ON THE SET OF SOLUTIONS OF THE NONNEGATIVE MATRIX FACTORIZATION PROBLEM

KLAUS NEYMEYR*,† AND MATHIAS SAWALL*

Abstract. The nonnegative matrix factorization (NMF) problem $D = XY^T$ for a given nonnegative matrix D and with nonnegative factors X and Y can have many solutions aside from trivial permutations or positive multiples of the columns of X and Y. The set of feasible solutions (SFS) is a low-dimensional representation of all possible columns of either X or Y in any NMF of D. The SFS provides important information on the possible ambiguity of the NMF. This paper conveys the SFS concept as developed in chemometrics to mathematics. Various properties of the SFS are proved. Numerical algorithms for the SFS computation are reviewed and tested for an application model problem from analytical chemistry.

Key words. Nonnegative matrix factorization, Perron-Frobenius theory, factor analysis.

1. Introduction. For a given nonnegative matrix $D \in \mathbb{R}^{k\times n}$ and a positive integer $s \leq \min\{k, n\}$ the *approximate nonnegative matrix factorization problem* is to determine the minimum

$$\min_{X\in\mathbb{R}^{k\times s},\, Y\in\mathbb{R}^{n\times s}} \|D - XY^T\|_F \quad \text{with} \quad X \geq 0,\ Y \geq 0. \tag{1.1}$$

Therein $\|\cdot\|_F$ is the Frobenius norm. If the rank of D equals s, then a factorization $D = XY^T$ with nonnegative factors $X \in \mathbb{R}^{k\times s}$ and $Y \in \mathbb{R}^{n\times s}$ can exist. Such a factorization makes the minimum (1.1) equal to 0. In general the nonnegative rank of D [9], denoted by $\text{rank}_+(D) = r$, is the smallest integer so that nonnegative factors $X \in \mathbb{R}^{k\times r}$ and $Y \in \mathbb{R}^{n\times r}$ exist. Here we assume that $\text{rank}_+(D) = \text{rank}(D) = s$. The focus of this paper is on these rank-s nonnegative matrix factorizations (NMF) of the form $D = XY^T$. The NMF constitutes the limit case $\varepsilon \to 0$ of approximate rank-s factorizations with $\text{rank}(D) > s$ but $\text{rank}_\varepsilon(D) = s$. Therein $\text{rank}_\varepsilon(D)$ is the number of singular values of D which are greater than or equal to ε. Rank-s approximations typically apply to cases where D contains noisy, perturbed or rounded numerical data, e.g., image data or spectral data.

A vast literature exists on nonnegative matrices and their nonnegative factorizations, see for example [2, 4, 8, 24]. The NMF problem has a strong application background, see [7, 26] and many others. The existence and uniqueness of NMFs is discussed e.g. in [18] and also in [10, 34] on the basis of a geometric cone theory. The numerical NMF computation is treated in [5, 15, 20, 21, 25] and many other publications.

Due to the non-convexity of

$$f(X, Y) = \|D - XY^T\|_F^2$$

in X and Y the problem (1.1) typically lacks a unique solution. However, there are trivial ambiguities how to generate from a given nonnegative factorization $D = XY^T$ further nonnegative factorizations:

1. Scaling ambiguity: If Δ is a nonnegative, invertible $s \times s$ diagonal matrix, then
$$D = (X\Delta)(\Delta^{-1}Y^T) \tag{1.2}$$
is a column-rescaled nonnegative matrix factorization of D.
2. Permutation ambiguity: If P is an $s \times s$ permutation matrix, then
$$D = (XP)(YP)^T \tag{1.3}$$
represents a simultaneous permutation of the columns of X and Y.

*Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18055 Rostock, Germany.
†Leibniz Institut für Katalyse, Albert-Einstein-Straße 29a, 18059 Rostock, Germany.
December 12, 2017.

The product $M := \Delta P$ is a *generalized permutation matrix* [24]. It is a well-known fact that M and M^{-1} are nonnegative if and only if M is a generalized permutation matrix, see Lemma 1.1 in [24].

1.1. Aim and organization of this paper. The aim is to determine a low-dimensional representation of the solutions X and Y of the NMF problem. Against the background of a spectroscopic application problem, see Section 2, we are interested in the sets of all possible columns of X respectively Y which arise in all nonnegative matrix factorizations $D = XY^T$. Such a representation of the factors has been developed in chemometrics [6, 12, 19, 27, 28]. Here our intention is to convey these ideas to the mathematical community and to discuss the underlying mathematical theory. The column-oriented approach to the set of the solutions of the NMF problem implicitly allows us to get rid of the permutation and the scaling ambiguities. The representing set for a rank-s matrix D is proved to be a bounded subset of the $\mathbb{R}^{s-1}$ with respect to a reduced basis of singular vectors. The case $s = 2$ was investigated by Lawton and Sylvestre [19] in 1971. See also [1, 35] on the non-uniqueness of such factorizations of D with a chemical application background and for which the term *rotational ambiguity* has been established. Here we call such a set the *set of feasible solutions* (SFS) for general s. For $s = 3$ common names for the SFS are *feasible region* [27] or *area of feasible solutions* [13]. For $s = 4$ the SFS is sometimes called the volume of feasible solutions [14]. The paper not only introduces the SFS and proves various of its properties, but also reports on methods for a numerical computation of the SFS. A major part of the analytic theory of the SFS rests upon the Perron-Frobenius spectral theory of nonnegative matrices.

The paper is organized as follows: First Sec. 2 presents a typical application problem from analytical chemistry which motivates and accompanies the following analytic part of the paper. Sec. 3 defines the SFS by means of an SVD of D. The analytic SFS theory is presented in Sec. 4. Sec. 5 shows the SFS for the model problem from Sec. 2. A stable and fast numerical algorithm to compute the SFS for the important case $s = 3$ is explained in Sec. 6. Finally, Sec. 7 proves an important property of the SFS which makes it possible to compute the SFS for $s \geq 3$.

2. An application in analytical chemistry. A significant field of application of nonnegative matrix factorizations is the so-called *pure component analysis* in analytical chemistry. Next we introduce a typical model problem, namely a three-component two-step consecutive reaction scheme

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \tag{2.1}$$

with the reaction rate constants k_1 and k_2. Throughout this paper we use this model problem to illustrate the theoretical results.

Next we intermediately assume that the pure component spectra for the three components A, B and C are given. Then the Lambert-Beer law [3] in its frequency and time-dependent form [22] allows us to construct a nonnegative matrix whose elements represent the idealized mixture spectra which can be taken from a spectroscopic observation of this reaction system. The steps of this matrix construction are explained later in this section. The matrix assembly phase is followed by a recovery phase we refer to the NMF problem. In this phase we try to recover *only* from the spectral data matrix all the initial data, namely the pure component spectra and also the concentration profiles of the pure components. The mathematical basis for the recovery phase is analyzed in this paper.

Next the spectral data matrix D is constructed: According to the kinetic model (2.1) the time-dependent concentration profiles $c(t) = (c_A(t), c_B(t), c_C(t))$ of the three species

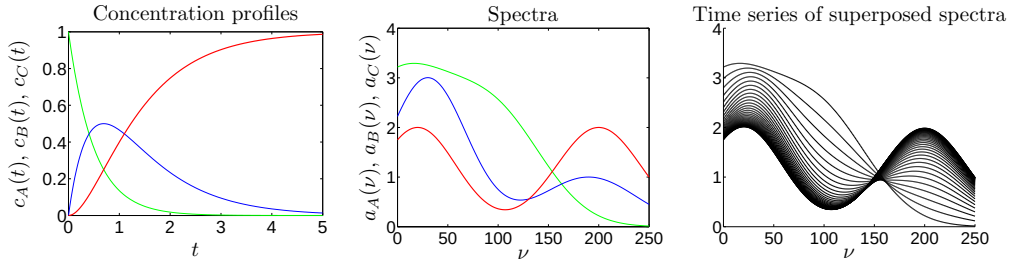

FIG. 2.1. *The three concentration profiles* (2.3) *(left subplot) and the associated spectral profiles* (2.4) *with* $z = 30$ *(centered subplot).Component A (green), component B (blue) and component C (red). Right: Plot of the series of superposed ("measured") spectra.*

satisfy the system of ordinary differential equations

$$\frac{d}{dt}c(t) = (-k_1 c_A(t),\ \ k_1 c_A(t) - k_2 c_B(t),\ \ k_2 c_B(t))\,. \tag{2.2}$$

With the initial concentration values $(c_A(0), c_B(0), c_C(0)) = (a_0, 0, 0)$ at $t = 0$ the system of ordinary differential equations (2.2) has the solution

$$c(t) = \left(a_0 \mathrm{e}^{-k_1 t},\ \ \frac{a_0 k_1}{k_2 - k_1}\left(\mathrm{e}^{-k_1 t} - \mathrm{e}^{-k_2 t}\right),\ \ a_0 - c_A(t) - c_B(t)\right). \tag{2.3}$$

For the limit case $k_1 = k_2 = k$ L'Hospital's rule results in $c_B(t) = a_0 k t\,\mathrm{e}^{-kt}$.

The associated frequency-dependent model spectra $a_A(\nu)$, $a_B(\nu)$ and $a_C(\nu)$ are each formed by a sum of two Gaussians with different centers, signal widths and amplitudes

$$\begin{aligned} a_A(\nu) &= 3\mathrm{e}^{-\nu^2/(200z)} + 2\mathrm{e}^{-(\nu-100)^2/(150z)}, \\ a_B(\nu) &= 3\mathrm{e}^{-(\nu-30)^2/(100z)} + \mathrm{e}^{-(\nu-190)^2/(150z)}, \\ a_C(\nu) &= 2\mathrm{e}^{-(\nu-20)^2/(100z)} + 2\mathrm{e}^{-(\nu-200)^2/(120z)}. \end{aligned} \tag{2.4}$$

The parameter $z > 0$ controls the signal width and thus the overlap of the Gaussians.

