Supplemental Problems
Convex Optimization (Winter 2018)

Here are a few homework problems from recent years that were not assigned this year.

Start with the optional non-required exercises in past homeworks, as well as suggested exercises from Boyd-Vandenberghe.

1 SDPs

1. Suppose we are given a set of symmetric matrices $A_0, A_1, A_2, ..., A_m$, and we want to choose a non-negative weighting $w_1, w_2, ..., w_m \geq 0$ such that the difference between the largest eigenvalue and smallest eigenvalue of $A(w) = A_0 + \sum_{i=1}^{m} w_i A_i$ is as small as possible.

   Write this as a semi-definite program, derive the dual, and write down the KKT conditions.

2. Show that linear programs are a special case of semi-definite programs. In particular, given a linear program

$$\begin{aligned} \min_x \ & c^\top x \\ \text{s.t. } & Ax \leq b \\ & x \geq 0 \end{aligned}$$

   write down an SDP in terms of A, b, c and describe how to convert a solution to this SDP into a solution for the LP.

2 Affine invariance of Newton - more detail

In this question we will formally prove the affine invariance of Newton's method. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and an affine transform $y \in \mathbb{R}^m \mapsto Ay + b$ where $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$. Define $g(y) = f(Ay + b)$.

1. For $x = Ay + b$, let $\triangle x$ and $\triangle y$ be the Newton steps for $f(x)$ and $g(y)$ respectively. Prove that $\triangle x$ = $A\triangle y$.

2. Prove that for any $t > 0$, the exit condition for backtracking linesearch on $f(x)$ in direction $\triangle x$ will hold if and only if the exit condition holds for $g(y)$ in direction $\triangle y$.

3. Consider running Newton's method on $g(\cdot)$ starting at some $y^{(0)}$ and on $f(\cdot)$ starting at $x^{(0)} = Ay^{(0)} + b$. Use the above to prove that the sequences of iterates obeys $x^{(k)} = Ay^{(k)} + b$ and $f(x^{(k)}) = g(y^{(k)})$.

4. Prove that Newton's decrement for $f(\cdot)$ at x is equal to Newton's decrement for $g(\cdot)$ at y, and so the stopping conditions are also identical.

5. Now consider a function $h(x) = cf(x)$ for some scalar $c > 0$. Prove that the Newton search directions $\triangle x$ and step sizes are the same when optimizing $h(\cdot)$ and $f(\cdot)$. Conclude that the sequence of Newton iterates is the same when optimizing $h(\cdot)$ or $f(\cdot)$.

6. Will the Newton decrement and the stopping condition also be the same for $h(\cdot)$ and $f(\cdot)$?

3 Quasi-Newton methods

In quasi-Newton methods the descent direction is given by:

$$\Delta x^{(k)} = -D^{(k)} \nabla f\left(x^{(k)}\right)$$

Let $s^{(k)} = x^{(k)} - x^{(k-1)}$ and $y^{(k)} = \nabla f(x^{(k)}) - \nabla f(x^{(k-1)})$. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method updates $D^{(k)}$ by making the smallest change, under some specific weighted norm, that agrees with the latest change in the gradient:

$$D^{(k+1)} = \underset{s^{(k)} = Dy^{(k)}}{\arg\min} \left\| W^{\frac{1}{2}} \left(D - D^{(k)}\right) W^{\frac{1}{2}} \right\|_F \tag{1}$$

where $\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2}$ is the Frobenius norm, W is any matrix such that $q^{(k)} = Wp^{(k)}$.

1. Show that the solution of equation (1) is given by:

$$D^{(k+1)} = \left(I - \rho^{(k)} y^{(k)} {s^{(k)}}^T\right) D^{(k)} \left(I - \rho^{(k)} s^{(k)} {y^{(k)}}^T\right) + \rho^{(k)} s^{(k)} {s^{(k)}}^T \tag{2}$$

   where $\rho^{(k)} = \frac{1}{\langle s^{(k)}, y^{(k)} \rangle}$ is a scalar.

2. Show that the following update is equivalent to the one in equation (2):

$$D^{(k+1)} = D^{(k-1)} + \frac{1 + \langle y^{(k)}, u^{(k)} \rangle}{\langle s^{(k)}, y^{(k)} \rangle} s^{(k)} {s^{(k)}}^T - u^{(k)} {s^{(k)}}^T - s^{(k)} {u^{(k)}}^T \tag{3}$$

   where $u^{(k)} = \frac{D^{(k-1)} y^{(k)}}{\langle s^{(k)}, y^{(k)} \rangle}$.

