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) problem under the hermitian positive definite (HPD) constraint. First, we find that the ML estimator is geodesic convex (g-convex) which allows global optimization. Therefore, lots of convex optimization algorithms on manifolds can be applied to this problem. We will apply many manifold optimization methods 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 communications, bioinformatics, and financial engineering.

Classical covariance estimation 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 holds neither of these assumptions.

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 [12], matrix penalties [14], and smoothness penalties [3]. Another approaches is the shrinkage technique, which includes shrinkage to identity [5], shrinkage to diagonal structure [8], and shrinkage to positive definite matrix [14].

For the situation of samples from the distribution except for Gaussian distribution. An alternative to maximizing the likelihood is proposed in [6], which does not require assuming a particular probability distribution for the data. Another common study [11] is to replace the Gaussian assumption with a more general scaled Gaussian model, e.g., elliptical distributions. Tyler's ML estimator [7], a well-studied estimator, performs robustly in scaled Gaussian distribution, the Tyler's estimator defined as the solution to the fixed-point equation

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

is a minimax estimator. On the other hand, Tyler's ML estimator can also combine with the shrinkage techniques [14].

In many scenarios, 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. For example, Kronecker structures, also known as separable models, transposable covariances models, or matrix-variate-normal models are typically used when dealing with random matrices. Covariance matrix estimation in Kronecker structures involves a non-convex optimization. But its maximum likelihood estimation function is g-convex [13]. It is concluded that any local minimum of the cost function on a group Kronecker constraint set is a global minimum. Many numerical algorithms such as the majorization-minimization algorithm is also proposed to solve the constrained minimization problem [10].

II. Problem Formulation

Suppose a column vector x is drawn from an N-dimensional Gaussian distribution with zero mean and covariance matrix $\mathbf{R}$, the corresponding probability density function is [4]

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

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 (2) as

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = (2\pi)^{-NM/2}\det(\mathbf{R})^{-M/2}e^{-\sum_{i=1}^{M}\frac{\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i}{2}}. \tag{3}$$

The R that has a special structure and maximizes (3) is the "maximum-likelihood" estimate of the covariance matrix. In order to simplify the equation (3), we take the logarithm of it, we get

$$-\frac{MN}{2}\log(2\pi) - \frac{M}{2}\log\det(\mathbf{R}) - \frac{1}{2}\sum_{m=1}^{M}x_m^T\mathbf{R}^{-1}x_m. \tag{4}$$

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

$$\begin{aligned} \underset{\mathbf{R}}{\text{minimize}} \quad & \frac{1}{M}\log\det(\mathbf{R}) + \sum_{m=1}^{M}\mathbf{x}_m^T\mathbf{R}^{-1}\mathbf{x}_m \\ \text{subject to} \quad & \mathbf{R}\in \text{HPD}. \end{aligned} \tag{5}$$

The objective function in (5) is non-convex in Euclidean spaces. 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 scalar 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 (6)

$$\phi(\gamma(t)) \le (1-t)\phi(\gamma(0)) + t\phi(\gamma(1)) = (1-t)\phi(x) + t\phi(y). \tag{6}$$

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, we first define the $\mathbb{P}_d$ as 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 $ds=\|\mathbf{A}^{\frac{1}{2}}d\mathbf{A}\mathbf{A}^{-\frac{1}{2}}\|_F$[2, 6.1.6]. For $\mathbf{A},\mathbf{B}\in\mathbb{P}_d$ there is a unique geodesic path[9]

$$\gamma(t) = \mathbf{A}\#_t\mathbf{B} := \mathbf{A}^{\frac{1}{2}}(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}})^t\mathbf{A}^{\frac{1}{2}}, \qquad t\in[0,1]. \tag{7}$$

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 $A\#B$.

According to [1], the map $(\mathbf{A},\mathbf{B})\to\mathbf{A}\#_t\mathbf{B}$, $\mathbf{A},\mathbf{B}\in\mathbb{P}_d$ is concave. Then we have

$$\mathbf{A}\#_t\mathbf{B} \preceq (1-t)\mathbf{A} + t\mathbf{B}. \tag{8}$$

(where $\preceq$ denotes the Löwner Partial order)

In order to prove the convexity of the objective function in (5), we firstly focus on the last part. We are going to prove that $f(\mathbf{A})=\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x}$ is g-convex. To prove this is equal to verify the midpoint convexity:

$$f(\mathbf{A}\#\mathbf{B}) \le \frac{f(\mathbf{A})+f(\mathbf{B})}{2}, \tag{9}$$

where $\mathbf{A},\mathbf{B}\in\mathbb{P}_d$. Since $(\mathbf{A}\#\mathbf{B})^{-1}=\mathbf{A}^{-1}\#\mathbf{B}^{-1}$ and $\mathbf{A}^{-1}\#\mathbf{B}^{-1}\preceq\frac{\mathbf{A}^{-1}+\mathbf{B}^{-1}}{2}$ [2, 4.16]. Therefore

$$f(\mathbf{A}\#\mathbf{B}) = \mathbf{x}^T(\mathbf{A}^{-1}\#\mathbf{B}^{-1})\mathbf{x} \le \mathbf{x}^T\frac{\mathbf{A}^{-1}+\mathbf{B}^{-1}}{2}\mathbf{x} = \frac{\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x}+\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x}}{2} = \frac{f(\mathbf{A})+f(\mathbf{B})}{2}. \tag{10}$$

We prove convexity of the first part of objective function then,

$$\begin{aligned} f(\mathbf{A}\#\mathbf{B}) &= \log\det(\mathbf{A}^{\frac{1}{2}}(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}})^t\mathbf{A}^{\frac{1}{2}}) \\ &= \log\det(\mathbf{A}) + t\log\det(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}}) \\ &= (1-t)\log\det(\mathbf{A}) + t\log\det(\mathbf{B}) \\ &= (1-t)f(\mathbf{A}) + tf(\mathbf{B}). \end{aligned}$$

Therefore, the whole objective function is g-convex.

IV. Optimization

References

[1] T. Ando. Concavity of certain maps on positive definite matrices and applications to hadamard products. *Linear Algebra and its Applications*, 26:203 – 241, 1979.

[2] Rajendra Bhatia. *Positive Definite Matrices*. Princeton University Press, Princeton, 10 Jan. 2009.

[3] 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.

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

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] David E. Tyler. A Distribution-Free M-Estimator of Multivariate Scatter. *The Annals of Statistics*, 15(1):234–251, March 1987.

[12] 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.

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

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