Equivariant Algorithms for Estimating the Strong-Uncorrelating Transform in Complex Independent Component Analysis

Scott C. Douglas[1], Jan Eriksson[2], and Visa Koivunen[2]

[1] Department of Electrical Engineering,
Southern Methodist University, Dallas, Texas 75275, USA
[2] Signal Processing Laboratory, SMARAD CoE,
Helsinki University of Technology,
Espoo 02015, Finland

Abstract. For complex-valued multidimensional signals, conventional decorrelation methods do not completely specify the covariance structure of the whitened measurements. In recent work [1,2], the concept of strong-uncorrelation and its importance for complex-valued independent component analysis has been identified. Few algorithms for estimating the strong-uncorrelating transform currently exist. This paper presents two novel algorithms for estimating and computing the strong uncorrelating transform. The first algorithm uses estimated covariance and pseudo-covariance matrices, and the second algorithm estimates the strong uncorrelating transform directly from measurements. An analysis shows that the only stable stationary point of both algorithms produces the strong uncorrelating transform when the circularity coefficients of the sources are distinct and positive. Simulations show the efficacy of the approach in a source clustering task for wireless communications.

1 Introduction

In most treatments of blind source separation and independent component analysis, the signals are assumed to be real-valued. In a number of practical applications, however, measurements are naturally represented using complex linear models. In wireless communications, multiantenna or multiple-input, multiple-output systems can be conveniently described using a complex-valued mixture model. Multiple-sensor recordings in various biological signal processing applications are also well-represented in complex form [3]. These applications motivate the study of m-dimensional complex-valued signal mixtures of the form

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{1}$$

where $\mathbf{A}$ is an arbitrary complex-valued $(m \times m)$ matrix and the source signal vector sequence $\mathbf{s}(k)$ contains statistically-independent complex-valued elements.

Recently, work in complex ICA has uncovered a statistical structure that is unlike the real-valued case [1,2]. In particular, it is possible in some cases to

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 57–65, 2006.

identify $\mathbf{A}$ in (1) using only second-order statistics from $\mathbf{x}(k)$ at time k, a situation that is distinct from the real-valued case. The key construct in these results is the *strong-uncorrelating transform,* which we now describe. Without loss of generality, assume that the source covariance and pseudo-covariance matrices are $E\{\mathbf{s}(k)\mathbf{s}^H(k)\} = \mathbf{I}$ and $E\{\mathbf{s}(k)\mathbf{s}^T(k)\} = \boldsymbol{\Lambda}$, respectively, where $\boldsymbol{\Lambda}$ is a diagonal matrix of ordered real-valued entries between zero and one called *circularity coefficients* $\{\lambda_i\}$, $i \in \{1, \ldots, m\}$. Define the covariance and pseudo-covariance matrices of $\mathbf{x}(k)$ as

$$\mathbf{R} = E\{\mathbf{x}(k)\mathbf{x}^H(k)\} = \mathbf{A}\mathbf{A}^H \qquad \text{and} \qquad \mathbf{P} = E\{\mathbf{x}(k)\mathbf{x}^T(k)\} = \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T, \quad (2)$$

respectively. Then, the strong-uncorrelating transform $\underline{\mathbf{W}}$ is a matrix satisfying

$$\underline{\mathbf{W}}\mathbf{R}\underline{\mathbf{W}}^H = \mathbf{I} \qquad \text{and} \qquad \underline{\mathbf{W}}\mathbf{P}\underline{\mathbf{W}}^T = \boldsymbol{\Lambda}. \quad (3)$$

If the $\{\lambda_i\}$ values are distinct, then a matrix $\mathbf{W}$ satisfying (3) is also a separating matrix for the mixing model in (1). Additional results for the strong-uncorrelating transform are in [1,2], and [9] uses the transform to derive kurtosis-based fixed-point algorithms for complex signal mixtures.

In [1], a technique for computing the strong uncorrelating transform for given values of $\mathbf{R}$ and $\mathbf{P}$ is described. This technique employs both an eigenvalue decomposition of a Hermitian symmetric matrix and the *Takagi factorization* of a complex symmetric matrix, the latter of which requires specialized numerical code [5]. A Jacobi-type rotation method for the Takagi factorization is outlined in [6], but its numerical and convergence properties are not established. Both of these methods are computationally-complex and not amenable to situations in which the second-order data statistics are slowly-varying. Since few methods for computing the strong-uncorrelation transform currently exist, it is of great interest to derive simple algorithms for the strong-uncorrelating transform that could be employed in adaptive estimation and tracking tasks.

