I. $\Delta$-Monotonicity of Vector-Valued Functions.

I.1 $\Delta$-Monotonicity of $F$ and Positive Semi-Definite $DF$.

Let $D$ be an open convex subset of $\mathbb{R}^n$, and let $F: D \to \mathbb{R}^n$.

**Proposition I.1:** Suppose that $F$ is differentiable. Then the following two conditions are equivalent:

(i) $[F(x') - F(x)][x' - x] \geq 0$ for every $x, x' \in D$,

(ii) the matrix $DF(x)$ is positive semi-definite for every $x \in D$.

**Proof:** Consider arbitrary $x, x' \in D$, and denote $x' - x$ by $z$. Further, define

$$x(\lambda) = \lambda x' + (1 - \lambda)x, \quad \text{for } \lambda \in [0, 1]$$

It holds, $x(0) = x$, $x(1) = x'$, and $x(\lambda) = x + \lambda z$. Next, define function $g: [0, 1] \to \mathbb{R}$ by

$$g(\lambda) = z[F(x(\lambda)) - F(x)].$$

Note that $g(0) = 0$, $g(1) = [x' - x][F(x') - F(x)]$, and $g'(\lambda) = zDF(x(\lambda))z$.

Suppose that (i) holds. Since $g(\lambda) = \frac{1}{\lambda}[x(\lambda) - x][F(x(\lambda)) - F(x)]$ for $\lambda > 0$, it follows that $g(\lambda) \geq 0$. Therefore, $g$ has a minimum at $\lambda = 0$. This implies $g'(0) \geq 0$, which is $zDF(x)z \geq 0$. Since $z$ was arbitrary, we obtain (ii).

Conversely, suppose that (ii) holds. Then $g'(\lambda) \geq 0$ for every $\lambda \in [0, 1]$. So function $g$ is non-decreasing and hence $g(1) \geq g(0) = 0$. This implies (i).

QED
Condition (i) of Proposition I.1 will be called **\Delta-monotonicity** for it can be imprecisely written as $\Delta F \Delta x \geq 0$. This is different from the usual condition of $F$ being nondecreasing. Function $F : D \to \mathbb{R}^n$ is **nondecreasing** on $D \subset \mathbb{R}^n$ if

$$x \leq x' \implies F(x) \leq F(x') \quad (1)$$

for every $x, x' \in D$. Both inequalities in (1) are vectorial inequalities in $\mathbb{R}^n$.

\Delta-monotonicity (i) and property (1) are unrelated (except for when $n = 1$). Neither (i) implies (1), nor the opposite.

If function $F$ is differentiable on $D$, then a necessary and sufficient condition for being nondecreasing in the sense of (1) is that

$$DF(x) \geq 0 \quad \text{for every } x, \quad (2)$$

i.e., that the matrix $DF(x)$ is positive. That is,

$$\frac{\partial F_i}{\partial x_j}(x) \geq 0 \quad \forall i, j, \quad \text{for every } x.$$

Note that both \Delta-monotonicity and nondecreasing imply that $\frac{\partial F_i}{\partial x_j}(x) \geq 0$, but that is about as much as they have in common.
I.2 Convexity of $f$ and $\Delta$-Monotonicity of $Df$.

Let $D$ be an open convex subset of $\mathbb{R}^n$, and let $f : D \to \mathbb{R}$ be a real-valued function. Suppose that $f$ is differentiable and let $Df(x)$ be the derivative of $f$. $Df$ is a function from $D$ to $\mathbb{R}^n$.

**Proposition I.2:** Suppose that $f : D \to \mathbb{R}$ is continuously differentiable. Then $f$ is convex if and only if $Df$ is $\Delta$-monotone.

**Proof:** First, assume that $f$ is convex. It follows that

$$Df(x)[y - x] \leq f(y) - f(x),$$

for every $x, y \in D$, see MWG, Theorem M.C.1. This implies

$$[Df(x) - Df(y)][x - y] \geq 0,$$

for every $x, y$. Hence, $Df$ is $\Delta$-monotone.