Figure 2.1 shows the concentration profiles for $t \in [0, 5]$ and the spectra for $\nu \in [0, 250]$, $z = 30$ and the rate constants $k_1 = 2$ and $k_2 = 1$. The spectroscopic measurement of this system takes k separate spectra each at n frequency values. The spectroscopic data are stored in a $k \times n$ spectral data matrix D. The time-series of superposed spectra are also shown in Figure 2.1.

The Lambert-Beer law of spectroscopy states that the measured absorbance data (i.e. the matrix elements of D) are linear superpositions of the contributions from the chemical components A, B and C. The measured absorbance at a certain frequency and for a fixed chemical component is proportional to the product of its concentration and its absorptivity. Therefore the Lambert-Beer law in matrix form [22] complies with the nonnegative matrix factorization $D = XY^T$. The three columns of X represent the concentration profiles (2.3) evaluated at these k points in time. Further, the columns of Y are the three spectra (2.4) evaluated at the given n frequency values.

In analytical chemistry the so-called *multivariate curve resolution problem* [22, 23] is to recover from given D the unknown matrix factors X and Y of the pure components. This is an NMF problem. The nonnegativity of the factors is a consequence of the physical fact that concentration values and absorptivity constants are nonnegative. Typically, many nonnegative matrix factorizations of D exist and the problem is to find the chemically correct factorization among all nonnegative factorizations. To find these true solutions chemical expertise is required. For instance, concentration profiles are often monotone

functions or have only one isolated maximum. Then all other concentration profiles can be ignored. Similarly, a chemist can identify those feasible nonnegative columns of Y which for physical reasons cannot be spectra of the expectable chemical components. The *mathematical challenge* is to determine in a first step the sets of all possible nonnegative factors of D in a certain low-dimensional representation. Here we do not treat the second step, namely to sort out all solutions which do not satisfy various chemical criteria. Next we introduce the low-dimensional representation of the sets of all possible columns of X and Y in nonnegative factorizations $D = XY^T$.

3. An SVD based representation of the factorization problem. The definition of the SFS requires a singular value decomposition (SVD) $D = U\Sigma V^T$ of the nonnegative matrix $D \in \mathbb{R}^{k\times n}$. The matrices U, V are orthogonal and Σ is a diagonal matrix whose diagonal contains the singular values σ_i. D is a rank-s matrix so that $\sigma_1 \geq \ldots \geq \sigma_s > 0 = \sigma_{s+1} = \sigma_{s+2} = \ldots = \sigma_{\min\{k,n\}}$. This justifies to work with the truncated SVD with $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_s) \in \mathbb{R}^{s\times s}$, $U \in \mathbb{R}^{k\times s}$ and $V \in \mathbb{R}^{n\times s}$.

Any matrix factorization $D = XY^T$ (not only the nonnegative ones) can be represented in terms of the bases of left and right singular vectors with an invertible matrix $T \in \mathbb{R}^{s\times s}$ according to

$$D = \underbrace{(U\Sigma T^{-1})}_{=:X}\underbrace{(TV^T)}_{=:Y^T}. \tag{3.1}$$

The s^2 matrix elements of T parametrize the factorizations of D and also its subset of nonnegative factorizations.

THEOREM 3.1. *Let $D \in \mathbb{R}^{k\times n}$ be a nonnegative matrix with $s = \operatorname{rank}(D) \geq 2$. If D^TD is an irreducible matrix, then $Vw \geq 0$ for a nonzero $w \neq 0$ implies that $w_1 \neq 0$. In other words, any nonzero linear combination of the columns $2, \ldots, s$ of V has at least one negative component.*

Proof. The matrix vector product Vw is written as

$$Vw = V(:,1)w_1 + V(:,2:s)(w_2, \ldots, w_s)^T. \tag{3.2}$$

The Perron-Frobenius theory applied to the irreducible nonnegative matrix D^TD guarantees that the Perron eigenvalue σ_1^2 is simple and that the associated eigenvector satisfies $V(:,1) > 0$ or $V(:,1) < 0$ componentwise. Forming the Euclidean product of $V(:,1)$ with (3.2) and the orthogonality of the columns of V show that

$$0 \neq \underbrace{V(:,1)^T}_{\leqslant 0}\underbrace{Vw}_{\neq 0} = w_1\|V(:,1)\|^2 + V(:,1)^TV(:,2:s)(w_2, \ldots, w_s)^T = w_1,$$

which proves the proposition. □

REMARK 3.2. *In the following we always assume D^TD to be an irreducible matrix. This is not a restrictive assumption for the general application background in analytical chemistry, see Section 2. Irreducibility of D^TD guarantees according to the Perron-Frobenius theory that the first right singular vector $V(:,1)$ is either componentwise positive or componentwise negative. Without loss of generality we also assume that the singular value decomposition $D = U\Sigma V^T$ results in $V(:,1) > 0$. This assumption avoids bothersome case distinctions. Otherwise, the simultaneous replacement of $V(:,1)$ and $U(:,1)$ by $-V(:,1)$ and $-U(:,1)$ yields the desired SVD.*

The following corollary is the analog of Thm. 3.1 for expansions of the columns of X in terms of left singular vectors.

COROLLARY 3.3. *On the assumptions of Theorem 3.1 and if DD^T is an irreducible matrix, then $U\Sigma w \geq 0$ for a nonzero $w \neq 0$ implies that $w_1 \neq 0$.*

Proof. The Perron vector $U(:,1)$ is a componentwise positive vector of the irreducible nonnegative matrix DD^T. The evaluation of the product $U(:,1)^T U\Sigma w$ as in the proof of Theorem 3.1 shows that $w_1 \neq 0$. □

Next we discuss how the scaling ambiguity and the permutation ambiguity can be taken into account in the nonnegative factorization problem: By (3.1) the ith column of Y reads

$$Y(:,i) = \sum_{\ell=1}^{s} V(:,\ell) t_{i\ell}. \tag{3.3}$$

Thm. 3.1 guarantees that $t_{1\ell} \neq 0$ for $\ell = 1, \ldots, s$. This allows us to insert the diagonal matrix $\Delta = \operatorname{diag}(t_{11}, \ldots, t_{s1})$ and its inverse in Eq. (1.2). Hence in

$$D = XY^T = (X\Delta)(\Delta^{-1}Y^T) = (X\Delta)(\Delta^{-1}TV^T)$$

the matrix $\Delta^{-1}T$ has the form

$$\Delta^{-1}T = \begin{pmatrix} 1 & t_{12}/t_{11} & \ldots & t_{1s}/t_{11} \\ 1 & t_{22}/t_{21} & \ldots & t_{2s}/t_{21} \\ \vdots & \vdots & & \vdots \\ 1 & t_{s2}/t_{s1} & \ldots & t_{ss}/t_{s1} \end{pmatrix} \in \mathbb{R}^{s\times s}.$$

Such a scaling of the factors X and Y does not imply any restriction for the application problem from Sec. 2. For structural elucidation purposes the spectra and concentration profiles can (in a first step) be determined qualitatively. A quantitative determination of the factors is not mandatory. Quantitative information by eliminating the influence of the scaling ambiguity is sometimes accessible in an optional second step, for instance, by considering a mass balance or by *a posteriori* calibration measurements.

The permutation ambiguity, see Eq. (1.3), permits a further simplification. The simultaneous permutation P of the columns of X and Y can be expressed by a permutation of the rows of $\Delta^{-1}T$. In other words and according to (1.3) any row of $\Delta^{-1}T$ can be transposed to its first row. Hence the matrix elements $\alpha_1, \ldots, \alpha_{s-1}$ in the first row of

$$P^T\Delta^{-1}T = \begin{pmatrix} 1 & \alpha_1 & \ldots & \alpha_{s-1} \\ 1 & w_{11} & \ldots & w_{1,s-1} \\ \vdots & \vdots & & \vdots \\ 1 & w_{s-1,1} & \ldots & w_{s-1,s-1} \end{pmatrix} \tag{3.4}$$

are the representatives of the possible columns of Y. The set of all possible first rows of $P^T\Delta^{-1}T =: \widetilde{T}$ represents the set of all possible and properly scaled columns of the factor Y. Without loss of generality we can drop the tilde superscript. Then the aim is to determine the set of all possible first row vectors $(1, \alpha_1, \ldots, \alpha_{s-1})$ of T so that T can be supplemented to an invertible matrix by proper matrix elements w_{ij} and that $X = U\Sigma T^{-1}$ and $Y = VT^T$ in (3.1) are nonnegative matrices. This construction is summarized in the following definition.

Definition 3.4 (Set of feasible solutions). *A vector* $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{s-1})^T \in \mathbb{R}^{s-1}$ *is called* feasible *if a matrix* $W \in \mathbb{R}^{s-1\times s-1}$ *exists so that*

$$T = \begin{pmatrix} 1 & \boldsymbol{\alpha}^T \\ e & W \end{pmatrix} \tag{3.5}$$

with the all-ones vector $e = (1, \ldots, 1)^T \in \mathbb{R}^{s-1}$ *is an invertible matrix and that*

$$X = U\Sigma T^{-1} \geq 0, \qquad Y = VT^T \geq 0. \tag{3.6}$$

The set of feasible solutions *(SFS) for the factor Y comprises all feasible vectors $\boldsymbol{\alpha}$*

$$\mathcal{M} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^{s-1} : \exists W \in \mathbb{R}^{s-1 \times s-1} \text{ in (3.5) so that } T \text{ is regular and } X, Y \geq 0 \right\}. \tag{3.7}$$

REMARK 3.5. *The SFS is the set of vectors $\boldsymbol{\alpha}$ which according to* (3.5)-(3.7) *represent the possible nonnegative columns of Y in a way that $\boldsymbol{\alpha}$ can be augmented to the matrix T by* (3.5) *so that $X, Y \geq 0$. The same construction applied to $D^T = YX^T$ serves to define the corresponding set of feasible solutions with respect to the left singular vectors of D. Then irreducibility of DD^T must be assumed. This leads to the corresponding representation of all possible nonnegative columns of the matrix factor X.*

4. Properties of the SFS. In this section various properties of the SFS are proved. First the superset $\mathcal{N}$ of the SFS $\mathcal{M}$ by (3.7) is introduced.

