M. Lazar, W. P. M. H. Heemels, and A. R. Teel. Lyapunov functions, stability and input-to-state stability subtleties for discrete-time discontinuous systems. *IEEE Trans. Auto. Cont.*, 54(10):2421–2425, 2009.

J. B. Rawlings and L. Ji. Optimization-based state estimation: Current status and some new results. *J. Proc. Cont.*, 22:1439–1444, 2012.

J. B. Rawlings and M. J. Risbeck. On the equivalence between statements with epsilon-delta and K-functions. Technical Report 2015–01, TWCCC Technical Report, December 2015.

J. B. Rawlings and M. J. Risbeck. Model predictive control with discrete actuators: Theory and application. *Automatica*, 78:258–265, 2017.

E. D. Sontag. *Mathematical Control Theory*. Springer-Verlag, New York, second edition, 1998a.

E. D. Sontag. Comments on integral variants of ISS. *Sys. Cont. Let.*, 34:93–100, 1998b.

E. D. Sontag and Y. Wang. On the characterization of the input to state stability property. *Sys. Cont. Let.*, 24:351–359, 1995.

E. D. Sontag and Y. Wang. Output-to-state stability and detectability of nonlinear systems. *Sys. Cont. Let.*, 29:279–290, 1997.

A. R. Teel. Discrete time receding horizon control: is the stability robust. In Marcia S. de Queiroz, Michael Malisoff, and Peter Wolenski, editors, *Optimal control, stabilization and nonsmooth analysis*, volume 301 of *Lecture notes in control and information sciences*, pages 3–28. Springer, 2004.

A. R. Teel and L. Zaccarian. On “uniformity” in definitions of global asymptotic stability for time-varying nonlinear systems. *Automatica*, 42:2219–2222, 2006.

C
Optimization

C.1 Dynamic Programming

The name *dynamic programming* dates from the 1950s when it was coined by Richard Bellman for a technique for solving dynamic optimization problems, i.e., optimization problems associated with deterministic or stochastic systems whose behavior is governed by differential or difference equations. Here we review some of the basic ideas behind dynamic programming (DP) Bellman (1957); Bertsekas, Nedic, and Ozdaglar (2001).

To introduce the topic in its simplest form, consider the simple routing problem illustrated in Figure C.1. To maintain connection with optimal control, each node in the graph can be regarded as a point (x, t) in a subset S of $X \times T$ where both the state space $X = \{a, b, c, \ldots, g\}$ and the set of times $T = \{0, 1, 2, 3\}$ are discrete. The set of permissible control actions is $\mathbb{U} = \{U, D\}$, i.e., to go "up" or "down." The control problem is to choose the lowest cost path from event $(d, 0)$ (state d at $t = 0$) to any of the states at $t = 3$; the cost of going from one event to the next is indicated on the graph. This problem is equivalent to choosing an open-loop control, i.e., a sequence $(u(0), u(1), u(2))$ of admissible control actions. There are 2^N controls where N is the number of stages, 3 in this example. The cost of each control can, in this simple example, be evaluated and is given in Table C.1.

There are two different *open-loop* optimal controls, namely (U, D, U) and (D, D, D), each incurring a cost of 16. The corresponding

control	UUU	UUD	UDU	UDD	DUU	DUD	DDU	DDD
cost	20	24	16	24	24	32	20	16

Table C.1: Control Cost.

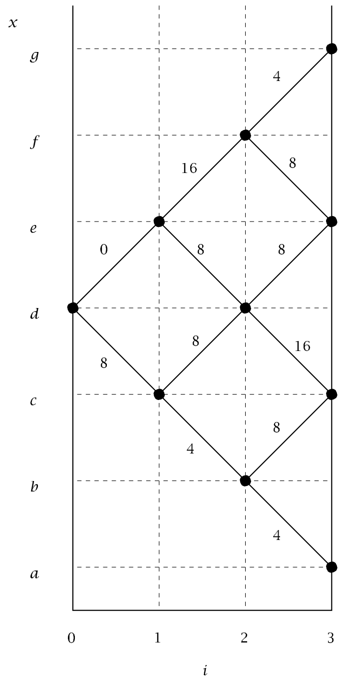


Figure C.1: Routing problem.

state trajectories are (d, e, d, e) and (d, c, b, a).

In discrete problems of this kind, DP replaces the N-stage problem by M single stage problems, where M is the total number of nodes, i.e., the number of elements in $S \subset X \times T$. The first set of optimization problems deals with the states b, d, f at time $N - 1 = 2$. The optimal decision at event $(f, 2)$, i.e., state f at time 2, is the control U and gives rise to a cost of 4. The optimal cost and control for node $(f,$ $2)$ are recorded; see Table C.2. The procedure is then repeated for states d and b at time $t = 2$ (nodes $(d, 2)$ and $(b, 2)$) and recorded as shown in Table C.2. Attention is next focused on the states e and c at $t = 1$ (nodes $(e, 1)$ and $(c, 1)$). The lowest cost that can be achieved at node $(e, 1)$ *if* control U is chosen, is $16 + 4$, the sum of the path cost 16 associated with the control U, and the *optimal* cost 4 associated with the node $(f, 2)$ that results from using control U at node $(e, 1)$. Similarly the lowest possible cost, if control D is chosen, is $8 + 8$. Hence

t	0	1		2		
state	d	e	c	f	d	b
control	U or D	D	D	U	U	D
optimal cost	16	16	8	4	8	4

Table C.2: Optimal Cost and Control

the optimal control and cost for node $(e, 1)$ are, respectively, D and 16. The procedure is repeated for the remaining state d at $t = 1$ (node $(d, 1)$). A similar calculation for the state d at $t = 0$ (node $(d, 0)$), where the optimal control is U or D, completes this backward recursion; this backward recursion provides the optimal cost and control for each (x, t), as recorded in Table C.2. The procedure therefore yields an optimal *feedback* control that is a function of $(x, t) \in S$. To obtain the optimal open-loop control for the initial node $(d, 0)$, the feedback law is obeyed, leading to control U or D at $t = 0$; if U is chosen, the resultant state at $t = 1$ is e. From Table C.2, the optimal control at $(e, 1)$ is D, so that the successor node is $(d, 2)$. The optimal control at node $(d, 2)$ is U. Thus the optimal open-loop control sequence (U, D, U) is re-obtained. On the other hand, if the decision at $(d, 0)$ is chosen to be D, the optimal sequence (D, D, D) is obtained. This simple example illustrates the main features of DP that we will now examine in the context of discrete time optimal control.

C.1.1 Optimal Control Problem

The discrete time system we consider is described by

$$x^+ = f(x, u) \tag{C.1}$$

where $f(\cdot)$ is continuous. The system is subject to the mixed state-control constraint

$$(x, u) \in \mathbb{Z}$$

where $\mathbb{Z}$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^m$ and $\mathcal{P}_u(\mathbb{Z})$ is compact where $\mathcal{P}_u$ is the projection operator $(x, u) \mapsto u$. Often $\mathbb{Z} = \mathbb{X} \times \mathbb{U}$ in which case the constraint $(x, u) \in \mathbb{Z}$ becomes $x \in \mathbb{X}$ and $u \in \mathbb{U}$ and $\mathcal{P}_u(\mathbb{Z}) = \mathbb{U}$ so that $\mathbb{U}$ is compact. In addition there is a constraint on the terminal state $x(N)$:

$$x(N) \in \mathbb{X}_f$$

where $\mathbb{X}_f$ is closed. In this section we find it easier to express the value function and the optimal control in terms of the current state and current time i rather than using time-to-go k. Hence we replace time-to-go k by time i where $k = N - i$, replace $V_k^0(x)$ (the optimal cost at state x when the time-to-go is k) by $V^0(x, i)$ (the optimal cost at state x, time i) and replace X_k by $X(i)$ where $X(i)$ is the domain of $V^0(\cdot, i)$).

The cost associated with an initial state x at time 0 and a control sequence $\mathbf{u} := (u(0), u(1), \ldots, u(N-1))$ is

$$V(x, 0, \mathbf{u}) = V_f(x(N)) + \sum_{i=1}^{N-1} \ell(x(i), u(i)) \tag{C.2}$$

where $\ell(\cdot)$ and $V_f(\cdot)$ are continuous and, for each i, $x(i) = \phi(i; (x, 0), \mathbf{u})$ is the solution at time i of (C.1) if the initial state is x at time 0 and the control sequence is $\mathbf{u}$. The optimal control problem $\mathbb{P}(x, 0)$ is defined by

$$V^0(x, 0) = \min_{\mathbf{u}} V(x, 0, \mathbf{u}) \tag{C.3}$$

subject to the constraints $(x(i), u(i)) \in \mathbb{Z}$, $i = 0, 1, \ldots, N-1$ and $x(N) \in \mathbb{X}_f$. Equation (C.3) may be rewritten in the form

$$V^0(x, 0) = \min_{\mathbf{u}} \{V(x, 0, \mathbf{u}) \mid \mathbf{u} \in \mathcal{U}(x, 0)\} \tag{C.4}$$

where $\mathbf{u} := (u(0), u(1), \ldots, u(N-1))$,

$$\mathcal{U}(x, 0) := \{\mathbf{u} \in \mathbb{R}^{Nm} \mid (x(i), u(i)) \in \mathbb{Z}, i = 0, 1, \ldots, N-1; x(N) \in \mathbb{X}_f\}$$

and $x(i) := \phi(i; (x, 0), \mathbf{u})$. Thus $\mathcal{U}(x, 0)$ is the set of admissible control sequences[1] if the initial state is x at time 0. It follows from the continuity of $f(\cdot)$ that for all $i \in \{0, 1, \ldots, N-1\}$ and all $x \in \mathbb{R}^n$, $\mathbf{u} \mapsto \phi(i; (x, 0), \mathbf{u})$ is continuous, $\mathbf{u} \mapsto V(x, 0, \mathbf{u})$ is continuous and $\mathcal{U}(x, 0)$ is compact. Hence the minimum in (C.4) exists at all $x \in \{x \in \mathbb{R}^n \mid \mathcal{U}(x, 0) \neq \emptyset\}$.

DP embeds problem $\mathbb{P}(x, 0)$ for a given state x in a whole family of problems $P(x, i)$ where, for each (x, i), problem $\mathbb{P}(x, i)$ is defined by

$$V^0(x, i) = \min_{\mathbf{u}^i} \{V(x, i, \mathbf{u}^i) \mid \mathbf{u}^i \in \mathcal{U}(x, i)\}$$

where

$$\mathbf{u}^i := (u(i), u(i+1), \ldots, u(N-1))$$

[1] An admissible control sequence satisfies all constraints.

$$V(x, i, \mathbf{u}^i) := V_f(x(N)) + \sum_{j=i}^{N-1} \ell(x(j), u(j)) \tag{C.5}$$

and

$$\mathcal{U}(x, i) := \{\mathbf{u}^i \in \mathbb{R}^{(N-i)m} \mid (x(j), u(j)) \in \mathbb{Z}, j = i, i+1, \ldots, N-1 \\ x(N) \in \mathbb{X}_f\} \tag{C.6}$$

In (C.5) and (C.6), $x(j) = \phi(j; (x, i), \mathbf{u}^i)$, the solution at time j of (C.1) if the initial state is x at time i and the control sequence is $\mathbf{u}^i$. For each i, $X(i)$ denotes the domain of $V^0(\cdot, i)$ and $\mathcal{U}(\cdot, i)$ so that

$$X(i) = \{x \in \mathbb{R}^n \mid \mathcal{U}(x, i) \neq \emptyset\}. \tag{C.7}$$

C.1.2 Dynamic Programming

One way to approach DP for discrete time control problems is the simple observation that for all (x, i)

$$\begin{aligned} V^0(x, i) &= \min_{\mathbf{u}^i}\{V(x, i, \mathbf{u}^i) \mid \mathbf{u}^i \in \mathcal{U}(x, i)\} \\ &= \min_{u}\{\ell(x, u) + \min_{\mathbf{u}^{i+1}} V(f(x, u), i+1, \mathbf{u}^{i+1}) \mid \\ &\qquad \{u, \mathbf{u}^{i+1}\} \in \mathcal{U}(x, i)\} \end{aligned} \tag{C.8}$$

where $\mathbf{u}^i = (u, u(i+1), \ldots, u(N-1)) = \left(u, \mathbf{u}^{i+1}\right)$. We now make use of the fact that $\{u, \mathbf{u}^{i+1}\} \in \mathcal{U}(x, i)$ if and only if $(x, u) \in \mathbb{Z}$, $f(x, u) \in X(i+1)$, and $\mathbf{u}^{i+1} \in \mathcal{U}(f(x, u), i+1)$ since $f(x, u) = x(i+1)$. Hence we may rewrite (C.8) as

$$\begin{aligned} V^0(x, i) = \min_{u}\{\ell(x, u) + V^0(f(x, u), i+1) \mid \\ (x, u) \in \mathbb{Z}, f(x, u) \in X(i+1)\} \end{aligned} \tag{C.9}$$

for all $x \in X(i)$ where

$$X(i) = \{x \in \mathbb{R}^n \mid \exists u \text{ such that } (x, u) \in \mathbb{Z} \text{ and } f(x, u) \in X(i+1)\} \tag{C.10}$$

Equations (C.9) and (C.10), together with the boundary condition

$$V^0(x, N) = V_f(x) \;\; \forall x \in X(N), \quad X(N) = \mathbb{X}_f$$

constitute the DP recursion for constrained discrete time optimal control problems. If there are no state constraints, i.e., if $\mathbb{Z} = \mathbb{R}^n \times \mathbb{U}$ where

$\mathbb{U} \subset \mathbb{R}^m$ is compact, then $X(i) = \mathbb{R}^n$ for all $i \in \{0, 1, \ldots, N\}$ and the DP equations revert to the familiar DP recursion:

$$V^0(x, i) = \min_u \{\ell(x, u) + V^0(f(x, u), i + 1)\} \ \forall x \in \mathbb{R}^n$$

with boundary condition

$$V^0(x, N) = V_f \ \forall x \in \mathbb{R}^n$$

We now prove some basic facts; the first is the well known *principle of optimality.*

