Convex Optimization: Problem Set 1

February 17, 2026

Some problems borrowed from Optimization course of Daniel Dadush.

1. Recall the definition of convex hull:

$$\text{conv}(S) := \left\{ \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \mid \mathbf{x}_i \in S, \lambda \geq 0, \sum_{i=1}^{N} \lambda_i = 1 \right\}.$$

   1. Prove that for finite $S, \text{conv}(S) = \cap_{K \supseteq S} K$ where the intersection is over all closed convex K containing S. Therefore $\text{conv}(S)$ is the smallest convex set containing S.
   2. Prove Jensen's inequality: for convex $f : \mathbb{R}^n \to \mathbb{R}$ and input $\mathbf{x} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i$ where $\lambda_i \geq 0, \sum_{i=1}^{N} \lambda_i = 1$

$$f(\mathbf{x}) \leq \sum_{i=1}^{N} \lambda_i f(\mathbf{x}_i).$$

   3. Show that if $\mathcal{C} \subseteq \mathbb{R}^n$ is a compact convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function then the **supremum** of f over $\mathcal{C}$ is attained at an extreme point of $\mathcal{C}$.

Solution:

1. Clearly any convex $K \supseteq S$ contains any convex combination $\sum_{i=1}^{N} \lambda_i \mathbf{x}_i$ where $\mathbf{x}_i \in S$, so $\text{conv}(S) \subseteq \cap_{K \supseteq S} K$. Conversely, if $x \notin \text{conv}(S)$ then there exists a separating hyperplane $\langle w, x \rangle > \nu := \sup_{y \in \text{conv}(S)} \langle w, y \rangle$ and so the halfspace

$$H := \langle y \mid \langle y, w \rangle \leq \nu \}$$

   is a convex set satisfying $H \supseteq S, y \notin H$.

2. Recall that Jensen's inequality states that for convex function $f : \mathcal{C} \to \mathbb{R}$ on convex set $\mathcal{C}$, set of points $\mathbf{x}_1, ..., \mathbf{x}_k \in \mathcal{C}$, and coefficients $\lambda \in \mathbb{R}_+^k$ with $\sum_{i=1}^{k} \lambda_i = 1$, we have

$$f\big(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i\big) \leq \sum_{i=1}^{k} \lambda_i \mathbf{x}_i.$$

We will prove the statement by induction on the number of points k. Notice that the case $k = 2$ is exactly the definition of convexity for f. Now for $k \geq 2$, given $k+1$ points $\mathbf{x}_1, ..., \mathbf{x}_{k+1}$ and non-zero coefficients $\lambda \in \mathbb{R}^{k+1}_{++}$ with $\sum_{i=1}^{k+1} \lambda_i = 1$, let $\tilde{\lambda} := \sum_{j=1}^{k} \lambda_j = 1 - \lambda_{k+1}$ and

$$\mathbf{x} := \sum_{i=1}^{k} \frac{\lambda_i}{\tilde{\lambda}} \mathbf{x}_i.$$

Note this is a convex combination of $\mathbf{x}_1, ..., \mathbf{x}_k$ as $\gamma_i := \frac{\lambda_i}{\tilde{\lambda}}$ satisfies $\gamma \in \mathbb{R}^k_+$ and $\sum_{i=1}^{k} \gamma_i = 1$. Therefore we can bound the function as

$$\begin{aligned} f\big(\sum_{i=1}^{k+1} \lambda_i \mathbf{x}_i\big) &= f(\tilde{\lambda}\mathbf{x} + (1-\tilde{\lambda})\mathbf{x}_{k+1}) \leq \tilde{\lambda} f(\mathbf{x}) + (1-\tilde{\lambda}) f(\mathbf{x}_{k+1}) \\ &= \tilde{\lambda} f\big(\sum_{i=1}^{k} \frac{\lambda_i}{\tilde{\lambda}} \mathbf{x}_i\big) + \lambda_{k+1} f(\mathbf{x}_{k+1}) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i) + \lambda_{k+1} f(\mathbf{x}_{k+1}), \end{aligned}$$

where in the first step we used the definitions of $\mathbf{x}, \tilde{\lambda}$, in the second step we used the base case, and in the final step we used the induction hypothesis.

