Advances in Applied Mathematics 应用数学进展, 2024, 13(8), 3694-3701
Published Online August 2024 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2024.138352

Hans汉斯

JBLD均值的影响函数及鲁棒性分析

冯　莉

大连交通大学理学院，辽宁　大连

收稿日期：2024年7月7日；录用日期：2024年8月2日；发布日期：2024年8月7日

摘　要

影响函数描述了正定矩阵流形上的数据被异常值污染时，异常值对平均值估计精度的影响。本文给出了Jensen-Bregman LogDet (JBLD)均值矩阵的影响函数及鲁棒性分析。首先概述正定矩阵流形的几何结构，包括欧氏度量和JBLD度量及它们的测地距离和均值矩阵，进而计算JBLD均值的影响函数，并分析均值的稳定性。

关键词

JBLD，欧氏度量，影响函数，正定矩阵流形

Influence Function and Robustness Analysis of JBLD Mean Value

Li Feng

School of Science, Dalian Jiaotong University, Dalian Liaoning

Received: Jul. 7th, 2024; accepted: Aug. 2nd, 2024; published: Aug. 7th, 2024

Abstract

The influence function describes the influence of outliers on the estimation accuracy of the mean value when the data on the positive definite matrix manifold is contaminated by outliers. In this paper, the influence function of Jensen-Bregman LogDet (JBLD) mean matrix is given and its robustness is analyzed. Firstly, the geometric structure of positive definite matrix manifold is summarized, including Euclidean metric and JBLD metric, their geodesic distance and mean matrix, and then the influence function of JBLD mean is calculated, and the stability of mean is analyzed.

文章引用: 冯莉. JBLD 均值的影响函数及鲁棒性分析[J]. 应用数学进展, 2024, 13(8): 3694-3701.
DOI: 10.12677/aam.2024.138352

Keywords

JBLD, Euclidean Metric, Influence Function, Positive Definite Matrix Manifold

1. 引言

正定矩阵流形作为一般线性群的一个经典子流，在最优控制、信号处理、经济学、金融学、计算机科学等多个领域中都展现出了其独特的应用价值[1]-[5]。通过利用其独特的性质和数学结构，可以为解决实际问题提供有效的数学方法和工具。在正定矩阵流形上，可以定义多种距离和散度来量化正定矩阵之间的差异。常用的距离和散度包括欧氏距离、仿射不变黎曼距离、对数欧氏距离、Kullback-Leibler divergence 散度和 Symmetry Kullback-Leibler divergence 散度等。由于正定矩阵流形是一个非线性的黎曼流形，需要使用流形上的几何结构来定义均值，不同的距离或散度可以诱导出不同的均值矩阵定义。影响函数(Influence Function)在分析统计估计量的稳健性时非常有用，它衡量了单个数据点对估计量的影响。在正定矩阵流形上，当数据(即正定矩阵)被异常值污染时，影响函数可以量化这些异常值如何影响平均值的估计精度。许多专家学者致力于研究不同度量下矩阵均值的影响函数。文献[6]计算了正定矩阵流形上算术均值的影响函数。文献[7]分别计算了仿射不变黎曼距离、Total Square Loss 散度、Total Log-determiant 散度和 Total Von Neumann 所对应的影响函数。本文研究当正定矩阵流形上的数据点被异常值污染时，JBLD 均值矩阵的影响函数和稳定性。首先介绍正定矩阵流形上常用的度量、距离和均值矩阵；进而给出 JBLD 均值矩阵的影响函数，并给出固定点迭代算法；最后利用数值模拟验证 JBLD 均值矩阵比算术均值矩阵更具有鲁棒性。

2. 欧氏度量和 JBLD 度量的几何结构

设 A 为 n 阶正定矩阵，即满足 $A^{\mathrm{T}}=A$， $x^{\mathrm{T}}Ax\ge 0$ 当且仅当 $x=0$ 等号成立。由 n 阶正定矩阵全体构成的微分流形被称为正定矩阵流形，记为 $SPD(n)$。可以在 $SPD(n)$ 上定义不同的度量，进而得到不同的几何结构[8] [9]。

