* $g(x) = \inf_{y \in C} f(x, y)$ *in convex on* X*. (In Heyman and Sobel, 1984:525, it state the same result with more strong condition by requiring* $f(x, y)$ *be convex)*

Theorem 64 *(Preservation under maximization)(Heyman and Sobel, 1984:525) Let* $f(x, y) : X \times Y \to R$*. If* Y *is non-empty and* X *is convex set,* $f(\cdot, y)$ *is convex function on a convex set* X *for each* $y \in Y$*. Then* $g(x) = \sup_{y \in Y} f(x, y)$ *in convex on* X*.*

Theorem 65 *(Envelope Theorem) For the following parameterized mathematical programming*

$$\left\{ \begin{array}{c} \underset{x}{Min}\ f(x, r) \\ s.t.\ \ g_i(x, r) = 0,\ \ i = 1, ..., M \end{array} \right\}$$

Let $v(r)$ *denotes the value function of this problem:* $v(r)$ *is the optimal value attained by* $f(\cdot)$ *when the parameter verctor is* r*. Let* $x(r)$ *denotes the optimal solution of this problem:* $x(r)$ *is the optimal solution which solve* $\min f(\cdot)$ *when the parameter verctor is* r*. Let* $\lambda(r)$ *be the lagrange multipliers associated with the minimizer solution* $x(r)$ *at* r*. If* $v(r)$ *is differentiable*[1]*, then*

$$\frac{\partial v(r)}{\partial r_i} = \frac{\partial f(x(r), r)}{\partial r_i} - \sum\nolimits_{m=1}^{M} \lambda_m(r) \frac{\partial g_m(x(r), r)}{\partial r_i} \text{ for } i = 1, ..., N$$

[1] The differentiability requires additional condition: e.g. $x(r)$ is singleton. For detailed treatment, refer to "Milgrom - 1999 - The Envelope Theorems"

(PS: for inequality constraint, this still holds due to the KKT condition that require CS condition)

Remark 66 *The continuity of the value function $v(r)$ is garrenteed by the fact that $f(x,r)$ is continuous in r.*

Theorem 67 *(Theorem 1 of Paul Milgrom 1999: The envelope theorems) $V(t) = \max_{x\in K} f(x,t)$ and $t \in [0,1]$. Let $K \subset X$ be non-empty and compact and suppose that for all t, $f(\cdot,t) : K \to R$ is upper semi-continuous. Further assume that the partial derivative $f_t(x,t)$ exists and is a continuous function of (x,t). Then*

a) V has bounded right-hand and left-hand derivatives on $[0,1)$ and $(0,1]$, respectively, and these are given by the formulas:

$$V_+^{'}(t) = \max_{x\in x^*(t)} f_t(x,t) \text{ and } V_-^{'}(t) = \min_{x\in x^*(t)} f_t(x,t)$$

b). V is almost everywhere differentiable on $(0,1)$ and whereever the derivative exists,

$$V'(t) = f_t(x(t),t) \text{ for any } x(t) \in x^*(t)$$

c) for every mt$\in [0,1]$ and any selection $x(t)$ from $x^*(t)$,

$$V(t) = V(0) + \int_0^t f_t(x(s),s)\,ds$$

Remark 68 *This theorem implies that whenever $x^*(t)$ is singleton (e.g. if $f(\cdot,t)$ is strict concave function, then $x^*(t)$ must be singleton/unique), then $V(t)$ is differentiable.*

Theorem 69 *(Corollary 5 of Paul Milgrom 1999: The envelope theorems) $V(t,s) = \max_x f(x,t)$ subject to $g(x,s) \geq 0$. Suppose that, for all s, $K(s)$, the feasible set of x, is non-empty and compact. Suppose that for all t and s, the functions $f(\cdot,t) : X \to R$ and $g(\cdot,s) : X \to R^N$ are continuous and concave. In addition, assume that (i) $f(\cdot,t)$ is strictly concave and (ii) the partial derivative $f_t(x,t)$ exists and is a continuous function of (x,t). Then, $V_t(t,s)$ exists and satisfies $V_t(t,s) = f_t(x^*(t,s),t)$.*

Theorem 70 *(Theorem 6 of Paul Milgrom 1999: The envelope theorems) $V(t,s) = \max_{x\in X} f(x,t)$ subject to $g(x,s) \geq 0$ (By saddle point theorem, $V(t,s) = \min_{\lambda\geq 0}\max_{x\in X}(f(x,t) + \lambda\cdot g(x,s))$). Suppose that, for all s, $K(s)$, the feasible set of x, is interior of X, non-empty, and compact. Suppose that for all t and s, the functions $f(\cdot,t) : X \to R$ and $g(\cdot,s) : X \to R^N$ are continuous and concave. In addition, assume that (i) $f(\cdot,t)$ is strictly concave; (ii) the partial derivative $g_s(x,s)$ exists and are a continuous function of (x,t); and (iii) for all s, $t \in (0,1)$ and $x \in X$, $f_x(x,t)$ and $g_x(x,s)$ exist and $g_x(x,s)$ has full row rank. Then, the solutions λ^* and x^* exist and are singletons, and $V_s(t,s)$ exists on the interval $(0,1)$ and satisfies $V_s(t,s) = \lambda^*(t,s)\cdot g_s(x^*(t,s),s)$, where $x^*(t,s) = x^*(\lambda^*(t,s),t,s)$.*

2.2 Solution Function properties

Theorem 71 *(P115, 7.6, Porteus 2002) Suppose that f_1 and f_2 are both differentiable functions defined on R that have finite minimizers, S_1 and S_2, respectively, and that $f_1^{'} \leq f_2^{'}$. Then,*

1. If S_1 and S_2 are unique minimizers, then $S_1 \geq S_2$;

2. In general, here exist minimizers, say S_1^* and S_2^*, of f_1 and f_2, respectively, that are finite and satisfy $S_1^* \geq S_2^*$.

Theorem 72 *(Theorem 2 of Nachbar Monotone Comparative statistics) Let $f : R^2 \to R$, let $C \subseteq R$, and for each $\theta \in R$, let $\phi(\theta)$ be the set of solutions, assumed non-empty, to the problem*

$$\max_{x\in C} f(x,\theta)$$

If f is supermodular then $\phi(\theta)$ is weakly increasing.

Theorem 73 *(Theorem 3 of Nachbar Monotone Comparative statistics: for general case: multivariable case) Let* $f : R^{N+M} \to R$, *let* $C \subseteq R^N$, *and for each* $\theta \in R^M$, *let* $\phi(\theta)$ *be the set of solutions, assumed non-empty, to the problem*

$$\max_{x \in C} f(x, \theta)$$

If f *is supermodular in* x *and exhibits increasing differences in* (x, θ) *then* $\phi(\theta)$ *is weakly increasing.*

Theorem 74 *(Theorem 4 of Nachbar Monotone Comparative statistics) Let* $f : R^{N+M} \to R$, *let* $C \subseteq R^N$, *and for each* $\theta \in R^M$, *let* $\phi(\theta)$ *be the set of solutions, assumed non-empty, to the problem*

$$\max_{x \in C} f(x, \theta)$$

If f *is quasi-supermodular in* x *and satisfies the single crossing property in* (x, θ) *then* $\phi(\theta)$ *is weakly increasing.*

Theorem 75 *(Theorem 5 of Nachbar Monotone Comparative statistics) Let* $f : R^2 \to R$, *let* $C \subseteq R$ *be an interval, and for each* $\theta \in R$, *let* $\phi(\theta)$ *be the set of solutions, assumed non-empty, to the problem*

$$\max_{x \in C} f(x, \theta)$$

If f *satisfies the interval order dominance then* $\phi(\theta)$ *is weakly increasing.*

Chapter 3

Separation Theorem

Theorem 76 *(Closest Point Theorem) Let* $x \subseteq R^n$, $S \neq \phi$, *and* S *closed and convex; let* $y \notin S$, *then* $\exists$ *a unique point* $\bar{x} \in S$ *that is a minimum distance from* y. *Furthermore,* $(y - \bar{x})^t(x - \bar{x}) \leq 0$ *for* $\forall\ x \in S$.