Lemma C.1 (Principle of optimality)**.** *Let $x \in X_N$ be arbitrary, let $\mathbf{u} := (u(0), u(1), \ldots, u(N-1)) \in \mathcal{U}(x, 0)$ denote the solution of $\mathbb{P}(x, 0)$ and let $(x, x(1), x(2), \ldots, x(N))$ denote the corresponding optimal state trajectory so that for each i, $x(i) = \phi(i; (x, 0), \mathbf{u})$. Then, for any $i \in \{0, 1, \ldots, N-1\}$, the control sequence $\mathbf{u}^i := (u(i), u(i+1), \ldots, u(N-1))$ is optimal for $\mathbb{P}(x(i), i)$ (any portion of an optimal trajectory is optimal).*

Proof. Since $\mathbf{u} \in \mathcal{U}(x, 0)$, the control sequence $\mathbf{u}^i \in \mathcal{U}(x(i), i)$. If $\mathbf{u}^i = (u(i), u(i+1), \ldots, u(N-1))$ is not optimal for $\mathbb{P}(x(i), i)$, there exists a control sequence $\mathbf{u}' = (u'(i), u'(i+1), \ldots, u(N-1)') \in \mathcal{U}(x(i), i)$ such that $V(x(i), i, \mathbf{u}') < V(x(i), \mathbf{u})$. Consider now the control sequence $\tilde{\mathbf{u}} := (u(0), u(1), \ldots, u(i-1), u'(i), u'(i+1), \ldots, u(N-1)')$. It follows that $\tilde{\mathbf{u}} \in \mathcal{U}(x, 0)$ and $V(x, 0, \tilde{\mathbf{u}}) < V(x, 0, \mathbf{u}) = V^0(x, 0)$, a contradiction. Hence $\mathbf{u}(x(i), i)$ is optimal for $\mathbb{P}(x(i), i)$. ■

The most important feature of DP is the fact that the DP recursion yields the optimal value $V^0(x, i)$ and the optimal control $\kappa(x, i) = \arg\min_u \{\ell(x, u) + V^0(f(x, u), i+1) \mid (x, u) \in \mathbb{Z}, f(x, u) \in X(i+1)\}$ for each $(x, i) \in X(i) \times \{0, 1, \ldots, N-1\}$.

Theorem C.2 (Optimal value function and control law from DP)**.** *Suppose that the function $\Psi : \mathbb{R}^n \times \{0, 1, \ldots, N\} \to \mathbb{R}$, satisfies, for all $i \in \{1, 2, \ldots, N-1\}$, all $x \in X(i)$, the DP recursion*

$$\Psi(x, i) = \min\{\ell(x, u) + \Psi(f(x, u), i+1) \mid (x, u) \in \mathbb{Z}, f(x, u) \in X(i+1)\}$$
$$X(i) = \{x \in \mathbb{R}^n \mid \exists u \in \mathbb{R}^m \text{ such that } (x, u) \in \mathbb{Z},\ f(x, u) \in X(i+1)\}$$

with boundary conditions

$$\Psi(x, N) = V_f(x) \ \forall x \in \mathbb{X}_f, \quad X(N) = \mathbb{X}_f$$

Then $\Psi(x, i) = V^0(x, i)$ for all $(x, i) \in X(i) \times \{0, 1, 2, \ldots, N\}$; the DP recursion yields the optimal value function and the optimal control law.

Proof. Let $(x, i) \in X(i) \times \{0, 1, \ldots, N\}$ be arbitrary. Let $\mathbf{u} = (u(i), u(i+1), \ldots, u(N-1))$ be an arbitrary control sequence in $\mathcal{U}(x, i)$ and let $\mathbf{x} = (x, x(i+1), \ldots, x(N))$ denote the corresponding trajectory starting at (x, i) so that for each $j \in \{i, i+1, \ldots, N\}$, $x(j) = \phi(j; x, i, \mathbf{u})$. For each $j \in \{i, i+1, \ldots, N-1\}$, let $\mathbf{u}^j := (u(j), u(j+1), \ldots, u(N-1))$; clearly $\mathbf{u}^j \in \mathcal{U}(x(j), j)$. The cost due to initial event $(x(j), j)$ and control sequence $\mathbf{u}^j$ is $\Phi(x(j), j)$ defined by

$$\Phi(x(j), j) := V(x(j), j, \mathbf{u}^j)$$

Showing that $\Psi(x, i) \leq \Phi(x, i)$ proves that $\Psi(x, i) = V^0(x, i)$ since $\mathbf{u}$ is an arbitrary sequence in $\mathcal{U}(x, i)$; because $(x, i) \in X(i) \times \{0, 1, \ldots, N\}$ is arbitrary, that fact that $\Psi(x, i) = V^0(x, i)$ proves that DP yields the optimal value function.

To prove that $\Psi(x, i) \leq \Phi(x, i)$, we compare $\Psi(x(j), j)$ and $\Phi(x(j), j)$ for each $j \in \{i, i+1, \ldots, N\}$, i.e., we compare the costs yielded by the DP recursion and by the arbitrary control $\mathbf{u}$ along the corresponding trajectory $\mathbf{x}$. By definition, $\Psi(x(j), j)$ satisfies for each j

$$\Psi(x(j), j) = \min_u \{\ell(x(j), u) + \Psi(f(x(j), u), j+1) \mid \\ (x(j), u) \in \mathbb{Z}, f(x(j), u) \in X(j+1)\} \quad \text{(C.11)}$$

To obtain $\Phi(x(j), j)$ for each j we solve the following recursive equation

$$\Phi(x(j), j) = \ell(x(j), u(j)) + \Phi(f(x(j), u(j)), j+1) \qquad \text{(C.12)}$$

The boundary conditions are

$$\Psi(x(N), N) = \Phi(x(N), N) = V_f(x(N)) \qquad \text{(C.13)}$$

Since $u(j)$ satisfies $(x(j), u(j)) \in \mathbb{Z}$ and $f(x(j), u(j)) \in X(j+1)$ but is not necessarily a minimizer in (C.11), we deduce that

$$\Psi(x(j), j) \leq \ell(x(j), u(j)) + \Psi(f(x(j), u(j)), j+1) \qquad \text{(C.14)}$$

For each j, let $E(j)$ be defined by

$$E(j) := \Psi(x(j), j) - \Phi(x(j), j)$$

Subtracting (C.12) from (C.14) and replacing $f(x(j), u(j))$ by $x(j+1)$ yields

$$E(j) \leq E(j+1) \;\; \forall j \in \{i, i+1, \ldots N\}$$

Since $E(N) = 0$ by virtue of (C.13), we deduce that $E(j) \leq 0$ for all $j \in \{i, i+1, \ldots, N\}$; in particular, $E(i) \leq 0$ so that

$$\Psi(x, i) \leq \Phi(x, i) = V(x, i, \mathbf{u})$$

for all $\mathbf{u} \in \mathcal{U}(x, i)$. Hence $\Psi(x, i) = V^0(x, i)$ for all $(x, i) \in X(i) \times \{0, 1, \ldots, N\}$. ∎

Example C.3: DP applied to linear quadratic regulator

A much used example is the familiar linear quadratic regulator problem. The system is defined by

$$x^+ = Ax + Bu$$

There are no constraints. The cost function is defined by (C.2) where

$$\ell(x, u) := (1/2)x'Qx + (1/2)u'Ru$$

and $V_f(x) = 0$ for all x; the horizon length is N. We assume that Q is symmetric and positive semidefinite and that R is symmetric and positive definite. The DP recursion is

$$V^0(x, i) = \min_u \{\ell(x, u) + V^0(Ax + Bu, i+1)\} \ \forall x \in \mathbb{R}^n$$

with terminal condition

$$V^0(x, N) = 0 \ \forall x \in \mathbb{R}^n$$

Assume that $V^0(\cdot, i+1)$ is quadratic and positive semidefinite and, therefore, has the form

$$V^0(x, i+1) = (1/2)x'P(i+1)x$$

where $P(i+1)$ is symmetric and positive semidefinite. Then

$$V^0(x, i) = (1/2) \min_u \{x'Qx + u'Ru + (Ax + Bu)'P(i+1)(Ax + Bu)\}$$

The right-hand side of the last equation is a positive definite function of u for all x, so that it has a unique minimizer given by

$$\kappa(x, i) = K(i)x \qquad K(i) := -(B'P(i+1)B + R)^{-1}B'P(i+1)$$

Substituting $u = K(i)x$ in the expression for $V^0(x, i)$ yields

$$V^0(x, i) = (1/2)x'P(i)x$$

where $P(i)$ is given by:

$$P(i) = Q + K(i)'RK(i) - A'P(i+1)B(B'P(i+1)B + R)^{-1}B'P(i+1)A$$

Hence $V^0(\cdot, i)$ is quadratic and positive semidefinite if $V^0(\cdot, i+1)$ is. But $V^0(\cdot, N)$, defined by

$$V^0(x, N) := (1/2)x'P(N)x = 0 \qquad P(N) := 0$$

is symmetric and positive semidefinite. By induction $V^0(\cdot, i)$ is quadratic and positive semidefinite (and $P(i)$ is symmetric and positive semidefinite) for all $i \in \{0, 1, \ldots, N\}$. Substituting $K(i) = -(B'P(i+1)B + R)^{-1}B'P(i+1)A$ in the expression for $P(i)$ yields the more familiar matrix Riccati equation

$$P(i) = Q + A'P(i+1)A - A'P(i+1)B(B'P(i+1)B + R)^{-1}BP(i+1)A$$

$\square$

C.2 Optimality Conditions

In this section we obtain optimality conditions for problems of the form

$$f^0 = \inf_u \{f(u) \mid u \in U\}$$

In these problems, $u \in \mathbb{R}^m$ is the *decision* variable,$f(u)$ the cost to be minimized by appropriate choice of $u$and $U \subset \mathbb{R}^m$ the constraint set. The value of the problem is f^0. Some readers may wish to read only Section C.2.2, which deals with convex optimization problems and Section C.2.3 which deals with convex optimization problems in which the constraint set U is polyhedral. These sections require some knowledge of tangent and normal cones discussed in Section C.2.1; Proposition C.7 in particular derives the normal cone for the case when U is convex.

C.2.1 Tangent and Normal Cones

In determining conditions of optimality, it is often convenient to employ approximations to the cost function $f(\cdot)$ and the constraint set U. Thus the cost function $f(\cdot)$ may be approximated, in the neighborhood of a point $\bar{u}$, by the first order expansion $f(\bar{u}) + \langle \nabla f(\bar{u}), (u - \bar{u}) \rangle$ or by the second order expansion $f(\bar{u}) + \langle \nabla f(\bar{u}), (u - \bar{u}) \rangle + (1/2)((u - \bar{u})'\nabla^2 f(\bar{x})(u - \bar{u}))$ if the necessary derivatives exist. Thus we see that

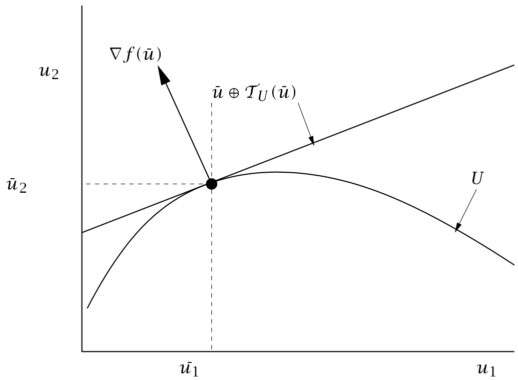


Figure C.2: Approximation of the set U.

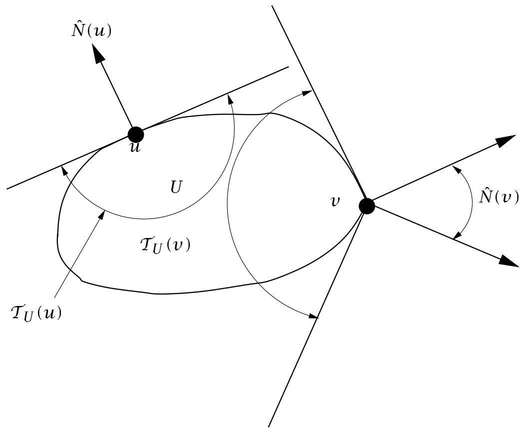


Figure C.3: Tangent cones.

in the unconstrained case, a necessary condition for the optimality of $\bar{u}$ is $\nabla f(\bar{u}) = 0$. To obtain necessary conditions of optimality for constrained optimization problems, we need to approximate the constraint set as well; this is more difficult. An example of U and its approximation is shown in Figure C.2; here the set $U = \{u \in \mathbb{R}^2 \mid g(u) = 0\}$ where $g : \mathbb{R} \to \mathbb{R}$ is approximated in the neighborhood of a point $\bar{u}$ satisfying $g(\bar{u}) = 0$ by the set $\bar{u} \oplus \mathcal{T}_U(\bar{u})$ where[2] the tangent cone $\mathcal{T}_U(\bar{u}) := \{h \in \mathbb{R}^2 \mid \nabla g(\bar{u}), u - \bar{u}\rangle = 0\}$. In general, a set U is approx-

[2]If A and B are two subsets of $\mathbb{R}^n$, say, then $A \oplus B := \{a + b \mid a \in A, b \in B\}$ and $a \oplus B := \{a + b \mid b \in B\}$.