For the converse, let $x, y$ be arbitrary and consider $x(\lambda) = \lambda y + (1 - \lambda)x$, for $\lambda \in [0, 1]$. It holds, $x(0) = x$, $x(1) = y$, and $x(\lambda) = x + \lambda[y - x]$. Define function $g : [0, 1] \to \mathbb{R}$ by

$$g(\lambda) = f(x(\lambda)).$$

Function $g$ is differentiable with $g'(\lambda) = Df(x(\lambda))[y - x]$. We shall show that $g'$ is nondecreasing, that is $g'(\lambda') \geq g'(\lambda)$ for $\lambda' > \lambda$. Indeed,

$$g'(\lambda') - g'(\lambda) = [Df(x(\lambda')) - Df(x(\lambda))[x(\lambda') - x(\lambda)] \frac{1}{(\lambda' - \lambda)} \geq 0.$$
where we used $\Delta$-monotonicity of $Df$. Since $g'$ is a nondecreasing function of single variable $\lambda$, it follows that $g$ is convex on $[0, 1]$. This imples that

$$f(x(\lambda)) \leq \lambda f(x) + (1 - \lambda) f(y),$$

for every $\lambda \in [0, 1]$. Thus $f$ is convex.
II. Theorem of the Maximum

There are two sets $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$. Further, there is a correspondence $\varphi$ mapping $S$ into the sets of subsets of $T$ and a function $f : S \times T \to \mathbb{R}$. That is, $\varphi(x)$ is a subset of $T$ for every $x \in S$, and $f(x, t)$ is a real number for every $x \in S$ and $t \in T$.

We are interested in the constrained maximization problem with $f$ as the objective function and $\varphi$ as the constraint. That is, given $x \in S$,

$$\max_t f(x, t)$$

subject to $t \in \varphi(x)$. 

We denote by $V(x)$ the value function of (3) and by $\mu(x)$ the set solutions. Formally,

$$V(x) = \max_{t \in \varphi(x)} f(x, t) \quad \text{and} \quad \mu(x) = \{t \in \varphi(x) : f(x, t) = V(x)\}. \quad (4)$$

**Interpretation:** Think about an economic agent whose environment is described by a vector $x \in S$. The agent’s set of actions is $T$, but when the environment is $x$, she is restricted to choose her action only from the subset $\varphi(x)$. Her utility of action $t$ is $f(x, t)$, when the environment is $x$. Her objective is to choose an action in $\varphi(x)$ to maximize her utility.
We shall assume that the set $T$ is **compact**.

Correspondence $\varphi$ is said to be **continuous** if it is lower hemi-continuous and upper hemi-continuous. These are defined as follows:

- **(LHC)** $\varphi$ is **lower hemi-continuous** at $x$ if for every sequence $\{x_n\}$ in $S$ converging to $x$ and every $t \in \varphi(x)$, there exists a sequence $\{t_n\}$ in $T$ such that $t_n \in \varphi(x_n)$ and $\{t_n\}$ converges to $t$.

- **(UHC)** $\varphi$ is **upper hemi-continuous** at $x$ if for every sequence $\{x_n\}$ in $S$ converging to $x$ and every sequence $\{t_n\}$ in $T$ converging to $t$, with $t_n \in \varphi(x_n)$, it holds that $t \in \varphi(x)$.

Our definition of UHC is the closed graph property. MasColell, Whinston and Green give definitions of LHC and UHC in Appendix M.H, pg. 949-951. Their definition of upper hemi-continuity is different, but if the range of $\varphi$ (i.e., the set $T$) is compact as assumed, then their definition is equivalent to the above one. Note that upper hemi-continuous correspondence $\varphi$ must have compact values $\varphi(x)$.

