# Numerical Methods, Spring 2023
# Assignment 8 [Optimization: Applications.]
# Total: 135, Deadline: 8 Apr

## I. SUGGESTED READING

- Chapters 5, 7 [1]

## II. EXERCISES

1. (70) We consider (two) the multivariate ($n$ variables) timeseries ($T$ steps) possibly with some of the observations missing (miss-rate $p$). These time series are sampled from multivariate normal distributions:

$$P(\mathbf{x}) = (2\pi)^{-n/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{\Sigma}^{-1}\mathbf{x}\right). \tag{1}$$

For each time series, the covariance matrix $\mathbf{\Sigma}$ is positive definite of a specific form:

$$\mathbf{\Sigma} = \boldsymbol{C}\boldsymbol{C}^T + \alpha\mathbf{1} \tag{2}$$

where $\boldsymbol{C}$ is a tall $n \times k$ matrix and $\alpha > 0$. You are given the timeseries-1 and timeseries-2, and your task is to estimate the matrix $\mathbf{\Sigma}$ as accurately as possible. In both cases, $k = 2$, $n = 15$ and $T = 20$. In the second time series, roughly 20% of observations are missing.

How do we estimate the covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ given the observations and additional information above? Before we switch to that, lets set up the scoring rule: how do we evaluate the accuracy of such estimation? From the probabilistic point of view, it makes sense to rely on the Kullback–Leibler divergence between the true covariance matrix and estimated one. Formally, for distributions $P$ and $Q$ of a continuous random variable, it is defined to be the integral:

$$D_{\mathrm{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx. \tag{3}$$

In the case of multivariate normal distributions with zero mean it becomes:

$$D_{\mathrm{KL}}(\Sigma_0 \parallel \Sigma_1) = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma_1^{-1}\Sigma_0\right) - n + \ln\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right)\right). \tag{4}$$

We will use this formula, having in mind that $\Sigma_0$ is the true covariance matrix and $\Sigma_1$ is the proposed estimation. Note that the smaller $D_{\mathrm{KL}}$ the better the estimation is and $D_{\mathrm{KL}} \geq 0$. In this exercise, your are given the timeseries and do not have access to the true covariance matrices, which were used to generate them. So, optimizing $D_{\mathrm{KL}}$ is not an option (but $D_{\mathrm{KL}}$ will be used to check your solution).

Your goal will be to find the most suitable matrix of form in Eq. (2). The most suitable means the one maximizing the likelihood of the observed time series:

$$L(\Sigma) = \Pi_{t=1}^{t=T} P(\mathbf{x}_t). \tag{5}$$

This equation can be computed only if no data is missing. In some of the observations are missing, the most systematic approach is to integrate out corresponding missing variables $x_{ti}$. For example, imagine that at $t = 1$ observations $x_{1,1}, x_{1,5}, x_{1,9}, x_{1,10}, x_{1,12}$ are missing, then the corresponding factor $t = 1$ in the likelihood Eq. (5) will look like:

$$\int P(x_{1,0}, x_{1,1}, ..., x_{1,14}) dx_{1,1} dx_{1,5} dx_{1,9} dx_{1,10} dx_{1,12}. \tag{6}$$

The integrals can be conveniently computed in terms of inverse of certain submatrices of $\Sigma$ and depend only on the observed values.

In practice, its more convenient to maximize the mean log-likelihood which in this case reads:

$$LL(\Sigma) = \frac{1}{T} \ln L(\Sigma) = \frac{1}{T} \sum_{t=1}^{T} \ln P(\mathbf{x}_t). \tag{7}$$

- (5) Compute empirical covariance matrices for both timeseries. Do they satisfy Eq. (2)?
- (10) Show that if there is no missing data, LL is a function of empirical covariance matrix $\Sigma_{\text{emp}}$ and of $\Sigma$.
- (15) For the timeseries-1, you thus have to maximize

$$LL(\boldsymbol{C}\boldsymbol{C}^T + \alpha \mathbf{1}, \Sigma_{\text{emp}}) \tag{8}$$

  over $\boldsymbol{C}$ and $\alpha > 0$. Set up the cost function in `JAX` and use some 2nd-order optimization algorithm. Don't forget to compile your cost function as well as its gradients for speed-up. For starting point of the optimization algorithm, use some heuristic decomposition of $\Sigma_{\text{emp}}$ to the form of Eq. (2).
- (10) To better understand the approach to the case of missing data, consider analytically the case of one-step timeseries with single vector observation $x = (x_1, \texttt{nan}, x_3)$ and proposed covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}.$$

  Integrate out unobserved value – compute $\int P(x_1, x_2, x_3) dx_2$ analytically.
- (30) Generalize the previous approach for arbitrary pattern of the missing data. Set up the corresponding cost function for the timeseries-2 and optimize it over $\Sigma$, similarly to the case of the timeseries-1.

