Global Mapping of the Surface Composition on an Exo-Earth Using Color Variability

Hajime Kawahara1,2

1 Department of Earth and Planetary Science, The University of Tokyo, 7-3-1, Hongo, Tokyo, Japan; kawahara@eps.s.u-tokyo.ac.jp
2 Research Center for the Early Universe, School of Science, The University of Tokyo, Tokyo 113-0033, Japan

Received 2020 March 1; revised 2020 April 3; accepted 2020 April 6; published 2020 May 6

Abstract

Photometric variation of a directly imaged planet contains information on both the geography and spectra of the planetary surface. We propose a novel technique that disentangles the spatial and spectral information from the multiband reflected light curve. This will enable us to compose a two-dimensional map of the surface composition of a planet with no prior assumption on the individual spectra, except for the number of independent surface components. We solve the unified inverse problem of the spin–orbit tomography and spectral unmixing by generalizing the nonnegative matrix factorization using a simplex volume minimization method. We evaluated our method on a toy cloudless Earth and observed that the new method could accurately retrieve the geography and unmixed spectral components. Furthermore, our method is also applied to the real-color variability of the Earth as observed by Deep Space Climate Observatory. The retrieved map explicitly depicts the actual geography of the Earth, and unmixed spectra capture features of the ocean, continents, and clouds. It should be noted that the two unmixed spectra consisting of the reproduced continents resemble those of soil and vegetation.

Unified Astronomy Thesaurus concepts: Exoplanet surface characteristics (496); Exoplanet surfaces (2118); Habitable planets (695); Direct imaging (387); Astronomical methods (1043); Exoplanet surface composition (222); Exoplanet surface variability (2023)

1. Introduction

Direct imaging of terrestrial planets around a nearby solar-type star is an important goal for future astronomy. In the 2020 decadal surveys, both HabEx and LUVOIR have shown the capability to search for these planets, even in the habitable zone. Direct imaging with spectroscopy provides information regarding the molecules in the atmosphere of the planet, which enables us to search for biosignatures such as oxygen, carbon dioxide, and water. Moreover, surface inhomogeneity can be explored with photometric monitoring of the reflected light as proposed by Ford et al. (2001). The color variability of the reflected light has been studied as a probe of surface composition (e.g., Cowan et al. 2009; Fuji et al. 2011). The spatial distribution of the planet surface can also be inferred from the photometric variation. Diurnal variation, due to the rotation of the planet, provides the spin rotation period and a one-dimensional distribution of the surface (Pallé et al. 2008; Cowan et al. 2009; Oakley & Cash 2009; Fuji et al. 2010, 2011; Lustig-Yaeger et al. 2018). Furthermore, the axial tilt can be obtained from the analysis of the frequency modulation of the periodicity (Kawahara 2016; Nakagawa et al. 2020). The analytic expression of the reflected light curves has been studied (Cowan et al. 2013; Haggard & Cowan 2018).

A full two-dimensional inversion technique called “spin–orbit tomography” (analogous to computer tomography) was proposed by Kawahara & Fuji (2010) and has been studied in terms of the inverse problem (Kawahara & Fuji 2011; Fuji & Kawahara 2012; Farr et al. 2018; Berdyugina & Kuhn 2019; Aizawa et al. 2020) and obliquity measurement (Schwartz et al. 2016; Farr et al. 2018). Recently, Luger et al. (2019) analyzed a single-band light curve of the Earth with Transiting Exoplanet Survey Satellite data and inferred a rough two-dimensional cloud distribution. Fan et al. (2019) successfully retrieved a global map that was analogous to the distribution of a continent, from data that was obtained by DSCOVR by monitoring the Earth for two years (Jiang et al. 2018). They used the second principle component (PC2) of a multicolor light curve. Aizawa et al. (2020) improved the retrieved map from DSCOVR using sparse modeling. These examples showed that a global map could be retrieved from a time series of a single band or PC. However, there exists a level of ambiguity when interpreting the derived maps when we do not have prior knowledge on the surface compositions.

Moreover, a blind retrieval of the reflectance spectra of the surface components from the integrated light is known as “spectral unmixing” in remote sensing. Cowan & Strait (2013) formulated the spectral unmixing as a disentanglement of geography by spin rotation. However, the longitudinal map inferred from the EPOXI data did not match the real geographies because of the degeneracy of the inferred geometric distribution and spectral components (Fuji et al. 2017). The ambiguity of spectral unmixing originates from the matrix factorization not being unique, which has been extensively studied in the field of remote sensing. These studies found that additional constraints such as the simplex volume minimization of spectral components guarantee a unique solution to the unmixed spectra (Craig 1994; Fu et al. 2015, 2019; Lin et al. 2015; Ang & Gillis 2019). In practice, nonnegative matrix factorization (NMF) with regularization terms easily retrieves the surface components in hyperspectral unmixing (Ang & Gillis 2019). These techniques in remote sensing are worth considering in their application to multicolor light curves of directly imaged exoplanets.

This paper aims to formulate a single inverse problem that unifies the spin–orbit tomography and spectral unmixing using a novel technique via remote sensing. To achieve this, we unify the NMF-based spectral unmixing technique and spin–orbit tomography to retrieve both the spectra and geographies of a disk-integrated light curve from an exoplanet. We demonstrate its capabilities using the simulated data and real data from Deep
Space Climate Observatory (DSCOVR). The rest of the paper is organized as follows. In Section 2, we first review the spin–orbit tomography and spectral unmixing. Next, we construct a unified retrieval model using NMF; the optimization scheme is also provided. In Section 3, we test the technique by applying it to a cloudless toy model. In Section 4, we demonstrate this new technique by applying it to real observational data of the Earth recorded by DSCOVR. Finally, in Section 5, we summarize our results.

2. Formulation of Spin–Orbit Tomography with Spectral Unmixing

2.1. Spin–Orbit Tomography

Space direct imaging in optical and near-infrared bands aims to detect the reflected light (or scattered light) from a planet. The reflected light is a summation of photons from the illuminated and visible side of a planet. This integrated-reflected light is expressed as

\[ f_p = \frac{f_s R_p^2}{\pi a^2} \int_{IV} d\Omega s R' (\theta_0, \vartheta_1, \varphi) \cos \vartheta_0 \cos \vartheta_1, \]  

where \( f_s \) is the stellar flux, \( R_p \) is the radius of the planet, \( a \) is the star–planet distance, \( IV \) is the illuminated and visible area, and \( \Omega_s \) is the solid angle of the planet’s sphere. \( R' (\theta_0, \vartheta_1, \varphi) \) represents the bidirectional reflectance distribution function (BRDF) of the surface element. \( \theta_0 \) is the solar zenith angle, \( \vartheta_1 \) is the zenith angle between the direction toward an observer and the normal vector of the surface, and \( \varphi \) is the relative azimuth angle between the line of sight and stellar direction. The derivation of Equation (1) is given in Appendix A. An isotropic approximation of the surface reflectance (the Lambert approximation), \( R' (\theta_0, \vartheta_1, \varphi) = R' \), significantly reduces the complexity of the problem. We also define the spherical coordinate fixed on the surface by \( (\theta, \phi) \) and express the surface component, \( s \), in spherical coordinates \( R = m(\theta, \phi) \) as the time-dependent quantity (static surface approximation). Then, we obtain

\[ f_p = \frac{f_s R_p^2}{\pi a^2} \int_{IV} d\Omega m(\theta, \phi) \cos \vartheta_0 \cos \vartheta_1. \]  

