Structured Covariance Estimation via HPD Manifold Optimization

Huanyu Zhang

Abstract

This paper considers the problem of estimating a structured covariance matrix with an elliptical contoured distributions (ECDs) with zero mean. This problem can always be formulated to optimize the maximum likelihood (ML) estimator under the hermitian positive definite (HPD) constraint. First, we find that the ML estimator is geodesic convexity (g-convexity) which allows global optimization. Therefore, lots of convex optimization algorithms on manifolds can be applied to this problem. We will apply kinds of manifold optimization method to handle this problem. In addition, Nesterov's accelerating gradient descent algorithms will be implemented. And we will compare with the fixed-point method.

I. Introduction

Estimating the covariance matrix is a common problem that is critically important in many fields such as signal processing, wireless communication, bioinformatics, and financial engineering.

Classical covariance estimation always has two basic assumptions: the number of samples is larger than the dimensions of the samples; the samples are drawn from an Gaussian distribution. However, in practice, there are many applications neither of these assumptions holds.

For the situation of high dimension covariance estimation using small number of samples, one kind of classical approaches is adding penalties to the cost function which can exploit prior knowledge of the unknown parameters, such as scalar penalties [9], matrix penalties [11], and smoothness penalties [1]. Another approaches is the shrinkage technique, which includes shrinkage into identity [3], shrinkage into diagonal structure [6], and shrinkage into positive definite matrix [11].

For the situation of samples from the distribution except for Gaussian distribution. [4] provides an alternative to maximizing the likelihood, which does not require assuming a particular probability distribution for the data. Another common study is to replace the Gaussian assumption with a more general scaled Gaussian model, e.g., elliptical distributions. Tyler's ML estimator [5], a well-studied estimator, performs robustly in scaled Gaussian distribution, the Tyler's estimator defined as the solution to the fixed-point equation

$$\mathbf{R}^{(k+1)} = \frac{K}{N}\sum_{i=1}^{N}\frac{\mathbf{x}_i\mathbf{x}_i^T}{\mathbf{x}_i^T\left(\mathbf{R}^{(k)}\right)^{-1}\mathbf{x}_i}$$

is a minimax robust estimator. Actually, Tyler's estimator does not depend on the distribution of data. Tyler's ML estimator can also combine with the shrinkage technique [11].

In many scenario, the covariance matrix has special structure naturally, which implies a reduction in the number of parameters to be estimated, and thus is beneficial to improving the estimation accuracy. Various types of structures have been studied. For example, the Toeplitz structure with applications in time series analysis and array signal processing. Kronecker structures, also known as separable models, transposable covariances models, or matrix-variate-normal models are typically used when dealing with random matrices (rather than random vectors). Covariance matrix estimation in Kronecker structures involves a non-convex optimization. But its maximum likelihood estimation function is fact g-convex [10]. It is concluded that any local minimum of the cost function on a group Kronecker constraint set is a global minimum. Many numerical algorithm such as majorization-minimization algorithm was also proposed to solve the constrained minimization problem[8].

Similarly, we will consider the HPD structures, *whose* maximum likelihood estimation function is also g-convex.

II. Problem Formulation

Suppose a column vector x is drawn from an N-dimensional Gaussian distribution with zero mean and covariance matrix R, the corresponding probability density function is[2]

$$p(x) = (2\pi)^{-N/2}\det(R)^{-1/2}e^{-\frac{x^TR^{-1}x}{2}}. \tag{1}$$

Now, instead of a single vector sample, suppose that we have M independent vector samples, x_i, $i = 1$ to M. The probability density for this set of vectors follows from (1) as

$$p(x_1, x_2, \dots, x_N) = (2\pi)^{-NM/2}\det(R)^{-M/2}e^{-\sum_{i=1}^{M}\frac{x_i^TR^{-1}x_i}{2}}. \tag{2}$$

Given a set of vector samples, x_m, $m = 1$ to M, the R that has a special structure and maximizes (2) is the "maximum-likelihood" estimate of the covariance matrix. In order to simplify the equation (2), we taking the logarithm of (2), we get

$$-(\frac{MN}{2})\log(2\pi) - (\frac{M}{2})\log\det(R) - (\frac{1}{2})\sum_{m=1}^{M} x_m^T R^{-1} x_m$$

Dropping the leading constant term and dividing through by $M/2$, we define our optimization problem to be:

$$\begin{aligned} \underset{R}{\text{minimize}} \quad & \log\det(R)(\frac{1}{M}) + \sum_{m=1}^{M} x_m^T R^{-1} x_m \\ \text{subject to} \quad & R \in \text{HPD} \end{aligned} \tag{3}$$

The objective function in (3) is non-convex in classical definition. However, from the perspective of geodesic convexity, the objective function can be viewed as g-convex, which will be proved in the next section.

III. Convexity of Objective Function

First, we will introduce the definition of g-convex sets and the g-convex functions.