DEFINITION 4.1. *The set*

$$\mathcal{N} = \{ \boldsymbol{\alpha} \in \mathbb{R}^{s-1} : V \begin{pmatrix} 1 \\ \boldsymbol{\alpha} \end{pmatrix} \geq 0 \}, \tag{4.1}$$

contains all coefficient vectors $\boldsymbol{\alpha}$ for which $V(1, \boldsymbol{\alpha}^T)^T$ is nonnegative. Borgen and Kowalski [6] in their work on the case $s = 3$ call the set (4.1) the *first polygon* or FIRPOL. Here this name is adopted for any integer number $s \geq 2$.

THEOREM 4.2. *The set $\mathcal{N}$ is a convex polyhedron. Hence, its subset $\mathcal{M}$ is also bounded.*

Proof. By definition, $\mathcal{N}$ is the intersection of the n half-spaces

$$H_i = \{ \boldsymbol{\alpha} \in \mathbb{R}^{s-1} : V(i, 2:s)\boldsymbol{\alpha} \geq -V(i,1) \} \quad i = 1, \ldots, n.$$

Hence $\mathcal{N}$ is a convex set. The origin $\boldsymbol{\alpha} = 0$ is contained in $\mathcal{N}$ since the Perron vector satisfies $V(:,1) > 0$, cf. Remark 3.2. Convexity of $\mathcal{N}$ implies that $\mathcal{N}$ is unbounded if it contains a semi-straight line $\boldsymbol{\alpha}^* \mathbb{R}_+$ for an $\boldsymbol{\alpha}^* \neq 0$. This means that for $i = 1, \ldots, n$

$$V(i, 2:s)(\omega \boldsymbol{\alpha}^*) \geq -V(i,1) \quad \text{for all } \omega \geq 0.$$

This can only be true if $V(:, 2:s)\boldsymbol{\alpha}^* \geq 0$. The latter inequality contradicts Theorem 3.1 for nonzero $\boldsymbol{\alpha}^*$. Hence $\mathcal{N}$ is a bounded convex polyhedron. The set definition (4.1) requires only nonnegativity of $X(:,1) = V(1, \boldsymbol{\alpha}^T)^T$ whereas (3.7) requires (even more than) $X \geq 0$. Hence $\mathcal{M}$ is a subset of $\mathcal{N}$ and is also bounded. ∎

In contrast to the polyhedron FIRPOL, the SFS does not contain the origin $\boldsymbol{\alpha} = 0$. This is proved next.

THEOREM 4.3. *Let $D \in \mathbb{R}^{k \times n}$ be a nonnegative matrix with $s = \operatorname{rank}(D) \geq 2$ so that $D^T D$ and DD^T are irreducible. Then $0 \notin \mathcal{M}$.*

Let $D = XY^T$ be a nonnegative factorization with $X \in \mathbb{R}^{k \times s}$ and $Y \in \mathbb{R}^{n \times s}$. Moreover, the Perron vector $V(:,1)$ of $D^T D$ is not collinear to any of the columns of Y and the Perron vector $U(:,1)$ of DD^T is not collinear to any of the columns of X.

Proof. For each $\alpha \in \mathcal{M}$ a nonnegative factorization $D = XY^T$ exists so that the first column of Y has the form

$$Y(:,1) = V \begin{pmatrix} 1 \\ \boldsymbol{\alpha} \end{pmatrix}. \tag{4.2}$$

Thus the representation of Y with respect to the right singular vectors as $Y = VT^T$ fulfills $T(1,:) = (1, \boldsymbol{\alpha}^T)$. The associated second column of X in $D = XY^T$ reads $X(:,2) = U\Sigma\upsilon$ for a $\upsilon \neq 0$. For the identity matrix I it holds that

$$0 = I_{1,2} = (TT^{-1})_{1,2} = (1, \boldsymbol{\alpha}^T)\upsilon,$$

which shows that $v_1 = -\boldsymbol{\alpha}^T v(2:s)$. Corollary 3.3 guarantees that $v_1 \neq 0$ so that $\boldsymbol{\alpha} \neq 0$ and $v(2:s,1) \neq 0$. This proves that $0 \notin \mathcal{M}$.

Since $\boldsymbol{\alpha} \neq 0$, Eq. (4.2) shows that $V(:,1)$ cannot be equal or collinear to a column of Y. Applying the same arguments to D^T proves that $U(:,1)$ is not (collinear to) a column of X either. □

DEFINITION 4.4. *The convex hull of the row vectors*

$$w_i := \frac{1}{(DV)_{i,1}}\left((DV)_{i,2},\ldots,(DV)_{i,s}\right) = \frac{1}{\sigma_1 U_{i,1}}(U\Sigma)(i,2:s) \in \mathbb{R}^{1\times s-1} \tag{4.3}$$

for $i = 1,\ldots,k$ is a polyhedron which is called INNPOL

$$\mathcal{I} := \text{convhull}\{w_i \,:\, i = 1,\ldots,k\}.$$

The name INNPOL was coined in [6] for the case $s = 3$; here we consider $s \geq 3$.

LEMMA 4.5. *Let DD^T be an irreducible matrix. Then the origin $\boldsymbol{\alpha} = 0$ is an interior point of $\mathcal{I}$. Thus the volume of the polyhedron $\mathcal{I}$ is nonzero.*

Proof. We assume $\boldsymbol{\alpha} = 0$ not to be an interior point of $\mathcal{I}$. Then convexity of $\mathcal{I}$ implies that $\mathcal{I}$ is a subset of the half plane

$$\mathcal{H} = \{z \in \mathbb{R}^{1\times s-1} :\ zy \geq 0\}$$

for a proper nonzero column vector $y \in \mathbb{R}^{s-1\times 1}$. With (4.3) it holds that

$$0 \leq w_i y = \frac{1}{\sigma_1 U_{i,1}} U\Sigma(i,2:s)y, \qquad i = 1,\ldots,k,$$

which reads in vectorial form

$$\frac{1}{\sigma_1}\text{diag}(1/U_{1,1},\ldots,1/U_{k,1})U\Sigma(:,2:s)y \geq 0.$$

Since $\sigma_1 > 0$ and $U(:,1) > 0$ by Thm. 3.1 and Remark 3.2, this means that $U\Sigma(:,2:s)y \geq 0$ for the given $y \neq 0$. This contradicts Corollary 3.3 for irreducible DD^T. □

THEOREM 4.6. *Let $D \in \mathbb{R}^{k\times n}$ be a nonnegative matrix with the rank s so that $D^T D$ and DD^T are irreducible. A vector $\boldsymbol{\alpha} \in \mathbb{R}^{s-1}$ satisfies $\boldsymbol{\alpha} \in \mathcal{M}$ if and only if $\boldsymbol{\alpha} \in \mathcal{N}$ and if the simplex being the convex hull of $\boldsymbol{\alpha}$ and other $s-1$ vectors in $\mathcal{N}$ contains the polyhedron $\mathcal{I}$.*

Proof. Let an element $\boldsymbol{\alpha} = \boldsymbol{\alpha}^{(1)}$ of $\mathcal{M}$ be given. Then, by definition, a nonnegative factorization $D = XY^T$ of D exists with $Y = VT^T$, $T(:,1) = e$ and $T(1,:) = (1,(\boldsymbol{\alpha}^{(1)})^T)$. The $s-1$ vectors $\boldsymbol{\alpha}^{(m)}$ with $T(m,:) = (1,(\boldsymbol{\alpha}^{(m)})^T)$ for $m = 2,\ldots,s$ are also elements of $\mathcal{M} \subset \mathcal{N}$. It remains to show that the simplex with the vertices $\boldsymbol{\alpha}^{(m)}$, $m = 1,\ldots,s$, includes $\mathcal{I}$. To show this we start with

$$DV = XY^T V = X(VT^T)^T V = XT.$$

The components $2:s$ of its ith row are

$$(DV)(i,2:s) = X(i,:)T(:,2:s) \quad i = 1,\ldots,k.$$

Division by $(DV)_{i,1} \neq 0$ together with (4.3) yield

$$w_i = \frac{(DV)(i,2:s)}{(DV)_{i,1}} = \frac{X(i,:)}{(DV)_{i,1}}T(:,2:s) \quad i = 1,\ldots,k. \tag{4.4}$$

This proves that the generating vectors w_i of $\mathcal{I}$ are linear combinations of the row vectors $T(m, 2:s) = (\boldsymbol{\alpha}^{(m)})^T$. The linear combinations are convex combinations if $X_{i,j}/(DV)_{i,1} \geq 0$ and $\sum_{j=1}^{s} X_{i,j}/(DV)_{i,1} = 1$. First, the inequality holds since $X \geq 0$ and $(DV)_{i,1} = D(i,:)V(:,1)$ with $D \geq 0$ and $V(:,1) > 0$. For the second property we start with

$$Xe = XTe_1 = XTV^TVe_1 = XY^TVe_1 = DVe_1,$$

whose ith component reads

$$(Xe)_i = \sum_{j=1}^{s} X_{i,j} = (DV)_{i,1}.$$

This proves the desired equality.