We note that the IV area, \( \cos \vartheta_0 \), and \( \cos \vartheta_1 \) are time-dependent. The terms of \( \cos \vartheta_0 \) and \( \cos \vartheta_1 \) also depend on the position of the planet surface, \( (\theta, \phi) \), and the axial tilt parameters, \( g = (\zeta, \Theta_{eq}) \), where \( \zeta \) is the planet obliquity and \( \Theta_{eq} \) is the orbital phase at the equinox. We define the geometric kernel introduced by Kawahara & Fujii (2010) as

\[ W_g (t, \theta, \phi) = \begin{cases} \frac{f_s R_p^2}{\pi a^2} \cos \vartheta_0 \cos \vartheta_1 & \text{for } \cos \vartheta_0, \cos \vartheta_1 > 0 \\ 0 & \text{otherwise} \end{cases} \]  

where the positive condition of \( \cos \vartheta_0 \) and \( \cos \vartheta_1 \) restricts the surface integral to pixels on the IV area. Assuming that \( g \) is fixed, we obtain the Fredholm integral equation of the first kind

\[ f_p (t) = \int d\Omega W_g (t, \theta, \phi) m(\theta, \phi). \]  

Discretization of the time \( t \rightarrow t_i \) and planetary surface \( (\theta, \phi) \rightarrow (\theta_i, \phi_j) \) reduces Equation (4) to the linear inverse problem

\[ d_i = \sum_j W_{ij} m_j, \]  

or, using the vector form, we can express it as

\[ d = W m. \]  

where \( d_i = f_g (t_i) \) for \( i = 0, 1, \ldots, N_i - 1 \) and \( m_j = m (\theta_i, \phi_j) \) for \( j = 0, 1, \ldots, N_j - 1 \). The explicit expression of the geometric kernel \( W_{ij} = W_g (t_i, \theta_i, \phi_j) \) in the spherical coordinate is given in Appendix A.

Because the inverse problem of Equation (6) is ill-posed, an additional constraint or regularization is needed to solve the problem. Various types of regularizations have been attempted so far. Kawahara & Fujii (2010) used nonnegative regularization and the requirement of an upper limit of albedo as regularization using the bounded variable least-squares solver (Lawson & Hanson 1995). Kawahara & Fujii (2011) used the Tikhonov regularization, which minimizes the cost function

\[ \min Q = \frac{1}{2} ||d - W m||_2^2 + \frac{\lambda_2}{2} ||m||_2^2, \]  

where \( \lambda_4 \) is the spatial regularization parameter, and \( || \cdot ||_2^2 \) is the squared L2 norm. To construct the model on the Bayesian framework, Farr et al. (2018) used a Gaussian process to regularize the map while Berdyugina & Kuhn (2019) used an Occamian approach algorithm. Recently, Aizawa et al. (2020) demonstrated that the L1 + total square variation (TSV) provided better results than a simple L2 (Tikhonov) regularization.

The value of \( d \) depends on which features we want to extract from the multicolor light curve. Kawahara & Fujii (2011) used a single-band light curve to retrieve a cloud map of the simulated Earth. They also demonstrated that a rough two-dimensional distribution of the continent or ocean can be retrieved from a color difference between 0.85 \( \mu \)m and 0.45 \( \mu \)m or 0.85 \( \mu \)m and 0.65 \( \mu \)m, owing to the near flatness of the cloud spectrum. Cowan et al. (2009) utilized principle component analysis (PCA) for their longitudinal mapping of EPOXI data. Fan et al. (2019) used the second component of PCA of the multicolor light curve of DSCOVR. Comparing with the ground truth, they found that the resultant map was similar to the global continent/ocean map of the Earth. However, these two examples required prior knowledge of the surface composition or the ground truth of the geography. The ambiguity in the interpretation of the map is a limitation of the spin–orbit tomography.

2.2. Spectral Unmixing

Spectral unmixing is a procedure that disentangles mixed spectra by finding the end members. The mixing model of the spectra of multiple surface compositions is required to unmix the spectra. The simplest model is the linear mixing model, expressed as

\[ d(t_i, \lambda_i) = D = \sum_k A_k X_{\lambda_i}, \]  

or simply

\[ D = AX, \]  

where \( A_k = a_k (t_i) \) is the contribution of the \( k \)-th component at time \( t = t_i \) to the intensity of light, and \( X_{\lambda_i} = x_k (\lambda_i) \) for

Kawahara

\[ \]
The Astrophysical Journal, 894:58 (14pp), 2020 May 1

Kawahara

\[ l = 0, 1, \ldots, N_l - 1 \] is the reflection spectra of the \( k \)-th component at wavelength \( \lambda_l \).\(^3\) We need to solve the matrix factorization of \( A \) and \( X \). Generally, the matrix factorization is formulated as the minimization of the cost function, where the cost function can either be the squared Euclidean distance or the Kullback–Leibler distance. In this paper, we use the squared Euclidean distance

\[ Q = \frac{1}{2} ||D - AX||_F^2 + R(A, X) \] (10)

where \( || \cdot ||_F^2 \) is the squared Frobenius norm defined by

\[ ||Y||_F^2 = \sum_{i,j} Y_{ij}^2, \] (11)

and \( R(A, X) \) is the regularization term.

2.2.1. Principle Component Analysis

PCA is a traditional technique used to disentangle the spectral components of multicolor light curves, as observed in Cowan et al. (2009). It was also used in a global map reconstruction of the Earth by Fan et al. (2019) and Aizawa et al. (2020). PCA can also be formulated as a minimization of the cost function, from the perspective of optimization,

\[ \text{minimize } Q = \frac{1}{2} ||D - AX||_F^2 \] (12)

subject to \( A^T A = \text{diag}(\sigma_A) = \Sigma_A, \)

\[ X^T X = \text{diag}(\sigma_X) = \Sigma_X, \] (14)

where \( \text{diag}(\sigma) \) is a diagonal matrix whose elements are \( \sigma_i \). The drawback of the PCA as a matrix factorization method is the strong assumption of orthogonality for \( A \) and \( X \). However, its orthogonality is useful for visualizing the simplex by reducing its dimensionality Cowan & Strait (2013). In this paper, we denote the orthogonal PCA basis by \( U_X = (\Sigma_X^{-1/2}X)^T \), i.e.,

\[ U_X^T U_X = I \] (\( I \) is an identity matrix). An arbitrary matrix \( M \) can be decomposed by row vectors of \( U_X \) as

\[ M = \sum_k p_k u_k, \] (15)

where \( u_k \) is the \( k \)-th row of \( U_X \). The projection of \( M \) onto \( PC_k \) is computed by

\[ p_k = M u_k. \] (16)

2.2.2. Nonnegative Matrix Factorization

In the field of remote sensing, a wide variety of spectral unmixing has been studied. Among these techniques, NMF decomposes a single matrix \( D \) to two matrices \( A \) and \( X \) whose elements are nonnegative; that is, \( D = AX \) (Paatero & Tapper 1994; Lee & Seung 1999). NMF can be defined by the minimization of the cost function. For instance, using the

\[ \text{minimize } Q = \frac{1}{2} ||D - AX||_F^2 + R(A, X) \] (17)

subject to \( A_{ik} \geq 0, X_{il} \geq 0. \) (18)

NMF is known to be NP-hard (Vavasis 2009); therefore, the optimization of NMF is difficult to achieve. Nevertheless, various efficient optimization methods have been proposed (Lee & Seung 1999; Cichocki et al. 2009, references therein).

In particular, NMF combined with the simplex volume minimization technique can accurately reproduce the high-resolution spectrum components from remote-sensing satellite data (Craig 1994; Fu et al. 2015, 2019; Lin et al. 2015; Ang & Gillis 2019). The concept of the simplex volume minimization can be summarized as follows: if the data are sufficiently spread in the convex hull defined by the end members, the data-enclosing simplex whose volume is minimized identifies the true end members.

In Figure 1, we plot three simplexes that enclose all of the data points. Each simplex provides its vertices as a solution of NMF. The simplex volume minimization choose the vertices of the volume-minimum simplex (dashed triangle) as the end members of NMF. When there is at least one pure pixel of each end member in the data, the volume-minimum simplex obviously identifies the true end members. Even in the case without pure pixels, Lin et al. (2015) showed that the true end members could be identified by the volume-minimum simplex under the condition of the pixel purity level that applies uniformly to all of the end members. In Figure 1, if the data points on the red dashed lines have high purity levels, that is, they are on the boundaries of the simplex defined by the true end members, then the volume-minimum simplex identifies the true end members.

