IMPROVED EIGENVALUE SHRINKAGE USING WEIGHTED CHEBYSHEV POLYNOMIAL APPROXIMATION

Masaki Onuki[†,a], *Yuichi Tanaka*[†,b], *and Masahiro Okuda*[‡]

[†]Graduate School of BASE, Tokyo University of Agriculture and Technology, Tokyo, Japan
[a]Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong
[b]PRESTO, Japan Science and Technology Agency, Saitama, Japan
[‡]Faculty of Environmental Engineering, The University of Kitakyushu, Fukuoka, Japan
Email: masaki.o@msp-lab.org, ytnk@cc.tuat.ac.jp, okuda-m@kitakyu-u.ac.jp

ABSTRACT

We propose an eigenvalue shrinkage method with a modified Chebyshev polynomial approximation (CPA). The eigenvalue shrinkage has been used in many fields of signal and image processing. However, the shrinkage takes enormous computation time especially in the case that a matrix constructed from a signal or image becomes very large, i.e., eigendecomposition can hardly be performed. The CPA is an approximation method of the shrinkage function that avoids the eigendecomposition of the matrix. Unfortunately, it is known that the CPA generates *Gibbs phenomenon* around points of discontinuity for approximating an ideal response. The Chebyshev-Jackson polynomial approximation (CJPA) will alleviate the problem, but the transition bandwidth becomes wide, which is an undesired characteristic for some applications. In this paper, we propose an eigenvalue shrinkage method with the reduced Gibbs phenomenon by modifying the CPA using the weighted least squares approach. Our method can reduce the error as well as the CJPA. Furthermore, it yields the narrow transition band. Some experimental results on spectral clustering validate the effectiveness of the method.

Index Terms— Chebyshev polynomial approximation, weighted least squares, Gibbs phenomenon

1. INTRODUCTION

An eigenvalue shrinkage method has been used in many research fields of signal and image processing. For example, this approach is very useful for improving the smoothing performance of image filtering [1, 2] such as bilateral filter [3–6] and non-local means [7]. The eigenvalue shrinkage is also applied to a spectral graph filtering in graph signal processing (GSP) [8–16].

It is well known that the eigenvalue shrinkage takes enormous computation time because a matrix constructed from a signal often becomes very large, i.e., eigendecomposition (EVD) can hardly be performed [17, 18]. Note that the eigenvalue shrinkage can actually be carried out by computing neither eigenvectors nor eigenvalues. Chebyshev polynomial approximation (CPA) [9, 19, 20] is effective method to achieve the approximate eigenvalue shrinkage without using EVD, which drastically reduces the computation time. In our previous work, the above CPA-based eigenvalue shrinkage was further extended to accelerate a singular value shrinkage, which is often used in a matrix rank minimization [21, 22].

This work was supported in part by MEXT Tenure-Track Promotion Program and JSPS Grants-in-Aid for JSPS fellows (15J08568).

The CPA is effective for approximating a smooth function, however, approximating an ideal function with discontinuities and/or sharp transition gives truncation errors so called *Gibbs phenomenon* [23] even in the higher-order approximations (see Fig. 1(a)). This is because the CPA is defined as the weighted sum of sinusoidal waves. To reduce the error, the Chebyshev-Jackson polynomial approximation (CJPA) [24–28] has recently been used in the GSP-related field. The approximation errors are reduced by using the CJPA with a drawback of a wide transition band (see Fig. 1(b)). As a result, the difference between the approximated response and the ideal one becomes large, which affects the performance of practical applications.

In this paper, we propose an eigenvalue shrinkage method using a modified CPA which reduces the Gibbs phenomenon without affecting the transition bandwidth. A weighted least squares (WLS) approach is a key tool to achieve it. In [29], the WLS method was applied to the CPA to efficiently design linear phase FIR filters. It enables us to flexibly determine the area of the filter response where we would like to reduce the errors. For our method, the eigenvalue shrinkage method is modified by using the weighted CPA[1] (WCPA) in order to reduce the Gibbs phenomenon. It leads to the matrix form of the WCPA, which is the main contribution in this paper. Among practical applications, we apply our method to a fast spectral clustering method proposed in [30] for example. We replace the polynomial approximation part of [30], which uses the CJPA, with our method. Some experimental results show that our method can efficiently approximate an ideal response which leads to superior performances than the other methods.