To prove the reverse direction we start with the $(s-1)$-dimensional vectors $\boldsymbol{\alpha}^{(m)} \in \mathcal{N}$, $m = 1, \ldots, s$, whose convex hull encloses $\mathcal{I}$. These vectors form an $(s \times s)$-matrix T in a way that $T(m,:) = (1, (\boldsymbol{\alpha}^{(m)})^T)$ for $m = 1, \ldots, s$. The matrix T is regular since the convex hull of the $\boldsymbol{\alpha}^{(m)}$ encloses $\mathcal{I}$ which is nonempty by Lemma 4.5. The inclusions $\boldsymbol{\alpha}^{(m)} \in \mathcal{N}$ for $m = 1, \ldots, s$ guarantee according to (4.1) that $Y = VT^T \geq 0$. It remains to show that $X \geq 0$. As $\mathcal{I}$ is a subset of the simplex with the vertices $\boldsymbol{\alpha}^{(m)}$, $m = 1, \ldots, s$, it holds for each $i \in \{1, \ldots, k\}$ that

$$w_i = \sum_{m=1}^{s} \mu_m^{(i)} \boldsymbol{\alpha}^{(m)} \tag{4.5}$$

with the coefficients $\mu_m^{(i)}$ of the convex combination satisfying $\mu_m^{(i)} \geq 0$ and $\sum_{m=1}^{s} \mu_m^{(i)} = 1$. Since $T(m, 2:s) = (\boldsymbol{\alpha}^{(m)})^T$ a comparison of (4.5) with (4.4) shows that

$$0 \leq \mu_m^{(i)} = X_{i,m}/(DV)_{i,1},$$

which proves $X \geq 0$. □

Next we show that the representing matrices T which belong to the SFS $\mathcal{M}$ are bounded away from singular matrices.

THEOREM 4.7. *Let $D \in \mathbb{R}^{k \times n}$ be a nonnegative matrix with the rank s so that D^TD and DD^T are irreducible. Then a $\sigma > 0$ exists so that*

$$\sigma_s(T) \geq \sigma > 0, \tag{4.6}$$

for each T belongs to a nonnegative factorization $D = XY^T$ with $X = U\Sigma T^{-1}$ and $Y = VT^T$. Therein $\sigma_s(T)$ is the smallest singular value of T.

Proof. We assume that only $\sigma = 0$ satisfies (4.6) and derive a contradiction. Then a sequence of factorizations $D = X^{(i)}(Y^{(i)})^T$, $i = 1, 2, \ldots$, with $X^{(i)}, Y^{(i)} \geq 0$ exists so that

$$(Y^{(i)})^T = T^{(i)}V^T \qquad \text{and} \qquad \lim_{i \to \infty} \sigma_s(T^{(i)}) = 0.$$

For any i with $T := T^{(i)}$ of the form

$$T = \begin{pmatrix} 1 & (\boldsymbol{\alpha}^{(1)})^T \\ \vdots & \vdots \\ 1 & (\boldsymbol{\alpha}^{(s)})^T \end{pmatrix}$$

it holds that

$$\det(T) = \begin{vmatrix} 1 & (\boldsymbol{\alpha}^{(1)})^T \\ 0 & (\boldsymbol{\alpha}^{(2)})^T - (\boldsymbol{\alpha}^{(1)})^T \\ \vdots & \vdots \\ 0 & (\boldsymbol{\alpha}^{(s)})^T - (\boldsymbol{\alpha}^{(1)})^T \end{vmatrix} = \begin{vmatrix} (\boldsymbol{\alpha}^{(2)})^T - (\boldsymbol{\alpha}^{(1)})^T \\ \vdots \\ (\boldsymbol{\alpha}^{(s)})^T - (\boldsymbol{\alpha}^{(1)})^T \end{vmatrix} = (s-1)!\,\mathrm{vol}(\mathcal{S}(T)),$$

where $\mathcal{S}$ is the $(s-1)$-dimensional simplex with the vertices 0 and $(\boldsymbol{\alpha}^{(j)})^T - (\boldsymbol{\alpha}^{(1)})^T$, $j = 2, \ldots, s$. By Theorem 4.6 the simplex $(\boldsymbol{\alpha}^{(1)})^T + \mathcal{S}$ encloses $\mathcal{I}$, i.e. INNPOL. The volume of $\mathcal{I}$ is bounded from below by the volume of a ball $B_\varepsilon(0)$ centered at the origin with the radius $\varepsilon > 0$ as 0 is an interior point of $\mathcal{I}$. This results in

$$\frac{1}{(s-1)!}\det(T) = \mathrm{vol}(\mathcal{S}(T)) \geq \mathrm{vol}(\mathcal{I}) \geq \mathrm{vol}(B_\varepsilon(0)) > 0.$$

For each i the absolute value of the determinant equals the product of all singular values σ_ℓ so that

$$\frac{1}{(s-1)!}|\det(T^{(i)})| = \frac{1}{(s-1)!}\prod_{\ell=1}^{s}\sigma_\ell(T^{(i)}) = \mathrm{vol}(\mathcal{S}(T^{(i)})) \geq \mathrm{vol}(B_\varepsilon(0)) > 0.$$

Thus the limit $\lim_{i\to\infty}\sigma_s(T^{(i)}) = 0$ results in a contradiction if the remaining singular values are bounded from above independently on i. To show this, it is sufficient to bound $\|T^{(i)}\|$ from above since

$$\|T^{(i)}\| = \sigma_1(T^{(i)}) \geq \sigma_2(T^{(i)}) \geq \ldots \geq \sigma_s(T^{(i)}).$$

Let $(1, \boldsymbol{\alpha}^T)$ be an arbitrary row of $T^{(i)}$. Then

$$\|(1, \boldsymbol{\alpha}^T)\| \leq (1 + \mathrm{diam}(\mathcal{N})^2)^{1/2} =: M$$

since $\alpha \in \mathcal{N}$ and where $\mathrm{diam}(\mathcal{N})$ is the diameter of the polyhedron FIRPOL. Hence M is a constant for the given matrix D since FIRPOL is a bounded set by Thm. 4.2. Direct computation shows that

$$\|T^{(i)}\| = \|(T^{(i)})^T\| = \max_{x\neq 0}\frac{\|(T^{(i)})^T x\|}{\|x\|} \leq \max_{x\neq 0}\frac{\|x\|_1}{\|x\|}M \leq \sqrt{s}M.$$

This concludes the proof. □

After these preparatory steps we can show that the SFS $\mathcal{M}$ is a closed set.

THEOREM 4.8. *On the assumptions of Thm. 4.7 the SFS $\mathcal{N}$ is a compact set.*

Proof. First we form the set

$$\mathcal{T} = \{T \in \mathbb{R}^{s\times s} \,:\, T(i, 2:s) = (1, (\boldsymbol{\alpha}^{(i)})^T),\ \boldsymbol{\alpha}^{(i)} \in \mathcal{N} \text{ for } i = 1, \ldots, s \text{ and } \sigma_s(T) \geq \sigma\}.$$

Compactness of $\mathcal{T}$ follows from the fact that the $\boldsymbol{\alpha}$ are from the compact set $\mathcal{N}$, see Thm. 4.2, and by $\sigma_s(T) \geq \sigma > 0$. The later inequality is guaranteed by Theorem 4.7 for any nonnegative factorization $D = XY^T$ for T in the sense of Eq. (3.6) and avoids that $\mathcal{T}$ contains a sequence of matrices which tends to a singular matrix. The image of $\mathcal{T}$ with respect to the continuous matrix inversion

$$\overline{\mathcal{T}} = \{T^{-1} \,:\, T \in \mathcal{T}\}$$

is also compact. Its subset of all T^{-1} which additionally satisfy that $U\Sigma T^{-1} \geq 0$

$$\widehat{\mathcal{T}} := \{T^{-1} \in \overline{\mathcal{T}} \,:\, U\Sigma T^{-1} \geq 0\}$$

is a closed set due to the greater-and-equal constraints. Hence the intersection

$$\mathcal{T}^* = \mathcal{T} \cap \{T \,:\, T^{-1} \in \widehat{\mathcal{T}}\}$$

is also compact. Thus for a $T \in \mathcal{T}^*$ it holds that $TV^T \geq 0$, $U\Sigma T^{-1} \geq 0$ and $T(:,1) = e$ which are the defining conditions of the SFS. Thus the SFS written as

$$\mathcal{M} = \{ \text{ rows of } T \in \mathcal{T}^* \text{ aside from the leading 1}\}$$

is also a compact set. □

5. The SFS for the model problem. This section presents the SFS for various parameter selections for the model problem as introduced in Section 2. First Figure 5.1 shows in its upper row the spectra (2.4) of the three components A, B and C for three values of the signal width parameter $z \in \{1, 30, 50\}$. The overlap of the spectra increases with an increasing signal width. The associated concentration profiles are given by (2.3) independently of z. The time interval $t \in [0, 5]$ is discretized by $k = 200$ equidistant points and the frequency interval $\nu \in [0, 250]$ uses $n = 251$ equidistant frequency values. According to the three possible values of z three matrices $D \in \mathbb{R}^{200\times 251}$ are formed. The lower row of Figure 5.1 presents the three associated SFS plots. For $z = 1$ with the lowest signal overlap the factorization $D = XY^T$ is nearly unique. Then the SFS consists of three almost point-shaped subsets which are drawn in Figure 5.1 by three points in red, blue and green color. For $z = 30$ the SFS shows that an increased signal width results in an increased size of the set of possible nonnegative factorizations $D = XY^T$. The SFS consists of three relatively large isolated subsets which are plotted in Figure 5.1 in red, blue and green color. Finally, for the largest signal width $z = 50$ the SFS is a single topologically connected set. These sets of feasible solutions confirm the result of Thm. 4.3 that the origin is never contained in the SFS.

With increasing z the SFS has an increasing area. Correspondingly the range of possible nonnegative factorizations $D = XY^T$ increases. Figure 5.2 shows for the three components A, B and C the possible series of spectra for the case $z = 30$. To this end the (α_1, α_2) plane of the SFS is covered by a quadratic mesh and for each grid point which falls into one of the three SFS subsets the function $V(1, \alpha_1, \alpha_2)^T$ is plotted. Additionally, the boundary curve of each subset of the SFS is covered with nodes which are equidistant with respect to the curve parameter. Again the associated solutions are plotted. The comparison with the original spectra in the centered subplot in the first row of Figure 5.1 shows that the series of feasible spectra contains (aside from scaling) the initial spectra in each case of the three components A, B and C.

Finally, Fig. 5.3 shows the SFS for $z \in \{1, 30, 50\}$ of the transposed matrix $D^T = YX^T$. These SFS sets represent the possible columns of X with respect to the left singular vectors of the SVD $D = U\Sigma V^T$. Once again, with increasing z the area of the SFS increases. The figure shows in its lower row for $z = 30$ the associated series of possible nonnegative columns of the matrix factor X, namely for the reactant A for the intermediate component B and the reaction product C. The original and true concentration profiles by Eq. (2.3) are (aside from scaling) contained in these series of spectra.

