Data Min Knowl Disc (2012) 24:325–354
DOI 10.1007/s10618-011-0216-z

Descriptive matrix factorization for sustainability Adopting the principle of opposites

Christian Thurau · Kristian Kersting · Mirwaes Wahabzada · Christian Bauckhage

Received: 5 June 2010 / Accepted: 5 February 2011 / Published online: 2 March 2011

Abstract Climate change, the global energy footprint, and strategies for sustainable development have become topics of considerable political and public interest. The public debate is informed by an exponentially growing amount of data and there are diverse partisan interest when it comes to interpretation. We therefore believe that data analysis methods are called for that provide results which are intuitively understandable even to non-experts. Moreover, such methods should be efficient so that non-experts users can perform their own analysis at low expense in order to understand the effects of different parameters and influential factors. In this paper, we discuss a new technique for factorizing data matrices that meets both these requirements. The basic idea is to represent a set of data by means of convex combinations of extreme data points. This often accommodates human cognition. In contrast to established factorization methods, the approach presented in this paper can also determine over-complete bases. At the same time, convex combinations allow for highly efficient matrix factorization. Based on techniques adopted from the field of distance geometry, we derive a linear time algorithm to determine suitable basis vectors for factorization. By means of the example of several environmental and developmental data sets we discuss the

Responsible editor: Katharina Morik, Kanishka Bhaduri and Hillol Kargupta.

C. Thurau (✉) · K. Kersting · M. Wahabzada · C. Bauckhage
Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, Germany
e-mail: christian.thurau@iais.fraunhofer.de

K. Kersting
e-mail: kristian.kersting@iais.fraunhofer.de

M. Wahabzada
e-mail: mirwaes.wahabzada@iais.fraunhofer.de

C. Bauckhage
e-mail: christian.bauckhage@iais.fraunhofer.de

performance and characteristics of the proposed approach and validate that significant efficiency gains are obtainable without performance decreases compared to existing convexity constrained approaches.

Keywords Matrix factorization · Convex combinations · Distance geometry · Large-scale data analysis

1 Introduction

Questions as to the sustainability of economic growth, energy consumption, and agricultural production have become the focal point of worldwide debate. Issues like globalization, climate change, and economic turmoil add heat to the public discussion and exegesis of the available scientific data differs considerably among different interest groups.

Following David MacKay's arguments (MacKay 2009), we believe that what the debate on sustainability needs is less emotions but a better understanding of data. This immediately begs the question of how the massive amounts of data that accumulate each day can be made widely understandable. Charts and diagrams may speak a clear language to the experts, but there is too much at stake to exclude the public from the discussion. Given these premises, we believe that data analysis methods are called that fulfill the following requirements:

(R1) They provide intuitively understandable results that are accessible even to non-experts.

(R2) They scale to massive data sets and thus allow for analysis at low expenses.

In this paper, we discuss a new technique for statistical data analysis that rises to these challenges.

Specifically, we address the problem of latent components analysis for very large data collections. For this purpose, matrix factorization algorithms have proven to be a viable tool. Similar to earlier work by Cutler and Breiman (1994) and Ding et al. (2010), our basic idea is to search for certain extremal elements in a set of data and to represent the data by means of convex combinations of these *archetypes*. The benefits of this archetype analysis are twofold:

- It accommodates human cognition, since memorable insights and experiences typically occur in form of extremes rather than as averages. Philosophers and Psychologists have noted this for long, since explanations of the world in terms of archetypes date back to Plato. According to C.G. Jung, it is the opposition that creates imagination. Every wish immediately suggests its opposite and in order to have a concept of good, there must be a concept of bad, just as there cannot be an idea of up without a concept of down. This principle of opposites is best summarized by Hegel's statement that "everything carries with it its own negation". The only way we can know anything is by contrast with an opposite. By focusing on extreme opposites, we simply enlarge the margin of what we know and in turn our chance to separate things.
- As we will show below, convex combinations allow for highly efficient data processing. One the key contributions of the present paper is to adopt techniques

from the field of distance geometry (Blumenthal 1953) to derive a linear time algorithm that determines suitable basis vectors for convexity constrained matrix factorization. In contrast to other factorization methods, this new approach can also determine over-complete bases where the number of basis elements exceeds the dimensionality of the embedding space. Over-complete representations have been advocated because they are more robust in the presence of noise, can be sparser, and may better capture internal structures of the data.

The novel algorithm to determine convexity constrained latent components we present is called *Simplex Volume Maximization*. Simplex Volume Maximization runs in linear time and is—to the best of our knowledge—the fastest algorithm to date for solving the task at hand. Further, it is the first algorithm in this area that does not require subsampling in order to handle gigantic matrices. With respect to common error measures such as the Frobenius norm, it shows a similar or even better performance than related methods. Moreover, as the algorithm only relies on iterative distance computations, it inherently allows for parallelization and is well suited for massive data analysis application. By means of the example of several environmental and developmental data sets we discuss the performance and characteristics of the proposed approach. Our exhaustive experimental results show that significant efficiency gains are obtainable without performance decreases compared to existing convexity constrained matrix factorization approaches.

