

Kernel Methods and Its Application in Wavefront Reconstruction

Zhiying Tan, Ying Chen, Kun She

School of Computer Science & Engineering, University
of Electronic Science & Technology of China,
Chengdu, China
e-mail: tanzhiying1010@gmail.com

Yong Feng

Chongqing Inst. of Green and Intelligent Technology,
Chinese Academy of Sciences,
Chongqing, China
e-mail: yongfeng@casit.ac.cn

Abstract—Kernel methods can effectively deal with the nonlinear problem. The methods not only can be used for data de-noising, also be effective for classification problems. Using kernel PCA method, we provide a more precise Zernike expansion, which can apparently improve the reconstruction accuracy. At the same time, explore learning the kernel function by the alignment. We verify that the alignment value and recognition rate is proportional relationship.

Keywords—Kernel PCA; Adaptive optics; Zernike polynomials; Alignment

I. INTRODUCTION

Kernel methods are a family of machine learning algorithms based on the statistical theory and kernel technology [1]. Kernel methods can discover more accurate dependencies among the data via constructing nonlinear mapping relations [2]. Kernel plays an important role in kernel methods, because the calculation result is completely determined by it.

The kernel alignment is a measurement which can measure the similarity between matrices and kernels. Alignment was first introduced by Cristianini et al. [3]. It can be seen as the generalized cosine between matrices. When the sample set is identified, kernel alignment can measure the similarity between the kernel functions. Another definition of kernel alignment based on the feature space was given by Cortes et al. [4]. And some kernel functions were learned by the centred alignment. Some multi-class problems and regression problems can be solved by the SVMs based on alignment [5]. We provide a learning method by the alignment.

Using the kernel PCA, the principal component can be effectively obtained. Some methods were provided to solve the pre-image problems. The standard gradient ascent methods were used to solve the optimization problem of minimizing the reconstruction error by the Mika et al. [6]. To obtain the wave front, we will also use the iterative method.

Adaptive optics (AO) can effectively improve the capability of optical systems by actively compensating for aberrations [7]. The atmospheric turbulence, optical fabrication errors and the thermally induced distortions may cause the aberrations. The residual wave front error in AO systems includes the wave front reconstruction errors, servo lag errors, and errors due to CCD noise [8]. There are three common methods for the wave front reconstruction task: the

zonal method [9], the modal method [10] and the direct gradient method [11].

In the modal method, the wavefront can be expanded into a set of orthogonal basis functions. And the expression coefficients can be estimated from the discrete phase slope measurements [12]. One often takes the Zernike polynomials as the wave front basis functions.

In the paper, we apply the learning methods of the wave front reconstruction of AO and the classification problems. The experiment results show that the nonlinear kernel can obtain very good wavefront reconstruction results. And verify that using the alignment can learn an optimal kernel function by numerical experiments.

II. ALGORITHM AND THEORETICAL RESULTS

A. Kernel PCA

For the sample set $S = \{x_1, x_2, \dots, x_N\} \subset R^n$, we first map the samples x_1, \dots, x_N into a feature space F by

$$\phi: R^n \rightarrow F, \quad x \mapsto X \quad (1)$$

Usually the function Φ is nonlinear.

The data can be centred by [6]

$$\tilde{\phi}(x_i) := \phi(x_i) - \frac{1}{N} \sum_{i=1}^N \phi(x_i) \quad (2)$$

The covariance matrix \tilde{K} , $\tilde{K}_{ij} = (\tilde{\phi}(x_i) \cdot \tilde{\phi}(x_j))$ in F .

In the paper, we choice the radial basis function

$$k(x, y) = e^{-\|x-y\|^2/2\sigma^2} \quad (3)$$

as the kernel.

For calculating the principal components in feature space F , we should firstly solve the characteristic equations $\tilde{K}\alpha = \lambda\alpha$. The solutions α^k ($k=1, 2, \dots, p$) belong to the nonzero eigenvalues λ_k ($k=1, 2, \dots, p$). Then the eigenvectors $V^k \in F \setminus \{0\}$ can be obtained

$$V^k = \sum_{i=1}^N \alpha_i^k \phi(x_i) \quad (4)$$

The solutions α^k can be normalized by requiring that the corresponding vector V^k is normalized, i.e. $(V^k, V^k) = 1$. For the convenient, we still use α^k to mark the normalized solutions.