5.1. NMF algorithms and the SFS. The SFS (by its column-oriented representation of all possible nonnegative factors of a given nonnegative matrix) provides the basis for a comparative study of the results of various NMF algorithms. Many numerical NMF algorithms have been devised, see for example Kim and Park [15–17] or Lee and Seung [20, 21].

Next we use the model data $D \in \mathbb{R}^{200\times 251}$ for $z = 30$ and apply to it four different NMF algorithms. Each algorithm is started 200 times. The NMF codes use random initial values so that the resulting factors are different. The representatives of the columns of

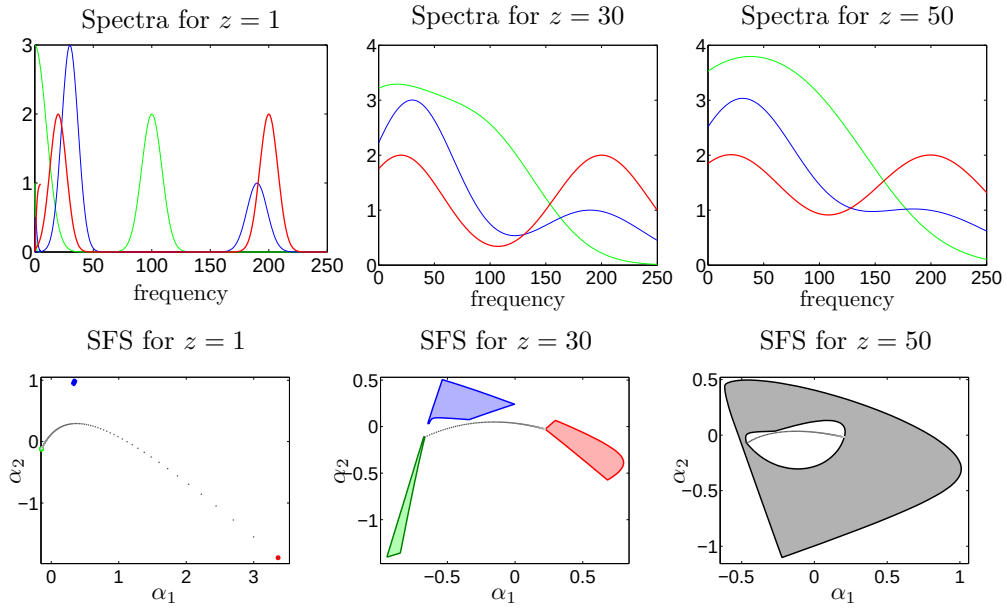

FIG. 5.1. *Upper row of 2D line plots: The spectra* (2.4) *are shown for* $z = 1$, $z = 30$ *and* $z = 50$. *With an increasing value of the peak width parameter* z *the spectra of the components* A *(green),* B *(blue) and* C *(red) show an increasing overlap. Lower row of plots: SFS plots according to* (3.7). *The left subplot for* $z = 1$ *shows three very small subsets of the SFS which indicates more-or-less unique solutions. The ambiguity increases for* $z = 30$ *(centered subplot) and* $z = 50$ *(right subplot). In each subplot* $k = 200$ *gray dots is plotted. The coordinates of these points are the expansion coefficients of the rows of* D *with respect to the right singular vectors with a scaling so that the expansion coefficient of* $V(:, 1)$ *equals 1.*

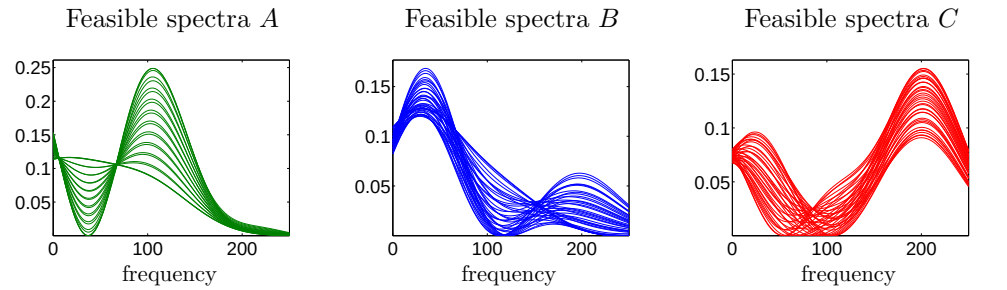

FIG. 5.2. *A series of possible nonnegative spectra is shown for* $z = 30$. *Left subplot (component A), centered subplot (component B) and right subplot (component C). A comparison with the original spectra in the centered subplot in the first row of Figure 5.1 shows that the series of feasible spectra for the components A, B and C (aside from scaling) include the original spectra.*

the computed factors X are marked in the corresponding SFS. The same is done for the columns of Y. Fig. 5.4 shows the results which have been computed by the *Nonnegative Matrix and Tensor Factorization Algorithms Toolbox* by Kim and Park [16, 17]. The following toolbox algorithms have been used:

1. Alternating non-negative least-squares (ANLS) with the block principal pivoting method (ANLS_BPP).
2. ANLS with the active set method and column grouping (ANLS_ASGROUP).
3. The hierarchical alternating least squares method (HALS).
4. The multiplicative updating method (MU).

The numerical results show that the different methods result in different scatter patterns in the SFS subsets. For all these computations the stopping tolerance parameter TOL has been decreased from 10^{-3} to 10^{-8} and the parameter MAX_ITER has been dou-

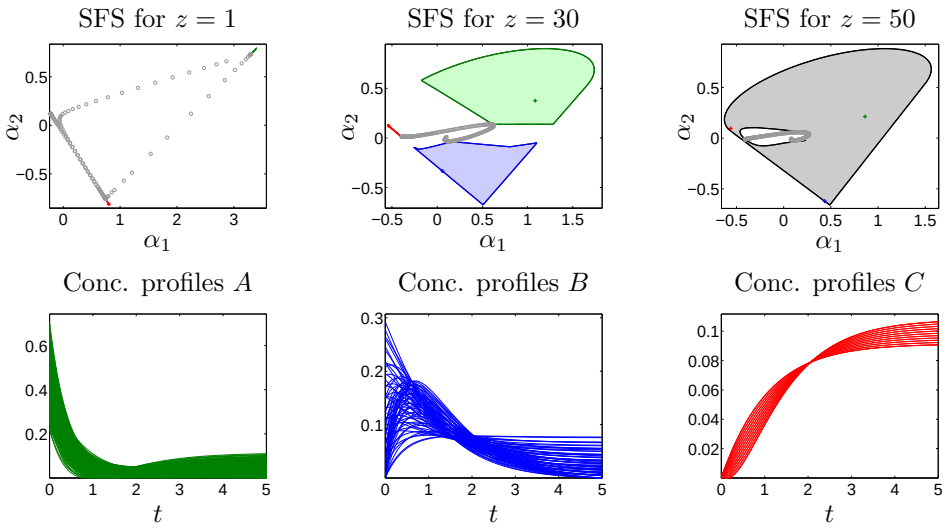

FIG. 5.3. *Upper row of SFS plots: SFS sets for $z \in \{1, 30, 50\}$ which represent the possible nonnegative columns of the factor X with respect to the left singular vectors of D. In each SFS a number of n small gray circles is plotted. The coordinates of these points are the expansion coefficients of the n columns of D with respect to the left singular vectors of D. Once again, these coefficients are rescaled so that the first coefficient equals 1 and is not plotted. Lower row of feasible concentration profiles for $z = 30$. The correct profiles by Eq.* (2.3) *are (aside from scaling) contained in these series of spectra.*

bled to a value of 1000 maximal iterations in order to avoid approximate factors whose representatives are not located in the SFS sets. Nevertheless there are still approximate factorizations which are not located in the SFS; see for example the + symbols outside the blue SFS subsets for the MU method.

6. Numerical approximation of the SFS for $s = 3$. This section briefly explains how the SFS-plots, as shown in Section 5, can be computed. In 1971 Lawton and Sylvestre [19] were the first to solve the problem of how to represent the possible columns of the nonnegative factors of a given matrix for the background problem of pure component decompositions in analytical chemistry, see Section 2. This first work is restricted to ($s = 2$)-component systems. In 1985 Borgen and Kowalski [6] and later Rajkó and István [27] extended these representations to three-component systems ($s = 3$) in terms of a geometry-based SFS construction, see also [29].

The first numerical SFS approximation method for $s = 3$ was suggested in 2011, see [13]. This algorithm aims at the enclosure of the boundary lines of the SFS by chains of equilateral triangles which cover the boundary. An extension of this algorithm to the case $s = 4$ appeared in [14]. Instead of representing the ambiguity of the NMF in the SFS, one can also plot series of possible columns of X and Y. A related analysis on upper and lower enclosing curves is given by Gemperline [11] and Tauler [33]. See also [12] for a review on recent AFS methods.