3. We focus on the case $\mathcal{C} = \text{conv}\{x_1, ..., x_N\}$ and show the maximum is achieved at some x_i. By continuity of f and compactness of $\mathcal{C}$, $x^* := \arg\max_{x \in \mathcal{C}} f(x)$ is attained at some point. Now let $x^* = \sum_{i=1}^{N} \lambda_i x_i$ where $\lambda \geq 0, \sum_{i=1}^{N} \lambda_i = 1$. Then by Jensen's inequality we have

$$f(x^*) \leq \sum_{i=1}^{N} \lambda_i f(x_i) \leq \max_{i \in [N]} f(x_i),$$

where the first step was by Jensen's inequality, in the second we used that λ is a convex combination. Since $x^* := \arg\max_{x \in \mathcal{C}} f(x)$ is a maximizer, the statement is shown.

We sketch some remarks for removing technical assumptions: it can be shown that convex f is always continuous at any $x \in \text{int}(\text{dom}(f))$ (see Theorem 10.1 and Cor 10.1.1 in Rockafellar). Since our function $f : \mathbb{R}^n \to \mathbb{R}$ never takes unbounded values, the domain is $\mathbb{R}^n$, so f is automatically continuous on $\mathcal{C}$. And for general compact $\mathcal{C}$, the Krein-Milman theorem tells us that $\mathcal{C}$ is the convex hull of its extreme points, i.e. for the optimizer x^*, there exists extreme points $\{x_1, ..., x_N\}$ and $\lambda \geq 0, \sum_{i=1}^{N} \lambda_i = 1$ such that $x^* = \sum_{i=1}^{N} \lambda_i x_i$.

2. Describe the set of boundary points for the following norm balls, i.e. points not in the interior. For each boundary point $\mathbf{x}$ give the set of supporting hyperplanes at $\mathbf{x}$.

1. $B_1^n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i| \leq 1\}$
2. $B_2^n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^n x_i^2} \leq 1\}$
3. $B_\infty^n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_\infty := \max_{i\in[n]} |x_i| \leq 1\}$.

Solution:

1. x is on the boundary iff $\|x\|_1 = 1$, and in this case the set of supporting hyperplanes have normal vectors
$$\{w \in [-1,1] \mid \forall i \in \mathrm{supp}(x) : w_i := \mathrm{sign}(x_i)\}.$$
2. x is on the boundary iff $\|x\|_2 = 1$, and in this case the supporting hyperplane has normal vector x.
3. x is on the boundary iff $\|x\|_\infty = 1$, and in this case the set of supporting hyperplanes have normal vectors
$$\mathrm{conv}\{e_i \mid |x_i| = 1\}.$$

3. For convex closed $f : \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathrm{dom}(f)$:

- Show for any subgradient $g \in \partial f(x)$, $(g, -1)$ gives a *supporting hyperplane* at $(x, f(x))$ for the epigraph
$$\mathrm{epi}(f) := \{(x,t) \mid f(x) \leq t\}.$$
- Show that $g \in \partial f(x)$ gives a *supporting hyperplane* at x for the sub-level set
$$L := \{y \in \mathbb{R}^n \mid f(y) \leq f(x)\}.$$

Solution:

-
$$\begin{aligned} g \in \partial f(x) &\iff \forall y : f(y) \geq f(x) + \langle g, y - x\rangle \\ &\iff \langle (g,-1),(x,f(x))\rangle \geq \sup_y \langle (g,-1),(y,f(y))\rangle = \sup_{(y,t)\in\mathrm{epi}(f)} \langle (g,-1),(y,t)\rangle \end{aligned}$$

where the first step was by definition, and in the last step we used $(y,t) \in \mathrm{epi}(f) \iff f(y) \leq t$.

