Computer Science Department, Ben Gurion University of the Negev

Optimization methods with Applications - Spring 2023

Home assignment 1

Due on 27/4/2023.

1. **Vector norms, matrix norms, and inner products (Chapter 2 in the NLA booklet).**

   (a) We have learned that for any matrix $A \in \mathbb{R}^{m\times n}$, we have the induced norms

   $$\|A\|_1 = \max_{\mathbf{x}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = \max_j \left(\sum_{i=1}^{m} |a_{i,j}| \right),$$

   and similarly

   $$\|A\|_\infty = \max_{\mathbf{x}} \frac{\|A\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_i \left(\sum_{j=1}^{n} |a_{i,j}| \right).$$

   For the matrix

   $$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & -4 & 8 \\ -5 & 4 & 1 & 5 \\ 5 & 0 & -3 & -7 \end{bmatrix}$$

   Explicitly find a vector $\mathbf{x}$ that maximizes the terms above and fulfils the definitions. Find one for ℓ_1 and one for ℓ_∞. No need to prove your solution for $\mathbf{x}$ - just write an intuitive explanation.

   (b) Repeat the previous section for the ℓ_2 norm:

   $$\|A\|_2 = \max_{\mathbf{x}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \sigma_{max}.$$

   where σ_{max} is the largest singular value of A (singular values are the square root of the eigenvalues of $A^T A$). Repeat the previous question for this norm as well (find $\mathbf{x}$ that maximizes the expression). It is recommended to use a computer program here, and submit the code.

2. **The existence of solution for LS**

   (a) Recall the definition of the null-space: $null(A) = \{\mathbf{x} : A\mathbf{x} = 0\}$. Show that $null(A^T A) = null(A)$. Hint: $A^T A\mathbf{x} = 0 \Rightarrow \mathbf{x}^T A^T A\mathbf{x} = 0$.

   (b) Recall the definition of range: $range(A) = \{\mathbf{y} : A\mathbf{x} = \mathbf{y}\}$. Show that $range(A^T A) = range(A^T)$.
   Hint: You may have learned that for every matrix B, $range(B^T) = null(B)^\perp$ where $^\perp$ denotes the complement of the subspace. In other words: $range(B^T) \cup null(B) = \mathbb{R}^n$ and $range(B^T) \cap null(B) = \{0\}$. Use this for $B = A^T A$ together with the previous section.

   (c) Using the previous section, conclude that the normal equation is always consistent - i.e. there is always at least one solution to the Least Squares problem.

3. **Minimization with different norms lead to different answer**
   (Q3 is not for submission - was in the IDS course). Others are encouraged to think about it. We are trying to approximate a vector $[x_1, x_2, x_3]$ by a constant c using ℓ_p norms. Assume $x_1 < x_2 < x_3$. Find the best approximation of the vector using a constant c in the following norms:

   (a) ℓ_2 norm (Least squares): $\min_c\{(c - x_1)^2 + (c - x_2)^2 + (c - x_3)^2\}$.

   (b) ℓ_∞ norm: $\min_c\{\max(|c - x_1|, |c - x_2|, |c - x_3|)\}$.

   (c) ℓ_1 norm: $\min_c\{|c - x_1| + |c - x_2| + |c - x_3|\}$.

   Clarification: the x_i's are given, and you need to find c.
   Hint for (b) and (c): the solution is obtained by logic, not by calculations as in (a).

4. **Least Squares**

   (a) First, as a preliminary task, write two programs, `BwdSub(U,b)` and `FwdSub(L,b)`, that solve upper and lower triangular linear systems, respectively.

   (b) Find the best approximation in a least square sense for satisfying the linear system $A\mathbf{x} = \mathbf{b}$ where

   $$A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \quad , \mathbf{b} = \begin{bmatrix} 6 \\ 1 \\ 5 \\ 2 \end{bmatrix}. \tag{1}$$

   Write the normal equations and solve the problem using the Cholesky factorization of $A^\top A$. You may use a computer and built-in functions for computing the Cholesky factorization, and provide the code. Use your functions for the forward and backward substitutions.

(c) Solve the least squares problem in subsection 4b using the (a) QR factorization of A, and (b) SVD factorization of A. Verify that you get the same answers. You may use a computer and built-in functions for QR, SVD, but not for the solution of the system given the factorizations.

(d) Compute the residual of the least squares system $\mathbf{r} = A\mathbf{x} - \mathbf{b}$, with $\mathbf{x}$ that you found in the previous section. Show that $A^\top \mathbf{r} = 0$. Is that surprising?
(Comment: numerically, the result will not really be zero because of roundoff errors, but it will be close to zero).