This paper is organized as follows. Section 2 describes some notations and definitions. Section 3 presents the WCPA. First, the WCPA for a real-valued function is introduced, and then it is formulated to its matrix form, which is a main contribution in this paper. In Section 4, our method is applied to the spectral clustering, and some experimental results are shown. Finally, Section 5 concludes the paper.

2. NOTATIONS AND PRELIMINARIES

2.1. Notations

Bold-face capital and small letters indicate a matrix and a vector, respectively. Superscript $\cdot^{\top}$ is the transpose of a matrix and a vector, and superscript $\cdot^{-1}$ is the inverse of a non-singular matrix. The matrices $\mathbf{I}$ and $\mathbf{O}$ are the identity and null matrices, respectively. The

[1]This method is the CPA modified by using the WLS.

ℓ_p norm for $p \geq 1$ is defined as $\|\mathbf{x}\|_p := (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ $(\forall \mathbf{x} \in \mathbb{R}^n)$.

2.2. Chebyshev Polynomial Approximation of Matrix Form

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a full rank matrix and $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ be its EVD, where $\mathbf{P} \in \mathbb{R}^{n \times n}$ is the matrix composed of eigenvectors and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_n)$ is the diagonal matrix with the corresponding eigenvalues. In this paper, $\mathbf{A}$ is assumed to be a real symmetric matrix. Additionally, we define that the eigenvalues are bounded between 0 and $\lambda_{\max}$, where $\lambda_{\max} > 1$. The eigenvalues of $\mathbf{A}$ are processed with the function $\mathcal{H}(\cdot)$ as

$$\mathcal{H}(\mathbf{A}) := \mathbf{P}\mathrm{diag}\big(h(\lambda_1), \ldots, h(\lambda_n)\big)\mathbf{P}^{-1}, \quad (1)$$

where $h(x)$ is the filter kernel defined in $x \in [0, \lambda_{\max}]$. In this subsection, we present the approximated solution of (1) using the CPA.

The CPA of the matrix form gives an approximated solution of the function $\mathcal{H}(\cdot)$ by using truncated Chebyshev series as

$$\widehat{\mathcal{H}}(\mathbf{A}) := \frac{1}{2}\widehat{c}_0\mathbf{I} + \sum_{k=1}^{\alpha-1} \widehat{c}_k\, \Psi_k(\widehat{\mathbf{A}}), \quad (2)$$

where $\widehat{c}_k$ and $\Psi_k(\widehat{\mathbf{A}})$ are Chebyshev coefficients and Chebyshev polynomials, respectively, which are defined later. Additionally, α is an arbitrary approximation order. In (2), $\widehat{\mathbf{A}}$ is the eigenvalue-shifted matrix given by

$$\widehat{\mathbf{A}} := \frac{2}{\lambda_{\max}}\mathbf{A} - \mathbf{I}, \quad (3)$$

whose eigenvalues are obviously within $[-1, 1]$. Here, the ith eigenvalue of $\widehat{\mathbf{A}}$ is represented as $\widehat{\lambda}_i$. Thanks to (3), the kth order Chebyshev polynomial of $\widehat{\mathbf{A}}$ is defined as

$$\begin{aligned} \Psi_k(\widehat{\mathbf{A}}) &:= \Psi_k\left(\tfrac{2}{\lambda_{\max}}\mathbf{A} - \mathbf{I}\right) \\ &= \mathbf{P}\,\Psi_k\left(\tfrac{2}{\lambda_{\max}}\mathbf{\Lambda} - \mathbf{I}\right)\mathbf{P}^{-1} \\ &= \mathbf{P}\mathrm{diag}(\cos k\theta_1, \ldots, \cos k\theta_n)\mathbf{P}^{-1} \\ &= \mathbf{P}\mathrm{diag}(\psi_k(\widehat{\lambda}_1), \ldots, \psi_k(\widehat{\lambda}_n))\mathbf{P}^{-1} \end{aligned} \quad (4)$$

