Appendix to Lecture 11, Proofs

In this appendix I provide proofs not covered in class for lack of time.

Lemma 11.12: Suppose that assumptions 5 and 8 apply. If \( x \in X \) and \( y \in \text{int} \, G(x) \), then there is a positive number \( \varepsilon \) such that \( y \in \text{int} \, G(x) \), if \( ||x - x|| < \varepsilon \) and \( x \in X \). If in addition \( x \in \text{int} \, X \), then \((x, y)\) belongs to the interior of the graph of \( G \).

Proof:

Let \( e = (1, \ldots, 1) \in \mathbb{R}^N \) and, for \( n = 1, \ldots, N \), let \( e_n \) be the nth standard basis vector of \( \mathbb{R}^N \). For any positive number \( \delta \),

\[
y = \frac{1}{2} \left( y - \frac{\delta}{2N} e \right) + \frac{1}{2} \sum_{n=1}^{N} \left( y - \frac{\delta}{2N} e + \frac{\delta}{2N} e_n \right)
\]

Let \( y^0 = y - \frac{\delta}{2N} e \) and let \( y^n = y - \frac{\delta}{2N} e + \frac{\delta}{2N} e_n \), for \( n = 1, \ldots, N \). Since \( y \in \text{int} \, G(x) \), we may choose \( \delta \) so small that \( y^0, y^1, \ldots, y^N \) all belong to \( \text{int} \, G(x) \). The construction is illustrated in the above figure for the case \( N = 2 \). The \( N+1 \)-vector \( z = \left( \frac{1}{2}, \frac{1}{2N}, \ldots, \frac{1}{2N} \right) \) satisfies the equation.
\[
\begin{pmatrix}
y_0^0 & y_1^1 & \cdots & y_N^N \\
y_0^1 & y_1^1 & \cdots & y_N^1 \\
\vdots & \vdots & \ddots & \vdots \\
y_0^N & y_1^N & \cdots & y_N^N \\
1 & 1 & \cdots & 1 \\
\end{pmatrix}
\begin{pmatrix}
y_0 \\
y_1 \\
\vdots \\
y_N \\
1 \\
\end{pmatrix}
= 
\begin{pmatrix}
y \\
\vdots \\
1 \\
\end{pmatrix}
\]

which we may write as

\[
B \mathbf{z} = \begin{pmatrix} y \\ \vdots \\ 1 \end{pmatrix}.
\]

The \((N+1)\times(N+1)\) matrix \(B\) is non-singular, for suppose that \(B\mathbf{q} = 0\). Then \(\sum_{n=0}^{N} q_n = 0\) and

\[
q_0 \left( y - \frac{\delta}{2N} \mathbf{e} \right) + \sum_{n=1}^{N} q_n \left( y - \frac{\delta}{2N} e_n + \delta e_n \right) = 0,
\]

so that

\[
\sum_{n=0}^{N} q_n \left( y - \frac{\delta}{2N} e_n \right) + \delta \sum_{n=1}^{N} q_n e_n = 0.
\]

Since \(\sum_{n=0}^{N} q_n = 0\), it follows that

\[
\sum_{n=0}^{N} q_n \left( y - \frac{\delta}{2N} e_n \right) = \left( \sum_{n=0}^{N} q_n \right) \left( y - \frac{\delta}{2N} \mathbf{e} \right) = 0
\]

and hence \(\sum_{n=1}^{N} q_n e_n = 0\). Because \(e_1, \ldots, e_N\) are independent, \(q_1 = q_2 = \ldots = q_N = 0\). Hence

\[
q_0 = -\sum_{n=1}^{N} q_n = 0,
\]

and so \(B\) is invertible. Since \(B\) has an inverse,

\[
\mathbf{z} = B^{-1} \begin{pmatrix} y \\ \vdots \\ 1 \end{pmatrix}
\]
Since the determinant is a continuous function and \( \det B \neq 0 \), it follows that \( \det D \neq 0 \), if the \((N+1) \times (N+1)\) matrix \( D \) is close enough to \( B \). Therefore there is a positive number \( \eta \) such that if \( \vec{y}^n \in \mathbb{R}^N \), for \( n = 0, \ldots, N \) and \( || \vec{y}^n - \vec{y}^0 || < \eta \), for all \( n \), then

\[
D = \begin{pmatrix}
\vec{y}^0 & \vec{y}^{-1} & \cdots & \vec{y}^N \\
\vec{y}_1 & \vec{y}_1 & \cdots & \vec{y}_1 \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\vec{y}_N & \vec{y}_N & \cdots & \vec{y}_N \\
1 & 1 & \cdots & 1
\end{pmatrix}
\]

is invertible. Furthermore because \( \vec{z} \gg 0 \), it follows that

\[
\vec{z} = D^{-1} \begin{pmatrix}
\vec{y} \\
\vdots \\
1
\end{pmatrix} \gg 0 ,
\]

provided \( \eta \) is sufficiently small. Of course,

\[
D \vec{z} = \begin{pmatrix}
\vec{y} \\
\vdots \\
1
\end{pmatrix}
\]

so that \( \sum_{n=0}^{N} \vec{z}_n = 1 \) and \( \vec{y} = \sum_{n=0}^{N} \vec{z}_n \vec{y}^n \). Since all the components of \( \vec{z} \) are positive and \( \vec{y} = \sum_{n=0}^{N} \vec{z}_n \vec{y}^n \), it follows that \( \vec{y} \) is a convex combination of \( \vec{y}^0, \vec{y}^{-1}, \ldots, \vec{y}^N \).