**Theorem II.1:** Suppose that the set $T$ is compact. If correspondence $\varphi$ is continuous on $S$ and function $f$ is continuous on $S \times T$, then

- (i) $V$ is continuous on $S$, and
- (ii) $\mu$ is an upper hemi-continuous correspondence on $S$. 
Proof: (i) Let \( \{x_n\} \) be a sequence of vectors in \( S \) converging to \( x \). We have to show that \( \lim_n V(x_n) = V(x) \). Since \( \varphi(x_n) \) is a compact set for every \( n \), there exist \( t_n \in \varphi(x_n) \) such that \( V(x_n) = f(x_n, t_n) \). Since the set \( T \) is compact, sequence \( \{t_n\} \) must have a convergent subsequence with a limit \( \bar{t} \in T \). We switch to that subsequence of \( \{t_n\} \), but we retain the same notation; i.e., we keep \( \{t_n\} \) and assume that it converges to \( \bar{t} \). Upper hemi-continuity of \( \varphi \) implies that \( \bar{t} \in \varphi(x) \). By continuity of \( f \), we have \( \lim_n f(x_n, t_n) = f(x, \bar{t}) \).

Since \( f(x, \bar{t}) \leq V(x) \), it follows that

\[
\lim_n V(x_n) \leq V(x).
\]

To prove the opposite inequality, we note that \( V(x) = f(x, t) \) for some \( t \in \varphi(x) \) since \( \varphi(x) \) is a compact set. Lower hemi-continuity of \( \varphi \) at \( x \) implies that there is sequence \( \{\tilde{t}_n\} \) converging to \( t \) such that \( \tilde{t}_n \in \varphi(x_n) \) for every \( n \). We have \( f(x_n, \tilde{t}_n) \leq V(x_n) \). Using continuity of \( f \), we obtain \( \lim_n f(x_n, \tilde{t}_n) = f(x, t) \).

Consequently

\[
\lim_n V(x_n) \geq V(x).
\]

This concludes the proof of (i)

(ii) Consider two sequences: \( \{x_n\} \) in \( S \) converging to \( x \), and \( \{t_n\} \) in \( T \) converging to \( t \) such that \( t_n \in \mu(x_n) \). We have to show that \( t \in \mu(x) \).

We first observe that upper hemi-continuity of \( \varphi \) implies that \( t \in \varphi(x) \). Next, consider arbitrary \( \bar{t} \in \varphi(x) \). Lower hemi-continuity of \( \varphi \) at \( x \) implies
that there is a sequence \( \{t_n\} \) converging to \( t \) such that \( t_n \in \varphi(x_n) \) for every \( n \). Clearly then \( f(x_n, t_n) \geq f(x_n, \tilde{t}_n) \). Passing to the limit with \( n \) and using continuity of \( f \), we obtain \( f(x, t) \geq f(x, \tilde{t}) \). Since \( \tilde{t} \) was arbitrary in \( \varphi(x) \), this implies that \( t \in \mu(x) \). This concludes the proof of (ii).

**Remarks:**

- The assumption that set \( T \) is compact can be dropped. Then the MWG definition of upper hemi-continuity has to be used. Note that that definition requires that correspondence \( \varphi \) be compact-valued.

- One application of the Theorem of the Maximum II.1 is in producer theory. We set \( S \) as the set of price vectors, \( T \) as the production set, i.e., \( T = Y \), function \( f \) as \( f(p, y) = py \), and correspondence \( \varphi \) as \( \varphi(p) = Y \). Assuming that \( Y \) is compact, Theorem II.1 implies continuity of the profit function and upper hemi-continuity of the supply correspondence (Proposition 6.3 (iii)).
III. Kuhn-Tucker Theorems

III.1 Constrained Maximization: Necessary Conditions.

Function $F : \mathbb{R}_+^n \to \mathbb{R}$ is the objective function; functions $g^j : \mathbb{R}_+^n \to \mathbb{R}$, for $j = 1, \ldots, k$, are constraint functions. Assume that $F$ and $g^j$ are differentiable, with partial derivatives $\frac{\partial F}{\partial x_i}$ and $\frac{\partial g^j}{\partial x_i}$ denoted by $\partial_i F$ and $\partial_i g^j$, respectively.

