On Variations of Power Iteration

Seungjin Choi

Department of Computer Science
Pohang University of Science and Technology
San 31 Hyoja-dong, Nam-gu
Pohang 790-784, Korea
seungjin@postech.ac.kr

Abstract. The power iteration is a classical method for computing the eigenvector associated with the largest eigenvalue of a matrix. The subspace iteration is an extension of the power iteration where the subspace spanned by n largest eigenvectors of a matrix, is determined. The natural power iteration is an exemplary instance of the subspace iteration, providing a general framework for many principal subspace algorithms. In this paper we present variations of the natural power iteration, where n largest eigenvectors of a symmetric matrix without rotation ambiguity are determined, whereas the subspace iteration or the natural power iteration finds an invariant subspace (consisting of rotated eigenvectors). The resulting method is referred to as *constrained natural power iteration* and its fixed point analysis is given. Numerical experiments confirm the validity of our algorithm.

1 Introduction

A symmetric eigenvalue problem where the eigenvectors of a symmetric matrix are required to be computed, is a fundamental problem encountered in a variety of applications involving the spectral decomposition. The power iteration is a classical and the simplest method for computing the eigenvector with the largest modulus. The subspace iteration is a natural generalization of the power iteration, where the subspace spanned by n largest eigenvectors of a matrix, is determined.

The natural power iteration [1] is an exemplary instance of the subspace iteration, that was investigated mainly for principal subspace analysis. In this paper we present variations of the natural power iteration and show that its fixed point is the n largest eigenvectors of a symmetric matrix up to a sign ambiguity, whereas the natural power iteration just finds a principal subspace (i.e., arbitrarily rotated eigenvectors). The resulting algorithm is referred to as *constrained natural power iteration*. Numerical experiments confirm the validity of our algorithm.

2 Natural Power Iteration

The power iteration is a classical method which finds the largest eigenvector (associated with the largest eigenvalue) of a matrix $\boldsymbol{C} \in \mathbb{R}^{m \times m}$ [2]. Given a

symmetric matrix $\boldsymbol{C} \in \mathbb{R}^{m\times m}$ (hence its eigenvalues are real), the power iteration starts from a nonzero vector $\boldsymbol{w}(0)$ and iteratively updates $\boldsymbol{w}(t)$ by

$$\widetilde{\boldsymbol{w}}(t+1) = \boldsymbol{C}\boldsymbol{w}(t), \tag{1}$$

$$\boldsymbol{w}(t+1) = \frac{\widetilde{\boldsymbol{w}}(t+1)}{\|\widetilde{\boldsymbol{w}}(t+1)\|_2}, \tag{2}$$

where $\|\cdot\|_2$ represents Euclidean norm. Combining (1) and (2) leads to the updating rule which has the form

$$\boldsymbol{w}(t+1) = \boldsymbol{C}\boldsymbol{w}(t)\left[\boldsymbol{w}^T(t)\boldsymbol{C}^2\boldsymbol{w}(t)\right]^{-\frac{1}{2}}. \tag{3}$$

Assume that $\boldsymbol{C}$ has an unique eigenvalue of maximum modulus λ_1 associated with the leading eigenvector $\boldsymbol{u}_1$. Then the power iteration (3) leads $\boldsymbol{w}(t)$ to converge to $\boldsymbol{u}_1$.

The subspace iteration [3] is a direct generalization of the power iteration, for computing several eigenvectors of $\boldsymbol{C}$. Starting from $\boldsymbol{W}(0) \in \mathbb{R}^{m\times n}$, the subspace iteration updates $\boldsymbol{W}(t)$ by

$$\boldsymbol{W}(t+1) = \boldsymbol{C}\boldsymbol{W}(t). \tag{4}$$

The space spanned by $\boldsymbol{W}(t)$ converges to invariant subspace determined by n largest eigenvectors of $\boldsymbol{C}$, provided that $|\lambda_n| > |\lambda_{n+1}|$ [3]. As in the power iteration, the subspace iteration requires the normalization or orthogonalization.

