4301 - HW3

November 2020

Problem 1: Eigenvalues

A real number $t \in \mathbb{R}$ is an eigenvalue of a symmetric matrix $A \in \mathbb{R}^{n\times n}$ if there exists a vector $x \neq 0 \in \mathbb{R}^n$ that $Ax = tx$. Here, x is called an eigenvector for the eigenvalue t. Consider the following optimization problem,

$$\max_{x\in\mathbb{R}^n} x^T Ax$$

such that

$$\|x\|_2^2 = 1$$

1. Construct a dual of this optimization problem using the method of Lagrange multipliers.

   Answer:

   First, slightly modify the primal problem as

   $$\min_{x\in\mathbb{R}^n} -x^T Ax$$

   such that

   $$x^T x - 1 = 0.$$

   Now, unsing the Lagrangian multipliers method, we obtain the objective function of the dual problem in the following.

   $$\begin{aligned} g(\lambda) &= \inf_{x\in\mathbb{R}^n} [L(x,\lambda)] \\ &= \inf_{x\in\mathbb{R}^n} \left[-x^T Ax + \lambda(x^T x - 1)\right] \\ &= \inf_{x\in\mathbb{R}^n} \left[x^T(\lambda I - A)x - \lambda\right] \\ &= \begin{cases} -\lambda & \lambda I - A \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

   Notice that if $\lambda I - A$ has even one negative eigenvalue, then $x^T(\lambda I - A)x$ can decrease unboundedly. However, if $\lambda I - A$ is positive semifedinite, then $x^T(\lambda I - A)x$ has the minimum of 0, when $x = \mathbf{0}$. So, the dual problem is

   $$\sup_{\lambda\in\mathbb{R}} -\lambda$$

   such that

   $$\lambda I - A \succeq 0.$$

2. Argue that strong duality holds regardless of whether or not the problem above is convex. What is the solution to the primal problem?

Answer:

Take Λ as a diagonal matrix that is formed with the eigenvalues of A. In addition, it is easy to show that if Λ contains eigenvalues of A then $\lambda I - \Lambda$ contains eigenvalues of $\lambda I - A$. As a result, we reduce the constraint of the dual problem to $\lambda I - \Lambda \succeq 0$, which holds only if λ is larger that the largest eigenvalue of A. So, the final answer of the optimization is exactly the minus largest eigenvalue of matrix A. This shows that the modified primal and its dual problem both have the same final answer and hence the duality gap is zero.

3. Use the method of Lagrange multipliers (for positive semidefinite constraints) to construct a dual of your dual.

Answer:

Modify the dual problem as

$$\inf_{\lambda \in \mathbb{R}} \lambda$$

such that

$$\Lambda - \lambda I \preceq 0.$$

So,

$$\begin{aligned} h(y) &= \inf_{\lambda \in \mathbb{R}} \left[L(\lambda, \nu) \right] \\ &= \inf_{\lambda \in \mathbb{R}} \left[\lambda + \sum_{i=1}^{n} \nu_i (\Lambda_{i,i} - \lambda) \right] \end{aligned}$$

such that

$$\nu_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}.$$

As always, take the derivative of the Lagrangian. Put it equal to zero. And get rid of the primal variable.

$$\frac{\partial}{\partial \lambda} \left(\lambda + \sum_{i=1}^{n} \nu_i (\Lambda_{i,i} - \lambda) \right) = 0 \rightarrow \sum_{i=1}^{n} \nu_i = 1 \rightarrow \lambda - \sum_{i=1}^{n} \nu_i \lambda + \sum_{i=1}^{n} \nu_i \Lambda_{i,i} = \sum_{i=1}^{n} \nu_i \Lambda_{i,i}.$$

As a result, the dual of dual is

$$\sup \sum_{i=1}^{n} \nu_i \Lambda_{i,i}$$

such that

$$\sum_{i=1}^{n} \nu_i = 1,$$

$$\nu_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}.$$

To see how this is equivalent the primal problem, notice that $\sum_{i=1}^{n} \nu_i \Lambda_{i,i}$ is a convex combination of all the eigenvalues. So its maximum value is trivially the maximum eigenvalue. Lets rewrite the dual of dual as in the following:

$$\max_{i \in \{1, \ldots, n\}} \Lambda_{i,i}$$

This problem does not have any constraint, because all the ν_is have been eliminated analytically.

Problem 2: Interior Point Methods

Interior point methods take optimization problems of the following form

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$Ax = b,$$

$$g_k(x) \leq 0 \text{ for all } k \in \{1, \ldots, K\}.$$

And approximate them using some $\delta > 0$ as

$$\min_{x \in \mathbb{R}^n \text{ s.t. } Ax=b} \left[\delta f_0(x) - \sum_{k=1}^{K} \log(-g_k(x)) \right].$$

As $\delta \to \infty$, this approximation becomes better and better. Given an initial $x^{(0)}, \delta^{(0)} > 0, \rho > 1$, the interior point method then iterates the following starting at $t = 0$.