The constrained maximization problem (with nonnegativity constraints) is

$$\max_x F(x) \quad (1)$$

subject to $g^1(x) \geq 0,$

.....,

$g^k(x) \geq 0,$

$x_1 \geq 0, \ldots, x_n \geq 0.$

We write the Lagrangian as

$$\mathcal{L}(\lambda^1, \ldots, \lambda^k, x) = F(x) + \sum_{j=1}^{k} \lambda_j g^j(x),$$

where $\lambda_j \geq 0$, for $j = 1, \ldots, k$, are the Lagrange multipliers. We use $\lambda$ to denote the $k$-vector of multipliers.
Kuhn-Tucker conditions for \( x^* \geq 0 \) and \( \lambda^* \geq 0 \) are:

for all \( i = 1, \ldots, n \) and \( j = 1, \ldots, k \),

\[
\partial_i F(x^*) + \sum_{j=1}^{k} \lambda^*_j \partial_i g^j(x^*) \leq 0, \quad \text{and if } x^*_i > 0, \text{ then } " = 0", \tag{2a}
\]
\[
g^j(x^*) \geq 0, \quad \text{and if } \lambda^*_j > 0, \text{ then } " = 0". \tag{2b}
\]

Where do these conditions come from? Think about maximizing Lagrangian \( L(\lambda, x) \) with respect to \( x \) and minimizing it with respect to \( \lambda \), unconstrained, except for \( x \geq 0 \) and \( \lambda \geq 0 \). This is the saddle-point. K-T conditions (2) are FOCs for such max-min (or saddle-point) problem.

**Theorem (Kuhn-Tucker):** If \( x^* \geq 0 \) is a solution to the constrained maximization problem, and the Constraint Qualification Condition holds, then \( x^* \) and some \( \lambda^* \geq 0 \) satisfy K-T conditions (2).

**Constraint Qualification Condition:**

(i) Kuhn-Tucker original – don’t touch it.

(ii) \( g^j \) concave for all \( j \), and Slater’s condition, that is, there is some \( x^0 \geq 0 \) with \( g^j(x^0) > 0 \) for all \( j \).

(iii) rank condition (see Takayama 1.D.4, or Varian, ch 27),

(iv) \( g^j \) linear for all \( j \), (Arrow-Hurwicz-Uzawa, see Takayama 1.D.4)
III.2 Sufficiency of Kuhn-Tucker Conditions.

The most standard theorem is:

**Theorem S1:** Suppose that $F$ and $g^1, \ldots, g^k$ are all concave functions. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions (2), then $x^*$ is a solution to the constrained maximization problem.

A better theorem is due to Arrow and Enthoven (1961).

**Theorem S2:** Suppose that $F$ and $g^1, \ldots, g^k$ are all quasi-concave functions and some “mild” condition holds. If $x^* \geq 0$ and $\lambda^* \geq 0$ satisfy K-T conditions (2), then $x^*$ is a solution to the constrained maximization problem.

The extra (“mild”) condition is not needed if $F$ is concave (and $g^1, \ldots, g^k$ are quasi-concave). See Takayama 1.E for three versions of the condition.

Quasi-concavity (and therefore also concavity) of functions $g^j$ implies that the constraint set, i.e. the set of $x \geq 0$ satisfying $g^1(x) \geq 0, \ldots, g^k(x) \geq 0$, is convex.
III.3 Constrained Minimization

The constrained minimization problem (with nonnegativity constraints) is

$$\min_x F(x)$$  \quad (3)$$

subject to $g^i(x) \leq 0$, \ldots, $g^k(x) \leq 0$,

$$x_1 \geq 0, \ldots, x_n \geq 0.$$  

The Lagrangian is

$$\mathcal{L}(\lambda, x) = F(x) + \sum_{j=1}^{k} \lambda_g^j g^j(x).$$

**Kuhn-Tucker conditions** for $x^* \geq 0$ and $\lambda^* \geq 0$ are,

for all $i = 1, \ldots, n$ and $j = 1, \ldots, k$,

$$\partial_i F(x^*) + \sum_{j=1}^{k} \lambda^*_g \partial_i g^j(x^*) \geq 0, \quad \text{and if } x^*_i > 0, \text{ then } " = 0", \quad (4a)$$

$$g^j(x^*) \leq 0, \quad \text{and if } \lambda^*_j > 0, \text{ then } " = 0". \quad (4b)$$