This paper describes two simple iterative procedures for computing the strong uncorrelating transform adaptively. Both procedures can be viewed as extensions of the method in [7]. The first procedure employs sample estimates of the covariance and pseudo-covariance matrices and is equivariant with respect to the mixing system $\mathbf{A}$ when sample-based averages of these matrices are used. The second equivariant procedure estimates the strong-uncorrelating transform directly from measurements. Both techniques have the significant advantage of not requiring estimates of the circularity coefficients $\{\lambda_i\}$ for their successful operation. Simulations show the abilities of the methods to perform strong-uncorrelation in a source clustering task for wireless communications.

2 An Adaptive Algorithm for the Strong Uncorrelating Transform

The simple algorithms described in this paper adapt a row-scaled version of $\underline{\mathbf{W}}$, termed $\mathbf{W}(k)$, to compute the strong uncorrelating transform. In the interest of algorithm simplicity, and because overall output signal scaling is often not an issue, we define the space of allowable solutions for $\mathbf{W}(k)$ as

$$\lim_{k\to\infty} \mathbf{W}(k)\mathbf{R}\mathbf{W}^H(k) = \widetilde{\mathbf{D}} \qquad \text{and} \qquad \lim_{k\to\infty} \mathbf{W}(k)\mathbf{P}\mathbf{W}^T(k) = \widetilde{\mathbf{D}}\boldsymbol{\Lambda}, \tag{4}$$

where $\widetilde{\mathbf{D}}$ is an arbitrary diagonal matrix of positively-valued diagonal entries. If $\mathbf{R}$ is available or can be estimated, then a $\mathbf{W}(k)$ satisfying (4) can be turned into a $\underline{\mathbf{W}}$ satisfying (3) using $\underline{\mathbf{W}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{W}(k)$.

Our first proposed algorithm for adaptively computing the strong-uncorrelating transform is

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu\Big(\mathbf{I} - \mathbf{W}(k)\widehat{\mathbf{R}}(k)\mathbf{W}^H(k) - \text{tri}[\mathbf{W}(k)\widehat{\mathbf{P}}(k)\mathbf{W}^T(k)]\Big)\mathbf{W}(k), \tag{5}$$

where $\widehat{\mathbf{R}}(k)$ and $\widehat{\mathbf{P}}(k)$ are sample estimates of $\mathbf{R}$ and $\mathbf{P}$ and tri$[\mathbf{M}]$ denotes a matrix whose lower triangular portion is identical to that of $\mathbf{M}$ and whose strictly-upper triangular portion is zero.

The following three theorems describe important theoretical and practical convergence properties of this algorithm, the proofs of which are in the Appendix.

Theorem 1. *The algorithm in (5) is equivariant with respect to the mixing matrix* $\mathbf{A}$ *under the data model in (1).*

Remark. Although the algorithm is equivariant with respect to the mixing matrix $\mathbf{A}$, its performance is affected by the values in $\boldsymbol{\Lambda}$ that depend on the sources. Thus, convergence of the algorithm may be fast or slow depending on $\boldsymbol{\Lambda}$.

Theorem 2. *The space of stationary points for the algorithm in (5) are* $\mathbf{W} = \mathbf{0}$ *and the set of matrices that satisfy*

$$\mathbf{W}\mathbf{R}\mathbf{W}^H = \mathbf{I} - \mathbf{D} \ \ \textit{and} \ \ \mathbf{W}\mathbf{P}\mathbf{W}^T = \mathbf{D}, \tag{6}$$

where $\mathbf{D}$ *is a diagonal matrix of real-valued unordered entries that are all less than or equal to one.*

Theorem 3. *Suppose the diagonal entries of* $\boldsymbol{\Lambda}$ *are distinct and positive. Then, the only locally-stable stationary point of the algorithm in (5) is the unique matrix* $\mathbf{W}$ *that yields the solution*

$$\mathbf{W}\mathbf{R}\mathbf{W}^H = (\mathbf{I} + \boldsymbol{\Lambda})^{-1} \ \ \text{and} \ \ \mathbf{W}\mathbf{P}\mathbf{W}^T = (\mathbf{I} + \boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}. \tag{7}$$

Remark. We could have $\lambda_i = 0$ or $\lambda_i = \lambda_j$ for some diagonal entries of $\boldsymbol{\Lambda}$. In such cases, there is not one unique stationary point for the algorithm. This situation is similar to that for the strong uncorrelated transform, in which a unique solution is not guaranteed. Experience shows that the algorithm still accurately computes a strong uncorrelating transform satisfying (4) despite the fact that this transform may not be unique.

