Final exam

By now, you know how it works, so we won’t repeat it here. (If not, see the instructions for the EE364a final exam.) Since you have 96 hours to work on the final, your solutions must be typeset using \LaTeX. We are expecting your solutions to be typo-free, clear, and correctly typeset. (And yes, we will deduct points for poor typesetting, typos, or unclear solutions.) All code submitted must be clear, commented and readable.

To download files containing problem data, you’ll have to type the whole URL given in the problem into your browser; there are no links on the course web page pointing to these files. To get a file called filename.foo, for example, you would retrieve

http://www.stanford.edu/class/ee364b/final_data/filename.foo

with your browser.

Please make sure each problem starts on a new page, say, by using the \clearpage command. (This generates a new page after printing out any figures that have floated forward.)

Upload your submissions to gradescope by Tuesday June 12th 6pm at the latest.
1. [10 points] Polyak stepping and linear convergence of subgradient methods

Let $X \subset \mathbb{R}^n$ be a compact convex set. Assume that the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, subdifferentiable, $G$-Lipschitz on $X$ (i.e. $g \in \partial f(x)$ implies $\|g\|_2 \leq G$ for $x \in X$) and minimized over $X$ at the point $x^* \in X$, and satisfies the growth condition

$$f(x) \geq f(x^*) + \lambda \|x - x^*\|_2 \quad \text{for } x \in X$$

for some $\lambda > 0$. Show that if

$$x_{k+1} = \Pi_X(x^{(k)} - \alpha_k g^{(k)})$$

where $g^{(k)} \in \partial f(x^{(k)})$, $\Pi_X$ denotes (Euclidean) projection onto the set $X$, and $\alpha_k$ is chosen according to the Polyak stepping rule

$$\alpha_k = \frac{f(x^{(k)}) - f(x^*)}{\|g^{(k)}\|_2^2},$$

then

$$\|x^{(k+1)} - x^*\|_2 \leq C \|x^{(k)} - x^*\|_2$$

for a constant $C < 1$. Specify your constant $C$. 
2. [15 points] Comparison of distances to intersections of half-spaces

Let \( H_1 \) and \( H_2 \) be two halfspaces in \( \mathbb{R}^n \), where

\[
H_1 = \{ x \in \mathbb{R}^n | a_1^T x \leq 0 \} \quad \text{and} \quad H_2 = \{ x \in \mathbb{R}^n | a_2^T x \leq 0 \}
\]

Assume that \( H_1 \) and \( H_2 \) have a nice intersection, meaning that

\[-1 < \frac{a_1^T a_2}{\|a_1\|_2 \|a_2\|_2} < 1\]

(i.e. \( a_1 \) and \( a_2 \) are not parallel). For a (convex) set \( C \) define the distance function

\[d_C(x) = \inf_{y} \{\|x - y\|_2 | y \in C\} .\]

Let \( (t)_+ = \max\{t, 0\} \) and \( (t)_- = \max\{-t, 0\} \) denote the positive and negative parts, respectively, of their arguments. Show that

\[
\frac{1}{2} d_{H_1}(x)^2 + \frac{1}{2} d_{H_2}(x)^2 \leq d_{H_1 \cap H_2}(x)^2 \leq \kappa \cdot [d_{H_1}(x)^2 + d_{H_2}(x)^2]
\]

for all \( x \in \mathbb{R}^n \), where

\[
\kappa \leq \frac{1}{1 - (a_1^T a_2)_- / (\|a_1\|_2 \|a_2\|_2)} = \frac{1}{1 - (-a_1^T a_2)_+ / (\|a_1\|_2 \|a_2\|_2)} .
\]

(Note: You may not use the results of Question (3c) for this problem.)
3. [27 points] Distances, regular intersection, and growth

Let $C$ be a closed convex set in $\mathbb{R}^n$, and recall the distance function $d_C(x) = \inf_{y \in C} \{\|y - x\|_2\}$. The normal cone to $C$ at the point $x$ is

$$\mathcal{N}_C(x) := \begin{cases} \emptyset & \text{if } x \notin C \\ \{v \in \mathbb{R}^n \mid v^T(y - x) \leq 0 \text{ for all } y \in C\} & \text{if } x \in C. \end{cases}$$