We proceed as follows. After touching upon related work, we provide several motivating examples for our work in Sect. 3. Then, after introducing our notation in Sect. 4 and formalizing the problem in Sect. 5, we devise a new algorithm for simplex volume maximization in Sects. 6 and 7. Before concluding, we present the results of our experimental evaluation. All major proofs can be found in the Appendix.

2 Related work

Understanding data by unmixing its latent components is a standard technique in data mining and pattern recognition. Latent factor models and component analysis have a particularly venerable tradition in psychology and sociology (Spearman 1904; Hotelling 1933) but are also commonly applied in disciplines such as physics (Aguilar et al. 1998; Chan et al. 2003), economics (Lucas et al. 2003), or geology (Chang et al. 2006; Miao and Qi 2007; Nascimento and Dias 2005).

The main idea is to acquire a descriptive representation by explaining a set of data as a linear combination of certain important latent components. In this paper, we consider representations that can be formulated as a matrix decomposition problem where a data matrix is approximated by the product of a factor matrix and a matrix of mixing coefficients. Given this basic setting, different methods for discovering latent components differ by the constraints they impose on the two matrix factors. For example, the k-means clustering algorithm can be understood as a matrix factorization problem where the data is supposed to be explicable by means of a single centroid or basis vector per data sample which corresponds to a unary constraint on the coefficient matrix. *Non-negative matrix factorization (NMF)* as popularized by Lee and Seung (1999) explains the data by means of a non-negative combination of non-negative basis vectors.

In this paper, we focus on a type of constraint that restricts the representation to *convex combinations* of latent components. Previous contributions that incorporate this kind of constraint include *Archetypal Analysis (AA)* according to Cutler and Breiman (1994), *Convex-NMF (C-NMF)* as introduced by Ding et al. (2010), *Convex-hull NMF (CH-NMF)* proposed by Thurau et al. (2009), and *Hierarchical Convex-hull NMF (HCH-NMF)* presented by Kersting et al. (2010).

Convexity constraints result in latent components that show interesting properties: First, the basis vectors are included in the data set and often reside on actual data points. Second, convexity constrained basis vectors usually correspond to the most extreme data points rather than to the most average ones. Both these properties typically cause the basis elements to be readily interpretable even to non-experts. The usefulness of convexity constrained latent component detection has been noted in various fields. For example, in geoscience it is also referred to as *Endmember Detection* and is used in the analysis of spectral images. The most commonly applied algorithm in this community is N-FINDR (Winter 1999). In economics, the latent components are often called archetypes in reference to the Archetypal Analysis algorithm due to Cutler and Breiman (1994). Unfortunately, both these algorithms are examples of brute force techniques that scale quadratically with the number of data and are thus too demanding w.r.t. computation time to allow for the processing of more than a few hundred samples.

Owing to the exponential increase of available data, techniques based on matrix factorization are now being tailored to the analysis of massive data sets. Recently, we introduced a new algorithm that is applicable to very large data sets (Thurau et al. 2009, 2010). By means of either informed or random projections of data into 2D subspaces, we perform an efficient subsampling of possible archetypal candidates. On the subsampled data, conventional AA or C-NMF can be run in reasonable time. A disadvantage of this approach lies in its sensitivity to high dimensional data. A dense subsampling based on many projections would be again too demanding, while a coarser subsampling might sacrifice possibly good basis vector candidates. Other recently proposed techniques for efficient matrix factorization suffer from similar problems: Using random matrices to compute low rank approximations based on theorems by Achlioptas and McSherry (2007) leads to basis vectors that lack interpretability as do the basis elements that result from the sampling based *CUR decomposition* introduced by Drineas et al. (2006).

3 Basic idea via motivating examples

The key technical problem we solve in this paper is finding latent components in (massive) data sets that are easy to interpret. We formulate the problem as a constrained matrix factorization problem aiming at minimizing the Frobenius norm between a data matrix and its approximation. Based on two theorems, which we will derive from principles of distance geometry, we then show that for convexity constrained factorizations minimizing the Frobenius norm is equivalent to maximizing the volume of a simplex whose vertices correspond to basis vectors. This basic idea is illustrated in Fig. 1. Because convexity constrained basis vectors usually correspond to the most extreme