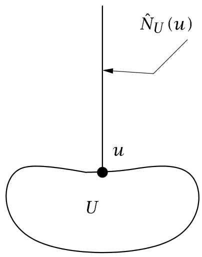


Figure C.4: Normal at u.

imated, near a point $\bar{u}$, by $\bar{u} \oplus \mathcal{T}_U(\bar{u})$ where its *tangent cone* $\mathcal{T}_U(\bar{u})$ is defined below. Following Rockafellar and Wets (1998), we use $u^\nu \xrightarrow[U]{} v$ to denote that the sequence $\{u^\nu \mid \nu \in \mathbb{I}_{\geq 0}\}$ converges to v as $\nu \to \infty$ while satisfying $u^\nu \in U$ for all $\nu \in \mathbb{I}_{\geq 0}$.

Definition C.4 (Tangent vector)**.** A vector $h \in \mathbb{R}^m$ is tangent to the set U at $\bar{u}$ if there exist sequences $u^\nu \xrightarrow[U]{} \bar{u}$ and $\lambda^\nu \searrow 0$ such that

$$[u^\nu - \bar{u}]/\lambda^\nu \to h$$

$\mathcal{T}_U(u)$ is the set of all tangent vectors.

Equivalently, a vector $h \in \mathbb{R}^m$ is tangent to the set U at $\bar{u}$ if there exist sequences $h^\nu \to h$ and $\lambda^\nu \searrow 0$ such that $\bar{u} + \lambda^\nu h^\nu \in U$ for all $\nu \in \mathbb{I}_{\geq 0}$. This equivalence can be seen by identifying u^ν with $\bar{u} + \lambda^\nu h^\nu$.

Proposition C.5 (Tangent vectors are closed cone)**.** *The set $\mathcal{T}_U(u)$ of all tangent vectors to U at any point $u \in U$ is a closed cone.*

See Rockafellar and Wets (1998), Proposition 6.2. That $\mathcal{T}_U(\bar{u})$ is a cone may be seen from its definition; if h is a tangent, so is αh for any $\alpha \geq 0$. Two examples of a tangent cone are illustrated in Figure C.3.

Associated with each tangent cone $\mathcal{T}_U(u)$ is a normal cone $\hat{N}(u)$ defined as follows Rockafellar and Wets (1998):

Definition C.6 (Regular normal)**.** A vector $g \in \mathbb{R}^m$ is a regular normal to a set $U \subset \mathbb{R}^m$ at $\bar{u} \in U$ if

$$\langle g, u - \bar{u} \rangle \leq o(|u - \bar{u}|) \ \forall u \in U \tag{C.15}$$

where $o(\cdot)$ has the property that $o(|u - \bar{u}|)/|u - \bar{u}| \to 0$ as $u \xrightarrow[U]{} \bar{u}$ with $u \neq \bar{u}$; $\hat{N}_U(u)$ is the set of all regular normal vectors.

Some examples of normal cones are illustrated in Figure C.3; here the set $\hat{N}_U(u) = \{\lambda g \mid \lambda \geq 0\}$ is a cone generated by a single vector g, say, while $\hat{N}_U(v) = \{\lambda_1 g_1 + \lambda_2 g_2 \mid \lambda_1 \geq 0, \lambda_2 \geq 0\}$ is a cone generated by two vectors g_1 and g_2, say. The term $o(|u - \bar{u}|)$ may be replaced by 0 if U is convex as shown in Proposition C.7(b) below but is needed in general since U may not be locally convex at $\bar{u}$ as illustrated in Figure C.4.

The tangent cone $\mathcal{T}_U(\bar{u})$ and the normal cone $\hat{N}_U(\bar{u})$ at a point $\bar{u} \in U$ are related as follows.

Proposition C.7 (Relation of normal and tangent cones)**.**

(a) At any point $\bar{u} \in U \subset \mathbb{R}^m$,

$$\hat{N}_U(\bar{u}) = \mathcal{T}_U(\bar{u})^* := \{g \mid \langle g, h\rangle \leq 0 \ \forall h \in \mathcal{T}_U(\bar{u})\}$$

where, for any cone V, $V^* := \{g \mid \langle g, h\rangle \leq 0 \ \forall h \in V\}$ denotes the polar cone of V.

(b) If U is convex, then, at any point $\bar{u} \in U$

$$\hat{N}_U(\bar{u}) = \{g \mid \langle g, u - \bar{u}\rangle \leq 0 \ \forall u \in U\} \tag{C.16}$$

Proof.

(a) To prove $\hat{N}_U(\bar{u}) \subset \mathcal{T}_U(\bar{u})^*$, we take an arbitrary point g in $\hat{N}_U(\bar{u})$ and show that $\langle g, h\rangle \leq 0$ for all $h \in \mathcal{T}(\bar{u})$ implying that $g \in \mathcal{T}_U^*(\bar{u})$. For, if h is tangent to U at $\bar{u}$, there exist, by definition, sequences $u^\nu \xrightarrow[U]{} \bar{u}$ and $\lambda^\nu \searrow 0$ such that

$$h^\nu := (u^\nu - \bar{u})/\lambda^\nu \to h$$

Since $g \in \hat{N}_U(\bar{u})$, it follows from (C.15) that $\langle g, h^\nu\rangle \leq o(|(u^\nu - \bar{u})|) = o(\lambda^\nu |h^\nu|)$; the limit as $\nu \to \infty$ yields $\langle g, h\rangle \leq 0$, so that $g \in \mathcal{T}_U^*(\bar{u})$. Hence $\hat{N}_U(\bar{u}) \subset \mathcal{T}_U(\bar{u})^*$. The proof of this result, and the more subtle proof of the converse, that $\mathcal{T}_U(\bar{u})^* \subset \hat{N}_U(\bar{u})$, are given in Rockafellar and Wets (1998), Proposition 6.5.

(b) This part of the proposition is proved in (Rockafellar and Wets, 1998, Theorem 6.9). ∎

Remark. A consequence of (C.16) is that for each $g \in \hat{N}_U(\bar{u})$, the half-space $H_g := \{u \mid \langle g, u - \bar{u}\rangle \leq 0\}$ supports the convex set U at $\bar{u}$, i.e., $U \subset H_g$ and $\bar{u}$ lies on the boundary of the half-space H_g.

We wish to derive optimality conditions for problems of the form $\mathbb{P} : \inf_u \{f(u) \mid u \in U\}$. The *value* of the problem is defined to be

$$f^0 := \inf_u \{f(u) \mid u \in U\}$$

There may not exist a $u \in U$ such that $f(u) = f^0$. If, however, $f(\cdot)$ is continuous and U is compact, there exists a minimizing u in U, i.e.,

$$f^0 = \inf_u \{f(u) \mid u \in U\} = \min_u \{f(u) \mid u \in U\}$$

The minimizing u, if it exists, may not be unique so

$$u^0 := \arg\min_u \{f(u) \mid u \in U\}$$

may be a set. We say u is feasible if $u \in U$. A point u is *globally optimal* for problem $\mathbb{P}$ if u is feasible and $f(v) \geq f(u)$ for all $v \in U$. A point u is *locally optimal* for problem $\mathbb{P}$ if u is feasible and there exists a $\varepsilon > 0$ such that $f(v) \geq f(u)$ for all v in $(u \oplus \varepsilon\mathcal{B}) \cap U$ where $\mathcal{B}$ is the closed unit ball $\{u \mid \min |u| \leq 1\}$.

C.2.2 Convex Optimization Problems

The optimization problem $\mathbb{P}$ is convex if the function $f : \mathbb{R}^m \to \mathbb{R}$ and the set $U \subset \mathbb{R}^m$ are convex. In convex optimization problems, U often takes the form $\{u \mid g_j(u) \leq 0, j \in \mathcal{J}\}$ where $\mathcal{J} := \{1, 2, \ldots, J\}$ and each function $g_j(\cdot)$ is convex. A useful feature of convex optimization problems is the following result:

Proposition C.8 (Global optimality for convex problems)**.** *Suppose the function $f(\cdot)$ is convex and differentiable and the set U is convex. Any locally optimal point of the convex optimization problem* $\inf_u \{f(u) \mid u \in U\}$ *is globally optimal.*

Proof. Suppose u is locally optimal so that there exists an $\varepsilon > 0$ such that $f(v) \geq f(u)$ for all $v \in (u \oplus \varepsilon\mathcal{B}) \cap U$. If, contrary to what we wish to prove, u is *not* globally optimal, there exists a $w \in U$ such that $f(w) < f(u)$. For any $\lambda \in [0, 1]$, the point $w_\lambda := \lambda w + (1 - \lambda)u$ lies in $[u, w]$ (the line joining u and w). Then $w_\lambda \in U$ (because U is convex) and $f(w_\lambda) \leq \lambda f(w) + (1 - \lambda)f(u) < f(u)$ for all $\lambda \in (0, 1]$ (because $f(\cdot)$ is convex and $f(w) < f(u)$). We can choose $\lambda > 0$ so that $w_\lambda \in (u \oplus \varepsilon\mathcal{B}) \cap U$ and $f(w_\lambda) < f(u)$. This contradicts the local optimality of u. Hence u is globally optimal. ∎

On the assumption that $f(\cdot)$ is differentiable, we can obtain a simple necessary and sufficient condition for the (global) optimality of a point u.

Proposition C.9 (Optimality conditions—normal cone). *Suppose the function $f(\cdot)$ is convex and differentiable and the set U is convex. The point u is optimal for problem $\mathbb{P}$ if and only if $u \in U$ and*

$$df(u; v - u) = \langle \nabla f(u), v - u \rangle \geq 0 \ \forall v \in U \tag{C.17}$$

or, equivalently

$$-\nabla f(u) \in \hat{N}_U(u) \tag{C.18}$$

Proof. Because $f(\cdot)$ is convex, it follows from Theorem 7 in Appendix A1 that

$$f(v) \geq f(u) + \langle \nabla f(u), v - u \rangle \tag{C.19}$$

for all u, v in U. To prove sufficiency, suppose $u \in U$ and that the condition in (C.17) is satisfied. It then follows from (C.19) that $f(v) \geq f(u)$ for all $v \in U$ so that u is globally optimal. To prove necessity, suppose that u is globally optimal but that, contrary to what we wish to prove, the condition on the right-hand side of (C.17) is not satisfied so that there exists a $v \in U$ such that

$$df(u; h) = \langle \nabla f(u), v - u \rangle = -\delta < 0$$

where $h := v - u$. For all $\lambda \in [0, 1]$, let $v_\lambda := \lambda v + (1 - \lambda)u = u + \lambda h$; because U is convex, each v_λ lies in U. Since

$$df(u; h) = \lim_{\lambda \searrow 0} \frac{f(u + \lambda h) - f(u)}{\lambda} = \lim_{\lambda \searrow 0} \frac{f(v_\lambda) - f(u)}{\lambda} = -\delta$$

there exists a $\lambda \in (0, 1]$ such that $f(v_\lambda) - f(u) \leq -\lambda\delta/2 < 0$ which contradicts the optimality of u. Hence the condition in (C.17) must be satisfied. That (C.17) is equivalent to (C.18) follows from Proposition C.7(b). ■

Remark. The condition (C.17) implies that the linear approximation $\hat{f}(v) := f(u) + \langle \nabla f(u), v - u \rangle$ to $f(v)$ achieves its minimum over U at u.

It is an interesting fact that U in Proposition C.9 may be replaced by its approximation $u \oplus \mathcal{T}_U(u)$ at u yielding

Proposition C.10 (Optimality conditions—tangent cone)**.** *Suppose the function $f(\cdot)$ is convex and differentiable and the set U is convex. The point u is optimal for problem $\mathbb{P}$ if and only if $u \in U$ and*

$$df(u; v - u) = \langle \nabla f(u), h \rangle \geq 0 \;\; \forall h \in \mathcal{T}_U(u)$$

or, equivalently

$$-\nabla f(u) \in \hat{N}_U(u) = \mathcal{T}_U^*(u).$$

Proof. It follows from Proposition C.9 that u is optimal for problem $\mathbb{P}$ if and only if $u \in U$ and $-\nabla f(u) \in \hat{N}_U(u)$. But, by Proposition C.7, $\hat{N}_U(u) = \{g \mid \langle g, h \rangle \leq 0 \;\; \forall h \in \mathcal{T}_U(u)\}$ so that $-\nabla f(u) \in \hat{N}_U(u)$ is equivalent to $\langle \nabla f(u), h \rangle \geq 0$ for all $h \in \mathcal{T}_U(u)$. ∎

C.2.3 Convex Problems: Polyhedral Constraint Set

The definitions of tangent and normal cones given above may appear complex but this complexity is necessary for proper treatment of the general case when U is not necessarily convex. When U is polyhedral, i.e., when U is defined by a set of linear inequalities

$$U := \{u \in \mathbb{R}^m \mid Au \leq b\}$$

where $A \in \mathbb{R}^{p \times m}$ and $b \in \mathbb{R}^p$, $\mathcal{I} := \{1, 2, \ldots, p\}$, then the normal and tangent cones are relatively simple. We first note that U is equivalently defined by

$$U := \{u \in \mathbb{R}^m \mid \langle a_i, u \rangle \leq b_i, \; i \in \mathcal{I}\}$$

where a_i is the ith row of A and b_i is the ith element of b. For each $u \in U$, let

$$\mathcal{I}^0(u) := \{i \in \mathcal{I} \mid \langle a_i, u \rangle = b_i\}$$

denote the index set of constraints *active* at u. Clearly $\mathcal{I}^0(u) = \emptyset$ if u lies in the interior of U. An example of a polyhedral constraint set is shown in Figure C.5. The next result shows that in this case, the tangent cone is the set of h in $\mathbb{R}^m$ that satisfy $\langle a_i, h \rangle \leq 0$ for all i in $\mathcal{I}^0(u)$ and the normal cone is the cone generated by the vectors a_i, $i \in \mathcal{I}^0(u)$; each normal h in the normal cone may be expressed as $\sum_{i \in \mathcal{I}^0(u)} \mu_i a_i$ where each $\mu_i \geq 0$.

