22494 Convex Optimization
July 20–21, 2020

Mojtaba Tefagh

Final Exam

This is a 24 hour take-home final. Please turn it in by uploading to Quera no more than 24 hours after you receive the final. We will *not* be very lenient with upload times: you must upload by the 11:00AM next-day deadline.

You may use any books, notes, or computer programs, but you may not discuss the exam with anyone until 11:00AM July 21, after everyone has finished the exam. The only exception is that you can ask us for clarification, via Quera. We've tried pretty hard to make the exam unambiguous and clear, so we're unlikely to say much.

Please submit your exam via Quera. 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 for all problems 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.

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 following location:

```
http://sharif.edu/~mtefagh/22494data/
```

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, which may be a while.

All problems have equal weight. Some are (quite) straightforward. 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. *Unbounded polyhedron.* Consider the polyhedron P defined by $Ax \leq b$ and $Cx = d$. Formulate a linear program to determine if P is nonempty and unbounded.

2. *A family of functions.* For constants α_1 and α_2, let $f_{\alpha_1,\alpha_2} : \mathbf{R}^2_+ \to \mathbf{R}_+$ be defined as

$$f_{\alpha_1,\alpha_2}(x_1, x_2) = x_1^{\alpha_1} x_2^{\alpha_2}, \qquad x_i \geq 0.$$

Identify and draw the regions in the (α_1, α_2) plane for which f_{α_1,α_2} is a

(a) convex function;

(b) concave function.

Extra credit. For a vector $\alpha \in \mathbf{R}^n$, let $f_\alpha : \mathbf{R}^n_+ \to \mathbf{R}_+$ be defined as

$$f_\alpha(x) = \prod_{i=1}^{n} x_i^{\alpha_i}, \qquad x_i \geq 0.$$

Generalized the previous result to this case.

3. *Reaction knockout in metabolic networks.* A metabolic network is the collection of all the biochemical reactions happening inside an organism, *e.g.*, a bacterium. Consider a metabolic network consisting of m metabolites and n reactions. Let S be an $m \times n$ matrix whose rows and columns correspond to the metabolites and reactions, respectively, and whose s_{ij} entry represents the relative rate of either consumption (if $s_{ij} \leq 0$) or production (if $s_{ij} \geq 0$) of the metabolite i in the reaction j. We call S the stoichiometric matrix as its columns characterize the stoichiometry of the reactions in the metabolic network.

If all the metabolites are in mass balance at specific densities, we say that the metabolic network is in the steady-state condition. This is equivalent to say that the densities of the metabolites do not change, *i.e.*, $Sv = 0$, where $v \in \mathbf{R}^n$ is the vector of the rates of reactions. Moreover, we assume that v satisfies $l \preceq v \preceq u$ where the lower and upper bounds are determined by the growth media and thermodynamic constraints.

In the given metabolic network, the last column of S specifies the amounts of metabolites consumed or produced per unit of cell growth rate, hence v_n corresponds to the biomass production. For a wild-type cell, *flux balance analysis* (FBA) predicts the wild-type $v^w \in \mathbf{R}^n$ by maximizing the rate of the biomass reaction, v_n, subject to the constraints described above.

For a given subsystem of reactions, the corresponding reaction knockout sets the lower and upper bounds of any reaction associated with that subsystem to zero. We can compute the updated $\tilde{l}$ and $\tilde{u}$ for a mutant cell with a known knocked-out subsystem by the subsystem association vector `subsystem`. Let $\tilde{v} \in \mathbf{R}^n$ denote the FBA prediction of the mutant activity rates which satisfies the steady-state constraint as well as the updated lower and upper bounds, *i.e.*, $\tilde{l} \preceq \tilde{v} \preceq \tilde{u}$.

The questions below pertain to the Recon3D model (`http://vmh.life`) found in `Recon3D.*`. The scripts `FBA.*` will help you import the data.

(a) What is the optimal biomass production rate for the wild-type cell?

(b) Find and report the relative change in biomass production, *i.e.*,

$$\frac{v_n^w - \tilde{v}_n}{v_n^w},$$

duo to subsystem knockout for the subsystems "Transport, nuclear" and "Fatty acid oxidation". Explain the difference.

(c) Redo the previous part for the case of single reaction knockout of each reaction in the "Transport, nuclear" subsystem. In this case, $\tilde{l}$ and $\tilde{u}$ are the same as l and u except for exactly one entry which is set to be zero and we try the entries corresponding to reactions in the "Transport, nuclear" subsystem one by one. Names of reactions can be found in the `name` vector. What are the two reactions whose deletion results in a relative change of at least 0.2 in biomass production?

Hint. This is a real-world problem with genome-scale data. If either your program runs for a long time or the optimizer terminates before returning the accurate solution, you need to change the solver you are using.

4. *Duality in metabolic networks.* In this problem, we will investigate some applications of Lagrange duality to demonstrate the fundamental chemical laws governing the stoichiometry of thousands of reactions in genome-scale metabolic networks. Don't worry if you haven't solved the previous problem. The only information you need from the previous problem is the definition of the stoichiometric matrix S.

(a) Suppose that in a metabolic network every reaction is thermodynamically irreversible, which means that it can be active in the forward direction spontaneously and in the backward direction only by getting lots of additional energy. Therefore, for any thermodynamically feasible $v \in \mathbf{R}^n$ representing the activation rates of the reactions we have

$$v \succeq 0.$$

Accordingly, we call a change in the densities of metabolites $c \in \mathbf{R}^m$ thermodynamically feasible if there exists $v \succeq 0$ for which

$$Sv = c.$$

If $\mu \in \mathbf{R}^m$ represents the potential of each different metabolite, it must satisfy the *second law of thermodynamics* which states that the total free energy is always reduced, *i.e.*,

$$S^T \mu \preceq 0.$$

For a fixed arbitrary c, prove that c is thermodynamically feasible if and only if for any such potential vector μ satisfying $S^T \mu \preceq 0$, it holds that $\mu^T c \leq 0$.

