EE364a: Convex Optimization I — N. Moehle
August 13–14 or August 14–15, 2016

Final Exam

This is a 24 hour take-home final. Please turn it in at Bytes Cafe in the Packard building, 24 hours after you pick it up.

You may use any books, notes, or computer programs, but you may not discuss the exam with anyone until August 16, after everyone has taken the exam. The only exception is that you can ask us for clarification, via the course staff email address. We've tried pretty hard to make the exam unambiguous and clear, so we're unlikely to say much.

Please make a copy of your exam, or scan it, before handing it in.

Please attach the cover page to the front of your exam. Assemble your solutions in order (problem 1, problem 2, problem 3, ...), starting a new page for each problem. Put everything associated with each problem (*e.g.*, text, code, plots) together; do not attach code or plots at the end of the final.

We will deduct points from long, needlessly complex solutions, even if they are correct. Our solutions are not long, so if you find that your solution to a problem goes on and on for many pages, you should try to figure out a simpler one. We expect neat, legible exams from everyone, including those enrolled Cr/N.

When a problem involves computation you must give all of the following: a clear discussion and justification of exactly what you did, the source code that produces the result, and the final numerical results or plots.

Files containing problem data can be found in the usual place,

`http://www.stanford.edu/~boyd/cvxbook/cvxbook_additional_exercises/`

Please respect the honor code. Although we allow you to work on homework assignments in small groups, you cannot discuss the final with anyone, at least until everyone has taken it.

All problems have equal weight. Some are easy. Others, not so much.

Be sure you are using the most recent version of CVX, CVXPY, or Convex.jl. Check your email often during the exam, just in case we need to send out an important announcement.

Some problems involve applications. But you do not need to know *anything* about the problem area to solve the problem; the problem statement contains everything you need.

1. *Funding an expense stream.* Your task is to fund an expense stream over n time periods. We consider an expense stream $e \in \mathbf{R}^n$, so that e_t is our expenditure at time t.

   One possibility for funding the expense stream is through our bank account. At time period t, the account has balance b_t and we withdraw an amount w_t. The value of our bank account accumulates with an interest rate ρ per time period, less withdrawals:

   $$b_{t+1} = (1+\rho)b_t - w_t.$$

   We assume the account value must be nonnegative, so that $b_t \geq 0$ for all t.

   We can also use other investments to fund our expense stream, which we purchase at the initial time period $t = 1$, and which pay out over the n time periods. The amount each investment type pays out over the n time periods is given by the *payout matrix* P, defined so that P_{tj} is the amount investment type j pays out at time period t per dollar invested. There are m investment types, and we purchase $x_j \geq 0$ dollars of investment type j. In time period t, the total payout of all investments purchased is therefore given by $(Px)_t$.

   In each time period, the sum of the withdrawals and the investment payouts must cover the expense stream, so that

   $$w_t + (Px)_t \geq e_t$$

   for all $t = 1, \ldots, n$.

   The total amount we invest to fund the expense stream is the sum of the initial account balance, and the sum total of the investments purchased: $b_1 + \mathbf{1}^T x$.

   (a) Show that the minimum initial investment that funds the expense stream can be found by solving a convex optimization problem.

   (b) Using the data in `expense_stream_data.*`, carry out your method in part (a). On three graphs, plot the expense stream, the payouts from the m investment types (so m different curves), and the bank account balance, all as a function of the time period t. Report the minimum initial investment, and the initial investment required when no investments are purchased (so $x = 0$).

2. *Approximations of the PSD cone.* A symmetric matrix is positive semidefinite if and only if all its principal minors are nonnegative. Here we consider approximations of the positive-semidefinite cone produced by partially relaxing this condition.