Definition 77 *Let* S_1, $S_2 \subseteq R^n$, *let* $H = \{x \in R^n | P^t x = \alpha\}$. H ***separate*** S_1 *and* S_2 *if* $P^t x \leq \alpha$ *for* $\forall$ $x \in S_1$ *and* $P^t x \geq \alpha$ *for* $\forall\ x \in S_2$. H *is called a* ***separate hyperplane.***

Definition 78 *Let* S_1, $S_2 \subseteq R^n$, *let* $H = \{x \in R^n | P^t x = \alpha\}$. H ***proper separate*** S_1 *and* S_2 *if* $P^t x < \alpha$ *for* $\forall\ x \in S_1$, $P^t x > \alpha$ *for* $\forall\ x \in S_2$, *and* $S_1 \cup S_2 \nsubseteq H$.

Definition 79 *Let* S_1, $S_2 \subseteq R^n$, *let* $H = \{x \in R^n | P^t x = \alpha\}$. H ***strict separate*** S_1 *and* S_2 *if* $P^t x < \alpha$ *for* $\forall\ x \in S_1$ *and* $P^t x > \alpha$ *for* $\forall\ x \in S_2$.

Definition 80 *Let* S_1, $S_2 \subseteq R^n$, *let* $H = \{x \in R^n | P^t x = \alpha\}$. H ***strong separate*** S_1 *and* S_2 *if* $P^t x \leq \alpha$ *for* $\forall\ x \in S_1$ *and* $P^t x \geq \alpha + \varepsilon$ *for* $\forall\ x \in S_2$.

Theorem 81 *Let* $x \subseteq R^n$, $S \neq \phi$, *and* S *closed and convex; let* $y \notin S$, *then* $\exists$ *a hyperplane that strongly separate* y *and* S.

Definition 82 H *is a supporting hyperplane to* S *at* $\bar{x}$ *if* $S \subseteq \{x \in R^n | P^t(x - \bar{x}) \leq 0\}$ *or* $S \subseteq \{x \in R^n | P^t(x - \bar{x}) \geq 0\}$.

Notes: the closest point theorem also constracte a supporting hyperplane to S at $\bar{x}$.

Theorem 83 *Let* $x \subseteq R^n$, $S \neq \phi$, *and* S *closed and convex; let* $\bar{x} \in \partial S$, *then* $\exists$ *a supporting hyperplane at* $\bar{x}$.

Theorem 84 *Let* $x \subseteq R^n$, $S \neq \phi$, *and* S *closed and convex; let* $y \notin S$, *then* $\exists$ *a hyperplane that separates* y *and* S.

Theorem 85 *(Farkas' Lemma)Exactly one of the following is ture:*
I: $Ax \leq 0$, $c^t x > 0$ *for some* $x \in R^n$. *(c does not lie in the cone generated by the rows of A)*
II: $A^t y = c$, $y \geq 0$ *for some* $y \in R^m$. *(c can be writen as non-negative combination of the columns of* A^t *or rows of A– lies in the cone generated by the row of A)*

Theorem 86 *(Gordarn's Theorem)Exactly one of the following is ture:*
I: $Ax < 0$ *for some* $x \in R^n$. *(*$\exists x$ *that made an obscule angle with each row of A)*
II: $A^t y = 0$, $y \geq 0$ *for some non-zero* $y \in R^m$.

Chapter 4

Optimal Control Theory

4.1 The Calculus of Variations (From Kirk 2004 and Friesz 2008)

4.1.1 Basics

Definition 87 *A **functional** J is a rule of correspondence that assigns to each function x in a certain class Ω a unique real number. Ω is called the **domain** of the functional, and the set of real numbers associated with the functions in Ω is called the **range** of the functional. (The domian of a functional is a class of functions; a functional is a "function of a function".)*

Definition 88 *J is a **linear functional** of x if and only if it satisfies the **principle of homogeneity***

$$J(\alpha x) = \alpha J(x)$$

*for all $x \in \Omega$ and for all real numbers α such that $\alpha x \in \Omega$, and the **principle of additivity***

$$J\left(x^{(1)} + x^{(2)}\right) = J\left(x^{(1)}\right) + J\left(x^{(2)}\right)$$

for all $x^{(1)}$, $x^{(2)}$, and $x^{(1)} + x^{(2)}$ in Ω.

Definition 89 *The **norm of a function** is a rule of correspondence that assigns to each function $x \in \Omega$, defined for $t \in [t_0, t_f]$, a real number. The **norm of** x, denoted by $\|x\|$, satisfies the following properties:*

1. *$\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x(t) = 0$ for all $t \in [t_0, t_f]$;*
2. *$\|\alpha x\| = |\alpha| \|x\|$ for all real numbers α;*
3. *$\left\|x^1 + x^2\right\| \leq \left\|x^1\right\| + \left\|x^2\right\|$.*

(to compare the closeness of two functions y and z that are defined for $t \in [t_0, t_f]$, let $x(t) = y(t) - z(t)$. if $\|x\|$ if zero/small/large, then two functions are identical/close/far-apart.)

(E.g. one acceptable norm for x can be defined as $\|x\| = \max_{t_0 \leq t \leq t_f} \{|x(t)|\}$)

Definition 90 *If x and $x + \delta x$ are functions for which the functional J is defined, then the **increment of** J, denoted by ΔJ, is*

$$\Delta J = \Delta J(x, \delta x) \triangleq J(x + \delta x) - J(x).$$

*Also, δx is called the **variation** of the function x.*

Remark 91 *1). δx does NOT mean $\delta \cdot x$. δ is not a scalar. Rather δx is a new admissible function near x, and the shape and property of δx and x can be different.*

2). $\delta x(t_f)$: the variation of $x(t_f)$: the change of x_{t_f} due to the change of function form of $x(\cdot)$, while keeping time t_f unchanged;

$\dot{x}(t_f)\,\delta t_f$: linear approximation of change of x_{t_f} due to the change in time t_f while keeping the function form $x(\cdot)$ unchanged;

δx_{t_f}: the linear approximation of the change of x_{t_f};

Hence, the relationship between those three terms are

$$\delta x_{t_f} = \delta x(t_f) + \dot{x}(t_f)\,\delta t_f \tag{4.1}$$

. (PS: In Friesz 2010, this relationship is writen as $dx(t_f) = \delta x(t_f) + \dot{x}(t_f)\,dt_f$.)