Let $C_1$ and $C_2$ be closed convex sets in $\mathbb{R}^n$. Then $C_1$ and $C_2$ have $c$-regular intersection at $x$ if

$$v_1 \in \mathcal{N}_{C_1}(x) \quad \text{and} \quad v_2 \in \mathcal{N}_{C_2}(x) \quad \text{implies} \quad -v_1^Tv_2 \leq c \|v_1\|_2 \|v_2\|_2,$$

and that they have $c$-regular intersection if the implication (2) holds for all $x \in \mathbb{R}^n$ (where we can ignore points $x \notin C_1 \cap C_2$). Note that $\mathcal{N}_C(x) = \{0\}$ if $x \in \text{int} C$.

(a) [4 points] Show that $\pi \in C$ is the projection of $x$ onto $C$ if and only if

$$x - \pi \in \mathcal{N}_C(\pi).$$

(b) [2 points] Draw two convex sets with $c < 1$-regular intersection, indicating the normal cones of the sets at a point at which they intersect.

(c) [15 points] Show that if $C_1$ and $C_2$ have $c < 1$-regular intersection, then

$$d_{C_1 \cap C_2}(x)^2 \leq \frac{1}{1 - c} \left[ d_{C_1}(x)^2 + d_{C_2}(x)^2 \right].$$

You may assume that $\mathcal{N}_{C_1 \cap C_2}(x) = \mathcal{N}_{C_1}(x) + \mathcal{N}_{C_2}(x)$ for all $x \in C_1 \cap C_2$, which is true, but we do not require you to prove it. (*Hint:* The results of question 2 may be helpful.)

(d) [6 points] Assume that $C_1$ and $C_2$ have $c < 1$-regular intersection. Prove that the alternating projections method converges linearly, that is, that for some constant $\kappa < 1$

$$d_{C_1 \cap C_2}(x^{(k+1)})^2 \leq \kappa d_{C_1 \cap C_2}(x^{(k)})^2.$$

For example, it is possible to show this result with $\kappa = \frac{1+c}{2}$. (Hint: use the representation of alternating projections as Polyak stepping on the function $f(x) = \max\{d_{C_1}(x), d_{C_2}(x)\}$.)
4. [15 points] Phase retrieval and nonconvex projections

In X-ray crystallography and Fourier ptychography, one shines light through an object whose structure one wishes to recover, such as a molecule, and observes the resulting pattern on a detector behind the object. Because of physical limitations, the detector collects only the amplitude of the light waves, rather than their phase, which makes the inverse recovery problem challenging. In a stylized version (ignoring the complex plane) of this problem, we have $x^* \in \mathbb{R}^n$, and for a sequence of vectors $a_i \in \mathbb{R}^n$, we observe

$$b_i = (a_i^T x^*)^2, \quad i = 1, \ldots, m.$$ 

A natural objective in this case is the convex composite objective

$$f(x) = \| (Ax)^2 - b \|_2,$$

where $(\cdot)^2$ denotes elementwise squaring and $A = [a_1 \cdots a_m]^T \in \mathbb{R}^{m \times n}$ is the data matrix. One can apply the compositional non-convex optimization approaches we develop in class. An alternative approach is to perform alternating projections onto the non-convex sets

$$C_i := \{ x \in \mathbb{R}^n \mid (a_i^T x)^2 = b_i \}. \quad (3)$$

(a) [5 points] Show how to compute the projection of a vector $x \in \mathbb{R}^n$ onto the set $C_i$ in Eq. (3).

