Log-Euclidean Kernels for Sparse Representation and Dictionary Learning

Peihua Li[1], Qilong Wang[2], Wangmeng Zuo[3,4], Lei Zhang[4]

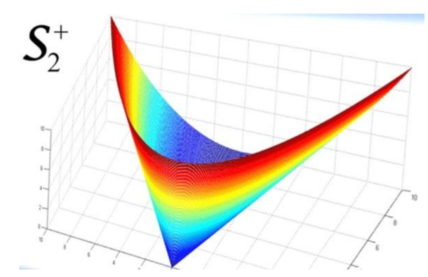
[1]Dalian University of Technology, [2]Heilongjiang University, [3]Harbin Institute of Technology, [4]The Hong Kong Polytechnic University

Introduction

Background

Recently there are growing interests in studying sparse representation (SR) and dictionary learning (DL) of symmetric positive definite (SPD) matrices.

The space of n-by-n SPD matrices S_n^+ is not a linear space but a Lie group that forms a Riemannian manifold.



Comparison of State-of-the-arts on SR and DL in S_n^+

Method	Representation given atoms	Riemannian Metric?	Riemannian atom update?	Mercer's condition?
TSC [ECCV10,ICCV11]	Linear in Euclidean space	No-LogDet divergence	No-Euclidean	N/A
GDL [ECML12]	Linear in Euclidean space	No-Frobenius norm	No-Euclidean	N/A
LogE-SR [ACCV 10]	Linear in Log-domain	Yes	No-Euclidean	N/A
RSR [ECCV 12]	Linear in RKHS	Approximation-Stein divergence	No-Euclidean	Satisfy-conditionally
Proposed method	Linear in RKHS	Yes	Yes-Riemannian	Satisfy

1. Previous work fails to exploit the geometry of S_n^+, using the Euclidean norm or Bregman divergence to evaluate the reconstruction error.
2. The dictionary atoms are updated without taking account of the geometric structure of S_n^+.
3. Linear decomposition makes sense in high- or infinite-dimensional RKHS In [Harandi *et al*. ECCV 12]; however, the Stein divergence is only an approximation of Riemannian metric and satisfy Mercer's condition under some restricted conditions.

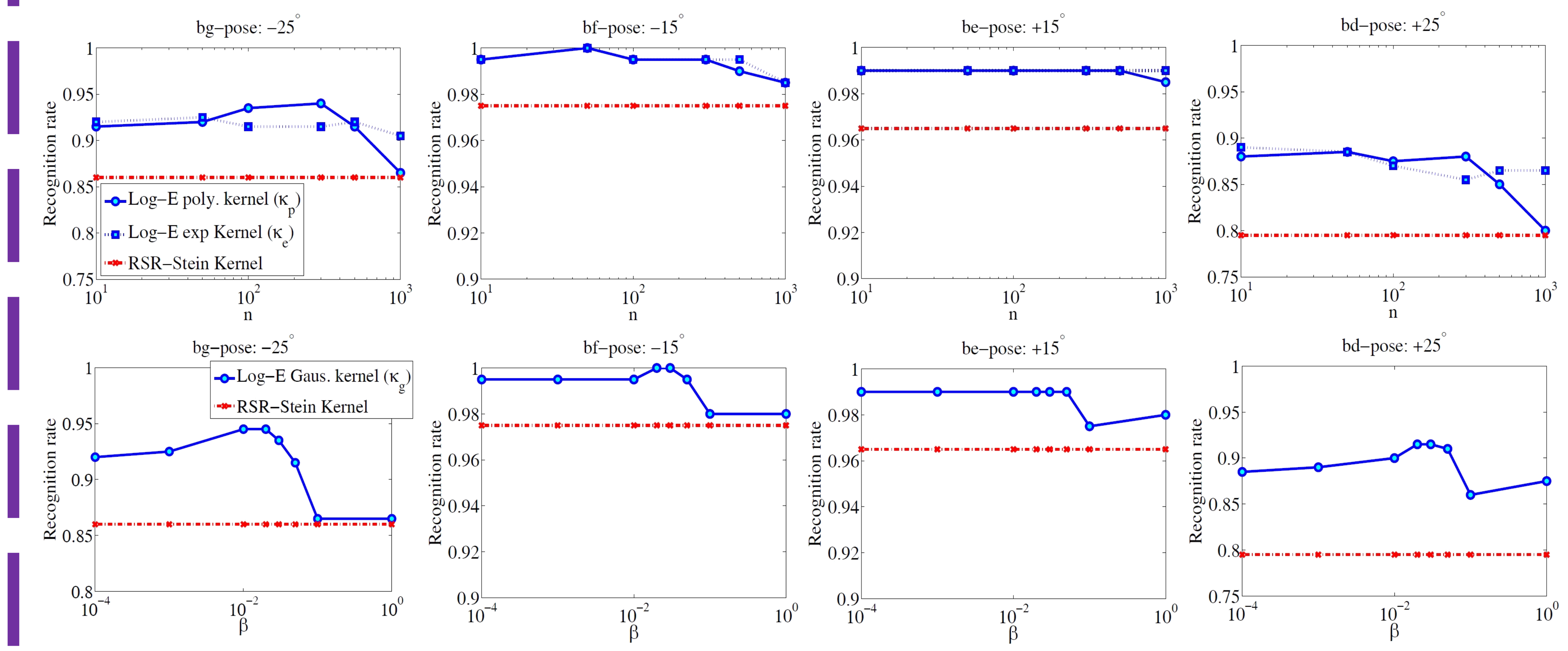
Key Idea

Our work is inspired by [Harandi *et al*. ECCV 12] and is also kernel-based. We develop a family of kernel functions based on the Log-Euclidean framework .