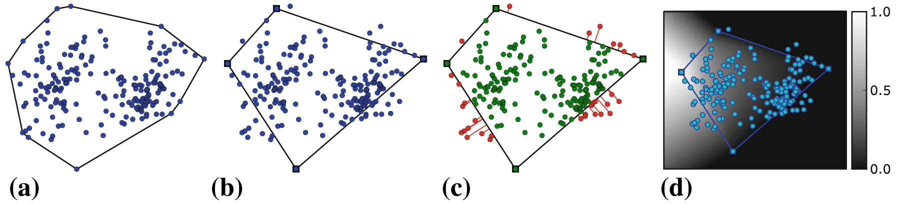

Fig. 1 Illustration of the approach proposed in this paper. Given a data matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, we determine a basis of extreme points $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$, $k \ll n$, to represent the data using convex combinations $\mathbf{V} \approx \mathbf{WH}$. The coefficients in $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$ result from solving constrained quadratic programs, its columns are non-negative and sum to one. This perfectly reconstructs those points of $\mathbf{V}$ inside the hull of $\mathbf{W}$; points on the outside are projected to the nearest point on the hull (best viewed in color). (**a**) data $\mathbf{V}$, (**b**) basis $\mathbf{W} = \mathbf{VG}$, (**c**) $\mathbf{V} \approx \mathbf{WG}$, (**d**) $\mathbf{pv}|\mathbf{W}_1$. (Color figure online)

data points and not to the most average ones, this increases interpretability. The rest of this section illustrates this increased interpretability on several environmental data sets. For a brief description of the used factorization methods as well as of the data sets, we refer to the following sections, in particular to the experimental evaluation in Sect. 8.

Computational sustainability problems arise in different domains, ranging from wildlife preservation and biodiversity, to balancing socio-economic needs and the environment, to large-scale deployment and management of renewable energy sources. Consider a rather classical domain: climate. As the climate system is characterized by highly complex interactions between a large number of physical variables, it is a challenging task for researchers to break up the complicated structures into a few significant modes of variability. Matrix factorization methods can be used to compute automatically a reduced-dimensional representation of large-scale non-negative data and, in turn, to extract underlying features. Consider for example Fig. 2. It shows the basis vectors (kind of cluster representatives) found by several non-negative matrix factorization[1] methods applied to the Historical Climatography Series (HCS) temperature normals for the period between January 1961 and December 2000 for U.S. States.[2] The data consists of monthly averaged temperatures for all U.S. States. Although classical NMF finds meaningful clusters (it seems to group together U.S. States with (a) hot summers and (b) cold winters), the clusters are difficult to understand without additional knowledge about the states: They do not correspond to actually observed data. In contrast, the results of the constraint NMF versions are readily interpretable even to non-expert users: they represent the data as convex combinations of extreme data points. That is, the maps in Fig. 2e–p show actual average temperatures for U.S. States. By assigning the corresponding month to each basis vector, we can easily verify that the basis vectors essentially span the four seasons: (e) cold winter, (f) spring, (g) hot summer, and (h) warm fall. Moreover, they characterize the variations among individual samples. For instance, the maps indicate that Maine "experiences a humid

[1] For a more detailed description of the techniques we refer to the subsequent sections. In particular, for "Robust" CH-NMF we refer to Sect. 8.

[2] Available as HCS 4-1 from http://cdo.ncdc.noaa.gov/cgi-bin/climatenormals/climatenormals.pl?directive=prod_select2&prodtype=HCS4&subrnum=.

