COL726 Assignment 5

16–30 April, 2021

Note: All answers should be accompanied by a rigorous justification, unless the question explicitly states that a justification is not necessary.

1. (a) Show that the convex hull of the unit sphere $\{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| = 1\}$ is the unit ball $\{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| \leq 1\}$.

   (b) Show that the convex hull of the set $\{\mathbf{u}\mathbf{u}^T : \mathbf{u} \in \mathbb{R}^n, \|\mathbf{u}\|_2 = 1\}$ is the set of all symmetric positive semidefinite $n \times n$ matrices with unit trace (i.e. sum of diagonal entries = 1).

2. A function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone if $(\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x}))^T(\mathbf{y} - \mathbf{x}) \geq 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Show that a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its gradient ∇f is monotone.

   Note: Do not assume that f is twice differentiable.

3. Given k points $\mathbf{p}_1, \ldots, \mathbf{p}_k \in \mathbb{R}^n$, I want to find a line segment of length $\leq \ell$ which passes as close as possible to each of them. Let us denote the line segment as $L(\mathbf{x}_1, \mathbf{x}_2)$, where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ are its endpoints.

   (a) Show that the problem

   $$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{k} \operatorname{dist}(\mathbf{p}_i, L(\mathbf{x}_1, \mathbf{x}_2)) \\ \text{subject to} \quad & \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq \ell, \end{aligned}$$

   where the optimization variables are $\mathbf{x}_1, \mathbf{x}_2$, is not a convex problem.

   Note: Minimizing over multiple variables $f(\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots)$ is equivalent to minimizing over a single vector $f([\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T, \cdots]^T)$, and convexity is with respect to this vector.

   (b) Show that the above problem is equivalent to

   $$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{k} \|\mathbf{p}_i - \theta_i \mathbf{x}_1 - (1 - \theta_i)\mathbf{x}_2\|_2 \\ \text{subject to} \quad & \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \leq \ell, \\ & 0 \leq \theta_i \leq 1 \qquad \forall i = 1, \ldots, k, \end{aligned}$$

   where the optimization variables are $\mathbf{x}_1, \mathbf{x}_2, \theta_1, \ldots, \theta_k$. Further show that this problem is convex in $\mathbf{x}_1, \mathbf{x}_2$ if we hold $\theta_1, \ldots, \theta_k$ fixed, and in $\theta_1, \ldots, \theta_k$ if we hold $\mathbf{x}_1, \mathbf{x}_2$ fixed.

4. Suppose I wish to minimize $f(\mathbf{x})$ subject to the constraint that $h(\mathbf{x}) = 0$, where both f and h are strictly convex. This is not a convex problem. However, suppose I also know that the function f has a unique global minimum at a point $\mathbf{x}_u^*$, and $h(\mathbf{x}_u^*) > 0$.

   Show that I can find a globally optimal point for the original problem, $\mathbf{x}_c^*$, by solving the convex problem of minimizing $f(\mathbf{x})$ subject to $h(\mathbf{x}) \leq 0$.

5. Consider minimizing the quadratic objective $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$, where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is positive definite, using steepest descent in the 1-norm with exact line search.

   (a) If $n = 2$, show that after the first iteration, the error decreases at a constant rate every two iterations; in particular, $\frac{\|\mathbf{x}^{(k+2)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} = \frac{p_{12}^2}{p_{11}p_{22}}$ for all $k \geq 1$.

   (b) Show that for any n, if $\mathbf{P}$ is diagonal then steepest descent in the 1-norm converges to $\mathbf{x}^*$ in only n iterations.

   Hint: Is the steepest descent method affected if you perform the change of variable $\mathbf{x} = \mathbf{z} + \mathbf{b}$ for some constant $\mathbf{b}$? What if $\mathbf{b} = -\mathbf{P}^{-1}\mathbf{q}$?

6. As mentioned in class, it is sometimes desirable to perform data using a norm other than the 2-norm. For example, let us consider minimizing $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p$ for some $p > 1$, or the equivalent problem of minimizing $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p^p = \sum_{i=1}^m |\mathbf{a}_i^T\mathbf{x} - b_i|^p$, where $\mathbf{a}_i \in \mathbb{R}^n$ are rows of $\mathbf{A}$.

   (a) Derive an expression for the gradient $\nabla f(\mathbf{x})$, or its components.

   Implement a Python function `gd(A, b, p, x0)` to find the optimal point $\mathbf{x}^*$ using gradient descent with backtracking line search. Terminate when $\frac{\|\nabla f(\mathbf{x}^{(k)})\|}{\|\nabla f(\mathbf{x}^{(0)})\|} < 10^{-6}$.

   (b) Implement either Newton's method (as a function `newton`) or BFGS (`bfgs`) to solve the same problem. Your function should use the same arguments and the same termination criterion as `gd`.

   You may use Numpy/Scipy's built-in functions for solving linear systems. For this assignment, you may perform a full $O(n^3)$ refactorization even in BFGS.

   (c) Generate some test data by choosing a degree-5 polynomial, sampling its values at many points, and adding some Gaussian noise to the values (using e.g. `numpy.random`). Solve the minimization problem for $p = 1.25, 2, 5$ using the zero polynomial as the initial guess. Plot the data, the original polynomial, and the three optimized polynomials on a single plot (similar to B&V Figure 6.5). For each value of p, make a plot of $\frac{\|\nabla f(\mathbf{x}^{(k)})\|}{\|\nabla f(\mathbf{x}^{(0)})\|}$ as a function of iteration number k for both your algorithms.

   As an additional exercise (not graded), you could try different distributions of noise other than Gaussian, and see if you can find some for which $p = 1.25$ or $p = 5$ do better than least-squares for recovering the original polynomial.

Collaboration policy: Refer to the policy on the course webpage.

If you collaborated with others to solve any question(s) of this assignment, give their names in your submission. If you found part of a solution using some online resource, give its URL.

Submission: Submit a PDF of your answers for all questions to Gradescope. Submit the code for Question 6 to Moodle. Both submissions must be uploaded before the assignment deadline.

Code submissions should contain a single .py file which contains all the necessary functions. Functions are permitted but not required to produce any side-effects like printing out values or drawing plots. Any results you are asked to show should go in the PDF.