where $\theta_i := \arccos\left(\frac{2}{\lambda_{\max}}\lambda_i - 1\right)$ and $\psi_k(\widehat{\lambda}_i) := \cos\big(k \arccos(\widehat{\lambda}_i)\big)$. The Chebyshev polynomials are obtained using the following recurrence relation:

$$\begin{aligned} &\Psi_k(\widehat{\mathbf{A}}) = 2\widehat{\mathbf{A}}\,\Psi_{k-1}(\widehat{\mathbf{A}}) - \Psi_{k-2}(\widehat{\mathbf{A}}), \\ &\Psi_0(\widehat{\mathbf{A}}) = \mathbf{I}, \quad \Psi_1(\widehat{\mathbf{A}}) = \widehat{\mathbf{A}}. \end{aligned} \quad (5)$$

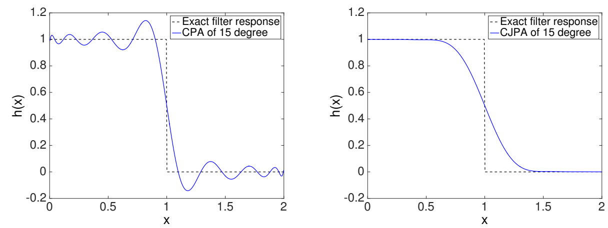
Then, the Chebyshev coefficients $\widehat{c}_k$ are derived by using the orthogonality of the cosine functions as follows:

$$\widehat{c}_k = \frac{2}{\alpha}\sum_{l=1}^{\alpha} \cos(k\theta_l) h\left(\frac{\lambda_{\max}}{2}(\cos\theta_l + 1)\right), \quad (6)$$

where $\theta_l := \frac{\pi(l - \frac{1}{2})}{\alpha}$. The term $h\left(\frac{\lambda_{\max}}{2}(\cos\theta_l + 1)\right)$ returns the shifted range back to the original range $[0, \lambda_{\max}]$. From (6), $\widehat{\mathcal{H}}(\mathbf{A})$ can also be represented using $\widehat{h}(\lambda_i)$ as

$$\widehat{\mathcal{H}}(\mathbf{A}) = \mathbf{P}\,\mathrm{diag}\big(\widehat{h}(\lambda_1), \ldots, \widehat{h}(\lambda_n)\big)\mathbf{P}^{-1}. \quad (7)$$

The function $\widehat{\mathcal{H}}(\cdot)$, which is referred to as the CPA-based eigenvalue shrinkage, results in the approximated function $\mathcal{H}(\cdot)$. The CPA-based eigenvalue shrinkage actually computes neither eigenvalues nor eigenvectors explicitly.



(a) (b)

Fig. 1. Examples of responses approximated from an ideal filter by using the CPA and the CJPA with the 15-th order approximation. In this figures, the black-dotted lines show ideal filter responses. (a) Response of CPA. (b) Response of CJPA.

2.3. Chebyshev-Jackson Polynomial Approximation

When the function $h(\cdot)$ in (1) is an ideal filter kernel, the function approximated by using the CPA shows the Gibbs phenomenon around the points of discontinuity, as shown in Fig. 1(a). The CJPA [24–28] reduces the errors by adding damping factors $d_{k,\alpha}$ to (2) as

$$\mathcal{H}_{\mathrm{J}}(\mathbf{A}) := \frac{1}{2}\widehat{c}_0\mathbf{I} + \sum_{k=1}^{\alpha-1} d_{k,\alpha}\widehat{c}_k\, \Psi_k(\widehat{\mathbf{A}}), \quad (8)$$

where

$$d_{k,\alpha} := \sum_{i=0}^{\alpha-k-1} a_{i,\alpha}a_{k+i,\alpha}, \quad (9)$$

in which

$$a_{i,\alpha} := \frac{U_i(\cos(\frac{\pi}{\alpha+1}))}{\sqrt{\sum_{j=0}^{\alpha-1} U_j(\cos(\frac{\pi}{\alpha+1}))}}. \quad (10)$$

The operator $U_i(\cdot)$ is the Chebyshev polynomials of the second kind defined as

$$\begin{aligned} U_i(x) &:= \tfrac{\partial}{\partial x}\tfrac{\psi_{k+1}(x)}{k+1} \\ &= \tfrac{\sin((k+1)\arccos(x))}{\sin(\arccos(x))}. \end{aligned} \quad (11)$$

As shown in Fig. 1(b), the CJPA reduces the Gibbs phenomenon but the CJPA also tends to have a large transition bandwidth.