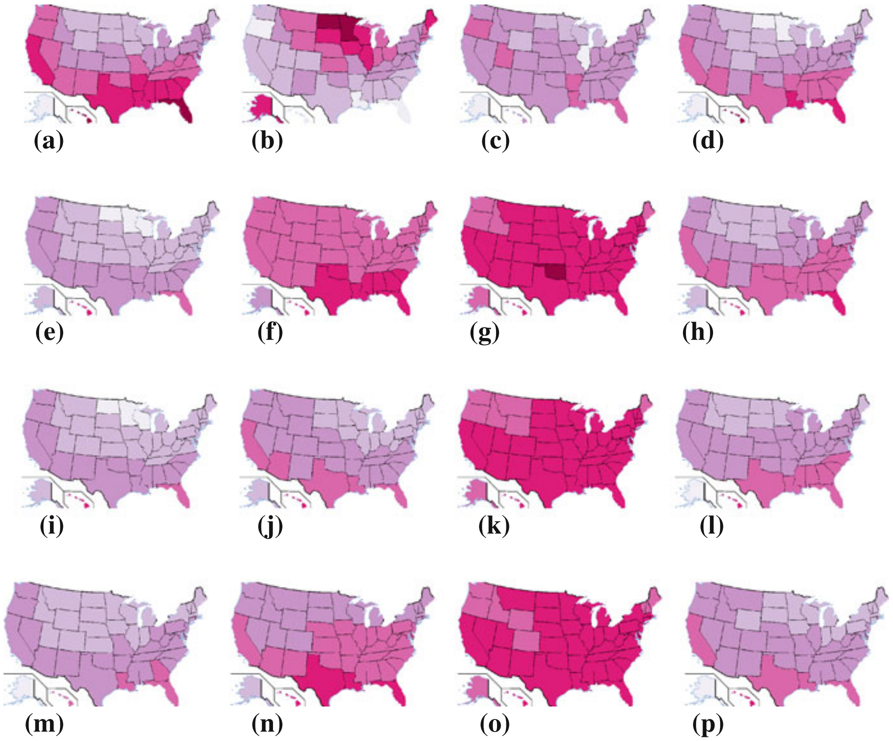

Fig. 2 Resulting basis vectors (BV) for NMF (**a**)–(**d**), SiVM (**e**)–(**h**) as proposed in this paper, CH-NMF (**i**)–(**l**), and robust CH-NMF (**m**)–(**p**) on the Historical Climatography Series (HCS) temperature normals for the period between January 1961 and December 2000 for U.S. States. The data consists of monthly averaged temperatures for all U.S. States. Triggered by the four seasons, we computed factorizations using 4 basis vectors. The *colors* vary smoothly from *lilac* (low values) to *magenta* (high values). Because SiVM and (robust) CH-NMF are constructed from actual data points, they are more easily interpretable by individuals than NMF. We could easily assign the month and years of basis vectors found. One can also see that robust CH-NMF covers the year more uniformly. In contrast, the NMF results are very different and require additional interpretation efforts: the values (*colors*) do not correspond to actual temperatures. It seems to group together U.S. States with hot summers (**a**) and cold winters (**b**) (best viewed in color). (**a**) NMF BV 1, (**b**) NMF BV 2, (**c**) NMF BV 3, (**d**) NMF BV 4, (**e**) SiVM Jan. '77, (**f**) SiVM May. '94, (**g**) SiVM Jul. '80, (**h**) SiVM Nov. '85, (**i**) CH-NMF Jan. '77, (**j**) CH-NMF Feb. '63, (**k**) CH-NMF Jul. '87, (**l**) CH-NMF Dec. '94, (**m**) Robust Jan '91, (**n**) Robust Apr '64, (**o**) Robust Aug '96, (**p**) Robust Dec. '62. (Color figure online)

continental climate, with warm (although generally not hot), humid summers. Winters are cold".[3]

As the next example, let us consider precipitation data. The basis vectors of several non-negative matrix factorization methods applied to the HCS precipitation normals for the period between January 1961 and December 2000 for U.S. States[4] are shown in Fig. 2. The data consists of monthly averaged precipitations for all U.S. States. Here,

[3] Wikipedia, read on June 4, 2010; http://en.wikipedia.org/wiki/Maine#Climate.

[4] Available as HCS 4-2 from http://cdo.ncdc.noaa.gov/cgi-bin/climatenormals/climatenormals.pl?directive=prod_select2&prodtype=HCS4&subrnum=.