2. (40) A signal emitted by a source at an unknown position $r \in R^2$ is received by $m$ sensors at known positions $r_1, ..., r_m \in R^2$ (see the file). From the strength of the received signals (see the file), the noisy estimates $d_k$ of the distances $\|r - r_k\|_2$ are obtained. We are interested in estimating the source position $r$ based on the measured distances $d_k$. Formally, we have to minimize:

$$L(r) = \sum_k \left( \|r - r_k\|_2^2 - d_k^2 \right)^2 \tag{9}$$

over $r$ at given $r_k$ and $d_k$.

- (5) Is the function in Eq. (9) convex?
- (25) Your goal in this task is to find the global minimum using the method of Lagrange multipliers. Consider rewriting the optimization problem identically to

$$\min_{r,t} \sum_k \left( t + \|r_k\|_2^2 - 2r \cdot r_k - d_k^2 \right)^2 \quad \text{s.t.} \quad \|r\|_2^2 = t \tag{10}$$

  This minimization under constraint is reduced to finding stationary points in $r = (x, y)$ and $t$ of the following Lagrangian

$$\bar{L}(x, y, t) = \sum_k \left( t + \|r_k\|_2^2 - 2r \cdot r_k - d_k^2 \right)^2 + \lambda \left( \|r\|_2^2 - t \right). \tag{11}$$

  Since this is just a quadratic form over $x, y, t$, the stationary point can be found explicitly as $x(\lambda), y(\lambda), t(\lambda)$ (express these functions in terms of $r_k$ and $d_k$). NB: `sympy` can be useful at some point.

- (10) Solve numerically the polynomial equation $x(\lambda)^2 + y(\lambda)^2 - t(\lambda) = 0$ over $\lambda$ to find the point, which delivers the global minimum for the given data. Plot the graph with (i) positions of the detectors, (ii) contour lines of $L(r)$, (iii) all locally extremal points and (iv) the global minimum. Characterize the locally extremal points (local maximum/minimum/saddle point).

3. (25) Consider trading a universe of $n$ stocks over $T$ days, with noisy predictions for the stock returns $p_{ti}$ available file (the rows are days, the data start with day 0). Assuming that we have the position evolving as $\pi_t$ ($\pi_t$ at each $t$ is a vector of $n$ components), expected risk-adjusted gain $G$ reads

$$G = \sum_t \left[ p_t \cdot \pi_t - \pi_t \cdot \Omega \cdot \pi_t - \gamma \sum_i |\pi_{t,i} - \pi_{t-1,i}| \right] \tag{12}$$

and has to be maximized over $\pi_t$ (pick $\gamma = 0.01$). The matrix $\Omega$ is available in the file.

- (5) Start with $t = 1$ ($\pi_0 = 0$ by definition). At this moment, you have access to $p_1$ and have to maximize:

$$p_1\pi_1 - \pi_1\Omega\pi_1 - \gamma \sum_i |\pi_{1,i} - 0| \longrightarrow \max$$

over $\pi_1$. Show that this is a concave function of $\pi_1$ and maximize it using `cvxpy`.

- (10) At $t = 2$ you already know $\pi_1$. At this moment, you have access to $p_2$ and have to maximize:

$$p_2\pi_2 - \pi_2\Omega\pi_2 - \gamma \sum_i |\pi_{2,i} - \pi_{1,i}| \longrightarrow \max$$

Repeat the process until you reach the end of the time series. The corresponding $\pi_{ti}$ should be stored as a file: this will be you 1st result in this problem.

- (5) considering the case of $\gamma = 0$: in this case, optimization of $G$ can be done directly. Make sure the result of such direct computation coincides with `cvxpy` result. The corresponding $\pi_{ti}$ should be stored as a file: this will be your 2nd result in this problem.
- (5) Compute expected gain/costs (the 1st and the 3rd terms in Eq. (12)) over the full period for two trading strategies computed above (note that the trading costs are present even if you decided to optimize at $\gamma = 0$).

*Interpretation.* In this problem, you are given the noisy estimates of the stock returns, $p_{it}$ and the true returns $r_{it}$ are known only to the problem author. If the true returns are known, the trading strategy ($\pi_t$) can be evaluated by computing the PnL:

$$\mathrm{PnL}_t = r_t \cdot \pi_t - \gamma \sum_i |\pi_{t,i} - \pi_{t-1,i}| \tag{13}$$

After the problem is solved, we will compute $\mathrm{PnL}_t$ and discuss the results.

## REFERENCES

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.