Supplementary Material
README file for the '`RMTest.m`' Matlab Code

Abstract

This document explains how to use the code implementing the improved RMT estimate of covariance matrix and precision matrix proposed in the core of the article.

1. Archive content

- The function implementing our method is called `RMTest.m` which takes as arguments the data matrix `X`, the `gradient_check` and `plot_cost` options, the initial point `C_0` and the distance under investigation. The function returns the estimated covariance matrix `C_est` and the cost function related to this estimated covariance.
- The main script comparing all algorithms for synthetic data is `CompareEst.m`.
- The main script comparing the estimation methods for the LDA/QDA application is `ML_applications.m`.
- folder `Manopt`: a Matlab toolbox for optimization on manifolds (Boumal et al., 2014). Some of the functions in this folder were adapted to better suit our present problem.
- folder `Othermethods`: containing alternative estimation methods among which the QuEST methods QuEST1 and QuEST2 from (Ledoit & Wolf, 2015; Ledoit et al., 2018), the Rao-Blackwell Ledoit-Wolf estimation methods (Chen et al., 2010) and the Oracle Approximating Shrinkage estimation method (Chen et al., 2010).
- folder `Utilities`: contains supplementary codes for the implementation of LDA,QDA.

2. Code `CompareEst.m`

The different options proposed to execute the script `CompareEst.m` comparing the different estimation algorithms are as follows:

- The range of n and the value of p
- The covariance matrices metric "`distance`" (among `Fisher, Battacharrya, KL, log, log1st, t`) or for the precision matrices (among `Inverse_Fisher, ,Inverse_Battacharrya,Inverse_KL, Inverse_log1st, Inverse_log,Inverse_t`)
- The target matrix "`Covariance`" (among `dirac,Wishart, toeplitz`) and their parameters "`param`" if needed (for the Wishart and Toeplitz cases)
- The initialization point for the gradient descent algorithm denoted "`initialization`" (linear shrinkage from (Ledoit & Wolf, 2004) denoted "`shrinkage`", shrinkage from (Chen et al., 2010) denoted "`alternative shrinkage`", QuEST denoted as "`ledoit-wolf`") or the identity denoted "`manual`"
- Other binary option can be chosen (among 0/1): (`gradient_check` to check if the gradient is correct, `plot_cost` to see the cost/real distance during iterations.

3. Code `ML_applications.m`

The different options proposed to execute the script `ML_applications.m` are as follows:

- the data on which the LDA/QDA are applied denoted "`dataset`" for which the options are `synthetic` for synthetic data and `eeg` for eeg dataset.
- The machine learning algorithms denoted "`application`" for which the options are `LDA` and `QDA`.
- For synthetic data, the examples of covariance under investigation. For both the covariance of the first and second class, `Wishart` and `toeplitz` are the two options.

4. Reproducing the results of the article

The following sections detail the parameter setting to reproduce the figures of the main article.

4.1. Figure 1

Script → `CompareEst.m`
covariance → toeplitz
parameter → 0.4
p → 512, n → 500
distance → Fisher
Initialisation → manual
plot_cost → 1

4.2. Figure 2

Script → `CompareEst.m`
covariance → Wishart/toeplitz(0.1)/toeplitz(0.9)/dirac
n → linspace(210,500,10), p → 200
distance → Fisher
Initialisation → shrinkage
plot_cost → 0

4.3. Figure 3

Script → `CompareEst.m`
covariance → Wishart/toeplitz(0.1)/toeplitz(0.9)/dirac
n → linspace(210,500,10), p → 200
distance → Inverse_Fisher
Initialisation → shrinkage
plot_cost → 0

4.4. Figure 4

Script → `ML_applications.m`
mu2 → mu1+80/p
application → LDA
covariance1 → Wishart/Wishart/toeplitz/
covariance2 → Wishart/toeplitz/toeplitz/
dataset → synthetic/synthetic/synthetic/eeg
p → linspace(500,200,10), n → 512
eeg dataset → $n\_train = 500$, $n\_test = 1000$,p=100

4.5. Figure 5

Script → `ML_applications.m`
mu2 → mu1+1/p except for the third one which is mu1+80/p
application → QDA
covariance1 → Wishart/Wishart/toeplitz/
covariance2 → Wishart/toeplitz/toeplitz/
dataset → synthetic/synthetic/synthetic/eeg
p → linspace(500,200,10), n → 512
eeg dataset → $n\_train = 150$, $n\_test = 1000$,p=100

References

Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014. URL `http://www.manopt.org`.

Chen, Y., Wiesel, A., Eldar, Y. C., and Hero, A. O. Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010.

Ledoit, O. and Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Ledoit, O. and Wolf, M. Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis*, 139: 360–384, 2015.

Ledoit, O., Wolf, M., et al. Optimal estimation of a large-dimensional covariance matrix under steins loss. *Bernoulli*, 24(4B):3791–3832, 2018.