Since \( G \) is lower semicontinuous by assumption 5, there is a positive number \( \epsilon \) so small that if \( || x - x_0 || < \epsilon \) and \( x \in X \), then there is, for each \( n = 0, 1, \ldots, N \), a \( \vec{y}^n \in G(x) \), such that \( || \vec{y}^n - \vec{y}^0 || < \eta \). Form \( \vec{y}^0, \vec{y}^{-1}, \ldots, \vec{y}^N \) and \( D \) and \( \vec{z} \) as above. Then \( \vec{y} \) is the convex combination, \( \sum_{n=0}^{N} \vec{z}_n \vec{y}^n \), of \( \vec{y}^0, \vec{y}^{-1}, \ldots, \vec{y}^N \). Since assumption 8 implies that \( G(x) \) is convex, \( \vec{y} \in G(x) \).

Because the linear function that carries \( \begin{pmatrix} y \\ \vdots \\ 1 \end{pmatrix} \) in \( \mathbb{R}^{N+1} \) to \( D^{-1} \begin{pmatrix} y \\ \vdots \\ 1 \end{pmatrix} \) in \( \mathbb{R}^{N+1} \) is continuous and
\[
D^{-1}\begin{pmatrix} y \\ 1 \end{pmatrix} >> 0, \text{ it follows that there is a positive number } \zeta \text{ so small that if } ||y - \underline{y}|| < \zeta, \text{ then } \\
Z = D^{-1}\begin{pmatrix} y \\ 1 \end{pmatrix} >> 0. \text{ Then } y \text{ is the convex combination, } y = \sum_{n=0}^{N} z^{\prime} \tilde{y}^{n} \text{ of } \tilde{y}^{0}, \tilde{y}^{1}, \ldots, \tilde{y}^{N} \text{ in } G(x) \\
\text{and so belongs to } G(x). \text{ Since } y \in G(x) \text{ if } ||y - \underline{y}|| < \zeta, \text{ it follows that } y \text{ belongs to the interior of } G(x), \text{ if } ||x - \underline{x}|| < \varepsilon. \text{ Here } x \text{ is held fixed and } \zeta \text{ depends on } x.
\]

We may go further with this line of argument by letting \( \zeta \) be independent of \( x \). Because the function \( D^{-1}\begin{pmatrix} y \\ 1 \end{pmatrix} \) is jointly continuous in \( D \) and \( y \), we may choose \( \eta \) and \( \zeta \) to be such small positive numbers that if \( ||\tilde{y}^{n} - \tilde{y}^{n}|| < \eta, \text{ for } n = 0, 1, \ldots, N, \text{ and } ||y - \underline{y}|| < \zeta, \text{ then } \\
Z = D^{-1}\begin{pmatrix} y \\ 1 \end{pmatrix} >> 0, \text{ when } D \text{ is formed from } \tilde{y}^{0}, \tilde{y}^{1}, \ldots, \tilde{y}^{N} \text{ as above. Let } \varepsilon \text{ be a positive number} \\
\text{so small that if } ||x - \underline{x}|| < \varepsilon, \text{ then, for } n = 0, 1, \ldots, N, \text{ there exist } \tilde{y}^{n} \in G(x) \text{ such that } \\
||\tilde{y}^{n} - \tilde{y}^{n}|| < \eta. \text{ Form } D \text{ from } \tilde{y}^{0}, \tilde{y}^{1}, \ldots, \tilde{y}^{N} \text{ and let } Z = D^{-1}\begin{pmatrix} y \\ 1 \end{pmatrix}. \text{ The components of } Z \text{ are} \\
\text{positive and sum to 1, so that } y = \sum_{n=0}^{N} z^{\prime} \tilde{y}^{n} \text{ is a convex combination of } \tilde{y}^{0}, \tilde{y}^{1}, \ldots, \tilde{y}^{N} \text{ and so} \\
\text{belongs to } G(x). \text{ That is, if } ||x - \underline{x}|| < \varepsilon \text{ and } ||y - \underline{y}|| < \zeta, \text{ then } y \in G(x), \text{ so that } (x, y) \text{ belongs to the interior of the graph of } G. \quad \blacksquare
\]

**Theorem 11.15:** Suppose that assumption 1 - 5, 8, and 9 apply, that \( x_{0} \in \text{int } X, \) and that \( (x_{1}, x_{2}, \ldots) \in \mathcal{F}(x_{0}) \) is optimal and is such that \( x_{t} \in \text{int } G(x_{t-1}) \), for \( t = 1, 2, \ldots \).