![Figure 1. Schematic picture of the simplex volume minimization, which is based on Figure 1 in Lin et al. (2015; see also Fujii et al. 2017). The black dots represent the observed data, and the three triangles indicate a simplex that encloses all of the data points. The dashed triangle is the simplex whose volume is minimized. The end members are defined by three vertices of the dashed triangle.](image-url)
As the regularization term for the simplex volume minimization, the Gram determinant of spectral components (VRDet)

\[
R(A, X) = \frac{\lambda_X}{2} \det(XX^T)
\]

was used (e.g., Schachtner et al. 2009; Zhou et al. 2011; Xiang et al. 2015; Ang & Gillis 2019; Fu et al. 2019), where \(\lambda_X\) is the spectral regularization parameter. The Gram determinant, Equation (19), is a surrogate of the volume of a convex hull of spectral vectors, \((x_0, x_1, \ldots, x_{Nt-1})\), where \(x_k = \{x_k(\lambda_i)\}\) for \(i = 0, 1, \ldots, N_t - 1\); such a convex hull is identifiable for its well-spread data (Lin et al. 2015). The minimization of the convex hull of spectral vectors can be achieved by minimizing Equation (19).

2.3. Unified Retrieval Method of Mapping and Spectra

Our task is to unify the spectral unmixing of Equation (9) and spin–orbit tomography of Equation (6). To achieve this, we assume a pixel-wise spectral unmixing

\[
m(\theta_j, \phi_j, \lambda_l) = m_{jl} = \sum_k A_{jk} X_{kl},
\]

where \(X_{kl}\) is the reflectivity of the \(k\)-th component at wavelength \(\lambda_l\), and \(A_{jk}\) is the surface distribution of the \(k\)-th surface component at the \(j\)-th pixel instead of time in Equation (9). Combining Equation (21) with the multicolor version of Equation (6), we obtain

\[
D_{jl} = \sum_j W_{lj} m_{jl} = \sum_j W_{lj} A_{jk} X_{kl}
\]

or simply,

\[
D = WAX.
\]

Equation (23) provides the general form of the spin–orbit tomography with spectral unmixing.

The two-dimensional mapping thus far estimated the spectra using PCA or a color difference prior to retrieving the geographic distribution. This “unmixing first” strategy could not feed back information on the fitting accuracy of the geographic retrieval to spectral unmixing. The improvement of Equation (23) over the spin–orbit tomography is that we fit both the spectral components and geography to data in a consistent manner.

Cowan & Strait (2013) solved an equation similar to Equation (23) for \(A\) and \(X\) using the multicolor light curve provided by the EPOXI satellite as \(D\). They retrieved a longitudinal map of the surface components from the diurnal rotation of the light curve. This procedure is referred to as “rotational unmixing.” They used the Markov Chain Monte Carlo method to determine the best parameters of \(X\) and \(D\). Their optimization corresponds to the minimization of

\[
Q = \frac{1}{2} ||D - WAX||_F^2 + R(A, X)
\]

subject to \(1 \geq \sum_k A_{jk} \geq 0, 1 \geq X_{kl} \geq 0\),

where \(W\) is the latitudinal average of the kernel. The minimization of Equation (24) under the constraint of Equation (25) is formally identical to the weighted NMF (we explain this in Section 2.4) with no regularization + the upper limits of \(A\) and \(X\).

In general, matrix factorization has a degeneracy of solutions. The transformation of \(A \leftarrow AG^{-1}\) and \(X \leftarrow GX\) for a regular matrix \(G\) under the constraint of Equation (25) does not change the value of the cost function, Equation (24). This means that if we change the spectral basis by a rotation of \(G\), then the inferred map should change too. This degeneracy is known in the field of blind signal separation (see, e.g., Cichocki et al. 2009). The nonuniqueness of the blind signal separation might explain the mismatch between the inferred longitudinal map by Cowan & Strait (2013) and the actual geography suggested by Fujii et al. (2017).

The nonuniqueness feature of NMF can be avoided by adding regularization when neglecting unavoidable scaling and permutation ambiguities (Cichocki et al. 2009; Pajic et al. 2015). Our task is to find a unique (identifiable) solution to the spectral unmixing and spin–orbit tomography. To achieve this, we consider the cost function with regularization for both \(X\) and \(A\). In this paper, we used the squared Euclidean distance as the cost function for Equation (23),

\[
Q = \frac{1}{2} ||D - WAX||_F^2 + R(A, X)
\]

where \(R(A, X)\) is the regularization term. Similar to the rotation unmixing for longitudinal mapping (Cowan & Strait 2013), we call the two-dimensional mapping+unmixing+regularization of Equation (26) “spin–orbit unmixing” in this paper.

2.4. Weighted Nonnegative Matrix Factorization

The nonnegative condition of Equation (26) yields an NMF version of the unified retrieval model

\[
\text{minimize } Q = \frac{1}{2} ||D - WAX||_F^2 + R(A, X)
\]

subject to \(A_{jk} \geq 0, X_{kl} \geq 0\).

As this formulation differs from a standard NMF, Equation (17), for a weight \(W\), we require an extension for the optimization of a standard NMF to the weighted NMF.

2.4.1. Regularization

The regularization term suppresses the instability of the retrieved map due to overfitting, otherwise known as over-training in machine learning. A Tikhonov regularization (or an L2 regularization) used in the original spin–orbit tomography (Kawahara & Fujii 2011) can be extended to the regularization term using the Frobenius norm, \(R(A, X) \propto ||A||_F^2\). Hence, a

\[
Q = \frac{1}{2} ||D - WAX||_F^2 + R(A, X)
\]

5 Besides the additional constraints on \(X\) and \(A\), the difference between the equation in Cowan & Strait (2013) and Equation (23) is \(W\). As rotational unmixing performs the longitudinal mapping according to spin rotation, the geometric kernel should be integrated unto the latitudinal direction, \(W = W(\theta, \phi)\). In the frame of the spin–orbit tomography, we need to use a two-dimensional discretization of a sphere, \(W(\theta, \phi)\).
simple extension of the spin–orbit tomography is expressed as

\[
R(A, X) = \frac{\lambda_A}{2} ||A||_F^2 + \frac{\lambda_X}{2} \det(XX^T) \quad (L2 - VRDet),
\]

where \(\lambda_A\) and \(\lambda_X\) are the regularization parameters for \(A\) and \(X\), respectively. We call this model “L2-VRDet”.

### 2.4.2. Optimization

The minimization of the cost function is performed by a block coordinate descent, which consists of two separate optimizations for \(A\) and \(X\) (see Kim et al. 2014, as a review paper). These optimizations are solved by minimizing the quadratic forms

\[
q_A = \frac{1}{2} a_k^T \mathcal{W}_k a_k - b_k^T a_k
\]

\[
q_X = \frac{1}{2} x_k^T \mathcal{W}_k x_k - b_k^T x_k.
\]

df of \(a_k\) (the \(k\)-th column vector of \(A\)) and \(x_k\) (the \(k\)-th row vector of \(X\)) for \(k = 0\) to \(N_k - 1\), iteratively (Zhou et al. 2011), where \(\mathcal{W}_k\), \(\mathcal{W}_X\), \(b_A\), and \(b_X\) are placeholders that depend on the cost function.

As the L2 regularization of \(A\), we minimize

\[
q_A = \frac{1}{2} a_k^T (\mathcal{L}_A + \mathcal{T}_A) a_k - b_k^T a_k \quad \text{(L2)}
\]

\[
\mathcal{L}_A = x_k^T x_k W^T W
\]

\[
\mathcal{T}_A = \lambda_A I
\]

where \(\Delta_{il} = D_{il} - \sum_{a=1}^{N_A} W_{il} A_{ia} \sum_{j=1}^{N_k} W_{lj} x_{ja}\), \(I_k \in \mathbb{R}^{N_k \times N_k}\) is an identity matrix. The \(X\) component of the quadratic problem for the VRDet model is calculated by

\[
q_X = \frac{1}{2} x_k^T (\mathcal{L}_X + \mathcal{D}_X) x_k - b_k^T x_k \quad \text{(VRDet)},
\]

\[
\mathcal{L}_X = ||W a_k||_F^2 I_k
\]

\[
\mathcal{D}_X = \lambda_X \det(\bar{X}_k \bar{X}_k^T [I_k - \bar{X}_k^T (\bar{X}_k \bar{X}_k^T)^{-1} \bar{X}_k]),
\]

where \(\bar{X}_k\) is a submatrix of \(X\) when the \(k\)-th row of \(X\) is removed, and \(I_k \in \mathbb{R}^{N_k \times N_k}\) is the identity matrix. The derivations of these terms are given in Appendix B.1.1.

