HW1 is due today!

HW2 coming out later today.

Project teams

(acknowledgments: R. Tibshirani (CMU) for parts of today’s lecture)
Tractable nonconvex problems

Not all non-convex problems are bad
Tractable nonconvex problems

Not all non-convex problems are bad

♠ Generalizing the notion of convexity
♠ Problems with hidden convexity
♠ Miscellaneous examples
♠ The list is much longer and growing!
Spectral problems
Simplest example: eigenvalues

Largest eigenvalue of a symmetric matrix

\[ Ax = \lambda_{\text{max}} x \iff \max_{x^T x = 1} x^T Ax. \]

Nonconvex problem, but we know how to solve it!
Simplest example: eigenvalues

Largest eigenvalue of a symmetric matrix

\[ Ax = \lambda_{\text{max}} x \quad \Leftrightarrow \quad \max_{x^T x = 1} x^T Ax. \]

Nonconvex problem, but we know how to solve it!

\[ \mathcal{L}(x, \theta) := -x^T Ax + \theta (x^T x - 1) \]

\[ -2Ax + 2\theta x = 0 \]

\[ Ax = \theta x \]

Thus, necessary condition for optimum asks for \((\theta, x)\) to be eigenpair, whereby clearly, \(x^T Ax\) is maximized by largest such pair.
Simplest example: eigenvalues

Largest eigenvalue of a symmetric matrix

\[ Ax = \lambda_{\text{max}} x \iff \max_{x^T x = 1} x^T Ax. \]

Nonconvex problem, but we know how to solve it!

\[
\mathcal{L}(x, \theta) := -x^T Ax + \theta(x^T x - 1)
\]

\[-2Ax + 2\theta x = 0 \]

\[ Ax = \theta x \]

Thus, necessary condition for optimum asks for \((\theta, x)\) to be eigenpair, whereby clearly, \(x^T Ax\) is maximized by largest such pair. Let \(A = UDU^*\), then \(\max_{x^T x = 1} x^T Ax = \max_{y^T y = 1} \sum \lambda_i y_i^2\), where \(y = U^* x\).
Simplest example: eigenvalues

Largest eigenvalue of a symmetric matrix

\[ Ax = \lambda_{\text{max}} x \iff \max_{x^T x = 1} x^T Ax. \]

Nonconvex problem, but we know how to solve it!

\[ \mathcal{L}(x, \theta) := -x^T Ax + \theta(x^T x - 1) \]
\[ -2Ax + 2\theta x = 0 \]
\[ Ax = \theta x \]

Thus, neccessary condition for optimum asks for \((\theta, x)\) to be eigenpair, whereby clearly, \(x^T Ax\) is maximized by largest such pair. Let \(A = UDU^*\), then \(\max_{x^T x = 1} x^T Ax = \max_{y^T y = 1} \sum \lambda_i y_i^2\), where \(y = U^* x\).

\[ \max_{y^T y = 1} \sum \lambda_i y_i^2 = \max_{z_1 = 1, z \geq 0} \sum \lambda_i z_i, \]

which is a convex optimization problem.
Generalized eigenvalues

\[
\min_{x \neq 0} \frac{x^T Ax}{x^T Bx}
\]

(more generally: \(Ax = \lambda Bx\), generalized eigenvectors)

**Exercise:** Study it’s Lagrangian formulation as well as a convex reformulation.
Trust region subproblem

\[
\begin{align*}
\min_{x} & \quad x^T Ax + 2b^T x + c \\
\text{s.t.} & \quad x^T Bx + 2d^T x + e \leq 0.
\end{align*}
\]

Here \( A \) and \( B \) are merely symmetric. Hence, nonconvex
Trust region subproblem

\[
\min_x \quad x^T A x + 2b^T x + c \\
\text{s.t.} \quad x^T B x + 2d^T x + e \leq 0. \\
\]

Here \( A \) and \( B \) are merely symmetric. Hence, nonconvex

The dual problem can be formulated as (Verify!)

\[
\max_{u, v \in \mathbb{R}} \quad u \\
\text{s.t.} \quad \begin{bmatrix} A + vB & b + vd \\ (b + vd)^T & c + ve - u \end{bmatrix} \succeq 0, \\
v \quad \geq 0. \\
\]

Importantly, strong duality holds (see Appendix B of BV). (alternatively: turns out SDP relaxation of the primal is exact)
Trust region subproblem

\[
\begin{align*}
\min_x & \quad x^T Ax + 2b^T x + c \\
\text{s.t.} & \quad x^T Bx + 2d^T x + e \leq 0.
\end{align*}
\]