对于 $SPD(n)$ 矩阵 $X_1,\cdots,X_m$，与几何测度相关的几何均值估计量是几何均值平方和的最小值的唯一解。

$$\bar{X}=\arcmin\frac{1}{m}\sum_{i=1}^{m}d^2\left(X,X_i\right)。$$

首先，可以在 $SPD(n)$ 上定义欧氏内积

$$g\left(X,Y\right)=tr\left(X^{\mathrm{T}}Y\right)。$$

式中 X,Y 为流形上任一点的切向量，$tr(\cdot)$ 表示矩阵的迹。在欧氏内积下，过 $A,B\in SPD(n)$ 两点的测地线。

$SPD(n)$ 可以表示为

$$\gamma(t)=(1-t)A+tB。$$

上式中 t 为参数。由此可知，测地线 $\gamma(t)$ 并不完全落到 $SPD(n)$ 上。测地距离是流形上连接两个矩阵的最

短曲线的长度。由欧氏内积的式子可以得到 A,B 之间的测地距离

$$d(A,B)=\|A-B\|_F,$$

其中 $\|\cdot\|$ 为 Frobenius 范数，$\|X\|=\sqrt{tr(X^{\mathrm{T}}X)}$。

对于欧氏度量，$X_1,\cdots,X_m$ 的均值矩阵即为算术均值 $\bar{X}:=\frac{1}{m}\sum_{i=1}^{m}X_i$。

其次，两个矩阵 $X,Y\in SPD(n)$ 的 JBLD 定义为[5]

$$d_j^2(X,Y)=\ln\det\left(\frac{X+Y}{2}\right)-\frac{1}{2}\ln\det(XY)。$$

对于 JBLD，$X_1,\cdots,X_m$ 的均值矩阵为[7]

$$\bar{X}_{t+1}=\left(\frac{1}{K}\sum_{i=1}^{K}\left(\frac{\bar{X}_t+X_i}{2}\right)^{-1}\right)^{-1},\quad t=0,1,\cdots。$$

3. JBLD 均值矩阵的影响函数

影响函数在许多领域有重要作用，例如物理、化学、生物学等。此外影响函数还可以用于预测和建模，通过理解和分析影响函数，可以预测当自变量发生变化时，因变量可能会如何变化。这对于制定决策、规划未来和解决问题都非常有用。本章主要研究 JBLD 均值矩阵对应的影响函数。

对于 m 个 $SPD(n)$ 上的矩阵 $X_1,\cdots,X_m$，$\bar{X}$ 表示 $X_1,\cdots,X_m$ 的均值矩阵。$\hat{X}$ 表示这 m 个正定矩阵受到 n 个异常值的污染之后的均值矩阵。这 n 个异常值记为 $P_1,\cdots,P_n$，其中异常值的权重为 ε。那么，$\hat{X}$ 可以被定义为一个扰动

$$\hat{X}=\bar{X}+\varepsilon H(\bar{X},P_1,\cdots,P_n)+O(\varepsilon^2)。$$

定理 1：JBLD 均值矩阵对应的影响函数满足下面矩阵方程

$$\frac{1}{n}\sum_j\left[\left(\frac{\bar{X}+P_j}{2}\right)^{-1}-\bar{X}^{-1}\right]+\frac{1}{m}\sum_i\left[-\left(\frac{\bar{X}+X_i}{2}\right)^{-1}\frac{H}{2}\left(\frac{\bar{X}+X_i}{2}\right)^{-1}+\bar{X}^{-1}H\bar{X}^{-1}\right]=0。$$

证明：为了计算相关的影响函数，定义 $\hat{X}$ 为 $m+n$ 个 $SPD(n)$ 矩阵的最小化目标函数，即

$$\hat{X}=\arcmin\left[(1-\varepsilon)\frac{1}{m}\sum_{i=1}^{m}d^2(X,X_i)+\varepsilon\frac{1}{n}\sum_{j=1}^{n}d^2(X,P_j)\right]。$$