Proposition C.11 (Representation of tangent and normal cones)**.** *Let $U := \{u \in \mathbb{R}^m \mid \langle a_i, u \rangle \leq b_i, \; i \in \mathcal{I}\}$. Then, for any $u \in U$:*

$$\mathcal{T}_U(u) = \{h \mid \langle a_i, h \rangle \leq 0, \; i \in \mathcal{I}^0(u)\}$$
$$\hat{N}_U(u) = \mathcal{T}_U^*(u) = \text{cone}\{a_i \mid i \in \mathcal{I}^0(u)\}$$

Proof. (i) Suppose h is any vector in $\{h \mid \langle a_i, h\rangle \le 0,\ i \in \mathcal{I}^0(u)\}$. Let the sequences u^ν and λ^ν satisfy $u^\nu = u + \lambda^\nu h$ and $\lambda^\nu \searrow 0$ with λ^0, the first element in the sequence λ^ν, satisfying $u + \lambda^0 h \in U$. It follows that $[u^\nu - u]/\lambda^\nu \equiv h$ so that from Definition C.4, h is tangent to U at u. Hence $\{h \mid \langle a_i, h\rangle \le 0,\ i \in \mathcal{I}^0(u)\} \subset \mathcal{T}_U(u)$. (ii) Conversely, if $h \in \mathcal{T}_U(u)$, then there exist sequences $\lambda^\nu \searrow 0$ and $h^\nu \to h$ such that $\langle a_i, u + \lambda^\nu h^\nu\rangle \le b_i$ for all $i \in \mathcal{I}$, all $\nu \in \mathbb{I}_{\ge 0}$. Since $\langle a_i, u\rangle = b_i$ for all $i \in \mathcal{I}^0(u)$, it follows that $\langle a_i, h^\nu\rangle \le 0$ for all $i \in \mathcal{I}^0(u)$, all $\nu \in \mathbb{I}_{\ge 0}$; taking the limit yields $\langle a_i, h\rangle \le 0$ for all $i \in \mathcal{I}^0(u)$ so that $h \in \{h \mid \langle a_i, h\rangle \le 0,\ i \in \mathcal{I}^0(u)\}$ which proves $\mathcal{T}_U(u) \subset \{h \mid \langle a_i, h\rangle \le 0,\ i \in \mathcal{I}^0(u)\}$. We conclude from (i) and (ii) that $\mathcal{T}_U(u) = \{h \mid \langle a_i, h\rangle \le 0,\ i \in \mathcal{I}^0(u)\}$. That $\hat{N}_U(u) = \mathcal{T}_U^*(u) = \operatorname{cone}\{a_i \mid i \in \mathcal{I}^0(u)\}$ then follows from Proposition C.7 above and Proposition 9 in Appendix A1. ■

The next result follows from Proposition C.5 and Proposition C.7.

Proposition C.12 (Optimality conditions—linear inequalities). *Suppose the function $f(\cdot)$ is convex and differentiable and U is the convex set $\{u \mid Au \le b\}$. Then u is optimal for $\mathbb{P} : \min_u\{f(u) \mid u \in U\}$ if and only if $u \in U$ and*

$$-\nabla f(u) \in \hat{N}_U(u) = \operatorname{cone}\{a_i \mid i \in \mathcal{I}^0(u)\}$$

Corollary C.13 (Optimality conditions—linear inequalities). *Suppose the function $f(\cdot)$ is convex and differentiable and $U = \{u \mid Au \le b\}$. Then u is optimal for $\mathbb{P} : \min_u\{f(u) \mid u \in U\}$ if and only if $Au \le b$ and there exist multipliers $\mu_i \ge 0, i \in \mathcal{I}^0(u)$ satisfying*

$$\nabla f(u) + \sum_{i \in \mathcal{I}^0(u)} \mu_i \nabla g_i(u) = 0 \tag{C.20}$$

where, for each i, $g_i(u) := \langle a_i, u\rangle - b_i$ so that $g_i(u) \le 0$ is the constraint $\langle a_i, u\rangle \le b_i$ and $\nabla g_i(u) = a_i$.

Proof. Since any point $g \in \operatorname{cone}\{a_i \mid i \in \mathcal{I}^0(u)\}$ may be expressed as $g = \sum_{i \in \mathcal{I}^0(u)} \mu_i a_i$ where, for each i, $\mu_i \ge 0$, the condition $-\nabla f(u) \in \operatorname{cone}\{a_i \mid i \in \mathcal{I}^0(u)\}$ is equivalent to the existence of multipliers $\mu_i \ge 0$, $i \in \mathcal{I}^0(u)$ satisfying (C.20). ■

The above results are easily extended if U is defined by linear equality and inequality constraints, i.e., if

$$U := \{\langle a_i, u\rangle \le b_i,\ i \in \mathcal{I},\ \langle c_i, u\rangle = d_i,\ i \in \mathcal{E}\}$$

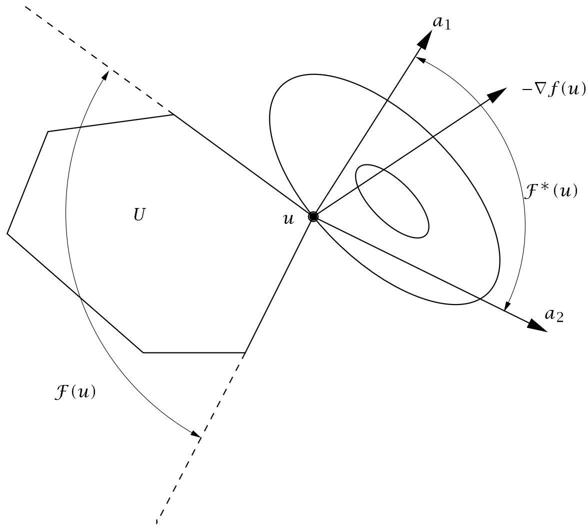


Figure C.5: Condition of optimality.

In this case, at any point $u \in U$, the tangent cone is

$$\mathcal{T}_U(u) = \{h \mid \langle a_i, h\rangle \leq 0,\ i \in \mathcal{I}^0(u),\ \langle c_i, h\rangle = 0,\ i \in \mathcal{E}\}$$

and the normal cone is

$$\hat{N}_U(u) = \{\sum_{i\in\mathcal{I}^0(u)} \lambda_i a_i + \sum_{i\in\mathcal{E}} \mu_i c_i \mid \lambda_i \geq 0\ \forall i \in \mathcal{I}^0(u), \mu_i \in \mathbb{R}\ \forall i \in \mathcal{E}\}$$

With U defined this way, u is optimal for $\min_u\{f(u) \mid u \in U\}$ where $f(\cdot)$ is convex and differentiable if and only if

$$-\nabla f(u) \in \hat{N}_U(u)$$

For each $i \in \mathcal{I}$ let $g_i(u) := \langle a_i, u\rangle - b_i$ and for each $i \in \mathcal{E}$, let $h_i(u) := \langle c_i, u\rangle - d_i$ so that $\nabla g(u_i) = a_i$ and $\nabla h_i = c_i$. It follows from the characterization of $\hat{N}_U(u)$ that u is optimal for $\min_u\{f(u) \mid u \in U\}$ if and only if there exist multipliers $\lambda_i \geq 0,\ i \in \mathcal{I}^0(u)$ and $\mu_i \in \mathbb{R},\ i \in \mathcal{E}$ such that

$$\nabla f(u) + \sum_{i\in\mathcal{I}^0(u)} \mu_i \nabla g_i(u) + \sum_{i\in\mathcal{E}} h_i(u) = 0 \tag{C.21}$$

C.2.4 Nonconvex Problems

We first obtain a necessary condition of optimality for the problem $\min\{f(u) \mid u \in U\}$ where $f(\cdot)$ is differentiable but not necessarily convex and $U \subset \mathbb{R}^m$ is not necessarily convex; this result generalizes the necessary condition of optimality in Proposition C.9.

Proposition C.14 (Necessary condition for nonconvex problem)**.** *A necessary condition for u to be locally optimal for the problem of minimizing a differentiable function $f(\cdot)$ over the set U is*

$$df(u;h) = \langle \nabla f(u), h \rangle \geq 0, \ \ \forall h \in \mathcal{T}_U(u)$$

which is equivalent to the condition

$$-\nabla f(u) \in \hat{N}_U(u)$$

Proof. Suppose, contrary to what we wish to prove, that there exists a $h \in \mathcal{T}_U(u)$ and a $\delta > 0$ such that $\langle \nabla f(u), h \rangle = -\delta < 0$. Because $h \in \mathcal{T}_U(u)$, there exist sequences $h^\nu \xrightarrow[U]{} h$ and $\lambda^\nu \searrow 0$ such that $u^\nu := u + \lambda^\nu h^\nu$ converges to u and satisfies $u^\nu \in U$ for all $\nu \in \mathbb{I}_{\geq 0}$. Then

$$f(u^\nu) - f(u) = \langle \nabla f(u), \lambda^\nu h^\nu \rangle + o(\lambda^\nu |h^\nu|)$$

Hence

$$[f(u^\nu) - f(u)]/\lambda^\nu = \langle \nabla f(u), h^\nu \rangle + o(\lambda^\nu)/\lambda^\nu$$

where we make use of the fact that $|h^\nu|$ is bounded for ν sufficiently large. It follows that

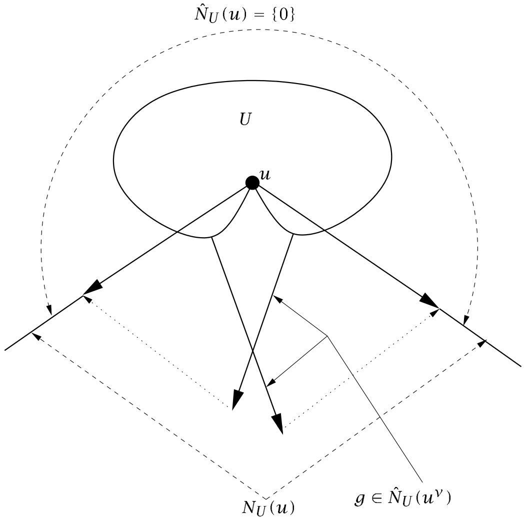
$$[f(u^\nu) - f(u)]/\lambda^\nu \to \langle \nabla f(u), h \rangle = -\delta$$

so that there exists a finite integer j such that $f(u^j) - f(u) \leq -\lambda^j \delta/2 < 0$ which contradicts the local optimality of u. Hence $\langle \nabla f(u), h \rangle \geq 0$ for all $h \in \mathcal{T}_U(u)$. That $-\nabla f(u) \in \hat{N}_U(u)$ follows from Proposition C.7. ■

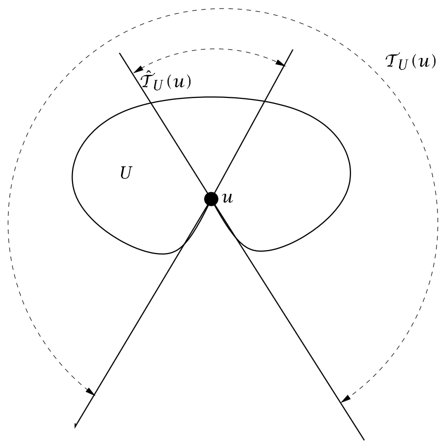
A more concise proof proceeds as follows Rockafellar and Wets (1998). Since $f(v) - f(u) = \langle \nabla f(u), v - u \rangle + o(|v - u|)$ it follows that $\langle -\nabla f(u), v - u \rangle = o(|v - u|) - (f(v) - f(u))$. Because u is locally optimal, $f(v) - f(u) \geq 0$ for all v in the neighborhood of u so that $\langle -\nabla f(u), v - u \rangle \leq o(|v - u|)$ which, by (C.15), is the definition of a normal vector. Hence $-\nabla f(u) \in \hat{N}_U(u)$.

C.2.5 Tangent and Normal Cones

The material in this section is *not* required for Chapters 1-7; it is presented merely to show that alternative definitions of tangent and normal cones are useful in more complex situations than those considered



(a) Normal cones.



(b) Tangent cones.

Figure C.6: Tangent and normal cones.

above. Thus, the normal and tangent cones defined in C.2.1 have some limitations when U is not convex or, at least, not similar to the constraint set illustrated in Figure C.4. Figure C.6 illustrates the type of difficulty that may occur. Here the tangent cone $\mathcal{T}_U(u)$ is not convex, as shown in Figure C.6(b), so that the associated normal cone

$\hat{N}_U(u) = \mathcal{T}_U(u)^* = \{0\}$. Hence the necessary condition of optimality of u for the problem of minimizing a differentiable function $f(\cdot)$ over U is $\nabla f(u) = 0$; the only way a *differentiable* function $f(\cdot)$ can achieve a minimum over U at u is for the condition $\nabla f(u) = 0$ to be satisfied. Alternative definitions of normality and tangency are sometimes necessary. In Rockafellar and Wets (1998), a vector $g \in \hat{N}_U(u)$ is normal in the *regular* sense; a normal in the *general* sense is then defined by:

Definition C.15 (General normal)**.** A vector g is normal to U at u in the general sense if there exist sequences $u^\nu \xrightarrow[U]{} u$ and $g^\nu \to g$ where $g^\nu \in \hat{N}_U(u^\nu)$ for all ν; $N_U(u)$ is the set of all general normal vectors.

The cone $N_U(u)$ of general normal vectors is illustrated in Figure C.6(a); here the cone $N_U(u)$ is the union of two distinct cones each having form $\{\alpha g \mid \alpha \geq 0\}$. Also shown in Figure C.6(a) are single elements of two sequences g^ν in $\hat{N}_U(u^\nu)$ converging to $N_U(u)$. Counter intuitively, the general normal vectors in this case point into the interior of U. Associated with $N_U(u)$ is the set $\hat{\mathcal{T}}_U(u)$ of regular tangents to U at u defined, when U is locally closed,[3] in (Rockafellar and Wets, 1998, Theorem 6.26) by:

Definition C.16 (General tangent)**.** Suppose U is locally closed at u. A vector h is tangent to U at u in the regular sense if, for all sequences $u^\nu \xrightarrow[U]{} u$, there exists a sequence $h^\nu \to h$ that satisfies $h^\nu \in \mathcal{T}_u(u^\nu)$ for all ν; $\hat{\mathcal{T}}_U(u)$ is the set of all regular tangent vectors to U at u.