The main differences:

- Characterizing the geodesic distance and so accurately measuring the reconstruction error;
- Satisfying Mercer's condition under broad conditions;
- DL that consider geometric structure of S_n^+ .

Parameters Evaluation



Parameters Evaluation on the FERET face dataset.

The classification rates of RSR that uses the Stein kernel [Harandi *et al.* ECCV 12] are shown as baseline (red dash-dotted line).

Method Description

S_n^+ is a complete inner product space

Corollary With two operations $\oplus$ and $\otimes$, the function from the product space of S_n^+ to the space R of real number

$$\langle \cdot,\cdot \rangle_{\log} : S_n^+ \times S_n^+ \mapsto \mathrm{R}$$

$$\langle \mathbf{S},\mathbf{T} \rangle_{\log} = \mathrm{tr}(\log(\mathbf{S})\log(\mathbf{T}))$$

is an inner product.

- $\langle \cdot,\cdot \rangle_{\log}$ satisfies the properties of symmetry, linearity, & non-negativity.
- The induced norm $\|\mathbf{S}\|_{\log} = \langle \mathbf{S},\mathbf{S} \rangle_{\log}^{1/2}$ can be used to define the distance that equals to the geodesic distance.
- S_n^+ is complete.
- $\langle \mathbf{S},\mathbf{T} \rangle_{\log,\mathbf{A}} = \mathrm{tr}(\log(\mathbf{S})\mathbf{A}\log(\mathbf{T}))$ is an inner product as well, where **A** is a SPD matrix.

$$\lambda \otimes \mathbf{S} = \exp(\lambda \log(\mathbf{S})) = \mathbf{S}^{\lambda} \quad \mathbf{S} \oplus \mathbf{T} = \exp(\log(\mathbf{S}) + \log(\mathbf{T}))$$

Kernelized SR & DL for SPD Matrices

$$f(\underbrace{\mathbf{S}_1,\ldots,\mathbf{S}_N}_{\text{Dictionary}}, \underbrace{\mathbf{x}_j,\ldots,\mathbf{x}_M}_{\text{Sparse code}}) = \sum_{j=1}^{M} \left\| \phi(\mathbf{Y}_j) - \sum_{i=1}^{N} x_{j,i}\phi(\mathbf{S}_i) \right\|_2^2 + \lambda \|\mathbf{x}_j\|_1$$

SPD matrices ($\mathbf{Y}_j$, $\mathbf{S}_i$)

$$\min_{\mathbf{x}\in \mathrm{R}^N} -2\sum_{i=1}^{N} x_i \kappa(\mathbf{Y}_i,\mathbf{S}_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \kappa(\mathbf{S}_i,\mathbf{S}_j) + \lambda\|\mathbf{x}\|_1$$

Let p_n *be* a polynomial of degree n≥1 with positive coefficients, we have p.d. kernels

- Log-E poly. kernel $\kappa_{p_n}(\mathbf{S},\mathbf{T}) = p_n(\langle \mathbf{S},\mathbf{T}\rangle_{\log})$
- Log-E exp. kernel $\kappa_{e_n}(\mathbf{S},\mathbf{T}) = \exp(p_n(\langle \mathbf{S},\mathbf{T}\rangle_{\log}))$
- Log-E Gaussian kernel $\kappa_g(\mathbf{S},\mathbf{T}) = \exp(-\mathrm{tr}((\log(\mathbf{S})-\log(\mathbf{T}))\mathbf{A}(\log(\mathbf{S})-\log(\mathbf{T}))))$

Atoms update

$$\mathbf{S}_r = \exp\left(\log(\mathbf{S}_r) + d_{\mathbf{S}_r}\log\left(-\varepsilon\left(\partial f / \partial \mathbf{S}_r\right)\right)\right)$$

$$\frac{\partial f}{\partial \mathbf{S}_r} = -2\beta\mathbf{S}_r^{-1}\left(\sum_{j=1}^{M} x_{j,r}\kappa(\mathbf{S}_r,\mathbf{Y}_j)(\log(\mathbf{S}_r)-\log(\mathbf{Y}_j)) - \sum_{j=1}^{M}\sum_{i=1}^{N} x_{j,r}x_{j,i}(\log(\mathbf{S}_r)-\log(\mathbf{S}_j))\right)$$

Main Contributions

(1) In the Log-Euclidean framework, we disclosed S_n^+ is a complete inner product space, and developed a broad family of p.d. kernels.

- Characterize the geodesic distance between SPD matrices.
- Satisfy the Mercer's condition in general conditions.

