Nonlinear Optimization (18799 B, PP)
IST-CMU PhD course
Spring 2017
Instructor: João Xavier (`jxavier@isr.ist.utl.pt`)
TA: Shanghang Zhang (`shzhang.pku@gmail.com`)

The homework is due May 24.

Homework 4

Problem A. *(Optimal choice of measurements)* We start by recalling some basic properties of maximum likelihood (ML) estimators in linear Gaussian models. Let $\theta \in \mathbf{R}^p$ be a vector of parameters of interest. We have access to a measurement

$$y = H\theta + w$$

where $H \in \mathbf{R}^{n\times p}$ is a known matrix and $w \sim \mathcal{N}(0, I)$, *i.e.*, w denotes zero-mean, Gaussian noise with unit covariance. Assume H has full column-rank. To estimate θ from y it is common to use the ML estimator

$$\widehat{\theta}_{\mathrm{ML}}(y) = \left(H^T H\right)^{-1} H^T y.$$

It is easy to show that the ML estimation error $\varepsilon_{\mathrm{ML}}(y) = \widehat{\theta}_{\mathrm{ML}}(y) - \theta$ has zero-mean (ML estimator is unbiased) and covariance

$$R = \mathbf{E}\left(\varepsilon_{\mathrm{ML}}(y)\varepsilon_{\mathrm{ML}}(y)^T\right) = \left(H^T H\right)^{-1}.$$

The matrix R quantifies the estimator error in each "direction". For example, if we wish to estimate some linear combination of the parameters, say, $u^T\theta$ for a given $u \in \mathbf{R}^p$, the unbiased estimator $u^T\widehat{\theta}_{\mathrm{ML}}(y)$ has variance $u^T R u$.

We now turn to our problem. Suppose that we have access to S independent measurement systems. Each system can be used several times. The mth usage of the sth system gives us the measurement

$$y_s(m) = H_s\theta + w_s(m)$$

where $H_s \in \mathbf{R}^{n_s\times p}$ is the measurement matrix (known, with full column-rank) of the sth system and $w_s(m) \sim \mathcal{N}(0, I)$. Assume that noise realizations are independent across systems and time. Performing one measurement with the sth system takes T_s seconds and costs C_s euros.

We have a total budget of C dollars and want to find how many times we should use each system to strike a good balance between the time needed to obtain the ML estimate and its accuracy. More specifically, assume that we use x_1 times system 1 , we use x_2 times system 2, and so on. This corresponds to collecting the measurement

$$y = H(x_1, \ldots, x_S)\theta + w$$

where

$$H(x_1, \ldots, x_S) = \begin{bmatrix} H_1 \\ \vdots \\ H_1 \\ H_2 \\ \vdots \\ H_S \end{bmatrix} \in \mathbf{R}^{(x_1 n_1 + \cdots + x_S n_S)\times p},$$

the matrix H_s appearing x_s times. Also, regardless of x_s, there holds $w \sim \mathcal{N}(0, I)$. It follows that the error covariance of the associated ML estimator is

$$R(x_1, \ldots, x_S) = \left(x_1 H_1^T H_1 + \cdots + x_S H_S^T H_S\right)^{-1}. \qquad (1)$$

Note that it takes $x_1T_1 + \cdots + x_ST_S$ seconds and it costs $x_1C_1 + \cdots + x_SC_S$ euros to obtain this estimate.

We want to solve the optimization problem

$$\begin{array}{ll} \underset{x_1,\ldots,x_S}{\text{minimize}} & \lambda_{\max}(R(x_1,\ldots,x_S)) + \rho(x_1T_1 + \cdots + x_ST_S) \\ \text{subject to} & x_1C_1 + \cdots + x_SC_S \leq C \\ & x_s \geq 0,\ s = 1,\ldots,S \\ & x_s \text{ is integer},\ s = 1,\ldots,S. \end{array} \tag{2}$$

The first term in the objective function gives the worst estimation accuracy (in estimating $u^T\theta$ with unit-norm u) and the second term penalizes large delays (the constant $\rho > 0$ is given).

Problem (2) involves integer variables. We relax it to continuous variables by dropping the last constraint. That is, as an approximation, we look instead at the problem

$$\begin{array}{ll} \underset{x_1,\ldots,x_S}{\text{minimize}} & \lambda_{\max}(R(x_1,\ldots,x_S)) + \rho(x_1T_1 + \cdots + x_ST_S) \\ \text{subject to} & x_1C_1 + \cdots + x_SC_S \leq C \\ & x_s \geq 0,\ s = 1,\ldots,S. \end{array} \tag{3}$$

Note that $R(x_1,\ldots,x_S)$ is given by (1). Formulate (3) as a semidefinite program (SDP). Hint: you need to learn about Schur complements.

