Definition 39 If $V$ is a vector space over $\mathbb{R}$, and $U \subset V$, the set $U$ is called convex if $\alpha u_1 + (1 - \alpha) u_2 \in U$ for all $u_1, u_2 \in U, 0 \leq \alpha \leq 1$.

For convex subsets in $\mathbb{R}^n$ (which is also an inner-product space) there is an important result called the separating hyperplane theorem. It is also true in more general inner product spaces, but the proof is more technical.

Lemma 40 Suppose $U \subset \mathbb{R}^n$ is closed, and convex, and $x \in \mathbb{R}^n$, with $x \notin U$. There is a vector $c \in \mathbb{R}^n$ ($c \neq 0_n$) and $c^T x < c^T u$ for all $u \in U$.

Proof: Let $\delta := \inf_{u \in U} \|u - x\|$. Since $U$ is closed, and $x \notin U$, it follows that $\delta > 0$. Also, since $U$ is closed, this infimum is actually achieved—that is, there is a point $u_0 \in U$ such that $\|u_0 - u\| = \delta$. Now, for any $u \in U$, and $\alpha \in [0,1]$, we have that $\alpha u + (1 - \alpha) u_0 \in U$. Hence,

$$\|u_0 - x\| \leq \|\alpha u + (1 - \alpha) u_0 - x\| = \|u_0 - x + \alpha (u - u_0)\|$$

for all $\alpha \in [0,1]$. Squaring both sides gives

$$0 \leq \alpha^2 \|u - u_0\|_2^2 + 2\alpha (u_0 - x)^T (u - u_0)$$

This is a quadratic function, which must be nonnegative for $\alpha \in [0,1]$, hence it must be that $(u_0 - x)^T (u - u_0) \geq 0$. Now, define $c := u_0 - x$. Note that $c \neq 0_n$, and for any $u \in U$, we have

$$c^T (u - x) = c^T (u - u_0 + u_0 - x)$$

$$= (u_0 - x)^T (u - u_0) + \|u_0 - x\|^2$$

$$\geq \|u_0 - x\|^2$$

$$= \delta$$

$$> 0, \text{ as desired.}$$

Remark: Once such a $c$ is found, for any positive scalar $\gamma > 0$, $\tilde{c} := \gamma c$ also satisfies the inequality.
More separating hyperplane theorems

**Theorem 41** Suppose that $U \subset \mathbb{R}^n$ is convex, and $x$ is a boundary point of $U$ (every open neighborhood containing $x$ contains points $y \neq x$ and $z \neq x$ such that $y \in U, z \notin U$). Then, there is a vector $c \in \mathbb{R}^n, c \neq 0_n$ such that $c^T x \leq c^T u$ for all $u \in U$.

**Proof:** Let $\{z_n\}_{n=1}^{\infty}$ be a sequence of points not in $U$, but converging to $x$. Apply Lemma 40 and the Remark to obtain a vector $c_n$ with the property that $\|c_n\| = 1$ and $c_n^T z_n < c_n^T u$ for all $u \in U$. Since $\{c_n\}_{n=1}^{\infty}$ is a bounded sequence in $\mathbb{R}^n$, it has a convergent subsequence $c_{n_k} \rightarrow_k c$. We know that for all $u \in U$, and all $k$

$$c_{n_k}^T z_{n_k} < c_{n_k}^T u$$

Taking limits as $k \rightarrow \infty$ gives $c^T x \leq c^T u$. This holds for any $u \in U$, as desired.
Convex Functions

Suppose that $V$ is a vector space, and $U \subset V$ is a convex set.

**Definition 42** A function $f : U \to \mathbb{R}$ is called convex if

$$f(\alpha u_1 + (1 - \alpha)u_2) \leq \alpha f(u_1) + (1 - \alpha)f(u_2)$$

for all $u_1, u_2 \in U, 0 \leq \alpha \leq 1$.

Convex functions on $\mathbb{R}^n$ are continuous on the interior of their domain.

**Theorem 43** Suppose $U \subset \mathbb{R}^n$ is convex, and $f : U \to \mathbb{R}$ is convex. Then $f$ is continuous at every $x$ in the interior of $U$.