Then there is a sequence \( (p_{0}, p_{1}, \ldots) \) in \( R^{N} \) such that, for all \( t, p_{t}, \) is a subgradient of \( V \) at \( x_{t} \) and \( (p_{t}, -\beta p_{t+1}) \) is a subgradient of \( u \) at \( (x_{t}, x_{t+1}) \).

The main step of the proof of this theorem is the next lemma.

**Lemma 11.22:** Under the assumptions of theorem 11.15, there exist, for each positive integer \( T \) a finite sequence \( (p_{0}, p_{1}, \ldots, p_{2T+1}) \) in \( R^{N} \) such that for \( t = 0, 1, \ldots, 2T, \\
(p_{t}, -\beta p_{t+1}) \) is a subgradient of \( u \) at \( (x_{t}, x_{t+1}) \) and, for \( t = 0, \ldots, 2T+1, p_{t} \) is a subgradient of
Proof: I apply the Minkowski separation theorem to the sets $\Gamma$ and $Q$ defined as follows.

$$\Gamma = \{ (x_0, x_1, \ldots, x_{2T+1}, x_{2T}) \mid x_0 \in X, x_1 \in G(x_0), x_2 \in G(x_1), \ldots, x_{2T+1} \in G(x_{2T}) \},$$

$$s \leq u(x_0, x_1) + \beta^0 u(x_2, x_3) + \beta^1 u(x_4, x_5) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}) \} \text{ and}$$

$$Q = \{ (x_0, x_1, \ldots, x_{2T+1}, x_{2T+1}, x_{2T+1}) \mid x_0 \in G(x_0), x_1 \in G(x_1), \ldots, x_{2T+1} \in G(x_{2T}) \},$$

$$s \geq V(x_0) - \beta u(x_1, x_2) - \beta^2 u(x_3, x_4) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T} V(x_{2T+1}) \},$$

where $V$ is the value function. By assumption 8, the graph of $G$ is convex. By assumption 9, $u$ is concave and by theorem 11.10, $V$ is concave. Therefore the function

$$f(x_0, x_1, \ldots, x_{2T+1})$$

$$= \beta u(x_1, x_2) + \beta^2 u(x_3, x_4) + \ldots + \beta^{2T-1} u(x_{2T-1}, x_{2T}) + \beta^{2T} V(x_{2T+1})$$

is concave and so $-f(x_0, x_1, \ldots, x_{2T+1})$ is a convex function. It follows that the set $Q$ is convex. Similarly $\Gamma$ is convex because $X$ and the graph of $G$ are convex and the function

$$g(x_0, x_1, \ldots, x_{2T+1})$$

$$= u(x_0, x_1) + \beta^2 u(x_2, x_3) + \beta^4 u(x_4, x_5) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1})$$

is concave.

The interior of $\Gamma$ is not empty, since it contains the point

$$\left( x_0, x_1, \ldots, x_{2T+1}, u(x_0, x_1) + \beta^2 u(x_2, x_3) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}) - 1 \right).$$

I next show that $Q$ does not intersect the interior of $\Gamma$. If $(x_0, x_1, \ldots, x_{2T+1}, s)$ belongs to the interior of $\Gamma$, then

$$s < u(x_0, x_1) + \beta^2 u(x_2, x_3) + \beta^4 u(x_4, x_5) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}).$$

Therefore if $(x_0, x_1, \ldots, x_{2T+1}, s) \in Q \cap (\text{int } \Gamma)$, then
\[ V(x_0) = \beta u(x_1, x_2) - \beta^3 u(x_3, x_4) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1}) \]
\[ \leq s < u(x_0, x_1) + \beta^2 u(x_2, x_3) + \beta^4 u(x_4, x_5) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}), \]
where \( x_t \in G(x_T) \) and \( x_t \in G(x_{t-1}) \), for \( t = 2, \ldots, 2T+1 \). Hence
\[ V(x_0) < u(x_0, x_1) + \beta u(x_1, x_2) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}) + \beta^{2T+1} V(x_{2T+1}), \]
which is impossible. Hence we conclude that \( Q \cap (\text{int } \Gamma) \) is empty.

Since all the assumptions of the Minkowski separation theorem apply to the sets \( Q \) and \( \Gamma \), that theorem implies that there exists a non-zero vector \( (v_0, v_1, \ldots, v_{2T+1}, r) \in \mathbb{R}^N \times \ldots \times \mathbb{R}^N \times \mathbb{R} \) such that
\[
\sum_{t=0}^{2T+1} v_t x_t + \sum_{t=1}^{2T+1} r s t \geq v_0 x_0 + \sum_{t=1}^{2T+1} v_t x_t + rs,
\]
for all \( (x_0, \ldots, x_{2T+1}, s) \in \Gamma \) and \( (\overline{x_0}, \overline{x_1}, \ldots, \overline{x_{2T+1}}, \overline{s}) \in Q \).