For a new sample $\phi(x)$, the projection can be expressed as

$$P_p \phi(x) = \sum_{k=1}^p w_k V^k \quad (5)$$

where $w_k = (\phi(x) \cdot V^k) = \sum_{m=1}^N \alpha_m^k (\phi(x) \cdot \phi(x_m))$, $k = 1, \dots, p$,

the projection operator P_p .

The pre-image problem can be solved by calculating an optimal approximation vector z in input space R^n .

$$\min_z \rho(z) = \min_z \|\phi(z) - P_p \phi(x)\|^2 \quad (6)$$

Replacing terms independent of z by Ω , we obtain

$$\begin{aligned} \rho(z) &= (\phi(z) - P_p \phi(x)) \cdot (\phi(z) - P_p \phi(x)) \\ &= (\phi(z) \cdot \phi(z)) - 2(\phi(z) \cdot P_p \phi(x)) + (P_p \phi(x) \cdot P_p \phi(x)) \quad (7) \\ &= k(z, z) - 2 \sum_{k=1}^p w_k \sum_{m=1}^N \alpha_m^k k(z, x_m) + \Omega \end{aligned}$$

Obviously, the radial basis function $k(x, x) \equiv const$ for all x . To solve the optimization problem (6), we utilize the standard gradient descent methods.

For a maximum value, the gradient with respect to z vanish

$$\nabla_z \rho(z) = -4 \sum_{m=1}^N \sum_{k=1}^p w_k \alpha_m^k k'(\|z - x_m\|^2)(z - x_m) = 0 \quad (8)$$

This leads to a necessary condition for the maxima

$$z = \frac{\sum_{m=1}^N \sum_{k=1}^p w_j \alpha_m^k k'(\|z - x_m\|^2) x_m}{\sum_{m=1}^N \sum_{k=1}^p w_j \alpha_m^k k'(\|z - x_m\|^2)}$$

According to the Hilbert space theory [14] [15], we can devise an iteration scheme for z by

$$z(t) = \frac{\sum_{m=1}^N \sum_{k=1}^p w_j \alpha_m^k k'(\|z(t-1) - x_m\|^2) x_m}{\sum_{m=1}^N \sum_{k=1}^p w_j \alpha_m^k k'(\|z(t-1) - x_m\|^2)} \quad (9)$$

with $z(0) = x$.

B. Alignment Definitions

For the sample set $S = \{x_1, x_2, \dots, x_N\} \subset R^n$, we denote the samples' label vector as $y = [y_1, y_2, \dots, y_N]^T$, where $y_i \in \{-1, +1\}$, $(i = 1, 2, \dots, N)$. Assume the samples obey the same probability distribution D . And the kernels k_1 and k_2 are positive semi-definite and symmetrical (PDS). The alignment can be written as [3]

$$\hat{A}_s(K_1, K_2) = \frac{\langle K_1, K_2 \rangle_F}{\sqrt{\langle K_1, K_1 \rangle_F \langle K_2, K_2 \rangle_F}} \quad (10)$$

where $K_1, K_2 \in R^{N \times N}$ denote the kernel matrices. And the inner production between matrices is defined by

$$\langle K_1, K_2 \rangle_F = \sum_{i,j=1}^N k_1(x_i, x_j) k_2(x_i, x_j) = trace(K_1^T K_2)$$

It is easy to find that alignment $\hat{A}_s(K_1, K_2) \in [-1, 1]$, especially $\hat{A}_s(K_1, K_2) \in [0, 1]$ when $K_1, K_2 \geq 0$.