**Proof:** The proof is complicated enough to skip - in all of the cases we deal with, continuity will be apparent, and not need to be deduced from convexity. The proof for $U \subset \mathbb{R}$ is not too bad (see Rudin, “Principles of Mathematical Analysis,” page 101, problem 23).
Sublevel Sets, and Epigraphs

For any $\beta \in \mathbb{R}$, define a set $L_\beta$, called a sublevel set of $f$, as $L_\beta := \{u \in U : f(u) \leq \beta\}$.

**Lemma 44** For any $\beta \in \mathbb{R}$, and any convex function $f : U \to \mathbb{R}$, the set $L_\beta$ is a convex set.

**Proof:** Let $u_1, u_2 \in L_\beta$. Hence, $f(u_1) \leq \beta, f(u_2) \leq \beta$. Take $0 \leq \alpha \leq 1$.

Then

$$f(\alpha u_1 + (1 - \alpha)u_2) \leq \alpha f(u_1) + (1 - \alpha)f(u_2)$$

$$\leq \alpha \beta + (1 - \alpha)\beta$$

$$= \beta$$

as desired. $\sharp$

The epigraph of $f$ is a subset of $V \times \mathbb{R}$, defined as

$$\text{epi}(f) := \{(u, \beta) : u \in U, \beta \geq f(u)\}$$

If $f$ is a convex function, and $U$ is a convex set, then $\text{epi}(f)$ is a convex set in $V \times \mathbb{R}$. 
Simple optimality

Suppose $U \subset V$ is a convex subset in a normed vector space $V$, and $f$ is a function mapping $U \rightarrow \mathbb{R}$.

**Definition 45** A point $u$ is called a local minimum of $f$ on $U$ if there exists an $\epsilon > 0$ such that for all $v \in U$, $\|v - u\| < \epsilon$, $f(v) \geq f(u)$.

**Definition 46** A point $u$ is called a global minimum of $f$ on $U$ if for all $v \in U$, $f(v) \geq f(u)$.

**Theorem 47** If $U$ is a convex set, and $f$ is a convex function, and $u$ is a local minimum of $f$ on $U$, then $u$ is a global minimum.

**Proof:** By contrapositive. Let $\epsilon > 0$. Now, assume that $u$ is not a global minimum of $f$. Then, there is a $\bar{u} \in U$, with $f(\bar{u}) < f(u)$. Choose

$$\alpha := \min \left\{ \frac{1}{2}, \frac{\epsilon}{2\|\bar{u} - u\|} \right\}$$

Note that $\|\alpha(\bar{u} - u)\| \leq \epsilon/2$. Define $u_\alpha := \alpha \bar{u} + (1 - \alpha)u$. Note that $u_\alpha \in U$, and

$$\|u_\alpha - u\| = \|\alpha \bar{u} + (1 - \alpha)u - u\|$$
$$= \|\alpha (\bar{u} - u)\|$$
$$\leq \frac{\epsilon}{2}$$
$$< \epsilon$$

Also,

$$f(u_\alpha) \leq \alpha f(\bar{u}) + (1 - \alpha) f(u)$$
$$= f(u) + \alpha [f(\bar{u}) - f(u)]$$
$$< f(u)$$

since $\alpha > 0$, and $f(\bar{u}) - f(u) < 0$. Hence, we have shown that for any $\epsilon > 0$, there is a point $u_\alpha$, $\|u_\alpha - u\| < \epsilon$, but $f(u_\alpha) < f(u)$. By definition, $u$ is not a local minimum of $f$. $\#$

---

*ME 234, UC Berkeley, Fall 2008, Packard*
Nonegative multiples of convex functions are convex.

**Lemma 48** If $f : U \to \mathbb{R}$ is convex, and $\gamma \in \mathbb{R}$, $\gamma \geq 0$, then the function $h$, defined by $h(v) := \gamma f(v)$, is convex.

Sums of convex functions are convex

**Lemma 49** Suppose $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ are convex. Then the function $h := f + g$ is convex.

