MEDICAL PHYSICS / BIOMEDICAL ENGINEERING 574

Homework Set 1

Diego Hernando

Due date: Friday, February 15, 2019
(Due by the end of the day)

Instructions

- *Homework solutions may be turned in either electronically or on paper.*
- *Make sure to hand in your complete solution to each of the questions, in order, ie: first your complete answer to problem 1, next your complete answer to problem 2, etc. Do not provide separate documents for the theoretical and computational components of each problem.*
- *Note that this homework set is posted before we are done covering all the required topics - this will take another week. However, your solutions are due two weeks after the homework set is posted, ie: about one week after we cover all the required topics. The goal of this timing is to keep the homework topics temporally close to the class materials, and to encourage you to read the homework questions before the materials are covered in class, in order to direct focus during lectures.*

1. (25 points) Vector Norms and Convexity

A. (5 points) Prove that the ℓ_1 norm is a valid vector norm.

B. (5 points) Prove that the ℓ_0 'metric' is not a valid vector norm.

Note: Remember that the ℓ_0 'metric' of a vector is defined as the number of non-zero entries in said vector.

C. (5 points) Prove that any valid vector norm $\|\cdot\|$ is a convex function

Note: In other words, prove that the function $f : \mathbb{R}^N \to \mathbb{R}$ defined as $f(\mathbf{x}) = \|\mathbf{x}\|$, where $\|\mathbf{x}\|$ calculates a given norm of $\mathbf{x}$ (for instance, the ℓ_2 aka Euclidean norm), is a convex function. This will be very important for optimization problems, as these problems (both the objective function as well as the inequality constraints) are often written as norms.

D. (5 points) Prove that the set $\|\mathbf{x}\|_2 \le b$, for some $b > 0$, is a convex set within $\mathbb{R}^N$.

E. (5 points) Plot or draw the following sets for a two-dimensional vector $\mathbf{x} \in \mathbb{R}^2$

i. $\|\mathbf{x}\|_\infty < 1$ (where $\|\mathbf{x}\|_\infty = \max_n |x_n|$

ii. $\|\mathbf{x}\|_2 < 1$

iii. $\|\mathbf{x}\|_1 < 1$

iv. $\|\mathbf{x}\|_{\frac{1}{2}} < 1$ (note that this one is not a norm, but we can define and plot the set just as well using the expression $\|\mathbf{x}\|_{\frac{1}{2}} = \left(\sum_n |x_n|^{1/2}\right)^2$))

Feel free to do this analytically (eg: drawing by hand for easy shapes), or computationally. For each of these plots, graphically indicate whether the corresponding set is convex or not (ie: if the set is convex, the line segment that joins any two points within the set will be completely inside the set, whereas if the set is non-convex you will be able to find two points within the set such that the segment that joins them is not completely within the set).

2. Solving Optimization Problems by Plotting and Staring (15 points)

A. (7.5 points) Minimum-norm solutions to a linear system of equations

Consider the following constrained optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_p, \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \tag{2.1}$$

where $\|\cdot\|_p$ represents the ℓ_p norm (if $p \geq 1$) or metric (if $p < 1$), and

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \end{bmatrix} \tag{2.2}$$

Solve this problem graphically (ie: find the optimum $\hat{\mathbf{x}}$ by plotting, on the space (x_1, x_2), the constraint region $\mathbf{A}\mathbf{x} = \mathbf{b}$ allowed by our equality constraints and finding the points with minimum ℓ_p norm for various p) for the cases where:

i. $p = 0$

ii. $p = 1$

iii. $p = 2$

iv. $p = \infty$

Note that, for some values of p, there may be several solutions - make sure to indicate all of them on the plot. You do not need to specify all the numerical values of the solutions (although this is OK too, in addition to the graphical display), but should display graphically the location of each solution $\hat{\mathbf{x}}$, labeling them with the corresponding value of p. *Hint: Feel free to use the shape of the sets* $\|\mathbf{x}\|_p \leq 1$ *derived in the previous problem as a guide to solving this problem.*

Additional question: As we will see later in the course, the ℓ_0 metric is the ideal metric for minimization if we seek sparse solutions to linear systems (which is often the case in image reconstruction and processing). Remember that a sparse solution is one where many (or most) of the entries in the solution $\hat{\mathbf{x}}$ are zero. In our two-dimensional case, we only have two entries so the sparsest solutions may have one non-zero entry, and one zero entry. However, minimizing the ℓ_0 metric subject to some constraints is computationally very challenging (as we will also see later in the course). Based on your experience with this problem, if you had to pick an actual norm to minimize and try to mimic the behavior of ℓ_0 minimization, would you pick the ℓ_1, ℓ_2, or ℓ_∞ norm?

