Non-negative Matrix Factorization for Self-calibration of Photometric Redshift Scatter in Weak-lensing Surveys

Le Zhang1, Yu Yu1,2, and Pengjie Zhang1,3,4

1 Department of Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, China; leizhang@sjtu.edu.cn
2 Key laboratory for research in galaxies and cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai, 200030, China
3 IFSA Collaborative Innovation Center, Shanghai Jiao Tong University, Shanghai 200240, China
4 Tsung-Dao Lee Institute, Shanghai 200240, China

Received 2016 December 13; revised 2017 August 28; accepted 2017 September 11; published 2017 October 10

Abstract

Photo-z error is one of the major sources of systematic degradations degrading the accuracy of weak-lensing cosmological inferences. Zhang et al. proposed a self-calibration method combining galaxy–galaxy correlations and galaxy–shear correlations between different photo-z bins. Fisher matrix analysis shows that it can determine the rate of photo-z outliers at a level of 0.01%–1% merely using photometric data and do not rely on any prior knowledge. In this paper, we develop a new algorithm to implement this method by solving a constrained nonlinear optimization problem arising in the self-calibration process. Based on the techniques of fixed-point iteration and non-negative matrix factorization, the proposed algorithm can efficiently and robustly reconstruct the scattering probabilities between the true-z and photo-z bins. The algorithm has been tested extensively by applying it to mock data from simulated stage IV weak-lensing projects. We find that the algorithm provides a successful recovery of the scatter rates at the level of 0.01%–1%, and the true mean redshifts of photo-z bins at the level of 0.001, which may satisfy the requirements in future lensing surveys.

Key words: cosmology: observations – gravitational lensing: weak – techniques: photometric

1. Introduction

Weak gravitational lensing is one of the most powerful cosmological probes to study the distribution of dark matter, the dynamics of dark energy and the formation of large scale structures (LSSs) in the universe, as well as the nature of gravity at cosmological scales (e.g., Refregier 2003; Albrecht et al. 2006; Jain & Zhang 2008; Weinberg et al. 2013; Hurtet al. 2015). By statistically measuring the cosmic shear signals, existing surveys, including CFHTLS5 (Heymans et al. 2013; Kilbinger et al. 2013; Fu et al. 2014) and SDSS (Mandelbaum et al. 2013; Huff et al. 2014), have already tightly constrained a combination of the amplitude of density fluctuation $\sigma_8$ and the matter density parameter $\Omega_m$. Ongoing weak-lensing surveys such as DES6 and HSC,7 along with the planned stage IV surveys such as Euclid8 and LSST,9 will further increase the survey depth to $z_S > 1$, and increase the available number of source galaxies by orders of magnitude.

Despite the great capability of these surveys to achieve high precision weak-lensing measurements, weak-lensing cosmology suffers from various sources of systematic errors such as shape measurement errors (Heymans et al. 2006; Massey et al. 2007; Mandelbaum et al. 2015) and intrinsic alignments of source galaxies (Hirata et al. 2004; Mandelbaum et al. 2006; Hirata et al. 2007; Okumura et al. 2009; Okumura & Jing 2009; Troxel & Ishak 2015). Another major systematic error is the photometric redshift (photo-z) errors (Ma et al. 2006; Bernstein & Huterer 2010). Even though the telescope technology developments have made rapid progress, spectroscopic redshift surveys of a large number of galaxies are still very time consuming especially for high redshift large scale galaxy surveys. All of the above-mentioned large galaxy surveys will thus have photometric rather than spectroscopic redshift identifications. This results in non-negligible photo-z errors. It is found that both the bias and scatter need to be controlled at the level of $10^{-3}$ to avoid a considerable degradation in cosmological parameter accuracies (Huterer et al. 2006; Ma et al. 2006; Kitching et al. 2008). Therefore, it is crucial to precisely determine the true redshift distribution of galaxies in a photometric weak-lensing survey.

Besides efforts of directly improving the photo-z estimation algorithms, attempts have been made to infer the true redshift distribution of source galaxies by the LSS statistics. For example, the cross-correlation techniques that measure redshift errors by cross correlating the photo-z samples with spectroscopic ones in overlapping survey areas have been proposed recently (Newman 2008; Matthews & Newman 2010, 2012; Schulz 2010; McQuinn & White 2013; Ménard et al. 2013; Schmidt et al. 2013). It has been applied in various surveys (e.g., Mitchell-Wynne et al. 2012; Choi et al. 2016; Johnson et al. 2017; Rahman et al. 2016a, 2016b; Scottez et al. 2016). This cross-calibration technique is powerful. However, it suffers from two potentially significant systematic errors. For spectroscopic galaxies in the redshift bin $z_S$, its cross-correlation with the photometric galaxies of a given photo-z bin is $\propto b_S(z_S) b_p(z_S) r_{SP}(z_S) P_{P_{\rightarrow S}}$. Here, $b_S$ is the bias of spectroscopic galaxies. $b_p(z_S)$ is the bias of photometric galaxies whose true redshifts fall in the redshift bin $z_S$, $P_{P_{\rightarrow S}}$ is the fraction of these photometric galaxies, and $r_{SP}(z_S)$ is the cross-correlation coefficient between these galaxies and the spectroscopic galaxies at $z_S$, $P_{P_{\rightarrow S}}$ quantifies the photo-z error, but it is degenerate with $b_p(z_S)$ and $r_{SP}(z_S)$. The degeneracy with $b_p$ has been investigated (de Putter et al. 2014; Rahman

5 http://www.cfht.hawaii.edu/Science/CFHTLS/
6 http://www.darkenergysurvey.org/
7 http://www.naoj.org/Projects/HSC/index.html
8 http://sci.esa.int/euclid/
9 http://www.lsst.org/lsst/
et al. 2015). It can result in significant systematic error in the determined photo-z or even complete failure (de Putter et al. 2014). The second degeneracy is seldom addressed in the literature. But even a small stochasticity \((1 - s_{sp} \sim O(0.01))\) can result in \(O(1\%)\) systematic error in \(P_{\ell}\). Therefore it severely limits the application of cross-calibration in the nonlinear regime where we expect significant stochasticity.

Meanwhile, the self-calibration techniques, which does not rely on external spectroscopic data, have been proposed and applied in real data (Schneider et al. 2006; Erben et al. 2009; Benjamin et al. 2010; Quadri & Williams 2010; Zhang et al. 2010). Under the Limber approximation, the galaxy–galaxy angular correlation between two different redshift bins should vanish if there are no photo-z errors. Therefore, the galaxy–galaxy correlation between redshift bins can be used to infer the photo-z outlier rate. The galaxy–galaxy correlation can be measured to high precision by future weak-lensing surveys, leading to robust constraint on photo-z error. However, an intrinsic degeneracy between up and down scatters would severely limit the accuracy of the determination of outlier rates, since a non-zero galaxy– galaxy correlation between redshift bins can be induced by both up and down scatters and galaxy–galaxy measurement alone cannot break this degeneracy (Benjamin et al. 2010; Zhang et al. 2010).