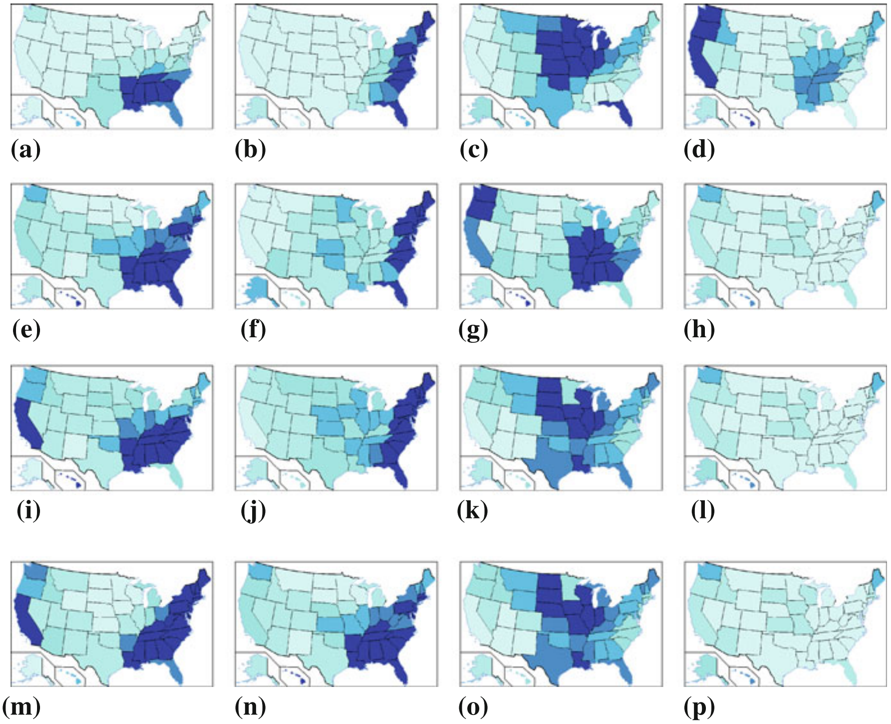

Fig. 3 Resulting basis vectors (BV) for NMF (**a**)–(**d**), SiVM (**e**)–(**h**) as proposed in this paper, CH-NMF (**i**)–(**l**), and robust CH-NMF (**m**)–(**p**) on the Historical Climatography Series (HCS) precipitation normals for the period between January 1961 and December 2000 for U.S. States. The data consists of monthly averaged precipitations for all U.S. States. Triggered by the four seasons, we computed factorizations using 4 basis vectors. The *colors* vary smoothly from *cyan* (low values) to *dark blue* (high values). The basis vectors of SiVM and CH-NMF are actual data points so that they are more easily interpretable by individuals than NMF: the values (*colors*) are actual precipitations. We could easily assign the month and years of basis vectors found. The robust variant again distributes the basis vectors more uniformly across the year. In contrast to the temperatures shown in Fig 2, NMF finds basis vectors similar to those from the other methods. Note, however, that the values (*colors*) do not correspond to actual precipitations; they have been re-scaled for the sake of visualization. More interestingly, NMF does not capture the variability of the data as well as convex counterparts: months with little precipitation across the country (**h, l, p**) are missing (best viewed in color). (**a**) NMF BV 1, (**b**) NMF BV 2, (**c**) NMF BV 3, (**d**) NMF BV 4, (**e**) SiVM Mar. '80, (**f**) SiVM Sept. '99, (**g**) SiVM Dec. '82, (**h**) SiVM Oct. '63, (**i**) CH-NMF Mar. '75, (**j**) CH-NMF Jun. '72, (**k**) CH-NMF Jun. '93, (**l**) CH-NMF Oct. '63, (**m**) Robust Jan '78, (**n**) Robust Mar. '80, (**o**) Robust Jun. '93, (**p**) Robust Oct. '63. (Color figure online)

the methods find similar results. However, we note that the NMF basis vectors do not correspond to actual data points and are not indicative of characteristic variations among individual samples. In contrast, Fig. 3g, m–p indicate that rain is common for Washington State in winter. This at least matches well Wikipedia's description of Seattle, Washington States's capital.[5] Nevada receives rather scarce precipitation

[5] Read on June 4, 2010; http://en.wikipedia.org/wiki/Seattle#Climate.

during the year.[6] In general, we can easily see that October can be a month with little precipitation across the country (h,l,p). It is difficult—if not impossible — to read all this off the NMF basis vectors without additional knowledge.

Finally, consider another classical domain: energy consumption. The International Energy Outlook[7] 2010 by the U.S. Energy Information Administration (EIA) expects that the "world marketed energy consumption increases by 49% from 2007 to 2035 in the Reference case. Total energy demand in the non-OECD countries increases by 84%, compared with an increase of 14% in the OECD countries." Rising energy consumption, however, grows the concern about global warming. Similar to climate data, energy consumption is characterized by highly complex interactions between a large number of variables, so that again it is a challenging task to break up the complicated structures into a few significant modes of variability. Figure 4 shows abundance maps for the basis vectors found by several non-negative matrix factorization approaches applied to the yearly total electricity consumption data for the world's countries in the period of 1980 till 2008 as reported by the EIA.[8] Here, an abundance map essentially shows how well a country's electricity consumption is explained by a single basis vector. As one can see, it is again difficult—if not impossible without additional efforts—to extract any meaningful insides from NMF's abundance maps (a,d,g,j). They simply represent the countries with the highest electricity consumption: USA, China, and Russia. In contrast, the convexity constraint allows us to associate a country with each basis vector. In turn, other countries' energy consumption can be explained in terms of the associated one. For instance, robust CH-NMF nicely groups the world's countries into (c) industrialized countries with high electricity consumption represented as Spain, (f) desert-like countries represented by Taiwan , (i) energy efficient Scandinavian-like countries represented by Poland, and (l) low energy consumption countries such as the African ones or Greenland represented by the Solomon Islands.

These simple examples expose requirement **R1** from the introduction: *data analysis methods should provide intuitively understandable results that are accessible even to non-experts*. As already envisioned by such eighteenth-century philosophers as Jean Jacques Rousseau, John Locke, and John Stuart Mill, government requires that everyone have the right to influence political and environmental decisions that affect them. A basic assumption is that everyone is—or should be—essentially equal, in both their concern for environmental issues and their competency to make decisions about them. However, in order to make these decisions, (at least informed) individuals need accurate and understandable models. In this sense, classical NMF might not always be the best choice for a problem at hand.

Moreover, computational sustainability problems may involve massive data sets. As an example that we will also investigate in our experimental evaluation consider hyperspectral images. They are often used in fields such as oceanography, environmental science, snow hydrology, geology, volcanology, soil and land management,

[6] Wikipedia, read on June 4, 2010; http://en.wikipedia.org/wiki/Nevada#climate.

[7] See http://www.eia.doe.gov/oiaf/ieo/highlights.html.

[8] Available from http://tonto.eia.doe.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=2&pid=2&aid=2.