- We claim $L \subseteq \{y \mid \langle g, y\rangle \le \langle g, x\rangle\} =: H$, which is clearly a supporting hyperplane at x. To show this, assume $y \notin H$ so $\langle g, y\rangle > \langle g, x\rangle$. Then by definition of subgradient we have

$$f(y) - f(x) \ge \langle g, y - x\rangle > 0,$$

i.e. $y \notin L$ the sub-level set.

4. Prove Exercise 2.27 in [BV]: Let $K \subseteq \mathbb{R}^n$ be closed, bounded, with non-empty interior, such that there exists a supporting hyperplane of K at every point of the boundary $x \in \partial K$. Show this implies K is convex.

Solution: We prove the contrapositive: if K is not convex, then there is a point on the boundary with no supporting hyperplane. K is not convex iff $\exists x, y \in K$ such that the line segment $[x, y]$ is not contained in K. We make the technical assumption athat $x, y \in \operatorname{int}(K)$ and describe how to extend to the general case at the end. As the line segment leaves K, there must be some $\lambda \in (0, 1)$ such that $x_\lambda := (1-\lambda)x + \lambda y \in \partial K$. Assume for contradiction $w \in \mathbb{R}^n$ gives a supporting hyperplane for K at x_λ. Then

$$\begin{aligned}\langle w, x_\lambda\rangle &= \max_{z\in K}\langle w, z\rangle \ge \max\{\langle w, x\rangle, \langle w, y\rangle\} \\ &= \langle w, x_\lambda\rangle + \max\{\lambda\langle w, x - y\rangle, (1-\lambda)\langle w, y - x\rangle\},\end{aligned}$$

and since $\lambda \in (0, 1)$, we must have $\langle w, x - y\rangle = 0$, so the whole line segment $[x, y]$ is on the boundary of the hyperplane. But this contradicts the fact that $x, y \in \operatorname{int}(K)$.

To remove the assumption, we use that K has non-empty interior, so we can take a small convex combination of x, y with $z \in \operatorname{int}(K)$ to push them into $\operatorname{int}(K)$.

5. Affine and quadratic functions are the most basic convex functions. We will prove some properties about them:

   - Show $h : \mathbb{R}^n \to \mathbb{R}$ is convex *and* concave (i.e. h is convex and $-h$ is convex) iff h is an affine function, i.e. $h(x) = \langle a, x\rangle + b$ for $a \in \mathbb{R}^n, b \in \mathbb{R}$.
   - Let $\tilde{q}(x) := \langle x, Qx\rangle + \langle a, x\rangle + b$ for symmetric matrix $Q \in \mathbb{R}^{n\times n}, a \in \mathbb{R}^n, b \in \mathbb{R}$. Show $\tilde{q}$ is convex iff $q(x) := \langle x, Qx\rangle$ is convex. (Hint: use first part)
   - Show $q(x) = \langle x, Qx\rangle$ is convex iff

   $$\forall v \in \mathbb{R}^n : \quad \langle v, Qv\rangle \ge 0,$$

   and similarly show it is strictly convex iff the inequality is strict for any $v \neq 0$. (These conditions are known as positive-semi-definiteness and positive-definiteness, and are denoted $Q \succeq 0, Q \succ 0$.)

- Find the optimizer and optimum value of strictly convex quadratic

$$\tilde{q}(x) := \langle x, Qx\rangle + \langle a, x\rangle + b.$$

Solution:

- Clearly affine h is convex and concave, so we focus on the converse. By convexity and concavity, respectively, we have

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

so in fact we must have equality for all $x, y \in \mathbb{R}^n, \lambda \in [0,1]$. We can use this to explicitly compute the affine function:

$$h(x) = h(0) + \sum_{i=1}^{n}(h(e_i) - h(0))x_i.$$

- We use that sums of convex functions are convex, and

$$\tilde{q}(x) = q(x) + h(x), \quad q(x) = \tilde{q}(x) - h(x)$$

where $h(x) := \langle a, x\rangle + b$ is an affine function, so convex and concave. So $\tilde{q}$ is convex iff q is.