**Proof:** Take any $u_1, u_2 \in U$, $0 \leq \alpha \leq 1$,

\[
h(\alpha u_1 + (1 - \alpha)u_2) = f(\alpha u_1 + (1 - \alpha)u_2) + g(\alpha u_1 + (1 - \alpha)u_2) \\
\leq [\alpha f(u_1) + (1 - \alpha) f(u_2)] + [\alpha g(u_1) + (1 - \alpha) g(u_2)] \\
= \alpha [f(u_1) + g(u_1)] + (1 - \alpha) [f(u_2) + g(u_2)] \\
= \alpha h(u_1) + (1 - \alpha) h(u_2)
\]
Pointwise maximums of collections of convex functions are convex.

**Lemma 50** Suppose that for each $\gamma \in \Gamma$, $f_\gamma : U \to \mathbb{R}$ is a convex function. Then the function $f$ defined by

$$f(x) := \sup_{\gamma \in \Gamma} f_\gamma(x)$$

is convex.

**Proof:** Take any $u_1, u_2 \in U$, $0 \leq \alpha \leq 1$. Now, choose $\epsilon > 0$. By definition, there is a $\tilde{\gamma} \in \Gamma$ such that

$$f(\alpha u_1 + (1 - \alpha)u_2) < f_{\tilde{\gamma}}(\alpha u_1 + (1 - \alpha)u_2) + \epsilon$$

But $f_{\tilde{\gamma}}$ is convex, so we have

$$f(\alpha u_1 + (1 - \alpha)u_2) \leq f_{\tilde{\gamma}}(\alpha u_1) + (1 - \alpha)f_{\tilde{\gamma}}(u_2) + \epsilon$$

$$\leq \alpha \sup_{\gamma \in \Gamma} f_\gamma(u_1) + (1 - \alpha)\sup_{\gamma \in \Gamma} f_\gamma(u_2) + \epsilon$$

$$= \alpha f(u_1) + (1 - \alpha) f(u_2) + \epsilon$$

(note that we used the fact that $\alpha \geq 0$ and $1 - \alpha \geq 0$). Hence, for every $\epsilon > 0$, we have shown that

$$f(\alpha u_1 + (1 - \alpha)u_2) < \alpha f(u_1) + (1 - \alpha)f(u_2) + \epsilon$$

which implies that

$$f(\alpha u_1 + (1 - \alpha)u_2) \leq \alpha f(u_1) + (1 - \alpha)f(u_2)$$

as desired. \#
**Linear and Affine functions**

**Definition 51** Suppose $V$ and $W$ are vector spaces, and $g : V \to W$. The function $g$ is **linear** if

\[
g(\alpha v_1 + \beta v_2) = \alpha g(v_1) + \beta g(v_2)
\]

for all $v_1, v_2 \in V$, $\alpha, \beta \in \mathbb{R}$,

**Definition 52** Suppose $V$ and $W$ are vector spaces, and $g : V \to W$. The function $g$ is **affine** if the function $h : V \to W$, defined as $h(v) := g(v) - g(0)$, is linear.

**Lemma 53** Suppose $V$ and $W$ are vector spaces, and $g : V \to W$ is affine. Then, for any $v_1, v_2 \in V$, $\alpha \in \mathbb{R}$,

\[
g(\alpha v_1 + (1 - \alpha)v_2) = \alpha g(v_1) + (1 - \alpha)g(v_2)
\]

**Proof:** Define $h$ as above, $h(v) := g(v) - g(0)$. Since $g$ is affine, $h$ is linear. Hence

\[
g(\alpha v_1 + (1 - \alpha)v_2) = h(\alpha v_1 + (1 - \alpha)v_2) + g(0)
\]

\[
= \alpha h(v_1) + (1 - \alpha)h(v_2) + g(0)
\]

\[
= \alpha [g(v_1) - g(0)] + (1 - \alpha) [g(v_2) + g(0)] + g(0)
\]

\[
= \alpha g(v_1) + (1 - \alpha)g(v_2) - (\alpha + 1 - \alpha)g(0) + g(0)
\]

\[
= \alpha g(v_1) + (1 - \alpha)g(v_2)
\]
Composing Convex with Affine

The composition of a convex function with an affine function is convex.

Lemma 54 Suppose $V$ and $W$ are vector spaces, and $g : V \rightarrow W$ is affine, and $h : W \rightarrow \mathbb{R}$ is convex. Then, the composition $f := h \circ g : V \rightarrow \mathbb{R}$ is convex.