Because $\delta x_{t_f} = \delta x(t_f) + \dot{x}(t_f)\,\delta t_f$, $\delta x(t_f)$ and dt_f is not free. In particular, $\delta x(t_f)$ depends on dt_f. (e.g. Kirk 2004 page 135-136 has example of linear relationship between those two.) (e.g. think this way: if we fix the total change constant, then $\delta x(t_f)$ will depends on dt_f.) However, δx_{t_f} and dt_f will be independent, and, hence, out of the integral we often collect terms in the form of δx_{t_f} and dt_f in calculus of variations.(e.g. Kirk 2004 page 139)

3). δx is used when analyize the variation of x which is a function in integral, such as if $J = \int_{t_0}^{t_f} f(x(t))\,dt$, then $\delta J = \int_{t_0}^{t_f} f_x(x(t))\,\delta x dt$. However, when x is a function of a point variable u, we will use du to represent the change of this point variable, such as if $J = g(x(u))$, then $\delta J = g_x(x(u))\,\frac{\partial x(u)}{\partial u} du$. Remember, we only express δx to represent the function form changes only when x is in an integral. (of course, δx may be transfered out by some methods such as integrate by part, e.g. Euler's equation or in the analysis below of "Continuous Time Optimal Control (From Friesz 2008)" section. However, after transfer δx out of the integral into a point function, we need to use $dx(u) = \delta x(u) + \dot{x}(u)\,du$ to express $\delta x(u)$ as $dx(u) - \dot{x}(u)\,du$, where $\dot{x}(u)\,du$ and $dx(u)$ are independent.)

Definition 92 *The **increment of a functional** can be written as*

$$\Delta J(x, \delta x) = \delta J(x, \delta x) + g(x, \delta x) \cdot \|\delta x\|,$$

where δJ is linear in δx. If

$$\lim_{\|\delta x\| \to 0} \{g(x, \delta x)\} = 0,$$

*then J is said to be **differentiable** on x and δJ is the **variation of** J evaluated for the function x.*

(The variation of J, δJ, is the linear approximation *to the difference in the functional J caused by two comparion curves, ΔJ.)*

Theorem 93 *Chain rule for the calculus of variations (T. Friesz's note: chapter 3): Let $J(X) = \int_{t_0}^{t_f} g\left(X(t), \dot{X}(t), t\right) dt$, then the variation of this functional will obeys*

$$\delta J(X) = \sum_{i=1}^{n} \int_{t_0}^{t_f} \left\{ \left[\frac{\partial}{\partial x_i} g\left(X(t), \dot{X}(t), t\right) \right] \delta x_i + \left[\frac{\partial}{\partial \dot{x}_i} g\left(X(t), \dot{X}(t), t\right) \right] \delta \dot{x}_i \right\} dt.$$

Definition 94 *A functional J with domain Ω has a relative extremum at x^* if there is an $\varepsilon > 0$ such that for all functions x in Ω which satisfy $\|x - x^*\| < \varepsilon$ the increment of J has the same sign. If*

$$\Delta J = J(x) - J(x^*) \geq 0,$$

*$J(x^*)$ is a **relative minimum**; if*

$$\Delta J = J(x) - J(x^*) \leq 0,$$

*$J(x^*)$ is a **relative maximum**.*

*If the condition is satisfied for arbitrarily large ε, then $J(x^*)$ is a **global**, or **absolute**, minimum/maximum. x^* is called an **extremal**, and $J(x^*)$ is referred to as an **extremum**.*

4.1.2 Necessary conditions in the Calculus of Variations

Theorem 95 *(The fundamental theorem of the calculus of variations) Let x be a vector function of t in the class Ω, and $J(x)$ be a differentiable functional of x. Assume that the functions in Ω are not constrained by any boundaries. If x^* is an extremal, the variation of J must vanish on x^*; that is,*

$$\delta J(x^*, \delta x) = 0 \text{ for all admissible } \delta x.^1$$

Lemma 96 *Vanishing integral property (T. Friesz's note: chapter 3): If $\psi \in C^0[a,b]$, continuous function in $[a,b]$, and if, for all $\phi \in C^1[a,b]$, function has first order derivative in $[a,b]$, such that $\phi(a) = \phi(b) = 0$, we have*

$$\int_a^b \psi(t) \frac{d\phi(t)}{dt} dt = 0,$$

then $\psi(t) = c$, a constant, for all $t \in [a,b] \in R^1$.

Lemma 97 *Fundamental lemma (T. Friesz's note: chapter 3): If $g \in C^0[a,b]$ and $h \in C^0[a,b]$, and if, for all $\phi \in C^1[a,b]$ such that $\phi(a) = \phi(b) = 0$, we have*

$$\int_a^b \left[g(t)\phi(t) + h(t)\dot{\phi}(t) \right] dt = 0,$$

then

$$g(t) - \frac{dh(t)}{dt} = 0, \text{ for } t \in [a,b].$$

Lemma 98 *(The fundamental lemma of the calculus of variations) If a function h is coninuous and*

$$\int_{t_0}^{t_f} h(t)\,\delta x(t)\,dt = 0$$

for every function δx that is continuous in the interval $[t_0, t_f]$, then h must be zero everywhere in the interval $[t_0, t_f]$.

Remark 99 *By adopting the Fundamental lemma, we have*

$$\int_{t_0}^{t_f} \left\{ \left[\frac{\partial}{\partial x} g(x(t), \dot{x}(t), t) \right] \delta x + \left[\frac{\partial}{\partial \dot{x}} g(x(t), \dot{x}(t), t) \right] \delta \dot{x} \right\} dt = 0 \Rightarrow$$
$$\frac{\partial}{\partial x} g(x(t), \dot{x}(t), t) - \frac{d}{dt} \left[\frac{\partial}{\partial \dot{x}} g(x(t), \dot{x}(t), t) \right] = 0$$

*where the last equation is a form of **Euler equation** (**Euler-Lagrange equation**). (When both boundary times and boundary values are fixed, then $\delta J(X) = 0$ is equivalent to the Euler equation.)*

Remark 100 *The **second form of the Euler euqation** (the second formEuler-Lagrange equation): By chain rule, we have*

$$\frac{dg}{dt} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial \dot{x}}\frac{d\dot{x}}{dt}$$
$$\frac{d}{dt}\left(\dot{x}\frac{\partial g}{\partial \dot{x}} \right) = \frac{\partial g}{\partial \dot{x}}\frac{d\dot{x}}{dt} + \dot{x}\frac{d}{dt}\left(\frac{\partial g}{\partial \dot{x}} \right)$$

. By combining those two equations, we have

$$\dot{x}\left[\frac{\partial g}{\partial x} - \frac{d}{dt}\left(\frac{\partial g}{\partial \dot{x}} \right) \right] + \frac{d}{dt}\left(\dot{x}\frac{\partial g}{\partial \dot{x}} \right) - \frac{dg}{dt} + \frac{\partial g}{\partial t} = 0.$$

Therefore, if x is a solution of the Euler equation, then

$$\frac{d}{dt}\left(\dot{x}\frac{\partial g}{\partial \dot{x}} - g \right) + \frac{\partial g}{\partial t} = 0,$$

which is known as the second form of the Euler equation.