计算 $\bar{X}$ 的目标函数 $F(X)$ 为

$$F(X)=\frac{1}{m}\sum_{i=1}^{m}d^2(X,X_i)。$$

计算 $\hat{X}$ 的目标函数 $G(X)$ (异常值权重为 ε)为

$$G(X)=(1-\varepsilon)\frac{1}{m}\sum_i d^2(X,X_i)+\varepsilon\frac{1}{n}\sum_j d^2(X,P_j)。$$

从上式可以看出，为了计算 $G(X)$ 的梯度 $\nabla G(X)$，只需计算 $F(X)$ 的梯度 $\nabla F(X)$。设 $Z(t):=X+tY$，有

$$
\begin{aligned}
\langle \nabla F(X), Y\rangle &:= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} F(X+tY) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} F(Z(t)) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \frac{1}{m}\sum_{i=1}^{m} d_j^2(Z(t), X_i) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \frac{1}{m}\sum_{i=1}^{m}\left(\ln\left|\frac{Z(t)+X_i}{2}\right| - \frac{1}{2}\ln\left(|Z(t)|\cdot|X_i|\right)\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\ln\left|\frac{Z(t)+X_i}{2}\right| - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\ln\left(|Z(t)|\cdot|X_i|\right)\right)
\end{aligned}
$$

首先计算第一项，有

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\ln\left|\frac{Z(t)+X_i}{2}\right| &= \frac{1}{\left|\frac{Z(t)+X_i}{2}\right|}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left|\frac{Z(t)+X_i}{2}\right| \\
&= \frac{1}{\left|\frac{Z(t)+X_i}{2}\right|}\left|\frac{Z(t)+X_i}{2}\right| tr\left(\left(\frac{Z(t)+X_i}{2}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\left|\frac{Z(t)+X_i}{2}\right|\right) \\
&= tr\left(\left(\frac{Z(t)+X_i}{2}\right)^{-1}\cdot\frac{1}{2}\cdot\frac{\mathrm{d}Z(t)}{\mathrm{d}t}\right)
\end{aligned}
$$

令 $t=0$ 时，有

$$
\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\ln\left|\frac{Z(t)+X_i}{2}\right| = \frac{1}{2}tr\left(\left(\frac{X+X_i}{2}\right)^{-1}Y\right)。
$$

其次计算第二项，有

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\ln\left(|Z(t)|\cdot|X_i|\right) &= \frac{1}{|Z(t)|\cdot|X_i|}\frac{\mathrm{d}}{\mathrm{d}t}\left(|Z(t)|\cdot|X_i|\right) \\
&= \frac{1}{|Z(t)|}\frac{\mathrm{d}|Z(t)|}{\mathrm{d}t} \\
&= \frac{1}{|Z(t)|}\cdot|Z(t)|tr\left(Z(t)^{-1}\frac{\mathrm{d}Z(t)}{\mathrm{d}t}\right) \\
&= tr\left(Z(t)^{-1}\frac{\mathrm{d}Z(t)}{\mathrm{d}t}\right)
\end{aligned}
$$

令 $t=0$ 时，有

$$
\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\ln\left(|Z(t)|\cdot|X_i|\right) = tr\left(X^{-1}Y\right)。
$$

故

$$
\langle \nabla F(X), Y\rangle := \left\langle \frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{2}\left(\frac{X+X_i}{2}\right)^{-1} - \frac{1}{2}X^{-1}\right], Y\right\rangle。
$$

所以，

$$\nabla F(X)=\frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{2}\left(\frac{X+X_i}{2}\right)^{-1}-\frac{1}{2}X^{-1}\right]。$$

同理，可以得到

$$\nabla G(X)=(1-\varepsilon)\frac{1}{m}\sum_{i}\left[\frac{1}{2}\left(\frac{X+X_i}{2}\right)^{-1}-\frac{1}{2}X^{-1}\right]+\varepsilon\frac{1}{n}\sum_{j}\left[\frac{1}{2}\left(\frac{X+P_j}{2}\right)^{-1}-\frac{1}{2}X^{-1}\right]。$$