Zhang et al. (2010) pointed out that the galaxy–shear angular correlation available in the same weak-lensing survey naturally breaks this degeneracy, leading to significant or even orders of magnitude improvement in the calibration accuracy. This establishes the galaxy–shear correlation as an indispensable part of the self-calibration technique. The self-calibration technique does not rely on any assumptions on cosmological priors, galaxy bias and parameterization of the photo-z probability distribution function. Through the measurements of galaxy–galaxy correlation and galaxy–shear correlation between different photo-z bins, it can accurately reconstruct the full relation between the photometric and spectroscopic redshift and, for a given photo-z bin, derive the fraction of galaxies which in fact come from a distinct true redshift bin. For a stage IV projects like LSST, it can detect outliers with rate as low as 0.01%–1%.

The theoretical reconstruction accuracy of such method is obtained from the Fisher matrix analysis (Zhang et al. 2010). In practice, however, implementing such calibration process requires solving the set of nonlinear constrained matrix equations, which is a computationally hard problem. So far, implementations in data analysis such as the pairwise analysis (Erben et al. 2009; Benjamin et al. 2010) are based on significant simplifications of these equations and therefore cannot meet the requirement of stage IV projects.

The exact and accurate implementation of the self-calibration is thus the focus of this paper. Unfortunately, none of the widely used algorithms proposed in the literature is capable of finding a reliable solution that are close enough to the simulation truth, e.g., the Powell’s method, the quasi-Newton method and expectation-maximization method. Even for Monte Carlo-based fitting methods, the estimation of a large number of parameters, typically in the order of 100, to the desired accuracy is still prohibitive computationally. The reason is not only the non-convex nonlinear optimization problem with numerous unknown parameters need to be solved, but also multiple constraints have to be imposed for those unknowns.

In this paper, we report on the development of a novel algorithm that can be implemented in the self-calibration process. The algorithm is based on the techniques of fixed-point iteration and non-negative matrix factorization (NMF). In Section 2, we briefly review the self-calibration theory and give a detailed description about the proposed algorithm. Section 3 presents the application of this algorithm to mock photometric data and shows the main results. Finally, Section 4 provides discussion and concluding remarks.

The notations and conventions used in this paper are as follows. The superscript “\(P\)” denotes a property in photo-z bins, and the superscript “\(R\)” denotes a corresponding property in true-z bins. The capital “\(G\)” denotes gravitational lensing, to more specific, the lensing convergence converted from the more direct observable cosmic shear, and the little “\(g\)” denotes galaxy number density (or overdensity). Furthermore, upper case letters are used to denote matrices, let \(A\) be a matrix. Then the \((i,j)\)th entry of a matrix \(A\) is referred to by either \(A_{ij}\) or \(a_{ij}\) and its transpose by \(A^T\).

### 2. Problem Formalization

As proposed by Zhang et al. (2010), the rate of photo-z outliers can be determined by using galaxy–galaxy and lensing-galaxy measurements in photometric data from the original lensing survey. Let us start with the measurement equations in terms of photo-z scatters,

\[
P_{ij}^{G\ell}(\ell) = \sum_k P_{ik} P_{kj} C_{kk}^{G\ell}(\ell) + \delta N_{ij}^{G\ell}(\ell),
\]

where \(P_{ij}\) represents the scattering probability between the \(i\)th true-z bin and the \(j\)th photo-z bin. To be exact, \(P_{ij} = N_{i-j}/N^j\), where \(N^j\) is the total number of galaxies in the \(j\)th photo-z bin and \(N_{i-j}\) of them come from the \(i\)th true-z bin. \(P_{ij}\) satisfies two conditions, \(P_{ij} \geq 0\) and \(\sum_j P_{ij} = 1\). Notice that the first subscript of \(P_{ij}\) labels the true-z bin and the second subscript denotes the label of photo-z bin. Therefore it is in general asymmetric \((P_{ij} \neq P_{ji})\). \(C_{ij}^{G\ell}(\ell)\) stands for the measured galaxy power spectrum between the \(i\)th and \(j\)th photo-z bins, and \(C_{ij}^{G\ell}(\ell)\) denotes the measured cross-correlation power spectrum between the lensing convergence in the \(i\)th photo-z bin and the galaxy number density in the \(j\)th photo-z bin. As defined above, \(C_{ij}^{G\ell}(\ell)\) and \(C_{ij}^{G\ell}(\ell)\) are the corresponding power spectra in the true redshift bins. For a given \(\ell\), \(C_{ij}^{G\ell}(\ell)\) is expected to be \(C_{ij}^{G\ell}(\ell) = 0\), as the galaxy cross-correlation between non-overlapping redshift bins would vanish under the Limber approximation. Furthermore, using the notation where a larger index corresponds to a higher redshift, we also expect \(C_{g\ell}^{G\ell}(\ell) = 0\) and \(C_{g\ell}^{G\ell}(\ell) = 0\). This is because that, in the absence of lensing magnification bias, only source galaxies behind a lens can be lensed, resulting in non-zero lensing-galaxy correlations. Therefore, we are only summing over the diagonal \((k = m)\) and lower triangular \((k \geq m)\) components in the above measurement equations.

In realistic cases, the measured power spectra are certainly contaminated by shot noise, so that we introduce the shot noise terms, \(\delta N_{ij}^{G\ell}(\ell)\) and \(\delta N_{ij}^{G\ell}(\ell)\), which represent the fluctuations of the associated shot noise in measurements. Note that here we have implicitly subtracted the ensemble average of shot noise out of
the observations as it cannot bias the estimate of outliers. That is
the reason why only the fluctuations \( \delta N_{\text{ng}}P \) and \( \delta N_{\text{gG}}P \) are taken
into account in the measurement equations.

For a given \( \ell \), if we split galaxies into \( n \) photo-z bins, the same
equations but in matrix notation read

\[
C_{\ell P}^{\text{gg}} = P^T C_{\ell}^{\text{gg}P} P + \delta N_{\text{ng}}^{\text{gg}},
\]

\[
C_{\ell P}^{\text{gG}} = P^T C_{\ell}^{\text{gG}P} P + \delta N_{\text{ng}}^{\text{gG}},
\]

where \( P \in \mathbb{R}^{n \times n} \) is a non-symmetric matrix, subject to the
column-sum-to-one constraint, i.e., \( \sum_i P_{ij} = 1 \), for all \( j \). The
matrix \( P \) is also a so-called non-negative matrix, that is, \( P_{ij} \geq 0 \)
for all \( (i, j) \). \( C_{\ell}^{\text{gg}P} \in \mathbb{R}^{n \times n} \) is a non-negative diagonal
matrix, and \( C_{\ell}^{\text{gG}P} \in \mathbb{R}^{n \times n} \) is a lower triangular matrix.