3. WEIGHTED CHEBYSHEV POLYNOMIAL APPROXIMATION

The WCPA is a method that resolves the problems in the CPA and CJPA. We firstly introduce the WCPA for real-valued functions, and then it is extended to the matrix form for eigenvalue shrinkage.

3.1. Weighted Least Squares

First, we introduce the WLS. Let $h_{\mathrm{a}}(x)$ be the approximated response of $h(x)$, and $w(x)$ be a weight function. The input x is bounded between -1 and 1. The weighted squared error by using these notations is defined as

$$\varepsilon := \sum_{l=0}^{L-1} w(\cos\theta_l)(h(\cos\theta_l) - h_{\mathrm{a}}(\cos\theta_l))^2, \quad (12)$$

where L is an arbitrary integer ($L \geq \alpha$), and the approximated response is defined as

$$h_{\mathrm{a}}(x) := \sum_{k=0}^{\alpha-1} a_k \cos(k \arccos(x)), \tag{13}$$

in which a_k is a coefficient. When setting the derivative of ε to zero, i.e., $\frac{\partial \varepsilon}{\partial a_r} = 0$, we obtain

$$\begin{aligned} &\sum_{l=0}^{L-1} w(\cos\theta_l) h(\cos\theta_l) \cos(r\theta_l) \\ &= \sum_{l=0}^{L-1} \sum_{k=0}^{\alpha-1} w(\cos\theta_l) \cos(r\theta_l) \cos(k\theta_l) a_k. \end{aligned} \tag{14}$$

where $r \in [0, \alpha - 1]$ is constant. For simplicity, the terms of (14) are denoted as

$$\begin{aligned} s_1(r) &:= \sum_{l=0}^{L-1} w(\cos\theta_l) h(\cos\theta_l) \cos(r\theta_l), \\ s_2(r, k_{\mathrm{c}}) &:= \sum_{l=0}^{L-1} w(\cos\theta_l) \cos(r\theta_l) \cos(k_{\mathrm{c}}\theta_l), \end{aligned} \tag{15}$$

where k_{c} is constant within $[0, \alpha - 1]$. Let $\mathbf{a} := [a_0, \dots, a_k, \dots, a_{\alpha-1}]^\top$ and $\mathbf{s}_1 := [s_1(0), \dots, s_1(r), \dots, s_1(\alpha - 1)]^\top$. Additionally, the matrix of $s_2(r, k_{\mathrm{c}})$ is defined as $\mathbf{S}_2 := [s_2(r, k_{\mathrm{c}})]_{\alpha \times \alpha}$, where $s_2(r, k_{\mathrm{c}})$ is the component of the matrix $\mathbf{S}_2 \in \mathbb{R}^{\alpha \times \alpha}$ in the r-th row and k_{c}-th column. The matrix $\mathbf{S}_2$ is a symmetric matrix from (15). According to the above, the coefficients $\mathbf{a}$ can be calculated as $\mathbf{a} = \mathbf{S}_2^{-1} \mathbf{s}_1$.

3.2. Relation Between Orthogonal Functions and WLS

In this paper, the inner product between arbitrary functions is defined as

$$\langle f, g \rangle := \sum_{l=0}^{L-1} w(\cos\theta_l) f(\cos\theta_l) g(\cos\theta_l). \tag{16}$$

The cost function in (12) is rewritten by using (16) as

$$\varepsilon = \langle h - h_{\mathrm{a}}, h - h_{\mathrm{a}} \rangle. \tag{17}$$