Following Ang & Gillis (2019), we use the accelerated projected gradient descent + restart (APG+restart) to optimize the quadratic problems with nonnegative conditions. The APG+restart is based on a projected gradient descent onto a positive orthant with Nesterov’s acceleration and the restarting method. The algorithm is summarized as follows.

\[\text{Algorithm 1.}\]

### 3. Testing the Spin–Orbit Unmixing Using a Cloudless Toy Model

We test the spin–orbit unmixing by using the volume-regularized NMF and applying it to a toy model. The toy model assumes three surface types on a planet: ocean, land, and vegetation. The reflection spectra for land and vegetation were taken from the ASTER spectral library (Baldridge et al. 2009), and the ocean albedo is from McLinden et al. (1997), as indicated by the gray lines in Figure 2. The input classification of the map is based on the moderate resolution imaging spectroradiometer classification map in 2008, as shown in the left panel in Figure 3. We use the geometric settings of Fuji & Kawahara (2012), an orbital inclination of 45°, an obliquity of 23.4°, and \(\theta_{eq} = 90°\). We assume that the spin rotation period is a sidereal day of Earth, 23.9344699/24.0 days, and we assume an orbital revolution period \(P_{orb}\) of 365 d. We took \(N_i = 512\) homogeneous samples over a year and injected a 1% Gaussian noise into the light curve.

---

6. https://github.com/HajimeKawahara/sot
For the retrieval, we use a HEALPix map (Górski et al. 2005) and $A_{k}$ with $j = 1, 2, ..., N_{\text{pix}} = 3072$ pixels. In this test, we assume that we know the number of spectral components, $N_{k} = 3$. Furthermore, we assume that we know the axial tilt parameters $g$ and set $10^{\text{5}}$ as the number of iterations for the optimization.

Figures 2 and 4 are examples of unmixed spectra and retrieved maps for the L2-VRDet model ($\lambda_{A} = 10^{-1}$ and $\lambda_{X} = 10^{2}$). Because the normalization of each component is arbitrary, we adjust the normalization of each component to the input spectra. In this case, the input spectra and geography are accurately reproduced by the unmixed spectra and their retrieved distributions of components 0, 1, and 2, which correspond to vegetation, land, and water, respectively. These results indicate that the spin–orbit unmixing using the volume-regularized NMF can infer the spectral components and their geography simultaneously.

The sparsity of the retrieved maps is a notable feature of the spin–orbit unmixing that utilizes NMF. Because of the nonnegative constraint, large parts of the maps remain zero. This feature was observed in the spin–orbit tomography using BVLS in Kawahara & Fujii (2010). In contrast, the spin–orbit tomography that uses the Tikhonov regularization does not exhibit such sparsity (Kawahara & Fujii 2011).

We made a color composite map from the three maps in Figure 4, as shown in the right panel of Figure 3. These results show that the L2-VRDet model with an appropriate regularization can infer a global composition map for the toy model.

We remind the reader that the spectra are well mixed even for the cloudless toy model. To illustrate how the spectra are unmixed, we project the input light curves and unmixed spectra as end members onto the PC1–PC2 plane in Figure 5. The PCA is computed using the input light curve. To draw this plot, we first compute the normalized light curve via the mixing matrix

$$\tilde{A} = WA;$$

that is,

$$D = \tilde{A} X.$$

The light curve is normalized as $\tilde{D}_{l} = D_{l}/\sum_{k} \tilde{A}_{k}$. We then derive PC1 and PC2 using $\tilde{D}_{l}$ and $X$ onto the PC1–PC2 plane, indicated by the orange crosses and red points, was computed using Equation (16). The light curve does not touch the boundary of the triangle, which is defined by end members; that is, the triangle is not a convex hull of the light curve. This is because the spectra of the light curve are well mixed, and the purity is low. In this case, the geometric disentanglement is essential for the spectra unmixing because the end members are far from the trajectory of the light curves.
The “disentangled spectra” of the light curve are defined by $\tilde{X} = AX$; that is,

$$D = W\tilde{X}. \quad (42)$$

The blue dots are the projection of the disentangled spectra normalized by $\sum A_{jk}$ onto the PC1–PC2 plane. The disentangled spectra are well spread in the triangle, and therefore, the triangle defines a convex hull of the disentangled spectra. The effect of the geometric disentanglement of the spectral unmixing is visualized as the expansion from orange crosses to blue dots in Figure 5.

### 3.1. Dependence on Regularization Parameters

The over-regularization of the simplex volume (i.e., large $\lambda_X$) induces a worse fit of the data. This is confirmed by the mean residual of fitting the model to the data.

$$\text{mean residual} = \frac{1}{D} \sqrt{\frac{||D - W\tilde{X}||_F^2}{N_lN_i}}, \quad (43)$$

where $D$ is the mean value of the data. The top panel of Figure 6 presents the mean residual as a function of $\lambda_X$. The mean residual gradually increases as the spectral regularization parameter increases. $\lambda_X = 10^{-2}$ and $10^{-1}$ have similar mean residuals, which indicates that the model fits the data well for small regularization parameters. A smaller spectral regularization provides fewer constraints on the simplex volume minimization. Hence, there is a trade-off relation between the models goodness of fit and volume minimization.

The second panel from the top illustrates a surrogate of the spectral of volume of normalized spectral components $\det(\hat{X}^T\hat{X})$, where $\hat{X}_{jk} = X_{jk}/\sum X_{jk}$. This quantity decreases as the spectral regularization parameter increases, which indicates that the spectral volume is minimized more as $\lambda_X$ increases. The surrogate of the normalized spectral volume roughly converges at $\lambda_X = 10^4$.

A direct comparison with the ground truth is useful to see how $\lambda_X$ affects the estimate of the spectral components and geography. To quantify the difference between the unmixed spectra and ground truth, we define the mean removed spectral angle (MRSA) between the two vectors $x$ and $y$ as

$$\text{MRSA}(x, y) = \frac{1}{\pi} \cos^{-1}\left(\frac{(x - \bar{x})^T(y - \bar{y})}{||x - \bar{x}||_2||y - \bar{y}||_2}\right). \quad (44)$$

where $\bar{x}$ and $\bar{y}$ are the mean of $x$ and $y$, respectively, and $\text{MRSA}(x, y) \in [0, 1]$. The two vectors perfectly match when the MRSA $(x, y) = 0$. The third panel shows the mean of MRSA over the components

$$\text{MRSA} = \sum_k \text{MRSA}(x_k, x_k^G) / N_k, \quad (45)$$

where $x_k^G$ is the ground truth (input spectrum). Moreover, Figure 7 shows the actual shapes of the unmixed spectra for different $\lambda_X$. The difference is observed for Component 2 (orange, land). The unmixed spectra when $\lambda_X \leq 10^0$ resulted in a worse fit to the ground truth. Interestingly, when $\lambda_X \geq 10^3$, the fit of the unmixed spectra to the ground truth also worsened. This is possibly due to the over-constraint on the spectral model, which in turn restricts its ability to explain the data accurately (large residuals); this results in an incomplete estimate of the spectral components.

The comparison between the retrieved map and ground truth is quantified by the Correct Pixel Rate (CPR), which is defined by the correct answer rate of the classification map. The bottom panel of Figure 6 shows its dependence on $\lambda_X$. Insufficient spectral regularization resulted in not only a worse mean MRSA but also a worse estimate of the geography. In regards to both the mean MRSA and CPR, an optimal range for $\lambda_X$ of $10^{-1}$ to $10^{-3}$ was observed.