Here we consider a more efficient numerical algorithm for the approximation of the boundary of the SFS by sequences of adaptively refined polygons. This algorithm is called the *polygon inflation algorithm* [30]. If the SFS consists of several isolated hole-free subsets, then the boundary of each subset is computed separately. For the other case that the SFS is a topologically connected set which contains a hole around the origin (see Thm. 4.3) we need a special approximation technique which allows the computation of the inner and the outer boundaries. For $s = 3$ further cases than the mentioned ones, e.g. more than one SFS subset which contains a hole, cannot occur. This can be understood by the

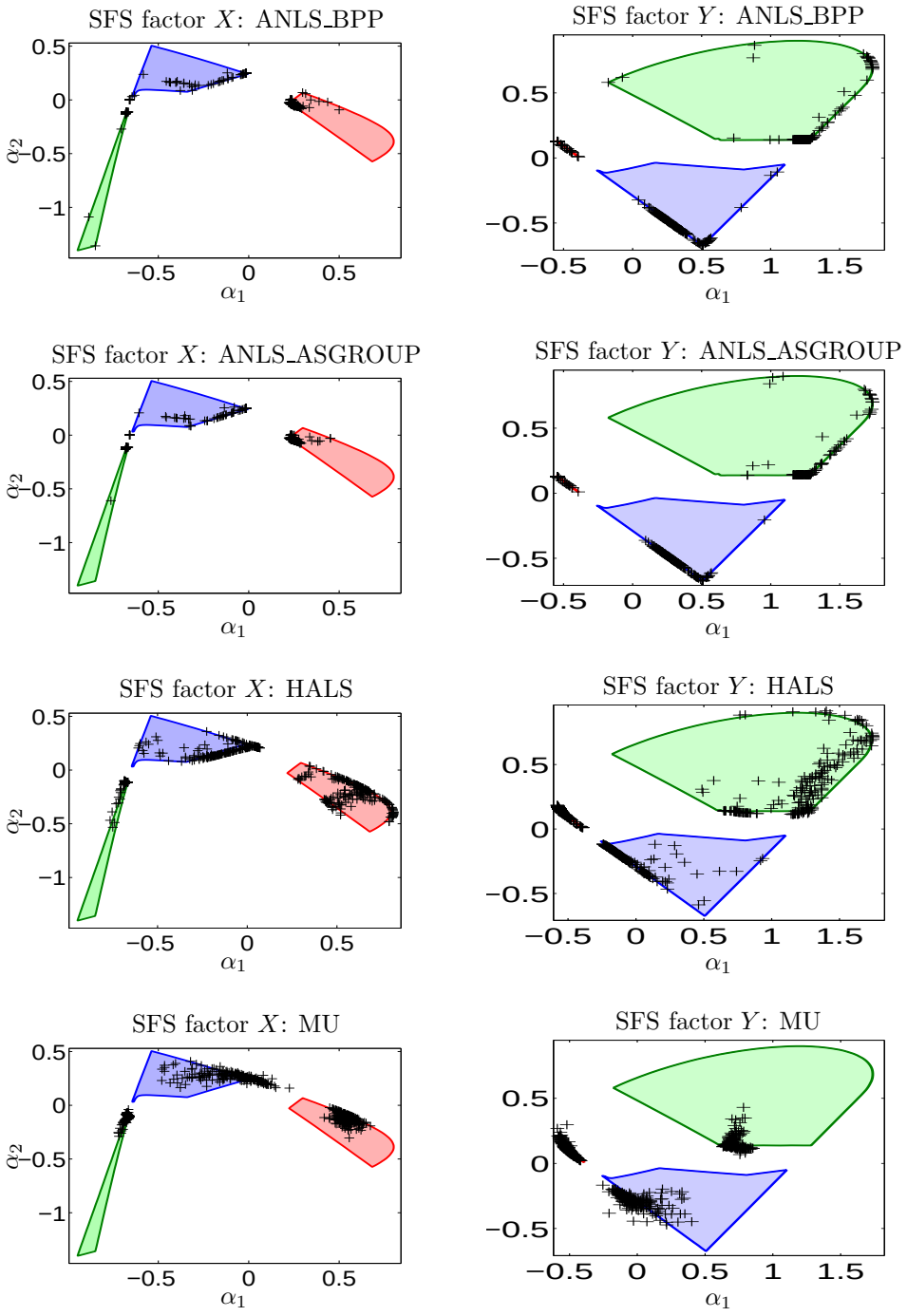

FIG. 5.4. *The nonnegative matrix $D \in \mathbb{R}^{200\times 251}$ is factorized by four different NMF algorithms. Each algorithm is started 200 times. The columns of X and Y are marked in the SFS plane by $+$ symbols. First row: Factorization by alternating non-negative least-squares (ANLS) with block principal pivoting method (ANLS_BPP). Second row: ANLS with the active set method and column grouping (ANLS_ASGROUP). Third row: Hierarchical alternating least squares method (HALS). Fourth row: Multiplicative updating method (MU). We have used the implementation in the* Nonnegative Matrix and Tensor Factorization Algorithms Toolbox *by Kim and Park [16, 17]. For all computations the default stopping tolerance parameter TOL has been decreased from 10^{-3} to 10^{-8} and the parameter MAX_ITER has been doubled to 1000 maximal iterations. In particular for the multiplicative updating method some of the factorizations are not sufficiently accurate so that the $+$ symbols are not located in the SFS.*

geometric SFS construction by the triangle-rotation algorithm in [6]. Hence, the SFS sets as shown in Figures 5.1 and 5.3 are typical representatives of the general case. However, sets of feasible solutions exist which consist of a multiple of 3 isolated hole-free subsets for properly constructed matrices D. In practical applications (analytical chemistry) only the cases of a connected SFS and of three isolated subsets are known. These typical cases are illustrated in Fig. 5.1.

The polygon inflation algorithm is based on a target function whose minimization allows to decide whether or not a certain point $\boldsymbol{\alpha} \in \mathbb{R}^2$ is contained in $\mathcal{M}$. A numerically efficient and stable SFS algorithm should be able to work with experimental, noisy or perturbed spectral data matrices D. Such matrices can even contain small negative matrix elements. The source of these negative matrix entries can be baseline correction steps or rank-reducing subtractions from the spectral data in order to remove background signals or already known components. A further source can be a low-rank approximation which can be used to reduce noise and other perturbations.

Hence a stable and practically usable numerical algorithm should be robust with respect to small negative entries of X and Y. To this end a small parameter $\varepsilon \geq 0$ is fixed so that

$$\frac{\min_i X_{ij}}{\max_i |X_{ij}|} \geq -\varepsilon, \quad \frac{\min_i Y_{ij}}{\max_i |Y_{ij}|} \geq -\varepsilon, \quad j = 1,2,3 \tag{6.1}$$

are lower bounds for the acceptable smallest (negative) entries in X and Y in terms of a relative measure. In order to decide if $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathcal{M}$ is true, we consider the target function $f : \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \to \mathbb{R}$ with

$$\begin{aligned} f(\boldsymbol{\alpha}, W) = &\sum_{i=1}^{k}\sum_{j=1}^{3} \min(0, \frac{X_{ij}}{\|X(:,j)\|_\infty} + \varepsilon)^2 + \sum_{i=1}^{n}\sum_{j=1}^{3} \min(0, \frac{Y_{ij}}{\|Y(:,i)\|_\infty} + \varepsilon)^2 \\ &+ \|I_3 - TT^+\|_F^2. \end{aligned} \tag{6.2}$$

Therein $\|\cdot\|_\infty$ is the maximum norm, $\|\cdot\|_F$ is the Frobenius norm, T^+ is the pseudo-inverse of T and $W \in \mathbb{R}^{2\times 2}$ is the $(2,2)$ block submatrix of T in (3.5). Numerically and in consideration of small negative entries in X and Y a point $\boldsymbol{\alpha}$ is contained in the SFS if the function

$$F : \mathbb{R}^2 \to \mathbb{R}, \quad F(\boldsymbol{\alpha}) = \min_{W \in \mathbb{R}^{2\times 2}} f(\boldsymbol{\alpha}, W) \tag{6.3}$$

satisfies $F(\boldsymbol{\alpha}) \leq \varepsilon_{\text{tol}}$ with a small ε_{tol}, for example $\varepsilon_{\text{tol}} = 100\text{eps}$ with the machine precision eps. Hence the numerical approximation $\widetilde{\mathcal{M}}$ of $\mathcal{M}$ is

$$\widetilde{\mathcal{M}} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^2 : \ F(\boldsymbol{\alpha}) \leq \varepsilon_{\text{tol}} \right\}. \tag{6.4}$$

With the target function f and the resulting function F the polygon inflation algorithm works as follows:

1. First construct an initial nonnegative matrix factorization $D = XY^T \in \mathbb{R}^{k\times n}$ by a proper numerical algorithm, e.g., by the algorithm of Lee and Seung [21] or the algorithm by Kim and Park [15]. According to $Y = VT^T$ the three rows of T result in three points in the SFS.
2. Starting from one of these points the value of α_1 is increased to a maximal value $\overline{\alpha}_1$ so that $(\overline{\alpha}_1, \alpha_2)$ is located on the boundary of $\widetilde{\mathcal{M}}$. Then α_1 is minimized so that $(\underline{\alpha}_1, \alpha_2)$ is also located on the boundary. Finally, a third point on the boundary of the SFS is determined by a minimization of α_2 which results in the boundary point $((\overline{\alpha}_1 + \underline{\alpha}_1)/2, \underline{\alpha}_2)$. These three boundary points of $\widetilde{\mathcal{M}}$ define an initial triangle.

3. Each of the three edges of the initial triangle is separated in the middle and the closest boundary point is determined as a new vertex of a refined polynomial. For each of the three edges the resulting gain-of-area in the refinement is stored as an attribute of the two new edges.
4. The edge partition and polygon refinement steps are repeated until the gain-of-area for each of the edges falls below a predefined threshold value.

The adaptive polygon inflation procedure is illustrated in Fig. 6.1, where it is used to approximate the boundary of a quarter-circle. The vertices of the initial triangle are on the boundary of the quarter circle. A small number of refinement steps is needed on the rectilinear parts of the boundary of the quarter circle and a larger number of refinement steps on the curvilinear boundary. Table 1 shows the Hausdorff distances of the polygons and the quarter circle against the number of vertices. These data validate the effectiveness of the adaptive procedure.

The approximation error of the piecewise linear interpolation of the boundary of $\mathcal{M}$ behaves like $\mathcal{O}(h^2)$ in the edge-length h. Together with the adaptiveness of the algorithm the boundary approximation of $\mathcal{M}$ is much more effective compared to the enclosure of the boundary by a sequence of equilateral triangles whose approximation error equals $\mathcal{O}(h)$.