3 A Simple Algorithm for Tracking the Strong Uncorrelating Transform

In many applications, tracking versions of algorithms are desired. We seek a simpler version of (5) for tracking a strong-uncorrelating transform solution given a measured sequence $\mathbf{x}(k)$. Our second proposed algorithm replaces $\widehat{\mathbf{R}}(k)$ and $\widehat{\mathbf{P}}(k)$ in (5) with their instantaneous values $\mathbf{x}(k)\mathbf{x}^H(k)$ and $\mathbf{x}(k)\mathbf{x}^T(k)$ to yield

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{8}$$

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)[\mathbf{W}(k) - \mathbf{y}(k)\mathbf{y}^H(k)\mathbf{W}(k) - \mathrm{tri}[\mathbf{y}(k)\mathbf{y}^T(k)]\mathbf{W}(k)]. \tag{9}$$

This algorithm is particularly simple, requiring approximately $5m^2$ complex-valued multiply/adds at each iteration if an order-recursive procedure is used to compute $\mathrm{tri}[\mathbf{y}(k)\mathbf{y}^T(k)]\mathbf{W}(k)$. As in all similar adaptive algorithms, the step size sequence $\mu(k)$ controls both the data-averaging of the $\mathbf{x}(k)$ terms and the convergence performance of $\mathbf{W}(k)$. Care must be taken in choosing $\mu(k)$.

The algorithm in (8)–(9) is equivariant with respect to the mixing matrix $\mathbf{A}$ in (1). Moreover, because the discrete-time and differential averaged versions of (8)–(9) are the same as those for the updates in (5) and (17), respectively, Theorems 2 and 3 also apply to (8)–(9). Provided a suitably small step size is chosen and $\mathbf{x}(k)$ is a stationary input signal with distinct non-zero circularity coefficients, the only stable stationary point of (8)–(9) satisfies (7).

Eqns. (8)–(9) are closely related to simple decorrelation methods for real-valued signals [8]. One could view (8)–(9) as the complex extension of the natural gradient method in [8], with the additional feature that it computes the strong uncorrelating transform if $\mathbf{P} \neq \mathbf{0}$. In situations where $\mathbf{x}(k)$ is circularly-symmetric (*i.e.* $\mathbf{P} = \boldsymbol{\Lambda} = \mathbf{0}$), then $E\{\mathrm{tri}[\mathbf{y}(k)\mathbf{y}^T(k)]\} \approx \mathbf{0}$, such that (8)–(9) becomes a natural gradient algorithm for ordinary whitening of complex signals.

For source separation or clustering based on non-circularity, both (5) and (8)–(9) have the nice property that the sources $\{s_i(k)\}$ are grouped in $\mathbf{y}(k)$ in the order of their decreasing circularity coefficients. This property is maintained despite the fact that *the algorithm does not estimate the circularity coefficients of the sources explicitly.* A similar feature was noted for the algorithm in [7].

4 Simulations

We now explore the behaviors of the two proposed algorithms via simulations. The first set of simulations illustrate the algorithms' convergence behaviors when $\mathbf{A}$ is identifiable through the strong-uncorrelating transform. Let $\mathbf{s}(k)$ contain six zero-mean, unit-variance, uncorrelated, and non-circular Gaussian sources with distinct circularity coefficients $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\}$=$\{1, 0.8, 0.6, 0.4, 0.2, 0.1\}$. One hundred simulations are run, in which $\mathbf{A}$ is generated as a random mixing matrix with jointly-Gaussian real and imaginary elements. Both exponential ($\alpha = 0.999$), denoted by 'exp', and growing-window, denoted by 'lin', averaging of the sequences $\mathbf{x}(k)\mathbf{x}^H(k)$ and $\mathbf{x}(k)\mathbf{x}^T(k)$ with $\widehat{\mathbf{R}}(0) = \widehat{\mathbf{P}}(0) = 0.01\mathbf{I}$ were