Although we cannot compute the MRSA and CPR for unknown geographies and surface spectra, these results suggest that a curve of a surrogate of normalized spectral volume, as a function of $\lambda_X$, can be used to determine the optimal value of $\lambda_X$. We suggest the following procedure: (1) Plot the mean residual and $\det(\hat{X}^T\hat{X})$ as a function of $\lambda_X$. (2) Observe the change of the spectral shape as a function of $\lambda_X$. (3) Use $\lambda_X$ at a turning point of the spectral shape and $\det(\hat{X}^T\hat{X})$ and avoid a large value for the mean residual.
Figure 8 shows similar plots to those in Figure 6 but for the spatial regularization $\lambda_A$. The over-constraint on $A$ (i.e., large $\lambda_A$) contributes to a bad fit of the data and a smaller volume of spectral components. Smaller spatial regularization parameters resulted in a bad estimate of the geography, as indicated by the CPR. This is because insufficient spatial regularization creates a noisy map due to the instability of the mapping (see Figure 9 as an example). Contrastingly, a large $\lambda_A$ will slightly decreases the CPR because it will imply that the inferred spectra are getting worse, and the spatial resolution of the map is decreasing in quality. This poor resolution will result in a large mean residual. Therefore, one should check both the residual and the surrogate of normalized spectral volume as a function of $\lambda_A$ because these quantities have a trade-off relation. We suggest choosing the optimal $\lambda_A$ as the smallest value that (1) keeps the noise in the inferred map nonsignificant and (2) avoids a large mean residual.

We have described how spatial and spectral regularization parameters affect the results; we have also discussed a guideline to follow when choosing the optimal spatial and spectral parameters. To find the optimal $\lambda_X$ (or $\lambda_A$), we fixed $\lambda_A$ (or $\lambda_X$) in Figure 6 (or Figure 8). In practice, this procedure should be iterative so that we can find the optimal set of $\lambda_A$ and $\lambda_X$. We recognize that our current guideline for choosing the optimal parameters is not quantitative. Ideally, the performance of the prediction can be used to choose the optimal parameters, such as cross validation. However, the large computational time of the optimization method is too long to perform a cross validation. Therefore, we postpone the quantitative criterion needed to choose the parameters for further study.

3.2. Choice of the Number of Spectral Components

So far, we have assumed the number of spectral components $N_k = 3$. Generally, $N_k$ should be one of the free parameters. Here, we consider the cases for when $N_k = 2$ (over-constrained) and $N_k = 4$ (under-constrained). The cost function for $N_k = 2$ ($Q = 6 \times 10^4$) is much larger than that of $N_k = 3$ ($Q = 2355$) and 4 ($Q = 2565$). This indicates that $N_k = 2$ is insufficient to explain the data. For $N_k = 4$, we could not reach the convergence of the cost function, Equation (26), with the regularization term of Equation (30), even though the number of iterations reached $10^6$, which was where we stopped.

Although information criteria such as the Akaike Information Criterion (AIC) are used as the model selection for a different number of free parameters, the degrees of freedom are not clear for the inverse problem. Ignoring this fact, if we evaluate AIC by $-2 \log \text{(Likelihood)} + 2 \text{(degree of freedom)} = |D - WAX|^2 / \sigma^2 + 2N_kN_f$, where $\sigma$ is the standard deviation of the input noise, we obtain AIC = 135158.8, 23141.4, and 28889.3 for $N_k = 2$, 3, and 4, respectively. These results might indicate that $N_k = 3$ is the optimal number for the components.

Another problem in real data is that it is often difficult to estimate the likelihood because we do not understand the statistical nature of the noise. In this case, cross validation is often used as the model selection. However, cross validation is unrealistic because of the high computational cost of the current scheme. We postpone the criterion that will allow us to choose the optimal number of surface components for further study. Hence, in this paper, we require the number of surface components as prior knowledge for mapping.

3.3. Comparison with Spectral Unmixing on Light Curves

So far, we have explained how geography is disentangled from spectra in spin–orbit unmixing. Here, we consider spectral unmixing on the light curve with no disentanglement of geometry and compare it with the unified model. By minimizing the cost function

$$Q = \frac{1}{2} ||D - \tilde{A}X||^2 + \frac{\lambda_X}{2} \det (XX^T) \quad \text{(46)}$$

subject to $\tilde{A}_{ik} \geq 0$, $X_{il} \geq 0$, \quad \text{(47)}

we obtain the unmixed spectral components for different spectral regularization as shown in Figure 10. We find that both components 0 and 1, which can be interpreted as surface components on continents, are sensitive to the volume regularization. These results

Figure 9. Example of the color composite map for insufficient spatial regularization ($\lambda_A = 10^{-5}$). We adopt $\lambda_A = 10^2$ in this figure.
show that the NMF with simplex volume minimization works even without geometric disentanglement.

Compared with the spin–orbit unmixing, the spectrum of soil (gray dashed) is less reproduced by component 1 (orange) even for the best case, $\lambda_X = 10^3$, and the results are more sensitive to the choice of $\lambda_X$ (see Figure 7 for comparison). This is likely because the geometric disentanglement is essential to sufficiently separate the spectrum of soil from that of vegetation.

### 4. Application to DSCOVR Data

In this section, we demonstrate our method using real multiband light curves of the Earth as observed by DSCOVR (Jiang et al. 2018). DSCOVR has been continuously monitoring our Earth from the L1 point since 2015. The geometry provided by DSCOVR is not the same as the geometry provided by direct imaging, because DSCOVR continuously looks almost at the dayside of Earth. However, the geometric kernel contains latitudinal information because of the axial tilt of the Earth. This enables us to do a two-dimensional mapping (Fan et al. 2019). We use seven optical bands ($0.388, 0.443, 0.552, 0.680, 0.688, 0.764, \text{and } 0.779 \ \mu m$) in DSCOVR filters ($N_l = 7$). The bandwidths are very narrow (0.8–3.9 nm), and there is strong oxygen B and A absorption in 0.688, 0.764 $\mu m$.

Owing to computational efficiency, we use one-fourth of the two-year data (i.e., one in each four bins) used in Fan et al. (2019), resulting in a number of $N_i = 2435$ time bins.

Figure 11 shows the unmixed spectra and color composite map when we assume that $N_k = 4$. For regularization parameters, we followed the procedure described in the previous section. Figure 12 shows the mean MRSA and surrogate of the normalized spectral volume. It was observed that $l = -10^4$ and $l = -10^{4.5}$ are the optimal values, because a significant increase in the mean residual at $l = -10^{1.5}$. Also, we take $10^{-4.5}$ as the optimal value of $\lambda_A$ because of a significant increase in the mean residual at $l = -10^4$.

Figure 11. Normalized unmixed spectra (top panel) and color composite map (bottom panel) for the DSCOVR data. In the top panel, both the 0.688 and 0.764 $\mu m$ bands are strongly affected by oxygen absorption (shaded by blue vertical lines). The bottom panel shows a color composite map. We use white, blue, green, and brown colors for components 0, 1, 2, and 3, respectively.

Figure 12. Mean residual and the surrogate of the normalized spectral volume as functions of $\lambda_A$ (top; $\lambda_X = 10^{-4.5}$) and $\lambda_X$ (bottom; $\lambda_A = 10^{-2}$). We take $10^{-2}$ as the optimal value of $\lambda_A$ because of a significant increase at $\lambda_A = 10^{-1.5}$. Also, we take $10^{-4.5}$ as the optimal value of $\lambda_X$ because of a significant increase in the mean residual at $\lambda_X = 10^{-4}$.

The Astrophysical Journal, 894:58 (14pp), 2020 May 1 Kawahara
Component 0 exhibited a flat spectrum except for strong oxygen absorption bands (0.688, 0.764 μm), reproducing the cloud or ice spectrum. Component 0 as well as component 1 are less sensitive to the choice of λχ compared with components 2 and 3. However, the distribution of component 0 is patchy, except for the localization at the North Pole. The patchy distribution probably reflects a temporal cloud distribution because real clouds do not have a static distribution, some of which might be from the ice near the pole.