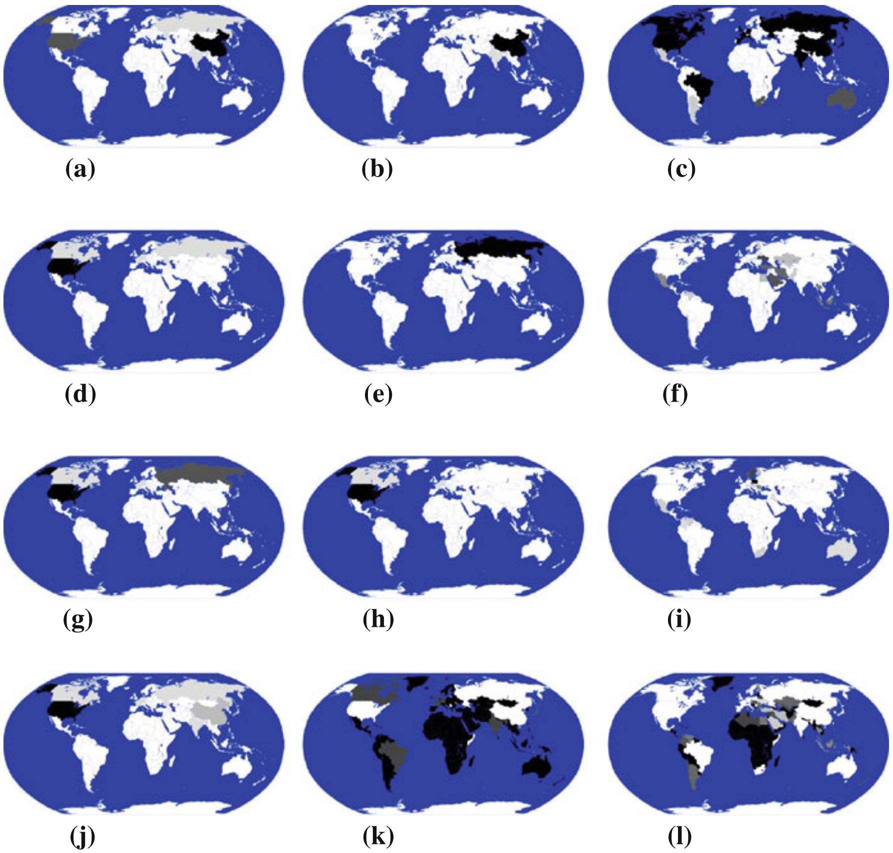

Fig. 4 Abundance maps for the basis vectors found by NMF (**a, d, g, j**), SiVM (**b, e, h, k**) as proposed in this paper, and robust CH-NMF (**c, f, i, l**) on the yearly total electricity consumption data for the world's countries in the period of 1980 till 2008 as reported by the EIA. The maps show how well each country is explained by a single basis vector; The shades vary smoothly from *white* (low fit) to *black* (high fit). The data consists of yearly electricity consumption (billion kilowatt-hours) for all countries of the world. We computed factorizations using 4 basis vectors. The basis vectors of SiVM are actual data points so that we can easily identify the corresponding countries: China (CHN), Russian (RUS) , USA (USA), and Saint Helena (SHN). For the robust variant, they are: Solomon Islands (SLB), Taiwan (TWN), Poland (POL), and Spain (ESP). NMF does not feature this (best viewed in color). (**a**) NMF BV 1, (**b**) SiVM CHN, (**c**) Robust ESP, (**d**) NMF BV 2, (**e**) SiVM RUS, (**f**) Robust TWN, (**g**) NMV BV 3, (**h**) SiVM USA, (**i**) Robust POL, (**j**) NMFBV 4, (**k**) SiVM SHN, (**l**) Robust SLB. (Color figure online)

atmospheric and aerosol studies, agriculture, and limnology to identify materials that make up a scanned area. Hyperspectral images easily produce gigantic matrices within hundreds of millions of entries. Thus, another requirement for "sustainable" matrix factorization is **R2** as already mentioned in the introduction: *data analysis methods should scale to massive data sets and thus allow for analysis at low expenses*.

In the following, we will present a novel matrix factorization method that meets **R1** and **R2**.

4 Notation and definitions

Throughout this paper, we denote vectors using bold lower case letters ($\mathbf{v}$); subscripted lower case italics (v_k) refer to the components of a vector. $\mathbf{0}$ is the vector of all zeros and $\mathbf{1}$ is the vector of all ones. We write $\mathbf{v} \succeq \mathbf{0}$ to indicate that $v_k \geq 0$ for all k. The inner product of two vectors $\mathbf{u}$ and $\mathbf{v}$ is written as $\mathbf{u}^T\mathbf{v}$. Consequently, the expression $\mathbf{1}^T\mathbf{v}$ is a shorthand for $\sum_k v_k$ and $\mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|^2$ is the squared Euclidean norm of $\mathbf{v}$.

Matrices are written using bold upper case letters ($\mathbf{M}$) and subscripted upper case italics (M_{ij}) denote individual matrix entries. In order to indicate that $\mathbf{M}$ is a real-valued $d \times n$ matrix, i.e. $\mathbf{M} \in \mathbb{R}^{d\times n}$, we may use the shorthand $\mathbf{M}^{d\times n}$. If the columns of a matrix are known, we also write $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n]$ where $\mathbf{m}_j \in \mathbb{R}^d$ is the jth column vector of $\mathbf{M}$. Finally, $\|\mathbf{M}\|^2 = \sum_{i,j} M_{ij}^2$ is the squared Frobenius norm of $\mathbf{M}$.

