Recap from last time

- Convex sets (distinguished convex, etc.)
- Merger convexity in metric spaces.
- 1-Open problem.

Let \((X,d)\) be a metric space.

- Complete if all Cauchy sequences in \(X\) converge to a point in \(X\) (so \(\mathbb{R}\) is complete but \(\mathbb{Q}\) is not complete)
- Locally compact: If every point of \(X\) has a compact nbhd (i.e., we can put an \(\varepsilon\)-ball around each point in \(X\))
- \(\text{D}\) not l.c. (around no point can we put a ball)

\[\{(0,0)\} \cup \{(x,0) \mid x > 0\} \subset \mathbb{R}^2\]

\(\{(0,0)\}\) does not have a compact nbhd.

Fundamental theorem of merger convexity:

**Theorem:** Suppose \((X,d)\) is complete and locally compact. Then, \((X,d)\) is merger convex ⇔

\[\forall x,y \in X, \exists m \in X \text{ s.t.} \quad d(x,m) = d(ym) = \frac{1}{2} d(x,y) \quad \text{(midpoint case)}\]

⇔ Any two \(x,y \in X\) are joined by a geodesic

**Definition:** A geodesic is a continuous path of shortest length between two points in \((X,d)\).

Let \(x,y \in X, \quad t \in [0,1]\)

\[\gamma(t) \quad \text{is such that} \quad \gamma(0) = x, \quad \gamma(1) = y \quad \text{and} \]

\[d(\gamma(t_1), \gamma(t_2)) = |t_1 - t_2| d(x,y), \quad \forall t_1, t_2 \in [0,1]\]
Example: \((\mathbb{R}^n, \| \cdot \|_2)\) is a geodesic metric space; here \(\| \cdot \|_2\) does not just ordinary convexity.

Question: Is \((\mathbb{R}^n, \| \cdot \|_2)\) complete?

Yes: each point in \(\mathbb{R}^n\) can be written as \(c u_i, u_i \in \mathbb{R}\)

and thus \(\| c u_i - u_j \| = \| c u_i - u_i \|, \text{but} \| R, 1 - 1 \|\) is complete.

---

**CONVEX FUNCTIONS**

- **Midpoint convex**

  Let \(x, y \in \mathbb{R}\), then \(f\) is called midpoint convex if

  \[
  f \left( \frac{x + y}{2} \right) \leq \frac{f(x) + f(y)}{2} \quad \forall x, y \in X
  \]

- **Jensen-convex, hence also just convex**

  \[
  f \left( (1 - x)u + x v \right) \leq (1 - x)f(u) + x f(v) \quad \forall x \in [0, 1].
  \]

Clearly, dom \(f\) better be convex.

Already mention the case \(m(\mathbb{R}, d)\), before going onto the next theorem.

**Theorem: (Jensen, 1905)**. Let \(f: \mathbb{R}^k \to \mathbb{R}\) be a continuous function. Then \(f\) is convex if it is midpoint

**Proof:** (In this case, general result); clearly, any sufficient data needs only one. Argue by contradiction.

Suppose

\[
\frac{f \left( \frac{x + y}{2} \right) - f(x)}{\left( \frac{x + y}{2} \right) - x} \text{ and } \frac{f \left( \frac{x + y}{2} \right) - f(y)}{\left( \frac{x + y}{2} \right) - y} \quad \forall x, y \in I,
\]

But then \(f \left( \frac{c(x + y)}{2} \right) \neq \text{ violates } c \).
then, consider the continuous function

\[ g(x) := f((1-x)x + dy) - (1-x)f(x) + f(y). \]

Since \( g \) is violated, \( \Rightarrow \) max value \( g' \) on \([0,1]\) will be \( \infty \). Hence new \( M > 0 \).

- Let \( \alpha_0 \) be the smallest \( \alpha \in (0,1) \) for which \( g(\alpha) = M \).
- Let \( \delta > 0 \) be small enough so that \( \alpha_0 \in [\alpha_0 - \delta, \alpha_0 + \delta] \subset [0,1] \).