(2) Dictionary atoms are updated in Riemannian space.

(3) Experiments have shown the superiority of our method to state-of-the-arts.

Classifiers

(1) The residual error approach for classification:

$$label(\mathbf{Y}) = \min_i \varepsilon_i(\mathbf{Y}) \qquad \varepsilon_i(\mathbf{Y}) = \left\|\phi(\mathbf{Y}) - \sum_{j=1}^{N} x_j\phi(\mathbf{S}_j)\delta_i(j)\right\|_2^2$$

$$\delta_i(j) = 1, \text{if } j \in \text{class } i; \text{otherwise } \delta_i(j) = 0.$$

(2) We learn the sparse codes obtained from the predefined atom matrices are used for classification with the nearest neighbor classifier or support machine vector (SVM).

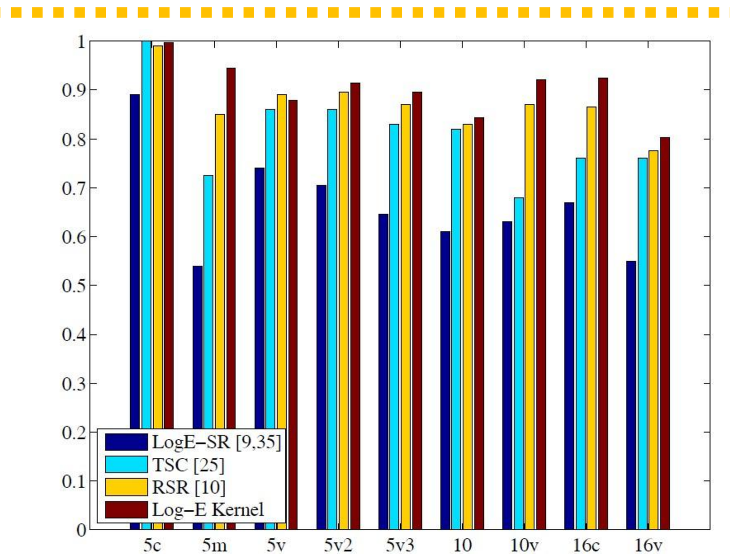
Reference V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. on Matrix Analysis and Applications, 2006.

M. T. Harandi, C. Sanderson, R. Hartley, and B. C. Lovell. Sparse coding and dictionary learning for symmetric positive definite matrices: a kernel approach. In ECCV(2), 2012.

Experimental Evaluation

Sparse Representation

Average rates on all nine mosaics from the Brodatz dataset are 0.66, 0.81, 0.87, and 0.92 for LogE-SR, TSC, RSR, and Log-E Kernel, respectively

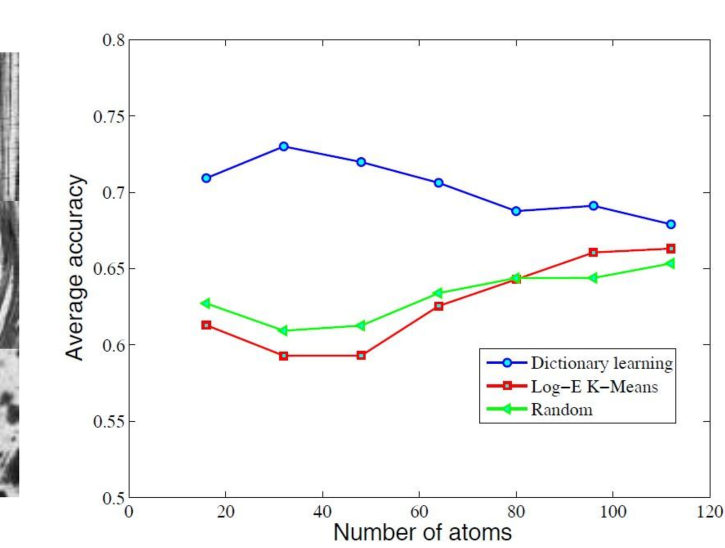


Classification rates on nine mosaics from the Brodatz dataset.

	SRC [PAMI 09]	GSRC [ECCV 10]	LogE-SR [ACCV 09]	TSC [ECCV 10]	RSR [ECCV 12]	Log-E Kernel		
						K_{pn}	K_{en}	K_g
bg	26.0	79.0	46.5	44.5	86.0	92.0	91.5	**94.5**
bf	61.0	97.0	91.0	73.5	97.5	100	99.5	**100**
be	55.5	93.5	81.0	73.0	96.5	99.0	99.0	**99.0**
bd	27.5	77.0	34.5	36.0	79.5	88.5	88.0	**91.5**
ave.	42.5	86.6	63.3	56.8	89.9	94.9	94.5	**96.3**

Comparison with state-of-the-arts on the FERET database

Dictionary Learning



Texture Classification on the Brodatz dataset

Num. of atoms	32	64	128
Random dictionary	44.80±0.90	57.64±0.59	62.25±0.65
LogE K-means	67.69±0.56	76.25±0.48	78.80±0.53
Dictionary learning	**75.84±0.64**	**79.27±0.65**	**80.92±0.44**

Scene Classification on the Scene15 dataset