因为 $\hat{X}$ 为 $\nabla G(X)$ 的最小值点，所以 $\nabla G(\hat{X})=0$，即

$$(1-\varepsilon)\frac{1}{m}\sum_{i}\left[\frac{1}{2}\left(\frac{\hat{X}+X_i}{2}\right)^{-1}-\frac{1}{2}\hat{X}^{-1}\right]+\varepsilon\frac{1}{n}\sum_{j}\left[\frac{1}{2}\left(\frac{\hat{X}+P_j}{2}\right)^{-1}-\frac{1}{2}\hat{X}^{-1}\right]=0。$$

因为

$$\hat{X}=\bar{X}+\varepsilon H(\bar{X},P_1,\cdots,P_n)+O(\varepsilon^2)。$$

所以为了求 H，上式对 ε 求导，再令 $\varepsilon=0$，有

$$\begin{aligned}&-\frac{1}{m}\sum_{i}\left[\left(\frac{\hat{X}+X_i}{2}\right)^{-1}-\hat{X}^{-1}\right]+(1-\varepsilon)\frac{1}{m}\sum_{i}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\left(\frac{\hat{X}+X_i}{2}\right)^{-1}-\hat{X}^{-1}\right]\\&+\frac{1}{n}\sum_{j}\left[\left(\frac{\hat{X}+P_j}{2}\right)^{-1}-X^{-1}\right]+\varepsilon\frac{1}{n}\sum_{j}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\left(\frac{\hat{X}+P_j}{2}\right)^{-1}-\hat{X}^{-1}\right]=0\end{aligned}$$

令 $\varepsilon=0$（当 $\varepsilon=0$ 时，$\hat{X}=\bar{X}$），

$$-\frac{1}{m}\sum_{i}\left[\left(\frac{\bar{X}+X_i}{2}\right)^{-1}-\bar{X}^{-1}\right]+\frac{1}{m}\sum_{i}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\left(\frac{\hat{X}+X_i}{2}\right)^{-1}-\hat{X}^{-1}\right]+\frac{1}{n}\sum_{j}\left[\left(\frac{\bar{X}+P_j}{2}\right)^{-1}-\bar{X}^{-1}\right]=0。$$

因为

$$\nabla F(\bar{X})=\frac{1}{m}\sum_{i=1}^{m}\left[\left(\frac{\bar{X}+X_i}{2}\right)^{-1}-\bar{X}^{-1}\right]=0。\tag{1}$$

所以(1)变为

$$\frac{1}{n}\sum_{j}\left[\left(\frac{\bar{X}+P_j}{2}\right)^{-1}-\bar{X}^{-1}\right]+\frac{1}{m}\sum_{i}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\left(\frac{\hat{X}+X_i}{2}\right)^{-1}-\hat{X}^{-1}\right]=0。$$

根据文献[10]中的求导法则，可得

$$\begin{aligned}0&=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}I\\&=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\left(\frac{\hat{X}+X_i}{2}\right)^{-1}\left(\frac{\hat{X}+X_i}{2}\right)\right]\\&=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\hat{X}+X_i}{2}\right)^{-1}\left(\frac{\hat{X}+X_i}{2}\right)+\left(\frac{\hat{X}+X_i}{2}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\hat{X}+X_i}{2}\right)\end{aligned}$$

所以

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\hat{X}+X_i}{2}\right)^{-1}=-\left(\frac{\hat{X}+X_i}{2}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\hat{X}+X_i}{2}\right)\left(\frac{\hat{X}+X_i}{2}\right)^{-1}。$$

即

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\hat{X}^{-1}=-\hat{X}^{-1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\hat{X}\hat{X}^{-1}。$$

得到

$$\frac{1}{n}\sum_j\left[\left(\frac{\bar{X}+P_j}{2}\right)^{-1}-\bar{X}^{-1}\right]+\frac{1}{m}\sum_i\left[-\left(\frac{\hat{X}+X_i}{2}\right)^{-1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\frac{\hat{X}+X_i}{2}\right)\left(\frac{\hat{X}+X_i}{2}\right)^{-1}+\hat{X}^{-1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\hat{X}\hat{X}^{-1}\right]=0。$$