Let $\phi_k(x)$ be orthogonal functions for the inner product (16) with $w(x)$, and this is represented as a polynomial form given by

$$\phi_k(x) := \sum_{t=0}^{k} b_{k,t} \cos^t(\arccos(x)), \tag{18}$$

where $b_{k,t}$ is a coefficient for the polynomial. Additionally, its orthonormal function is defined as $\bar{\phi}_k(x) := \frac{\phi_k(x)}{\langle \phi_k, \phi_k \rangle^{1/2}}$. Here, the approximated response $h_{\mathrm{a}}(x)$ is redefined by using $\bar{\phi}_k(x)$:

$$h_{\mathrm{a}}(\cos\theta_l) := \sum_{k=0}^{\alpha-1} \bar{g}_k \bar{\phi}_k(\cos\theta_l), \tag{19}$$

where $\bar{g}_k$ is a weight. According to the above definition, the equations in (15) can also be redefined as

$$\begin{aligned} s_1(r) &:= \sum_{l=0}^{L-1} w(\cos\theta_l) h(\cos\theta_l) \bar{\phi}_r(\cos\theta_l), \\ s_2(r, k_{\mathrm{c}}) &:= \sum_{l=0}^{L-1} w(\cos\theta_l) \bar{\phi}_r(\cos\theta_l) \bar{\phi}_{k_{\mathrm{c}}}(\cos\theta_l). \end{aligned} \tag{20}$$

Since $\bar{\phi}_k(x)$ is an orthonormal function, $s_2(k_1, k_2) = 1$ if $k_1 = k_2$ and $s_2(k_1, k_2) = 0$ otherwise. Therefore, the coefficient $\bar{g}_k$ can simply be calculated as

$$\bar{g}_k = s_1(k). \tag{21}$$

Equation (21) means that all $\bar{\phi}_k(x)$'s are required in order to obtain the coefficients $\bar{g}_k$. To obtain $\bar{\phi}_k(x)$ efficiently, we use the idea of the recurrence relation in the CPA.

3.3. Weighted Chebyshev Polynomial Approximation [29]

In [29], the following definitions are derived by satisfying the orthonormality of the functions $\bar{\phi}_k(x)$, i.e., $s_2(k_1, k_2) = 1$ if $k_1 = k_2$, $s_2(k_1, k_2) = 0$ otherwise. Its details are omitted in this paper due to limitation of the space.

The recurrence relation is defined as

$$\begin{aligned} &\phi_{k+1}(\cos\theta_l) := \frac{\cos\theta_l - \gamma_k}{\beta_k} \phi_k(\cos\theta_l) - \frac{\beta_k}{\beta_{k-1}} \phi_{k-1}(\cos\theta_l), \\ &\phi_{-1}(\cos\theta_l) := 0, \quad \phi_0(\cos\theta_l) := 1, \end{aligned} \tag{22}$$

where

$$\begin{aligned} \beta_k &:= \left\{ \sum_{l=0}^{L-1} w(\cos\theta_l) \phi_k^2(\cos\theta_l) \right\}^{\frac{1}{2}}, \\ \gamma_k &:= \frac{1}{\beta_k^2} \sum_{l=0}^{L-1} w(\cos\theta_l) \cos\theta_l \phi_k^2(\cos\theta_l). \end{aligned} \tag{23}$$

The coefficient $w(\cos\theta_l)$ controls some "don't care" points, e.g., $w(\cos\theta_l) = 0$ in transition bands and $w(\cos\theta_l) = 1$ otherwise. Note that the orthonormal function $\bar{\phi}_k(x)$ is given by $\frac{\phi_k(x)}{\beta_k}$ since $\beta_k = \langle \phi_k, \phi_k \rangle^{1/2}$ is considered as a norm from (23). According to the fact, the weighted Chebyshev series are given by

$$h_{\mathrm{a}}(\cos\theta_l) := \sum_{k=0}^{\alpha-1} \frac{\bar{g}_k}{\beta_k} \phi_k(\cos\theta_l). \tag{24}$$

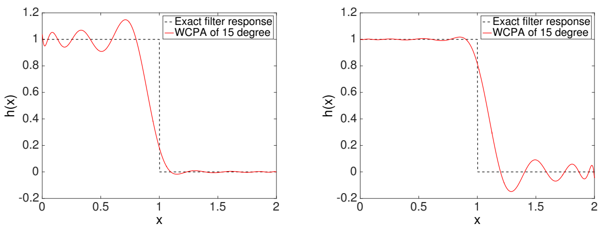
Additonally, the weighted Chebyshev coefficient $\bar{g}_k$ can be calculated from (20) and (21) as

$$\bar{g}_k = \frac{1}{\beta_k} \sum_{l=0}^{L-1} w(\cos\theta_l) h(\cos\theta_l) \phi_k(\cos\theta_l). \tag{25}$$