B. (7.5 points) Approximate solutions based on ℓ_p norms with additional constraints

Consider the following constrained optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p, \text{ subject to } \mathbf{C}\mathbf{x} = \mathbf{d} \tag{2.3}$$

where $\|\cdot\|_p$ represents the ℓ_p norm (if $p \geq 1$) or metric (if $p < 1$), and

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 3 & -1 \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 0 \end{bmatrix} \tag{2.4}$$

Solve this problem graphically (ie: find the optimum $\hat{\mathbf{x}}$ by plotting, on the space (b_1, b_2), the 'data' point $\mathbf{b}$ in 2D, as well as the space of possible vectors $\mathbf{Ax}$ allowed by our equality constraints $\mathbf{Cx} = \mathbf{d}$) for the cases where:

i. $p = 0$

ii. $p = 1$

iii. $p = 2$

iv. $p = \infty$

Note that, for some values of p, there may be several solutions - make sure to indicate all of them on the plot. You do not need to specify all the numerical values of the solutions (although this is OK too, in addition to the graphical display), but should display graphically the location of each solution $\mathbf{A}\hat{\mathbf{x}}$, labeling them with the corresponding value of p.

3. Reformulate LS Problems (15 points)

A. (7.5 points) Simplifying a two-component objective function

Reformulate the following formulation with two components

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{Cx}\|_2^2 \right\} \tag{3.1}$$

for some $\lambda > 0$, as a single matrix, single vector LS formulation.
In other words, rewrite the expression in Equation 3.1 above as the equivalent formulation $\min \|\mathbf{Dx} - \mathbf{e}\|_2^2$, and determine the matrix $\mathbf{D}$ and the vector $\mathbf{e}$ in terms of $\mathbf{A}$, $\mathbf{b}$, $\mathbf{C}$, and λ in order for the equivalence to hold.

B. (7.5 points) Simplifying a constrained optimization problem

Consider the following constrained optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \tag{3.2}$$

$$\text{where } \mathbf{x} = \mathbf{x_0} + \mathbf{Cy} \tag{3.3}$$

for some vector $\mathbf{y}$, where $\mathbf{x} \in \mathbb{R}^{20}$ and $\mathbf{y} \in \mathbb{R}^{10}$, and the remaining matrices and vectors have suitable dimensions. Note that we do not know $\mathbf{y}$ - all we know is that there needs to exist a $\mathbf{y}$ such that $\mathbf{x}$ satisfies the condition above.

Reformulate the optimization problem above as an unconstrained optimization problem in terms of $\mathbf{y}$ instead of $\mathbf{x}$, written as $\min \|\mathbf{Dy} - \mathbf{e}\|_2^2$ for some matrix $\mathbf{D}$ and vector $\mathbf{e}$, such that $\hat{\mathbf{x}}$ can be obtained readily from the optimum $\hat{\mathbf{y}}$. Express $\mathbf{D}$ and $\mathbf{e}$ in terms of $\mathbf{A}$, $\mathbf{b}$, $\mathbf{x_0}$, and $\mathbf{C}$.

4. Convexity (20 points)

Which of the following functions are convex? Prove it or disprove it for each function, based on the definition of convexity and/or its properties (eg: operations that preserve convexity):

i. $f(x) = 0$

ii. $f(x) = \log x$ (for $x > 0$)

iii. $f(x) = |x|^3$

iv. $f(x) = f_1(x) + f_2(x)$, where both f_1 and f_2 are convex.

v. $f(x) = ag(x) + b$, where g is convex, and a and b are scalars (which can be positive or negative).

vi. $f(x) = g(ax + b)$, where g is convex, and a and b are scalars (which can be positive or negative).

vii. $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2$

viii. $f(x) = |e^{ix} - b|^2$ (for $x \in \mathbb{R}$, and where $i = \sqrt{-1}$)

ix. $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \|\mathbf{Fx}\|_0$

Note: In the cases that include unspecified matrices and vectors, you can assume that the matrices are identity matrices and the vectors are all ones.