Proof: Choose any $v_1, v_2 \in V$, $\alpha \in [0, 1]$.

\[
 f(\alpha v_1 + (1 - \alpha)v_2) := (h \circ g)(\alpha v_1 + (1 - \alpha)v_2) \\
= h(g((\alpha v_1 + (1 - \alpha)v_2))) \\
= h(\alpha g(u_1) + (1 - \alpha)g(u_2)) \\
\leq \alpha h(g(u_1)) + (1 - \alpha)h(g(u_2)) \\
= \alpha f(u_1) + (1 - \alpha)f(u_2)
\]

Example: Let $S_{n \times n}$ be the set of real, symmetric matrices. Define $f : S_{n \times n} \rightarrow \mathbb{R}$,

\[
f(W) := \lambda_{\text{max}}(W)
\]

The function $f$ is convex. How? For each $v \in \mathbb{R}^n, \|v\| = 1$, define $f_v : S_{n \times n} \rightarrow \mathbb{R}$ as $f_v(W) := v^T W v$. This is a linear function whose range is $\mathbb{R}$, and hence is convex. Moreover

\[
\lambda_{\text{max}}(W) = \max_{v \in \mathbb{R}^n, \|v\|=1} f_v(W)
\]

and hence is convex.
Subgradients of Convex functions

**Definition 55** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex. Pick $x \in \mathbb{R}^n$. Suppose that $v \in \mathbb{R}^n$ satisfies
\[
f(y) \geq f(x) + v^T(y - x)
\]
for all $y \in \mathbb{R}^n$. Note that the right-hand side of (3.9) is an affine function of $y$, and equals $f(x)$ at $y = x$. Then $v$ is called a subgradient of $f$ at $x$. We use $\partial f_x$ to denote the set of all such vectors $v$.

**Lemma 56** If $f$ is convex, then for each $x$, the set $\partial f_x$ is nonempty, closed, bounded and convex.

**Proof:** Since $f$ is convex, the set $\text{epi}(f)$ is a convex set. Pick $x \in \mathbb{R}^n$, and take the point $(x, f(x)) \in \text{epi}(f) \subset V \times \mathbb{R} = \mathbb{R}^{n+1}$. This point is on the boundary of $\text{epi}(f)$, and hence by the separating hyperplane theorem, there is a vector $c \in \mathbb{R}^{n+1}, c \neq 0_{n+1}$ such that
\[
\begin{bmatrix}
c_1^T & c_2
\end{bmatrix}
\begin{bmatrix}
y \\
\beta
\end{bmatrix}
\geq
\begin{bmatrix}
c_1^T & c_2
\end{bmatrix}
\begin{bmatrix}
x \\
f(x)
\end{bmatrix}
\]
for all $(y, \beta) \in \text{epi}(f)$. Equivalently,
\[
\begin{bmatrix}
c_1^T & c_2
\end{bmatrix}
\begin{bmatrix}
y \\
f(y) + \gamma
\end{bmatrix}
\geq
\begin{bmatrix}
c_1^T & c_2
\end{bmatrix}
\begin{bmatrix}
x \\
f(x)
\end{bmatrix}
\]
for all $y \in \mathbb{R}^n, \gamma \geq 0$. This simplifies to
\[
c_2 \gamma \geq c_1^T (x - y) + c_2 [f(x) - f(y)]
\]
for all $\gamma \geq 0$. Hence, $c_2 \geq 0$. Can $c_2 = 0$? Suppose it does, then $c_1^T (x - y) \geq 0$ for all $y \in \mathbb{R}^n$, which implies $c_1 = 0_n$. But $c \neq 0$, hence $c_2 > 0$. 