Here \( A \) and \( B \) are merely symmetric. Hence, nonconvex

The dual problem can be formulated as (Verify!)

\[
\begin{align*}
\max_{u,v \in \mathbb{R}} & \quad u \\
\text{s.t.} & \quad \begin{bmatrix} A + vB & b + vd \\ (b + vd)^T & c + ve - u \end{bmatrix} \succeq 0, \\
v & \geq 0.
\end{align*}
\]

Importantly, strong duality holds (see Appendix B of BV). (alternatively: turns out SDP relaxation of the primal is exact)

Let $A$ be a complex, square matrix. Its *numerical range* is

$$W(A) := \{ x^*Ax \mid \|x\|_2 = 1, x \in \mathbb{C}^n \}.$$
Toeplitz-Hausdorff Theorem

Let $A$ be a complex, square matrix. Its *numerical range* is

$$W(A) := \{x^*Ax \mid \|x\|_2 = 1, x \in \mathbb{C}^n\}.$$ 

**Theorem.** The set $W(A)$ is convex.
Let $A$ be a complex, square matrix. Its **numerical range** is

$$W(A) := \{x^*Ax \mid \|x\|_2 = 1, x \in \mathbb{C}^n\}.$$ 

**Theorem.** The set $W(A)$ is convex.

Say $A$ is Hermitian, then clearly $W(A) = [\lambda_{\text{min}}, \lambda_{\text{max}}]$, which is convex. If $A$ is normal (i.e., $AA^* = A^*A$) then $W(A) = \text{conv}(\lambda_i(A))$. But more generally?
Let $A \in \mathbb{R}^{n \times p}$. Consider the nonconvex problem

$$\min_X \|A - X\|_F^2 \quad \text{s.t.} \quad \text{rank}(X) = k.$$
Let $A \in \mathbb{R}^{n \times p}$. Consider the nonconvex problem
\[
\min_X \| A - X \|_F^2 \quad \text{s.t.} \quad \text{rank}(X) = k.
\]

Well-known Eckart-Young-Mirsky theorem shows that
\[
X^* = U_k \Sigma_k V_k^T
\]
based on the SVD $A = U \Sigma V^T$. 

Another characterization of SVD (nonconvex prob)

\[
\min_{Z = Z^T} \| A - AZ \|_F^2, \quad \text{s.t. } \text{rank}(Z) = k, \text{Z is a projection}
\]

\[
\Leftrightarrow \max_{Z = Z^T} \langle A^T A, Z \rangle, \quad \text{s.t. } \text{rank}(Z) = k, \text{Z is a projection}.
\]
Another characterization of SVD (nonconvex prob)

\[
\begin{align*}
\min_{Z=Z^T} \|A - AZ\|_F^2, \quad & \text{s.t. } \text{rank}(Z) = k, Z \text{ is a projection} \\
\Leftrightarrow \max_{Z=Z^T} \langle A^T A, Z \rangle, \quad & \text{s.t. } \text{rank}(Z) = k, Z \text{ is a projection}.
\end{align*}
\]

Optimal solution here is \( Z = V_k V_k^T \), the top-\( k \) evecs of \( A^T A \).
Another characterization of SVD (nonconvex prob)

\[
\min_{Z=Z^T} \|A - AZ\|_F^2, \quad \text{s.t.} \quad \text{rank}(Z) = k, Z \text{ is a projection}
\]

\[
\Leftrightarrow \max_{Z=Z^T} \langle A^T A, Z \rangle, \quad \text{s.t.} \quad \text{rank}(Z) = k, Z \text{ is a projection}.
\]

Optimal solution here is \( Z = V_k V_k^T \), the top-\( k \) evecs of \( A^T A \)

**Equivalent convex problem!**

First, write constraint set \( C \) as

\[
C = \\left\{ Z = Z^T \mid \text{rank}(Z) = k, Z \text{ is a projection} \right\}
\]
Another characterization of SVD (nonconvex prob)

\[
\min_{Z = Z^T} \|A - AZ\|_F^2, \quad \text{s.t. } \text{rank}(Z) = k, \text{Z is a projection}
\]

\[
\Leftrightarrow \max_{Z = Z^T} \langle A^T A, Z \rangle, \quad \text{s.t. } \text{rank}(Z) = k, \text{Z is a projection}.
\]

Optimal solution here is \(Z = V_k V_k^T\), the top-\(k\) evcs of \(A^T A\)