(e) Find the least squares solution of the system in Eq. (1), but now find a solution for which the first equation is almost exactly satisfied satisfied (let's say, such that $|r_1| < 10^{-3}$). Hint: use weighted least squares. You may solve the problem anyway you want.

5. QR **factorization**

(a) Write two programs for getting the QR factorization using Gram Schmidt and Modified Gram Schmidt as presented in class.

(b) Using the two codes you programmed, find the QR factorization of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{bmatrix}$$

for $\epsilon = 1, 1\text{e-}10$, for both algorithms (that is, perform four factorizations).

(c) For all factorizations, calculate $\|Q^\top Q - I\|_F$. Which of the algorithms produces a better QR factorization? Explain.

6. **Regularized Least Squares and SVD** (Sections 5.4 - 5.5 in NLA.pdf.)

For this question we assume that we have a full rank matrix $A \in \mathbb{R}^{m\times n}$ for which we have an SVD factorization $A = U\Sigma V^\top$, and denote the singular triplets as $(\mathbf{u}_i, \sigma_i, \mathbf{v}_i)$.

We saw in class, that using the SVD, the solution of the LS problem $\arg\min_{\mathbf{x}}\{\|A\mathbf{x} - \mathbf{b}\|_2^2\}$, is given by

$$\hat{\mathbf{x}} = (A^\top A)^{-1}A^\top \mathbf{b} = V\Sigma^{-1}U^\top \mathbf{b}.$$

(a) (Not for submission) Show that $V\Sigma^{-1}U^\top = \sum_{i=1}^{\min(m,n)} \sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i^\top$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of U and V respectively.
Guidance: this equality holds simply from the definition of matrix multiplication. Show that the (i, j) entry of both sides are identical.

(b) Using the information above show:

$$\hat{\mathbf{x}} = \sum_{i=1}^{\min(m,n)} \frac{1}{\sigma_i}(\mathbf{u}_i^\top \mathbf{b})\mathbf{v}_i.$$

(c) Consider a full SVD, where $U \in \mathbb{R}^{m\times m}$ so we have m vectors $\mathbf{u}_i$ that span the whole R^m space. Assume that $\mathbf{b} = \sum_{i=1,..,m} \alpha_i \mathbf{u}_i$ for some α_i. Show that

$$\hat{\mathbf{x}} = \sum_{i=1}^{\min(m,n)} \frac{1}{\sigma_i}\alpha_i \mathbf{v}_i.$$

(d) Now we consider the regularized LS problem $\arg\min\{\|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2\}$. Show that the solution of this problem is given by the solution of the system

$$(A^\top A + \lambda I)\mathbf{x} = A^\top \mathbf{b}.$$

Also, show that the matrix $(A^\top A + \lambda I)$ is always positive definite for $\lambda > 0$, even if A is not full rank (use the definition of SPD matrices).

(e) Using the same notation of the previous sections show that the solution of the regularized LS problem is given by

$$\hat{\mathbf{x}} = \sum_{i=1}^{\min(m,n)} \frac{\sigma_i}{\sigma_i^2 + \lambda}\alpha_i \mathbf{v}_i.$$

(f) Read the deblurring example shown in the tutorials (Section 5.9.1) and try to explain it using the results above. Assume that the noise in the image corresponds to a singular vector $\mathbf{u}_i$ with a small corresponding singular value σ_i.

7. **Camera Calibration**

Given 3D world coordinates (x, y, z), their projection according to the pinhole camera model on an image plane (u, v) is:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Where

$$K = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \end{bmatrix}$$

The parameters f_x, f_y, x_0, y_0 are called the *intrinsic parameters* of the camera.
Particularly, f_x and f_y denote the focal length of the camera, and x_0, y_0 are the offset of the principal point relative to the origin.
For more information regarding the *pinhole camera* model, please refer to:

`https://en.wikipedia.org/wiki/Pinhole_camera_model`

You are given n correspondences, i.e., for every $i \in \{1, ..., n\}$, you are given the pair:

$$\left(\begin{bmatrix} u_i \\ v_i \end{bmatrix}, \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \right)$$

(a) What is the minimal number of correspondences required to find K?

(b) Assuming the number of correspondences, n, is larger than the minimum number you found in the previous section, how would you solve the problem to obtain a good solution for K?
Show all your computations and derivations for the optimal solution in a least squares sense.