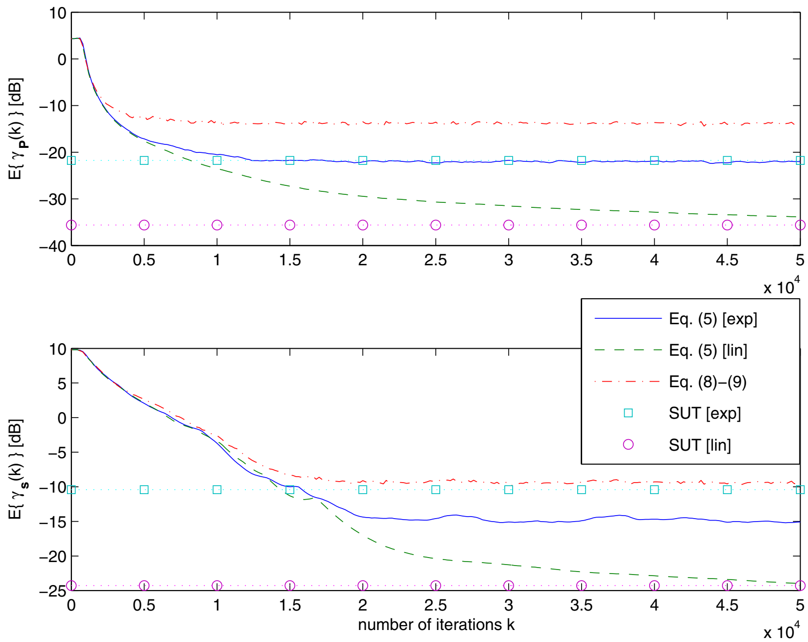

Fig. 1. Convergence of $E\{\gamma_{\mathbf{P}}(k)\}$ and $E\{\gamma_{\mathbf{s}}(k)\}$ in the first simulation example showing the proposed algorithms' successful estimation of the strong-uncorrelating transform

used to estimate $\widehat{\mathbf{R}}(k)$ and $\widehat{\mathbf{P}}(k)$ for two versions of (5). The combined system coefficient vector $\mathbf{C}(k) = \mathbf{W}(k)\mathbf{A}$ is computed and used to evaluate two metrics:

1. Pseudo-covariance Diagonalization: This cost verifies that the algorithms diagonalize the pseudo-covariance and is given by

$$\gamma_{\mathbf{P}}(k) = \frac{||\mathbf{C}(k)\boldsymbol{\Lambda}\mathbf{C}^T(k) - \mathrm{diag}[\mathbf{C}(k)\boldsymbol{\Lambda}\mathbf{C}^T(k)]||_F^2}{||\mathrm{diag}[\mathbf{C}(k)\boldsymbol{\Lambda}\mathbf{C}^T(k)]||_F^2}. \quad (10)$$

2. Source Separation Without De-rotation: This cost is the average of the inter-channel interferences of the combined system matrices $\mathbf{C}(k)$ and $\mathbf{C}^T(k)$, as

$$\gamma_{\mathbf{s}}(k) = \frac{1}{2m}\left(\sum_{i=1}^{n}\sum_{l=1}^{n}\frac{|c_{il}(k)|^2}{\max_{1\le i\le n}|c_{il}(k)|^2} + \frac{|c_{il}(k)|^2}{\max_{1\le l\le n}|c_{li}(k)|^2}\right) - 1. \quad (11)$$

Shown in Figure 1(a) and (b) are the evolutions of $E\{\gamma_{\mathbf{P}}(k)\}$ and $E\{\gamma_{\mathbf{s}}(k)\}$ for the various algorithms with their associated data averaging methods, where $\mu = \mu(k) = 0.007$ for (5) and (8)–(9). As can be seen, all versions of the algorithms diagonalize the pseudo-covariance matrix over time, and they also perform source separation for this scenario.

We now illustrate the behaviors of the simple algorithm in (8)–(9) in a more-practical setting. Let $\mathbf{s}(k)$ contain two BPSK and one 16-QAM source signals.