I show that \( r \leq 0 \). The vector
\[
(\overline{x_0}, \overline{x_1}, \ldots, \overline{x_{2T+1}}, V(x_0) - \beta u(x_1, x_2) - \beta^3 u(x_3, x_4) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1}) + 1)
\]
belongs to \( Q \). The vector
\[
(\overline{x_0}, \overline{x_1}, \ldots, \overline{x_{2T+1}}, u(x_0, x_1) + \beta^2 u(x_2, x_3) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}))
\]
belongs to \( \Gamma \). Therefore by inequality 11.14,
\[
\sum_{t=0}^{2T+1} v_t x_t + r[\beta u(x_0, x_1) + \beta^2 u(x_2, x_3) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1})] \]
\[
\geq \sum_{t=0}^{2T+1} v_t x_t + r[V(x_0) - \beta u(x_1, x_2) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1})] + r,
\]
so that
\[ r \leq r[-V(x) + \sum_{l=0}^{2T} \beta^l u(x, x_{-l}) + \beta^{2T+1} V(x_{-2T+1})] = 0. \]

I show that \( r < 0 \). Suppose that it is not true that \( r < 0 \). Then since \( r \leq 0 \), \( r = 0 \). Since \((v_0, v_1, ..., v_{2T+1}, r) \neq 0\), it follows that \((v_0, v_1, ..., v_{2T+1}) \neq 0\). For some number \( s \),

\[
(x_0, x_1, ..., x_{2T+1}, s) \in Q. \text{ Let } (x_0, x_1, ..., x_{2T+1}) \text{ be any sequence such that } x_0 \in X \text{ and } x_1 \in G(x_{-1}) \text{, for } t = 1, ..., 2T+1. \text{ Then for some number } s, (x_0, x_1, ..., x_{2T+1}, s) \in \Gamma. \text{ By inequality 11.15 and because } r = 0,
\]

\[
\sum_{t=0}^{2T+1} v_t x_t \geq \sum_{t=0}^{2T+1} v_t x_t. \tag{11.15}
\]

Since \( x_0 \in \text{int } X \) and \( x_1 \in \text{int } G(x_{-1}) \), for \( t \geq 1 \), lemma 11.12 implies that if the positive number \( \varepsilon \) is sufficiently small, \( x_0 - \varepsilon v_0 \in X \text{ and } x_1 - \varepsilon v_1 \in G(x_{-1} - \varepsilon v_{t-1}) \), for \( t = 1, ..., 2T+1 \). Therefore, we may let \( x_0 = x_0 - \varepsilon v_0 \), for all \( t \) in inequality 11.15. Hence

\[
\sum_{t=0}^{2T+1} v_t x_t - \varepsilon \sum_{t=0}^{2T+1} v_t x_t \geq \sum_{t=0}^{2T+1} v_t x_t,
\]

which is impossible. We conclude that \( r \neq 0 \) and so \( r < 0 \).

Let \( -w = \frac{1}{-r} v \), for \( t = 1, ..., 2T+1 \). Then the vector \((-w_0, ..., -w_{2T+1}, -1)\)

separates \( \Gamma \) from \( Q \) in the same way that \((v_0, v_1, ..., v_{2T+1}, r)\) does. That is,

\[
-w_0 x_0 - \sum_{t=1}^{2T+1} w_t x_t - s \geq -w_0 x_0 - \sum_{t=1}^{2T+1} w_t x_t - s, \tag{11.16}
\]

for all \((x_0, ..., x_{2T+1}, s) \in \Gamma\) and \((x_0, x_1, ..., x_{2T+1}, s) \in Q\).

I use inequality 11.16 to show that \((-w_0, -w_1)\) is a subgradient of \( u \) at \((x_0, x_1)\). By lemma 11.12, if \((x_0, x_1)\) is close enough to \((x_0, x_1)\), then \( x_1 \in \text{G}(x_0) \) and \( x_2 \in \text{G}(x_1) \), so that

\[
(x_0, x_1, x_2, ..., x_{2T+1} u(x_0, x_1) + \beta^1 u(x_2, x_1) + ... + \beta^{2T} u(x_{-2T+1}, x_{-2T+1}))\]

belongs to \( \Gamma \). We know that
\[
\left( x_0, x_1, \ldots, x_{2T+1}, V(x_0) - \beta u(x_0, x_2) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1}) \right)
\]

belongs to \( Q \). Hence inequality 11.16 implies that

\[
-w \cdot x - w \cdot x - \sum_{t=1}^{2T+1} w_t \cdot x - u(x_0, x_1) - \beta^2 u(x_0, x_2) - \ldots - \beta^{2T} u(x_{2T-1}, x_{2T+1})
\]

\[
\geq -w \cdot x - w \cdot x - \sum_{t=2}^{2T+1} w_t \cdot x - V(x_0) + \beta u(x_0, x_1) + \ldots
\]

\[
= \beta^{2T-1} u(x_{2T-1}, x_{2T}) + \beta^{2T} V(x_{2T+1}).
\]