The subspace iteration

$$\widetilde{\boldsymbol{W}}(t+1) = \boldsymbol{C}\boldsymbol{W}(t), \tag{5}$$

followed by an orthogonalization

$$\boldsymbol{W}(t+1) = \widetilde{\boldsymbol{W}}(t+1)\left[\widetilde{\boldsymbol{W}}^T(t+1)\widetilde{\boldsymbol{W}}(t+1)\right]^{-\frac{1}{2}}, \tag{6}$$

leads to

$$\boldsymbol{W}(t+1) = \underbrace{\boldsymbol{C}\boldsymbol{W}(t)}_{\text{power term}} \underbrace{\left[\boldsymbol{W}^T(t)\boldsymbol{C}^2\boldsymbol{W}(t)\right]^{-\frac{1}{2}}}_{\text{normalizer}}, \tag{7}$$

which is known as the *natural power iteration* proposed in [1].

Denote the eigendecomposition of the symmetric matrix $\boldsymbol{C} \in \mathbb{R}^{m\times m}$ of rank $r(> n)$ as

$$\boldsymbol{C} = \begin{bmatrix}\boldsymbol{U}_1 \, \boldsymbol{U}_2\end{bmatrix} \begin{bmatrix}\boldsymbol{\Lambda}_1 & 0 \\ 0 & \boldsymbol{\Lambda}_2\end{bmatrix} \begin{bmatrix}\boldsymbol{U}_1 \, \boldsymbol{U}_2\end{bmatrix}^T, \tag{8}$$

where $\boldsymbol{U}_1 \in \mathbb{R}^{m\times n}$ contains n largest eigenvectors, $\boldsymbol{U}_2 \in \mathbb{R}^{m\times(m-n)}$ consists of the rest of eigenvectors, and associated eigenvalues are in $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ with $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$. The key result in regards to the natural power iteration is summarized in the following theorem

Theorem 1 (Y. Hua *et al.* [1]). *The weight matrix $\boldsymbol{W}(t) \in \mathbb{R}^{m\times n}$ in the natural power iteration (10) globally and exponentially converges to $\boldsymbol{W} = \boldsymbol{U}_1\boldsymbol{Q}$ where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is an arbitrary orthogonal matrix, provided that the nth and $(n+1)$th eigenvalues of $\boldsymbol{C}$ are distinct and the initial weight matrix $\boldsymbol{W}(0)$ meets a mild condition, saying that there exists a nonsingular matrix $\boldsymbol{L} \in \mathbb{R}^{(m-n)\times n}$ such that $\boldsymbol{U}_2^T\boldsymbol{W}(0) = \boldsymbol{L}\boldsymbol{U}_1^T\boldsymbol{W}(0)$ for a randomly chosen $\boldsymbol{W}(0)$.*

The natural power iteration was mainly studied for principal subspace analysis where $\boldsymbol{C} = E\{\boldsymbol{x}(t)\boldsymbol{x}(t)\}$ is the covariance matrix of m-dimensional stationary vector sequences, $\boldsymbol{x}(t)$, with zero mean. In such a case, the matrix $\boldsymbol{C}$ is symmetric as well as positive semidefinite. For the case of principal subspace analysis, the weight vector $\boldsymbol{W}(t)$ of the natural power iteration (7) converges to n principal arbitrary rotated eigenvectors of $\boldsymbol{C}$. A variety of algorithms, including Oja's subspace rule [4], PAST [5], OPAST [6], can be viewed as the implementations of the natural power iteration [1]. However, all these algorithms belong to the principal subspace method where arbitrarily rotated eigenvectors are determined, unless the deflation method was used to extract principal components one by one. Next section describes a simple variation of the natural power iteration, incorporating the upper-triangularization operator into the normalizer in (7). This variation is referred to as a *constrained natural power iteration*. It is shown here that a fixed point of the constrained natural power iteration is $\boldsymbol{W} = \boldsymbol{U}_1$ (up to a sign ambiguity). Thus, the constrained natural power iteration computes the exact eigenvectors of a given symmetric matrix, whereas the natural power method finds a principal subspace.

