PURDUE UNIVERSITY · CS 52000
COMPUTATIONAL OPTIMIZATION

HOMEWORK 3
Mathieu Gaillard
February 9, 2019

Problem 0: Homework checklist

I was out of the university until Saturday, I could not collaborate with anybody. I mainly used the class material, Google, Wikipedia and my previous knowledge.

Problem 1: Convexity and least squares

Question 1:

Show that $f(x) = \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|^2$ is a convex function.

We write f in matrix form:

$$\begin{aligned} f(x) &= (b - Ax)^T(b - Ax) \\ &= b^T b - b^T Ax - (Ax)^T b + x^T A^T Ax \\ &= b^T b - 2b^T Ax + x^T A^T Ax \end{aligned}$$

In the previous equation, we notice that $b^T Ax = (Ax)^T b$. Both these terms are the scalar product between the vectors b and Ax. Because the scalar product is commutative, they are equal.

We already know that the sum of two convex functions is convex. Because $b^T b - 2b^T Ax$ is an affine transformation, it is simultaneously convex and concave. As a consequence, we just need to prove that $x^T A^T Ax$ is convex.

Through $x^T A^T Ax = (Ax)^T(Ax) = \|Ax\|^2 \geq 0$, we know that $A^T A$ is positive semi-definite. It is also symmetric because $(A^T A)^T = A^T(A^T)^T = A^T A$. Since in the last homework we proved that it is convex, f is convex.

Q.E.D.

Another way of proving this result is to use the fact that every norm is convex, because x^2 is also convex and monotonically increasing on $\mathcal{R}^+$, f is convex.

Question 2:

Show that the null-space of a matrix is a convex set. For every pair of points in the set, any point on the line joining those points is also in the set.

$$ker(A) = \{x \in \mathcal{R}^n | Ax = 0\}$$

Let $\alpha \in ker(A)$ and $\beta \in ker(A)$, the points on the line joining those points are:

$$\forall t \in [0, 1], t\alpha + (1 - t)\beta$$

Because $A\alpha = 0$ and $A\beta = 0$:

$$A(t\alpha + (1-t)\beta) = tA\alpha + (1-t)A\beta = 0$$

Q.E.D.

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.

Question 1:

Show that this problem always has a unique solution, for *any* $\boldsymbol{A}$ if $\lambda > 0$.

$$f(\mathbf{x}) = \|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$$

We know that if a function is convex, any local minimizer is a global minimizer. If in addition the function is differentiable, then any stationary point is a global minimizer. Thus, we first show that f is convex and then we show that there is a unique stationary point.

We already know that $\|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2$ is convex and that the sum of two convex functions is also convex. So, we just need to show that $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_2^2$ is convex if $\lambda > 0$.

We want to prove:

$$\forall t \in [0,1], \forall(\mathbf{x},\mathbf{y}) \in \mathcal{R}^n, g(t\mathbf{x} + (1-t)\mathbf{y}) - tg(\mathbf{x}) - (1-t)g(\mathbf{y}) \leq 0$$

$$\begin{aligned}
&g(t\mathbf{x} + (1-t)\mathbf{y}) - tg(\mathbf{x}) - (1-t)g(\mathbf{y}) \\
&\quad = \lambda(t\mathbf{x} + (1-t)\mathbf{y})^T(t\mathbf{x} + (1-t)\mathbf{y}) - t\lambda\mathbf{x}^T\mathbf{x} - \lambda(1-t)\mathbf{y}^T\mathbf{y} \\
&\quad = \lambda\left(t^2\mathbf{x}^T\mathbf{x} + (1-t)^2\mathbf{y}^T\mathbf{y} + 2t(1-t)\mathbf{x}^T\mathbf{y} - t\mathbf{x}^T\mathbf{x} - (1-t)\mathbf{y}^T\mathbf{y}\right) \\
&\quad = \lambda t(1-t)\left(2\mathbf{x}^T\mathbf{y} - \mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{y}\right) \\
&\quad = -\lambda t(1-t)\|\mathbf{x} - \mathbf{y}\|^2
\end{aligned}$$

This last expression is negative only if $\lambda > 0$.

$$\begin{aligned}
\lambda > 0 &\implies -\lambda t(1-t)\|\mathbf{x} - \mathbf{y}\|^2 \leq 0 \\
&\implies f \text{ is convex}
\end{aligned}$$

We now know that f is convex. We compute its gradient:

$$\nabla f(\mathbf{x}) = 2\left(A^T(A\mathbf{x} - \mathbf{b}) + \lambda\mathbf{x}\right)$$

We look for stationary points:

$$\nabla f(\mathbf{x}) = 0 \iff \left(A^T(A\mathbf{x} - \mathbf{b}) + \lambda \mathbf{x}\right) = 0$$

$$\iff \left(A^T A + \lambda I\right)\mathbf{x} = A^T \mathbf{b}$$

This last expression is another least square problem, which has a unique solution. $(A^T A + \lambda I)$ is square and invertible since we assume A has independent columns. So, f has a unique stationary point and is convex, therefore it has a unique global minimizer. **Q.E.D.**

Question 2:

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

To get the solution we need to find the stationary point.