If we cancel like terms on both sides of the inequality and rearrange the remaining terms, we find that

\[
u(x_0, x_1) \leq V(x_0) - \beta u(x_0, x_2) - \beta^2 u(x_0, x_3) - \ldots
\]

\[
= u(x_0, x_1) - w_0 \cdot (x_0 - x_1) - w_1 \cdot (x_0 - x_1)
\]

Since this inequality holds for any \( (x_0, x_1) \) close enough to \( (x_0, x_1) \), lemma 11.13 implies that \(-w, -w_1\) is a subgradient of \( u \) at \( (x_0, x_1) \).

A similar argument may be used to prove that \(-w_{2t}, -w_{2t+1}\) is a subgradient of \( \beta^t u(x_{2t}, x_{2t+1}) \) at \( (x_{2t}, x_{2t+1}) \), for \( t = 1, \ldots, T \). If \( (x_0, x_1) \) is close enough to \( (x_{2t}, x_{2t+1}) \), then \( x_0 \in G(x_{2t}) \) and \( x_{2t+1} \in G(x_{2t+1}) \). Then

\[
(\ldots, x_{2T-1}, x_{2T}, x_{2T+1}, x_{2T+2}) - \beta u(x_0, x_1) + \beta^2 u(x_0, x_2) + \ldots
\]

belongs to \( \Gamma \). Apply inequality 11.16 to this point in \( \Gamma \) and to the point

\[
(\ldots, x_{2T-1}, x_{2T+1}, V(x_0) - \beta u(x_0, x_2) - \ldots - \beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1}) \in Q \)

in \( Q \) and proceed as in the previous paragraph.

I next use inequality 11.16 to show that \( (w, w_1) \) is a subgradient of \( \beta u \) at \( (x_0, x_1) \).
By lemma 11.12, if \((x_1, x_2)\) is close enough to \((x_0, x_2)\), then \(x_1 \in G(x_0), x_2 \in G(x_1)\), and \(x_3 \in G(x_2)\). Hence

\[
\begin{align*}
(x_0, x_1, x_2, x_3, \ldots, x_{2T+1}, V(x_0) - \beta u(x_1, x_2) - \beta^2 u(x_2, x_3) - \ldots - \\
\beta^{2T-1} u(x_{2T-1}, x_{2T}) - \beta^{2T+1} V(x_{2T+1})
\end{align*}
\]

belongs to \(Q\). Clearly

\[
\begin{align*}
(x_0, x_1, x_2, \ldots, x_{2T+1}, u(x_0, x_1) + \beta^2 u(x_2, x_3) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1})
\end{align*}
\]

belongs to \(\Gamma\). Therefore inequality 11.16 implies that

\[
\begin{align*}
-w_0 x_0 - w_1 x_1 - w_2 x_2 - \ldots - w_{2T} x_{2T} - \\
-w^2 u(x_0, x_1) - \beta^2 u(x_2, x_3) - \ldots - \beta^{2T} u(x_{2T}, x_{2T+1})
\end{align*}
\]

\[
\geq -w_0 x_0 - w_1 x_1 - w_2 x_2 - \ldots - w_{2T} x_{2T} - \\
-w^2 u(x_0, x_1) + \beta^2 u(x_2, x_3) + \ldots + \beta^{2T} u(x_{2T}, x_{2T+1}) + \beta^{2T+1} V(x_{2T+1}),
\]

so that

\[
\begin{align*}
w_1(x_1 - x_2) + w_2(x_2 - x_3) + \beta u(x_1, x_2)
\end{align*}
\]

\[
= w_1(x_1 - x_2) + w_2(x_2 - x_3) + V(x_0) - u(x_0, x_1) - \beta^2 u(x_2, x_3) - \\
\beta^2 u(x_2, x_3) - \ldots - \beta^{2T} u(x_{2T}, x_{2T+1}) - \beta^{2T+1} V(x_{2T+1}) - \beta u(x_1, x_2).
\]

Since this inequality holds for any \((x_1, x_2)\) close enough to \((x_0, x_2)\), lemma 11.13 implies that \((w_1, w_2)\) is a subgradient of \(\beta u\) at \((x_0, x_2)\).