3 Constrained Natural Power Iteration

We impose a constraint in the normalization term in the natural power method (7), through an upper-triangularization operator $\mathcal{U}_T[\cdot]$ which sets all elements of its matrix argument that are below the diagonal to zero, i.e., $\mathcal{U}_T[\boldsymbol{Y}]$ for an arbitrary matrix $\boldsymbol{Y} \in \mathbb{R}^{n\times n}$ gives

$$\mathcal{U}_T[y_{ij}] = \begin{cases} 0 & \text{if } i > j \\ y_{ij} & \text{if } i \le j \end{cases}, \tag{9}$$

where y_{ij} is the (i,j)-element of $\boldsymbol{Y}$. The constrained natural power iteration updates the weight matrix by

$$\boldsymbol{W}(t+1) = \boldsymbol{C}\boldsymbol{W}(t)\left\{\mathcal{U}_T\left[\boldsymbol{W}^T(t)\boldsymbol{C}^2\boldsymbol{W}(t)\right]\right\}^{-\frac{1}{2}}. \tag{10}$$

Only difference between the constrained natural power iteration (10) and the natural power method (7) lies in the presence of $\mathcal{U}_T$ in the normalization term. As will be shown below, the operator $\mathcal{U}_T$ leads the algorithm (10) to find exact principal eigenvectors of $\boldsymbol{C}$ up to a sign ambiguity under mild conditions that are generally required for power iteration. That is, the fixed point of (10) satisfies $\boldsymbol{W} = \boldsymbol{U}_1\overset{\circ}{\boldsymbol{I}}$ where $\overset{\circ}{\boldsymbol{I}}$ is a diagonal matrix with its diagonal entries being 1 or -1, whereas the fixed point of (7) is $\boldsymbol{U}_1\boldsymbol{Q}$ for an arbitrary orthogonal matrix $\boldsymbol{Q}$.

Theorem 2. *The fixed point $\boldsymbol{W}$ of the constrained natural power iteration (10) satisfies $\boldsymbol{W} = \boldsymbol{U}_1 \overset{\circ}{\boldsymbol{I}}$, under the same conditions as Theorem 1.*

Proof. We define $\boldsymbol{\Phi}(t) = \boldsymbol{U}_1^T \boldsymbol{W}(t)$ and $\boldsymbol{\Omega}(t) = \boldsymbol{U}_2^T \boldsymbol{W}(t)$. With this definition, pre-multiplying both sides of (10) by $[\boldsymbol{U}_1 \, \boldsymbol{U}_2]^T$ leads to

$$\begin{bmatrix} \boldsymbol{\Phi}(t+1) \\ \boldsymbol{\Omega}(t+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 \\ 0 & \boldsymbol{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}(t) \\ \boldsymbol{\Omega}(t) \end{bmatrix} \boldsymbol{Z}(t), \tag{11}$$

where

$$\boldsymbol{Z}(t) = \left\{ \mathcal{U}_T \left[\boldsymbol{\Phi}^T(t) \boldsymbol{\Lambda}_1^2 \boldsymbol{\Phi}(t) + \boldsymbol{\Omega}^T(t) \boldsymbol{\Lambda}_2^2 \boldsymbol{\Omega}(t) \right] \right\}^{-\frac{1}{2}}. \tag{12}$$

As in the convergence proof of the natural power iteration in [1], one can show that $\boldsymbol{\Omega}(t)$ goes to zero. Assume that $\boldsymbol{\Phi}(0) \in \mathbb{R}^{n \times n}$ is a nonsingular matrix, then it implies that $\boldsymbol{\Omega}(0) = \boldsymbol{L}\boldsymbol{\Phi}(0)$ for some matrix $\boldsymbol{L}$. Then it follows from (11) that we can write

$$\boldsymbol{\Omega}(t) = \boldsymbol{\Lambda}_2^t \boldsymbol{L} \boldsymbol{\Lambda}_1^{-t} \boldsymbol{\Phi}(t). \tag{13}$$

The assumption that first n eigenvalues of $\boldsymbol{C}$ are strictly larger than the others, together with (13), implies that $\boldsymbol{\Omega}(t)$ converges to zero and is asymptotically in the order of $|\lambda_{n+1}/\lambda_n|^t$ where $|\lambda_n|$ and $|\lambda_{n+1}|$ ($< |\lambda_n|$) are nth and $(n+1)$th largest eigenvalues of $\boldsymbol{C}$.