For a given data set, \( C = \{ C_{\ell}^{\text{gg}P}, C_{\ell}^{\text{gG}P} \} \), the parameters,
\( \theta = \{ P, C_{\ell}^{\text{gg}P}, C_{\ell}^{\text{gG}P} \} \), can be estimated by the posterior
parameter distribution \( P(\theta | C) \). Using Bayes’ rule, \( P(\theta | C) \)
can be expressed by \( P(\theta | C^D) \propto L(C^D \theta) P(\theta) \), where the
likelihood function of the data is \( L(C^D \theta) \) and the parameter
prior is \( P(\theta) \). Assuming a flat prior on the parameters, we will
see that the best-fitted parameters can be derived by maximizing
the likelihood function.

As the fluctuations are induced by shot noise which is
assumed to be uncorrelated at different \( \ell \) modes and \( \ell \) bins,
different power spectrum measurements thus are assumed to be
uncorrelated and can be well approximated by a Gaussian
distribution if there are sufficiently large number of indepen-
dent \( \ell \) modes in each multipole \( \ell \) bin. With these assumptions,
the likelihood function then becomes a simple multivariate
Gaussian distribution with the covariance matrix \( \Sigma_N \), i.e.,
\( P(C^D \theta) \propto N(C^D - \hat{C}(\theta), \Sigma_N) \). Each element in \( \Sigma_N \) is
evaluated by \( \langle \delta n_{\gamma P}(\ell) \delta n_{\gamma P}(\ell') \rangle = \langle \sigma_{\gamma P}^2 \rangle \delta_{\ell \ell'} \delta_{\ell \ell'}, \) with
\( \alpha, \beta = gg, \ Gg, \) and

\[
\langle \sigma_{\gamma P}^2 \rangle = \frac{1}{2\ell + 1} \Delta f_{\text{sky}} \frac{1}{\hat{n}_i} (1 + \delta_{ij}),
\]

\[
\langle \sigma_{\gamma P}^2 \rangle = \frac{1}{2\ell + 1} \Delta f_{\text{sky}} \frac{\gamma_{\text{rms}}^2}{\hat{n}_i \hat{n}_j},
\]

for data in bins of width \( \Delta f \), where \( f_{\text{sky}} \) is the fraction of the sky
observed, \( \hat{n}_i \) is the mean galaxy surface density in the \( i \)th redshift
bin per steradian and \( \gamma_{\text{rms}} \) denotes the rms dispersion in the shear
measurement induced by the galaxy intrinsic ellipticities.

It is straightforward to show that the logarithm of the likelihood
function can be finally simplified to

\[
\ln L \propto - \sum_{\ell} \left[ \sum_{i} \left( \sigma_{\gamma P}^2(\ell) \right)^{-2} \left( C_{\ell P}^{\gamma P} (\ell) - \hat{C}_{\ell P}^{\gamma P} (\ell) \right)^2 \right. \\
+ \sum_{i} \left( \sigma_{\gamma P}^2(\ell) \right)^{-2} \left( C_{\ell P}^{\gamma P} (\ell) - \hat{C}_{\ell P}^{\gamma P} (\ell) \right)^2 \right],
\]

where \( \hat{C}_{\ell P}^{\gamma P} (\ell) \equiv P^T C_{\ell}^{\gamma P} P \) and \( \hat{C}_{\ell P}^{\gamma P} (\ell) \equiv P^T C_{\ell}^{\gamma P} P \) are
the estimated quantities in terms of \( \{ P, C_{\ell}^{\gamma P}, C_{\ell}^{\gamma P} \} \).

Note that, in an ideal case that the noise is completely
neglected, the model-independent self-calibration technique
requires solving the set of nonlinear matrix equations given by
Equations (3) and (4) for all \( P, C_{\ell}^{\gamma P}, \) and \( C_{\ell}^{\gamma P} \) simultaneously
with the required constraints. In the next section, we will first
present a new algorithm for solving such nonlinear problem in
the idealized noise-free case, and then will come back to
the realistic situation in Section 2.2.

### 2.1. Fixed-point-based Algorithm

In the absence of shot noise, the data depend deterministi-
cally on the matrices \( P, C_{\ell}^{ggP} \) and \( C_{\ell}^{gGP} \). As the number of
measurements in many multipole \( \ell \) bins is always much larger
than that of the unknown parameters, in principle, the unique
exact solution exists for Equations (3) and (4). However, due to
quadratic dependence on \( P \) and linear on \( C_{\ell}^{ggP} \) and \( C_{\ell}^{gGP} \), the
analytical solution remains unknown for such complex system.
We therefore resort to a numerical method.

The observation that, the problem can be split into two
subproblems, suggests a method for solving the two subproblems
dependently. According to the form of Equation (3) where
\( C_{\ell}^{ggP} \) is a diagonal matrix, one can find a matrix, which has rows of “true” \( P \) but in permuted sequence, also satisfies Equation (3).
On the other hand, Equation (4) alone can uniquely determine
the order of the rows since only the “true” \( P \) will make the
reconstructed \( \hat{C}_{\ell}^{gGP} \) to be lower triangular matrices, which
implies all of the remaining elements, i.e., \( \{ C_{\ell}^{gGP} \}_{ij} \) with \( i < j \),
are exactly zeros. The reconstruction error thus can be
quantitatively measured by

\[
ea_{\ell P} \equiv \sum_{i-j} \left( \langle C_{\ell P}^{gGP} \rangle_{ij} - \hat{C}_{\ell P}^{gGP} (\ell) \right)^2 \]  

where \( \{ C_{\ell}^{gGP} \}_{ij} = \sum_{k} \{ P \}_{ik} \{ C_{\ell}^{gGP} \}_{kj} \{ P \}_{lj} \}_{k} \). For these concerns, we
develop a novel fixed-point-based algorithm, which is capable of solving such high dimensional nonlinear problem iteratively. We first solve for the relatively simpler problem of Equation (3), and then the true \( P \) can be obtained by utilizing Equation (4) to break down the permutation degeneracy. The procedure of the algorithm is summarized in Algorithm 1.