A similar argument may be used to prove that \((w_{2T}, w_{2T+1})\) is a subgradient of \(\beta^{2T+1} u\) at \((x_{2T+1}, x_{2T+2})\), for \(t = 1, \ldots, T-1\). If \((x_{2T}, x_{2T+1})\) is close enough to \((x_{2T+1}, x_{2T+2})\), then \(x_{2T+1} \in G(x_{2T}), x_{2T+2} \in G(x_{2T+1}), \) and \(x_{2T+3} \in G(x_{2T+2})\). Then

\[
\begin{align*}
(x_0, x_1, \ldots, x_{2T}, x_{2T+1}, x_{2T+2}, x_{2T+3}, \ldots, x_{2T+1}, V(x_0) - \beta u(x_1, x_2) - \ldots - \\
\beta^{2T} u(x_{2T}, x_{2T+1}).
\end{align*}
\]
\[-\beta_T u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta_{T+1} u(\frac{x_{2T+1}}{2T}, \frac{x_{2T+2}}{2T}) - \beta_{T+3} u(\frac{x_{2T+3}}{2T}, \frac{x_{2T+4}}{2T}) - \ldots \]

\[-\beta_t u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta_{T+1} V(\frac{x_{2T+1}}{2T}) \]

belongs to \(Q\). Apply inequality 11.16 to this point in \(Q\) and to the point

\[
\left( \frac{x_0}{2T}, \frac{x_1}{2T}, \ldots, \frac{x_{2T+1}}{2T}, \frac{x_{2T}}{2T} \right) \in G(\frac{x_0}{2T})
\]

in \(\Gamma\) and proceed as in the previous paragraph.

The last use of inequality 11.16 is to show that \(w\) is a subgradient of \(\beta_{T+1} V\) at \(\frac{x_{2T+1}}{2T}\).

If \(x_{2T+1}\) is close enough to \(\frac{x_{2T+1}}{2T}\), then \(x_{2T+1} \in G(\frac{x_{2T}}{2T})\), so that

\[
\left( \frac{x_0}{2T}, \frac{x_1}{2T}, \ldots, \frac{x_{2T+1}}{2T}, \frac{x_{2T}}{2T}, V(\frac{x_0}{2T}) - \beta u(\frac{x_1}{2T}, \frac{x_2}{2T}) - \beta^3 u(\frac{x_3}{2T}, \frac{x_4}{2T}) - \ldots \right.
\]

\[-\beta_t u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta_{T+1} V(\frac{x_{2T+1}}{2T}) \]

belongs to \(Q\). If we apply inequality 11.16 to this point and to the point

\[
\left( \frac{x_0}{2T}, \frac{x_1}{2T}, \ldots, \frac{x_{2T+1}}{2T}, \frac{x_{2T}}{2T} \right) \in G(\frac{x_0}{2T})
\]

in \(\Gamma\), we find that

\[-w, x_{2T+1} - w, x_{2T} - w, x_{2T+2} - u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta^3 u(\frac{x_3}{2T}, \frac{x_4}{2T}) - \ldots \]

\[-\beta^T u(\frac{x_{2T-1}}{2T}, \frac{x_{2T}}{2T}) \]

\[\geq -w, x_{2T+1} - w, x_{2T} - w, x_{2T+2} - u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta u(\frac{x_1}{2T}, \frac{x_2}{2T}) + \ldots \]

\[+ \beta^T u(\frac{x_{2T-1}}{2T}, \frac{x_{2T}}{2T}) + \beta^T V(\frac{x_{2T+1}}{2T}) \]

Hence

\[w_{2T+1}(x_{2T+1} - x_{2T+1}) + \beta^T V(\frac{x_{2T+1}}{2T}) \]

\[= w_{2T+1}(x_{2T+1} - x_{2T+1}) + V(\frac{x_0}{2T}) - u(\frac{x_0}{2T}, \frac{x_1}{2T}) - \beta u(\frac{x_1}{2T}, \frac{x_2}{2T}) - \ldots - \beta^T u(\frac{x_{2T}}{2T}, \frac{x_{2T+1}}{2T}) \]

\[\geq \beta^T V(\frac{x_{2T+1}}{2T}) \]

\[\geq \frac{\beta^T V(\frac{x_{2T+1}}{2T})}{\beta^T V(\frac{x_{2T+1}}{2T})} \]

\[\geq \beta^T V(\frac{x_{2T+1}}{2T}) \]
Since this inequality holds for any \( x_{2T+1} \) close enough to \( x_{2T+1} \), lemma 11.13 implies that \( w_{2T+1} \)
is a subgradient of \( \beta^{2T+1}V \) at \( x_{2T+1} \).