Remark 101 *If $\frac{\partial g}{\partial t}=0$, then, form the second form of the Euler equation, it is immediate that* $\frac{d}{dt}\left(\dot{x}\frac{\partial g}{\partial \dot{x}}-g\right)=0$. *In other words,* $\dot{x}\frac{\partial g}{\partial \dot{x}}-g=c_0$, *a constant, and this expression is* ***Beltrami's identity****.*

Lemma 102 *Nonegativity of a functional (T. Friesz's note: chapter 3): Let $g\in C^0[a,b]$, $h\in C^0[a,b]$, and $\phi\in C^1[a,b]$; suppose also that $\phi(a)=\phi(b)=0$. A necessary condition for the functional*

$$F=\int_a^b\left[g(t)\{\phi(t)\}^2+h(t)\left\{\dot{\phi}(t)\right\}^2\right]dt$$

to be nonnegative for all ϕ is that

$$h(t)\geq 0,\ \text{for } \forall t\in[a,b].$$

Theorem 103 *(T. Friesz's note: chapter 3) A necessary condtion for x to minimize $J(x)$ in the problem defined by*

$$\left\{\begin{array}{c}\min\left\{J(x)=\int_{t_0}^{t_f}g(x(t),\dot{x}(t),t)\,dt\right\}\\ s.t. t_0,\ t_f,\ x(t_0)=x_0,\ \text{and } x(t_f)=x_f\ \text{are fixed}\end{array}\right\}$$

is that

1). Legendre's condition: $\dot{x}(t)$ is continuouse,

$$\frac{\partial^2 g(x(t),\dot{x}(t),t)}{\partial(\dot{x}(t))^2}\geq 0,\ \text{for all } t\in[t_0,t_f].$$

or 2). Euler's equation: $\dot{x}(t)$ is continuouse,

$$\frac{\partial}{\partial x}g(x(t),\dot{x}(t),t)-\frac{d}{dt}\left[\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]=0,\ \text{for all } t\in[t_0,t_f].$$

or 3). The second form of Euler's equation: $\dot{x}(t)$ is continuouse,

$$\frac{d}{dt}\left(\dot{x}\frac{\partial g}{\partial \dot{x}}-g\right)+\frac{\partial g}{\partial t}=0,\ \text{for all } t\in[t_0,t_f].$$

or 4). Weierstrass condition: $\dot{x}(t)$ is continuouse, and, for all admissible y and all $t\in[t_0,t_f]$,

$$g(x(t),\dot{y}(t),t)-g(x(t),\dot{x}(t),t)-\left(\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right)(\dot{y}(t)-\dot{x}(t))\geq 0$$

or 5) Weierstrass-Erdman conditions (corner conditions): when $\dot{x}(t)$ has a jump discontinuity at time t_1, then, for all $t\in[t_0,t_f]$, we must have

$$\begin{aligned}\left[\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_1^-}&=\left[\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_1^+}\\ \left[g(x(t),\dot{x}(t),t)-\dot{x}\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_1^-}&=\left[g(x(t),\dot{x}(t),t)-\dot{x}\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_1^+}\end{aligned}$$

Remark 104 *When the boundary conditions are free, e.g. $x(t_0)$ or/both $x(t_f)$, when, besides the Euler-equation, we need include* ***free endpoint conditions (naural boundary conditions)*** $\left[\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_f}=0$ *or/and* $\left[\frac{\partial}{\partial \dot{x}}g(x(t),\dot{x}(t),t)\right]_{t=t_0}=0$.

4.1.3 Sufficient conditions in the Calculus of Variations

Theorem 105 *(T. Friesz's note: chapter 3) A sufficient condtion for x to minimize $J(x)$ in the problem defined by*

$$\left\{ \begin{array}{c} \min \left\{ J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt \right\} \\ s.t. t_0,\ t_f,\ x(t_0) = x_0, \text{ and } x(t_f) = x_f \text{ are fixed} \end{array} \right\}$$

is that

1. $g(x(t), \dot{x}(t), t)$ is convex with respect to $(x(t), \dot{x}(t))$ for all $t \in [t_0, t_f]$;

and 2. $x(t)$ is an admissible function, and satisfies the Euler equation every where except possibly at points of jump discontinuity of its time derivative wherer it satisfies the Weierstrass-Erdman conditions;

Then $x(t)$ is a solution to the problem.

Remark 106 *When the boundary conditions are free, e.g. $x(t_0)$ or/both $x(t_f)$, when, besides the Euler-equation, we need include* ***free endpoint conditions (naural boundary conditions)*** $\left[\frac{\partial}{\partial \dot{x}} g(x(t), \dot{x}(t), t)\right]_{t=t_f} = 0$ *or/and* $\left[\frac{\partial}{\partial \dot{x}} g(x(t), \dot{x}(t), t)\right]_{t=t_0} = 0$.

4.1.4 Solutions for unconstraint optimization problem

Solution 107 *(Single function) if $J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt$, then a necessary condition for optimality will be*

$$\begin{aligned} & 0 = \delta J(x^*, \delta x) \\ & \Leftrightarrow \\ & 0 = \left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right] \delta x(t)\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \left(\frac{\partial g}{\partial x} g(x(t), \dot{x}(t), t) - \frac{d}{dt}\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right]\right) \delta x(t)\, dt. \end{aligned}$$

where

$$\frac{\partial g}{\partial x} g(x(t), \dot{x}(t), t) - \frac{d}{dt}\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right] = 0$$

is oftern called the ***Euler equation****, which always satisified under optimality. Other* ***boundary conditions*** *includes:*

1. when t_0, t_f, $x(t_0) = x_0$, and $x(t_f) = x_f$ are fixed: $\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right] \delta x(t)\big|_{t_0}^{t_f} = 0$ automatically, and, besides the Euler equation, we need boundary conditions: $x(t_0) = x_0$ and $x(t_f) = x_f$;

2. when t_0, t_f, and $x(t_0) = x_0$ are fixed, and $x(t_f)$ is free: $\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right] \delta x(t)\big|_{t_0}^{t_f} = \left[\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}}\right] \delta x(t_f)$, and, besides the Euler equation, we need boundary conditions: $x(t_0) = x_0$ and $\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}} = 0$;

3. when t_0, $x(t_0) = x_0$, and $x(t_f) = x_f$ are fixed, and t_f is free:

$$\delta J(x^*, \delta x) = \left\{ \begin{array}{c} \left\{ g(x(t_f), \dot{x}(t_f), t_f) - \left[\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}}\right] \dot{x}(t_f) \right\} \delta t_f \\ + \int_{t_0}^{t_f} \left(\frac{\partial g}{\partial x} g(x(t), \dot{x}(t), t) - \frac{d}{dt}\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right]\right) \delta x(t)\, dt \end{array} \right\} = 0$$

, so, besides the Euler equation, the boundary conditions will be: $x(t_0) = x_0$, $x(t_f) = x_f$, and $g(x(t_f), \dot{x}(t_f), t_f) - \left[\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}}\right] \dot{x}(t_f) = 0$.

4. when t_0 and $x(t_0) = x_0$ are fixed, and t_f and $x(t_f)$ are free:

$$\delta J(x^*, \delta x) = \left\{ \begin{array}{c} \left[\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}}\right] \delta x_f \\ + \left\{ g(x(t_f), \dot{x}(t_f), t_f) - \left[\frac{\partial g(x(t_f), \dot{x}(t_f), t_f)}{\partial \dot{x}}\right] \dot{x}(t_f) \right\} \delta t_f \\ + \int_{t_0}^{t_f} \left(\frac{\partial g}{\partial x} g(x(t), \dot{x}(t), t) - \frac{d}{dt}\left[\frac{\partial g(x(t), \dot{x}(t), t)}{\partial \dot{x}}\right]\right) \delta x(t)\, dt \end{array} \right\} = 0$$

, so, besides the Euler equation, the boundary conditions will be: $x(t_0) = x_0$, $g(x(t_f), \dot{x}(t_f), t_f) - \left[\frac{\partial g(x(t_f),\dot{x}(t_f),t_f)}{\partial \dot{x}}\right]\dot{x}(t_f) = 0$, *and* $\frac{\partial g(x(t_f),\dot{x}(t_f),t_f)}{\partial \dot{x}} = 0$. *The last condition, the transversality condition, implicitly assume* t_f *and* $x(t_f)$ *is unrelated. However if* t_f *and* $x(t_f)$ *is related by* $x(t_f) = \theta(t_f)$, *then the transversality condition is* $\frac{\partial g(x(t_f),\dot{x}(t_f),t_f)}{\partial \dot{x}}\left[\frac{d\theta}{dt}(t_f) - \dot{x}(t_f)\right] + g(x(t_f), \dot{x}(t_f), t_f) = 0$.