**Equivalent convex problem!**

First, write constraint set \(C\) as

\[
C = \left\{ Z = Z^T \mid \text{rank}(Z) = k, \text{Z is a projection} \right\}
\]

\[
= \left\{ Z = Z^T \mid \lambda_i(Z) \in \{0, 1\}, \text{Tr}(Z) = k \right\}.
\]
Now consider convex hull: $C = \text{conv } C$
Now consider convex hull: $C = \text{conv } C$

\[
C = \left\{ Z = Z^T \mid \lambda_i(Z) \in [0, 1], \text{Tr}(Z) = k \right\}
\]
Now consider convex hull: $C = \text{conv} \ C$

\[
C = \left\{ Z = Z^T \mid \lambda_i(Z) \in [0, 1], \text{Tr}(Z) = k \right\}
\]

\[
= \left\{ Z = Z^T \mid 0 \preceq Z \preceq I, \text{Tr}(Z) = k \right\}
\]

The set $C$ is called the **Fantope** (named after Ky Fan).
Now consider convex hull: $C = \text{conv } C$

$$
C = \left\{ Z = Z^T \mid \lambda_i(Z) \in [0, 1], \text{Tr}(Z) = k \right\}
$$

$$
= \left\{ Z = Z^T \mid 0 \preceq Z \preceq I, \text{Tr}(Z) = k \right\}.
$$

The set $C$ is called the **Fantope** (named after Ky Fan).

**Exercise:** Now invoke the “maximize a convex function” idea from yesterday to claim that the convex problem

$$
\max_{Z = Z^T} \langle A^T A, Z \rangle \text{ s.t. } Z \in C
$$

solves the original problem.
Sparsity
The $\ell_0$-quasi-norm is defined as

$$\|x\|_0 := \text{card} \{x_i \mid x_i \neq 0\}.$$
Nonconvex Sparse optimization

The $\ell_0$-quasi-norm is defined as

$$\|x\|_0 := \text{card } \{ x_i \mid x_i \neq 0 \}.$$ 

**Projection onto $\ell_0$-ball**

$$\min \, \frac{1}{2} \|x - y\|_2^2, \quad \text{s.t.} \quad \|x\|_0 \leq k.$$
The $\ell_0$-quasi-norm is defined as

$$\|x\|_0 := \text{card } \{x_i \mid x_i \neq 0\}.$$ 

**Projection onto $\ell_0$-ball**

$$\min \frac{1}{2}\|x - y\|_2^2, \quad \text{s.t.} \quad \|x\|_0 \leq k.$$ 

Nonconvex but tractable: If $\|y\|_0 \leq k$, then clearly $x = y$. Otherwise, pick the $k$ largest entries of $|y|$, and set the rest to 0.

**Exercise:** Prove the above claim.

**Exercise:** Similarly solve

$$\frac{1}{2}\|x - y\|_2^2 + \lambda \|x\|_0.$$ 

Used in so-called “Iterative Hard Thresholding” algorithms.
NonconvexSparse optimization

The $\ell_0$-quasi-norm is defined as

\[ \|x\|_0 := \text{card} \{ x_i \mid x_i \neq 0 \} . \]

Projection onto $\ell_0$-ball

\[ \min \quad \frac{1}{2} \|x - y\|_2^2, \quad \text{s.t.} \quad \|x\|_0 \leq k. \]

Nonconvex but tractable: If $\|y\|_0 \leq k$, then clearly $x = y$. Otherwise, pick the $k$ largest entries of $|y|$, and set the rest to 0.

**Exercise:** Prove the above claim.

**Exercise:** Similarly solve $\frac{1}{2} \|x - y\|_2^2 + \lambda \|x\|_0$

Used in so-called “Iterative Hard Thresholding” algorithms
Compressed Sensing

\[
\min \quad ||x||_0 \quad \text{s.t.} \quad Ax = b
\]
Compressed Sensing

\[ \begin{align*}
\min & \quad \|x\|_0 \\
\text{s.t.} & \quad Ax = b
\end{align*} \]

If the “measurement matrix” \( A \) satisfies so-called restricted isometry condition with the constant \( \delta_s \in (0, 1) \)

\[(1 - \delta_s)\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_s)\|x\|^2, \quad \text{if } x \text{ is } s\text{-sparse}, \]

then the \( \ell_1 \)-convex relaxation is exact.