We can learn the kernel by maximizing the objective function (11)

$$\begin{aligned} \hat{A}_s(K, yy') &= \frac{\langle K, yy' \rangle_F}{\sqrt{\langle K, K \rangle_F \langle yy', yy' \rangle_F}} \\ &= \frac{\langle K, yy' \rangle_F}{N \sqrt{\langle K, K \rangle_F}} \end{aligned} \quad (11)$$

C. Wave Front Reconstruction

The Zernike polynomials are defined here by [13]

$$\begin{aligned} Z_{even j} &= \sqrt{n+1} R_n^m(r) \sqrt{2} \cos(m\theta) \\ Z_{odd j} &= \sqrt{n+1} R_n^m(r) \sqrt{2} \sin(m\theta) \end{aligned} \quad \left. \vphantom{\begin{aligned} Z_{even j} \\ Z_{odd j} \end{aligned}} \right\} m \neq 0 \quad (12)$$

$$Z_j = \sqrt{n+1} R_n^0(r), \quad m = 0$$

where

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s! [(n+m)/2-s]! [(n-m)/2-s]!} r^{n-2s} \quad (13)$$

The values of n and m are always integral and satisfy $m < n$, $n - |m| = even$. The index j is a mode ordering number and is a function of n and m .

A wave front phase $\phi(r)$ over a circular aperture of unit radius can be expressed as follows:

$$\phi(r) = \sum_k a_k Z_k(r) \quad (14)$$

where k is the conventional ordering of the Zernike basis.

In this paper the discrete phase slope measurements contain some white noises. Those make the Zernike expression coefficients contain large errors. Kernel PCA is powerful in de-noising [6]. We use it to remove some noises in the discrete phase slope measurements. And in modal wave front reconstruction, we use the least squares criterion to derive the optimal Zernike expression.

III. LEARNING COMBINED KERNEL

In the paper, we take the Gaussian kernels, because their performance is relatively stable. There is only one important parameter in the Gaussian function. In this section, we will learn the classifier function from the space $H = \{uk_1 + \sqrt{1-u^2}k_2 : u \in [0, 1]\}$, where the two kernels k_1 and k_2 are the PDS, and they are the Gaussian function with the undefined bandwidths σ_1 and σ_2 . Denote the classifier function as $k = uk_1 + \sqrt{1-u^2}k_2$, which has the equivalent form with $k = u_1k_1 + u_2k_2$, where $u_1^2 + u_2^2 = 1$ and $u_1 \in [0, 1]$. This means that the kernels k_1 and k_2 have the symmetry position. Then we can assign the parameter σ_1 as a small fixed value. The parameter $\sigma_2 \in (\sigma_1, +\infty)$ is unknown and need to be selected for the Gaussian kernel.

Then the function in space H only include two variables $u \in [0,1)$ and $\sigma_2 \in (\sigma_1, +\infty)$.

$$F(u) = \frac{y'(uK_1 + \sqrt{1-u^2}K_2)y}{N\sqrt{\text{tr}\left((uK_1 + \sqrt{1-u^2}K_2)^2\right)}} \quad (15)$$

For the binary classification problem, the optimal bandwidth of Gaussian and the combination coefficient u can be obtained by calculating the maximum of function (15).

IV. EXPERIMENTS

A. Wave Front Reconstruction

We use the kernel methods to de-noising for wave front of the AO. In Fig.1 shows the Zernike expressions under the conditions of the wave front with no-noises, noises and de-noises. In order to test the effect of de-noising, the numerical simulation experiments are carried out under the different conditions. In the experiments, we get the Zernike expressions including 27 items by calculating the phase screen (256*256). Firstly, we get 100 samples by random sampling. And then trained the 100 samples and extracted the nonlinear features by kernel PCA analysis. We performed kernel PCA with these data using Gaussian kernels. Lastly, we applied our algorithm for de-noising.

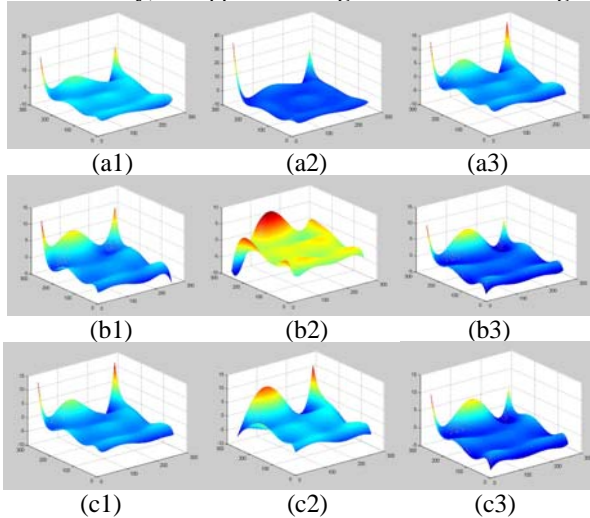


Fig. 1 Result of the de-noising experiments: (a1), (b1) and (c1): No-noise phase screens, (a2), (b2) and (c2): Having noise phase screens, (a3), (b3) and (c3): De-noising phase screens by kernel PCA.