These patchy pixels have values that are roughly α = 5 times larger compared to those of other components. Also, the unmixed spectrum of component 0 is β = 50 times higher on wavelength average than the total value of those of other components. The fraction of the patchy pixels is about γ ~ 1/100. Multiplying α, β, and γ, we find that the power of component 0 in the patchy pixels is roughly several times higher than the total power of other components. This value is consistent with the contribution of clouds on reflected light on Earth. The fact that the cloud component is localized in these patchy pixels represents a limitation of the current method, which assumes that all of the components have a static distribution over the observation period. Further improvement is needed so that non-static components can be included to the model.

5. Summary and Discussion

In this paper, we constructed a unified retrieval model for spectral unmixing and spin–orbit tomography (spin–orbit unmixing) using nonnegative matrix factorization and L2 and volume regularization. The spin–orbit unmixing works on the cloudless toy model and real multicolor light curves by DSCOVR. Here, we raise several remaining issues that we did not consider in this study.

The simultaneous estimate of the axial tilt parameters g is first. For simple two-dimensional mapping, Schwartz et al. (2016) analyzed how the axial tilt parameters are inferred from amplitude modulation, and Farr et al. (2018) constructed a Bayesian framework to estimate the parameters. Similar work should also be done in the spin–orbit unmixing. However, the computational cost will be an issue that would need to be addressed, as the optimization of NMF requires a high numerical cost.

Moreover, the clock setting problem still remains. Thus far, all of the works done on two-dimensional mapping assume that we know the exact phase of the geometric kernel. The spin rotation period, derived by the auto-correlation function, was assumed to be used (Fujii & Kawahara 2012). However, as Kawahara (2016) pointed out, the apparent periodicity of the photometric variability is not identical to the spin rotation period. The frequency modulation analysis provides the spin rotation period. For instance, we need to check if the spin–orbit unmixing works well when we use an inferred spin rotation period from the frequency modulation. Otherwise, a technique with a simultaneous estimate of the spin might be required.

The next challenge is how non-static compositions such as clouds can be included in the model (see Luger et al. 2019, as an attempt of the time-dependent mapping). This will become vitally important when we apply this technique to gaseous planets.

Another challenge is the dependency on results of various types of regularization. For instance, Aizawa et al. (2020) reported that the L1+TSV regularization provided better results than the Tikhonov regularization. Furthermore, several other types of volume regularization have been proposed in the field of remote sensing (e.g., Ang & Gillis 2019); therefore, a comparative study of regularization is required.

Additionally, a more quantitative criterion is needed to select the optimal number of surface components and the regularization parameters. Because of the high computational cost, cross validation is unrealistic for the current scheme. An objective criterion to select these parameters will help us to apply the technique to unknown exoplanets where the ground truth is not known.

The author is grateful to the DSCOVR team for making the data publicly available. I deeply appreciate Siteng Fan and Yuk L. Yung for providing the processed light curves and their geometric kernel from the DSCOVR data set. I would also like to thank Masatako Aizawa, Kento Masuda, and Nick Cowan for their insightful discussions. I would also like to thank the anonymous reviewer for a careful reading and constructive suggestions. This work was supported by JSPS KAKENHI grant Nos. JP17K14246, JP18H04577, JP18H01247, and JP20H00170. This work was also supported by the JSPS Core-to-Core Program Planet2 and SATELLITE Research from Astrobiology center (AB022006).

### Appendix A

#### Geometric Kernel of the Spin–Orbit Tomography

##### A.1. Disk-integrated Scattered Light

Here, we summarize the computation of the reflection light from a planet to an observer. The outward energy from a facet dA to a direction with a solid angle dΩ (the left panel in Figure A1) is expressed as

\[ dE = L_\parallel \cos \vartheta_1 dAdΩdλ, \]  

(A1)

where \( L_\parallel \) is the upward radiance, and \( \vartheta_1 \) is a zenith angle between a direction and a normal vector. Let us assume that we observe flux from a planet at a distance of d using a telescope with an effective area \( A_{\text{tel}} \), then light in a cone with a solid angle \( dΩ = dA_{\text{tel}}/d^2 \) contributes to the flux. Therefore, the flux from a facet \( dA \) on a plane to the telescope area \( dA_{\text{tel}} \) can be written as

\[ \Delta EdA_{\text{tel}} = L_\parallel \cos \vartheta_1 dΩdA \frac{L_\parallel}{d^2} \cos \vartheta_1 dAdA_{\text{tel}}. \]  

(A2)
Thus, we obtain the total flux from a planet as

$$f_p = \int_{\text{planet}} \Delta E = \int_{\text{planet}} dA \frac{L_\star}{\pi d^2} \cos \vartheta_1.$$  (A3)

The BRDF of the surface element $s$ is defined by the ratio of the outward radiance to the inward irradiance,

$$R^c(\vartheta_0, \varphi_0, \vartheta_1, \varphi_1) \equiv \frac{L_1(\vartheta_1, \varphi_1)}{E_i(\vartheta_0, \varphi_0)}$$  (A4)

where $\vartheta_0$ and $\varphi_0$ are the solar zenith angle and azimuth angle, respectively, and $\varphi_1$ is the azimuth angle to an observer (see Figure A1).\footnote{We inserted a factor of $\pi$ so that the BRDF becomes identical to the reflectivity when the scattering is isotropic.} For most surface types, the BRDF almost solely depends on a relative azimuth angle $\varphi = \varphi_1 - \varphi_0$ instead of each azimuth angle as

$$R^c(\vartheta_0, \varphi_0, \vartheta_1, \varphi_1) = R^c(\vartheta_0, \vartheta_1, \varphi).$$  (A5)

The stellar irradiance is expressed as

$$E_i(\vartheta_0) = \frac{L_\star}{4 \pi a^2} \cos \vartheta_0 = \frac{f_s d^2}{a^2} \cos \vartheta_0,$$  (A6)

where $a$ is the star–planet distance, and $L_\star$ and $f_s$ are the stellar luminosity and flux, respectively. The flux from a planet is expressed as

$$f_p = \int_{\text{IV}} dA \frac{E_i(\vartheta_0)}{\pi d^2} R^c(\vartheta_0, \vartheta_1, \varphi) \cos \vartheta_1 = \frac{f_s R_p^2}{\pi a^2} \int_{\text{IV}} d\Omega R^c(\vartheta_0, \vartheta_1, \varphi) \cos \vartheta_0 \cos \vartheta_1,$$  (A7)

where IV is the illuminated and visible region as shown in the right panel of Figure A1.

Assuming an isotropic reflection $R^c(\vartheta_0, \vartheta_1, \varphi) = m(\theta, \phi)$, we obtain,

$$f_p = \int d\Omega W_g(t, \theta, \phi) m(\theta, \phi)$$  (A8)

where $W_g(t, \theta, \phi)$ is the geometric kernel for the Lambert approximation.

$$W_g(t, \theta, \phi) = \begin{cases} \frac{f_s R_p^2}{\pi a^2} \cos \vartheta_0 \cos \vartheta_1 \cos \vartheta_0 \cos \vartheta_1 > 0, \\ 0 \text{ otherwise}, \end{cases}$$  (A9)

Here, we define the three fundamental vectors, $e_S$, $e_O$, and $e_R$, which are the unit vector from the planet center to the stellar center, from the planet center to the observer, and the normal unit vector at the planet surface, respectively. Using them, we can rewrite $\vartheta_0 = e_s \cdot e_R$ and $\vartheta_1 = e_o \cdot e_R$. Using the orbital phase $\Theta$ and an orbital inclination $i$, we obtain

$$e_S = (\cos(\Theta - \Theta_{\text{eq}}), \sin(\Theta - \Theta_{\text{eq}}), 0)^T, \quad (A10)$$

$$e_O = (\sin i \cos \Theta_{\text{eq}}, -\sin i \sin \Theta_{\text{eq}}, \cos i)^T, \quad (A11)$$

where $\Theta_{\text{eq}}$ is the orbital phase at equinox.