- We first note that

$$\begin{aligned}
&\lambda xx^T + (1-\lambda)yy^T - (\lambda x + (1-\lambda)y)(\lambda x + (1-\lambda)y)^T \\
&\qquad = (\lambda - \lambda^2)xx^T + ((1-\lambda) - (1-\lambda)^2)yy^T - \lambda(1-\lambda)(xy^T + yx^T) \\
&\qquad = \lambda(1-\lambda)[xx^T + yy^T - xy^T - yx^T] = \lambda(1-\lambda)(x-y)(x-y)^T.
\end{aligned}$$

Therefore we can rewrite

$$\begin{aligned}
&\lambda q(x) + (1-\lambda)q(y) - q(\lambda x + (1-\lambda)y) \\
&\qquad = \langle Q, \lambda xx^T + (1-\lambda)yy^T - (\lambda x + (1-\lambda)y)(\lambda x + (1-\lambda)y)^T\rangle \\
&\qquad = \lambda(1-\lambda)\langle Q, (x-y)(x-y)^T\rangle.
\end{aligned}$$

Now q is convex iff the above is non-negative for all $x, y \in \mathbb{R}^n, \lambda \in [0,1]$, which is equivalent to the condition $\langle v, Qv\rangle \geq 0$ for $v = x - y$.

- We compute the gradient and check where it vanishes:

$$0 = \nabla\tilde{q}(x^*) = 2Qx + a \quad \iff \quad x^* = -\frac{Q^{-1}a}{2}$$

where we used that $\tilde{q}$ is strictly convex so Q is invertible. Substituting gives

$$\tilde{q}(x^*) = \frac{\langle Q^{-1}a, Q(Q^{-1}a)\rangle}{4} - \frac{\langle a, Q^{-1}a\rangle}{2} + b = -\frac{\langle a, Q^{-1}a\rangle}{4} + b.$$

6. An Ellipsoid is an affine image of the Euclidean ball

$$\mathcal{E} = c + AB_2^n \quad \text{where} \quad B_2^n := \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\},$$

for some $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ invertible.

- Let $q(x) := \langle x, Qx\rangle + \langle d, x\rangle + e$ be *strictly convex*. Show any sub-level set

$$L_t := \{x \in \mathbb{R}^n \mid q(x) \leq t\}$$

is either empty or an Ellipsoid (i.e. find c, A such that $L_t = c + AB_2$).

- Conversely, given Ellipsoid $\mathcal{E} = c + AB_2$ as above, find convex quadratic $q(x) := \langle x, Qx\rangle + \langle d, x\rangle + e$ such that

$$\mathcal{E} = \{x \in \mathbb{R}^n \mid q(x) \leq 1\}.$$

Solution:

- By the last part of the above question, we can compute the optimizer

$$x^* := \arg\min_x q(x) = -\frac{Q^{-1}d}{2} \quad \text{with} \quad q(x^*) = e - \frac{\langle d, Q^{-1}d\rangle}{4}.$$

so clearly L_t is empty for $t < q(x^*)$. We can rewrite q in terms of the optimizer

$$q(x) = \langle (x - x^*), Q(x - x^*)\rangle + q(x^*),$$

as can be verified directly. Therefore, for any $t \geq q(x^*)$ we have

$$\begin{aligned} x \in L_t &\iff \langle (x - x^*), Q(x - x^*)\rangle \leq t - q(x^*) \\ &\iff \|Q^{1/2}(x - x^*)\|_2 \leq \sqrt{t - q(x^*)} \\ &\iff x - x^* \in \sqrt{t - q(x^*)}Q^{-1/2}B_2^n, \end{aligned}$$

where in the last step we used that Q is invertible. This gives the required form for the Ellipsoid.

- We reverse the sequence of equivalences

$$\begin{aligned} x \in \mathcal{E} = c + AB_2^n &\iff x - c \in AB_2^n \iff A^{-1}(x - c) \in B_2^n \\ &\iff 1 \geq \|A^{-1}(x - c)\|_2^2 = \langle (x - c), (AA^T)^{-1}(x - c)\rangle, \end{aligned}$$

so $q(x) := \langle (x - c), (AA^T)^{-1}(x - c)\rangle$ is the required quadratic.