In Fig.1, the figures (a2), (b2) and (c2) have some larger errors than the figures (a1), (b1) and (c1), and the figures (a3), (b3) and (c3) are obtained by de-noise using Kernel PCA.

In Table.I, S1-S5 are 5 different wave fronts. From the results we can find that the de-noise effect is very obvious.

TABLE. I RESULT OF THE DE-NOISING EXPERIMENTS (MSE): MSE INDICATES THE MEAN SQUARED ERROR BETWEEN THE RECONSTRUCTED PHASE SCREEN AND THE ORIGINAL PHASE SCREEN.

Norm	S1	S2	S3	S4	S5
MSE (noise)	17.82	27.49	20.34	35.32	20.06
MSE (de-noise)	4.42	3.37	2.36	2.353	5.58

B. Classification Problems

To verify the selected parameters' generalization capability, we do some numerical experiments on some data set in Fig 2. The data set is from some printed matter.

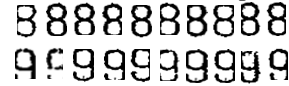


Fig. 2 Training samples being used in the experiment

In the experiments, we take the constant $\sigma_1 = 50$ and the vector $\sigma_2 = 100:50:4e3$. In Fig 4 shows the training sample set.

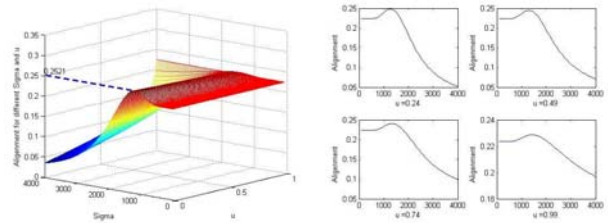


Fig. 3 Alignment values vary with parameter u and bandwidth σ

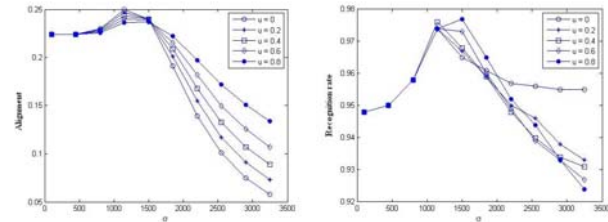


Fig. 4 Alignment values and recognition rates vary with parameter u and bandwidth σ

In Fig. 3 and Fig. 4, show the alignment values and recognition rates vary with parameter u and bandwidth σ . In Table. II and Table. III, provide more detailed experiment results. The results clearly show that the alignment values and recognition rates are proportional.

V. CONCLUSION

Kernel PCA is one of the most powerful techniques for feature extraction and de-noising. We can find that it is very effective in wave front reconstruction from the experiments in section IV. And the kernel can be learned by the alignment, because the alignment values and recognition rates are proportional.

ACKNOWLEDGMENT

This work was partially supported by the National Science and Technology Support Program (2012BAH29F00), the National Natural Science Foundation of China (61105111), and the Jiangsu Province Science and Technology Support Program (BE2012062).