The corresponding saddle-point problem is to *minimize* Lagrangian $\mathcal{L}(\lambda, x)$ with respect to $x$ and *maximize* it with respect to $\lambda$ for $x \geq 0$ and $\lambda \geq 0$.

The Kuhn-Tucker Theorem holds with no change for the constrained minimization problem. However, in constraint qualification conditions concavity of functions $g^j$, if present, has to be replaced by their convexity. This guarantees convexity of the constraint set described here by inequalities $g^j(x) \leq 0$.  

12
Theorems S1 and S2 continue to hold with concavity (quasi-concavity) of functions $F$ and $g^j$ replaced by their convexity (quasi-convexity, respectively).

III.4 Remarks:

- **Applications** of K-T theorems in microeconomics:
  
  (i) Consumer theory: utility maximization subject to budget constraint, and expenditure minimization.

  (ii) Welfare economics: Characterization of Pareto optimal allocations as solutions to maximization of a welfare function subject to resource constraints, or maximization of one agent’s utility subject to constraints on other agents’ utilities and resource constraints.

  (iii) Producer theory: cost minimization.

- There are versions of K-T theorems for maximization and minimization with mixed constraints, i.e., when some constraints are of the equality form, $g^j(x) = 0$. See Sundaram [2], Section 6.4.

- K-T theorems hold for local maxima (minima) as well.

III.5 Example: Consider the following constrained maximization problem:

$$\begin{align*}
\text{maximize} & \quad \ln(x_1 + 1) + \ln(x_2 + 1) \\
\text{subject to} & \quad p_1 x_1 + p_2 x_2 \leq m \\
& \quad x_1 \geq 0, \quad x_2 \geq 0,
\end{align*}$$

where $p_1 > 0$, $p_2 > 0$ and $m > 0$.

In order to derive the solution (as a function of parameters $p_1, p_2$ and $m$) we write the Kuhn-Tucker first-order conditions (2) as

$$(1) \quad \frac{1}{x_1^* + 1} - \lambda^* p_1 \leq 0, \quad \text{and if} \quad x_1^* > 0, \quad \text{then} \quad "= 0".$$ 

$$(2) \quad \frac{1}{x_2^* + 1} - \lambda^* p_2 \leq 0, \quad \text{and if} \quad x_2^* > 0, \quad \text{then} \quad "= 0".$$ 

$$(3) \quad p_1 x_1^* + p_2 x_2^* \leq m, \quad \text{and if} \lambda^* > 0, \quad \text{then} \quad "= 0".$$ 

with $x^* \geq 0$ and $\lambda^* \geq 0$.

Note that (3) holds with equality since it follows from (1) that $\lambda^* > 0$.

We solve inequalities (1-3) by considering cases:

**Case 1.** \(x_1^* > 0, \ x_2^* > 0\).

Then (1) and (2) hold with equalities. Solving (1), (2) and (3) we find 

$$x_1^* = \frac{m + p_2 - p_1}{2p_1}, \quad \text{and} \quad x_2^* = \frac{m + p_1 - p_2}{2p_2}, \quad \text{and} \quad \lambda^* = \frac{2}{m + p_1 + p_2}.$$ 

For \(x_1^*\) and \(x_2^*\) to be strictly positive, it has to be that \(m + p_2 > p_1\) and \(m + p_1 > p_2\). Thus Case 1 applies with \(x_1^*\) and \(x_2^*\) as listed above if \(m + p_2 > p_1\) and \(m + p_1 > p_2\).
Case 2. $x_1^* > 0, x_2^* = 0$.