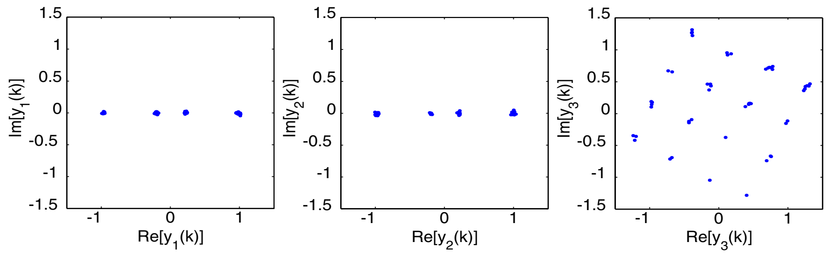

Fig. 2. Output signal constellations obtained by (8)–(9) for a source clustering task in wireless communications

The circularity coefficients in this situation are $\{\lambda_1, \lambda_2, \lambda_3\} = \{1, 1, 0\}$. The strong-uncorrelating transform applied to mixtures of these sources creates a combined system matrix $\mathbf{C}(k) = \mathbf{W}(k)\mathbf{A}$ in which the first two rows (resp. columns) are nearly orthogonal to the third row (resp. column). Thus, $y_1(k)$ and $y_2(k)$ largely contain mixtures of the two real-valued BPSK sources, and $y_3(k)$ largely contains the 16-QAM source. Shown in Figure 2 are the output signal constellations from $y_i(k)$, $i \in \{1, 2, 3\}$, $20000 \leq n \leq 25000$, obtained by applying (8)–(9) with $\mu = 0.0001$ to noisy mixtures of these sources, in which $\mathbf{A}$ contains jointly circular Gaussian entries with variance 2 and the (complex circular Gaussian) additive noise has variance 0.001. The first two outputs clearly show mixtures of the two real BPSK sources, whereas the last output contains the 16-QAM source.

5 Conclusions

The strong-uncorrelating transform is an important linear transform in complex independent component analysis. This paper describes two simple algorithms for adaptively estimating the strong-uncorrelating transform from known covariance and pseudo-covariance matrices and from measured signals, respectively. The algorithms are equivariant to the mixing system, and local stability analyses verify that they perform strong-uncorrelation reliably. Simulations illustrate their performances in separation and source clustering tasks.

References

1. J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," *Proc. IEEE Workshop Machine Learning Signal Processing,* Sao Luis, Brazil, pp. 183-191, Oct. 2004.
2. J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inform. Theory,* accepted.
3. V. D. Calhoun, T. Adali, G. D. Pearlson, P. C. van Zijl, and J. J. Pekar, "Independent component analysis of fMRI data in the complex domain," *Magn Reson. Med.*, vol. 48, pp. 180-192, 2002.

4. G.H. Golub and C.F. Van Loan, *Matrix Computations,* 3rd. ed. (Baltimore: Johns Hopkins Press, 1996).
5. C. Guo and S. Qiao, "A stable Lanczos tridiagonalization of complex symmetric matrices," Tech. Rep. CAS 03-08-SQ, Dept. of Comput. Software Engr., McMaster Univ., June 2003.
6. L. De Lathauwer and B. De Moor, "On the blind separation of non-circular sources," *Proc. XI European Signal Processing Conf.,* Toulouse, France, vol. II, pp. 99-102, Sept. 2002.
7. S.C. Douglas, "Simple algorithms for decorrelation-based blind source separation," *Proc. IEEE Workshop Neural Networks Signal Processing,* Martingy, Switzerland, pp. 545-554, Sept. 2002.
8. S.C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing,* vol. 45, pp. 2829-2842, Nov. 1997.
9. S.C. Douglas, "Fixed-point FastICA algorithms for the blind separation of complex-valued signal mixtures," *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2005.

Appendix

Proof of Theorem 1. Substituting the expressions for $\mathbf{R}$ and $\mathbf{P}$ in (2) for $\widehat{\mathbf{R}}(k)$ and $\widehat{\mathbf{P}}(k)$ in (5) and defining $\mathbf{C}(k) = \mathbf{W}(k)\mathbf{A}$, an equivalent expression for (5) is

$$\mathbf{C}(k+1) = \mathbf{C}(k) + \mu \left(\mathbf{I} - \mathbf{C}(k)\mathbf{C}^H(k) - \mathrm{tri}[\mathbf{C}(k)\boldsymbol{\Lambda}\mathbf{C}^T(k)]\right)\mathbf{C}(k), \quad (12)$$

which does not depend on $\mathbf{W}(k)$ or $\mathbf{A}$ individually.