**Algorithm 1.** The fixed-point-based algorithm for solving
Equations (3) and (4) in the absence of shot noise.

---

**Step 1:** solving for Equation (3) ---

**Input:** Non-negative data matrices \( C_{\ell P}^{ggP} \) for all \( \ell \).

**initialize:**

(a) Assign \( P \) to a random positive matrix with

\[ \sum_{j} P_{ij} = 1, \text{ for all } i \]

(b) Set \( C_{\ell}^{gGP} = P^T C_{\ell}^{ggP} P^{-1} \), for all \( \ell \)

**repeat** for \( \ell = 1 \) to \( n \) do

\( Q_{\ell} = C_{\ell}^{gGP} P^{-1} \)

\( P_{\ell} = \text{Abs} \left( \sum_{j} Q_{\ell j} (\sum_{j} C_{\ell}^{gGP})^{-1} \right) \)

\( v_{ij} = \sum_{j} P_{\ell ji} \), for all \( i, j \) % normalization

end for

\( C_{\ell}^{gGP} = \text{Abs} \left( \text{Diag} \{ C_{\ell}^{gGP} \} \right) \), for all \( \ell \)

until a convergence criterion is satisfied

**Output:** \( P \)

---

**Step 2:** determine the row order of \( P \) ---

**Input:** Data matrices \( C_{\ell}^{gGP} \) for all \( \ell \), and \( P \) from Step 1

**repeat** 1. Randomly swap two rows of \( P \)

2. Compute \( e_{\ell P} \) with Equation (8)

until \( e_{\ell P} \leq \epsilon \) (a typical value of \( \epsilon = 10^{-9} \))

**Output:** \( P \)
Note that, there are two important tricks implemented into Algorithm 1. One trick is that $P$ is successively updated by averaged $Q_i$ and $C_{\ell}^{gg,R}$ over all $\ell$, similar to the successive over relaxation method. The averaging scheme allows a remarkable convergence rate, and can make a solution significantly stabilized to desired precision after some iterations, which would otherwise diverge.

The other one is to introduce the vectors $v_{\ell}$, which is based on fixed-point method and is essential for this algorithm. This method has two-fold advantages as follows.

First, in order to deal with the constraint in $P$, the commonly used normalization step can be imposed during iterations to enforce the unitary sum for each column of $P$. However, we find such normalization step may cause unstable and slow convergence. Instead, we introduce vectors $v_{\ell}$ for rescaling each element of columns in $P$ with different values, which can be regarded as slight perturbations in $P$. It is straightforward to see that all of the elements in the rescaling vectors are unity when $P$ satisfies Equation (3) with the column-sum-to-one constraint. For this reason, the perfect recovery is necessarily a fixed point of the update rules.

Second, as we know, the solution obtained by an iterative process may prematurely get stuck in an undesirable local maximum before reaching the global maximum (i.e., the true solution), especially for a constrained optimization problem with a large number of parameters. Since the sum of each column of $P$ is not exactly unity before $P$ converges to the true value, the vectors $v_{\ell}$ can slightly relax the column-sum-to-one constraint to overcome local maxima.

Simulation results are given in Section 3 and will show that the fixed-point-based algorithm converges fast, and can efficiently solve the set of nonlinear matrix equations in an iterative way until the desired accuracy is attained.

### 2.2. Optimization Algorithm

In reality, observations are corrupted by measurement noise, and hence the fixed-point-based algorithm may not be desirable for a direct application to noisy data.

The standard approach for parameter estimate under a given noisy data model is by maximizing the likelihood function. Thanks to the fact that a maximum-likelihood estimate of the parameters is equivalent to a least-squares estimate as long as the errors for all data points are independent and belong to a same Gaussian distribution. The measurement noise in weak-lensing surveys is usually assumed to be a shot noise well approximated by a Gaussian distribution. If we appropriately choose the bin widths of multipole $\ell$ and redshift $z$, the noise levels $\sigma_{\ell}^{gg}(\ell)$ in the measured galaxy–galaxy power spectrum would be approximatively identical over all $\ell$ bins. From Equation (5), we can find that the estimated noise levels have a factor of 2 difference between the diagonal ($i=j$) and off-diagonal ($i \neq j$) components, which would make the least-squares estimate slightly deviate from the maximum-likelihood estimate. However, such deviation is expected to be negligible as the noise levels are very small quantities compared to $C_{\ell}^{gg,R}$, if one choose bin widths of $\Delta \ell$ and $\Delta z$ sufficiently large. Therefore, the problem in Equation (3) can be straightforwardly reformulated as the following optimization problem:

\[
\min_{C_{\ell}^{gg,R}} J(P; C_{\ell}^{gg,R}) = \frac{1}{2} \sum_\ell ||C_{\ell}^{gg,P} - P^T C_{\ell}^{gg,R} P||_F^2, \quad (9)
\]

where $||.||_F$ is the Frobenius norm. We aim at minimizing the objective function $J$, which measures a accumulated decomposition error across all data matrices. Since we seek to represent each non-negative data matrix as the product of non-negative matrices, NMF is well-suited to solve this constrained optimization problem.

#### 2.2.1. Non-negative Matrix Factorization

In this section, we will first briefly review the background of the NMF method, and then describe the proposed approach based on NMF for solving the optimization problem in Equation (9).

Given an arbitrary $m \times n$ non-negative matrix $V$, NMF finds non-negative matrices $W$ and $H$ such that

\[
\min_{W,H} ||V - WH||_F^2 \quad \text{s.t.} \; W, H \geq 0, \quad (10)
\]

where $W \in \mathbb{R}^{m \times r}, H \in \mathbb{R}^{r \times n}$. Lee & Seung (2001) has found the following “multiplicative update rules” to minimize this conventional least-squares error and also the error is not non-increasing during iterations under these update rules:

\[
W_{ij} \leftarrow W_{ij} \left( V H_{ij} \right)^{W H_{ij}} H_{ij} \left( W^T V \right)^{W H_{ij}}. \quad (11)
\]

For constructing the multiplicative update rules adapted to the optimization problem in Equation (9), the most difficult part is that the objective is quartic with respect to $P$. To circumvent this situation, the natural way to update $P$ is to translate this tri-factor NMF to the standard bi-factor NMF. For example, one can use a simple so-called “splitting” technique. That is, the two appearances of $P$ in Equation (9) are represented by two different matrices, say $P_L$ and $P_R$, which are optimized independently. The optimization problem now amounts to minimizing

\[
\frac{1}{2} \sum_\ell ||C_{\ell}^{gg,P} - P_L^T C_{\ell}^{gg,R} P_R||_F^2. \quad (12)
\]

Therefore, one can update $P_L$ and $P_R$ alternatively and iteratively. After the convergence of a series of iterations, it is hoped that those two matrices obtained happen to be equal. However, we find that this is not always the case in practice for this specific problem.

Since we know $P_L$ and $P_R$ should be equal when the correct solution is obtained, enforcing $P_L = P_R$ after the update of either $P_L$ or $P_R$ at each iteration is suitable for this purpose. Note that this cannot guarantee monotonic decreasing of the objective function in general, which may lead to non-optimal solutions. However, the simulation results show that this approach is fairly robust and effective. The successful recovery of $P$ can be achieved as long as the initial guess for $P$ is not far from the optimal one. Since Algorithm 1 can provide an estimate quite close to the true value, the requirement for the initial guess for $P$ thus would be met using the fixed-point-based algorithm as an initialization.
Now we turn to develop the update rules for \( P_L, P_R \), and \( C_{gg}^{\ell R} \) by alternately updating each matrix while keeping the other matrices fixed. For details on the derivation of these update rules we refer to the Appendix. Finally, the NMF-based algorithm dedicated to self-calibrating the photo-z scatters from realistic measurements is specified in Algorithm 2.