   Denote by $K_{1,n}$ the cone of matrices whose 1×1 principal minors (*i.e.*, diagonal elements) are nonnegative, so that

$$K_{1,n} = \{X \in \mathbf{S}^n \mid X_{ii} \geq 0 \text{ for all } i\}.$$

   Similarly, denote by $K_{2,n}$ the cone of matrices whose 1×1 and 2×2 principal minors are nonnegative:

$$K_{2,n} = \left\{ X \in \mathbf{S}^n \;\middle|\; \begin{bmatrix} X_{ii} & X_{ij} \\ X_{ij} & X_{jj} \end{bmatrix} \succeq 0, \text{ for all } i \neq j \right\},$$

   i.e., the cone of symmetric matrices with positive semidefinite 2×2 principal submatrices. These two cones are convex (and in fact, proper), and satisfy the relation:

$$K_{1,n}^* \subseteq K_{2,n}^* \subseteq \mathbf{S}_+^n \subseteq K_{2,n} \subseteq K_{1,n},$$

   where $K_{1,n}^*$ and $K_{2,n}^*$ are the dual cones of $K_{1,n}$ and $K_{2,n}$, respectively. (The last two inclusions are immediate, and the first two inclusions follow from the second bullet on page 53 of the text.)

   (a) Give an explicit characterization of $K_{1,n}^*$.

   (b) Give an explicit characterization of $K_{2,n}^*$.
   Hint: You can use the fact that if $K = K_1 \cap \cdots \cap K_m$, then $K^* = K_1^* + \cdots + K_m^*$.

   (c) Consider the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, CX \\ \text{subject to} & \mathbf{Tr}\, AX = b \\ & X \in K \end{array}$$

   with variable $X \in \mathbf{S}^n$. The problem parameters are $C \in \mathbf{S}^n$, $A \in \mathbf{S}^n$, $b \in \mathbf{R}$, and the cone $K \subseteq \mathbf{S}^n$. Using the data in `psd_cone_approx_data.*`, solve this problem five times, each time replacing K with one of the five cones $K_{1,n}$, $K_{2,n}$, $\mathbf{S}_+^n$, $K_{2,n}^*$, and $K_{1,n}^*$. Report the five different optimal values you obtain.
   Note: Python users who run into numerical difficulties might want to use the SCS solver by using `prob.solve(solver=cvxpy.SCS)`.

   Note: For parts (a) and (b), the shorter and clearer your description is, the more points you will receive. At the very least, it should be possible to implement your description in CVX.*.

3. *Direct standardization.* Consider a random variable $(x, y) \in \mathbf{R}^n \times \mathbf{R}$, and N samples $(x_1, y_1), \ldots, (x_N, y_N) \in \mathbf{R}^n \times \mathbf{R}$, which we will use to estimate the (marginal) distribution of y. If the given samples were chosen according to the joint distribution of (x, y), a reasonable estimate for the distribution of y would be the uniform empirical distribution, which takes on values $y_1, \ldots, y_N$ each with probability $1/N$. (If y is Boolean, *i.e.*, $y \in \{0, 1\}$, we are using the fraction of samples with $y = 1$ as our estimate of $\mathbf{Prob}(y = 1)$.)

   The bad news is that the samples $(x_1, y_1), \ldots, (x_N, y_N) \in \mathbf{R}^n \times \mathbf{R}$ were *not* chosen from the distribution of (x, y), but instead from another (unknown, but presumably similar) distribution. The good news is that we know $\mathbf{E}\, x$, the expected value of x. We will use our knowledge of $\mathbf{E}\, x$, together with the samples, to estimate the distribution of y. *Direct standardization* replaces the uniform empirical distribution with a weighted one, which takes on values y_i with probability π_i, where $\pi \succeq 0$, $\mathbf{1}^T \pi = 1$. The weights or sample probabilities π are found by maximizing the entropy $-\sum_{i=1}^N \pi_i \log \pi_i$, subject to the requirement that the weighted sample expected value of x matches the known probabilities of x in the distribution, $\mathbf{E}\, x$. This can be expressed as $\sum_{i=1}^N \pi_i x_i = \mathbf{E}\, x$. (Both x_i and $\mathbf{E}\, x$ are known.)