In applications, structured matrices $A \in \mathbb{R}^{m \times n}$ are the only possible measurements. We consider Hadamard matrices, where the $k$th Hadamard matrix is defined recursively by

$$H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_k = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix}.$$ 

Let $n = 2^k$ for some $k$. We assume oversampling by a factor of $L > 1$, where we have $L$ random sign vectors $s_l \in \{-1, 1\}^n$, and $A \in \mathbb{R}^{Ln \times n}$ is

$$A = \begin{bmatrix} H_k \text{diag}(s_1) \\ \vdots \\ H_k \text{diag}(s_L) \end{bmatrix} \in \mathbb{R}^{Ln \times n}. \quad (4)$$

(b) [10 points] We consider recovering a $32 \times 32$ (grayscale) image represented by $x \in \mathbb{R}^n$ with $n = 32^2 = 1024$, using randomized Hadamard measurements with $L = 6$-times oversampling. The files prdata-S.dat and prdata-b.dat contain random signs $s_l$ and measurements $b \in \mathbb{R}^{Ln}$. prdata-S.dat contains an $n \times L$ matrix of random signs (rows are separated by semicolons), where the $l$th column of prdata-S.dat is the $l$th vector $s_l$; prdata-b.dat contains an $Ln = 6144$ length vector of measurements $b = (Ax^*)^2$ for matrix $A$ of the form (4).

With this data, implement a (randomized) alternating projections procedure, which projects onto a randomly chosen set $C_i$ of Eq. (3) at each iteration, and run your
procedure for $T = 12Ln$ steps. Plot the function value $f(x)$ every $n$ iterations. Do this experiment a few times, and on one of the executions in which the function value $f(x^{(T)})$ (the last iteration) is smallest, display the resulting image $x^{(T)}$ that you recover. (To do this, you will need to reshape $x \in \mathbb{R}^n$ into a $32 \times 32$ matrix, then display it as an image.) Include this figure in your submission as well. Your solution should include your code and the two plots above.
5. [33 points] Inventory management and robust optimization

In an inventory management problem with \( n \) items, we have a so-called two-stage optimization problem (though in this case, it can be reduced to a single convex objective). In the first stage, we order \( x \in \mathbb{R}_+^n \) units, where \( x_j \) is the amount of item \( j \) ordered. In the second stage, we receive a random demand \( d \in \mathbb{R}_+^n \), and then must either re-order more units of unmet demand or suffer losses for unsold inventory. We formulate this as a stochastic optimization problem as follows: before the realized demand appears, there is a fixed cost vector \( c \in \mathbb{R}_+^n \) for ordering items. The realized scenario is a vector \( \xi = (d, r, u) \), where \( d \in \mathbb{R}_+^n \) is the demand for each of the \( n \) items, \( r \in \mathbb{R}_+^n \) is the (per-unit) price for reordering items, and \( u \in \mathbb{R}_+^n \) is the cost of unsold inventory (because of storage or destruction fees). We may write this optimization problem as

\[
\begin{align*}
\text{minimize} \quad & c^T x + \mathbb{E}[f(x; \xi)] \\
\text{subject to} \quad & x \succeq 0,
\end{align*}
\]

with

\[
f(x; \xi) = f(x; d, r, u) = r^T (d - x) + u^T (x - d),
\]

where \((\cdot)_+\) denotes the elementwise positive part of its argument.

We consider two approaches to this problem: a Monte-Carlo based approach and one using robust optimization. For the robust optimization approach, we consider a robustness set defined over (empirical) distributions. Recall that the \( \phi \)-divergence between vectors \( p, q \in \mathbb{R}_+^m \) with \( p^T 1 = q^T 1 = 1 \) is

\[
\mathcal{D}_\phi (p\|q) = \sum_{i=1}^m \phi \left( \frac{p_i}{q_i} \right) q_i,
\]

where \( \phi : \mathbb{R}_+ \to \mathbb{R} \) is a closed convex function with \( \phi(1) = 0 \) and \( \phi(t) = +\infty \) for \( t < 0 \). Now, for a fixed value \( \rho > 0 \), consider the set \( \mathcal{P}_m \) defined by

\[
\mathcal{P}_m := \{ p \in \mathbb{R}_+^m \mid \mathcal{D}_\phi (p\|(1/m)1) \leq \rho, p^T 1 = 1 \}.
\]

(a) [5 points] Show that for any vector \( z \in \mathbb{R}^m \), we have

\[
\sup_{p \in \mathcal{P}_m} p^T z = \inf_{\lambda \geq 0, \eta} \left\{ \frac{\lambda}{m} \sum_{i=1}^m \phi^* \left( \frac{z_i - \eta}{\lambda} \right) + \rho \lambda + \eta \right\},
\]

where \( \phi^*(s) = \sup_{t \geq 0} \{ st - \phi(t) \} \) is the conjugate of \( \phi \). (Remark: You may not simply cite the lecture slides for this result. You may ignore the case \( \lambda = 0 \) in your result if it is convenient.)

