**Theorem:** Let \( x^* \) be a local minimizer of a convex function \( f \) on a convex set \( S \) (e.g., \( f \) and \( S \) come from a CP). Then \( x^* \) is a global minimizer of \( f \) on \( S \). Moreover, \( x^* \) is a strict global minimizer of \( f \) on \( S \) if either (a) \( f \) is strictly convex on \( S \), or (b) \( x^* \) is a strict local minimizer of \( f \) on \( S \).

**Proof:** Since \( x^* \) is a local minimizer there is \( \varepsilon > 0 \) so that
\[
 f(x) \geq f(x^*) \quad \text{whenever} \quad x \in S \quad \text{and} \quad \|x - x^*\| < \varepsilon.
\]
Now consider any \( y \in S \) with \( y \neq x^* \); we must show that \( f(y) \geq (> f(x^*) \). Since \( S \) is convex, the line segment between \( y \) and \( x^* \) lies in \( S \). We can choose a point \( x \) on this line segment not equal to \( x^* \) (or \( y \)) with \( \|x - x^*\| < \varepsilon \). Then
\[
 x = \alpha x^* + (1 - \alpha)y \quad \text{for some} \quad \alpha, \; 0 \leq (\leq) \alpha < 1.
\]
In general \( \alpha \) will be close to \( 1 \). Hence, by (strict) convexity, and since \( \|x - x^*\| < \varepsilon \),
\[
 \alpha f(x^*) + (1 - \alpha)f(y) \geq ((a) >) f(x) \geq ((b) >) f(x^*)
\]
and so
\[
 (1 - \alpha)f(y) \geq (>) f(x^*) - \alpha f(x^*) = (1 - \alpha)f(x^*)
\]
and since \( 1 - \alpha > 0 \), \( f(y) \geq (>) f(x^*) \) as required. (Modifications for strict global minimizer are shown.) \( \blacksquare \)

Results for maximizers: change convex to concave.

**Note:** a convex program may not have any minimizers, e.g.
\[
 \min e^x + e^y, \quad (x, y) \in \mathbb{R}^2.
\]
Convex (will show later) but no minimizers. So theorem does not say anything about existence of minimizers.

Previous optimality conditions involved derivatives. Now we can
(a) get conditions involving convexity and derivatives;
(b) relate convexity itself to (second) derivatives.

(a) Conditions involving convexity and derivs: **Theorem:** Suppose \( f \) is convex and has continuous partial derivatives on an open convex set \( S \). Then \( x^* \) is a global minimizer of \( f \) on \( S \) \( \iff \nabla f(x^*) = 0 \). So necessary condition \( \nabla f = 0 \) is also sufficient here.

Seems bit specialised but actually has wider implications.

**Corollary:** Suppose \( f \) is convex on a convex set \( S \), and has continuous partial derivatives in \( B(x^*, \varepsilon) \subseteq S \) for some \( \varepsilon > 0 \). Then \( x^* \) is a global minimizer of \( f \) on \( S \) \( \iff \nabla f(x^*) = 0 \).

(b) Relate convexity to 2nd deriv: **Function of one variable** \( f(x) \) continuous on nontrivial (not a single point) interval \( I \), \( f'' \) exists at every interior point of \( I \). Then \( f'' \geq 0 \) on interior of \( I \) \( \iff \ f \) convex (concave up) on \( I \).
\[
 \Rightarrow: \text{use MVT at two levels (assign problem sometimes)};
\]
\[
 \Leftarrow: \text{Limits and inequalities.}
\]

**Example:** \( f(x) = x^3 \), \( f'(x) = 3x^2 \), \( f''(x) = 6x \).

Not convex on \( (-\infty, \infty) \).

Convex on \( [0, \infty) \) since \( f'' = 6x \geq 0 \) on \( (0, \infty) \).

Also: \( f'' > 0 \) on interior of \( I \) \( \Rightarrow \ f \) strictly convex on \( I \).
E.g. \( f(x) = x^3 \), \( f''(x) = 6x > 0 \) on \( (0, \infty) \), so \( f \) actually strictly convex on \( [0, \infty) \).
But \( \neq \), e.g., \( f(x) = x^4 \).

Now consider \( f \) on \( \mathbb{R}^n \). Interval in \( \mathbb{R} \leftrightarrow \text{convex set in} \ \mathbb{R}^n \), \( f'' \geq 0 \leftrightarrow \nabla^2 f \) is positive semidefinite.
**Theorem:** Suppose $S \subseteq \mathbb{R}^n$ is a convex set with nonempty interior, and $f : S \to \mathbb{R}$ is continuous on $S$ and has continuous second partial derivatives on the interior of $S$. Then

(i) $\nabla^2 f$ is positive semidefinite on the interior of $S$ $\iff$ $f$ is convex on $S$.

(ii) $\nabla^2 f$ is positive definite on the interior of $S$ $\Rightarrow$ $f$ is strictly convex on the interior of $S$. 