We also define the spherical coordinate fixed on the planet surface,

$$e_p(\phi, \theta) = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)^T.$$  (A12)

Applying a spin rotation around $\Phi$ and a rotation matrix $R(\zeta)$ as a function of a planet’s obliquity $\zeta$, we get

$$e_R = R(\zeta) e_p(\phi + \Phi, \theta) = \begin{pmatrix} \cos(\Phi + \zeta) \sin \theta + \sin(\Phi + \zeta) \cos \zeta \cos \theta \\ \cos(\Phi + \zeta) \sin \sin \theta + \sin(\Phi + \zeta) \cos \zeta \cos \theta \\ -\sin(\Phi + \zeta) \sin \theta + \cos(\Phi + \zeta) \cos \zeta \cos \theta \end{pmatrix}. \quad (A13)$$

The geometric weight is given by

$$W_g(t, \theta, \phi) = \begin{cases} \frac{f_s R_p^2}{\pi a^2} (e_S \cdot e_R)(e_O \cdot e_R) \text{for } e_S \cdot e_R > 0, e_R \cdot e_O > 0, \\ 0 \text{ otherwise}. \end{cases} \quad (A14)$$

In addition, we consider the case where the reflectivity is constant and isotropic over the surface, $R(\vartheta_0, \vartheta_1, \varphi) = R$ (the Lambert approximation). We take $e_o = (1, 0, 0)^T$ and define the phase angle $\beta = e_S \cdot e_o$, i.e., $e_o = (\cos \beta, \sin \beta, 0)^T$.

These definitions yield Equation (A7):

$$f_p = \frac{f_s R_p^2 R}{\pi a^2} \int_{-\pi/2}^{\pi/2} \sin^2 \theta \cos \phi \int_0^\pi \cos \theta \cos \phi$$

$$\times (\cos \beta \cos \phi \sin \theta + \sin \beta \sin \phi \sin \theta).$$  (A15)
\[ f_\alpha = \frac{2R}{3} \phi_p(\beta) \left( R x \frac{a}{\alpha} \right)^2, \]  
(A16)

where

\[ \phi_p(\beta) = \frac{1}{\pi} [\sin \beta + (\pi - \beta) \cos \beta], \]  
(A17)

is the Lambert phase function.

**Appendix B**

**Optimization of the Weighted NMF by a Block Coordinate Descent**

The block coordinate descent (e.g., Zhou et al. 2011; Kim et al. 2014; Ang & Gillis 2019) consists of the following two subproblems:

1. QP(A): optimization of a quadratic form for \( a_k \) (the column vector of \( A \))
2. QP(X): optimization of a quadratic form for \( x_k \) (the row vector of \( X \))

The block coordinate descent solves these quadratic problems (QP(A) and QP(X)) iteratively using a nonnegative least square (NNLS) scheme. In this appendix, we derive the quadratic forms and then explain the projected gradient descent and its accelerated versions as the NNLS solver.

**B.1. Quadratic Programming**

**B.1.1. Quadratic Form for \( a_k \)**

A (log) likelihood term for the weighted NMF can be rewritten in the quadratic form

\[ \frac{1}{2} ||D - WAX||_F^2 = \frac{1}{2} \sum_i \sum_j \left( \Delta_{il} - \sum_j W_{ij} A_{jk} X_{kl} \right)^2 \]  
(B1)

\[ = \frac{1}{2} \sum_i X_{il} A_{ik}^T \left( \sum_j W_{ij} A_{jk} \right) A_{ik} \]  
\[ - \sum_i \sum_j \Delta_{il} \sum_j W_{ij} A_{jk} X_{kl} + \frac{1}{2} ||\Delta||_F^2 \]  
(B2)

\[ = \frac{1}{2} \sum_i X_{il} A_{ik}^T \left( \sum_j W_{ij} A_{jk} \right) A_{ik} \]  
\[ - \sum_j \left( \sum_i X_{il} \left( \sum_k \Delta_{kl} W_{kj} \right) A_{jk} + \frac{1}{2} ||\Delta||_F^2 \right) \]  
(B3)

\[ = \frac{1}{2} a_k^T \mathcal{L}_A a_k - I^T_k a_k + \text{const} \]  
(B4)

where

\[ \mathcal{L}_A \equiv x^T_k x_k W^T W \]  
(B5)

\[ I_k \equiv W^T \Delta x_k, \]  
(B6)

and \( \Delta = \Delta(k) \) is defined by \( \Delta_{kl} \equiv D_{kl} - \sum_i W_{ij} A_{jk} X_{il} \).

The penalty of the Tikhonov regularization (L2 term) is

\[ \frac{1}{2} \lambda \| A \|_F^2 = \frac{1}{2} a_k^T T_A a_k + \text{const} \]  
(B7)

\[ T_A \equiv \lambda A. \]  
(B8)

Thus, the quadratic programming for the weighted NMF with a spatial Tikhonov regularization minimizes

\[ q_A = \frac{1}{2} a_k^T (\mathcal{L}_A + T_A) a_k - I^T_k a_k. \]  
(B9)

**B.1.2. Quadratic Form for \( x_k \)**

Likewise, we obtain the (log) likelihood term as a quadratic form of \( x_k \) from Equation (B3) as

\[ \frac{1}{2} ||D - WAX||_F^2 = \frac{1}{2} x_k^T L_X x_k - I^T_k x_k + \text{const} \]  
(B10)

\[ L_X \equiv \| W a_k \|_F^2 \mathbb{I} \]  
(B11)

\[ I_k \equiv \Delta^T W a_k. \]  
(B12)

The volume regularization of the Gram determinant term can be written in the quadratic form of \( x_k \)

\[ \frac{1}{2} \lambda_x \det(X^T X) = \frac{1}{2} x_k^T D_X x_k \]  
(B13)

\[ D_X \equiv \lambda_x \det(\bar{X}_k X_k^T) - I^T_k \bar{X}_k \bar{X}_k^T x_k \]  
(B14)

where \( \bar{X}_k \) is a submatrix of \( X \) when we remove the \( k \)-th row of \( X \). The derivation of Equation (B13) is given in Zhou et al. (2011).

In Table B1, we summarize the quadratic terms for different regularization types. This list also includes the log-determinant type of the volume regularization (Ang & Gillis 2018) and a simple L2 term for \( x_k \).

**B.2. Projected Gradient Descent**

The projected gradient descent (PG) based methods to solve a quadratic problem,

\[ q = x^T W x - b^T x \]  
(B15)

are described. The gradient descent with a nonnegative condition is given by

\[ x^{(t+1)} = \mathcal{P}[x^{(t)} - \eta \nabla q] = \mathcal{P}[x^{(t)} - \eta (W x^{(t)} - b)], \]  
(B16)

where the projection operator on a nonnegative orthant is defined by \( \mathcal{P}[x] = \{ \max(x, 0) \} \). We obtain the PG algorithm by adopting the inverse of the Lipschitz constant \( L \) to \( \eta \).

As the Lipschitz constant, one can use the 2-norm of the 2-norm is more efficient than the Frobenius norm, especially for a large matrix.\(^8\)

---

**Projected Gradient Descent (PG)**

Minimization of \( q = x^T W x - b^T x \)

**Initialization:** \( T = I - W/L, s = b/L, x_0 \)

**while Condition do**

\[ x^{(t+1)} = \mathcal{P}[Tx^{(t)} + s] \]

**end while**

---

\(^8\) Therefore, we use a 2-norm for \( X \) and a Frobenius norm for \( A \).
is called the accelerated projected gradient descent. The APG algorithm is summarized as follows.

**Accelerated Projected Gradient Descent (APG)**

Minimization of $q = x^T A x - b^T x$

Initialization: $T = I - W/L$, $s = b/L$, $x^{(0)} = x^{(0)}$, $\alpha_0 = 0.9$

while Condition do
  $x^{(i+1)} = T[y^{(i)} + s]$
  $\alpha_{i+1} = (\alpha_i^2 + 4\alpha_i^2 - \alpha_i^2)/2$
  $\beta_{i+1} = \alpha_i(1 - \alpha_i)/(\alpha_i + \alpha_i^2)$
  $y^{(i+1)} = x^{(i+1)} + \beta_{i+1}(x^{(i+1)} - x^{(i)})$
end while

A residual curve as a function of iteration using Nesterov’s acceleration is not monotonic. Restarting Nesterov’s acceleration when the residual increases significantly improves the convergence rate (O’donoghue & Candès 2015).