(3) implies that $x_1^* = \frac{m}{p_1}$. Since (1) holds with equality, we solve it for $\lambda^* = \frac{1}{m + p_1}$. Next we need to verify inequality (2). It states

$$1 - \frac{p_2}{m + p_1} \leq 0,$$

and it holds if $p_2 \geq m + p_1$. Thus Case 2 applies (with $x_1^* = \frac{m}{p_1}, x_2^* = 0$) if $p_2 \geq m + p_1$.

Case 3. $x_1^* = 0, x_2^* > 0$.

This case is very similar to Case 2. From (3) and (2) we obtain $x_1^* = \frac{m}{p_2}$, $\lambda^* = \frac{1}{m + p_2}$. Verifying inequality (1), we obtain $p_1 \geq m + p_2$. Thus Case 3 applies (with $x_1^* = 0, x_2^* = \frac{m}{p_2}$) if $p_1 \geq m + p_2$.

The case $x_1^* = x_2^* = 0$ cannot hold since it violates equation (3). This concludes our solution to the K-T conditions.

Since utility function is concave and the constraint function is concave (in fact, it is linear) K - T conditions are sufficient (Theorem S1). Hence, the solution to K-T conditions is a constrained maximizer. Further, since the Slater’s condition holds, every constrained maximizer has to satisfy K - T conditions.
Mathematical Appendix IV

IV.1 Proof of Theorem 15.3: We first prove the following

Lemma 15.4: If $u$ is concave and supermodular, then

$$u(\lambda[x \lor y] + (1 - \lambda)y) - u(y) \geq u(x) - u(\lambda[x \land y] + (1 - \lambda)x),$$  \hspace{1cm} (47)

for every $x, y \in \mathbb{R}^L_+$ and $0 \leq \lambda \leq 1$.

Proof of Lemma 15.4 The following two inequalities follow from concavity of $u$

$$u(\lambda[x \lor y] + (1 - \lambda)y) \geq \lambda u(x \lor y) + (1 - \lambda)u(y),$$  \hspace{1cm} (48)

$$u(\lambda[x \land y] + (1 - \lambda)x) \geq \lambda u(x \land y) + (1 - \lambda)u(x)$$  \hspace{1cm} (49)

Also, because of supermodularity (34),

$$u(x \lor y) + u(x \land y) \geq u(x) + u(y).$$  \hspace{1cm} (50)

If we multiply (50) by $\lambda$ and sum the resulting inequality side-by-side with (48) and (49), we obtain (47).

We return to the proof of Theorem 15.3. Of course, we only need to consider $w' > w$. Let $y = x^*(p, w)$ and $x = x^*(p, w')$. Since $u$ is l.n.s., we have $py = w$ and $px = w'$. Clearly, $p[x \land y] \leq w$. Since $px > w$, there exists $0 \leq \lambda < 1$ such that $p(\lambda[x \land y] + (1 - \lambda)x) = w$. Denote $\lambda[x \land y] + (1 - \lambda)y$ by $\tilde{z}_\lambda$ and $\lambda[x \lor y] + (1 - \lambda)x$ by $\bar{z}^\lambda$. Since $\tilde{z}_\lambda + \bar{z}^\lambda = x + y$ (this follows from (32)), we have $p\bar{z}^\lambda = w'$.  

1
Since $y$ is the unique utility maximizer at $w$ and $p_{z\lambda} = w$, we have $u(y) \geq u(\bar{z}_\lambda)$. Lemma 15.4 implies that $u(\bar{z}^\lambda) \geq u(x)$. Since $x$ is the unique utility maximizer at $w'$ and $p_{\bar{z}\lambda} = w'$, it must be $\bar{z}^\lambda = x$. Then also $\bar{z}_\lambda = y$. It can be shown (see Figure 1) that $\bar{z}^\lambda = x$ if and only if $x = x \lor y$. Similarly, $\bar{z}_\lambda = y$ if and only if $y = x \land y$. But if $x = x \lor y$ and $y = x \land y$, then $y \leq x$. This concludes the proof.