REFERENCES

- [1] Z. Ghahramani, Unsupervised learning. Adv. Lect. Machine Learning, 72 (2004).
- [2] L. K. Saul and S. T. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds. J. Mach. Learn. Res. 4, 119 (2003).
- [3] N. Cristianini, J. Shawe-Taylor, Elisseeff A, et al. On kernel-target alignment [J]. In Advances in Neural Information Processing Systems (NIPS). 2002, 14: 367–373.
- [4] C. Cortes, M. Mohri, A. Roatamizadeh. Two-stage learning kernel algorithms. Proc. of the 27th IICML (2010).
- [5] H. Saigo, J. P. Vert, et al. Protein homology detection using string alignment kernels. Bioinformatics, 2004, 20(11): 1682-1689.
- [6] S. Mika, B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch. Kernel PCA and de-noising in feature spaces. In Advances in Neural Information Processing Systems 11, 1999,783-789.
- [7] R. K. Tyson. Adaptive optics engineering handbook. (CRC Press 1999)
- [8] A. Vyas, M. B. Roopashree, B. R. Prasad. Extrapolating Zernike Moments to Predict Future Optical Wave-fronts in Adaptive Optics Using Real Time Data Mining,(2010)
- [9] H. E. Rushmeier, K. E. Torrance. The zonal method for calculating light intensities in the presence of a participating (1987)
- [10] G. M. Dai. Modal wave-front reconstruction with Zernike polynomials and Karhunen-Loève functions.(1996)
- [11] W. H. Jiang, H. G. Li. Hartmann-Shack wavefront sensing and wavefront control algorithm. SPIE., 1271: 8293(1990)
- [12] M. Stangalini, D. D. Moro, F. Berrilli, O. Lühe. Zernike basis optimization for solar adaptive optics by using information theory. (2010)
- [13] R. J. Noll. Zernike polynomials and atmospheric turbulence. 1976.
- [14] B. Schölkopf, A. Smola, K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. (1998)
- [15] S. Saitoh. Theory of reproducing kernels and its Applications. Longman Scientific & Technical, Harlow, England (1988)
- [16] B. Schölkopf. Support vector learning. Oldenbourg Verlag, Munich,(1997)

TABLE. II CALCULATED ALIGNMENT VALUES WITH DIFFERENT PARAMETERS u AND BANDWIDTH σ

$u \backslash \sigma$	100	450	800	1150	1500	1850	2200	2550	2900	3250
0	0.224	0.224	0.230	0.250	0.240	0.191	0.139	0.101	0.075	0.058
0.1	0.224	0.224	0.230	0.248	0.240	0.196	0.147	0.109	0.083	0.065
0.2	0.224	0.224	0.229	0.247	0.240	0.201	0.155	0.117	0.091	0.073
0.3	0.224	0.224	0.229	0.245	0.240	0.205	0.162	0.125	0.099	0.083
0.4	0.224	0.224	0.228	0.243	0.240	0.209	0.168	0.133	0.107	0.089
0.5	0.224	0.224	0.228	0.242	0.240	0.212	0.175	0.142	0.116	0.097
0.6	0.224	0.224	0.227	0.24	0.239	0.215	0.182	0.150	0.126	0.107
0.7	0.224	0.224	0.227	0.239	0.238	0.219	0.189	0.160	0.137	0.119
0.8	0.224	0.224	0.226	0.236	0.237	0.222	0.197	0.172	0.151	0.134
0.9	0.224	0.224	0.226	0.234	0.235	0.225	0.207	0.189	0.172	0.157

TABLE. III CALCULATED RECOGNITION RATES WITH DIFFERENT PARAMETERS u AND BANDWIDTH σ

$u \backslash \sigma$	100	450	800	1150	1500	1850	2200	2550	2900	3250
0	0.948	0.950	0.958	0.974	0.965	0.961	0.957	0.956	0.955	0.955
0.1	0.948	0.950	0.958	0.975	0.966	0.958	0.953	0.949	0.945	0.941
0.2	0.948	0.950	0.958	0.974	0.967	0.959	0.950	0.946	0.938	0.933
0.3	0.948	0.950	0.958	0.974	0.967	0.959	0.949	0.944	0.935	0.932
0.4	0.948	0.950	0.958	0.976	0.968	0.959	0.948	0.94	0.934	0.931
0.5	0.948	0.950	0.958	0.974	0.970	0.958	0.948	0.938	0.934	0.931
0.6	0.948	0.950	0.958	0.974	0.973	0.960	0.949	0.939	0.933	0.927
0.7	0.948	0.950	0.958	0.974	0.974	0.961	0.950	0.940	0.932	0.924
0.8	0.948	0.950	0.958	0.974	0.977	0.965	0.952	0.944	0.933	0.924
0.9	0.948	0.950	0.958	0.972	0.979	0.97	0.958	0.946	0.939	0.925