Solution 108 *(Multiple independent functions: there is no constraint relationship among functions.) When functionals contain several independent functions and their first derivatives*

$$J(x_1, ..., x_n) = \int_{t_0}^{t_f} g(x_1(t), ..., x_n(t), \dot{x}_1(t), ..., \dot{x}_n(t), t)\, dt,$$

the optimality conditions will include the Euler equation

$$\begin{aligned}&\frac{\partial}{\partial x_i} g(x_1(t), ..., x_n(t), \dot{x}_1(t), ..., \dot{x}_n(t), t)\\ &- \frac{d}{dt}\left[\frac{\partial}{\partial \dot{x}_i} g(x_1(t), ..., x_n(t), \dot{x}_1(t), ..., \dot{x}_n(t), t)\right]\\ &= 0 \text{ for all } t \in [t_0, t_f] \text{ and } i = 1, ..., n.\end{aligned}$$

and boundary conditions in the following table:

Remark 109 *If* $\dot{x}(t)$ *can be discountinuous, then we need to include the Weierstrass-Erdmann corner conditions, besides Euler equation and boundary conditions.*

4.1.5 Solutions for constraint optimization problem

Assume that the admissible curves are smooth.

Point constraints.

$$\left\{\begin{array}{c}\min\left\{J(W) = \int_{t_0}^{t_f} g\left(W(t), \dot{W}(t), t\right) dt\right\}\\ s.t.\ F(W(t), t) = 0, \text{ for } t \in [t_0, t_f]\end{array}\right\}$$

where W is an vector of $(m+n)$ functions, and $F(W(t), t)$ is an vector of n functionals.

In order to solve this problem, we need to define a Lagrange functional:

$$\begin{aligned}J_a(W, P) &= \int_{t_0}^{t_f} g_a\left(W(t), \dot{W}(t), P(t), t\right) dt\\ &\doteq \int_{t_0}^{t_f} \left\{g\left(W(t), \dot{W}(t), t\right) + P^T(t)\left[F(W(t), t)\right]\right\} dt\end{aligned}$$

The variation of this functional J_a can be writen as

$$\begin{aligned}&\delta J_a(W, \delta W, P, \delta P)\\ &= \int_{t_0}^{t_f}\left\{\begin{array}{c}\left[\frac{\partial}{\partial W} g_a^T\left(W(t), \dot{W}(t), P(t), t\right) - \frac{d}{dt}\left[\frac{\partial}{\partial \dot{W}} g_a^T\left(W(t), \dot{W}(t), P(t), t\right)\right]\right]\delta W(t)\\ +\left[F^T(W(t), t)\right]\delta P(t)\end{array}\right\} dt\\ &= \int_{t_0}^{t_f}\left\{\begin{array}{c}\left[\begin{array}{c}\frac{\partial}{\partial W} g^T\left(W(t), \dot{W}(t), t\right) + P^T(t)\left[\frac{\partial}{\partial W} F\left(W(t), \dot{W}(t), t\right)\right]\\ -\frac{d}{dt}\left[\frac{\partial}{\partial \dot{W}} g^T\left(W(t), \dot{W}(t), t\right) + P^T(t)\left[\frac{\partial}{\partial \dot{W}} F\left(W(t), \dot{W}(t), t\right)\right]\right]\end{array}\right]\delta W(t)\\ +\left[F^T(W(t), t)\right]\delta P(t)\end{array}\right\} dt\end{aligned}$$

Table 4-1 DETERMINATION OF BOUNDARY-VALUE RELATIONSHIPS

Problem description	*Substitution*	*Boundary conditions*	*Remarks*
1. $\mathbf{x}(t_f)$, t_f both specified (*Problem 1*)	$\delta\mathbf{x}_f = \delta\mathbf{x}(t_f) = \mathbf{0}$ $\delta t_f = 0$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \mathbf{x}_f$	$2n$ equations to determine $2n$ constants of integration
2. $\mathbf{x}(t_f)$ free; t_f specified (*Problem 2*)	$\delta\mathbf{x}_f = \delta\mathbf{x}(t_f)$ $\delta t_f = 0$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = \mathbf{0}$	$2n$ equations to determine $2n$ constants of integration
3. t_f free; $\mathbf{x}(t_f)$ specified (*Problem 3*)	$\delta\mathbf{x}_f = \mathbf{0}$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \mathbf{x}_f$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)$ $- \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \dot{\mathbf{x}}^*(t_f) = 0$	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f
4. t_f, $\mathbf{x}(t_f)$ free and independent (*Problem 4*)	—	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = \mathbf{0}$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = 0$	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f
5. t_f, $\mathbf{x}(t_f)$ free but related by $\mathbf{x}(t_f) = \boldsymbol{\theta}(t_f)$ (*Problem 4*)	$\delta\mathbf{x}_f = \frac{d\boldsymbol{\theta}}{dt}(t_f)\,\delta t_f$†	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \boldsymbol{\theta}(t_f)$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)$ $+ \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \left[\frac{d\boldsymbol{\theta}}{dt}(t_f) - \dot{\mathbf{x}}^*(t_f)\right] = 0$†	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f

† $\frac{d\boldsymbol{\theta}}{dt}$ denotes the $n \times 1$ column vector $\left[\frac{d\theta_1}{dt} \quad \frac{d\theta_2}{dt} \quad \cdots \quad \frac{d\theta_n}{dt}\right]^T$.

Hence, the necessary condition for optimality will be:

1.Point Constraints (n equations): $F\left(W\left(t\right),\dot{W}\left(t\right),t\right)=0$

2.Euler's Equation ($m+n$ equations): $\frac{\partial}{\partial W}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial\dot{W}}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)\right]=0$

for all $t\in[t_0,t_f]$.

Differential Equation Constraints.

$$\left\{\begin{array}{l}\min\left\{J\left(W\right)=\int_{t_0}^{t_f}g\left(W\left(t\right),\dot{W}\left(t\right),t\right)dt\right\}\\ s.t.\ F\left(W\left(t\right),\dot{W}\left(t\right),t\right)=0,\text{ for }t\in[t_0,t_f]\end{array}\right\}$$

where W is an vector of $(m+n)$ functions, and $F\left(W\left(t\right),\dot{W}\left(t\right),t\right)$ is an vector of n functionals.

In order to solve this problem, we need to define a Lagrange functional:

$$\begin{aligned}J_a\left(W,P\right)&=\int_{t_0}^{t_f}g_a\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)dt\\ &\doteq\int_{t_0}^{t_f}\left\{g\left(W\left(t\right),\dot{W}\left(t\right),t\right)+P^T\left(t\right)\left[F\left(W\left(t\right),\dot{W}\left(t\right),t\right)\right]\right\}dt\end{aligned}$$

The variation of this functional J_a can be writen as

$$\begin{aligned}&\delta J_a\left(W,\delta W,P,\delta P\right)\\ &=\int_{t_0}^{t_f}\left\{\begin{array}{c}\left[\frac{\partial}{\partial W}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial\dot{W}}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)\right]\right]\delta W\left(t\right)\\ -\left[F^T\left(W\left(t\right),t\right)\right]\delta P\left(t\right)\end{array}\right\}dt\\ &=\int_{t_0}^{t_f}\left\{\begin{array}{c}\left[\frac{\partial}{\partial W}g^T\left(W\left(t\right),\dot{W}\left(t\right),t\right)+P^T\left(t\right)\left[\frac{\partial}{\partial W}F\left(W\left(t\right),t\right)\right]-\frac{d}{dt}\left[\frac{\partial}{\partial\dot{W}}g^T\left(W\left(t\right),\dot{W}\left(t\right),t\right)\right]\right]\delta W\left(t\right)\\ -\left[F^T\left(W\left(t\right),t\right)\right]\delta P\left(t\right)\end{array}\right\}dt\end{aligned}$$

Hence, the necessary condition for optimality will be:

1.Point Constraints (n equations): $F\left(W\left(t\right),t\right)=0$

2.Euler's Equation ($m+n$ equations): $\frac{\partial}{\partial W}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial\dot{W}}g_a^T\left(W\left(t\right),\dot{W}\left(t\right),P\left(t\right),t\right)\right]=0$

for all $t\in[t_0,t_f]$.