   (a) Explain why choosing π is a convex optimization problem.

   (b) Consider the simple case with $n = 1$, and $x \in \{0, 1\}$, so $\mathbf{E}\, x = \mathbf{Prob}(x = 1)$. Find the optimal sample weights $\pi_i^\star$ (analytically). Explain your solution in the following case. The samples are people, with $x = 0$ meaning the person is male, and $x = 1$ meaning the person is female. The overall population is known to have equal numbers of females and males, but in the sample population the male : female proportions are 0.7 : 0.3.

   (c) The data in `direct_std_data.*` contain the samples $x^{(i)}$ and $y^{(i)}$, as well as $\mathbf{E}\, x$. Find the weights $\pi^\star$, and report the weighted empirical distribution. On the same plot, compare the cumulative distributions of

   - the uniform empirical distribution,
   - the weighted empirical distribution using $\pi^\star$, and
   - the true distribution of y.

   The true and empirical distributions are provided in the data file. (For example, the 20 elements of `p_true` give $\mathbf{Prob}(y = 1)$ up to $\mathbf{Prob}(y = 20)$, in order).

   Note: Julia users might want to use the ECOS solver, by including `using ECOS`, and solving by using `solve!(prob, ECOSSolver())`.

   Note: You don't need to know this to solve the problem, but the data for part (c) are real. The random variable x is a vector of a student's gender, age, and mother's and father's educational attainment, and y is the student's score on a standardized test.

4. *Thermodynamic potentials.* We consider a mixture of k chemical species. The *internal energy* of the mixture is

$$U(S, V, N_1, \ldots, N_k),$$

where S is the entropy of the mixture, V is the volume occupied by the mixture, and N_i is the quantity (in moles) of chemical species i. We assume the function U is convex. (Real internal energy functions satisfy this and other interesting properties, but we won't need any others for this problem.) The *enthalpy* H, the *Helmoltz free energy* A, and the *Gibbs free energy* G are defined as

$$\begin{aligned} H(S, P, N_1, \ldots, N_k) &= \inf_V U(S, V, N_1, \ldots, N_k) - PV, \\ A(T, V, N_1, \ldots, N_k) &= \inf_S U(S, V, N_1, \ldots, N_k) + TS, \\ G(T, P, N_1, \ldots, N_k) &= \inf_{S,V} U(S, V, N_1, \ldots, N_k) + TS - PV. \end{aligned}$$

The variables T and P can be interpreted physically as the temperature and pressure of the mixture. These four functions are called *thermodynamic potentials.* We refer to the arguments S, V, and $N_1, \ldots N_k$ as the extensive variables, and the arguments T and P as the intensive variables.

(a) Show that H, A, and G are convex in the extensive variables, when the intensive variables are fixed.

(b) Show that H, A, and G are concave in the intensive variables, when the extensive variables are fixed.

(c) We consider a simple reaction involving three species,

$$2[\text{species } 1] \rightleftharpoons [\text{species } 2] + [\text{species } 3],$$

carried out at temperature T_{react} and volume V_{react}. The Helmholtz free energy of the mixture is

$$A(T, V, N_1, N_2, N_3) = T \sum_{j=1}^{3} N_j (s_{0,j} - R c_j) + TR \sum_{j=1}^{3} N_j \log \left(N_j \left(\frac{V_0}{V} \right) \left(\frac{T_0}{T} \right)^{c_j} \right),$$

where R, V_0, T_0, $s_{0,j}$, and c_j, for $j = 1, \ldots, k$, are known, positive constants. The equilibrium molar quantities $N_1^\star$, $N_2^\star$, and $N_3^\star$ of the three species are those that minimize $A(T_{\text{react}}, V_{\text{react}}, N_1, N_2, N_3)$ subject to the stoichiometry constraints

$$N_1 = N_{1,\text{init}} - 2z, \quad N_2 = N_{2,\text{init}} + z, \quad N_3 = N_{3,\text{init}} + z,$$

where $N_{j,\text{init}}$ is the initial quantity of species j, and the variable z gives the amount of the reaction that has proceeded. For the values of T_{react}, V_{react}, R, V_0, T_0, $s_{0,j}$, and c_j given in `thermo_potentials_data.*`, report the equilibrium molar quantities $N_1^\star$, $N_2^\star$, and $N_3^\star$.