即，

$$\frac{1}{n}\sum_j\left[\left(\frac{\bar{X}+P_j}{2}\right)^{-1}-\bar{X}^{-1}\right]+\frac{1}{m}\sum_i\left[-\left(\frac{\bar{X}+X_i}{2}\right)^{-1}\frac{H}{2}\left(\frac{\bar{X}+X_i}{2}\right)^{-1}+\bar{X}^{-1}H\bar{X}^{-1}\right]=0。$$

证明完毕。

注意到定理 1 中 JBLD 均值矩阵的影响函数是隐式表达式，所以给出下面的固定点迭代算法。

算法 1：计算 JBLD 均值矩阵的影响函数的迭代算法为

$$H_{t+1}=\frac{2}{m}\sum_i\bar{X}\left(\bar{X}+X_i\right)^{-1}H_t\left(\bar{X}+X_i\right)^{-1}\bar{X}+\bar{X}-\frac{2}{n}\sum_j\bar{X}\left(\bar{X}+P_j\right)^{-1}\bar{X}$$

其中 $t=0,1,\cdots$，取 $H_0=E$。

4. 数值结果和讨论

下面通过数值实验比较欧氏度量和 JBLD 对应的影响函数和稳定性。在这些模拟中，$SPD(n)$ 上的正定矩阵产生的公式为：

$$\exp\left(\frac{V+V^{\mathrm{T}}}{2}\right)$$

其中 $V\in\mathbb{R}^{n\times n}$ 是指由 MATLAB 生成的随机矩阵。显然，$\frac{V+V^{\mathrm{T}}}{2}$ 是对称矩阵。

为了影响函数的计算不是一般性，考虑 100 个随机生成的对称正定矩阵使用算法 1 进行数值模拟，如图 1 所示。可以看出，几何度量 JBLD 对应的影响函数的范数对异常值的变化不敏感，它一直保持在接近 0.2 的范围内。另一方面，欧氏度量影响函数的范数波动显著。因此，正定矩阵流形上的 JBLD 均值比算术均值更稳定。

图 2 为图 1 中欧氏度量和 JBLD 度量的 100 次数值模拟的平均值，提供了一个更直观的表示。从图 2 可以看出，与使用 JBLD 度量获得的影响函数值相比，欧氏度量的影响函数具有更大的范数值。这进一步支持了 JBLD 均值比算术均值更具鲁棒性。