But $\nabla^2 f$ (as before).

Need to restrict to interior because of possibility of chords lying entirely in boundary.

Making $S$ strictly convex avoids this, then can give result for all of $S$.

**Example:** $f(x, y) = e^x + e^y$, $(x, y) \in \mathbb{R}^2$.

$f_x = e^x$, $f_y = e^y$,

$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} e^x & 0 \\ 0 & e^y \end{bmatrix}$, diagonal, eigenvalues $e^x, e^y > 0$

so positive definite

so $f$ strictly convex on interior of $\mathbb{R}^2$, i.e. on all $\mathbb{R}^2$.

Convexity is a fundamental concept for theoretical work on optimization – Berkovitz and Borwein & Lewis both have ‘convex’ in title. We will take a more practical, algorithmic approach from now on and won’t explicitly mention convexity much.

**Convergence analysis**

First some notation we’ve seen informally.

**$O$-notation:** Given set $S$ and two functions $f, g : S \to \mathbb{R}$ (or $\mathbb{C}$), write $f(x) = O(g(x))$ on $S$ if there is a constant $c$ such that $|f(x)| \leq c|g(x)| \forall x \in S$.

Can also apply to vector-valued functions; then take magnitude $\| \cdot \|$ instead of absolute value.

Often use in situations with a limit, e.g.,

$f(x) = O(g(x))$ as $x \to a$ means for $x$ sufficiently close to $a$;

$f_k = O(g_k)$ as $k \to \infty$ means for integers $k$ sufficiently large.

**Example:** Taylor series for $f$ with continuous third derivative at $x_0$:

$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2f''(x_0) + \mathcal{O}((x - x_0)^3)$ as $x \to x_0$.

Also have $n$-dimensional version with gradient and Hessian matrix.

Now, want to look at iterative methods of finding minimizer. Want to measure how fast they converge to a solution.

Iterative method: $x_0, x_1, x_2, \ldots$ converging to $x^*$: $\lim_{k \to \infty} x_k = x^*$. (Limit in $\mathbb{R}^n$.)

Errors: $e_k = x_k - x^*$: $\lim_{k \to \infty} e_k = 0$.

**Q-rate convergence** (Q comes from ‘quotient’): say that $x_k \to x^*$ with $Q$-rate $r$ (or just rate $r$)

if there is a constant $c$ such that

$\|x_k + 1 - x^*\| \leq c\|x_k - x^*\|^r$ for all sufficiently large $k$,

i.e. $\|e_{k+1}\| \leq c\|e_k\|^r$ as $k \to \infty$.

Then for large $k$,

$-\log_{10} \|x_{k+1} - x^*\| \geq r(-\log_{10} \|x_k - x^*\|) - \log_{10} C$.

$-\log_{10}$ measures number of decimal places of accuracy, increases by factor of $r$ each time.

(Q-)$\text{linear convergence}$: rate $r = 1$, then need $c < 1$;

e.g. $1, 1 \cdot \frac{1}{2}, 1 \cdot \frac{1}{4}, 1 \cdot \frac{1}{16}, \ldots \to 0$ with rate 1, constant $c = \frac{1}{2}$.

(Q-)$\text{quadratic convergence}$: rate $r = 2$,

e.g. $10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, \ldots \to 0$ with rate 2, $c = 1$. 
(Q-)superlinear convergence: \[ \|x_{k+1} - x^*\| \leq c_k \|x_k - x^*\| \] where \( c_k \to 0 \) as \( k \to \infty \).

Faster than linear because not just fixed \( c \), have \( c_k \to 0 \). But cannot say has rate \( r > 1 \).

E.g. \( \frac{1}{k!} \to 0 \) superlinearly (\( c_k = \frac{1}{k+1} \to 0 \)), but does not have rate \( > 1 \).

Sometimes easiest to prove Q-rate via a lemma (sufficient condition, but not necessary).

**Q-rate Limit Lemma:** Suppose that \( r \geq 1 \) and \( L = \lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} \) exists (is finite), with \( L < 1 \) if \( r = 1 \) (important!). Then \( x_k \to x^* \) with Q-rate \( r \) (and constant \( c > L \)).

**Proof:** For \( \varepsilon > 0 \) and large enough \( k \), we have

\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^r} \leq L + \varepsilon \quad \text{so that} \quad \|x_{k+1} - x^*\| \leq (L + \varepsilon)\|x_k - x^*\|^r
\]

– the definition of Q-rate \( r \) with \( c = L + \varepsilon \) (when \( r = 1 \) choose \( \varepsilon \) so that \( c = L + \varepsilon < 1 \)).

**Note:** Can have Q-rate convergence even when no limit,

e.g. \( 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{12}, \frac{1}{36}, \frac{1}{72}, \frac{1}{216}, \ldots \to 0 \) with Q-rate 1 with \( c = \frac{1}{2} \)

but the limit in the lemma does not exist.