A vector $\mathbf{v} \in \mathbb{R}^d$ is a *convex combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_l \in \mathbb{R}^d$, if $\mathbf{v} = \sum_i \lambda_i \mathbf{v}_i$ where $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Using matrix notation, we write $\mathbf{v} = \mathbf{V}\boldsymbol{\lambda}$ where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_l]$ and $\boldsymbol{\lambda} \in \mathbb{R}^l$ such that $\mathbf{1}^T\boldsymbol{\lambda} = 1$ and $\boldsymbol{\lambda} \succeq \mathbf{0}$.

The *convex hull* C of a set $S \subset \mathbb{R}^d$ is the set of all convex combinations of points in S. An *extreme point* of a convex set C is any point $\mathbf{v} \in C$ that is not a convex combination of other points in C, i.e. if $\mathbf{v} = \lambda\mathbf{u} + (1-\lambda)\mathbf{w}$ for $\mathbf{u}, \mathbf{w} \in C$ and $\lambda \in [0, 1]$ then $\mathbf{v} = \mathbf{u} = \mathbf{w}$.

A *polytope* is the convex hull of finitely many points, i.e. it is the set $C(S)$ for $|S| < \infty$. The extreme points of a polytope are also called *vertices*. We use $V(S)$ to denote the set of all vertices of a polytope. Note that every point inside a polytope can be expressed as a convex combination of the points in V.

5 Problem formulation: constrained matrix factorization

We consider a data representation problem where the data $\mathbf{V} \in \mathbb{R}^{d\times n}$ is expressed by means of convex combinations of certain points in $\mathbf{V}$. The underlying problem can be formulated as a matrix factorization of the form

$$\mathbf{V} \approx \mathbf{VGH} \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{n\times k}, \mathbf{H} \in \mathbb{R}^{k\times n}$ are coefficient matrices such that $\mathbf{H}$ is restricted to convexity and $\mathbf{G}$ is restricted to unary column vectors, i.e.,

$$\begin{aligned} \mathbf{1}^T\mathbf{h}_j &= 1,\ \mathbf{h}_j \succeq \mathbf{0} \\ \mathbf{1}^T\mathbf{g}_i &= 1,\ \mathbf{g}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T. \end{aligned}$$

In other words, the factorization (1) approximates $\mathbf{V}$ using convex combinations where the basis vectors $\mathbf{W} = \mathbf{VG}$ are data points selected from $\mathbf{V}$. The goal now is to determine a basis that minimizes the Frobenius norm

$$E = \|\mathbf{V} - \mathbf{VGH}\|^2 = \|\mathbf{V} - \mathbf{WH}\|^2. \tag{2}$$

When minimizing (2), we have to simultaneously optimize $\mathbf{W}$ and $\mathbf{H}$ which is generally considered a difficult problem known to suffer from many local minima. Archetypal Analysis (AA) and Convex-NMF (C-NMF) are well understood examples of approaches that attempt a simultaneous optimization. AA as introduced by Cutler and Breiman (1994) applies an alternating least squares procedure where each iteration requires the solution of a constrained quadratic optimization problems of the order of $n \times n$ where n is the size of the data set. It should be noted that it solves the arguably more difficult case where the matrix $\mathbf{G}$ is also restricted to convexity instead of unarity. However, Cutler and Breiman report that the resulting $\mathbf{G}$ typically consists of unary columns anyway. C-NMF according to Ding et al. (2010) uses iterative update rules which require the computation of intermediate matrices of size $n \times n$. Both approaches do not scale to gigantic data matrices.

In order to avoid problems due to the simultaneous estimation of $\mathbf{W}$ and $\mathbf{H}$, other approaches attempt to determine suitable matrices $\mathbf{W}$ and $\mathbf{H}$ in a successive manner. Once $\mathbf{W}$ has been estimated, it is straightforward to determine $\mathbf{H}$. In fact, the coefficient vectors $\mathbf{h}_j$ can then be computed in parallel for all $\mathbf{v}_j \in \mathbf{V}$. For our problem of determining suitable basis vectors for convex combinations such a successive scheme is indeed well justifiable.

Cutler and Breiman (1994) prove that optimal basis vectors for a convex factorization of the data reside on the data convex hull. In other words, under the constraint $\mathbf{W} = \mathbf{VG}$, an optimal choice of the basis vectors will correspond to a subset of the vertices of $C(\mathbf{V})$. This has already been exploited in Convex-hull NMF as introduced in Thurau et al. (2009) as well as in various methods related to endmember detection for hyperspectral imaging (Nascimento and Dias 2005; Chang et al. 2006; Miao and Qi 2007). Nevertheless, estimating $\mathbf{W}$ remains a difficult problem. On the one hand, computing the vertices $V(\mathbf{V})$ of the convex hull of many (high-dimensional) data points $\mathbf{V}$ is itself a demanding problem. On the other hand, it is not immediately evident which points to select from $V(\mathbf{V})$.