(b) [5 points] Consider the function \( \phi(t) = \frac{1}{2} t^2 - \frac{1}{2} \) (for \( t \geq 0 \), while \( \phi(t) = +\infty \) for \( t < 0 \)). Show that with this choice in the definition of the uncertainty set \( \mathcal{P}_m \),

\[
\sup_{p \in \mathcal{P}_m} p^T z = \inf_{\eta} \sqrt{1 + \frac{2 \rho}{\eta} \left( \frac{1}{m} \sum_{i=1}^m (z_i - \eta)_+^2 \right)^{1/2}} + \eta.
\]

(Remark: You may not simply cite the lecture slides for this result.)
(c) [5 points] For the two-stage problem (5), show how to write the problem as an SOCP when the expectation $E$ is taken over a finite sample $\xi_1, \ldots, \xi_m$ of vectors. That is, write the problem

$$\text{minimize } c^T x + \sup_{p \in \mathcal{P}} \sum_{i=1}^m p_i f(x; \xi_i) \text{ subject to } x \succeq 0$$

as an SOCP.

(d) [18 points] Now for values of $m = 20, 40, 80, 160$, we will perform a simulation experiment. Let $n = 5$, so that we consider a 5 item problem, and define the cost vector $c$ to have entries $c_j = 1/j$, $j = 1, \ldots, n$. In each simulation, perform the following steps.

i. generate $m$ random realizations $\xi_i$ according to the following distribution. Let $\zeta_i \in [0, 3]^n$ have independent entries uniform on $[0, 3]$, set the demand $d_i = \zeta_i$, and set the reorder price $r_i = (1 + \zeta_i) \odot c$, where $\odot$ denotes elementwise multiplication. Set the unsold price $u_i = \text{Uniform}([0, c])$, that is, with each entry $u_{ij}$ uniform in $[0, c_j]$.

ii. For your data $\{\xi_i\}_{i=1}^m$ and robustness level $\rho = 3/m$, use CVX to solve the robust problem from part (5c). Let $\hat{x}_{ro}$ denote this solution. Use CVX to also solve the problem (5) with the expectation $E$ taken as the Monte-Carlo average over your sample, i.e. $\frac{1}{m} \sum_{i=1}^m f(x; \xi_i)$. Let $\hat{x}_{mc}$ denote this solution.

iii. Now generate $N = 200$ additional samples $\xi_i$ according to the distribution over $\xi_i$ above. On this new set, compute

$$f_{ro} = c^T \hat{x}_{ro} + \frac{1}{N} \sum_{i=1}^N f(\hat{x}_{ro}; \xi_i), \quad f_{mc} = c^T \hat{x}_{mc} + \frac{1}{N} \sum_{i=1}^N f(\hat{x}_{mc}; \xi_i)$$

and the extreme values

$$L = c^T \hat{x} + \min_{i \leq N} f(\hat{x}; \xi_i) \quad \text{and} \quad U = c^T \hat{x} + \max_{i \leq N} f(\hat{x}; \xi_i)$$

for both the Monte-Carlo solution $\hat{x}_{mc}$ and robust solution $\hat{x}_{ro}$.

Perform steps i–iii above $T = 40$ times for each value of $m = 20, 40, 80, 160$. Plot the mean of the values $f_{ro}$ and $f_{mc}$ for each of these sample sizes (mean taken across the $T$ tests), as well as the mean values for $L$ and $U$ for each of these $m$. Include plots in your solution as well as your code. What do you observe?