Problem B. *(Core ellipses)* We want to find the center c of the ellipse

$$\mathcal{E}(c,A) = \left\{x \in \mathbf{R}^2 \,:\, (x-c)^T A^{-1}(x-c) \leq 1\right\}$$

that maximizes the probability of containing the realizations of a random vector X. The matrix $A \succ 0$ is given and the random vector X takes values in a given finite set $\mathcal{X} = \{x_1,\ldots,x_K\} \subset \mathbf{R}^2$ with $p_k = \mathbf{Prob}(X = x_k)$ known for $k = 1,\ldots,K$. We formulate the problem

$$\underset{c}{\text{maximize}} \quad \mathbf{Prob}(X \in \mathcal{E}(c,A)). \tag{4}$$

The objective can be expressed as $\mathbf{Prob}(X \in A) = \sum_{k=1}^K p_k 1_{\mathcal{E}(c,A)}(x_k)$ where

$$1_{\mathcal{E}(c,A)}(x_k) = \begin{cases} 1 & ,\ \text{if } x_k \in \mathcal{E}(c,A) \\ 0 & ,\ \text{otherwise.} \end{cases}$$

It is clear that the objective is not a convex function of the optimization variable c; in fact, it is not even a continuous function. In this problem, we will explore two convex approximations for the difficult problem (4).

(a) In the first approach we minimize the expected distance from X to $\mathcal{E}(c,A)$:

$$\underset{c}{\text{minimize}} \quad \mathbf{E}\left(d(X,\mathcal{E}(c,A))\right). \tag{5}$$

Note that

$$\mathbf{E}\left(d(X,\mathcal{E}(c,A))\right) = \sum_{k=1}^K p_k d\left(x_k,\mathcal{E}(c,A)\right)$$

where, for a point x and a set S, $d(x,S) = \min\{\|y-x\| \,:\, y \in S\}$ denotes the distance from x to S. Express (5) as a convex problem (LP, QP, SOCP or SDP).

(b) In the second approach we formulate the optimization problem

$$\underset{c}{\text{minimize}} \quad \mathbf{E}\left(\left((X-c)^T A^{-1}(X-c) - 1\right)_+\right). \tag{6}$$

Note that if a realization of X belongs to $\mathcal{E}(c,A)$ then $\left((X-c)^T A^{-1}(X-c) - 1\right)_+ = 0$, and no penalization is incurred; otherwise, the objective penalizes how much the realization violates the inequality $(x-c)^T A^{-1}(x-c) \leq 1$ characterizing points $x \in \mathcal{E}(c,A)$. We can also interpret $\left((X-c)^T A^{-1}(X-c)-1\right)_+$ as a surrogate for $d(X,\mathcal{E}(c,A))$. Note that

$$\mathbf{E}\left(\left((X-c)^T A^{-1}(X-c)-1\right)_+\right) = \sum_{k=1}^{K} p_k \left((x_k-c)^T A^{-1}(x_k-c)-1\right)_+ .$$

Express (6) as a convex problem (LP, QP, SOCP or SDP).

(c) Use CVX to test your formulations: insert your code into the MATLAB file `hw4pB.m` and print the corresponding figure. A sample is given in figure 1.

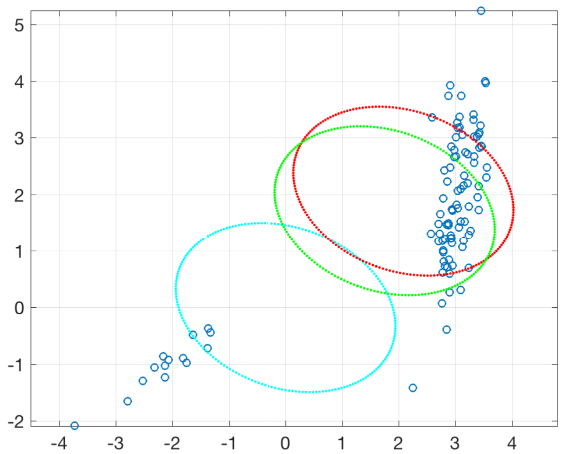

Figure 1: Set of possible realizations $\mathcal{X}$ (blue) on which we assume the uniform distribution ($p_k = 1/K$). We want to translate the ellipsoid at the origin (cyan). The optimal ellipsoids for parts (a) and (b) are in red and green, respectively.