(b) The *law of mass conservation* states that mass cannot be created by any combination of a set of reactions. Hence, for any $v \in \mathbf{R}^n$ representing the activation rates of reactions

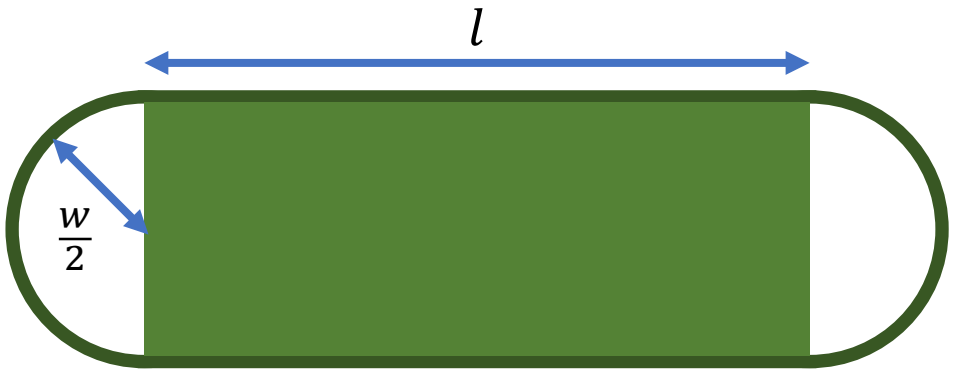
$$Sv \succeq 0 \Longrightarrow Sv = 0.$$

Under the above assumption, prove that there exists a vector $m \in \mathbf{R}^m$ representing the mass of each different metabolite which satisfies the following constraints:

(a) For all metabolites, the mass is always strictly positive, *i.e.*, $m \succ 0$.

(b) For all reactions, the mass of the products must equal the mass of the reactants, *i.e.*, $m^T S = 0$.

5. *Designing face masks.* In response to the dramatic increase in face mask demand due to the COVID-19 pandemic, a company wants to mass-produce face masks with the minimum cost possible. Although not to scale, the following diagram shows a proposed face mask design where the filter is shown by light green and the elastic cord is shown by dark green.



The basic idea is that the elastic cord loop goes through the top and bottom edge of the filter and forms two half-circles at its opposing ends. The filter is of width w and length l and must have a minimum area of $300cm^2$ according to the standards. Moreover, they impose the following upper and lower bounds on the aspect ratio,

$$2 \geq \frac{l}{w} \geq 1.$$

The minimum and maximum accepted width are $10cm$ and $20cm$ and the minimum and maximum accepted length are $20cm$ and $30cm$. Suppose that each cm^2 of filter costs twice each cm of elastic band and that the company wants to minimize the production cost of the design which is equal to the price that must be paid for the required filter and elastic band per mask (manufacturing cost is negligible).

(a) Reformulate this face mask design problem as a convex optimization problem.

(b) What is the size of the filter in the optimal design for mass production?

6. *Population dynamics of COVID-19 outbreak.* In epidemiology, the next-generation matrix A for a multi-group population with n groups is defined as

$$A = D + KG$$

where D and K are nonnegative $n \times n$ diagonal matrices and G is a nonnegative $n \times n$ topology matrix whose entries G_{ij} show how much interaction group i has with group j. In other words, G is the adjacency matrix of the contact network between the groups. We assume that this contact network is strongly connected which means that if the infection starts withing any group, all the other groups will eventually get infected after long enough time.

Roughly speaking, D_{ii} shows how fast an infectious disease spreads within group i and K_{ii} shows how fast an infectious disease spreads from the other groups to group i. Hence, reducing D_{ii} is possible by reducing the number of times people within group i come into close contact with each other, and reducing K_{ii} is possible by non-pharmaceutical interventions like quarantine to isolate group i from the other groups. We want to design a strategy by determining D and K to minimize the basic reproduction number R_0, which is the largest eigenvalue of A, given G and the lower bound l on $\gamma\,\mathbf{tr}(D) + \mathbf{tr}(K)$. Conceptually, the objective function R_0 is the asymptotic growth (decay) rate of the number of infected individuals and the given constant γ is the relative socioeconomic cost of isolating people within a group in comparison to isolating groups from each other.

(a) Formulate this problem as a convex optimization problem.

(b) Solve this problem for $n = 3$, $l = \gamma = 2$, and $G = \frac{1}{2}\big(\mathbf{1}\mathbf{1}^T - I\big)$ where $\mathbf{1}$ is the all-ones vector.

7. *Social distancing.* For the next time, "22494 Convex Optimization" lectures will be hopefully in-class lectures but with physical distancing measures intended to prevent the spread of the novel coronavirus. Assume that the sides of a rectangular classroom are w and l meters long, and according to the latest guidelines of WHO the safe distance between the students should be at least 2 meters. Here, the goal is to compute the maximum number of students that fit in this classroom and plot a seating arrangement that achieves this number of seats. Suppose that you measure the distance between students in the l_p norm. For each of the following cases, provide a nontrivial bound on how many convex optimization problems are enough to solve this global optimization problem.

(a) $p = 1$

(b) $p = \infty$

Hint. This problem is NP-hard which means very hard if you don't know the exact definition. Therefore, your answer is probably exponential in some variables and you definitely can't give a polynomial bound.

Extra credit. Solve this problem (approximately) for $w = l = 7$. The method you have given above is probably too slow to return a solution in 24 hours so if you don't have a good heuristic method, don't try it! Your heuristic method should be based on convex optimization and you can't use nonconvex solvers.