Note: Julia users might want the ECOS solver. Include `using ECOS`, and solve by using `solve!(prob, ECOSSolver())`.

5. *Controlling a switched linear system via duality.* We consider a discrete-time dynamical system with state $x_t \in \mathbf{R}^n$. The state propagates according to the recursion

$$x_{t+1} = A_t x_t, \quad t = 0, 1, \ldots, T-1,$$

where the matrices A_t are to be chosen from a finite set $\mathcal{A} = \{A^{(1)}, \ldots, A^{(K)}\}$ in order to control the state x_t over a finite time horizon of length T. More formally, the switched-linear control problem is

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T} f(x_t) \\ \text{subject to} & x_{t+1} = A^{(u_t)} x_t, \quad \text{for } t = 0, \ldots, T-1 \end{array}$$

The problem variables are $x_t \in \mathbf{R}^n$, for $t = 1, \ldots, T$, and $u_t \in \{1, \ldots, K\}$, for $t = 0, \ldots, T-1$. We assume the initial state, $x_0 \in \mathbf{R}^n$ is a problem parameter (*i.e.*, is known and fixed). You may assume the function f is convex, though it isn't necessary for this problem.

Note that, to find a feasible point, we take any sequence $u_0, \ldots, u_{T-1} \in \{1, \ldots, K\}$; we then generate a feasible point according to the recursion

$$x_{t+1} = A^{(u_t)} x_t, \quad t = 0, 1, \ldots, T-1.$$

The switched-linear control problem is *not* convex, and is hard to solve globally. Instead, we consider a heuristic based on Lagrange duality.

(a) Find the dual of the switched-linear control problem explicitly in terms of x_0, $A^{(1)}, \ldots, A^{(K)}$, the function f, and its conjugate f^*. Your formulation cannot involve a number of constraints or objective terms that is exponential in K or T. (This includes minimization or maximization with an exponential number of terms.)

(b) Given optimal dual variables $\nu_1^\star, \ldots, \nu_T^\star$ corresponding to the T constraints of the switched-linear control problem, a heuristic to choose u_t is to minimize the Langrangian using these optimal dual variables:

$$(\tilde{u}_0, \ldots, \tilde{u}_{T-1}) \in \underset{u_0, \ldots, u_{T-1} \in \{1, \ldots, K\}}{\operatorname{argmin}} \inf_{x_1, \ldots, x_T} L(x_1, \ldots, x_T, u_0, \ldots, u_{T-1}, \nu_1^\star, \ldots, \nu_T^\star),$$

Given the optimal dual variables, show (explicitly) how to find $\tilde{u}_0, \ldots, \tilde{u}_{T-1}$.

(c) Consider the case $f(x) = (1/2) x^T Q x$. with $Q \in \mathbf{S}_{++}^n$. For the data given in `sw_lin_sys_data.*`, solve the dual problem and report its optimal value $d^\star$, which is a lower bound on $p^\star$. (As a courtesy, we also included $p^\star$ in the data file, so you can check your bound.)

Note: Julia users might want to use the ECOS solver, by including `using ECOS`, and solving by using `solve!(prob, ECOSSolver())`.

(d) Using the same data as is part (c), carry out the heuristic method of part (b) to compute $\tilde{u}_0, \ldots \tilde{u}_{T-1}$. Use these values to generate a feasible point. Report the value of the objective at this feasible point, which is an upper bound on $p^\star$.