**Algorithm 2.** The NMF-based algorithm to estimate \( P \) from noisy data by solving for the problem of Equation (9). This algorithm should be run over many times with different starting points in order to find the globally optimal estimate for \( P \). Note that the Appendix elaborates on the notations and conventions.

---

**Input:** Non-negative data matrices \( C_{gg}^{\ell R} \) for all \( \ell \). 

**Initialize:** 
1. Assign \( P \) from Algorithm 1
2. Compute \( C_{gg}^{\ell R} \) with Equation (27)

**Repeat**
1. Update \( W \): fixing \( H_L \), updating \( W \) with Equation (25)
2. Update \( H_L \): imposing \( H_L = C_{gg}^{\ell R} W^T \)
3. Update \( C_{gg}^{\ell R} \): fixing \( W \), updating \( C_{gg}^{\ell R} \) with Equation (27)

**Until** convergence criterion is satisfied (e.g., the change of each element in \( W \) is below \( 10^{-4} \) or the maximum number of iterations is reached, \( n_{out} = 3 \times 10^3 \))

**Output:** \( P \) (as \( P = W^2 \)) and \( C_{gg}^{\ell R} \).

We note that the optimization problem of Equation (9) is not convex in \( P \) and \( C_{gg}^{\ell R} \) together, which means that our proposed algorithm can only guarantee, if at all, to converge to a local minimum. Hence, the estimated parameters are somewhat dependent on initial conditions, and good initializations can significantly eliminate unwanted solutions that can lead to large decomposition errors. Thanks to Algorithm 1, which can achieve a near global minimization successfully, in practice, we try to run this NMF-based optimization solver over many times with different initializations, and use the best minimum found (i.e., the minimum \( J \)) as the global minimum to determine the globally optimal estimate of \( P \).

**3. Tests with Mock Data**

In this section, we test our proposed algorithms through mock data mimicking stage IV weak-lensing experiments in order to demonstrate their effectiveness and robustness in self-calibrating photo-z scatter in weak-lensing surveys. There exist several exciting galaxy survey proposals in “Stage IV” lensing surveys such as Euclid, LSST, and WFIRST. We mostly follow the data set and experimental settings in Zhang et al. (2010; and references therein), whose survey area, galaxy number density, and redshift distribution are similar to that of LSST.

Following Huterer et al. (2006) and Zhan & Knox (2006), the fiducial galaxy redshift distribution \( n(z_p) \) is chosen to have the form

\[
n(z_p) = x^2 \exp(-x)dx/2,
\]

where \( x = z_p / z_0 \) with \( z_0 = 0.45 \).

As mentioned in Section 2.2, we have to appropriately choose the bin widths of \( \Delta \ell \) and \( \Delta z \) such that an almost identical shot noise level distributes both at each \( \ell \)-bin and each \( z \)-bin in the measured galaxy–galaxy power spectrum. Although the number of \( z \)-bins can be chosen arbitrarily, Ma et al. (2006) shows that we cannot improve the statistical errors on cosmological parameter constraints with \( n \geq 5 \). Hence, we choose the following binning scheme throughout the paper. Using Equation (13), we divide the galaxies into 5 \( z \)-bins at the range of \( 0 \leq z < 4 \) as \([0, 0.688), [0.688, 1.023), [1.023, 1.389), [1.389, 1.906), [1.906, 4.0]\), each bin containing an equal number of galaxies, while we divide the total \( \ell \)-range \( 20 \leq \ell < 1000 \) into \( N_\ell = 6 \) spectra bands: \([20, 408), [409, 577), [578, 707), [708, 816), [817, 913), [814, 1000]\), approximately with a constant \( \ell\Delta \ell \approx 82729 \) for each bin.

To determine the noise covariance matrix \( \sigma_{gg}^{\ell R} \) and \( \sigma_{gg}^{\ell R} \) with Equation (13), we adopt a fiducial value \( \bar{n}_g = 40 \) per arcmin\(^2\) and \( f_{sky} = 0.5 \). The rms dispersion in the shear measurement induced by the galaxy intrinsic ellipticities is adopted as \( \gamma_{int} = 0.2 \). For each simulation, the realizations of shot noise fluctuations are generated using such a noise model.

The fiducial scatters are obtained based on the simulated data from Bernstein & Huterer (2010), which is produced for a SNAP-like survey using the method described in Jouvel et al. (2009). Figure 1 illustrates the original photo-z scatters in this data and the corresponding binned scatters adopted for this study. The left panel shows the \( z_p-z_S \) distribution produced by the data having 177210 galaxies, where \( z_S \) denotes the spectroscopic redshift. As we can see, the photometric redshifts in some island-shaped regions are grossly misestimated, leading to catastrophic redshift errors. For example, an “island” is at \( z_p > 2.5, z_S < 0.6 \), which is probably due to confusion between high-\( z \) Lyman breaks and low-\( z \) 400 nm breaks.

According to the above binning scheme, the corresponding binned scatters \( P_{ij} \) are shown in the right panel in Figure 1. We have discarded the galaxy samples with \( z_p > 4 \) in the data since we know that such galaxies have large photo-z catastrophic errors, while such galaxies merely account for \( <1 \% \) of the total and neglecting them should not lead to loss of information. The purpose of this study is to reconstruct all \( P_{ij} \) using our proposed algorithms.

To generate the fiducial power spectra \( C_{gg}^{\ell R} \) and \( C_{gg}^{\ell R} \), we use the public code CLASS (Lesgourgues 2011) and adopt a flat \( \Lambda \)CDM cosmology with \( \Omega_m = 0.268 \), \( \Omega_\Lambda = 1 - \Omega_m \), \( \Omega_b = 0.045 \), \( h_8 = 0.85 \), \( n_s = 1 \), and \( h = 0.71 \). For simplicity, we assume the fiducial galaxy bias to be \( b_g = 1 \).

Under the above simulation settings, an example of the resulting cross-power spectrum and its statistical uncertainty arising from the shot noise fluctuations is shown in Figure 2. As expected, the photo-z errors induce a non-zero, but quite small cross-correlations between galaxies at different redshift bins. The associated uncertainties across all \( \ell \)-bins are almost identical as desired, with \( \sigma_{g g z}^2 = 3.68 \times 10^{-11} \). However, the resulting signal-to-noise ratio, which is \([91.9, 27.8, 18.9, 14.2, 11.7, 9.9]\) in each bin, tends to be low, since the power spectrum of the signal increases with multiple \( \ell \).

**3.1. Simulation Results**

Here, we present the simulation results that were performed on a PC with 2.8 GHz CPU and 16 GB RAM in python.