Proof of Theorem 2. The stationary points of the algorithm are defined by

$$\left(\mathbf{I} - \mathbf{W}\mathbf{R}\mathbf{W}^H - \mathrm{tri}[\mathbf{W}\mathbf{P}\mathbf{W}^T]\right)\mathbf{W} = \mathbf{0}. \quad (13)$$

Clearly, $\mathbf{W} = \mathbf{0}$ defines one stationary point. The other stationary points are determined by the solutions of $\mathbf{M} = \mathbf{0}$, where

$$\mathbf{M} = \mathrm{tri}[\mathbf{W}\mathbf{P}\mathbf{W}^T] + \mathbf{W}\mathbf{R}\mathbf{W}^H - \mathbf{I}. \quad (14)$$

Consider the symmetric and anti-symmetric parts of $\mathbf{M}$ separately. The anti-symmetric part of $\mathbf{M}$ is

$$\mathbf{M}_a = \frac{1}{2}(\mathbf{M} - \mathbf{M}^H) \; = \; \frac{1}{2}\left(\mathrm{tri}[\mathbf{W}\mathbf{P}\mathbf{W}^T] - \mathrm{tri}[\mathbf{W}\mathbf{P}\mathbf{W}^T]^H\right). \quad (15)$$

For $\mathbf{M}_a = \mathbf{0}$, we must have that $\mathbf{W}\mathbf{P}\mathbf{W}^T = \mathbf{D}$, where $\mathbf{D}$ has real-valued but potentially-unordered entries. Under this condition, the symmetric part of $\mathbf{M}$ is

$$\mathbf{M}_s = \frac{1}{2}(\mathbf{M} + \mathbf{M}^H) \; = \; \mathbf{W}\mathbf{R}\mathbf{W}^H - \mathbf{I} + \mathbf{D}. \quad (16)$$

For $\mathbf{M}_s = \mathbf{0}$ to hold, we must have $\mathbf{W}\mathbf{R}\mathbf{W}^H = \mathbf{I} - \mathbf{D}$, which verifies (6). Moreover, since $\mathbf{R}$ is non-negative definite, the diagonal entries of $\mathbf{I} - \mathbf{D}$ are non-negative, and the diagonal entries of $\mathbf{D}$ must satisfy $0 < d_i \leq 1$.

Proof of Theorem 3. Consider the differential form of the update in (5):

$$\frac{d\mathbf{W}}{dt} = \mathbf{W} - \mathbf{W}\mathbf{R}\mathbf{W}^H\mathbf{W} - \mathrm{tri}[\mathbf{W}\mathbf{P}\mathbf{W}^T]\mathbf{W}. \tag{17}$$

Substituting the expressions for $\mathbf{R}$ and $\mathbf{P}$ in (2) into (17) and post-multiplying both sides of (17) by $\mathbf{A}$, we re-write (17) in the combined matrix $\mathbf{C} = \mathbf{W}\mathbf{A}$ as

$$\frac{d\mathbf{C}}{dt} = \mathbf{C} - \mathbf{C}\mathbf{C}^H\mathbf{C} - \mathrm{tri}[\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^T]\mathbf{C}. \tag{18}$$

First, assume that $\mathbf{C}$ is near a stationary point corresponding to $\mathbf{W} = 0$, and let $\mathbf{C} = \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is a matrix of small complex-valued entries. Then, we can rewrite the update in (18) in the entries of $\boldsymbol{\Delta}$ as

$$\frac{d\boldsymbol{\Delta}}{dt} = \boldsymbol{\Delta} + \mathcal{O}(\Delta_{ij}^2) \tag{19}$$

where $\mathcal{O}(\Delta_{ij}^2)$ denotes terms that are second and higher-order in the entries of $\boldsymbol{\Delta}$. Eq. (19) is exponentially unstable; $\mathbf{W} = \mathbf{0}$ is not a stable stationary point.