Isoperimetric Constraints.

$$\left\{\begin{array}{c}\min\left\{J\left(W\right)=\int_{t_0}^{t_f}g\left(W\left(t\right),\dot{W}\left(t\right),t\right)dt\right\}\\ s.t.\ \int_{t_0}^{t_f}e_i\left(W\left(t\right),\dot{W}\left(t\right),t\right)dt=c_i,\text{ for }t\in[t_0,t_f]\ \ (i=1,...,r)\end{array}\right\}$$

where W is an vector of $(m+n)$ functions.

To begin with, define new variables

$$z_i\left(t\right)=\int_{t_0}^{t}e_i\left(W\left(t\right),\dot{W}\left(t\right),t\right)dt,\ i=1,...,r,$$

and with boundary condition

$$z_i\left(t_f\right)=c_i,\ i=1,...,r.$$

Hence, constraints can be rewriten as

$$e_i\left(W(t),\dot{W}(t),t\right)-\dot{z}_i(t)=0,\text{ for }t\in[t_0,t_f]\ \ (i=1,...,r)$$
$$z_i(t_f)-c_i=0,\text{ for }t\in[t_0,t_f]\ \ (i=1,...,r)$$

Define $E\left(W(t),\dot{W}(t),t\right)$ and $\dot{Z}(t)$ to be an vector of (r) functionals of $e_i\left(W(t),\dot{W}(t),t\right)$ and $\dot{z}_i(t)$. In order to solve this problem, we need to define a Lagrange functional:

$$\begin{aligned}J_a(W,P)&=\int_{t_0}^{t_f}g_a\left(W(t),\dot{W}(t),P(t),\dot{Z}(t),t\right)dt\\&\doteq\int_{t_0}^{t_f}\left\{g\left(W(t),\dot{W}(t),t\right)+P^T(t)\left[E\left(W(t),\dot{W}(t),t\right)-\dot{Z}(t)\right]\right\}dt\end{aligned}$$

Hence, the necessary condition for optimality will be:

1.Variable Constraints ($r+r$ equations): $\left\{\begin{array}{c}E\left(W(t),\dot{W}(t),t\right)-\dot{Z}(t)=0,\text{ for }t\in[t_0,t_f]\\Z(t_f)-C=0\end{array}\right\}$

2.Euler's Equation ($m+n+r$ equations): $\left\{\begin{array}{c}\frac{\partial}{\partial W}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial \dot{W}}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)\right]=0\\\frac{\partial}{\partial Z}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial \dot{Z}}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)\right]=0\end{array}\right\}$

for all $t\in[t_0,t_f]$.

In which

$$\frac{\partial}{\partial Z}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)-\frac{d}{dt}\left[\frac{\partial}{\partial \dot{Z}}g_a^T\left(W(t),\dot{W}(t),P(t),t\right)\right]=0$$
$$\Leftrightarrow$$
$$\dot{P}(t)=0$$

4.2 Continuous Time Optimal Control: analysis of variation approach (Pontryagin minimum principle) (From Friesz 2008)

Consider the following canonical form of the continuous time optimal control problem

$$\left\{\begin{array}{c}\min\left\{J(x(t),u(t))=K(x(t_f),t_f)+\int_{t_0}^{t_f}f_0(x(t),u(t),t)\,dt\right\}\\s.t.\\\text{State dynamics: }\dot{x}(t)=f(x(t),u(t),t)\\\text{Initial conditions: }x(t_0)=x^0\in R^m\\\text{Terminal conditions: }\Psi(x(t_f),t_f)=0\\\text{Control constraints: }u(t)\in\Omega\end{array}\right\}OCP$$

where $x(t)=(x_1(t),...,x_n(t))^T$, $u(t)=(u_1(t),...,u_m(t))^T$, $f_0:R^n\times R^m\to R^1$, $f:R^n\times R^m\times R^1\to R^n$, $K:R^n\times R^1\to R^1$, and $\Psi:R^n\times R^1\to R^1$. (*the initial time, terminal time, and their corresponding values may be unknowns.*) Also, we assume this OCP problem is regular.

Definition 110 *Regularity for OCP problem: We shall say optimal control problem OCP defined above is* ***regular*** *provided $f_0(\cdot)$, $\Psi(\cdot)$, $K(\cdot)$, and $f(\cdot)$ are everywhere once continuously differentiable with respect to their arguments.*

Definition 111 *Admissible control trajectory: we say that the control trajectory $u(t)$ is admisible relative to OCP if it is piecewise continuous for all time $t\in[t_0,t_f]$ and $u\in\Omega$.*

First need to obtain the Lagrangean by endogenize those constrains:

$$L = K\left(x\left(t_f\right), t_f\right) + v^T \Psi\left(x\left(t_f\right), t_f\right) + \int_{t_0}^{t_f} \left\{f_0\left(x, u, t\right) + \lambda^T\left[f\left(x, u, t\right) - \dot{x}\right]\right\} dt$$

Using the calculus of variations, the variation of the Lagrangean L will be

$$\begin{aligned}
\delta L = & \Phi_t\left(t_f\right) dt_f + \Phi_x\left(t_f\right) dx\left(t_f\right) && \text{Variations of the constraint} \\
& + f_0\left(t_f\right) dt_f - f_0\left(t_0\right) dt_0 && \text{Variations of integral part w.r.t. time} \\
& + \int_{t_0}^{t_f} \left[H_x \delta x + H_u \delta u - \lambda^T \delta \dot{x}\right] dt && \text{Variations of integral part w.r.t. functions}
\end{aligned}$$

where

$$\begin{aligned}
H\left(x, u, \lambda, t\right) &\equiv f_0\left(x, u, t\right) + \lambda^T f\left(x, u, t\right) \\
\Phi\left(t_f\right) &= K\left(x\left(t_f\right), t_f\right) + v^T \Psi\left(x\left(t_f\right), t_f\right) \\
f_0\left(t_0\right) &= f_0\left(x\left(t_0\right), u\left(t_0\right), t_0\right) \\
f_0\left(t_f\right) &= f_0\left(x\left(t_f\right), u\left(t_f\right), t_f\right)
\end{aligned}$$

In δL, the term $\int_{t_0}^{t_f} \left(-\lambda^T \delta \dot{x}\right) dt$ can be transformed into the following by using integrating by parts and equation (4.1):

$$\begin{aligned}
\int_{t_0}^{t_f} \left(-\lambda^T \delta \dot{x}\right) dt &= \lambda^T\left(t_0\right) \delta x\left(t_0\right) - \lambda^T\left(t_f\right) \delta x\left(t_f\right) + \int_{t_0}^{t_f} \left(\frac{d\lambda^T}{dt} \delta x\right) dt \\
&= \lambda^T\left(t_0\right)\left(dx\left(t_0\right) - \dot{x}\left(t_0\right) dt_0\right) \\
&\quad - \lambda^T\left(t_f\right)\left(dx\left(t_f\right) - \dot{x}\left(t_f\right) dt_f\right) \\
&\quad + \int_{t_0}^{t_f} \left(\frac{d\lambda^T}{dt} \delta x\right) dt
\end{aligned}$$