- \( \bar{x} := (1 - \alpha_0 - \delta)x + (\alpha_0 + \delta)y \)
- \( \bar{y} := (1 - \alpha_0 + \delta)x + (\alpha_0 - \delta)y \)

- \( \frac{\bar{y}}{2} \)

\[ f\left( \frac{\bar{x} + \bar{y}}{2} \right) = f\left( \frac{(1-\alpha_0)x + \alpha_0 y}{2} \right) \leq f((1-\alpha_0)x) + f(\alpha_0 y) \]

- \( g(\alpha_0) \leq g(\alpha_0 - \delta) + g(\alpha_0 + \delta) \frac{2}{2} < M, \) a contradiction.

\[ \frac{g(\alpha_0)}{g(\alpha_0 - \delta)} = \frac{f((1-\alpha_0)x + \alpha_0 y) - ((1-\alpha_0)f(x) - \alpha_0 f(y))}{g(\alpha_0 - \delta)} \]

(Cfr: Kranosbel's theorem for more.)
Convex functions in $\mathbb{R}^d$ are a good topic of study.

Let $Y(x) = (1-d)x \circ \Delta y$ denote a geodesic between $x, y \in \mathbb{R}^d$.

If $f: \mathbb{R} \to \mathbb{R}$ is called geodesically convex

if $f((1-d)x \circ \Delta y) \leq (1-d)f(x) + df(y)$

We will return to such geodesically convex functions later when studying "geometric optimization".

Some ways to recognize convex functions:

- Continuously differentiable convex
- If $f'$ is diff it is convex iff
  \[ f(x) \geq f(y) + \langle \nabla f(y), x-y \rangle \quad \forall x \in \text{dom } f \]
- If $f'$ is twice diff, convex iff
  \[ \nabla^2 f(x) \succeq 0 \quad \forall x \in \text{dom } f \]

- $f: \mathbb{R} \to \mathbb{R}$ convex iff its derivative $f'$ is $\uparrow$.
- $f: \mathbb{R}^n \to \mathbb{R}$ convex iff "" is a monotone operator
  i.e. \[ \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq 0 \quad \forall x, y \]

- We'll see other methods later on.

- Restating the line:
  \[ g(t) = f(x+ty) \quad \text{for almost all } x, t \in \mathbb{R} \quad (\text{subject to suitable}) ]
# We'll allow $f$ to take on $+\infty$ as a value.
# This allows us to just write $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

**Important conventions**

- $\infty + \infty = \infty + \infty$ for $-\infty < x < \infty$
- $\infty - \infty = -\infty + \infty$ for $-\infty \leq x < \infty$
- $x - \infty = 0$ for $0 < x \leq \infty$
- $0 - \infty = -\infty = 0$
- $-\infty = +\infty$
- $\inf \phi = +\infty$; $\sup \phi = -\infty$.
- Avoid $0 - \infty, -\infty + \infty$; akin to div. by zero.

# Proper cut for $f$.$f > -\infty$ everywhere and $< +\infty$ at at least one point.

($\exists \text{ dom } f = \{ x \in \mathbb{R}^n \mid f(x) < \infty \} \neq \emptyset$
and $f(\text{ dom } f$ is finite).

# With this, E.g.

$f(x) = \begin{cases} \frac{1}{x} & x > 0 \\ +\infty & x \leq 0 \end{cases}$

$f(x) = \begin{cases} -\log x & x > 0 \\ +\infty & x \leq 0 \end{cases}$

# Some important convex functions
Examples:

5. Indicator function: let \( C \subseteq \mathbb{R}^n \) be a nonempty set.

\[
S_C(x) := \begin{cases} 
0, & x \in C \\
+\infty, & x \notin C.
\end{cases}
\]

Verify: \( S_C \) is closed and concave iff \( C \) is closed and convex.

By closed we mean:

\[
\text{epi} f := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} | f(x) \leq t\}
\]

is a closed set.

6. Support function:

Let \( C \subseteq \mathbb{R}^n \) be a nonempty set.

\[
\sigma_C(x) := \sup_{z \in C} z^T x
\]

More generally, let \( f(x,y) \) be a family of functions indexed by \( y \in Y \) such that \( f(x,y) \) is concave in \( x \) for each \( y \). Then

\[
f(x) := \sup_{y \in Y} f(x,y) \text{ is cvx.}
\]

(Nota: for words of cvx., akin to "marginalizing out").

7. Consequence: Fenchel– conjugate:

Let \( f(x) : \mathbb{R}^n \to \mathbb{R} \) be any function. Then

\[
f^*(z) := \sup_{x \in \mathbb{R}^n} \langle z, x \rangle - f(x) \text{ is cvx.}
\]
For: cangiuli

Job Number: 23

Printing date: 12/17/2013 4:58:46 PM

Printer: GHC9206BW (interpreter version 3010.107)

\((\star)\) Let \( \mathcal{Y} \) be a nonempty convex set. Let 
\[ h(y, y') \] be jointly convex, i.e. \( \forall (x, y) \).