4 Practice KKT, primal-dual for a QP

Consider the quadratic program:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x'Hx + c'x \\ \text{s.t.} \quad & Qx \leq g \\ & Ax = b \end{aligned} \tag{QP}$$

where $H \in \mathbb{S}^n$, $Q \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}^m$, $A \in \mathbb{R}^{p\times n}$ and $b \in \mathbb{R}^p$.

1. Write down the KKT conditions for (QP), with the complimentary slackness constraint relaxed to a centering constraint with parameter t, as in the central path and primal-dual interior point methods.

2. Consider the Newton iterations when using the central path method to solve the quadratic program. Using the relaxed KKT conditions you wrote, explicitly derive the linear equations defining the Newton search direction $\triangle x$, as a function of the current iterate x and the centering parameter t. Explain how these can be obtained from the relaxed KKT conditions. Write the equations explicitly as $M \begin{bmatrix} \triangle x \\ \nu \end{bmatrix} = v$ with an explicit matrix M and vector v.

3. Now consider using the primal-dual interior point method presented in class. Write down the linear equations defining the search direction $(\triangle x, \triangle\lambda, \triangle\nu)$ as a function of the current iterate (x, λ, ν) and the centering parameter t. Write the equations explicitly as $M \begin{bmatrix} \triangle x \\ \triangle\lambda \\ \triangle\nu \end{bmatrix} = v$ with an explicit matrix M and vector v.

5 Primal-dual methods

Consider the primal-dual method presented in class and let $(x^{(k)}, \lambda^{(k)}, \nu^{(k)})$ be the iterates obtained at each iteration of the method. Recall that as in the central-path method, the primal and dual *inequality* constraints are maintained at each iteration, and we have:

$$f_i(x^{(k)}) < 0 \qquad \text{and} \qquad \lambda_i^{(k)} > 0$$

for all i at every iteration k. However, the primal *equality* constraints can be violated, allowing us to start at an infeasible starting point, and the dual objective $g(\lambda^{(k)}, \nu^{(k)})$ might be infinite, i.e. $(\lambda^{(k)}, \nu^{(k)})$ might be dual infeasible. In this question we will investigate the primal and dual feasibility of the iterates.

1. Prove that if some iterate x is primal feasible, then the next iterate x^+ will also be primal feasible. Conclude that if some iterate $x^{(k)}$ is primal feasible then all subsequent iterates $x^{(k')}$, $k' > k$, will also be primal feasible, and in particular if we start from a primal feasible point $x^{(0)}$, we never loose primal feasibility.

2. Recall that at each iteration of the primal-dual method we calculate a search direction $(\triangle x, \triangle\lambda, \triangle\nu)$ by solving the linearized KKT conditions, and then take a step with a step-size determined by backtracking line-search. Prove that regardless of the feasibility of the iterate x, if we take a step with step-size one, then the new iterate x^+ will be primal feasible (this happens if the initial step-size choice of one satisfies the backtracking line-search exit condition, as in the quadratic convergent phase of Newton's method, *and* it also satisfies the inequality constrains $f_i(x) < 0$ and $\lambda_i > 0$). That is, even if we start at an infeasible point, after taking a step with step-size one, all subsequent iterations will be primal feasible.

3. Now consider the primal-dual method applied to the quadratic programing problem (QP). Prove that after a step with step-size one is taken, and in all subsequent iterations, the dual point (λ, ν) is feasible, i.e. $g(\lambda, \nu)$ if finite. (Hint: prove that after such a step, $\nabla_x L(x, \lambda, \nu) = 0$ and use this to obtain an explicit expression for $g(\lambda, \nu)$).

4. What changes if the problem has a non-quadratic objective or non-linear constraints? Will (λ, ν) still be dual-feasible after a step with step-size one? Why or why not? If the dual-feasibility residual is zero for some iterate, $r_{\text{dual}}(x, \lambda, \nu) = 0$, will it stay zero? Why or why not?

6 Multi-Criterion Optimization

Consider the dual-criterion problem:

$$\begin{array}{rl} \text{minimize} & F_1(x), F_2(x) \\ \text{s.t.} & f_i(x) \le 0 \quad i = 1, \ldots, m, \end{array} \tag{M}$$

and for any $b \in \mathbb{R}$ consider the constraint problem:

$$\begin{array}{rl} \text{minimize} & F_1(x) \\ \text{s.t.} & f_i(x) \le 0 \quad i = 1, \ldots, m. \\ & F_2(x) \le b \end{array} \tag{M_b}$$

1. Is (M_b) always feasible?

2. Is every optimum of (M_b), for every $b \in \mathbb{R}$, Pareto optimal? Prove your answer or provide a counterexample.

3. Is every Pareto optimal point an optimum of (M_b) for some $b \in \mathbb{R}$? Prove your answer or provide a counterexample.