3.4. Weighted Chebyshev Polynomial Approximation for Matrix Form

Here, the matrix form of the WCPA is derived. The weighted Chebyshev polynomials for a matrix are defined as $\Phi_k(\widehat{\mathbf{A}}) := \mathbf{X}_{\mathbf{A}} \mathrm{diag}(\phi_k(\widehat{\lambda}_1), \dots, \phi_k(\widehat{\lambda}_i), \dots, \phi_k(\widehat{\lambda}_n)) \mathbf{X}_{\mathbf{A}}^{-1}$, whose diagonal matrix is represented as $\mathcal{D}(\phi_k(\widehat{\lambda}_i))$ in this subsection. Additionally, initial sets are defined as $\Phi_{-1}(\widehat{\mathbf{A}}) := \mathbf{O}$ and $\Phi_0(\widehat{\mathbf{A}}) := \mathbf{I}$. From the above, the $(k+1)$-th polynomial $\Phi_{k+1}(\widehat{\mathbf{A}})$ can be calculated as

$$\begin{aligned} \Phi_{k+1}(\widehat{\mathbf{A}}) &:= \mathbf{X}_{\mathbf{A}} \mathcal{D}(\phi_{k+1}(\widehat{\lambda}_i)) \mathbf{X}_{\mathbf{A}}^{-1} \\ &= \mathbf{X}_{\mathbf{A}} \mathcal{D}\Big(\tfrac{\widehat{\lambda}_i - \gamma_k}{\beta_k} \phi_k(\widehat{\lambda}_i) - \tfrac{\beta_k}{\beta_{k-1}} \phi_{k-1}(\widehat{\lambda}_i) \Big) \mathbf{X}_{\mathbf{A}}^{-1} \\ &= \tfrac{1}{\beta_k} (\widehat{\mathbf{A}} - \gamma_k \mathbf{I}) \Phi_k(\widehat{\mathbf{A}}) - \tfrac{\beta_k}{\beta_{k-1}} \Phi_{k-1}(\widehat{\mathbf{A}}), \end{aligned} \tag{26}$$



(a) (b)

Fig. 2. Examples of approximated responses by using the WCPA with the 15th order approximation. (a) $w(x) = 1$ if $x < 0.95$, $w(x) = 100$ if $x > 1.05$ and $w(x) = 0$ otherwise. (b) $w(x) = 100$ if $x < 0.95$, $w(x) = 1$ if $x > 1.05$ and $w(x) = 0$ otherwise.

where β_k and γ_k have already been derived in (23). By using the above recurrence relation, the weighted Chebyshev series for a matrix is defined as

$$\mathcal{H}_{\mathrm{a}}(\mathbf{A}) := \textstyle\sum_{k=0}^{\alpha-1} \frac{\widehat{g}_k}{\beta_k} \Phi_k(\widehat{\mathbf{A}}), \tag{27}$$

where the Chebyshev coefficient $\widehat{g}_k$ can be calculated as

$$\widehat{g}_k = \frac{1}{\beta_k} \sum_{l=0}^{L-1} w\Big(\frac{\lambda_{\max}}{2}(\cos\theta_l + 1)\Big) h\Big(\frac{\lambda_{\max}}{2}(\cos\theta_l + 1)\Big) \phi_k(\cos\theta_l). \tag{28}$$

According to the series, we can manipulate the eigenvalues of a matrix in the sense of the WLS.

As shown in Fig. 2, our method smoothes the part of the Gibbs phenomenon which we would like to reduce intensively. The advantage of our method is to keep the narrower transition bandwidth than that of the CJPA instead of removing all the errors.

4. EXPERIMENTS

4.1. Application to Spectral Clustering

Our method is applied to the fast spectral clustering method proposed in [30]. Here, we only describe how the polynomial approximation methods are applied to the spectral clustering method due to limitation of the space.