**3.1.1. Noise-free Case**

We first evaluate the Algorithm 1 on the idealized noise-free data, which is based on fixed-point iterations. To illustrate its fast convergence, Figure 3 checks the relation between the running time and the objective value. As can be seen, there are
significant oscillations at the beginning stage of iteration, since we randomly generate a positive dense matrix $P$ as an initialization. Shortly after that, the updates become very efficient, yielding a sharp decrease on $\mathcal{J}$, and then the algorithm quickly decreases the objective function almost monotonically. Finally, the objective function converges to $\mathcal{J}^{1/2} \lesssim 5.8 \times 10^{-10}$ by taking about 400 s and 300,000 iterations, which makes an accurate recovery of each element of $P$ at the level of $\sim 10^{-4}$ compared with its true value. Figure 3 therefore clearly demonstrates that, in the absence of shot noise, the fixed-point-based algorithm can efficiently solve the self-calibration problem exactly. Furthermore, it also illustrates that this algorithm can be successfully used in noisy data. Although this algorithm is designed for noise-free data, it can still provide a near global optimum solution with $\mathcal{J}^{1/2} \simeq 10^{-6}$, which may explain why this algorithm is applied for initialization of Algorithm 2 when estimating parameters from realistic measurements.

3.1.2. Realistic Case

We now report performance and efficiency of Algorithm 2 for noisy data. In order to investigate the reconstruction accuracy for the scattering matrix $P$, we have applied this algorithm to 50 realizations of simulated data, each with an independent random noise realization and identical simulated signal power spectra. For each realization, we ran this NMF-based algorithm 500 times with random initializations, and picked the result with the smallest $\mathcal{J}$ as its best estimate. The ensemble-average scattering matrix, $\langle P \rangle$, is obtained by averaging over such best estimates of the scatters from those realizations, and the associated standard deviation $\sigma_{P}$ is obtained from its dispersion. The corresponding absolute bias, $\langle |P - P_{\text{true}}| \rangle$, can be computed easily by comparing with the simulation truth. We aggregate all information about the scatters in Figures 4 and 5 to produce the average value, the absolute bias, and standard deviation.

We observe that the differences between the reconstructed matrix elements $P_{ij}$ and their respective true values are significantly small, with about half of the elements reaching...
an accuracy level of $<1 \times 10^{-4}$. Since the true values of the diagonal elements are close to unity, much larger than off-diagonals, the diagonals have some relatively large biases. Even for the worst case where the bias reaches the maximum value of 0.006, the reconstruction accuracy is still much less than 1%. More interestingly, the scatters for some bins with $P_{\text{true}} < 10^{-3}$ can be reconstructed almost exactly, with extremely small reconstruction errors of $<10^{-5}$–$10^{-6}$. Furthermore, the mean of all the derived biases (25 elements in total) is about 0.001, which is sufficiently small and indicates that the proposed NMF-based algorithm almost provides an unbiased estimate for scatters in the average sense.

In addition, the statistical errors of the reconstructed scatters, $\sigma_P$, are also quite small, reaching the accuracy of 0.003 on average. Since the shot noise fluctuations at diagonals are larger than those of the off-diagonals by a factor 2, as expected by Equation (5) in our fiducial model, the uncertainties of diagonal elements are somewhat larger than those of off-diagonals.

Finally, we conclude that the NMF-based algorithm is able to successfully detect the scattering probabilities with high accuracy at the level of $10^{-3}$, which may meet the accuracy requirement for the fiducial “stage IV” lensing survey.

Moreover, we can use a single convenient number, the density-weighted true redshift for each photo-$z$ bin, to quantify the quality of the reconstruction for photo-$z$ errors (see Zhang et al. 2010 for details). It can be well approximated by

$$\langle z_i \rangle \approx \sum_j P_{ji} \langle z_j^P \rangle = \sum_j P_{ji} \frac{\int n(z) dz}{\int n(z) dz_{P}},$$

(14)

where $\langle z_i \rangle$ is the density-weighted true redshift of the $i$th photo-$z$ bin, $\langle z_j^P \rangle$ stands for the density-weighted average photo-$z$ of the $j$th photo-$z$ bin, and $n(z_{P})$ is defined in Equation (13). Based on the reconstructed scatters $P_{ji}$ for the simulation data, the corresponding averaged bias and the statistical error in $\langle z_i \rangle$ can be then computed accordingly. The results are summarized in Table 1 and illustrated in Figure 6. We can see that the bias value in each photo-$z$ bin is negligibly small, ranging from $2 \times 10^{-4}$ to 0.002, which is of order 0.03%–0.3% of the true redshift. Meanwhile, the associated statistical error appears nearly 3–15 times greater than the bias value in each bin, but still well under the percent accuracy. Both biases and statistical errors slowly vary with the redshifts $z_{P}$.

Note that all the simulation results are based on the fiducial data at the low-$\ell$ range, $20 \leq \ell < 1000$, whereas collecting more information on high multipoles would improve the determination of $P$. If high-$\ell$ measurements of the power spectra are taken into account, there will be more parameters from high-$\ell$ bins to be estimated, while high-$\ell$ data will tend to give a low signal-to-noise ratio. When solving the non-convex optimization problem numerically, those two factors might decrease the reconstruction accuracy in practice and increase the computational time. More detailed tests for the inclusion of high-$\ell$ data will be made in future work.

4. Conclusions

Based on the fixed-point and the NMF methods, we have developed an iteration algorithm to solve the constrained nonlinear optimization problem arising from the studies of the self-calibration of photo-$z$ scatters in weak-lensing surveys. The algorithm was applied to the mock data mimicking a stage IV project, including galaxy density-density correlations and the density-shear cross-correlations that are contaminated by shot noise.

Our proposed algorithms exploits the “fixed-point” iteration and “multiplicative update rules” to efficiently minimize the objective function, leading to globally optimal estimates. The algorithm has the great virtue of being remarkably stable, robust, and fast. The typical run for finding the optimal solution takes about 15 minutes for the noisy data.

The results are very promising. For the noise level in a stage IV survey like LSST, photo-$z$ outlier rates can be determined at an accuracy level of $10^{-3}$ on average, even if we only use the information at $\ell < 1000$. This leads to nearly unbiased estimates of the true mean redshift of each photo-$z$ bin, at the level of $10^{-3}$ for the bins over the redshift range from 0 to 4. Such precision would almost attain the desirable statistical accuracy of the future “stage IV” projects, allowing for precise weak-lensing cosmology.

The high quality of the reconstruction indicates that the algorithm proposed in this paper is quite suitable for weak-lensing cosmology where one wants to accurately determine and understand the catastrophic photo-$z$ error rates, and for understanding the photometric redshift. The algorithm thus far only applies to data at $\ell < 1000$. When necessary, we should also include the information at $\ell > 1000$ to further improve photo-$z$ self-calibration. Since it is computationally challenging, we will leave that investigation to other works.
Of course, much work still has to be performed to better examine the robustness and effectiveness of those algorithms. It will be tested under more realistic simulated observations, while taking into account some possible extra error sources that might slightly degrade the reconstruction accuracy. These include the magnification bias and the size bias in the galaxy number distribution, unphysical correlation induced by observational selection effect, and the intrinsic cross-correlation between adjacent galaxy bins. For example, the magnification bias can also cause spatial correlation between foreground and background galaxies, which may be misinterpreted as photo-z errors.