- Definition 1 (g-convex sets). Let $\mathcal{M}$ be a d-dimensional connected C^2 Riemannian manifold. A set $\mathcal{X} \subset \mathcal{M}$ is called geodesically convex if any two points of $\mathcal{X}$ are joined by a geodesic lying in $\mathcal{X}$. That is, if $x, y \in \mathcal{X}$, then there exists a shortest path $\gamma : [0,1] \to \mathcal{X}$ such that $\gamma(0) = x$ and $\gamma(1) = y$
- Definition 2 (g-convex functions). Let $\mathcal{X} \subset \mathcal{M}$ be a g-convex set. A function $\phi : \mathcal{X} \to \mathbb{R}$ is called geodesically convex if for any $x, y \in \mathcal{X}$, we have the inequality (4)

  $$\phi(\gamma(t)) \le (1-t)\phi(\gamma(0)) + t\phi(\gamma(1)) = (1-t)\phi(x) + t\phi(y) \tag{4}$$

  where $\gamma(\cdot)$ is the geodesic $\gamma : [0,1] \to \mathcal{X}$ with $\gamma(0) = x$ and $\gamma(1) = y$.

To define g-convex functions on HPD matrices, recall that $\mathbb{P}_d$ is a differentiable Riemannian manifold where geodesics between points are available in closed form. At any point $A \in \mathbb{P}_d$, its Riemannian metric is defined as $ds = \|A^{\frac{1}{2}} \cdot dA \cdot A^{-\frac{1}{2}}\|_F$. For $A, B \in \mathbb{P}_d$ there is a unique geodesic path[7]

$$\gamma(t) = A\#_t B := A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t A^{\frac{1}{2}}, \qquad t \in [0,1].$$

The midpoint of this path, namely, $A\#_{1/2}B$, is called the matrix geometric mean, we drop the $1/2$ and denote it simply by $A\#B$.

Recall the fundamental operator inequality (where $\preceq$ denotes the Löwner Partial order):

$$A\#_t B \preceq (1-t)A + tB \tag{5}$$

The formula (5) uses the operator inequality

In order to prove the convexity of objective function, we take it apart. Firstly, we focus on the last part. we are going to prove that $f(A) = x^T A^{-1} x$ is g-convex. To prove this is equal to verify the midpoint convexty:

$$f(A\#B) \le \frac{f(A) + f(B)}{2},$$

where $A, B \in \mathbb{P}_d$. Since $(A\#B)^{-1} = A^{-1}\#B^{-1}$ and $A^{-1}\#B^{-1} \le \frac{A^{-1}+B^{-1}}{2}$[cite]. Therefore

$$f(A\#B) = x^T(A^{-1}\#B^{-1})x \le x^T \frac{A^{-1}+B^{-1}}{2} x = \frac{x^T A^{-1} x + x^T B^{-1} x}{2} = \frac{f(A)+f(B)}{2}.$$

To prove convexity of the first part of objective function, some definition and theorem will be introduced:

- Definition 3 (positive linear map). A linear map Φ from Hilbert space $\mathcal{H}_1$ to a Hilbert space $\mathcal{H}_2$ is called positive if for $0 \preceq A \in \mathcal{H}_1, \Phi(A) \succeq 0$. It is called strictly positive if $\Phi(A) \succ 0$ for $A \succ 0$; finally, it is called unital if $\Phi(I) = I$.
- Theorem 1. Let $h : \mathbb{P}_k \to \mathbb{R}$ be nondecreasing (in Löwner order) and $g-$ convex. Let $r \in \{\pm 1\}$, and let Φ be a positive linear map. Then, $\phi(S) = h\left(\Phi\left(S^r\right)\right)$ is g-convex.

Since log det is monotonic and determinants are multiplicative, let $h = \log\det(X)$ and $\Phi(X) = X$. Then $\phi(X) = \log\det(X)$ is g-convex. Therefore, the objective function in equation (3) is g-convex.

IV. Optimization

References

[1] T. Bucciarelli, P. Lombardo, and S. Tamburrini. Optimum CFAR detection against compound Gaussian clutter with partially correlated texture. *IEE Proceedings - Radar, Sonar and Navigation*, 143(2):95, 1996.

[2] J.P. Burg, D.G. Luenberger, and D.L. Wenger. Estimation of structured covariance matrices. *Proceedings of the IEEE*, 70(9):963–974, 1982.

[3] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, October 2010.

[4] Antoni Maria Musolas Otaño. *Covariance estimation on matrix manifolds*. PhD thesis, Massachusetts Institute of Technology, 2020.

[5] F. Pascal, Y. Chitour, J-P. Ovarlez, P. Forster, and P. Larzabal. Covariance Structure Maximum-Likelihood Estimates in Compound Gaussian Noise: Existence and Algorithm Analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48, January 2008.

[6] Juliane Schäfer and Korbinian Strimmer. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), January 2005.

[7] Suvrit Sra and Reshad Hosseini. Conic Geometric Optimization on the Manifold of Positive Definite Matrices. *SIAM Journal on Optimization*, 25(1):713–739, January 2015.

[8] Ying Sun, Prabhu Babu, and Daniel P. Palomar. Robust Estimation of Structured Covariance Matrix for Heavy-Tailed Elliptical Distributions. *IEEE Transactions on Signal Processing*, 64(14):3576–3590, July 2016.

[9] J. Wang, A. Dogandzic, and A. Nehorai. Maximum Likelihood Estimation of Compound-Gaussian Clutter and Target Parameters. *IEEE Transactions on Signal Processing*, 54(10):3884–3898, October 2006.

[10] A. Wiesel. Geodesic Convexity and Covariance Estimation. *IEEE Transactions on Signal Processing*, 60(12):6182–6189, December 2012.

[11] Ami Wiesel. Unified Framework to Regularized Covariance Estimation in Scaled Gaussian Models. *IEEE Transactions on Signal Processing*, 60(1):29–38, January 2012.