Let $\mathbf{V}_i \in \mathbb{R}^{m \times n}$ be the ith data of an arbitrary data set. The entire data are rearranged into a matrix $\mathbf{V} \in \mathbb{R}^{mn \times K}$. Let $\mathcal{G} := \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$ be an undirected graph with the set of vertices $\mathcal{V}$, the set of edges $\mathcal{E}$, and the adjacency matrix $\mathbf{W} \in \mathbb{R}^{K \times K}$ which encodes the weights of edges. The adjacency matrix is defined as $\mathbf{W} := [w_{ij}]_{K \times K}$, where w_{ij} is the weight on the edge. The edge weight w_{ij} is calculated from $\mathbf{V}_i$ and $\mathbf{V}_j$, whose details are indicated in [30]. Then, the diagonal degree matrix $\mathbf{M}$ is defined as $\mathbf{M} := \mathrm{diag}(m_i)$, where $m_i := \sum_j w_{ij}$. From the above definitions, the graph Laplacian matrix is defined as $\boldsymbol{\mathcal{L}} := \mathbf{M}^{-\frac{1}{2}}(\mathbf{M} - \mathbf{W})\mathbf{M}^{-\frac{1}{2}}$.

When the data set has k clusters, the k smallest eigenvalues of $\boldsymbol{\mathcal{L}}$ are remained as is when performing the spectral clustering as

$$h(\lambda_i^{\mathcal{L}}) = \begin{cases} 1 \text{ if } \lambda_i^{\mathcal{L}} \leq \lambda_k^{\mathcal{L}}, \\ 0 \text{ otherwise,} \end{cases} \tag{29}$$

where $\lambda_i^{\mathcal{L}}$ is the ith eigenvalue of the matrix $\boldsymbol{\mathcal{L}}$. The matrix whose eigenvalues are shrunk by (29) is defined as $\mathcal{H}(\boldsymbol{\mathcal{L}})$. In [30], the clustering is derived by using the matrix $\mathcal{H}(\boldsymbol{\mathcal{L}})\mathbf{R}$, where the matrix $\mathbf{R} \in \mathbb{R}^{K \times p}$ ($p \ll K$) is a random matrix whose components are independent random Gaussian variables. Our method and other polynomial approximation methods used the same eigenvalue shrinkage function shown in (29).

Table 1. Experimental Results (Average of Ten Executions): ARSI and Computation Time (s).

Approx. order	Evaluation index	CPA	CJPA	WCPA
10th order	ARSI	0.926	0.616	**0.938**
	Computation time	10.38	11.72	10.13
20th order	ARSI	0.945	0.929	**0.950**
	Computation time	17.87	18.24	17.88
30th order	ARSI	**0.949**	0.937	**0.949**
	Computation time	24.41	24.38	24.42

4.2. Experimental Results

The application was implemented with MATLAB R2015b and run on a 4-GHz Intel Core i7 processor with 32GB RAM. All condition was determined in the code provided by the authors of [30][2]. In our method, the weight function $w(x)$ in (28) is defined as $w(x) = 1$ if $x \leq \lambda_k^{\mathcal{L}} - 0.05$, $w(x) = 100$ if $x \geq \lambda_k^{\mathcal{L}} + 0.05$ and $w(x) = 0$ otherwise. The k-th eigenvalue $\lambda_k^{\mathcal{L}}$ is estimated as described in [30]. The value L in (23) and (28) was defined as $L = 1000$. The approximation precision of these methods was compared by using the 10th, 20th, and 30th order approximation whose values were used for α in (27). Additionally, the clustering performance was indicated by using the adjusted rand similarity index (ARSI) [31] between the ground truth and the obtained partitions.

Table 1 shows the experimental results regarding the approximation precision and computation time[3]. The clustering performance by using our method is better than the others even in the low approximation order. Especially, the performance of our method is much better than that of the CJPA. This is because the proposed method can use lower order approximation than the conventional method in order to present the comparable performance.