Alternatively, a vector h is tangent to U at u in the regular sense if, for all sequences $u^\nu \xrightarrow[U]{} u$ and $\lambda^\nu \searrow 0$, there exists a sequence $h^\nu \to h$ satisfying $u^\nu + \lambda^\nu h^\nu \in U$ for all $\nu \in \mathbb{I}_{\geq 0}$. The cone of regular tangent vectors for the example immediately above is shown in Figure C.6(b). The following result is proved in Rockafellar and Wets (1998), Theorem 6.26:

Proposition C.17 (Set of regular tangents is closed convex cone)**.** *At any $u \in U$, the set $\hat{\mathcal{T}}_U(u)$ of regular tangents to U at u is a closed convex cone with $\hat{\mathcal{T}}_U(u) \subset \mathcal{T}_U(u)$. Moreover, if U is locally closed at u, then $\hat{\mathcal{T}}_U(u) = N_U(u)^*$.*

[3] A set U is locally closed at a point u if there exists a closed neighborhood $\mathcal{N}$ of u such that $U \cap \mathcal{N}$ is closed; U is locally closed if it is locally closed at all u.

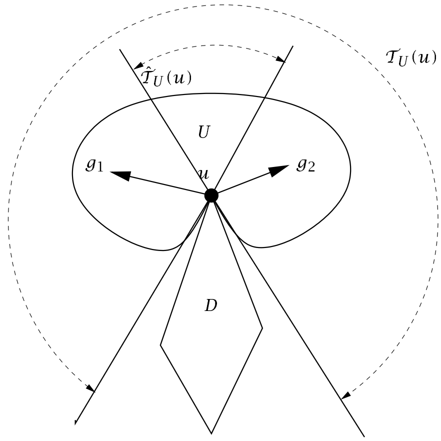


Figure C.7: Condition of optimality.

Figure C.7 illustrates some of these results. In Figure C.7, the constant cost contour $\{v \mid f(v) = f(u)\}$ of a *nondifferentiable* cost function $f(\cdot)$ is shown together with a sublevel set D passing through the point u: $f(v) \le f(u)$ for all $v \in D$. For this example, $df(u;h) = \max\{\langle g_1, h\rangle, \langle g_2, h\rangle\}$ where g_1 and g_2 are normals to the level set of $f(\cdot)$ at u so that $df(u;h) \ge 0$ for all $h \in \hat{\mathcal{T}}_U(u)$, a necessary condition of optimality; on the other hand, there exist $h \in \mathcal{T}_U(u)$ such that $df(u;h) < 0$. The situation is simpler if the constraint set U is *regular* at u.

Definition C.18 (Regular set)**.** A set U is regular at a point $u \in U$ in the sense of Clarke if it is locally closed at u and if $N_U(u) = \hat{N}_U(u)$ (all normal vectors at u are regular).

The following consequences of Clarke regularity are established in Rockafellar and Wets (1998), Corollary 6.29:

Proposition C.19 (Conditions for regular set)**.** *Suppose U is locally closed at $u \in U$. Then U is regular at u is equivalent to each of the following.*

(a) $N_U(u) = \hat{N}_U(u)$ *(all normal vectors at u are regular).*

(b) $\mathcal{T}_U(u) = \hat{\mathcal{T}}_U(u)$ *(all tangent vectors at u are regular).*

(c) $N_U(u) = \mathcal{T}_U(u)^*$.

(d) $\mathcal{T}_U(u) = N_U(u)^*$.

(e) $\langle g, h\rangle \le 0$ *for all* $h \in \mathcal{T}_U(u)$, *all* $g \in N_U(u)$.

It is shown in Rockafellar and Wets (1998) that if U is regular at u and a constraint qualification is satisfied, then a necessary condition of optimality, similar to (C.21), may be obtained. To obtain this result, we pursue a slightly different route in Sections C.2.6 and C.2.7.

C.2.6 Constraint Set Defined by Inequalities

We now consider the case when the set U is specified by a set of differentiable inequalities:

$$U := \{u \mid g_i(u) \le 0 \ \forall i \in \mathcal{I}\} \tag{C.22}$$

where, for each $i \in \mathcal{I}$, the function $g_i : \mathbb{R}^m \to \mathbb{R}$ is differentiable. For each $u \in U$

$$\mathcal{I}^0(u) := \{i \in \mathcal{I} \mid g_i(u) = 0\}$$

is the index set of active constraints. For each $u \in U$, the set $\mathcal{F}_U(u)$ of feasible variations for the *linearized* set of inequalities; $\mathcal{F}_U(u)$ is defined by

$$\mathcal{F}_U(u) := \{h \mid \langle \nabla g_i(u), h\rangle \le 0 \ \forall i \in \mathcal{I}^0(u)\} \tag{C.23}$$

The set $\mathcal{F}_U(u)$ is a closed, convex cone and is called a cone of first order feasible variations in Bertsekas (1999) because h is a descent direction for $g_i(u)$ for all $i \in \mathcal{I}^0(u)$, i.e., $g_i(u + \lambda h) \le 0$ for all λ sufficiently small. When U is polyhedral, the case discussed in C.2.3, $g_i(u) = \langle a_i, u\rangle - b_i$ and $\nabla g_i(u) = a_i$ so that $\mathcal{F}_U(u) = \{h \mid \langle a_i, h\rangle \le 0 \ \forall i \in \mathcal{I}^0(u)\}$ which was shown in Proposition C.11 to be the tangent cone $\mathcal{T}_U(u)$. An important question whether $\mathcal{F}_U(u)$ is the tangent cone $\mathcal{T}_U(u)$ for a wider class of problems because, if $\mathcal{F}_U(u) = \mathcal{T}_U(u)$, a condition of optimality of the form in (C.20) may be obtained. In the example in Figure C.8, $\mathcal{F}_U(u)$ is the horizontal axis $\{h \in \mathbb{R}^2 \mid h_2 = 0\}$ whereas $\mathcal{T}_U(u)$ is the half-line $\{h \in \mathbb{R}^2 \mid h_1 \ge 0, \ h_2 = 0\}$ so that in this case, $\mathcal{F}_U(u) \ne \mathcal{T}_U(u)$. While $\mathcal{F}_U(u)$ is always convex, being the intersection of a set of half-spaces, the tangent cone $\mathcal{T}_U(u)$ is not necessarily convex as Figure C.6b shows. The set U is said to be *quasiregular* at $u \in U$ if $\mathcal{F}_U(u) = \mathcal{T}_U(u)$ is which case u is said to be a quasiregular point Bertsekas (1999). The next result, due to Bertsekas (1999), shows that $\mathcal{F}_U(u) = \mathcal{T}_U(u)$, i.e., U is quasiregular at u, when a certain constraint qualification is satisfied.

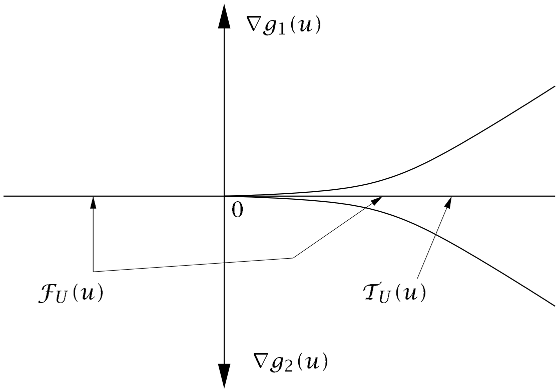


Figure C.8: $\mathcal{F}_U(u) \neq \mathcal{T}_U(u)$.

Proposition C.20 (Quasiregular set)**.** *Suppose $U := \{u \mid g_i(u) \leq 0 \; \forall i \in \mathcal{I}\}$ where, for each $i \in \mathcal{I}$, the function $g_i : \mathbb{R}^m \to \mathbb{R}$ is differentiable. Suppose also that $u \in U$ and that there exists a vector $\bar{h} \in \mathcal{F}_U(u)$ such that*

$$\langle \nabla g_i(u), \bar{h} \rangle < 0, \;\; \forall \; i \in \mathcal{I}^0(u) \tag{C.24}$$

Then

$$\mathcal{T}_U(u) = \mathcal{F}_U(u)$$

i.e., U is quasiregular at u.

Equation (C.24) is the constraint qualification; it can be seen that it precludes the situation shown in Figure C.8.

Proof. It follows from the definition (C.23) of $\mathcal{F}_U(u)$ and the constraint qualification (C.24) that:

$$\langle \nabla g_i(u), h + \alpha(\bar{h} - h) \rangle < 0, \;\; \forall h \in \mathcal{F}_U(u), \alpha \in (0,1], i \in \mathcal{I}^0(u)$$

Hence, for all $h \in \mathcal{F}_U(u)$, all $\alpha \in (0,1]$, there exists a vector $h_\alpha := h + \alpha(\bar{h} - h)$, in $\mathcal{F}_U(u)$ satisfying $\langle \nabla g_i(u), h_\alpha \rangle < 0$ for all $i \in \mathcal{I}^0(u)$. Assuming for the moment that $h_\alpha \in \mathcal{T}_U(u)$ for all $\alpha \in (0,1]$, it follows, since $h_\alpha \to h$ as $\alpha \to 0$ and $\mathcal{T}_U(u)$ is closed, that $h \in \mathcal{T}_U(u)$, thus proving $\mathcal{F}_U(u) \subset \mathcal{T}_U(u)$. It remains to show that h_α is tangent to U at u. Consider the sequences h^ν and $\lambda^\nu \searrow 0$ where $h^\nu := h_\alpha$ for all $\nu \in \mathbb{I}_{\geq 0}$. There exists a $\delta > 0$ such that $\langle \nabla g_i(u), h_\alpha \rangle \leq -\delta$ for all $i \in \mathcal{I}^0(u)$ and $g_i(u) \leq -\delta$ for all $i \in \mathcal{I} \setminus \mathcal{I}^0(u)$. Since

$$g_i(u + \lambda^\nu h^\nu) = g_i(u) + \lambda^\nu \langle \nabla g_i(u), h_\alpha \rangle + o(\lambda^\nu) \leq -\lambda^\nu \delta + o(\lambda^\nu)$$

for all $i \in \mathcal{I}^0(u)$, it follows that there exists a finite integer N such that $g_i(u + \lambda^\nu h^\nu) \le 0$ for *all* $i \in \mathcal{I}$, all $\nu \ge N$. Since the sequences $\{h^\nu\}$ and $\{\lambda^\nu\}$ for all $\nu \ge N$ satisfy $h^\nu \to h_\alpha$, $\lambda^\nu \searrow 0$ and $u + \lambda^\nu h^\nu \in U$ for all $i \in \mathcal{I}$, it follows that $h_\alpha \in \mathcal{T}_U(u)$, thus completing the proof that $\mathcal{F}_U(u) \subset \mathcal{T}_U(u)$.

Suppose now that $h \in \mathcal{T}_U(u)$. There exist sequences $h^\nu \to h$ and $\lambda^\nu \to 0$ such that $u + \lambda^\nu h^\nu \in U$ so that $g(u + \lambda^\nu h^\nu) \le 0$ for all $\nu \in \mathbb{I}_{\ge 0}$. Since $g(u + \lambda^\nu h^\nu) = g(u) + \langle \nabla g_j(u), \lambda^\nu h^\nu \rangle + o(\lambda^\nu |h^\nu|) \le 0$, it follows that $\langle \nabla g_j(u), \lambda^\nu h^\nu \rangle + o(\lambda^\nu) \le 0$ for all $j \in \mathcal{I}^0(u)$, all $\nu \in \mathbb{I}_{\ge 0}$. Hence $\langle \nabla g_j(u), h^\nu \rangle + o(\lambda^\nu)/\lambda^\nu \le 0$ for all $j \in \mathcal{I}^0(u)$, all $\nu \in \mathbb{I}_{\ge 0}$. Taking the limit yields $\langle \nabla g_j(u), h^\nu \rangle \le 0$ for all $j \in \mathcal{I}^0(u)$ so that $h \in \mathcal{F}_U(u)$ which proves $\mathcal{T}_U(u) \subset \mathcal{F}_U(u)$. Hence $\mathcal{T}_U(u) = \mathcal{F}_U(u)$. ∎

The existence of a $\bar{h}$ satisfying (C.24) is, as we have noted above, a constraint qualification. If u is locally optimal for the inequality constrained optimization problem of minimizing a differentiable function $f(\cdot)$ over the set U defined in (C.22) and, if (C.24) is satisfied thereby ensuring that $\mathcal{T}_U(u) = \mathcal{F}_U(u)$, then a condition of optimality of the form (C.20) may be easily obtained as shown in the next result.