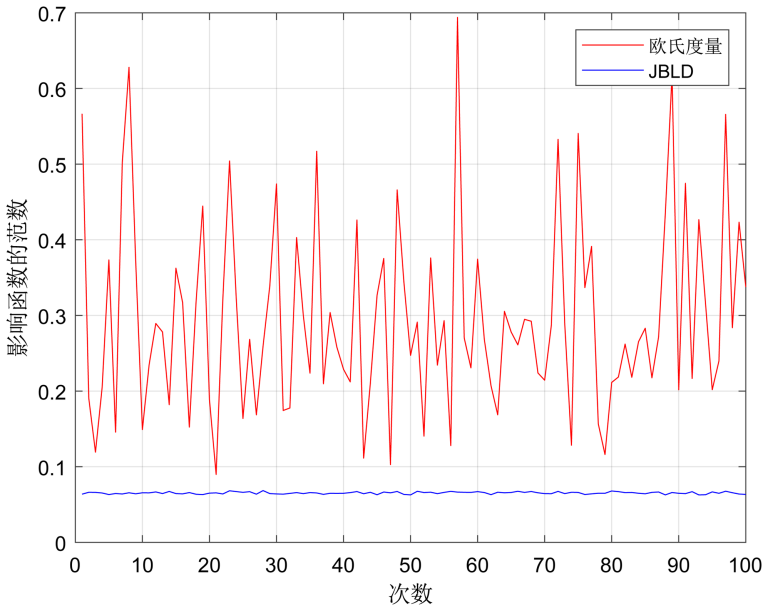

Figure 1. Norm of influence function
图 1. 影响函数的范数

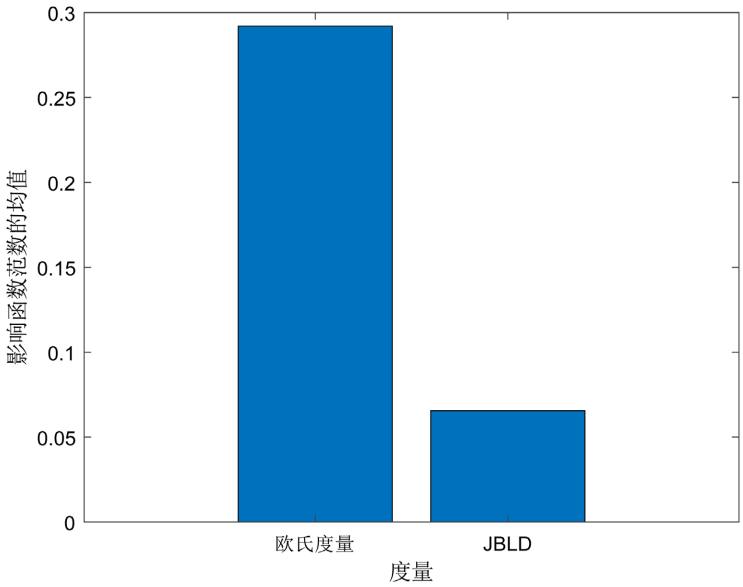

Figure 2. Mean of the influence function
图 2. 影响函数的均值

5. 结论

本文首先概述正定矩阵流形的几何结构，包括欧氏度量和 JBLD 度量及它们的测地距离和均值矩阵，进而计算 JBLD 均值的影响函数，并给出固定点迭代算法，JBLD 度量比欧氏度量在正定矩阵流形上增加了计算的难度，但最后利用数值模拟验证 JBLD 均值矩阵比算术均值矩阵更具有鲁棒性。

基金项目

国家自然科学基金项目(No. 61401058)；辽宁省教育厅科学研究经费项目(JYTMS20230010)。

参考文献

[1] Ono, Y. and Peng, L. (2022) Towards a Median Signal Detector through the Total Bregman Divergence and Its Robustness Analysis. *Signal Processing*, **201**, Article 108728. https://doi.org/10.1016/j.sigpro.2022.108728

[2] 孙华飞, 张真宁, 彭林玉, 等. 信息几何导引[M]. 北京: 科学出版社, 2016.

[3] Hua, X., Ono, Y., Peng, L., Cheng, Y. and Wang, H. (2021) Target Detection within Nonhomogeneous Clutter via Total Bregman Divergence-Based Matrix Information Geometry Detectors. *IEEE Transactions on Signal Processing*, **69**, 4326-4340. https://doi.org/10.1109/tsp.2021.3095725

[4] 孙华飞, 宋扬, 罗翼昊, 等. 一种基于统计流形的聚类算法[J]. 北京理工大学学报, 2021, 41(2): 226-230.

[5] Hua, X. and Peng, L. (2021) MIG Median Detectors with Manifold Filter. *Signal Processing*, **188**, Article 108176. https://doi.org/10.1016/j.sigpro.2021.108176

[6] Hua, X., Ono, Y., Peng, L., Cheng, Y. and Wang, H. (2021) Target Detection within Nonhomogeneous Clutter via Total Bregman Divergence-Based Matrix Information Geometry Detectors. *IEEE Transactions on Signal Processing*, **69**, 4326-4340. https://doi.org/10.1109/tsp.2021.3095725

[7] Hua, X., Cheng, Y., Wang, H. and Qin, Y. (2018) Robust Covariance Estimators Based on Information Divergences and Riemannian Manifold. *Entropy*, Article 219. https://doi.org/10.3390/e20040219

[8] Duan, X., Sun, H. and Peng, L. (2020) Application of Gradient Descent Algorithms Based on Geodesic Distances. *Science China Information Sciences*, **63**, Article No. 152201. https://doi.org/10.1007/s11432-019-9911-5

[9] Moakher, M. (2005) A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. *SIAM Journal on Matrix Analysis and Applications*, **26**, 735-747. https://doi.org/10.1137/s0895479803436937

[10] 张贤达. 矩阵分析与应用[M]. 北京: 清华大学出版社, 2014.