This problem of magnification bias can be dealt with by utilizing the unique flux dependence of magnification bias on galaxy flux (Zhang et al. 2010). For a given photo-z bin, we further split galaxies into several flux bins. The magnification bias in these flux bins is $\propto g(F)$, which is an observable determined by the observed galaxy flux distribution function. Since this $g(F)$ dependence is in general different from the dependence of galaxy bias on $F$, we can separate the magnification bias from the galaxy intrinsic clustering (Yang et al. 2017). A simpler but less optimal alternative is to weigh galaxies by their flux appropriately. As long as we require $\sum_i w_i g_i = 0$, the weighted overdensity $\delta^W_i$ is free of magnification bias. Here, $g_i$, $\delta^W_i$, and $w_i$ are the $g$ factor, galaxy number overdensity, and the weighting factor of the $i$th flux bin. The algorithm developed in the current paper then directly applies to $\delta^W_i$. Nevertheless, the existence of magnification bias (and size bias) complicates the

Figure 4. Average value, absolute bias, and standard deviation of the reconstructed scattering matrix $P$ obtained by averaging over the estimates from 50 realizations of noisy data based on Algorithm 2. To show the bias and its standard deviation more clearly, the same information is also provided in Figure 5.
photo-z self-calibration. It is an important issue for further investigation. Also, in future works it will be necessary to test it against and apply it to real data.

This work was supported by the National Science Foundation of China (11433001, 11320101002, 11621303, 11403071), the National Basic Research Program of China (2015CB85701), a grant from the Science and Technology Commission of Shanghai Municipality (grant No. 16DZ2260200), the Key Laboratory for Particle Physics, Astrophysics and Cosmology, Ministry of Education, and Shanghai Key Laboratory for Particle Physics and Cosmology (SKLPPC). We would like to thank the anonymous referee for valuable suggestions that helped us to significantly improve this paper.

Appendix

Here, we present a derivation of the expressions for the update rules of $P_L$, $P_R$, and $C_{ℓ,gg}^R$ in Equation (12). Those rules will be constructed by alternately updating each matrix while keeping the other matrices fixed.

A.1. Derivation of the Update Rules for $P_L$, $P_R$

In the following analysis, we will use the alternative notations $V_{ℓ} ≔ C_{ℓ,gg}^{ℓ,gr}$, $W ≔ P_L^T$, and $H_{ℓ} ≔ C_{ℓ,gg}^{ℓ,gr}P_R$ to clearly describe the algorithm without a loss of generality. We seek to update the rules for decreasing the objective function:

$$\mathcal{J} = \frac{1}{2} \sum_{ℓ} \| V_ℓ - WH_ℓ \|^2.$$

(15)

Based on the gradient-descent technique similar to that used in Lee & Seung (2001), in what follows we show the update rule for $W$, while keeping $H_ℓ$ fixed. By Taylor-expanding the objective function $\mathcal{J}$ with respect to the element $w_{ab}$ in $W$ about the point $w_{ab}^t$, we obtain

$$\mathcal{J}(w) = \mathcal{J}(w_{ab}^t) + \mathcal{J}_{w_{ab}}'(w - w_{ab}^t) + \frac{1}{2} \mathcal{J}_{w_{ab}}''(w - w_{ab}^t)^2,$$

(16)

where $\mathcal{J}_{w_{ab}}'$ and $\mathcal{J}_{w_{ab}}''$ are the first and second derivatives with respect to $w_{ab}$, respectively, and the high-order derivatives are zeros. Thus, it is easy to check that

$$\frac{∂\mathcal{J}}{∂w_{ab}} = -\sum_ℓ [(V_ℓ - WH_ℓ)H_ℓ^T]_{lab},$$

(17)

$$\frac{∂^2\mathcal{J}}{∂^2w_{ab}} = \sum_ℓ [H_ℓH_ℓ^T]_{bb},$$

(18)

which, for positive elements, has the upper bound

$$\sum_ℓ [H_ℓH_ℓ^T]_{bb} ≤ \sum_ℓ \frac{\|WH_ℓH_ℓ^T\|_2}{w_{ab}}.$$  

Figure 5. Same as Figure 4, but with a different view for the mean bias (represented as cross symbol) and its standard deviation of each element in the reconstructed scattering matrix $P$.

Figure 6. Bias and associated standard deviation in the true mean redshift ($z$) for the fiducial SNAP-like lensing survey. Both bias and standard deviation are determined from the reconstructed photo-z scatters through Equation (14), based on the 50 realizations of the simulated data.

Table 1

| $z_P$ Range | $\langle z_P \rangle$ | $\langle z^{\text{true}} \rangle$ | $\langle z - z^{\text{true}} \rangle$ | $σ(z)$ |
|------------|----------------|-------------------------------|-------------------|--------|
| [0.0–0.688] | 0.472          | 0.521                         | 0.472             | 0.003  |
| [0.688–1.023] | 0.857          | 0.861                         | 0.270             | 0.003  |
| [1.023–1.389] | 1.199          | 1.232                         | 0.000             | 0.005  |
| [1.389–1.906] | 1.625          | 1.693                         | 0.002             | 0.009  |
| [1.906–4.0] | 2.485          | 2.378                         | 0.000             | 0.002  |
Then, we can define an auxiliary function for the update rule. The auxiliary function regarding $w_{ab}$ is defined as

$$G(w, w_{ab}) = \mathcal{J}(w_{ab}) + \mathcal{J}_{wab}(w_{ab})(w - w_{ab}) + \frac{1}{2} \sum_{\ell} \left[WHH^T \right]_{lab} (w - w_{ab})^2.$$  

To make $\mathcal{J}$ be non-increasing under the update, we construct

$$\mathcal{J}(w_{ab}^{(+\ell)}) \leq G(w_{ab}^{(-\ell)}, w_{ab}^{(+\ell)}) = \mathcal{J}(w_{ab}^{(+\ell)}),$$

where the first inequality comes from the upper bound and the second is the result of minimization. Setting

$$\frac{\partial G(w, w_{ab})}{\partial w} = 0,$$

we finally obtain the update rule for $w_{ab}$:

$$w_{ab} = w_{ab} \sum_{\ell} [V_HH^T]_{lab} \sum_{\ell} \{WHH^T\}_{lab},$$

which is consistent with the “multiplicative update rules” for the standard NMF, when $n_t = 1$.