Proposition C.21 (Optimality conditions nonconvex problem)**.** *Suppose u is locally optimal for the problem of minimizing a differentiable function $f(\cdot)$ over the set U defined in* (C.22) *and that $\mathcal{T}_U(u) = \mathcal{F}_U(u)$. Then*

$$-\nabla f(u) \in \text{cone}\{\nabla g_i(u) \mid i \in \mathcal{I}^0(u)\}$$

and there exist multipliers $\mu_i \ge 0$, $i \in \mathcal{I}^0(u)$ satisfying

$$\nabla f(u) + \sum_{i \in \mathcal{I}^0(u)} \mu_i \nabla g_i(u) = 0 \tag{C.25}$$

Proof. It follows from Proposition C.14 that $-\nabla f(u) \in \hat{\mathcal{N}}_U(u)$ and from Proposition C.7 that $\hat{\mathcal{N}}_U(u) = \mathcal{T}_U^*(u)$. But, by hypothesis, $\mathcal{T}_U(u) = \mathcal{F}_U(u)$ so that $\hat{\mathcal{N}}_U(u) = \mathcal{F}_U^*(u)$, the polar cone of $\mathcal{F}_U(u)$. It follows from (C.23) and the definition of a polar cone, given in Appendix A1, that

$$\mathcal{F}_U^*(u) = \text{cone}\{\nabla g_i(u) \mid i \in\mid I^0(u)\}$$

Hence

$$-\nabla f(u) \in \text{cone}\{\nabla g_i(u) \mid i \in \mathcal{I}^0(u)\}$$

The existence of multipliers μ_i satisfying (C.25) follows from the definition of a cone generated by $\{\nabla g_i(u) \mid i \in \mathcal{I}^0(u)\}$. ∎

C.2.7 Constraint Set Defined by Equalities and Inequalities

Finally, we consider the case when the set U is specified by a set of differentiable equalities *and* inequalities:

$$U := \{u \mid g_i(u) \leq 0 \ \forall i \in \mathcal{I},\ h_i(u) = 0 \ \forall i \in \mathcal{E}\}$$

where, for each $i \in \mathcal{I}$, the function $g_i : \mathbb{R}^m \to \mathbb{R}$ is differentiable and for each $i \in \mathcal{E}$, the function $h_i : \mathbb{R}^m \to \mathbb{R}$ is differentiable. For each $u \in U$

$$\mathcal{I}^0(u) := \{i \in \mathcal{I} \mid g_i(u) = 0\}$$

the index set of active inequality constraints is defined as before. We wish to obtain necessary conditions for the problem of minimizing a differentiable function $f(\cdot)$ over the set U. The presence of equality constraints makes this objective more difficult than for the case when U is defined merely by differentiable inequalities. The result we wish to prove is a natural extension of Proposition C.21 in which the equality constraints are included in the set of active constraints:

Proposition C.22 (Fritz-John necessary conditions)**.** *Suppose u is a local minimizer for the problem of minimizing $f(u)$ subject to the constraint $u \in U$ where U is defined in* (C.22)*. Then there exist multipliers μ_0, μ_i, $i \in \mathcal{I}$ and λ_i, $i \in \mathcal{E}$, not all zero, such that*

$$\mu_0 \nabla f(u) + \sum_{i \in \mathcal{I}} \mu_i \nabla g_i(u) + \sum_{j \in \mathcal{E}} \lambda_j \nabla h_j(u) = 0 \tag{C.26}$$

and

$$\mu_i g_i(u) = 0 \ \forall i \in \mathcal{I}$$

where $\mu_0 \geq 0$ and $\mu_i \geq 0$ for all $i \in \mathcal{I}^0$.

The condition $\mu_i g_i(u) = 0$ for all $i \in \mathcal{I}$ is known as the *complementarity* conditions and implies $\mu_i = 0$ for all $i \in \mathcal{I}$ such that $g_i(u) < 0$. If $\mu_0 > 0$, then (C.26) may be normalized by dividing each term by μ_0 yielding the more familiar expression

$$\nabla f(u) + \sum_{i \in \mathcal{I}} \mu_i \nabla g_i(u) + \sum_{j \in \mathcal{E}} \nabla h_j(u) = 0$$

We return to this point later. Perhaps the simplest method for proving Proposition C.22 is the penalty approach adopted by Bertsekas (1999), Proposition 3.3.5. We merely give an outline of the proof. The constrained problem of minimizing $f(v)$ over U is approximated, for each

$k \in \mathbb{I}_{\geq 0}$ by a penalized problem defined below; as k increases the penalized problem becomes a closer approximation to the constrained problem. For each $i \in \mathcal{I}$, we define

$$g_i^+(v) := \max\{g_i(v), 0\}$$

For each k, the penalized problem $\mathbb{P}^k$ is then defined as the problem of minimizing $F^k(v)$ defined by

$$F^k(v) := f(v) + (k/2)\sum_{i\in\mathcal{I}}(g_i^+(v))^2 + (k/2)\sum_{j\in\mathcal{E}}(h_j(v))^2 + (1/2)|v-u|^2$$

subject to the constraint

$$S := \{v \mid |v-u| \leq \varepsilon\}$$

where $\epsilon > 0$ is such that $f(u) \leq f(v)$ for all v in $S \cap U$. Let v^k denote the solution of $\mathbb{P}^k$. Bertsekas shows that $v^k \to u$ as $k \to \infty$ so that for all k sufficiently large, v^k lies in the interior of S and is, therefore, the unconstrained minimizer of $F^k(v)$. Hence for each k sufficiently large, v^k satisfies $\nabla F^k(v^k) = 0$, or

$$\nabla f(v^k) + \sum_{i\in\mathcal{I}} \bar{\mu}_i^k \nabla g(v^k) + \sum_{i\in\mathcal{E}} \bar{\lambda}_i^k \nabla h(v^k) = 0 \tag{C.27}$$

where

$$\bar{\mu}_i^k := k g_i^+(v^k), \quad \bar{\lambda}_i^k := k h_i(v^k)$$

Let μ^k denote the vector with elements μ_i^k, $i \in \mathcal{I}$ and λ^k the vector with elements λ_i^k, $k \in \mathcal{E}$. Dividing (C.27) by δ^k defined by

$$\delta^k := [1 + |\mu^k|^2 + |\lambda^k|^2]^1/2$$

yields

$$\mu_0^k \nabla f(v^k) + \sum_{i\in\mathcal{I}} \mu_i^k \nabla g(v^k) + \sum_{j\in\mathcal{E}} \lambda_j^k \nabla h(v^k) = 0$$

where

$$\mu_0^k := \bar{\mu}_i^k/\delta^k, \quad \mu_i^k := \bar{\mu}_i^k/\delta^k, \quad \lambda_j^k := \bar{\lambda}_i^k/\delta^k$$

and

$$(\mu_0^k)^2 + |\mu^k|^2 + |\lambda^k|^2 = 1$$

Because of the last equation, the sequence $\left(\mu_0^k, \mu^k, \lambda^k\right)$ lies in a compact set, and therefore has a subsequence, indexed by $K \subset \mathbb{I}_{\geq 0}$, converging to some limit (μ_0, μ, λ) where μ and λ are vectors whose elements are, respectively, $\mu_i,\ i \in \mathcal{I}$ and $\lambda_j,\ j \in \mathcal{E}$. Because $v^k \to u$ as $k \in K$ tends to infinity, it follows from (C.27) that

$$\mu_0 \nabla f(u) + \sum_{i \in \mathcal{I}} \mu_i \nabla g_i(u) + \sum_{j \in \mathcal{E}} \lambda_j \nabla h_j(u) = 0$$

To prove the complementarity condition, suppose, contrary to what we wish to prove, that there exists a $i \in \mathcal{I}$ such that $g_i(u) < 0$ but $\mu_i > 0$. Since $\mu_i^k \to \mu_i > 0$ and $g_i(v^k) \to g_i(u)$ as $k \to \infty$, $k \in K$, it follows that $\mu_i \mu_i^k > 0$ for all $k \in K$ sufficiently large. But $\mu_i^k = \bar{\mu}_i^k / \delta^k = k g_i^+(v^k)/\delta^k$ so that $\mu_i \mu_i^k > 0$ implies $\mu_i g_i^+(v^k) > 0$ which in turn implies $g_i^+(v^k) = g_i(v^k) > 0$ for all $k \in K$ sufficiently large. This contradicts the fact that $g_i(v^k) \to g_i(u) < 0$ as $k \to \infty$, $k \in K$. Hence we must have $g_i(u) = 0$ for all $i \in \mathcal{I}$ such that $\mu_i > 0$.

The Fritz-John condition in Proposition C.22 is known as the Karush-Kuhn-Tucker (KKT) condition if $\mu_0 > 0$; if this is the case, μ_0 may be normalized to $\mu_0 = 1$. A constraint qualification is required for the Karush-Kuhn-Tucker condition to be a necessary condition of optimality for the optimization problem considered in this section. A simple constraint qualification is linear independence of $\{\nabla g_i(u),\ i \in \mathcal{I}^0(u),\ \nabla h_j(u),\ j \in \mathcal{E}\}$ at a local minimizer u. For, if u is a local minimizer and $\mu_0 = 0$, then the Fritz-John condition implies that $\sum_{i \in \mathcal{I}^0(u)} \mu_i \nabla g_i(u) + \sum_{j \in \mathcal{E}} \lambda_j \nabla h_j(u) = 0$ which contradicts the linear independence of $\{\nabla g_i(u),\ i \in \mathcal{I}^0(u),\ \nabla h_j(u),\ j \in \mathcal{E}\}$ since not all the multipliers are zero. Another constraint qualification, used in Propositions C.20 and C.21 for an optimization problem in which the constraint set is $U := \{u \mid g_i(u) \leq 0,\ i \in \mathcal{I}\}$, is the existence of a vector $\bar{h}(u) \in \mathcal{F}_U(u)$ such that $\langle \nabla g_i(u), \bar{h} \rangle < 0$ for all $i \in \mathcal{I}^0(u)$; this condition ensures $\mu_0 = 1$ in C.25. Many other constraint qualifications exist; see, for example, Bertsekas (1999), Chapter 3.

C.3 Set-Valued Functions and Continuity of Value Function

A set-valued function $U(\cdot)$ is one for which, for each value of x, $U(x)$ is a set; these functions are encountered in parametric programming. For example, in the problem $\mathbb{P}(x) : \inf_u \{f(x,u) \mid u \in U(x)\}$ (which has the same form as an optimal control problem in which x is the state and u is a control sequence), the constraint set U is a set-valued function of the state. The solution to the problem $\mathbb{P}(x)$ (the value of u that achieves the minimum) can also be set-valued. It is important to

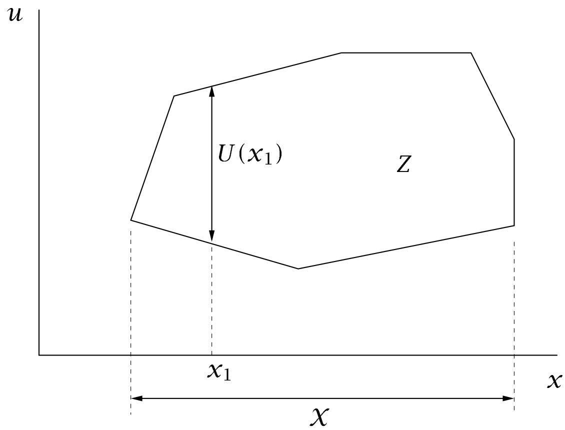
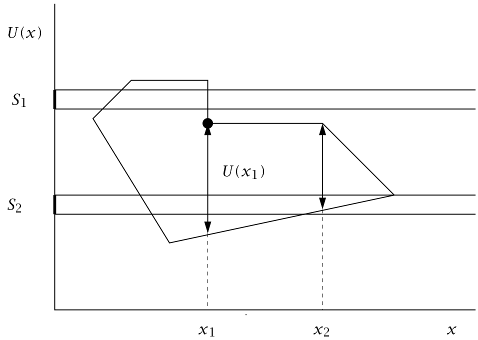


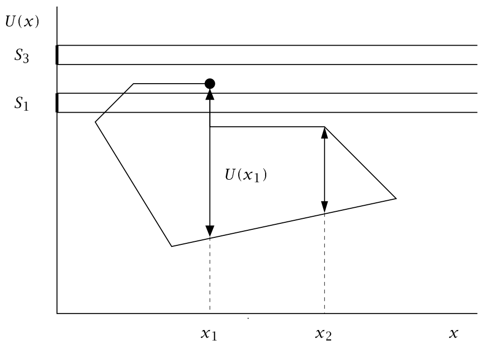
Figure C.9: Graph of set-valued function $U(\cdot)$.

know how smoothly these set-valued functions vary with the parameter x. In particular, we are interested in the continuity properties of the value function $x \mapsto f^0(x) = \inf_u\{f(x,u) \mid u \in U(x)\}$ since, in optimal control problems we employ the value function as a Lyapunov function and robustness depends, as we have discussed earlier, on the continuity of the Lyapunov function. Continuity of the value function depends, in turn, on continuity of the set-valued constraint set $U(\cdot)$. We use the notation $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ to denote the fact that $U(\cdot)$ maps points in $\mathbb{R}^n$ into subsets of $\mathbb{R}^m$.

The *graph* of a set-valued functions is often a useful tool. The graph of $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is defined to be the set $Z := \{(x,u) \in \mathbb{R}^n \times \mathbb{R}^m \mid u \in U(x)\}$; the *domain* of the set-valued function U is the set $\mathcal{X} := \{x \in \mathbb{R}^n \mid U(x) \neq \emptyset\} = \{x \in \mathbb{R}^n \mid \exists u \in \mathbb{R}^m$ such that $(x,u) \in Z\}$; clearly $\mathcal{X} \subset \mathbb{R}^n$. Also $\mathcal{X}$ is the *projection* of the set $Z \subset \mathbb{R}^n \times \mathbb{R}^m$ onto $\mathbb{R}^n$, i.e., $(x,u) \in Z$ implies $x \in \mathcal{X}$. An example is shown in Figure C.9. In this example, $U(x)$ varies continuously with x. Examples in which $U(\cdot)$ is discontinuous are shown in Figure C.10. In Figure C.10(a), the set $U(x)$ varies continuously if x increases from its initial value of x_1, but jumps to a much larger set if x decreases an infinitesimal amount (from its initial value of x_1); this is an example of a set-valued function that is inner semicontinuous at x_1. In Figure C.10(b), the set $U(x)$ varies continuously if x decreases from its initial value of x_1, but jumps to a much smaller set if x increases an infinitesimal amount (from its initial value of x_1); this is an example of a set-valued function that is



(a) Inner semicontinuous set-valued function.



(b) Outer semicontinuous set-valued function.

Figure C.10: Graphs of discontinuous set-valued functions.