MR 234, UC Berkeley, Fall 2008, Packard

58
So now we can divide by $c_2$ to get

$$f(y) \geq f(x) + \frac{-1}{c_2} c_1^T (y - x)$$

for all $y \in \mathbb{R}^n$. This means that

$$\frac{-1}{c_2} c_1 \in \partial f_x$$

showing that $\partial f_x$ is not empty. Next, suppose that $\{v_i\}_{i=1}^{\infty}$ is a sequence of points $v_i \in \partial f_x$, with $v_i \to v$. Pick any $z \in \mathbb{R}^n$. For each $i$, we have

$$f(z) \geq f(x) + v_i^T (z - x)$$

Taking limits gives $f(z) \geq f(x) + v(z - x)$, which holds for any $z$, and hence $v \in \partial f_x$, so that $\partial f_x$ is closed. Suppose that $\partial f_x$ is not bounded. Then, there exist a sequence $\{v_i\}_{i=1}^{\infty}$ such that $\|v_i\| > i$ and $v_i \in \partial f_x$. Now, set $z_i := x + \frac{1}{\|v_i\|} v_i$. Note that

$$f(z_i) \geq f(x) + \|v_i\| > f(x) + i$$

Hence, $f$ is unbounded on the closed unit ball centered at $x$. But this contradicts that $f$ is continuous. Hence $\partial f_x$ is bounded. Finally, if $v_1, v_2 \in \partial f_x$ and $0 \leq \alpha \leq 1$. Take any $z \in \mathbb{R}^n$, which gives

$$f(z) \geq f(x) + v_1^T (z - x)$$
$$f(z) \geq f(x) + v_2^T (z - x)$$

Multiply the first by $\alpha$ and the second by $1 - \alpha$, and add, leaving

$$f(z) \geq f(x) + [\alpha v_1 + (1 - \alpha) v_2]^T (z - x)$$

which holds for any $z$. Hence $\alpha v_1 + (1 - \alpha) v_2 \in \partial f_x$. #
Subgradients for functions defined on $\mathbb{R}^{n \times m}$

We worked out definitions and formula for subgradients of convex functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$. It is straightforward to extend this to convex functions $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$. Let $\Psi : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{nm}$ and $\psi : \mathbb{R}^{nm} \rightarrow \mathbb{R}^{n \times m}$ be stacking and unstacking operators (reshape in MatLab) with

$$
\Psi \left( \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1m} \\
a_{21} & a_{22} & \cdots & a_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nm}
\end{bmatrix} \right) := \begin{bmatrix}
a_{11} \\
a_{21} \\
\vdots \\
a_{n1} \\
a_{12} \\
a_{22} \\
\vdots \\
a_{n2} \\
\vdots \\
a_{1m} \\
a_{2m} \\
\vdots \\
a_{nm}
\end{bmatrix}
$$

and

$$
\psi \left( \begin{bmatrix}
v_1 \\
v_2 \\
v_3 \\
\vdots \\
v_{n-m} 
\end{bmatrix} \right) := \begin{bmatrix}
v_1 & v_{n+1} & \cdots & v_{n*(m-1)+1} \\
v_2 & v_{n+2} & \cdots & v_{n*(m-1)+2} \\
\vdots & \vdots & \ddots & \vdots \\
v_n & v_{2n} & \cdots & v_{nm}
\end{bmatrix}
$$

It is easy to verify that $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ is convex if and only if $\tilde{f} : \mathbb{R}^{nm} \rightarrow \mathbb{R}$, defined as

$$
\tilde{f}(v) := f (\psi(v))
$$

is convex. Its also simple to verify that for $A, B \in \mathbb{R}^{n \times m}$

$$
\text{tr} (A^T B) = [\Psi(A)]^T \Psi(B)
$$
Hence, if \( f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \) is convex, then for each \( X \in \mathbb{R}^{n \times m} \), there is an element \( V \in \mathbb{R}^{n \times m} \) such that for all \( Z \in \mathbb{R}^{n \times m} \),

\[
f(Z) \geq f(X) + \text{tr} [V^T (Z - X)]
\]

\( V \) is called a subgradient of \( f \) at \( X \), and written \( V \in \partial f_X \).

Similar ideas hold for symmetric matrices, stacking and unstacking into vectors. However, viewing (for instance) \( 3 \times 3 \) symmetric matrices as \( 6 \times 1 \) vectors (as opposed to \( 9 \times 1 \)), there are some factors of 2 which need to be bookkept to get the trace formula for inner product to be the same. More generally, for \( V \) an inner product space, then for a convex function \( f \), and a point \( x \in V \), an element \( v \in V \) is a subgradient of \( f \) at \( x \) if

\[
f(z) \geq f(x) + \langle v, z - x \rangle
\]

for all \( z \in V \).