7. Recall the GLS oracle model for convex sets. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex closed function with $\mathrm{dom}(f) = \mathbb{R}^n$. We investigate natural function oracles in terms of the epigraph:

- Show MEM(epi(f)) can be implemented using an EVALuation oracle for f; show EVAL(f) can be approximately implemented using MEM(epi(f)), i.e. for input $x \in \mathbb{R}^n$ compute $t \in \mathbb{R}$ such that $|t - f(x)| \leq \varepsilon$.
- Show SEP(epi(f)) can be implemented using an GRADient (or subgradient) and EVALuation oracles for f; show GRAD(f) can be approximately implemented using SEP(epi(f)), i.e. for input $x \in \mathbb{R}^n$ compute $g \in \mathbb{R}^n$ such that

$$\forall y \in \mathbb{R}^n : \quad f(y) \geq f(x) - \varepsilon + \langle g, y - x \rangle.$$

- Relate OPT(epi(f)) and EVAL and GRAD for the Fenchel dual

$$f^*(w) := \sup_{x \in \mathbb{R}^n} \langle w, x \rangle - f(x).$$

- Relate MEM(epi(f^*)) and VAL(epi(f)).

Solution:

- $(x, t) \in \text{epi}(f) \iff f(x) \leq t$, so for query MEM(epi($f$), (x, t)) we can compute EVAL(f, x) and output YES iff $f(x) \leq t$. For the opposite direction, we can apply binary search on t to find an ε-approximation for $f(x)$ in $\log \frac{f_{\max} - f_{\min}}{\varepsilon}$ time, where in order for this to be finite, we need the assumption that we have some bounds $f_{\min} \leq f(x) \leq f_{\max}$.

- The membership part of SEP(epi) is implemented using the EVALuation oracle for f as above. So assume query $(x, t) \notin \text{epi}(f)$. Then by part (1) of question 3 above, we have that for any $g \in \partial f(x)$ $(g, -1)$ is a supporting hyperplane for the epigraph at $(x, f(x))$, so we can use the GRADient oracle to output this separating hyperplane.

  For the reverse direction, we can first use our SEParation oracle to compute t such that $f(x) > t \geq f(x) - \varepsilon$ using binary search. Further, $(x, t) \notin \text{epi}(f)$ so the oracle outputs separating hyperplane $(\tilde{g}, -\nu)$ such that

$$\langle (\tilde{g}, -\nu), (x, t) \rangle = \langle \tilde{g}, x \rangle - \nu t > \sup_{(y,s) \in \text{epi}(f)} \langle (g, -\nu), (y, s) \rangle.$$

  Since epi(f) is upwards closed, we must have $\nu \geq 0$. Further, since dom(f) $= \mathbb{R}^n$, we must in fact have $\nu > 0$. Therefore, by normalizing $g := \tilde{g}/\nu$ we have

$$\langle g, x \rangle - t > \sup_y \langle g, y \rangle - f(y),$$

  so by rearranging and using that $f(x) > t \geq f(x) - \varepsilon$ we have that g is an approximate subgradient.

- $f^*(w) = h_{\mathrm{epi}(f)}(w, -1)$ so the value output by $\mathrm{OPT}(\mathrm{epi}(f), (w, -1))$ is equivalent to $\mathrm{EVAL}(f^*, w)$. Further it can be directly verified that

$$x \in \partial f^*(w) \iff f^*(w) = \langle w, x \rangle - f(x),$$

so the optimizer output by $\mathrm{OPT}(\mathrm{epi}(f), (w, -1))$ is equivalent to $\mathrm{GRAD}(f^*, w)$.

- $(w, \nu) \in \mathrm{epi}(f^*) \iff f^*(w) \leq \nu \iff h_{\mathrm{epi}(f)}(w, -1) \leq \nu \iff \langle (w, -1), \cdot \rangle \leq \nu$ if VALid for $\mathrm{epi}(f)$.