6. *Elastic stored energy in a spring.* A spring is a mechanical device that exerts a force F that depends on its extension x: $F = \phi(x)$, where $\phi : \mathbf{R} \to \mathbf{R}$. The domain $\mathbf{dom}\,\phi$ is an interval $[x^{\min}, x^{\max}]$ containing 0, where $x^{\min}$ ($x^{\max}$) is the minimum (maximum) possible extension of the spring. When $x > 0$, the spring is said to be extended, and when $x < 0$, it is said to be in compression. The force exerted by the spring must be *restoring*, which means that $F \geq 0$ when $x \geq 0$, and $F \leq 0$ when $x \leq 0$. (Our sign convention is that a positive force F opposes a positive extension x.) This implies that $F = 0$ when $x = 0$, *i.e.*, zero force is developed when the spring is not extended or compressed.

   The simplest spring is a Hooke (linear) spring, with $\phi(x) = Kx$, where $K > 0$ is the *spring constant*. (The constant $1/K$ is called the spring *compliance*.)

   A spring is called *monotonic* if the function ϕ is nondecreasing, *i.e.*, larger extension leads to a stronger restoring force. Many, but not all, springs are monotonic. A classic example is a compound bow, which has a force that first increases with x, and then decreases to a small value at the extension x where it is fully drawn. (This decrease in force from the maximum is called the *let off* of the bow.)

   The elastic stored energy in the spring is

$$E(x) = \int_0^x \phi(x)\, dx,$$

   with domain $[x^{\min}, x^{\max}]$.

   Show that E is quasi-convex. Show that E is convex if and only if the spring is monotonic. You may assume ϕ is differentiable.

7. *Recovering latent periodic signals.* First, a definition: a signal $x \in \mathbf{R}^n$ is *p-periodic* with $p < n$ if $x_{i+p} = x_i$ for $i = 1, \ldots, n-p$.

   In this problem, we consider a noisy, measured signal $y \in \mathbf{R}^n$ which is (approximately) the sum of a several periodic signals, with unknown periods. Given only the noisy signal y, our task is to recover these latent periodic signals. In particular, y is given as

$$y = v + \sum_{p \in \mathcal{P}} x^{(p)},$$

   where $v \in \mathbf{R}^n$ is a (small) random noise term, and $x^{(p)}$ is a p-periodic signal. The set $\mathcal{P} \subset \{1, \ldots, p_{\max}\}$ contains the periods of the latent periodic signals that compose y.

   If $\mathcal{P}$ were known, we could approximately recover the latent periodic signals $x^{(p)}$ using, say, least squares. Because $\mathcal{P}$ is *not* known, we instead propose to recover the latent periodic signals $x^{(p)}$ by solving the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{p=1}^{p_{\max}} w_p \|\hat{x}^{(p)}\|_2 \\ \text{subject to} & \hat{y} = \sum_{p=1}^{p_{\max}} \hat{x}^{(p)} \\ & \hat{x}^{(p)} \text{ is } p\text{-periodic, for } p = 1, \ldots, p_{\max}. \end{array}$$

   The variables are $\hat{y}$ and $\hat{x}^{(p)}$, for $p = 1, \ldots, p_{\max}$. The first sum in the objective penalizes the squared deviation of the measured signal y from our estimate $\hat{y}$, and the second sum is a heuristic for producing vectors $\hat{x}^{(p)}$ that contain only zeros. The weight vector $w \succeq 0$ is increasing in its indices, which encodes our desire that the latent periodic signals have small period.

   (a) Explain how to solve the given optimization problem using convex optimization, and how to use it to (approximately) recover the set $\mathcal{P}$ and the latent periodic signals $x^{(p)}$, for $p \in \mathcal{P}$.

   (b) The file `periodic_signals_data.*` contains a signal y, as well as a weight vector w. Return your best guess of the set $\mathcal{P}$. plot the measured signal y, as well as the different periodic components that (approximately) compose it. (Use separate graphs for each signal, so you should have $|\mathcal{P}| + 1$ graphs.)