Then 
\[ f(x) = \inf_{y \in \mathcal{Y}} h(x, y) \] is a convex function \( \forall g \in \mathcal{Y} \) (provided \( f(x) > -\infty \)).

**Proof:**

\[ \inf_{y} h(x, y) \leq h(y_x, y + \lambda y_y, y) \leq h(0, y + \lambda y_y), \lambda (1-\lambda) h(y_x, y_y) + \lambda (1-\lambda) y_y \]

\[ \leq \]

Example: \( d(x, C) = \inf_{y \in \mathcal{Y}} |x - y| \) \( y \in C \subseteq \mathbb{R}^N \).
Norms on $\mathbb{R}^n$:

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a function that satisfies

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}^n$
- $f(x) = 0$ if $x = 0$ (the def.)
- $f(\lambda x) = |\lambda| f(x)$ (the homogeneity)
- $f(x+y) \leq f(x) + f(y)$ (subadditivity)

Then $f$ is called a norm.

We denote such an $f$ as $\| f \|_1$.

Thm: All norms are ordered.

Examples:
- $l_p$ norm: $\| x \|_p = \left[ \sum_{i=1}^{n} |x_i|^p \right]^{1/p}$, $1 \leq p < \infty$
- $l_\infty$ norm: $\| x \|_\infty = \max_{1 \leq i \leq n} |x_i|$.

Exercise: Verify that $\| x+y \|_p \leq \| x \|_p + \| y \|_p$.

Mixed-norm: $l_{p,q}$-norm: Let $x \in \mathbb{R}^{n_1+n_2+\ldots+n_m}$ be partitioned into subvectors $(x_{1}, \ldots, x_{n})$.

Let $1 \leq p, q \leq \infty$.

- $l_{p,q}$-norm: $\| x \|_{p,q} = \left( \sum_{i=1}^{m} \| x_i \|_p^q \right)^{1/q}$.

(Shawn up freq. in ML.)
Matrix Norms

- Frobenius norm: \( A \in \mathbb{C}^{m \times n} \)
  \[
  \|A\|_F := \sqrt{\text{tr}(A^*A)}
  \]
  Well-based because: \( \|A\|_F := \sqrt{\sum_{ij} |a_{ij}|^2} \).

- Induced norm: Let \( A \in \mathbb{R}^{m \times n} \); \( \|A\|_\infty \) any
  vector norm,
  \[
  \|A\|_{\infty} := \sup_{\|x\|_{\infty} = 1} \|Ax\|_{\infty}
  \]
  Eg: \( \|A\|_1 := \text{matrix 1-norm} \)
  \( \|A\|_{\infty} := \text{oo-norm} \)    \( \|A\|_1 \) : NP-hard to compute.

- Schatten \( p \)-norm: \( \|A\|_{(p)} = \|A\|_p \) (don't confuse with above)
  \[
  := \|
  \sigma(A)\|_p \quad \sigma(A) \text{ singular values}
  \]

- Ky-Fan norm (for vectors)
  \[
  \|x\|_{(t)} := \sum_{i=1}^t |x_i| \quad \text{(Sorted in descending order)}
  \]
  \[
  \|A\|_{(t)} := \sum_{i=1}^t \sigma_i(A) \quad \sigma_i(A) \text{ singular values}
  \]

- Operator 2-norm: \( A \in \mathbb{C}^{m \times n} \)
  \[
  \|A\|_2 := \sup_{\|x\|_2 = 1} \|Ax\|_2
  \]
  Claim: \( \|A\|_2 := \sigma_1(A) \) (Prove it!!)
Notes:

- Dual-norms.

Let $\| \cdot \|$ be any norm on $\mathbb{R}^n$. Its dual norm is

$$
\| u \|_* := \sup \{ u(x) \mid \| x \| \leq 1 \}.
$$

(Notice this is the support function of the unit norm ball.)

- Minkowski gauge functional.

(Norms are awesome because their epi is a cone)

- Let $C$ be any nonempty convex set.

  **Gauge:** $k(x) := \inf \{ \mu > 0 \mid x \in \mu C \}$

  **Verify:** $k$ is non-negative; $k(\alpha x) = \alpha k(x)$, $\alpha > 0$.

  $k(0) = 0$.

  If $k$ is additive, $k(x+y) = \inf \{ \mu > 0 \mid x \in \mu C, y \in \mu C \}$.

  Notice if $x \notin \mu C$ for any $\mu > 0$, by our convention $k(x) = +\infty$.

