PURDUE UNIVERSITY · CS 52000
COMPUTATIONAL OPTIMIZATION

HOMEWORK
David F. Gleich
February 5, 2023

Please answer the following questions in complete sentences in a clearly prepared manuscript and submit the solution by the due date on Gradescope.

Remember that this is a graduate class. There may be elements of the problem statements that require you to fill in appropriate assumptions. You are also responsible for determining what evidence to include. An answer alone is rarely sufficient, but neither is an overly verbose description required. Use your judgement to focus your discussion on the most interesting pieces. The answer to "should I include 'something' in my solution?" will almost always be: Yes, if you think it helps support your answer.

Problem 0: Homework checklist

- Please identify anyone, whether or not they are in the class, with whom you discussed your homework. This problem is worth 1 point, but on a multiplicative scale.
- Make sure you have included your source-code and prepared your solution according to the most recent Piazza note on homework submissions.

Problem 1: Convexity and least squares

1. Show that $f(x) = \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|^2$ is a convex function. Feel free to use the result proved on the last homework.

   First, we can expand $f(x) = \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|^2$ as: $f(x) = \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|^2 = (\mathbf{b} - \boldsymbol{A}\mathbf{x})^T(\mathbf{b} - \boldsymbol{A}\mathbf{x}) = \mathbf{b}^T\mathbf{b} - 2\mathbf{b}^T\boldsymbol{A}\mathbf{x} + \mathbf{x}^T\boldsymbol{A}^T\boldsymbol{A}\mathbf{x}$, which is a combination of a constant, a linear term and a quadratic term. For the linear term and constant term $l(\mathbf{x}) = \mathbf{b}^T\mathbf{b} - 2\mathbf{b}^T\boldsymbol{A}(x)$, we can easily prove that $l(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) = \mathbf{b}^T\mathbf{b} - 2\mathbf{b}^T\boldsymbol{A}(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) = \alpha(\mathbf{b}^T\mathbf{b} - 2\mathbf{b}^T\boldsymbol{A}\mathbf{y}) + (1-\alpha)(\mathbf{b}^T\mathbf{b} - 2\mathbf{b}^T\boldsymbol{A}x) = \alpha l(\mathbf{y}) + (1-\alpha)l(\mathbf{x})$ holds for any $\mathbf{x}, \mathbf{y}$ and $\alpha \in (0, 1)$.

   As for the quadratic term, it is obvious that $\boldsymbol{A}^T\boldsymbol{A}$ is positive semi-definite since for any vector x, $q(\mathbf{x}) = \mathbf{x}^T\boldsymbol{A}^T\boldsymbol{A}\mathbf{x} = \|\boldsymbol{A}\mathbf{x}\|^2 \geq 0$. And we can easily prove that $q(\mathbf{x}) = \mathbf{x}^T\boldsymbol{A}^T\boldsymbol{A}\mathbf{x}$ is convex by the conclusion from the last homework.

   Therefore,

   $$\begin{aligned} &f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) = l(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) + q(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) \\ &\leq \alpha l(\mathbf{y}) + (1-\alpha)l(\mathbf{x}) + \alpha q(\mathbf{y}) + (1-\alpha)q(\mathbf{x}) \\ &= \alpha f(\mathbf{y}) + (1-\alpha)f(\mathbf{x}) \end{aligned} \tag{1}$$

   Thus we have shown that $f(x)$ is convex.

2. Show that the null-space of a matrix is a convex set. (A convex set satisfies the condition that, for every pair of points in the set, any point on the line joining those points is also in the set.)

   Since the null space for matrix $\boldsymbol{A}$ contains all and only the $\mathbf{x}$ such that $\mathrm{N}(\boldsymbol{A}) = \{\mathbf{x} | \boldsymbol{A}\mathbf{x} = 0\}$. For any $\mathbf{x}, \mathbf{y} \in \mathrm{N}(\boldsymbol{A})$, we have $\boldsymbol{A}\mathbf{x} = \boldsymbol{A}\mathbf{y} = 0$. Thus, for any $\alpha \in (0, 1)$, $\boldsymbol{A}(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) = \alpha\boldsymbol{A}\mathbf{y} + (1-\alpha)\boldsymbol{A}\mathbf{x} = 0$ holds. Thus $\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})$ is also in $\mathrm{N}(\boldsymbol{A})$. Therefore we have shown that $\mathrm{N}(\boldsymbol{A})$ is convex set.