**Search keywords:** compressed sensing, sparse recovery, restricted isometry
Generalized convexity
Geometric programming

**Monomial:** $g : \mathbb{R}^n_{++} \rightarrow \mathbb{R}$ of the form

$$g(x) = \gamma x_1^{a_1} \cdots x_n^{a_n}, \quad \gamma > 0, a_i \in \mathbb{R}.$$

**Posynomial:** Sum of monomials, e.g., $f(x) = \sum_j g_j(x)$
**Geometric programming**

**Monomial:** $g : \mathbb{R}_+^n \to \mathbb{R}$ of the form

$$g(x) = \gamma x_1^{a_1} \cdots x_n^{a_n}, \quad \gamma > 0, a_i \in \mathbb{R}.$$

**Posynomial:** Sum of monomials, e.g.,

$$f(x) = \sum_j g_j(x)$$

**Geometric Program**

$$\begin{align*}
\min_x & \quad f(x) \\
\text{s.t.} & \quad f_i(x) \leq 1, \quad i \in [m] \\
& \quad g_j(x) = 1, \quad j \in [r],
\end{align*}$$

where $f_i$ are posynomials and $g_j$ are monomials.

Clearly, nonconvex.
Geometric programming

Make change of variables: $y_i = \log x_i$ (recall $x_i > 0$). Then,

$$f(x) = f(e^y) = \gamma (e^{y_1})^{a_1} \cdots (e^{y_n})^{a_n} = e^{a^T y + b},$$

for $b = \log y$. Thus, after taking logs, geometric program is

$$\min_y \log \left( \sum_k e^{a_{0k}^T y + b_{0k}} \right)$$
$$\text{s.t. } \log \left( \sum_k e^{a_{0k}^T y + b_{0k}} \right) \leq 0, \ i \in [m]$$
$$c_j^T y + d_j = 0, \ j \in [r],$$

for suitable sets of vectors $\{a_{ik}\}$, and $\{c_j\}$.
Geometric programming

Make change of variables: \( y_i = \log x_i \) (recall \( x_i > 0 \)). Then,

\[
f(x) = f(e^y) = \gamma (e^{y_1})^{a_1} \cdots (e^{y_n})^{a_n} = e^{a^T y + b},
\]

for \( b = \log y \). Thus, after taking logs, geometric program is

\[
\begin{align*}
\min_{y} & \quad \log \left( \sum_k e^{a_{0k}^T y + b_{0k}} \right) \\
\text{s.t.} & \quad \log \left( \sum_k e^{a_{0k}^T y + b_{0k}} \right) \leq 0, \ i \in [m] \\
\end{align*}
\]

\[
c_j^T y + d_j = 0, \ j \in [r],
\]

for suitable sets of vectors \( \{a_{ik}\} \), and \( \{c_j\} \).
Recall, log-sum-exp is convex, so above is a convex opt.

Ref: See Chapter 8.8 of BV; search online for “geometric programming”
Generalized convexity

- Quasiconvexity: If level sets $L_t(f) = \{x \mid f(x) \leq t\}$ are convex, we say $f$ is quasiconvex.

Arcwise Convexity:

Let $\gamma(t) = x + t(y - x)$ be an arc joining point $x$ to point $y$. Then

$$f(\gamma(t)) \leq (1 - t)f(x) + tf(y),$$

Exercise: Suppose a set $X$ is arcwise convex, and $f: X \to \mathbb{R}$ is an arcwise convex function. Prove that a local optimum of $f$ is also global (assume regularity as needed).
Generalized convexity

- Quasiconvexity: If level sets $L_t(f) = \{x \mid f(x) \leq t\}$ are convex, we say $f$ is **quasiconvex**

- Arcwise Convexity: $f(\gamma_{xy}(t)) \leq (1 - t)f(x) + tf(y)$, where \textit{arc} $\gamma : [0, 1] \to X$ joins point $x$ to point $y$. 

Several other notions of generalized convexity exist (see also: genconv.org!)

Exercise: Suppose a set $X$ is arcwise convex, and $f : X \to \mathbb{R}$ is an arcwise convex function. Prove that a local optimum of $f$ is also global (assume regularity as needed).

Exercise: View GP as arcwise convexity using: $\gamma(t) = x(1-t) + ty$.
Generalized convexity

- **Quasiconvexity**: If level sets $L_t(f) = \{x | f(x) \leq t\}$ are convex, we say $f$ is *quasiconvex*.

- **Arcwise Convexity**: $f(\gamma_{xy}(t)) \leq (1 - t)f(x) + tf(y)$, where \(\text{arc } \gamma : [0, 1] \to X\) joins point $x$ to point $y$.

- Several other notions of generalized convexity exist (see also: genconv.org!)