Whenever the algorithm detects that the SFS consists of only a single topologically connected set, which then contains a hole around the origin according to Thm. 4.3 , then an alternative algorithm is employed. This algorithm is called *inverse polygon inflation*, see [31]. This algorithm is computationally somewhat more expensive (less than the factor of 2) and also applicable to the general case of any SFS. The idea of inverse polygon inflation is to compute first the boundary of the bounded set FIRPOL as introduced in Def. 4.1. The numerical test for $\boldsymbol{\alpha} \in \mathcal{N}$ does not require the solution of an optimization problem as only the nonnegativity of $V(1, \alpha^T)^T$ is tested. Hence the boundary approximation of $\mathcal{N}$ is computationally cheap. Then a second set

$$\mathcal{M}^* = \{\boldsymbol{\alpha} \in \mathbb{R}^2 : \min_W \|g(\boldsymbol{\alpha}, W)\|_2^2 = 0\} \tag{6.5}$$

is considered which works with the modification g of the f defined by (6.2) in a way that the nonnegativity constraint of the first column of Y (which is already tested for the membership to the set $\mathcal{N}$) is omitted. Hence in (6.2) the substitution

$$\sum_{i=1}^{n}\sum_{j=1}^{3} \min(0, \frac{Y_{ij}}{\|Y(:,i)\|_\infty} + \varepsilon)^2 \qquad \rightarrow \qquad \sum_{i=1}^{n}\sum_{j=2}^{3} \min(0, \frac{Y_{ij}}{\|Y(:,i)\|_\infty} + \varepsilon)^2$$

results in the target function g. Then the intersection of the boundary of $\mathcal{M}^*$ with the interior of the set $\mathcal{N}$ is computed by polygon inflation. Finally, the intersection $\mathcal{N} \cap \mathcal{M}^*$ equals $\mathcal{M}$.

7. An intersection property of the SFS. This section proves a further property of the SFS for $s \geq 2$, namely that rays which start at the origin intersect the SFS at most in a line segment [32]. The intersection may also be empty or degenerated to a single point which is then located on the boundary of the set FIRPOL (4.1). This line segment intersection property is the basis for an efficient numerical computation of the SFS for higher dimensions s.

LEMMA 7.1. *Let $X \in \mathbb{R}^{k\times s}$, $X \geq 0$ and*

$$M = \begin{pmatrix} \mu & m^T \\ 0 & I \end{pmatrix} \in \mathbb{R}^{s\times s} \tag{7.1}$$

with $\mu > 0$, $m \in \mathbb{R}^{s-1\times 1}$ and $m \leq 0$. Then $XM^{-1} \geq 0$. Further, nonnegativity of $Y \in \mathbb{R}^{n\times s}$ and $M(1,:)Y^T \geq 0$ imply that $MY^T \geq 0$.

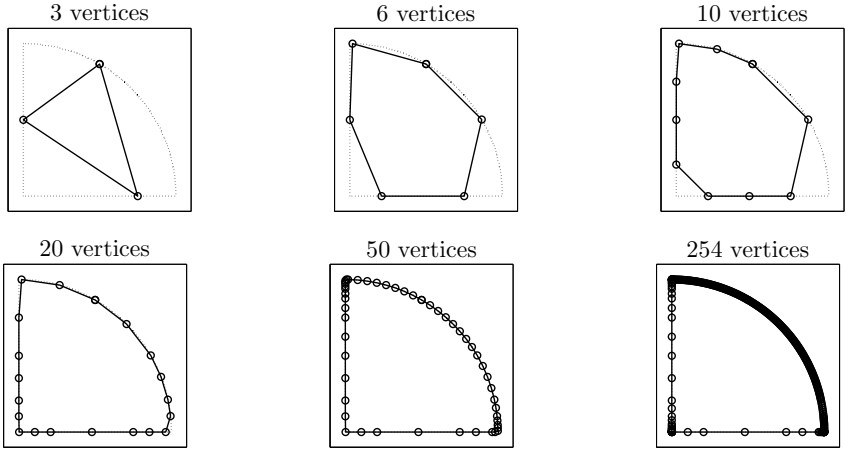

FIG. 6.1. *Approximation of the boundary of a quarter circle with the radius 1 by a series of adaptively refined polygons by the polygon inflation algorithm.*

Vertices of polygon	Hausdorff distance of polygon and quarter circle
3	$4.160 \cdot 10^{-1}$
6	$2.437 \cdot 10^{-1}$
10	$2.437 \cdot 10^{-1}$
20	$3.560 \cdot 10^{-2}$
50	$7.354 \cdot 10^{-3}$
150	$2.818 \cdot 10^{-4}$
254	$1.563 \cdot 10^{-4}$

TABLE 1

Approximation of a quarter circle with the radius 1 by polygon inflation. The Hausdorff distances of the boundaries of the approximating polygons and of the quarter circle are tabulated. The adaptive approximation scheme results in a fast boundary approximation in straight-line regions of the boundary.

Proof. M is invertible. Its inverse on the assumptions on μ and m satisfies

$$M^{-1} = \begin{pmatrix} 1/\mu & -m^T/\mu \\ 0 & I \end{pmatrix} \geq 0.$$

Thus $XM^{-1} \geq 0$. Moreover,

$$MY^T = \begin{pmatrix} M(1,:)Y^T \\ Y^T(2:s,:) \end{pmatrix} \geq 0$$

holds since $M(1,:)Y^T \geq 0$ and $Y \geq 0$. ∎

Lemma 7.1 and Corollary 3.3 are needed for the proof of the next theorem on the intersection of the SFS with semi-straight lines starting at the origin.

THEOREM 7.2. *Let $D \in \mathbb{R}^{k\times n}$ be a nonnegative matrix,* $\operatorname{rank}(D) = s$ *and let D^TD and DD^T be irreducible matrices. If $\boldsymbol{\alpha} \in \mathcal{M}$, then a maximal number ρ^* exists so that $\rho^*\boldsymbol{\alpha} \in \mathcal{M} \cap \partial\mathcal{N}$ where $\partial\mathcal{N}$ is the boundary of FIRPOL* (4.1)*. Further a minimal number $0 < \rho_*$ exists so that $\rho_*\boldsymbol{\alpha} \in \mathcal{M}$. The intersection of $\mathcal{M}$ with the semi-straight line $\omega\boldsymbol{\alpha}$, $\omega \geq 0$ is the (full) line segment $\{\rho\boldsymbol{\alpha} : \rho \in [\rho_*, \rho^*]\}$.*

Proof. By Theorem 4.2 the set $\mathcal{N}$ is bounded which together with $\mathcal{M} \subset \mathcal{N}$ proves the existence of a supremum ρ^*. By Theorem 4.3 it holds that $0 \notin \mathcal{M}$ so that $\rho_* > 0$. The

compactness of $\mathcal{M}$ due Thm. 4.8 with $\varepsilon = 0$ guarantees the existence of a maximum and a minimum.

Let $\boldsymbol{\alpha} \in \mathcal{M}$ be given. Without loss of generality we can assume that $\rho_* \boldsymbol{\alpha} = \boldsymbol{\alpha}$ (namely that $\boldsymbol{\alpha}$ is the point closest to the origin in $\boldsymbol{\alpha}\mathbb{R}_+ \cap \mathcal{M}$). Then according to (3.5) and (3.7) this $\boldsymbol{\alpha}$ is associated with

$$T_\rho := \begin{pmatrix} 1 & \rho\boldsymbol{\alpha}^T \\ e & W \end{pmatrix} \text{ with } \rho = 1 \text{ and } X = U\Sigma T_1^{-1} \geq 0, \quad Y = VT_1^T \geq 0.$$

Therein $D = U\Sigma V^T$ is a truncated SVD with $V(:,1) > 0$ and also $U(:,1) > 0$. It remains to show that $\widehat{X} = U\Sigma T_\rho^{-1} \geq 0$ and $\widehat{Y} = VT_\rho^T \geq 0$ for any $\rho \in (1,\, \rho^*]$. These matrices can be written as

$$\widehat{X} = \underbrace{(U\Sigma T^{-1})}_{X}\underbrace{(TT_\rho^{-1})}_{M_\rho^{-1}} \quad \text{and} \quad \widehat{Y}^T = \underbrace{(T_\rho T^{-1})}_{M_\rho}\underbrace{(TV^T)}_{Y^T}$$

with $M_\rho = T_\rho T^{-1}$. We show that M_ρ satisfies the assumptions of Lemma 7.1. The inverse of T with the Schur complement $\mathcal{S} = W - e\boldsymbol{\alpha}^T$ is given by

$$T^{-1} = \begin{pmatrix} 1 + \boldsymbol{\alpha}^T\mathcal{S}^{-1}e & -\boldsymbol{\alpha}^T\mathcal{S}^{-1} \\ -\mathcal{S}^{-1}e & \mathcal{S}^{-1} \end{pmatrix} \tag{7.2}$$

and $\mathcal{S}^{-1} = W^{-1} + (1/1 - \boldsymbol{\alpha}^T W^{-1}e)W^{-1}e\boldsymbol{\alpha}^T W^{-1}$ by the Sherman-Morrison formula. Corollary 3.3 applied to each column of $U\Sigma T^{-1} \geq 0$ guarantees that the first row of T^{-1} in (7.2) is strictly positive. Thus

$$\boldsymbol{\alpha}^T\mathcal{S}^{-1} < 0 \quad \text{and} \quad \boldsymbol{\alpha}^T\mathcal{S}^{-1}e < 0 \tag{7.3}$$

where $\boldsymbol{\alpha}^T\mathcal{S}^{-1}e$ is the sum of the components of $\boldsymbol{\alpha}^T\mathcal{S}^{-1}$. Hence $M_\rho = T_\rho T^{-1}$ reads

$$M_\rho = \begin{pmatrix} 1 - (\rho-1)\boldsymbol{\alpha}^T\mathcal{S}^{-1}e & (\rho-1)\boldsymbol{\alpha}^T\mathcal{S}^{-1} \\ 0 & I \end{pmatrix}.$$

Then (7.3) together with $\rho \geq 1$ prove that $m = (\rho-1)\boldsymbol{\alpha}^T\mathcal{S}^{-1} \leq 0$, which is the first premise of Lemma 7.1. Further, $\mu = 1 - (\rho-1)\boldsymbol{\alpha}^T\mathcal{S}^{-1}e \geq 1$ since $\boldsymbol{\alpha}^T\mathcal{S}^{-1}e < 0$ by (7.3) and $\rho \geq 1$. Thus $\widehat{X} = XM_\rho^{-1} \geq 0$ by Lemma 7.1.