Now, assume that $\mathbf{C}$ is near a stationary point such that $\mathbf{C}_s\mathbf{C}_s^H = \mathbf{I} - \mathbf{D}$ and $\mathbf{C}_s\boldsymbol{\Lambda}\mathbf{C}_s^T = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix of real-valued scaling factors $\{d_i\}$ satisfying $0 < d_i \leq 1$, and let $\mathbf{C} = (\mathbf{I} + \boldsymbol{\Delta})\mathbf{C}_s$, where $\boldsymbol{\Delta}$ is a matrix of small complex-valued entries. Then, we can rewrite the update in (18) in the entries of $\boldsymbol{\Delta}$ as

$$\frac{d\boldsymbol{\Delta}}{dt} = -\boldsymbol{\Delta}(\mathbf{I} - \mathbf{D}) - (\mathbf{I} - \mathbf{D})\boldsymbol{\Delta}^H - \mathrm{tri}[\boldsymbol{\Delta}\mathbf{D} + \mathbf{D}\boldsymbol{\Delta}^T] + \mathcal{O}(\Delta_{ij}^2). \tag{20}$$

Ignoring second and higher-order terms, the diagonal entries of $\boldsymbol{\Delta}$ evolve as

$$\frac{d\Delta_{ii}}{dt} = -2\Delta_{ii}, \tag{21}$$

and they are exponentially convergent. The off-diagonal entries of $\boldsymbol{\Delta}$ evolve in a pairwise coupled manner and for $i < j$ satisfy

$$\frac{d\Delta_{ij}}{dt} = (-1 + d_j)\Delta_{ij} + (-1 + d_i)\Delta_{ji}^* \tag{22}$$

$$\frac{d\Delta_{ji}}{dt} = -\Delta_{ij} + (-1 + d_i)\Delta_{ji}^* - d_i\Delta_{ji} \tag{23}$$

Considering the real and imaginary parts of $\Delta_{ij} = \Delta_{R,ij} + j\Delta_{I,ij}$ and $\Delta_{ji} = \Delta_{R,ji} + j\Delta_{I,ji}$ jointly, we have

$$\frac{d}{dt}\begin{bmatrix} \Delta_{R,ij} \\ \Delta_{R,ji} \\ \Delta_{I,ij} \\ \Delta_{I,ji} \end{bmatrix} = \begin{bmatrix} -1 + d_j & -1 + d_i & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & -1 + d_j & 1 - d_i \\ 0 & 0 & -1 & 1 - 2d_i \end{bmatrix} \begin{bmatrix} \Delta_{R,ij} \\ \Delta_{R,ji} \\ \Delta_{I,ij} \\ \Delta_{I,ji} \end{bmatrix}. \tag{24}$$

For these terms to be convergent, the (2×2) dominant sub-matrices in the above transition matrix must have negative real parts. Recall that $0 < d_l \leq 1$ for all

$1 \leq l \leq m$ by the stationary point condition. Then, the eigenvalue of the first dominant (2×2) matrix with the largest real part is

$$r_{max}^{(1)} = \frac{2-d_j}{2}\left(-1+\sqrt{1-4\frac{d_i-d_j}{(2-d_j)^2}}\right). \tag{25}$$

For $\Re e[r_{max}^{(1)}] < 0$, we require that $d_i > d_j$. With this result, the eigenvalue of the second dominant (2×2) matrix with the largest real part is

$$r_{max}^{(2)} = \frac{2d_i-d_j}{2}\left(-1+\sqrt{1-4\frac{d_i+d_j-2d_id_j}{(2d_i-d_j)^2}}\right), \tag{26}$$

which for $d_i > d_j$ is guaranteed to satisfy $\Re e[r_{max}^{(2)}] < 0$. Thus, the only stable stationary point of the algorithm is when $d_1 > d_2 > \cdots > d_m$.

Now, consider the only stable stationary point solution in (6). Define $\underline{\mathbf{W}} = (\mathbf{I}-\mathbf{D})^{-1/2}\mathbf{W}$ such that

$$\underline{\mathbf{W}}\mathbf{R}\underline{\mathbf{W}}^H = \mathbf{I} \text{ and } \underline{\mathbf{W}}\mathbf{P}\underline{\mathbf{W}}^T = (\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}. \tag{27}$$

It is straightforward to show that $d_i > d_j$ implies $d_i/(1-d_i) > d_j/(1-d_j)$, such that $(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D}$ has ordered entries. Eqn. (27) is exactly the strong uncorrelating transform, such that $(\mathbf{I}-\mathbf{D})^{-1}\mathbf{D} = \boldsymbol{\Lambda}$, or $\mathbf{D} = (\mathbf{I}+\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}$ and $\mathbf{I}-\mathbf{D} = (\mathbf{I}+\boldsymbol{\Lambda})^{-1}$. This proves the theorem.