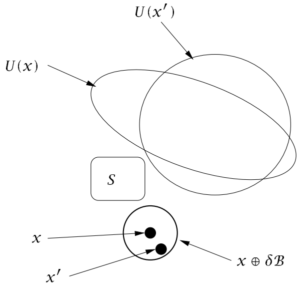
outer semicontinuous at x_1. The set-valued function is continuous at x_2 where it is both outer and inner semicontinuous.

We can now give precise definitions of inner and outer semicontinuity.

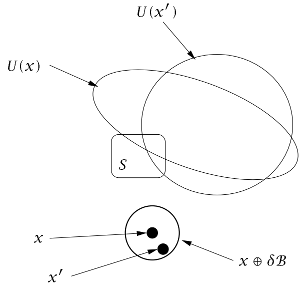
C.3.1 Outer and Inner Semicontinuity

The concepts of inner and outer semicontinuity were introduced by Rockafellar and Wets (1998, p. 144) to replace earlier definitions of lower and upper semicontinuity of set-valued functions. This section is based on the useful summary provided by Polak (1997, pp. 676-682).

Definition C.23 (Outer semicontinuous function)**.** A set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is said to be outer semicontinuous (osc) at x if $U(x)$



(a) Outer semicontinuity.



(b) Inner semicontinuity.

Figure C.11: Outer and inner semicontinuity of $U(\cdot)$.

is closed and if, for every compact set S such that $U(x) \cap S = \emptyset$, there exists a $\delta > 0$ such that $U(x') \cap S = \emptyset$ for all $x' \in x \oplus \delta\mathcal{B}$.[4] The set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is outer semicontinuous if it is outer semicontinuous at each $x \in \mathbb{R}^n$.

Definition C.24 (Inner semicontinuous function)**.** A set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is said to be inner semicontinuous (isc) at x if, for every open set S such that $U(x) \cap S \neq \emptyset$, there exists a $\delta > 0$ such that $U(x') \cap S \neq \emptyset$ for all $x' \in x \oplus \delta\mathcal{B}$. The set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is inner semicontinuous if it is inner semicontinuous at each $x \in \mathbb{R}^n$.

These definitions are illustrated in Figure C.11. Roughly speaking, a set-valued function that is outer semicontinuous at x cannot explode as x changes to x' arbitrarily close to x; similarly, a set-valued function that is inner semicontinuous at x cannot collapse as x changes to x' arbitrarily close to x.

Definition C.25 (Continuous function)**.** A set-valued function is continuous (at x) if it is both outer and inner continuous (at x).

If we return to Figure C.10(a) we see that $S_1 \cap U(x_1) = \emptyset$ but $S_1 \cap U(x) \neq \emptyset$ for x infinitesimally less than x_1 so that $U(\cdot)$ is not outer semicontinuous at x_1. For all S_2 such that $S_2 \cap U(x_1) \neq \emptyset$, however, $S_2 \cap U(x) \neq \emptyset$ for all x in a sufficiently small neighborhood of x_1 so that $U(\cdot)$ is inner semicontinuous at x_1. If we turn to Figure C.10(b) we see that $S_1 \cap U(x_1) \neq \emptyset$ but $S_1 \cap U(x) = \emptyset$ for x infinitesimally greater than x_1 so that in this case $U(\cdot)$ is not inner semicontinuous at x_1. For all S_3 such that $S_3 \cap U(x_1) = \emptyset$, however, $S_3 \cap U(x) = \emptyset$ for

[4]Recall that $\mathcal{B} := \{x \mid |x| \leq 1\}$ is the closed unit ball in $\mathbb{R}^n$.

all x in a sufficiently small neighborhood of x_1 so that $U(\cdot)$ is outer semicontinuous at x_1.

The definitions of outer and inner semicontinuity may be interpreted in terms of infinite sequences (Rockafellar and Wets, 1998, p. 152), (Polak, 1997, pp. 677-678).

Theorem C.26 (Equivalent conditions for outer and inner semicontinuity)**.**

(a) A set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is outer semicontinuous at x if and only if for every infinite sequence (x_i) converging to x, any accumulation point[5] *u of any sequence (u_i), satisfying $u_i \in U(x_i)$ for all i, lies in $U(x)$ ($u \in U(x)$).*

(b) A set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is inner semicontinuous at x if and only if for every $u \in U(x)$ and for every infinite sequence (x_i) converging to x, there exists an infinite sequence (u_i), satisfying $u_i \in U(x_i)$ for all i, that converges to u.

Proofs of these results may be found in Rockafellar and Wets (1998); Polak (1997). Another result that we employ is:

Proposition C.27 (Outer semicontinuity and closed graph)**.** *A set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is outer semicontinuous in its domain if and only if its graph Z is closed in $\mathbb{R}^n \times \mathbb{R}^m$.*

Proof. Since $(x, u) \in Z$ is equivalent to $u \in U(x)$, this result is a direct consequence of the Theorem C.26. ∎

In the above discussion we have assumed, as in Polak (1997), that $U(x)$ is defined everywhere in $\mathbb{R}^n$; in constrained parametric optimization problems, however, $U(x)$ is defined on $\mathcal{X}$, a closed subset of $\mathbb{R}^n$; see Figure C.9. Only minor modifications of the above definitions are then required. In definitions C.23 and C.24 we replace the closed set $\delta\mathcal{B}$ by $\delta\mathcal{B} \cap \mathcal{X}$ and in Theorem C.26 we replace "every infinite sequence (in $\mathbb{R}^n$)" by "every infinite sequence in $\mathcal{X}$." In effect, we are replacing the topology of $\mathbb{R}^n$ by its topology relative to $\mathcal{X}$.

C.3.2 Continuity of the Value Function

Our main reason for introducing set-valued functions is to provide us with tools for analyzing the continuity properties of the value function and optimal control law in constrained optimal control problems.

[5]Recall, u is the limit of (u_i) if $u_i \to u$ as $i \to \infty$; u is an accumulation point of (u_i) if it is the limit of a subsequence of (u_i).

These problems have the form

$$V^0(x) = \min\{V(x,u) \mid u \in U(x)\} \tag{C.28}$$

$$u^0(x) = \arg\min\{V(x,u) \mid u \in U(x)\} \tag{C.29}$$

where $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is a set-valued function and $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous; in optimal control problems arising from MPC, u should be replaced by $\mathbf{u} = (u(0), u(1), \ldots, u(N-1))$ and m by Nm. We are interested in the continuity properties of the value function $V^0 : \mathbb{R}^n \to \mathbb{R}$ and the control law $u^0 : \mathbb{R}^n \to \mathbb{R}^m$; the latter may be set-valued (if the minimizer in (C.28) is not unique).

The following max problem has been extensively studied in the literature

$$\phi^0(x) = \max\{\phi(x,u) \mid u \in U(x)\}$$

$$\mu^0(x) = \arg\max\{\phi(x,u) \mid u \in U(x)\}$$

If we define $\phi(\cdot)$ by $\phi(x,u) := -V(x,u)$, we see that $\phi^0(x) = -V^0(x)$ and $\mu^0(x) = u^0(x)$ so that we can obtain the continuity properties of $V^0(\cdot)$ and $u^0(\cdot)$ from those of $\phi^0(\cdot)$ and $\mu^0(\cdot)$ respectively. Using this transcription and Corollary 5.4.2 and Theorem 5.4.3 in Polak (1997) we obtain the following result:

Theorem C.28 (Minimum theorem)**.** *Suppose that $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous, that $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is continuous, compact-valued and satisfies $U(x) \subset \mathbb{U}$ for all $x \in \mathcal{X}$ where $\mathbb{U}$ is compact. Then $V^0(\cdot)$ is continuous and $u^0(\cdot)$ is outer semicontinuous. If, in addition, $u^0(x) = \{\mu^0(x)\}$ (there is a unique minimizer $\mu^0(x)$), then $\mu^0(\cdot)$ is continuous.*

It is unfortunately the case, however, that due to state constraints, $U(\cdot)$ is often not continuous in constrained optimal control problems. If $U(\cdot)$ is constant, which is the case in optimal control problem if state or mixed control-state constraints are absent, then, from the above results, the value function $V^0(\cdot)$ is continuous. Indeed, under slightly stronger assumptions, the value function is Lipschitz continuous.

Lipschitz continuity of the value function. If we assume that $V(\cdot)$ is Lipschitz continuous and that $U(x) \equiv U$, we can establish Lipschitz continuity of $V^0(\cdot)$. Interestingly the result does not require, nor does it imply, Lipschitz continuity of the minimizer $u^0(\cdot)$.

Theorem C.29 (Lipschitz continuity of the value function, constant U)**.** *Suppose that $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is Lipschitz continuous on bounded sets*[6] *and that $U(x) \equiv U$ where U is a compact subset of $\mathbb{R}^m$. Then $V^0(\cdot)$ is Lipschitz continuous on bounded sets.*

Proof. Let S be an arbitrary bounded set in $\mathcal{X}$, the domain of the value function $V^0(\cdot)$, and let $R := S \times \mathbb{U}$; R is a bounded subset of $\mathcal{Z}$. Since R is bounded, there exists a Lipschitz constant L_S such that

$$\left| V(x', u) - V(x'', u) \right| \leq L_S |x' - x''|$$

for all $x', x'' \in S$, all $u \in U$. Hence,

$$V^0(x') - V^0(x'') \leq V(x', u'') - V(x'', u'') \leq L_S |x' - x''|$$

for all $x', x'' \in S$, any $u'' \in u^0(x'')$. Interchanging x' and x'' in the above derivation yields

$$V^0(x'') - V^0(x') \leq V(x'', u') - V(x', u') \leq L_S |x'' - x'|$$

for all $x', x'' \in S$, any $u' \in u^0(x')$. Hence $V^0(\cdot)$ is Lipschitz continuous on bounded sets. ∎

We now specialize to the case where $U(x) = \{u \in \mathbb{R}^m \mid (x, u) \in \mathcal{Z}\}$ where $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{R}^m$; for each x, $U(x)$ is a polytope. This type of constraint arises in constrained optimal control problems when the system is linear and the state and control constraints are polyhedral. What we show in the sequel is that, in this special case, $U(\cdot)$ is continuous and so, therefore, is $V^0(\cdot)$. An alternative proof, which many readers may prefer, is given in Chapter 7 where we exploit the fact that if $V(\cdot)$ is strictly convex and quadratic and $\mathcal{Z}$ polyhedral, then $V^0(\cdot)$ is piecewise quadratic and continuous. Our first concern is to obtain a bound on $d(u, U(x'))$, the distance of any $u \in U(x)$ from the constraint set $U(x')$.

A bound on $d(u, U(x')), u \in U(x)$. The bound we require is given by a special case of a theorem due to Clarke, Ledyaev, Stern, and Wolenski (1998, Theorem 3.1, page 126). To motivate this result, consider a differentiable convex function $f : \mathbb{R} \to \mathbb{R}$ so that $f(u) \geq f(v) + \langle \nabla f(v), u - v \rangle$ for any two points u and v in $\mathbb{R}$. Suppose also that there exists a nonempty interval $U = [a, b] \subset \mathbb{R}$ such that $f(u) \leq 0$ for

[6] A function $V(\cdot)$ is Lipschitz continuous on bounded sets if, for any bounded set S, there exists a constant $L_S \in [0, \infty)$ such that $|V(z') - V(z)| \leq L_S |z - z'|$ for all $z, z' \in S$.

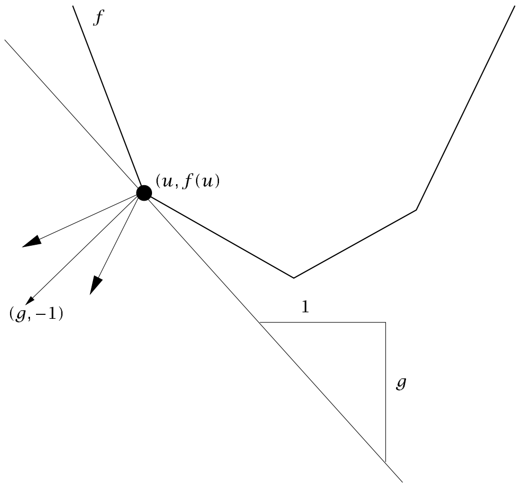


Figure C.12: Subgradient of $f(\cdot)$.

all $u \in U$ and that there exists a $\delta > 0$ such that $\Delta f(u) > \delta$ for all $u \in \mathbb{R}$. Let $u > b$ and let $v = b$ be the closest point in U to u. Then $f(u) \geq f(v) + \langle \nabla f(v), u - v \rangle \geq \delta |v - u|$ so that $d(u, U) \leq f(u)/\delta$. The theorem of Clarke et al. (1998) extends this result to the case when $f(\cdot)$ is not necessarily differentiable but requires the concept of a subgradient of a convex function

Definition C.30 (Subgradient of convex function)**.** Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is convex. Then the subgradient $\delta f(u)$ of $f(\cdot)$ at u is defined by

$$\delta f(u) := \{g \mid f(v) \geq f(u) + \langle g, v - u \rangle \;\forall v \in \mathbb{R}^m\}$$

Figure C.12 illustrates a subgradient. In the figure, g is one element of the subgradient because $f(v) \geq f(u) + \langle g, v - u \rangle$ for all v; g is the slope of the line passing through the point $(u, f(u))$. To obtain a bound on $d(u, U(x))$ we require the following result which is a special case of the much more general result of the theorem of Clarke *et al.*:

Theorem C.31 (Clarke et al. (1998))**.** *Take a nonnegative valued, convex function $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. Let $U(x) := \{u \in \mathbb{R}^m \mid \psi(x, u) = 0\}$ and $\mathcal{X} := \{x \in \mathbb{R}^n \mid U(x) \neq \emptyset\}$. Assume there exists a $\delta > 0$ such that*

$$u \in \mathbb{R}^m, \; x \in \mathcal{X}, \psi(x, u) > 0 \text{ and } g \in \partial_u \psi(x, u) \implies |g| > \delta$$

where $\partial_u \psi(x, u)$ denotes the convex subgradient of ψ with respect to *the variable u. Then, for each $x \in \mathcal{X}$, $d(u, U(x)) \le \psi(x,u)/\delta$ for all $u \in \mathbb{R}^m$.*

The proof of this result is given in the reference cited above. We next use this result to bound the distance of u from $U(x)$ where, for each x, $U(x)$ is polyhedral.