**APG + restart**

Minimization of $q = x^T A x - b^T x$

Initialization: $T = I - W/L$, $s = b/L$, $x^{(0)} = x^{(0)}$, $\alpha_0 = 0.9$

while Condition do
  $x^{(i+1)} = T[y^{(i)} + s]$
  $\alpha_{i+1} = (\alpha_i^2 + 4\alpha_i^2 - \alpha_i^2)/2$
  $\beta_{i+1} = \alpha_i(1 - \alpha_i)/(\alpha_i + \alpha_i^2)$
  $y^{(i+1)} = x^{(i+1)} + \beta_{i+1}(x^{(i+1)} - x^{(i)})$
  if $q^{(i+1)} > q^{(i)}$ then
    $x^{(i+1)} = T[x^{(i)} + s]$
  end if
  $y^{(i+1)} = x^{(i+1)}$, $\alpha_i + 1 = \alpha_0$
end while

Figure C1 shows a comparison of the above three algorithms for a randomly generated matrix $A$ and a vector $p$ as a quadratic problem ($x^T A x - b^T x$) for $W = A^T A$ $(100 \times 100)$ matrix and $b = A^T p$. The residual after the $t$-th iteration is defined by $\|A x^{(t)} - b\|^2$, where $x^{(t)}$ is the estimated value after $t$ iterations.

The drawback of the PG and APG method is that convergence is sensitive to the initial point. When all of the components of $T[x^0 - \eta \Delta Q]$ are zero, the algorithm fails.

**Appendix C**

**Optimization of Weighted NMF by Multiplicative Update**

The multiplicative iterative algorithm (Lee & Seung 1999) is often used to minimize the cost function of the standard NMF, given in Equation (17). It can be directly derived from the cost function and the Karush–Kuhn–Tucker first-order optimal conditions (Cichocki et al. 2009). We need to extend the standard multiplicative iterative algorithm to include the geometric kernel $W$ in Equation (27).

First, let us explain the algorithm for the weighted NMF with no regularization term ($R(A) = 0$). Following the derivation of the multiplicative iterative algorithm, we compute the derivative of the cost function of Equation (27) as

$$\nabla_A Q = W^T W A X X^T - W^T D X^T$$  \hspace{1cm} (C1)

$$\nabla_X Q = A^T W^T W A - A^T W^T D.$$ \hspace{1cm} (C2)

To ensure the nonnegativity, we divide the derivative of the cost function into the positive terms and negative terms

$$\nabla Q = [\nabla Q]_+ - [\nabla Q]_- = 0,$$  \hspace{1cm} (C3)

where $[\nabla Q]_+ \geq 0$, $[\nabla Q]_- \geq 0$. The multiplicative update is an operation that multiplies $[\nabla Q]_+$ by $A$ or $X$. This procedure can be interpreted as the steepest gradient descent

$$A \leftarrow A - \eta_A \odot [\nabla_A Q],$$  \hspace{1cm} (C4)

$$\eta_A = A \odot [\nabla_A Q]_+.$$  \hspace{1cm} (C5)
We found that the constraint of Equation (D1) functions as a form of regularization if we combine the constraint with the volume-regularization term. When we use the constraint of Equation (D1) with the volume-regularization term, the effect of the volume-regularization vanishes. We do not recommend the use of the constraint of Equation (D1) in our case.

**References**

Aizawara, M., Kawahara, H., & Fan, S. 2020, arXiv:2004.03941

Ang, M. A., & Gillis, N. 2018, in 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Volume Regularized Non-negative Matrix Factorizations (Piscataway, NJ: IEEE), I

Ang, M. S., & Gillis, N. 2019, JSTARS, 12, 4843

Baldridge, A. M., Hook, S., Grove, C., & Rivera, G. 2009, RSEnv, 113, 711

Berdyugina, S. V., & Kuhn, J. R. 2019, AJ, 158, 246

Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S-I. 2009, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation (New York: Wiley)

Cowan, N. B., Agol, E., Meadows, V. S., et al. 2009, ApJ, 700, 915

Cowan, N. B., Fuentes, P. A., & Haggard, H. M. 2013, MNRAS, 434, 2465

Cowan, N. B., & Strait, T. E. 2013, ApJL, 765, L17

Craig, M. D. 1994, ITGRS, 32, 542

Fan, S., Li, C., Li, J.-Z., et al. 2019, ApJL, 882, L1

Farr, B., Farr, W. M., Cowan, N. B., Haggard, H. M., & Robinson, T. 2018, AJ, 156, 146

Ford, E. B., Seager, S., & Turner, E. L. 2001, Natur, 412, 885

Fu, X., Huang, K., Sidiropoulos, N. D., & Ma, W.-K. 2019, ISPM, 36, 59

Fu, X., Ma, W.-K., Huang, K., & Sidiropoulos, N. D. 2015, ITSP, 63, 2306

Fujii, Y., & Kawahara, H. 2012, ApJ, 755, 101

Fujii, Y., Kawahara, H., Suto, Y., et al. 2010, ApJ, 715, 866

Fujii, Y., Kawahara, H., Suto, Y., et al. 2011, ApJ, 738, 184

Fujii, Y., Lustig-Yaeger, J., & Cowan, N. B. 2017, AJ, 154, 189

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Haggard, H. M., & Cowan, N. B. 2018, MNRAS, 478, 371

Jiang, J. H., Zhai, A. J., Herman, J., et al. 2018, AJ, 156, 26

Kawahara, H. 2016, ApJ, 822, 112

Kawahara, H., & Fujii, Y. 2010, ApJ, 720, 1333

Kawahara, H., & Fujii, Y. 2011, ApJL, 739, L62

Kim, J., He, Y., & Park, H. 2014, Journal of Global Optimization, 58, 285

Lawson, C. L., & Hanson, R. J. 1995, Solving Least Squares Problems, Vol. 15 (Philadelphia: SIAM)

Lee, D. D., & Seung, H. S. 1999, Natur, 401, 788

Lin, C.-H., Ma, W.-K., Li, W.-C., Chi, C.-Y., & Ambikapathi, A. 2015, ITGRS, 53, 5530

Luger, R., Bedell, M., Vanderspek, R., & Burke, C. J. 2019, arXiv:1903.12182

Lustig-Yaeger, J., Meadows, V. S., Tovar Mendoza, G., et al. 2018, AJ, 156, 301

McLinden, C. A., McConnell, J. C., Griffioen, E., McElroy, C. T., & Pfister, L. 1997, JGR, 102, 18,801

Nakagawa, Y., Kodama, T., & Ishiwatari, M. 2020, submitted

Nesterov, Y. E. 1983, Dokl. akad. nauk Sssr, 269, 543

Oakley, P. H. H., & Cash, W. 2009, ApJ, 700, 1428

O’donoghue, B., & Candes, E. 2015, Foundations of computational mathematics, 15, 715

Paatero, P., & Tapper, U. 1994, Environmetrics, 5, 111

Pallé, E., Ford, E. B., Seager, S., Montañés-Rodríguez, P., & Vazquez, M. 2008, ApJ, 676, 1319

Schachtnet, R., Pöppel, G., Tomé, A. M., & Lang, E. W. 2009, Independent Component Analysis and Signal Separation (Berlin: Springer), 106

Schwartz, J. C., Sekowski, C., Haggard, H. M., Pallé, E., & Cowan, N. B. 2016, MNRAS, 457, 926

Vavasis, S. A. 2009, SIAM Journal on Optimization, 20, 1364

Xiang, Y., Peng, D., & Yang, Z. 2015, Blind Source Separation: Dependent Component Analysis (Berlin: Springer)

Zhou, G., Xie, S., Yang, Z., Yang, J.-M., & He, Z. 2011, ITNN, 22, 1626