So,

$$\begin{aligned}
\delta L = & \left[\Phi_t\left(t_f\right) + f_0\left(t_f\right) + \lambda^T\left(t_f\right) \dot{x}\left(t_f\right)\right] dt_f && \text{Variations of the terminal time} \\
& + \left[\Phi_x^T\left(t_f\right) - \lambda^T\left(t_f\right)\right] dx\left(t_f\right) && \text{Variations of the terminal value} \\
& - \left[f_0\left(t_0\right) + \lambda^T\left(t_0\right) \dot{x}\left(t_0\right)\right] dt_0 && \text{Variations of the initial time} \\
& + \lambda^T\left(t_0\right) dx\left(t_0\right) && \text{Variations of the initial value} \\
& + \int_{t_0}^{t_f} \left[\left[H_x + \dot{\lambda}\right] \delta x\right] dt && \text{Variations of function } x \\
& + \int_{t_0}^{t_f} \left[\left[H_u\right] \delta u\right] dt && \text{Variations of function } u
\end{aligned}$$

Remark 112 *The above analysis of δL does not assume fixed boundary time and value. However, if the boundary times t_0 or/and t_f are fixed, then dt_0 or/and dt_f will be zero. If the boundary values are fixed $x\left(t_0\right)$ or/and $x\left(t_f\right)$ are fixed, then $dx\left(t_0\right)$ or/and $dx\left(t_f\right)$ will be zero.*

By the fundamental theorem of the calculus of variations, the neccessary condition for optimality is that δL need to vanish for admissible x and u, which requires coefficients of each variations to be zero. Hence, we have the following necessary conditions for optimality:

Theorem 113 *When there is no explicit constraint on control variable u (e.g. $\Omega = R^m$), the **necessary***

conditions *for optimality are*

Necessary Conditions	*Corresponding terms*	*Names in Literature*
$\Phi_t(t_f) + f_0(t_f) + \lambda^T(t_f)\dot{x}(t_f) = 0$	*Free terminal time*	*Terminal time condition 2*
$\Phi_x^T(t_f) - \lambda^T(t_f) = 0$	*Free terminal value*	*Transversality condition*
$f_0(t_0) + \lambda^T(t_0)\dot{x}(t_0) = 0$	*Free initial time*	*Initial time condition 1*
$\lambda^T(t_0)\,dx(t_0) = 0$	*Free initial value*	*Initial time condition 2*
$H_x + \dot{\lambda} = 0$	*Function x*	*Adjoint equations*
$H_u = 0$	*Function u*	*Minimum principle*
$H_{uu} \geq 0$		*Legendre-Clebsch condition*

, and together with constraints:

Constraints of the original problem	*Constraints' names*
$\dot{x}(t) = f(x(t), u(t), t)$	*State dynamics*
$x(t_0) = x^0$	*Initial conditions*
$\Psi(x(t_f), t_f) = 0$	*Terminal conditions*

Remark 114 *1). When the control variable u is unconstraint (e.g. $\Omega = R^m$),* ***Legendre-Clebsch condition****, $H_{uu} \geq 0$, must be satisfied as well. Intuitively, this means that the second derivative of local minimue point must be positive.*

2). If the boundary times t_0 or/and t_f are fixed, then we do not need "Initial time condition 1" or/and "Terminal time condition 2".

3). If the boundary values $x(t_0)$ or/and $x(t_f)$ are fixed, then we do not need "Initial time condition 2" or/and "Transversality condition".

Theorem 115 *When there is explicit and pure control constraints on control variable u such that $u \in \Omega$ and Ω is convex, the* ***necessary conditions*** *for optimality are*

Necessary Conditions	*Corresponding terms*	*Names in Literature*
$\Phi_t(t_f) + f_0(t_f) + \lambda^T(t_f)\dot{x}(t_f) = 0$	*Free terminal time*	*Terminal time condition 2*
$\Phi_x^T(t_f) - \lambda^T(t_f) = 0$	*Free terminal value*	*Transversality condition*
$f_0(t_0) + \lambda^T(t_0)\dot{x}(t_0) = 0$	*Free initial time*	*Initial time condition 1*
$\lambda^T(t_0)\,dx(t_0) = 0$	*Free initial value*	*Initial time condition 2*
$H_x + \dot{\lambda} = 0$	*Function x*	*Adjoint equations*
$H_u(u - u^*) \geq 0$, *for* $\forall u \in \Omega$	*Function u*	*Minimum principle / Variational inequality*

, and together with constraints:

Constraints of the original problem	*Constraints' names*
$\dot{x}(t) = f(x(t), u(t), t)$	*State dynamics*
$x(t_0) = x^0$	*Initial conditions*
$\Psi(x(t_f), t_f) = 0$	*Terminal conditions*
$u(t) \in \Omega$	*Control constraints*

Remark 116 *1). If the boundary times t_0 or/and t_f are fixed, then we do not need "Initial time condition 1" or/and "Terminal time condition 2".*

2). If the boundary values $x(t_0)$ or/and $x(t_f)$ are fixed, then we do not need "Initial time condition 2" or/and "Transversality condition".

3). From dynamic programming approach, $\lambda(t) = V_x(x,t)|_{x=x^*(t)}$, where $V(x,t)$ is the value function starting with state x and from time t onward. Hence, $\lambda(t)$ can be interpreted as per unit change in the objective function for a small change in $x^*(t)$.

4). From dynamic programming approach, the Hamilton-Jacobi-Bellman equation,HJB equation, is $0 = \min_{u(s)\in\Omega(s)} \{H(x,u,V_x,t) + V_t\}$, which can be restated as $H(x^*(t), u^*(t), \lambda(t), t) \leq H(x^*(t), u, \lambda(t), t)$ for all $u \in \Omega(t)$, which is equivalent to the variational inequality we defined above.

Theorem 117 *Restricted Mangasarian* ***sufficiency*** *theorem: Suppose 1). there is no terminal time conditions (no* $\Psi(x(t_f), t_f) = 0$*); 2).* $K(x(t_f), t_f) = 0$*; 3).* t_0*,* t_f*, and* x_0 *are fixed; 4). The Hamiltonian* H *is jointly convex in* x *and* u *for all admissible solutions; 5). the set of feasible controls* Ω *is convex; 6). Regularity condition is satisfied. Then any solution of the continuous-time optimal control necessary conditions is a global minimum.*

Remark 118 *This sufficient condition can be extended to include terminal condition and non-trivial salvage function.*

Theorem 119 *Arrow Sufficiency conditions (Sethi 2005 & Friesz 2010): Let* $u^*(t)$*, and the corresponding* $x^*(t)$ *and* $\lambda^*(t)$ *satisfy the necessary conditions, for all* $t \in [0, T]$*. Then* u^* *is an optimal control if* $H^0(x, \lambda(t), t) = \min_{u(s)\in\Omega(s)} H(x, u, \lambda(t), t)$ *is convex in* x *for each* t *and* $K(x(t_f), t_f)$ *is convex in* x*. (?if there is no Terminal conditions?)*

4.3 Continuous Time Optimal Control: Dynamic programming approach (HJB equation)

The Pontryagin minimum principle approach gives necessary condition for optimality, but the dynamic programming approach gives a sufficient (and often necessary) condition for optimality. However, in the dynamic programming approach, we need to solve HJB equation, which is partial differential equation and is very difficulty to solve. While, in Pontryagin minimum principle approach, we only need to solve ordinary differential equations, which are comparitively simple than partial differential equaiton.