**Exercise**: Suppose a set $X$ is arcwise convex, and $f : X \to \mathbb{R}$ is an arcwise convex function. Prove that a local optimum of $f$ is also global (assume regularity as needed).

**Exercise**: View GP as arcwise convexity using: $\gamma(t) = x - ty$.
Generalized convexity

- Quasiconvexity: If level sets \( L_t(f) = \{x \mid f(x) \leq t\} \) are convex, we say \( f \) is quasiconvex.

- Arcwise Convexity: \( f(\gamma_{xy}(t)) \leq (1 - t)f(x) + tf(y) \), where \( arc \, \gamma : [0, 1] \rightarrow X \) joins point \( x \) to point \( y \).

- Several other notions of generalized convexity exist (see also: genconv.org!)

**Exercise:** Suppose a set \( X \) is arcwise convex, and \( f : X \rightarrow \mathbb{R} \) is an arcwise convex function. Prove that a local optimum of \( f \) is also global (assume regularity as needed).

**Exercise:** View GP as arcwise convexity using: \( \gamma(t) = x^{1-t}y^t \)
Linear fractional programming

\[
\begin{align*}
\text{min} \quad & \frac{a^T x + b}{c^T x + d} \\
\text{s.t.} \quad & Gx \leq h, \quad c^T x + d > 0, \quad Ex = f.
\end{align*}
\]

This problem is nonconvex, but it is quasiconvex.
Linear fractional programming

\[
\begin{align*}
\min & \quad \frac{a^T x + b}{c^T x + d} \\
\text{s.t.} & \quad Gx \leq h, c^T x + d > 0, Ex = f.
\end{align*}
\]

This problem is nonconvex, but it is quasiconvex. Provided it is feasible, it is equivalent to the LP

\[
\begin{align*}
\min_{y,z} & \quad a^T y + bz \\
\text{s.t.} & \quad G_y - h z \leq 0, z \geq 0 \\
& \quad E_y = b z, c^T y + dz = 1.
\end{align*}
\]
Linear fractional programming

\[
\begin{align*}
\min \quad & a^T x + b \\
\text{s.t.} \quad & Gx \leq h, \ c^T x + d > 0, \ Ex = f.
\end{align*}
\]

This problem is nonconvex, but it is quasiconvex. Provided it is feasible, it is equivalent to the LP

\[
\begin{align*}
\min_{y,z} \quad & a^T y + bz \\
\text{s.t.} \quad & Gy - hz \leq 0, \ z \geq 0 \\
& Ey = bz, \ c^T y + dz = 1.
\end{align*}
\]

These two problems connected via the transformation

\[
y = \frac{x}{c^T x + d}, \quad z = \frac{1}{c^T x + d}.
\]

See BV Chapter 4 for details.
Generalized Perron-Frobenius

Let $A, B \in \mathbb{R}^{m \times n}$.

$$
\begin{align*}
\max_{x, \lambda} & \quad \lambda \\
\text{s.t.} & \quad \lambda Ax \leq Bx, x^T1 = 1, x \geq 0.
\end{align*}
$$
Generalized Perron-Frobenius

Let $A, B \in \mathbb{R}^{m \times n}$.

$$\begin{align*}
\max_{x, \lambda} & \quad \lambda \\
\text{s.t.} & \quad \lambda Ax \leq Bx, x^T1 = 1, x \geq 0.
\end{align*}$$

**Exercise:** Try solving it directly somehow.
Generalized Perron-Frobenius

Let $A, B \in \mathbb{R}^{m \times n}$.

$$\begin{align*}
\max_{x, \lambda} & \quad \lambda \\
\text{s.t.} & \quad \lambda Ax \leq Bx, \ x^T1 = 1, \ x \geq 0.
\end{align*}$$

**Exercise:** Try solving it directly somehow.

**Exercise:** Cast this as an (extended) linear-fractional program.
Challenge: Simplex convexity

Let $\Delta_n$ be the probability simplex, i.e., set of vectors $x = (x_1, \ldots, x_n)$ such that $x_i \geq 0$ and $x^T 1 = 1$. Assume that $n \geq 2$. Prove that the following “entropy”

$$g(x) = \sum_i x_i \log \frac{1}{x_i} + (1 - x_i) \log(1 - x_i),$$

is concave on $\Delta_n$. 
Submodular optimization (later in course)
More generally, any combinatorial problem whose convex relaxation is tight
More hidden convex problems
Problems free of local minima: matrix completion, deep linear neural networks, tensor factorization, etc. See PhD thesis: “When are nonconvex optimization problems not scary?” by Ju Sun, Columbia University, 2016.