  So gauges can be infinite even though $x \in \mathbb{R}^n$ is not.
If gauge is finite, symmetric, and \( \gamma \),

\# In the paper "Convex geometry", charles gronberg, 1972

V. Chandok et al. introduced "alternative norms",

by writing them as gauges of suitably convex hulls.

\( \mathbb{S} \) is finite, symmetric, and \( \gamma \),

Consider two cases: \( \| f \|_1 \) and \( \| f \|_2 \leq 1 \)

To solve:

Prop. Let \( \| f \|_1 \) be any norm. Then \( \| f \|_1 \) super-

Note: \( \| f \|_1 \) is always finite.
Claim: Let \( f(x) = \|x\|_1 \); Then \( f^*(z) = s_{\|z\|_1} (z) \).

Proof: Consider two cases: (i) \( \|z\|_1 > 1 \) and (ii) \( \|z\|_1 \leq 1 \).

(i) \( \|z\|_1 > 1 \):
\[
\|z\|_1 = \min \left\{ u \mid z^T u = 1, u \geq 0 \right\}
\]
\[
f^*(z) = \sup \left\{ z^T x - f(x) \mid x \geq 0 \right\}
\]
Select \( x = \alpha u \) and let \( \alpha \to \infty \)
\[
\alpha z^T u = \alpha \|u\|_1 = \alpha (z^T u - \|u\|_1) \to \infty \quad (\text{since } z^T u > 1)
\]

(ii) \( \|z\|_1 \leq 1 \):
\[
\|z\|_1 = \min \left\{ u \mid z^T u = 1, u \geq 0 \right\}
\]
\[
m [\text{why conclude with } \alpha \|u\|_1 = \alpha (z^T u - 1) \leq 0] \Rightarrow z^T u - \|u\|_1 \leq \|z\|_1 (\|z\|_1 - 1) \leq 0
\]
So \( \alpha = 0 \) makes this here.
\[
f^*(z) = 0 \quad \text{in this case.}
\]
Thus:
\[
f^*(z) = \begin{cases} 0 & \text{if } \|z\|_1 \leq 1 \\ +\infty & \text{otherwise} \end{cases}
\]

---

Challenge Problem: Let \( x_1, x_2, \ldots, x_n > 0 \).
\( h_\alpha (x) = g_U := h_n (x + tv) \)

\[
\frac{1}{n} \left( \frac{1}{x_1 + tv} - \frac{1}{x_1 + tv} \right) \sum_{i=2}^{n} \frac{1}{x_i + tv} + \sum_{i=1}^{n-1} \frac{1}{x_i + x_{i+1} + tv} \quad \text{... eh. isn't?}
\]
Challenger 2

$$F_{\mu}(x) := \frac{1}{\mu H} \int_{-\infty}^{\infty} e^{tH} dt$$

Claim: \( \left[ F_{\mu}(x) \right] \frac{1}{\mu H} \)

\(-\mu H (-e^x) = F_{\mu}(x)\)

Claim \( \left( \frac{1}{\mu H} \right)^{\frac{x}{\mu H}} = \left[ F_{\mu}(x) \right] \frac{1}{\mu H} \quad \mu \geq 0 \quad x \in \mathbb{R}\)

\(F_0 = \log (1 + e^x) \in \mathbb{C}\)
Proof: we have equivalents that
\[ \left| \frac{x+y}{2} \right| \geq |x| |y|^{1/2} \] (AM-GM inequality)

Divide both sides by $|x|$, this reduces to
\[ \left| \frac{x-x^{-1}y}{2} \right| \geq |x^{-1}y|^{1/2} \]

Let $\lambda_i$ be the eigenvalue of $X^{-1}Y$, some $det = Prod$ yields
\[ det(z) = \prod_i \lambda_i(z) \]
\[ = \prod_i \left( 1 + \lambda_i \right) \geq \prod_i \sqrt{\lambda_i} \]

When is always true

Non-trivial fact used

\[ A, B > 0 \Rightarrow \lambda(AB) > 0 \]
\[ \lambda(AB) = \lambda(B^{1/2}AB^{1/2}) \]
\[ \lambda(AB) = \lambda(B^{1/2}AB^{-1}B^{1/2}) \]
\[ = \lambda(B^{1/2}AB^{1/2}) > 0 \]

If $f: \mathbb{R}^n \to \mathbb{R}$ is $C^1$

then $A: \mathbb{R}^m \to \mathbb{R}^n$ is an affine map

then $f \circ A$ is $C^1$: Ultimative

$\star$ $f(A(x)) = f(A(x)) = x \in \mathbb{R}^m$