Problem 2: Ridge Regression

The Ridge Regression problem is a variant of least squares:

$$\text{minimize} \quad \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2.$$

This is also known as Tikhonov regularization.

1. Show that this problem always has a unique solution, for *any* $\boldsymbol{A}$ if $\lambda > 0$, using the theory discussed in class so far.

    > This property is one aspect of the reason that Ridge Regression and used. It is also a common regularization method that can help avoid overfitting in a regression problem.

    First, it is obvious that $f(x)\|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$ is convex since $\|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2$ is convex(from Problem 1.1) and $\lambda\|\mathbf{x}\|_2^2$ is also strictly convex($\lambda\|\mathbf{x}\|_2^2 = \mathbf{x}^T(\lambda\boldsymbol{I})\mathbf{x}$, $\lambda\boldsymbol{I}$ is positive definite).

    Now we will show that $f'(\mathbf{x}) = 0$ always has a solution, then we can sufficiently prove that

    $$\text{minimize} \quad \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2.$$

    always have a unique solution.

    Thus

    $$\begin{aligned} &f'(\mathbf{x}) = -2A^T(\mathbf{b} - \boldsymbol{A}\mathbf{x}) + 2\lambda\mathbf{x} = 0 \\ &\Rightarrow (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})\mathbf{x} = \boldsymbol{A}^T\mathbf{b} \end{aligned} \tag{2}$$

    Since $\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I}$ is positive definite and has full rank, there always exists a unique solution for eq. 2 which is $\mathbf{x}^* = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\mathbf{b}$

    Therefore we have shown that $\mathbf{x}^*$ is the unique solution to ridge regression.

2. Use the SVD of $\boldsymbol{A}$ to characterize the solution as a function of λ.

    For any $\boldsymbol{A}$ that is a $m \times n$ matrix, the SVD decomposition on A can be written as: $\boldsymbol{A} = U\Sigma V^T$ where $\boldsymbol{U}$ is a $m \times n$ unitary matrix, Σ is a $n \times n$ diagonal matrix and $\boldsymbol{V}$ is a $n \times n$ unitary matrix.

    Then we can expand the solution from Problem 2.1 as:

    $$\begin{aligned} \mathbf{x}^* &= (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\mathbf{b} \\ &= (\boldsymbol{V}\Sigma\boldsymbol{U}^T\boldsymbol{U}\Sigma\boldsymbol{V}^T + \lambda\boldsymbol{I})^{-1}\boldsymbol{V}\Sigma\boldsymbol{U}^T\mathbf{b} \\ &= (\boldsymbol{V}\Sigma^2\boldsymbol{V}^T + \lambda\boldsymbol{I})^{-1}\boldsymbol{V}\Sigma\boldsymbol{U}^T\mathbf{b} \\ &= [\boldsymbol{V}(\Sigma^2 + \lambda\boldsymbol{I})\boldsymbol{V}^T]^{-1}\boldsymbol{V}\Sigma\boldsymbol{U}^T\mathbf{b} \\ &= \boldsymbol{V}(\Sigma^2 + \lambda\boldsymbol{I})^{-1}\boldsymbol{V}^T\boldsymbol{V}\Sigma\boldsymbol{U}^T\mathbf{b} \\ &= \boldsymbol{V}(\Sigma^2 + \lambda\boldsymbol{I})^{-1}\Sigma\boldsymbol{U}^T\mathbf{b} \end{aligned} \tag{3}$$

    Let's assume that $\boldsymbol{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n), \boldsymbol{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n), \Sigma = \text{diag}\{s_1, \ldots, s_n\}$ and $\mathbf{b} = (b_1, \ldots, b_m)^T$, we can express equation (2) as $\mathbf{x}^* = \sum_{i=1}^{q} \frac{s_i}{s_i^2+\lambda}\mathbf{v}_i\mathbf{u}_i^T b_i$

3. What is the solution when $\lambda \to \infty$?
    What is the solution when $\lambda \to 0$.

    When $\lambda \to \infty$, $lim_{\lambda\to\infty}\mathbf{x}^* = \sum_{i=1}^{q} \frac{s_i}{s_i^2+\lambda}\mathbf{v}_i\mathbf{u}_i^T b_i \to 0$