Consider the following **free-end-point problems** (terminal values of the state variables are not constrained, but they assume the boundary times $t_0 = 0$ and $t_f = T$ are fixed. Also, assume that the initial boundary value is fixed by fixing x_0.):

$$\left\{ \begin{array}{c} \max \left\{ J = \underbrace{\int_0^T F(x(t), u(t), t)dt}_{\text{Profit}} + \underbrace{S[x(T), T]}_{\text{Salvage Value}} \right\} \\ \text{subject to} \\ \text{State Dynamics: } \dot{x}(t) = f(x(t), u(t), t) \\ \text{Initial conditions: } x(0) = x_0 \\ \text{Admissible Control: } u(t) \in \Omega(t), \forall t \in [0, T] \end{array} \right\} \tag{(OCP1)}$$

where State equation is the combination of state dynamics and initial conditions; admissible control is called control constrains as well

Let

$$\left\{ \begin{array}{c} \text{Value Function: } V(x, t) = \max_{u(s)\in\Omega(s)} \left\{ J = \int_t^T F(x(s), u(s), s)ds + S[x(T), T] \right\} \\ \text{subject to} \\ \dot{x}(s) = f(x(s), u(s), s) \\ x(t) = x \end{array} \right\} \tag{4.2}$$

Assuming value function exist for every x and t, and also assuming value function is continuously differentiable function of its arguments, then

$$\begin{aligned}
V(x,t) &= \max_{u(s)\in\Omega(s)} \{F(x,u,t)dt + V(x+\dot{x}(t)dt, t+dt)\} \\
&\Leftrightarrow \\
V(x,t) &= \max_{u(s)\in\Omega(s)} \{F(x,u,t)dt + V(x,t) + V_x(x,t)\,\dot{x}(t)dt + V_t(x,t)\,dt\} \\
&\Leftrightarrow \\
0 &= \max_{u(s)\in\Omega(s)} \{F(x,u,t)dt + V_x(x,t)\,f(x(t),u(t),t)dt + V_t(x,t)\,dt\} \\
&\Leftrightarrow \\
-V_t(x,t) &= \max_{u(s)\in\Omega(s)} \{F(x,u,t) + V_x(x,t)\,f(x(t),u(t),t)\}
\end{aligned}$$

The last partial differential equation is called The Hamilton–Jacobi–Bellman (HJB) equation:

$$-V_t(x,t) = \max_{u(s)\in\Omega(s)} \{F(x,u,t) + V_x(x,t)f(x,u,t)\} \tag{4.3}$$

with boundary condition

$$V(x,T) = S(x,T) \tag{4.4}$$

Remark 120 *Sometimes in solving this HJB equation, we can transfer those partial differentical equations into ordinary equations by collection terms respect to quadratic, linear and constant in state variable x. E.g., if the optimal u can be choose easily so that the HJB equation can be stated as $f_0(x,t) + f_1(x,t)\,x + f_2(x,t)\,x^2 + f_3(x,t)\,x^3 + ... = 0$, then we must simultaniously have $f_0(x,t) = 0$, $f_1(x,t) = 0$, $f_2(x,t) = 0$, ... This is because the HJB equation must be satisified for any realized state x.*

4.3.1 Derive The Pontryagin minimum principle from HJB equation (Sethi 2005)

Define adjoint vector λ, shadow price, as the following:

$$\lambda(t) = V_x(x,t)|_{x=x^*(t)} \tag{4.5}$$

Hence, $\lambda(t)$ can be interpreted as per unit change in the objective function for a small change in $x^*(t)$. Also, define the so-called Hamiltonian

$$H(x,u,V_x,t) = F(x,u,s) + V_x(x,t)f(x,u,t) = F(x,u,s) + \lambda(t)f(x,u,t) \tag{4.6}$$

Then, put the adjoint vector into equation 4.3, and we have the Hamilton-Jacobi-Bellman equation, or simply the HJB equation.

$$\text{HJB equation: } 0 = \max_{u(s)\in\Omega(s)} \{H(x,u,V_x,t) + V_t\} \tag{4.7}$$

Next, we will show how to solve, or restate, the HJB equation to maximum principle:

First, by analyzing value function and Hamiltonian, we can draw a connection for adjoint vector:

$$\text{Adjoint equation:} \dot{\lambda} = -H_x \tag{4.8}$$

With the definition of adjoint vector, equation 4.5, and boundary condition, equation 4.4, we have the so called transversality condition:

$$\text{Transversality condition: } \lambda(T) = \frac{\partial S(x,T)}{\partial x}|_{x=x(T)} \tag{4.9}$$

Hence, Adjoint equation and Transversality condition together will determine the adjoint variables. Meanwhile, the state dynamics and initial condition together will determine the states.

$$\left\{ \begin{array}{c} \dot{x}(s) = H_\lambda,\ x(0) = x_0 \\ \dot{\lambda} = -H_x,\ \lambda(T) = \frac{\partial S(x,T)}{\partial x}|_{x=x(T)} \end{array} \right\} \tag{4.10}$$

From HJB equation, equation 4.7, it can be restated as

$$H(x^*(t), u^*(t), \lambda(t), t) \geq H(x^*(t), u, \lambda(t), t) \text{ for all } u \in \Omega(t)$$

Together with canonical adjoints, equation 4.10, the maximum principle can be stated as following:

$$\left\{ \begin{array}{c} \dot{x}(s) = H_\lambda,\ x(0) = x_0 \\ \dot{\lambda} = -H_x,\ \lambda(T) = \frac{\partial S(x,T)}{\partial x}|_{x=x(T)} \\ H(x^*(t), u^*(t), \lambda(t), t) \geq H(x^*(t), u, \lambda(t), t) \text{ for all } u \in \Omega(t) \end{array} \right\} \tag{4.11}$$

Theorem 121 *The necessary condition for optimal control problem, Problem (OCP1), is the optimal control $u^*(t)$ satisfies maximum principle, equation 4.11*

Theorem 122 *Sufficiency conditions: Let $u^*(t)$, and the corresponding $x^*(t)$ and $\lambda(t)$ satisfy the maximum principle, equation 4.11, for all $t \in [0, T]$. Then u^* is an optimal control if $H^0(x, \lambda(t), t) = \max_{u(s)\in\Omega(s)} H(x, u, \lambda(t), t)$ is concave in x for each t and $S(x, T)$ is concave in x.*

4.3.2 Stochastic optimal control

When the state variable is subject to stochasticity, we can still use the dynamic programming approach to solve this stochastic optimal control problem. However, the HJB equation will need to be revised because of the Taylor expansion of $V(x_{t+dt}, t+dt)$ will need to be taken in the form of stochastic differential equations. E.g. if the state dynamics is in the form of

$$dx = f\left(x\left(t\right), u\left(t\right), t\right) dt + dW$$

where dW is a Wiener processes with $\langle dW_i dW_j \rangle = v_{ij}\left(t, x, u\right) dt$ a symmetric positive definite matrix, then by using stochastic calculus we have

$$-V_t\left(x, t\right) = \max_{u(s)\in\Omega(s)} \left\{ F(x, u, s) + V_x(x, t) f(x, u, t) + \frac{1}{2} v\left(t, x, u\right) V_{xx}(x, t) \right\}$$

, which is called the stochastic Hamilton-Jacobi-Bellman Equation with boundary condition $V(x, T) = S(x, T)$.

(Similarly, the stochastic Pontryagin minimum principle can be derived from from stochastic HJB equation. E.g. see the note of "Stochastic_Optimal_Control.pdf")