$$\nabla f(\mathbf{x}) = 0 \iff \mathbf{x} = \left(A^T A + \lambda I\right)^{-1} A^T b$$

With the SVD ($A = v\Sigma u^T$), we have:

$$\begin{aligned}
\mathbf{x} &= \left(A^T A + \lambda I\right)^{-1} A^T b \\
&= \left(v\Sigma^2 v^T + \lambda I\right)^{-1} v\Sigma^T u^T b \\
&= \left(v\Sigma^2 v^T + \lambda v v^T\right)^{-1} v\Sigma^T u^T b \\
&= v\left(\Sigma^2 v^T + \lambda I\right)^{-1} v^T v\Sigma^T u^T b \\
&= v\left(\Sigma^2 v^T + \lambda I\right)^{-1} \Sigma^T u^T b
\end{aligned}$$

It would look like:

$$\mathbf{x} = \sum \mathbf{v}_i \frac{\sigma_i}{\sigma_i^2 + \lambda} \mathbf{u}_i^T \mathbf{b}_i$$

Question 3:

When $\lambda \to \infty$, the solution is $\mathbf{x} \to 0$. The regularization is too high and the only solution is a vector whose norm is very small. When $\lambda \to 0$, the solution is $A^{-1}b$, which is the same solution as if there were not regularization.

Question 4:

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.

We add a Tikhonov matrix Γ so that the problem becomes:

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

The matrix contains only a 1 on the top left corner so that $\Gamma\mathbf{x} = x_1$.

$$\Gamma = \begin{bmatrix} 1 & 0 & \dots \\ 0 & 0 & \\ \vdots & & \ddots \end{bmatrix}$$

Then the solution becomes:

$$\begin{aligned} \mathbf{x} &= \left(A^T A + \lambda \Gamma^T \Gamma\right)^{-1} A^T b \\ &= v\left(\Sigma^2 v^T + \lambda \Gamma^T \Gamma\right)^{-1} \Sigma^T u^T b \end{aligned}$$

In particular the diagonal matrix in the middle would be:

$$\left(\Sigma^2 v^T + \lambda \Gamma^T \Gamma\right)^{-1} \Sigma^T = \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\lambda} & 0 & \dots & \\ 0 & \frac{1}{\sigma_2} & & \\ \vdots & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix}$$

When $\lambda \to \infty$, the optimization favors a solution where $x_1 \to 0$.

Problem 3: Alternate formulations of Least Squares

Consider the constrained least squares problem:

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

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

Question 1:

Convert this problem into the standard constrained least squares form.

The constraint is under-determined, so the constraint leads to infinite solutions.
We rewrite the constraint:

$$\mathbf{r} = \mathbf{b} - \boldsymbol{C}\mathbf{y} \iff \boldsymbol{C}\mathbf{y} + \mathbf{r} - \mathbf{b} = 0$$

We concatenate the vectors $\mathbf{y}$ and $\mathbf{r}$:

$$\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{r} \end{bmatrix}$$

The constraint becomes:

$$\boldsymbol{C}\mathbf{y} + \mathbf{r} - \mathbf{b} = 0 \iff \boldsymbol{D}\mathbf{x} - \mathbf{b} = 0$$

With

$$\boldsymbol{D} = \begin{bmatrix} C & I_m \end{bmatrix}$$

The term to optimize becomes:

$$\mathbf{r} = \boldsymbol{A}\mathbf{x}$$

With

$$\boldsymbol{A} = \begin{bmatrix} 0_n & I_m \end{bmatrix}$$

The problem is now:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \frac{1}{2}\|\boldsymbol{A}\mathbf{x}\|^2 \\ \text{subject to} & \boldsymbol{D}\mathbf{x} = \mathbf{b} \end{array}$$

Note that we add a $\frac{1}{2}$ in the function to minimize. It is equivalent and it simplifies the

Question 2:

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

Sorry, I couldn't attend class this week, and the videos are not online, here is my attempt.

$$\begin{aligned} (L)(\mathbf{x}, \lambda) &= \frac{1}{2}\|\boldsymbol{A}\mathbf{x}\|^2 + \lambda^T(\boldsymbol{D}\mathbf{x} - \mathbf{b}) \\ &= \mathbf{x}^T\boldsymbol{A}^T\boldsymbol{A}\mathbf{x} + \lambda^T(\boldsymbol{D}\mathbf{x} - \mathbf{b}) \end{aligned}$$

Optimality conditions are:

$$\nabla_x(L)(\mathbf{x}, \lambda) = \boldsymbol{A}^T\boldsymbol{A}\mathbf{x} - \boldsymbol{D}^T\lambda = 0$$

$$\nabla_\lambda(L)(\mathbf{x}, \lambda) = \boldsymbol{D}\mathbf{x} - \mathbf{b} = 0$$

And the augmented system is:

$$\begin{bmatrix} \boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{D}^T \\ \boldsymbol{D} & 0 \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ \lambda \end{pmatrix} = \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix} \quad \Longleftrightarrow \quad \begin{bmatrix} 0 & 0 & C^T \\ 0 & I_m & I_m \\ C & I_m & 0 \end{bmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{r} \\ \lambda \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{b} \end{bmatrix}$$

Question 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.