For \( t = 0, \ldots, T \), let \( p_{2t} = -\beta^{2t}w_{2t} \) and let \( p_{2t+1} = \beta^{2(t+1)}w_{2t+1} \). I show that \( p_{2T+1} \) is a
subgradient of \( V \) at \( x_{2T+1} \) and, for \( t = 0, 1, \ldots, 2T \), \((p_t, -\beta p_t)\) is a subgradient of \( u \) at
\((x, x_{t+1})\). Since \( w_{2T+1} \) is a subgradient of \( \beta^{2T+1}V \) at \( x_{2T+1} \), it follows that \( p_{2T+1} = \beta^{2(t+1)}w_{2T+1} \)
is a subgradient of \( V \) at \( x_{2T+1} \). By definition,

\[
(p_{2t}, -\beta p_{2t}) = (-\beta^{2t}w_{2t}, -\beta^{2t}w_{2t}) = \beta^{2t}(-w_{2t}, -w_{2t}).
\]

Since \((-w_{2t}, -w_{2t})\) is a subgradient of \( \beta^{2t}u \) at \((x, x_{2t+1})\), it follows that \((p_{2t+1}, -\beta p_{2t+1})\) is a
subgradient of \( u \) at \((x, x_{2t+1})\). Again by definition,

\[
(p_{2t+1}, -\beta p_{2t+1}) = (\beta^{2(t+1)}w_{2t+1}, \beta^{2(t+1)}w_{2t+2}) = \beta^{2(t+1)}(w_{2t+1}, w_{2t+2}).
\]

Since \((w_{2t+1}, w_{2t+2})\) is a subgradient of \( \beta^{2t+1}u \) at \((x, x_{2t+1})\), it follows that \((p_{2t+1}, -\beta p_{2t+1})\)
is a subgradient of \( u \) at \((x, x_{2t+1})\).

I next show that for \( t = 0, \ldots, 2T \), \( p_t \) is a subgradient of \( V \) at \( x_t \). Given \( t \) and \( x_t \in X \),
let \((x_{t+1}, \ldots, x_{2T+1})\) be such that \( x_{t+s+1} \in G(x_t) \), for all \( s = t, \ldots, 2T \), and

\[
V(x_t) = u(x, x_{t+1}) + \beta u(x, x_{t+1}) + \ldots + \beta^{2T-t}u(x, x_{2T+1}) + \beta^{2T+1}V(x_{2T+1}).
\]

By the subgradient properties of the \( p_s \),

\[
V(x_t) = u(x, x_{t+1}) + \beta u(x, x_{t+1}) + \ldots + \beta^{2T-t}u(x, x_{2T+1}) + \beta^{2T+1}V(x_{2T+1})
\]

\[
\leq u(x, x_{t+1}) + p_t(x - x_{t+1}) - \beta p_{t+1}(x_{t+1} - x_{t+1})
\]

\[
+ \beta u(x, x_{t+1}) + \beta p_{t+1}(x_{t+1} - x_{t+1}) - \beta^{2t+1}p_{t+2}(x_{t+2} - x_{t+2}) + \ldots
\]

\[
+ \beta^{2T-t}u(x_{2T+1} - x_{2T+1}) + \beta^{2T+1}p_{2T+1}(x_{2T+1} - x_{2T+1})
\]

\[
+ \beta^{2T+1}V(x_{2T+1}) + \beta^{2T+1}p_{2T+1}(x_{2T+1} - x_{2T+1})
\]

11
\[= u(\underbrace{x_t, x_{-t+1}}_{t,T}) + \ldots + \beta^{2T-t}u(\underbrace{x_{-2T}, x_{-2T+1}}_{-2T+1}) + \beta^{2T+1-t}V(\underbrace{x_{-2T+1}}_{-2T+1}) = p_{t,T}(x_t - x_{-t})
\]
\[= V(x_t) + p_{t,T}(x_t - x_{-t}).\]

The second equation is obtained by canceling terms. Hence \(p_{t,T}\) is a subgradient of \(V\) at \(x_{-t}\). \(\blacksquare\)

**Proof of theorem 11.15:** For each \(T = 1, 2, \ldots \), let \((p_{0,T}, p_{1,T}, \ldots, p_{2T+1,T})\) be a sequence of subgradients as in lemma 11.22. I find bounds on the components of \(p_{T,t}\) for each \(t\), and then apply a Cantor diagonal argument to define the infinite sequence \((p_0, p_1, \ldots)\).

We know that \(x_{-t} \in \text{int } X\) and, for \(t = 1, 2, \ldots\), \(x_{-t} \in \text{int } G(\underbrace{x_{-t}}_{-t-1})\). Therefore there is a positive number \(\varepsilon\) such that for each \(n = 1, \ldots, N\), \(x_{n}^{+} = x_{-t} = \underbrace{-\varepsilon e}_{0} + \underbrace{e}_{0} \) and \(x_{n}^{-} = x_{-t} = \underbrace{-\varepsilon e}_{0} + \underbrace{e}_{0}\) belong to \(X\), where \(e\) is the \(n\)th standard basis vector of \(\mathbb{R}^N\). Similarly, for each \(t = 1, 2, \ldots\), there is a positive number \(\varepsilon\) such that \(x_{n}^{+} = x_{-t} = \underbrace{-\varepsilon e}_{0} + \underbrace{e}_{0}\) and \(x_{n}^{-} = x_{-t} = \underbrace{-\varepsilon e}_{0} + \underbrace{e}_{0}\) belong to \(G(\underbrace{x_{-t}}_{-t-1})\).