5. CONCLUSION

In this paper, we proposed an eigenvalue shrinkage method with the alleviated Gibbs phenomenon by using the weighted least squares method. The CPA can be redefined by using the weighted least squares method for reducing the Gibbs phenomenon. Thanks to this, the Gibbs phenomenon can be selectively attenuated while keeping the narrow transition band. We applied the WCPA to the eigenvalue shrinkage by formulating its matrix form to utilize the advantage. Our eigenvalue shrinkage enables us to efficiently approximate an ideal function in the sense of the WLS. Our method can be applied to many applications in various research fields including signal and image processing and machine learning. Among practical applications, our method was applied to the spectral clustering for example. The proposed method presented better clustering performances than those of the conventional methods even if we used the low order approximation.

[2] Available at `http://cscbox.gforge.inria.fr/`

[3] The computation time was measured in the entire spectral clustering algorithms.

6. REFERENCES

[1] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, 2014.

[2] M. Onuki, S. Ono, K. Shirai, and Y. Tanaka, "Non-local/local image filters using fast eigenvalue filtering," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2015, pp. 4659–4663.

[3] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 1998, pp. 839–846.

[4] S. Paris, P. Jirboribst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 12, pp. 2324–2333, 2009.

[5] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic range images," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 21, no. 3, pp. 257–266, 2002.

[6] S. Fleishman, I. Drori, and D. Cohen-Or, "Bilateral mesh denoising," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 22, no. 3, pp. 950–953, 2003.

[7] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, vol. 2, pp. 60–65.

[8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[9] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied Computational Harmonic Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[10] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3357–3367, 2013.

[11] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2014, pp. 3892–3896.

[12] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, 2012.

[13] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.

[14] Y. Tanaka and A. Sakiyama, "M-channel oversampled graph filter banks," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3578–3590, 2014.

[15] S. Ono, I. Yamada, and I. Kumazawa, "Total generalized variation for graph signals," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015, pp. 5456–5460.

[16] M. Onuki, S. Ono, M. Yamagishi, and Y. Tanaka, "Graph signal denoising via trilateral filter on graph spectral domain," *IEEE Trans. Signal Info. Process. Netw.*, vol. 2, no. 2, pp. 137–148, 2016.

[17] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*, Cambridge Univ. Press, New York, NY, USA, 1986.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[19] G. M. Phillips, *Interpolation and approximation by polynomials*, CMS Books Mathematics. Springer-Verlag, 2003.

[20] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*, Chapman and Hall/CRC, 2002.

[21] M. Onuki, S. Ono, K. Shirai, and Y. Tanaka, "Image colorization based on ADMM with fast singular value thresholding by Chebyshev polynomial approximation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2016, pp. 4762–4766.

[22] M. Onuki, S. Ono, K. Shirai, and Y. Tanaka, "Fast singular value shrinkage with Chebyshev polynomial approximation based on signal sparsity," *submitted to IEEE Trans. Signal Process.*

[23] P. P Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[24] D. Jackson, *The Theory of Approximation*, vol. 11, American Mathematical Society, 2005.

[25] R. N. Silver, H. Roeder, A. F. Voter, and J. D. Kress, "Kernel polynomial approximations for densities of states and spectral functions," *J. Comp. Phys.*, vol. 124, no. 48, pp. 115–130, 1996.

[26] R. N. Silver and H. Röder, "Calculation of densities of states and spectral functions by Chebyshev recursion and maximum entropy," *Phys. Rev. E*, vol. 56, no. 4, pp. 4822–4829, 1997.

[27] L. Lin, Y. Saad, and C. Yang, "Approximating spectral densities of large matrices," *SIAM Rev.*, vol. 58, no. 1, pp. 34–65, 2016.

[28] L. O. Jay, H. Kim, Y. Saad, and J. R. Chelikowsky, "Electronic structure calculations in plane-wave codes without diagonalization," *Comput. Phys. Commun.*, vol. 118, no. 1, pp. 21–30, 1999.

[29] M. Okuda, M. Ikehara, and S. Takahashi, "Fast and stable least-squares approach for the design of linear phase FIR filters," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1485–1493, 1998.

[30] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst, "Compressive spectral clustering," in *Proc. Int. Conf. Mach. Learn.*, 2016.

[31] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, no. 1, pp. 192–218, 1985.