Next, this update rule can be easily extended to accommodate the constraint of the row-wise unitary sum for $W$ (recall $W \equiv P_t^T$ and the column-wise unitary sum for $P_L$). Introducing a set of Lagrangian multipliers, $\lambda_i$ with $i = 1, \ldots, n$, we can rewrite the objective function as

$$\bar{\mathcal{J}} = \mathcal{J} + \sum_{i} \lambda_i (1 - \sum_{j} W_{ij}).$$

Similar to the proof by Zhu et al. (2013), we find that the objective function is non-increasing using the following multiplicative update rule:

$$W_{ij} \leftarrow \frac{\nabla_{W_{ij}} A_{ij} + 1 - B_{ij}}{\nabla_{W_{ij}} A_{ij}}$$

with $\nabla_{ij} = \sum_{\ell} [V_HH^T]_{ij}, \nabla_{ij}^+ = \sum_{\ell} \{WHH^T\}_{ij}, A_{ij} = \sum_{\ell} W_{ib}/\nabla^+_{ib}, B_{ij} = \sum_{\ell} W_{ib}/\nabla^+_{ib}$. Note that there is a negative term $-B_{ij}$ in the numerator, which may cause negative entries in $W$ during updates. Hence, we apply the “moving term” trick (Yang & Oja 2010) to overcome this problem if any negative entries in $W$ appear, which gives $W_{ij} \leftarrow \frac{\nabla_{W_{ij}} A_{ij} + 1}{\nabla_{W_{ij}} A_{ij} + B_{ij}}$.

Once the update for $W$ (i.e., $P_t^T$) is obtained, we update $P_R$ by imposing $P_L = W^T$ such that the constraint $P_L = P_R$ is implemented during iterations.

The simulation tests show that the use of the following convergence criterion is appropriate: the change of each element in $W$ per iteration satisfies $|W_{ij}^{(\ell+1)} - W_{ij}^{(\ell)}| < 10^{-8}$ or the maximum number of iterations $n_{\text{iter}} = 3 \times 10^5$ is reached.

### A.2. Derivation of the Update Rules for $C_{\ell}^{gg,R}$

Let us now consider the update rule for $C_{\ell}^{gg,R}$ with $W$ fixed. Based on the fact that $C_{\ell}^{gg,R}$ for all $\ell$ are diagonal matrices,

$$\mathcal{J} = \sum_{\ell} \| V_{\ell} - WC_{\ell}^{gg,R} W^T \|^2_F = \sum_{\ell} \| \text{Vec}[V_{\ell}] - Uc_{\ell} \|^2_F,$$

where the “vectorization” operator (vec[-]) is used for converting the matrix into a column vector by stacking the columns into a long column vector, and $c_{\ell}$ collects all of the diagonal elements of $C_{\ell}^{gg,R}$, i.e., $c_{\ell} = [C_{\ell}^{gg,R}]_{ii}$ for all $i$. Here, $U$ is the $n^2 \times n$ matrix, consisting of the Kronecker products, denoted by $\otimes$, on the columns of $W$ as its columns, i.e., $U = [w_1 \otimes w_2 \ldots w_n \otimes w_1]$, where $w_i$ is the $i$th column of $W$. According to a well-known solution for this linear least-squares problem, the optimal $c_{\ell}$ can be found as

$$c_{\ell} = (U^T U)^{-1} U^T \text{Vec}[V_{\ell}], \quad \text{for all } \ell,$$

which determines the optimal $C_{\ell}^{gg,R}$, and in other words inequivalent to its update rule. To ensure the non-negativity, an additional step is added after each such update to project all negative elements of $C_{\ell}^{gg,R}$ to their absolute values. Note that it is not necessary to use the NMF technique for updating $C_{\ell}^{gg,R}$, since the objective function is convex in $C_{\ell}^{gg,R}$ and this update rule is stable and efficient for decreasing an objective function monotonically.

### ORCID iDs

Le Zhang https://orcid.org/0000-0002-1182-8888
Yu Yu https://orcid.org/0000-0002-9359-7170

### References

Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:0609591
Benjamin, J., van Waerbeke, L., Ménard, B., & Kilbinger, M. 2010, MNRAS, 408, 1168
Bernstein, G., & Huterer, D. 2010, MNRAS, 401, 1399
Choi, A., Heymans, C.,Blake, C., et al. 2016, MNRAS, 463, 3737
de Putter, R., Doré, O., & Das, S. 2014, ApJ, 780, 185
Erben, T., Hildebrandt, H., Lerchster, M., et al. 2009, A&A, 493, 1197
Fu, L., Kilbinger, M., Erben, T., et al. 2014, MNRAS, 441, 2725
Heymans, C., Grocutt, E., Heavens, A., et al. 2013, MNRAS, 432, 2433
Heymans, C., Van Waerbeke, L., Bacon, D., et al. 2006, MNRAS, 368, 1323
Hirata, C. M., Mandelbaum, R., Ishak, M., et al. 2007, MNRAS, 381, 1197
Hirata, C. M., Mandelbaum, R., Seljak, U., et al. 2004, MNRAS, 353, 529
Huff, E. M., Eifler, T., Hirata, C. M., et al. 2014, MNRAS, 440, 1322
Huterer, D., Kirkby, D., Bean, R., et al. 2015, ApJ, 63, 23
Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, MNRAS, 366, 101
Jain, B., & Zhang, P. 2008, PhilD, 78, 063503
Johnson, A., Blake, C., Amon, A., et al. 2017, MNRAS, 465, 4118
Jouvel, S., Kneib, J.-P., Ilbert, O., et al. 2009, A&A, 504, 359
Kilbinger, M., Fu, L., Heymans, C., et al. 2013, MNRAS, 430, 2200
Kitching, T. D., Taylor, A. N., & Heavens, A. F. 2008, MNRAS, 389, 173
Lee, D. D., & Seung, H. S. 2001, NIPS (Cambridge, MA: MIT Press) 556–62
Lesgourgues, J. 2011, arXiv:1104.2932
Ma, Z., Hu, W., & Huterer, D. 2006, ApJ, 636, 21
Mandelbaum, R., Hirata, C. M., Ishak, M., Seljak, U., & Brinkmann, J. 2006, MNRAS, 367, 611
Mandelbaum, R., Rowe, B., Armstrong, R., et al. 2015, MNRAS, 450, 2963
Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544
Massey, R., Heymans, C., Berge, J., et al. 2007, MNRAS, 376, 15
Matthews, D. J., & Newman, J. A. 2010, ApJ, 721, 456
Matthews, D. J., & Newman, J. A. 2012, ApJ, 745, 180
McQuinn, M., & White, M. 2013, MNRAS, 433, 2857
Ménard, B., Scramon, R., Schmidt, S., et al. 2013, arXiv:1303.4722
Mitchell-Wynne, K., Cooray, A., Gong, Y., et al. 2012, ApJ, 753, 23
Newman, J. A. 2008, ApJ, 684, 88
Okumura, T., & Jng, Y. P. 2009, ApJL, 694, L83