Taking into account that $\boldsymbol{\Omega}(t)$ goes to zero, the fixed point $\boldsymbol{\Phi}$ of (11) satisfies

$$\boldsymbol{\Phi} \left\{ \mathcal{U}_T \left[\boldsymbol{\Phi}^T \boldsymbol{\Lambda}_1^2 \boldsymbol{\Phi} \right] \right\}^{\frac{1}{2}} = \boldsymbol{\Lambda}_1 \boldsymbol{\Phi}. \tag{14}$$

Note that $\boldsymbol{\Lambda}_1$ is a diagonal matrix with diagonal entries λ_i for $i = 1, \ldots, n$. Thus, one can easily see that $\boldsymbol{\Phi}$ is the eigenvector matrix of $\left\{ \mathcal{U}_T \left[\boldsymbol{\Phi}^T \boldsymbol{\Lambda}_1^2 \boldsymbol{\Phi} \right] \right\}^{\frac{1}{2}}$ with associated eigenvalues in $\boldsymbol{\Lambda}_1$. Note that the eigenvalues of an upper-triangular matrix are the diagonal elements. Then it follows from (14) that we have a set of equations

$$\left(\boldsymbol{\varphi}_i^T \boldsymbol{\Lambda}_1^2 \boldsymbol{\varphi}_i \right)^{\frac{1}{2}} = \lambda_i, \quad i = 1, \ldots, n. \tag{15}$$

where $\boldsymbol{\varphi}_i$ is the ith column vector of $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1 \, \boldsymbol{\varphi}_2 \, \cdots \, \boldsymbol{\varphi}_n]$. We can re-write (15) as

$$\sum_{i=1}^{n} \lambda_i^2 \varphi_{ij}^2 = \lambda_j^2, \quad j = 1, \ldots, n, \tag{16}$$

where φ_{ij} is the (i,j)-element of $\boldsymbol{\Phi}$. Assume $n \leq \mathrm{rank}(\boldsymbol{C})$, then $\lambda_i \neq 0$, $i = 1, \ldots, n$. For non-zero λ_i, the only $\boldsymbol{\Phi}$ satisfying (16) is $\boldsymbol{\Phi} = \overset{\circ}{\boldsymbol{I}}$. Therefore, $\boldsymbol{W} =$

$\boldsymbol{U}_1 \overset{\circ}{\boldsymbol{I}}$, implying that the fixed point of (10) is the true eigenvector matrix $\boldsymbol{U}_1$ up to a sign ambiguity. ■

Based on the result in Theorem 2, we can consider a variation of the constrained natural power iteration (10), described by

$$\boldsymbol{W}(t+1) = \boldsymbol{C}\boldsymbol{W}(t) \left\{ \mathcal{U}_T \left[\boldsymbol{W}^T(t) \boldsymbol{C}^2 \boldsymbol{W}(t) \right] \right\}^{-1}. \tag{17}$$

Following Theorem 2, one can easily see that the weight matrix $\boldsymbol{W}(t)$ in (17) also converges to the scaled eigenvector matrix of $\boldsymbol{C}$. Algorithms (10) and (17) have a difference in their normalizers. The matrix inverse requires less complexity, compared to the square-root-inverse of a matrix, although (17) finds scaled eigenvectors.

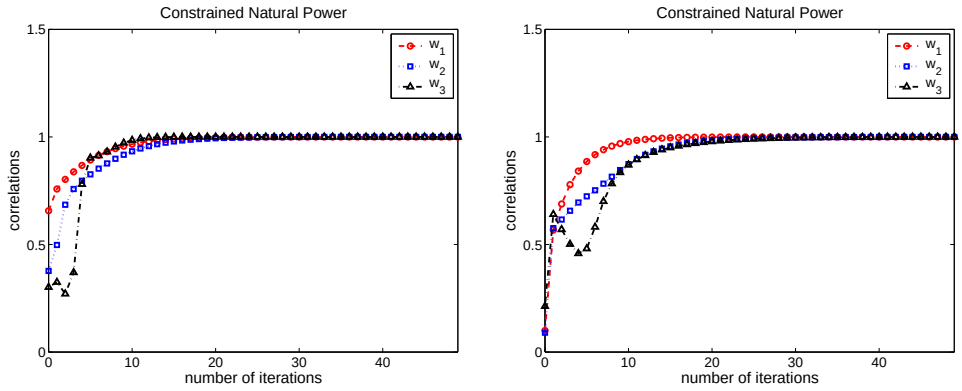
4 Numerical Experiments

Two simple numerical examples are shown in order to verify that the weight matrix $\boldsymbol{W}(t)$ converges to true eigenvectors of a given symmetric matrix $\boldsymbol{C}$. The first experiment was carried out with a symmetric matrix $\boldsymbol{C} \in \mathbb{R}^{5\times 5}$ whose eigenvalues are $2.48, -2.18, 1.20, -0.50, 0.34$. Fig. 1 (a) shows the the evolution of $\left| \boldsymbol{w}_i^T \boldsymbol{u}_i \right|$ for $i = 1, 2, 3$, where $\boldsymbol{w}_i$ is the ith column vector of $\boldsymbol{W}$ and $\boldsymbol{u}_i$ are true eigenvectors computed by SVD in Matlab.