- Find an $x^* \in \operatorname{argmin}_{x \in \mathbb{R}^n \text{ s.t. } Ax=b} \left[\delta^{(t)} f_0(x) - \sum_{k=1}^{K} \log(-g_k(x)) \right]$ by using Newton's method starting at $x^{(t)}$.
- Set $x^{(t+1)} = x^*$.
- Update $\delta^{(t+1)} = \rho \delta^{(t)}$.

Recall the problem of computing the projection of a query point q onto the convex hull of the given $x^{(1)}, \ldots, x^{(M)} \in \mathbb{R}^n$ from Homework 2.

$$\min_{\lambda \in \mathbb{R}^M} \frac{1}{2} \|q - \sum_{m=1}^{M} \lambda_m x^{(m)}\|_2^2$$

such that

$$\sum_{m=1}^{M} \lambda_m = 1,$$

$$\lambda \succeq 0.$$

Our Notation:

Following the above notation, we have $f_0(\lambda) = \frac{1}{2}\|q - \sum_{m=1}^{M} \lambda_m x^{(m)}\|_2^2$, $g_k(\lambda) = -\lambda_k$, $A = \mathbf{1}$, and $b = 1$. Given δ, the objective is to find λ that minimizes the following problem.

$$\min_{\lambda \in \mathbb{R}^n \text{ s.t. } A\lambda=b} \left[\frac{\delta}{2} \|q - \sum_{m=1}^{M} \lambda_m x^{(m)}\|_2^2 - \sum_{m=1}^{M} \log(\lambda_m) \right]$$

such that

$$\sum_{m=1}^{M} \lambda_m = 1.$$

1. Explain why the subproblem solved by the interior point method is a convex optimization problem.

   Answer:

   Since $-\log$ and $\|\cdot\|_2^2$ are both convex, the whole objective function is the summation of convex functions, hence convex.

2. Colab

3. How should you assess the convergence of the interior point method? In particular, what would be a good stopping criteria?

   Answer:

   The answers may vary. Two possible answers could be $\dfrac{\sum_{k=1}^{K} \log(-g_k(x))}{\delta f_0(x)} \leq \epsilon$ or $\|\lambda^{\text{new}} - \lambda^{\text{old}}\| \leq \epsilon$, for some $\epsilon > 0$.

Problem 3: Power Iteration

The power iteration method attempts to find the eigenvalue of maximum magnitude a matrix $A \in \mathbb{R}^{n \times n}$ by iterating the following steps:

- Pick an initial $x^{(0)} \in \mathbb{R}^n$.
- Repeat until convergence:
  - $x^{(t+1)} = \frac{Ax^{(t)}}{\|Ax^{(t)}\|_2}$.

1. Colab

2. How should you pick the initial vector? Explain.

   Answer:

   Consider the eigendecomposition of $A = U\Lambda U^T$. Since at each iteration, we multiply $x^{(t)}$ to A, at iteration $t+1$ we have

$$\begin{aligned} x^{(t+1)} &= \frac{Ax^{(t)}}{\|Ax^{(t)}\|_2} = \frac{A^2x^{(t-1)}}{\|A^2x^{(t-1)}\|_2} = \cdots = \frac{A^{t+1}x^{(0)}}{A^{t+1}x^{(0)}} \\ &= \frac{(U\lambda U^T)^{t+1}x^{(0)}}{\|(U\lambda U^T)^{t+1}x^{(0)}\|_2} \\ &= \frac{U\Lambda^{t+1}U^Tx^{(0)}}{\|U\Lambda^{t+1}Ux^{(0)}\|} \end{aligned}$$

   So, if $x^{(0)}$ is orthogonal to the eigevnvector corresponding to the maximum eigenvalue, denoted as $\nu_{\max}$, then $x^{(t)}$ keeps being orthogonal to $\nu_{\max}$ even if $t \to \infty$. There is no other concern regarding how to pick the initial point.

   One approach to guarantee the desirable result could be taking n mutually orthogonal vectors, and keep repeating the whole process by taking one vector as the initial point at a time. The maximum obtained answer is the maximum eigenvalue.

3. **Assume that A is symmetric**. Explain how to use the same approach to find the eigenvalue with second largest magnitude of a matrix by only changing the way in which you pick the initial vector.

   Answer:

   Matrix $A \in \mathbb{R}^{n \times n}$ as a symmetric real matrix has n orthogonal eigenvectors. So by following the process explained in the previous part 2, all the eigenvalues of the matrix A are among those n different outputs. Which makes the second largest eigenvalue the second largest output.

4. Is it possible that the power iteration method is ill-defined, i.e., can the denominator in the update become zero?

   Answer:

   If A has a zero eigenvalue, then there is a possible scenario that we mistakenly pick $x^{(0)}$ equal to the eigenvector corresponding to the zero eigenvalue. In this very special case $\|Ax^{(0)}\|_2 = 0$. Other than that, if $Ax^{(t)} \neq \mathbf{0}$ then it is guaranteed that $Ax^{(t+1)} \neq \mathbf{0}$. (Note that $\|x\|_2 = 0$ if and only if $x = \mathbf{0}$.)