When $\lambda \rightarrow 0$, $lim_{\lambda\rightarrow 0}\mathbf{x}^* = \sum_{i=1}^{q} \frac{s_i}{s_i^2+\lambda}\mathbf{v}_i\mathbf{u}_i^T b_i \rightarrow \sum_{i=1}^{q} s_i^{-1} v_i u_i^T b_i = (A^T A)^{-1} A^T b$

4. Give the solutions of the least squares problem using the sports teams data when $\lambda = 0$ and $\lambda = \infty$. (Here, you will not need the constraint that the sum of entries is 1.) Any notable differences from the ranking giving in class?

5. Suppose that you only want to regularize one component of the solution, say, x_1, so that your optimization problem is

$$\text{minimize} \quad \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2 + \lambda x_1^2.$$

Show how to adapt your techniques in this problem to accomplish this goal.

Problem 3: Thinking about constraints

Find the minimizer of $f(x, y) = x^2 + 2y^2$ where also $y = 5x + 2$ (that is, we have added one linear constraint) and justify, as concisely and correctly as you can, that you have found a solution. Do the terms local and global minimizer apply? If so, use one of them to describe your solution. Furthermore, evaluate the gradient of f at your solution. What is notable about the gradient at your solution in comparison with the solution where x, y have no constraints?

First, let's substitute $y = 5x + 2$ into the function $f(x, y)$, we have

$$f(x, y) = x^2 + 2(5x + 2)^2 = 51x^2 + 40x + 8 = 51\left(x + \frac{20}{51}\right)^2 + \frac{8}{51} \tag{4}$$

It is clear that the function has a global minimum at $\left(-\frac{20}{51}, \frac{2}{51}\right)$, with a minimum value of $8/51$. It is a global minimum since $f(x, y)$ is convex, and any notable minimum would be the global minimum.

The gradient at $\left(-\frac{20}{51}, \frac{2}{51}\right)$ is $\left(-\frac{40}{51}, \frac{8}{51}\right)$.

However, if no constraint is impulsed, the minimizer would be $(0, 0)$ with gradient of $(0, 0)$. Therefore, it can be concluded that the optimality of the minimizer might be hampered when constrains are impulsed, resulting in a gradient that may not be zero.

Problem 4: Alternate formulations of Least Squares

Consider the constrained least squares problem:

$$\begin{array}{ll} \underset{\mathbf{r},\mathbf{y}}{\text{minimize}} & \|\mathbf{r}\|_2 \\ \text{subject to} & \mathbf{r} = \mathbf{b} - \boldsymbol{C}\mathbf{y} \end{array}$$

where $\boldsymbol{C} \in \mathbb{R}^{m\times n}$, $n \leq m$ and rank n.

1. Convert this problem into the standard constrained least squares form.

   Reformulate the optimization problem as a standard constrained least squares form: $\min_y \frac{1}{2}\|b - Cy\|_2^2 \quad \text{s.t. } Cy = b - r$

2. Form the augmented system from the Lagrangian as we did in class.

   The unconstrained Lagrangian of the original problem can be written as:

$$\mathcal{L}(y; \lambda) = \frac{1}{2}\|b - Cy\|_2^2 - \lambda^T(Cy - b + r)$$

Take the derivative of $\mathcal{L}(y;\lambda)$, we have

$$\frac{\partial \mathcal{L}}{\partial y} = -C^T(b - Cy) - C^T\lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = Cy + r - b = 0$$

thus the augmented system is

$$\begin{pmatrix} C^TC & C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} y \\ -\lambda \end{pmatrix} = \begin{pmatrix} C^Tb \\ b - r \end{pmatrix}$$

Since C^TC is invertible. We can write the augmented matrix of the linear system as:

$$A = \begin{pmatrix} C^TC & C^T & C^Tb \\ C & 0 & b - r \end{pmatrix}$$

.

3. Manipulate this problem to arrive at the normal equations for a least-squares problem: $\boldsymbol{C}^T\boldsymbol{C}\mathbf{y} = \boldsymbol{C}^T\mathbf{b}$.
   Discuss any advantages of the systems at intermediate steps.