Corollary C.32 (A bound on $d(u, U(x'))$ for $u \in U(x)$)**.** [7] *Suppose $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{R}^m$ and let $\mathcal{X}$ denote its projection on $\mathbb{R}^n$ ($\mathcal{X} = \{x \mid \exists u \in \mathbb{R}^m$ such that $(x,u) \in \mathcal{Z}\}$). Let $\mathcal{U}(x) := \{u \mid (x, u) \in \mathcal{Z}\}$. Then there exists a $K > 0$ such that for all $x, x' \in \mathcal{X}$, $d(u, U(x')) \le K|x' - x|$ for all $u \in U(x)$ (or, for all $x, x' \in \mathcal{X}$, all $u \in U(x)$, there exists a $u' \in U(x')$ such that $|u' - u| \le K|x' - x|$).*

Proof. The polyhedron $\mathcal{Z}$ admits the representation $\mathcal{Z} = \{(x,u) \mid \langle m^j, u\rangle - \langle n^j, x\rangle - p^j \le 0,\ j \in \mathcal{J}\}$ for some $m^j \in \mathbb{R}^m$, $n^j \in \mathbb{R}^n$ and $p^j \in \mathbb{R}$, $j \in \mathcal{J} := \{1, \ldots, J\}$. Define $\mathcal{D}$ to be the collection of all index sets $I \subseteq \mathcal{J}$ such that $\sum_{j \in I} \lambda^j m^j \neq 0,\ \forall \lambda \in \Lambda_I$ in which, for a particular index set I, Λ_I is defined to be $\Lambda_I := \{\lambda \mid \lambda^j \ge 0,\ \sum_{j\in I} \lambda^j = 1\}$. Because $\mathcal{D}$ is a finite set, there exists a $\delta > 0$ such that for all $I \in \mathcal{D}$, all $\lambda \in \Lambda_I$, $|\sum_{j \in I} \lambda^j m^j| > \delta$. Let $\psi(\cdot)$ be defined by $\psi(x,u) := \max\{\langle m^j, u\rangle - \langle n^j, x\rangle - p^j, 0 \mid j \in \mathcal{J}\}$ so that $(x,u) \in \mathcal{Z}$ (or $u \in \mathcal{U}(x)$) if and only if $\psi(x,u) = 0$. We now claim that, for every $(x,u) \in \mathcal{X} \times \mathbb{R}^m$ such that $\psi(x,u) > 0$ and every $g \in \partial_u \psi(x,u)$, the subgradient of ψ with respect to u at (x,u), we have $|g| > \delta$. Assuming for the moment that the claim is true, the proof of the Corollary may be completed with the aid of Theorem C.31. Assume, as stated in the Corollary, that $x, x' \in \mathcal{X}$ and $u \in \mathcal{U}(x)$; the theorem asserts

$$d(u, \mathcal{U}(x')) \le (1/\delta)\psi(x', u),\ \forall x' \in \mathcal{X}$$

But $\psi(x,u) = 0$ (since $u \in \mathcal{U}(x)$) so that

$$d(u, \mathcal{U}(x')) \le (1/\delta)[\psi(x', u) - \psi(x,u)] \le (c/\delta)|x' - x|$$

where c is the Lipschitz constant for $x \mapsto \psi(x,u)$ ($\psi(\cdot)$ is piecewise affine and continuous). This proves the Corollary with $K = c/\delta$.

It remains to confirm the claim. Take any $(x,u) \in \mathcal{X} \times \mathbb{R}^m$ such that $\psi(x,u) > 0$. Then $\max_j\{\langle m^j, u\rangle - \langle n^j, x\rangle - p^j, 0 \mid j \in \mathcal{J}\} > 0$. Let

[7] The authors wish to thank Richard Vinter and Francis Clarke for providing this result.

$I^0(x,u)$ denote the active constraint set (the set of those constraints at which the maximum is achieved). Then

$$\langle m^j, u\rangle - \langle n^j, x\rangle - p^j > 0, \ \forall j \in I^0(x,u)$$

Since $x \in \mathcal{X}$, there exists a $\bar{u} \in \mathcal{U}(x)$ so that

$$\langle m^j, \bar{u}\rangle - \langle n^j, x\rangle - p^j \leq 0, \ \forall j \in I^0(x,u)$$

Subtracting these two inequalities yields

$$\langle m^j, u - \bar{u}\rangle > 0, \ \forall j \in I^0(x,u)$$

But then, for all $\lambda \in \Lambda_{I^0(x,u)}$, it follows that $|\sum_{j \in I^0(x,u)} \lambda^j m^j (u - \bar{u})| > 0$, so that

$$\sum_{j \in I^0(x,u)} \lambda^j m^j \neq 0$$

It follows that $I^0(x,u) \in \mathcal{D}$. Hence

$$\Big|\sum_{j \in I^0(x,u)} \lambda^j m^j\Big| > \delta, \ \forall \lambda \in \Lambda_{I^0(x,u)}$$

Now take any $g \in \partial_u f(x,u) = \text{co}\{m^j \mid j \in I^0(x,u)\}$ (co denotes "convex hull"). There exists a $\lambda \in \Lambda_{I^0(x,u)}$ such that $g = \sum_{j \in I^0(x,u)} \lambda^j m_j$. But then $|g| > \delta$ by the inequality above. This proves the claim and, hence, completes the proof of the Corollary. ■

Continuity of the value function when $U(x) = \{u \mid (x,u) \in \mathcal{Z}\}$. In this section we investigate continuity of the value function for the constrained linear quadratic optimal control problem $\mathbb{P}(x)$; in fact we establish continuity of the value function for the more general problem where the cost is continuous rather than quadratic. We showed in Chapter 2 that the optimal control problem of interest takes the form

$$V^0(x) = \min_u \{V(x,u) \mid (x,u) \in \mathcal{Z}\}$$

where $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{U}$ where $\mathbb{U} \subset \mathbb{R}^m$ is a polytope and, hence, is compact and convex; in MPC problems we replace the control u by the sequence of controls $\mathbf{u}$ and m by Nm. Let $u^0 : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ be defined by

$$u^0(x) := \arg\min_u \{V(x,u) \mid (x,u) \in \mathcal{Z}\}$$

and let $\mathcal{X}$ be defined by

$$\mathcal{X} := \{x \mid \exists u \text{ such that } (x,u) \in \mathcal{Z}\}$$

so that $\mathcal{X}$ is the projection of $\mathcal{Z} \subset \mathbb{R}^n \times \mathbb{R}^m$ onto $\mathbb{R}^n$. Let the set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ be defined by

$$U(x) := \{u \in \mathbb{R}^m \mid (x, u) \in Z\}$$

The domain of $V^0(\cdot)$ and of $U(\cdot)$ is $\mathcal{X}$. The optimization problem may be expressed as $V^0(x) = \min_u\{V(x, u) \mid u \in U(x)\}$. Our first task is establish the continuity of $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$.

Theorem C.33 (Continuity of $U(\cdot)$). *Suppose $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{U}$ where $\mathbb{U} \subset \mathbb{R}^m$ is a polytope. Then the set-valued function $U : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ defined above is continuous in $\mathcal{X}$.*

Proof. By Proposition C.27, the set-valued map $U(\cdot)$ is outer semicontinuous in $\mathcal{X}$ because its graph, $\mathcal{Z}$, is closed. We establish inner semicontinuity using Corollary C.32 above. Let x, x' be arbitrary points in $\mathcal{X}$ and $U(x)$ and $U(x')$ the associated control constraint sets. Let S be any open set such that $U(x) \cap S \neq \emptyset$ and let u be an arbitrary point in $U(x) \cap S$. Because S is open, there exist an $\varepsilon > 0$ such that $u \oplus \varepsilon\mathcal{B} \subset S$. Let $\varepsilon' := \varepsilon/K$ where K is defined in Corollary 1. From Corollary C.32, there exists a $u' \in U(x')$ such that $|u' - u| \leq K|x' - x|$ which implies $|u' - u| \leq \varepsilon$ ($u' \in u \oplus \varepsilon\mathcal{B}$) for all $x' \in \mathcal{X}$ such that $|x' - x| \leq \varepsilon'$ ($x' \in (x \oplus \varepsilon'\mathcal{B}) \cap \mathcal{X}$). This implies $u \in \mathcal{U}(x') \cap S$ for all $x' \in \mathcal{X}$ such that $|x' - x| \leq \varepsilon'$ ($x' \in (x \oplus \varepsilon'\mathcal{B}) \cap \mathcal{X}$). Hence $\mathcal{U}(x') \cap S \neq \emptyset$ for all $x' \in (x \oplus \varepsilon'\mathcal{B}) \cap \mathcal{X}$, so that $\mathcal{U}(\cdot)$ is inner semicontinuous in $\mathcal{X}$. Since $U(\cdot)$ is both outer and inner semicontinuous in $\mathcal{X}$, it is continuous in $\mathcal{X}$. ∎

We can now establish continuity of the value function.

Theorem C.34 (Continuity of the value function). *Suppose that $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous and that $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{U}$ where $\mathbb{U} \subset \mathbb{R}^m$ is a polytope. Then $V^0 : \mathbb{R}^n \to \mathbb{R}$ is continuous and $u^0 : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is outer semicontinuous in $\mathcal{X}$. Moreover, if $u^0(x)$ is unique (not set-valued) at each $x \in \mathcal{X}$, then $u^0 : \mathbb{R}^n \to \mathbb{R}^m$ is continuous in $\mathcal{X}$.*

Proof. Since the real-valued function $V(\cdot)$ is continuous (by assumption) and since the set-valued function $U(\cdot)$ is continuous in $\mathcal{X}$ (by Theorem C.33), it follows from Theorem C.28 that $V^0 : \mathbb{R}^n \to \mathbb{R}$ is continuous and $u^0 : \mathbb{R}^n \rightsquigarrow \mathbb{R}^m$ is outer semicontinuous in $\mathcal{X}$; it also follows that if $u^0(x)$ is unique (not set-valued) at each $x \in \mathcal{X}$, then $u^0 : \mathbb{R}^n \to \mathbb{R}^m$ is continuous in $\mathcal{X}$. ∎

Lipschitz continuity when $U(x) = \{u \mid (x, u) \in \mathcal{Z}\}$. Here we establish that $V^0(\cdot)$ is Lipschitz continuous if $V(\cdot)$ is Lipschitz continuous and $U(x) := \{u \in \mathbb{R}^m \mid (x, u) \in \mathcal{Z}\}$; this result is more general than Theorem C.29 where it is assumed that U is constant.

Theorem C.35 (Lipschitz continuity of the value function—$U(x)$)**.** *Suppose that $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous, that $\mathcal{Z}$ is a polyhedron in $\mathbb{R}^n \times \mathbb{U}$ where $\mathbb{U} \subset \mathbb{R}^m$ is a polytope. Suppose, in addition, that $V : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is Lipschitz continuous on bounded sets.*[8] *Then $V^0(\cdot)$ is Lipschitz continuous on bounded sets.*

Proof. Let S be an arbitrary bounded set in $\mathcal{X}$, the domain of the value function $V^0(\cdot)$, and let $R := S \times \mathbb{U}$; R is a bounded subset of $\mathcal{Z}$. Let x, x' be two arbitrary points in S. Then

$$V^0(x) = V(x, \kappa(x))$$
$$V^0(x') = V(x', \kappa(x'))$$

where $V(\cdot)$ is the cost function, assumed to be Lipschitz continuous on bounded sets, and $\kappa(\cdot)$, the optimal control law, satisfies $\kappa(x) \in U(x) \subset \mathbb{U}$ and $\kappa(x') \in U(x') \subset \mathbb{U}$. It follows from Corollary C.32 that there exists a $K > 0$ such that for all $x, x' \in \mathcal{X}$, there exists a $u' \in U(x') \subset \mathbb{U}$ such that $|u' - \kappa(x)| \leq K|x' - x|$. Since $\kappa(x)$ is optimal for the problem $\mathbb{P}(x)$, and since $(x, \kappa(x))$ and (x', u') both lie in $R = S \times \mathbb{U}$, there exists a constant L_R such that

$$\begin{aligned} V^0(x') - V^0(x) &\leq V(x', u') - V(x, \kappa(x)) \\ &\leq L_R(|(x', u') - (x, \kappa(x))|) \\ &\leq L_R|x' - x| + L_R K|x' - x| \\ &\leq M_S|x' - x|, \quad M_S := L_R(1 + K) \end{aligned}$$

Reversing the role of x and x' we obtain the existence of a $u \in \mathcal{U}(x)$ such that $|u - \kappa(x')| \leq K|x - x'|$; it follows from the optimality of $\kappa(x')$ that

$$\begin{aligned} V^0(x) - V^0(x') &\leq V(x, u) - V(x', \kappa(x')) \\ &\leq M_S|x - x'| \end{aligned}$$

where, now, $u \in U(x)$ and $\kappa(x') \in U(x')$. Hence $|V^0(x') - V^0(x)| \leq M_S|x - x'|$ for all x, x' in S. Since S is an arbitrary bounded set in $\mathcal{X}$, $V^0(\cdot)$ is Lipschitz continuous on bounded sets. ■

[8] A function $V(\cdot)$ is Lipschitz continuous on bounded sets if, for any bounded set S, there exists a constant $L_S \in [0, \infty)$ such that $|V(z') - V(z)| \leq L_S|z - z'|$ for all $z, z' \in S$.