Secondly, we prove that $\widehat{Y}^T = M_\rho Y^T \geq 0$. By Lemma 7.1 only $M_\rho(1,:)Y^T \geq 0$ is to be shown. It holds that

$$M_\rho(1,:)Y^T = y_1^T - (\rho-1)\underbrace{\left[\boldsymbol{\alpha}^T\mathcal{S}^{-1}ey_1^T - \boldsymbol{\alpha}^T\mathcal{S}^{-1}(Y([2:s],:))^T\right]}_{=:r^T}$$

with the first column y_1 of Y for $\rho = 1$. If $\rho = \rho^* = 1$, nothing remains to prove. Next the remaining case $1 < \rho < \rho*$ with $\rho^*\boldsymbol{\alpha} \in \partial\mathcal{N}$ is analyzed. It holds that

$$M_{\rho^*}(1,:)Y^T = y_1^T - (\rho^*-1)r^T \geq 0.$$

Multiplication of the last inequality with $(\rho-1)/(\rho^*-1) > 0$ results in

$$(\rho-1)/(\rho^*-1)y_1^T - (\rho-1)r^T \geq 0.$$

Since $\omega := 1 - ((\rho-1)/(\rho^*-1)) \geq 0$ for $1 \leq \rho \leq \rho^*$ an addition of $\omega y_1^T \geq 0$ to the left-hand side of the last inequality results in

$$y_1^T - (\rho-1)r^T \geq 0$$

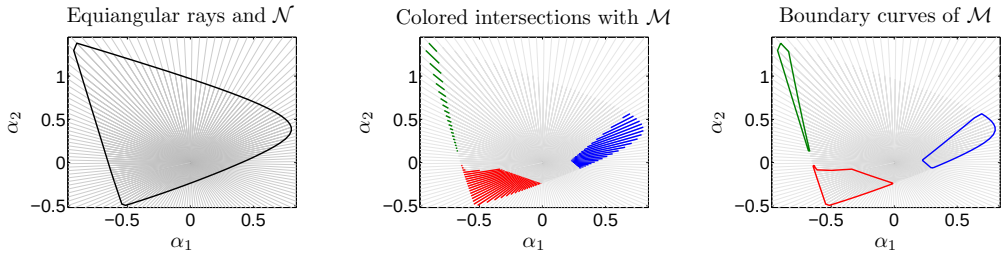

FIG. 7.1. *Construction of the two-dimensional SFS for the three-component model problem from Section 2. Left: Equiangular rays starting at the origin. The bold black line marks the boundary of $\mathcal{N}$. Center: The intersections of these rays with the SFS $\mathcal{M}$ are colored. Right: Boundary curves of the three subsets of the SFS.*

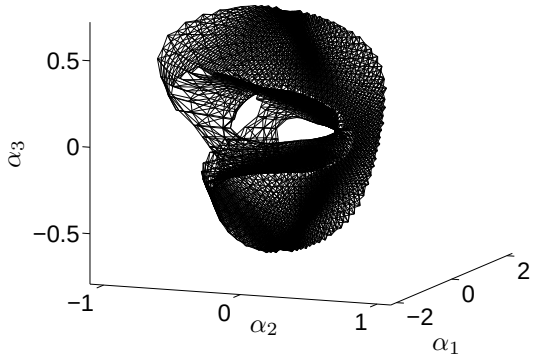

FIG. 7.2. *The plot shows a typical three-dimensional SFS. The surface triangle mesh of this topologically connected object with holes can be constructed from the set of the endpoints of the line segments $\boldsymbol{\alpha}\mathbb{R}_+ \cap \mathcal{M}$. Evenly distributed directions $\boldsymbol{\alpha}$ can be constructed by using spherical coordinates.*

which shows that $M_\rho(1,:)Y^T \geq 0$. Lemma 7.1 proves $\widehat{Y} \geq 0$. □

REMARK 7.3. *The proof of Theorem 7.2 is based on matrix theory. Alternatively, the proof can be based on the geometric construction of the SFS as introduced in 4.6. Then the intersection property of Theorem 7.2 is equivalent to the possibility to move one or more vertices of the polyhedron in $\mathcal{N}$ to the boundary $\partial\mathcal{N}$ without breaking the inclusion condition for the representatives of the rows of the matrix D.*

8. Conclusion. H. Minc in the foreword of his monograph "Nonnegative Matrices" [24] formulated the "aim of the book to provide the reader with a rigorous study of the Perron-Frobenius theory". He criticized other mathematical textbooks on nonnegative matrices for their inclusion of "all kinds of applications to as many cognate and unrelated fields as possible". Bearing these words in mind, we have isolated the analytic representation problem of all NMF solutions from a field of application and have imported it to the mathematical theory of nonnegative matrices.

The set of NMF solutions in its SFS representation has a rich mathematical structure and various of its properties are largely based on the Perron-Frobenius theory. In view of the field of application, a deeper mathematical understanding of these sets and also the development of efficient numerical algorithms for their computation can considerably support the structure elucidation in analytical chemistry. We hope that this contribution at the interface of mathematics and a challenging field of application in chemistry can inspire further analytic developments and open new SFS applications for NMF analyses.

REFERENCES

[1] H. Abdollahi and R. Tauler. Uniqueness and rotation ambiguities in Multivariate Curve Resolution methods. *Chemom. Intell. Lab. Syst.*, 108(2):100–111, 2011.

[2] R.B. Bapat and T.E.S. Raghavan. Nonnegative matrices and applications, volume 64 of Encyclopedia of Mathematics and its Applications, 1997.

[3] A. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik*, 162(5):78–88, 1852.

[4] A. Berman and R.J. Plemmons. Nonnegative matrices. *The Mathematical Sciences, Classics in Applied Mathematics*, 9, 1979.

[5] M.W. Berry, M. Browne, A.N. Langville, V.P. Pauca, and R.J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, 52:155–173, 2007.

[6] O.S. Borgen and B.R. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta*, 174:1–26, 1985.

[7] A. Cichocki, M. Mørup, P. Smaragdis, W. Wang, and R. Zdunek. Advances in nonnegative matrix and tensor factorization. *Computational intelligence and neuroscience: CIN*, 2008.

[8] A. Cichocki, R. Zdunek, A.H. Phan, and S. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[9] J. E. Cohen and U. G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.*, 190:149–168, 1993.

[10] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? Advances in Neural Information processes, NIPS 2003, Cambridge, 2003.

[11] P.J. Gemperline. Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Anal. Chem.*, 71(23):5398–5404, 1999.

[12] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, and R. Tauler. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data. *Anal. Chim. Acta*, 911:1–13, 2016.

[13] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of Rotational Ambiguity for Three-Component Systems. *Anal. Chem.*, 83(3):836–841, 2011.

[14] A. Golshan, M. Maeder, and H. Abdollahi. Determination and visualization of rotational ambiguity in four-component systems. *Anal. Chim. Acta*, 796(0):20–26, 2013.

[15] H. Kim and H. Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.*, 30:713–730, 2008.

[16] H. Kim and H. Park. Fast nonnegative matrix factorization: An active-set-like method and comparisons. *SIAM J. Sci. Comput.*, 33(6):3261–3281, 2011.

[17] H. Kim and H. Park. Algorithms for nonnegative natrix and tensor factorizations: A unified view based on block coordinate descent framework. *J. Global Optim.*, 58(2):285–319, 2014.

[18] H. Laurberg, M.G. Christensen, M.D. Plumbley, L.K. Hansen, and S.H. Jensen. Theorems on Positive Data: On the Uniqueness of NMF. *Computational Intelligence and Neuroscience*, 2008:9 pages, 2008.

[19] W.H. Lawton and E.A. Sylvestre. Self modelling curve resolution. *Technometrics*, 13:617–633, 1971.

[20] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[21] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13:556–562, 2001.

[22] M. Maeder and Y.M. Neuhold. *Practical data analysis in chemistry*. Elsevier, Amsterdam, 2007.

[23] E. Malinowski. *Factor analysis in chemistry*. Wiley, New York, 2002.

[24] H. Minc. *Nonnegative matrices*. John Wiley & Sons, New York, 1988.

[25] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994.

[26] V. P. Pauca, J. Piper, and R.J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra Appl.*, 416(1):29–47, 2006.

[27] R. Rajkó and K. István. Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. *J. Chemom.*, 19(8):448–463, 2005.

[28] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. *On the analysis and computation of the area of feasible solutions for two-, three- and four-component systems*, volume 30 of Data Handling in Science and Technology, "Resolving Spectral Mixtures", Ed. C. Ruckebusch, chapter 5, pages 135–184. Elsevier, Cambridge, 2016.

[29] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. Simultaneous construction of dual Borgen plots. I: The case of noise-free data. Technical report, To appear in J. Chemom., DOI: 10.1002/cem.2954., 2017.

[30] M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr. A fast polygon inflation algorithm to

compute the area of feasible solutions for three-component systems. I: Concepts and applications. *J. Chemom.*, 27:106–116, 2013.

[31] M. Sawall and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. II: Theoretical foundation, inverse polygon inflation, and FAC-PACK implementation. *J. Chemom.*, 28:633–644, 2014.

[32] M. Sawall and K. Neymeyr. A ray casting method for the computation of the area of feasible solutions for multicomponent systems: Theory, applications and FACPACK-implementation. *Anal. Chim. Acta*, 960:40–52, 2017.

[33] R. Tauler. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemom.*, 15(8):627–646, 2001.

[34] L. B. Thomas. Rank factorization of nonnegative matrices (A. Berman). *SIAM Review*, 16(3):393–394, 1974.

[35] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, and M. Maeder. On rotational ambiguity in model-free analyses of multivariate data. *J. Chemom.*, 20(6-7):302–310, 2006.