5. ML and MAP Estimation (25 points)

A. (8 points) Formulate a ML estimation problem in the presence of Gaussian noise as a LS optimization problem

Suppose we have a measuring device (eg: an imaging system) that takes measurements from an unknown object determined by a vector $\mathbf{x}$, as follows:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \eta \tag{5.1}$$

where $\mathbf{b} \in \mathbb{R}^M$ is our measurement vector, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the system matrix, $\mathbf{x} \in \mathbb{R}^N$ is the unknown vector that we would like to estimate, and $\eta \in \mathbb{R}^M$ is a noise vector containing independent, identically distributed (iid) noise in each of its entries. Further, assume that the noise distribution is Gaussian with mean zero and standard deviation σ_1.
Derive the maximum-likelihood (ML) estimate for $\mathbf{x}$ *as a least-squares (LS) optimization problem, and describe whether this optimization problem depends on* σ *(ie: will the ML solution* $\hat{\mathbf{x}}_{ML}$ *change based upon the noise level?).*
Hint: Note that the ML estimate $\hat{\mathbf{x}}_{ML}$ corresponds to maximizing the probability of observing $\mathbf{b}$ given $\mathbf{x}$, ie:

$$\hat{\mathbf{x}}_{ML} = \arg\max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x}) \tag{5.2}$$

where, in the presence of Gaussian noise, $P(\mathbf{b}|\mathbf{x})$ is given by:

$$P(\mathbf{b}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2}{2\sigma_1^2}} \tag{5.3}$$

B. (8 points) Let's turn our ML estimation into a MAP estimation problem

Now let us assume that we know the a priori statistical distribution of $\mathbf{x}$, which happens to be a Gaussian iid distribution with standard deviation σ_2. In other words, in the absence of any data, the probability density function of $\mathbf{x}$ is given by:

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \sigma_2^{2N}}} e^{-\frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2}} \tag{5.4}$$

Now that we have both an a priori distribution as well as some noisy data, let's formulate the maximum-a-posteriori (MAP) estimation of $\mathbf{x}$ (which combines both sources of information), as follows:

$$\begin{aligned}
\hat{\mathbf{x}}_{MAP} &= \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{b}) && (5.5)\\
&= \arg\max_{\mathbf{x}} \frac{P(\mathbf{b}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{b})} && (5.6)\\
&= \arg\max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x})P(\mathbf{x}) && (5.7)
\end{aligned}$$

where the last step is warranted since $P(\mathbf{b})$ does not depend on $\mathbf{x}$.
Formulate the MAP estimation above as a LS optimization problem of the form:

$$\hat{\mathbf{x}}_{MAP} = \arg\min_{\mathbf{x}} \|\mathbf{E}\mathbf{x} - \mathbf{f}\|_2^2 + \lambda \|\mathbf{G}\mathbf{x}\|_2^2 \tag{5.8}$$

and determine the matrices, vectors, and scalar in this LS optimization ($\mathbf{E}$, $\mathbf{f}$, λ, $\mathbf{G}$), in terms of the parameters of our problem ($\mathbf{A}$, $\mathbf{b}$, σ_1, σ_2, etc).

C. (9 points) A specific example of ML and MAP estimation.

Assuming that $\sigma_1 = 1$, $\sigma_2 = 1$, and:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{5.9}$$

$$\mathbf{b} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \tag{5.10}$$

find the ML and MAP estimates $\hat{\mathbf{x}}_{ML}$ and $\hat{\mathbf{x}}_{MAP}$.

Next, assume that the a priori distribution of $\mathbf{x}$ is 'tighter', eg: $\sigma_2 = 0.5$. What is the new value of $\hat{\mathbf{x}}_{MAP}$? How about if $\sigma_2 = 0.2$? Can you guess what is the limit of $\hat{\mathbf{x}}_{MAP}$ as $\sigma_2 \to 0$?

Next, assume that the a priori distribution of $\mathbf{x}$ is'broader', eg: $\sigma_2 = 2$. What is the new value of $\hat{\mathbf{x}}_{MAP}$? How about if $\sigma_2 = 5$? Can you guess what is the limit of $\hat{\mathbf{x}}_{MAP}$ as $\sigma_2 \to \infty$?
MatLab note: For the LS problems considered in this exercise, if we want to solve

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \tag{5.11}$$

we can simply run "`xhat = A\b`" on MatLab. Note that for the problems that have several terms (eg: the MAP estimation in Equation 5.8), we can rewrite them as a single-term LS problem as shown in Question 3 in this homework set.