If \(t = 0, 1, \ldots, T\), then by the subgradient property of \(p_{T,t}\),
\[u(x_{n}^{+}, x_{-t+1}) \leq u(x_{-t}, x_{-t+1}) + p_{T,t}(x_{n}^{+} - x_{-t}) = u(x_{-t}, x_{-t+1}) - \varepsilon p_{T,t},\]
so that
\[p_{T,t} \leq e^{-\varepsilon}[u(x_{-t}, x_{-t+1}) - u(x_{n}^{+}, x_{-t+1})] = b_{tn}.\]

Notice that \(b_{tn}\) may be negative, since it has not been assumed that \(u(x, y)\) is non-decreasing with respect to \(x\). Again by the subgradient property of \(p_{T,t}\),
\[u(x_{n}^{+}, x_{-t+1}) \leq u(x_{-t}, x_{-t+1}) + p_{T,t}(x_{n}^{+} - x_{-t}) = u(x_{-t}, x_{-t+1}) + \varepsilon p_{T,t},\]
so that
\[p_{T,t} \geq e^{-\varepsilon}[u(x_{n}^{+}, x_{-t+1}) - u(x_{-t}, x_{-t+1})] = b_{tn}.\]

In summary, for each non-negative integer \(t\), there are numbers \(b_{tn}\) and \(b_{tn}\) such that
\[b_{tn} \leq p_{T,t} \leq b_{tn}.\]
for all n and for T such that $2T + 1 \geq t$. (The concavity of $u$ implies that $b_{-ln} \leq b_{-ln}$.)

I now apply the Cantor diagonal argument. Because the sequence $p^T_0, T = 1, 2, \ldots$, belongs to a compact set, by the Bolzano-Weierstrass theorem 4.12 there is a subsequence of the sequence $T_0$, call it $T^0_k$, where $k = 1, 2, \ldots$, such that $p^T_{0k}$ converges. Similarly there exists a subsequence of $T^0_k$, call it $T^1_{1k}$, such that $p^T_{1k}$ converges. Suppose that subsequences $T^0_k, T^1_1, \ldots, T^{s-1}_{sk}$ have been defined such that $T^{s-1}_{sk}$ is a subsequence of $T^s_k$, for $s = 0, 1, \ldots, S - 1$, and such that $p^T_{sk}$ converges, for $s = 0, 1, \ldots, S$, as $k$ goes to infinity. Then because the $p^T_{sk}$ belong to a compact set, there exists a subsequence of $T^s_k$, call it $T^{s+1}_{sk}$, such that $p^T_{sk}$ converges. By induction on $S$, I have defined an infinite sequence of subsequences, $(T^s_k)_{k=1}^{S+1}$, for $S = 0, 1, 2, \ldots$, such that, for all $S$, $T^{S+1}_{sk}$ is a subsequence of $T^0_k$, and $p^T_{sk}$ converges as $k$ goes to infinity.

Let $T = T^s_k$, for $k = 1, 2, \ldots$. Since $T^s_k$ is a subsequence of $T^s_k$ and $p^T_{sk}$ converges, it follows that $p^T_{sk}$ converges, for all $S$, as $k$ goes to infinity. Let $p = \lim_{k \to \infty} p^T_{sk}$, for $t = 0, 1, 2, \ldots$. Since $b_{-ln} \leq p_{-ln} \leq b_{-ln}$, for all $n$, $p_t$ is a well-defined vector in $R^n$.

For each $t = 0, 1, \ldots$, we know from lemma 11.22 that

$$u(x_t, x_{t+1}) \leq u(x_{-l_t}, x_{-l_t}) + p^T_{sk}(x_t - x_{-l_t}) - \beta p^T_{sk}(x_{t+1} - x_{-l_t+1})$$

provided $0 \leq t < 2T + 1$. We also know that for $t = 0, 1, \ldots, 2T + 1$,

$$V(x_t) \leq V(x_{-l_t}) + p^T_{sk}(x_t - x_{-l_t}).$$

Passing to the limit in these inequalities, we see that

$$u(x_t, x_{t+1}) \leq u(x_{-l_t}, x_{-l_t}) + p_t(x_t - x_{-l_t}) - \beta p_{t+1}(x_{t+1} - x_{-l_t+1})$$

and

$$V(x_t) \leq V(x_{-l_t}) + p_t(x_t - x_{-l_t}).$$

That is, $(p_t, -\beta p_{t+1})$ is a subgradient of $u$ at $(x_{-l_t}, x_{-l_t+1})$ and $p_t$ is a subgradient of $V$ at $x_{-l_t}$. $\blacksquare$

13