The second experiment is related to principal component analysis. We generated 100-dimensional data vectors of length 1000, $\boldsymbol{x}(t) \in \mathbb{R}^{100}$, $t = 1, \ldots, 1000$, through linearly transforming 5-dimensional Gaussian vectors, $\boldsymbol{s}(t) \in \mathbb{R}^5$, with zero mean and unit variance, i.e., $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)$ where $\boldsymbol{A} \in \mathbb{R}^{100\times 5}$ and its elements were randomly drawn from Gaussian distribution. We applied the constrained natural power iteration (10) with a weight matrix $\boldsymbol{W}(t) \in \mathbb{R}^{100\times 3}$ to estimate first 3 eigenvectors of $\boldsymbol{C} = \frac{1}{1000} \sum_{t=1}^{1000} \boldsymbol{x}(t)\boldsymbol{x}^T(t)$. Fig. 1 (b) shows the evolution of $\left| \boldsymbol{w}_i^T \boldsymbol{u}_i \right|$ for $i = 1, 2, 3$, where $\boldsymbol{u}_i$ are true eigenvectors computed by SVD in Matlab.

5 Discussions

We have presented the constrained natural power iteration and have shown that its fixed point corresponded to the exact eigenvectors of a given symmetric matrix, up to sign ambiguity. Its slight variation was also discussed. Numerical experiments confirmed that the constrained natural power iteration successfully first n eigenvectors of $\boldsymbol{C}$. The constrained natural power iteration will be useful, especially for the case where a few eigenvectors are required to be determined from very high-dimensional data. Constrained natural power iteration could be viewed as a recognition model counterpart of the generative model-based methods in [7, 8] where EM optimization were used. The constrained natural power iteration has an advantage over EM algorithms in [7, 8], in the sense that the former involves a single-step updating whereas the latter needs two-step updating (E and M steps), although both share a similar spirit.



(a) (b)

Fig. 1. Convergence of $\boldsymbol{W} = [\boldsymbol{w}_1\, \boldsymbol{w}_2\, \boldsymbol{w}_3]$ in the constrained natural power iteration, is shown in terms of the absolute values of the inner product between these weight vectors and first three true eigenvectors of $\boldsymbol{C}$: (a) experiment 1; (b) experiment 2.

Acknowledgments

This work was supported by ITEP Brain Neuroinformatics program, KISTEP International Cooperative Research Program, Systems Bio-Dynamics Research Center under NCRC progrm, KOSEF 2000-2-20500-009-5, and Basic Research Fund in POSTECH.

References

1. Hua, Y., Xiang, Y., Chen, T., Abed-Meraim, K., Miao, Y.: A new look at the power method for fast subspace tracking. Digital Signal Processing **9** (1999) 297–314
2. Golub, G.H., Loan, C.F.V.: Matrix Computations. 2 edn. Johns Hopkins (1993)
3. Heath, M.T.: Scientific Computing: An Introductory Survey. 2 edn. McGraw-Hill, New York (2002)
4. Oja, E.: Neural networks, principal component analysis, and subspaces. International Journal of Neural Systems **1** (1989) 61–68
5. Yang, B.: Projection approximation subsapce tracking. IEEE Trans. Signal Processing **43** (1995) 95–107
6. Abed-Meraim, K., Chkeif, A., Hua, Y.: Fast orthonormal PAST algorithm. IEEE Signal Processing Letters **7** (2000) 60–62
7. Ahn, J.H., Oh, J.H.: A constrained EM algorithm for principal component analysis. Neural Computation **15** (2003) 57–65
8. Ahn, J.H., Choi, S., Oh, J.H.: A new way of PCA: Integrated-squared-error and EM algorithms. In: Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Montreal, Canada (2004)