Our contribution in this paper is a novel, highly efficient algorithm for estimating $\mathbf{W} = \mathbf{VG}$. It is based on the observation that, if $\mathbf{v}_j$ is expressed as a convex combination $\mathbf{v}_j = \mathbf{W}\mathbf{h}_j$, the coefficient vectors $\mathbf{h}_j$ reside in a $(k-1)$-simplex whose k vertices correspond to the basis vectors in $\mathbf{W}^{d \times k}$. Because of this duality, we may use the terms polytope and simplex interchangeably in the following.

6 Simplex volume

If we assume that the basis vectors $\mathbf{W}^{d \times k}$ for a convex combination are selected from actual data samples $\mathbf{v}_j \in \mathbf{V}$, we can proof the following Theorem; the proof is given in the Appendix.

Theorem 1 *Extending a given simplex $\mathbf{W}^{d \times k}$ by adding a vertex $\mathbf{w}_{k+1}$ sampled from a data matrix $\mathbf{V}^{d \times n}$ will not increase the Frobenius norm of the optimal convex approximation of the data. That is*

$$\left\| \mathbf{V}^{d \times n} - \mathbf{W}^{d \times (k+1)} \mathbf{H}^{(k+1) \times n} \right\|^2 \leq \left\| \mathbf{V}^{d \times n} - \mathbf{W}^{d \times k} \mathbf{H}^{k \times n} \right\|^2$$

if $\mathbf{H}^{k \times n}$ *and* $\mathbf{H}^{(k+1) \times n}$ *are convexity constrained coefficient matrices that result from solving constrained quadratic optimization problems.*

Theorem 1 hints at the idea of volume maximization for matrix factorization. Any increase of the volume of the k-simplex encoded in $\mathbf{W}$ will reduce the overall residual of the reconstruction. But why should maximizing the simplex volume be advantageous over minimizing the Frobenius norm? The answer is computational efficiency. Next, we derive a highly efficient volume maximization algorithm that determines a suitable basis $\mathbf{W}$ for convex reconstruction of a set of data. It is rooted in the notion of distance geometry.

Distance geometry studies sets of points based only on the distances between pairs of points. It plays an important role, for instance, in three dimensional molecular modeling from connectivity data in chemistry, or various applications in geography and physics (Crippen 1988; Sippl and Sheraga 1986). In the following, we will denote the distance between two points $\mathbf{v}_i$ and $\mathbf{v}_j$ as $d_{i,j}$.

Distance geometry draws heavily on the notion of the *Cayley-Menger determinant* (CMD) (Blumenthal 1953) which indicates the volume of a polytope or simplex. Given the lengths $d_{i,j}$ of the edges between the $k+1$ vertices of an k-simplex S, its volume is given by

$$\mathrm{Vol}(S)_k^2 = \frac{-1^{k+1}}{2^k (k!)^2} \det(\mathbf{A}) \tag{3}$$

where

$$\det(\mathbf{A}) = \begin{vmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{1,1}^2 & d_{1,2}^2 & \dots & d_{1,k+1}^2 \\ 1 & d_{1,1}^2 & 0 & d_{2,2}^2 & \dots & d_{2,k+1}^2 \\ 1 & d_{1,2}^2 & d_{2,2}^2 & 0 & \dots & d_{3,k+1}^2 \\ \vdots & & & & \ddots & \vdots \\ 1 & d_{1,k+1}^2 & d_{2,k+1}^2 & d_{3,k+1}^2 & \dots & 0 \end{vmatrix} \tag{4}$$

is the Cayley-Menger determinant.

With respect to data analysis, our goal is to select vertices $\{\mathbf{w}_1, \dots, \mathbf{w}_k\} \in \mathbf{V}$ such that they maximize the volume of the corresponding simplex. If a number of vertices has already been acquired in a sequential manner, we can prove the following Theorem; the proof is given in the Appendix.

Theorem 2 *Let* S *be a* k*-simplex. Suppose that the vertices* $\mathbf{w}_1, \dots, \mathbf{w}_k$ *are equidistant and that this distance is* a*. Also, suppose that the distances between vertex* $\mathbf{w}_{k+1}$ *and the other vertices are given by* $\{d_{1,k+1}, \dots, d_{n,k+1}\}$*, then the volume of* S *is determined by*

$$Vol(S)_k^2 = \frac{a^{2k}}{2^k (k!)^2} \left[\frac{2}{a^4} \sum_{i=1}^{k} \sum_{j=i+1}^{k} d_{i,k+1}^2 d_{j,k+1}^2 + \frac{2}{a^2} \sum_{i=1}^{k} d_{i,k+1}^2 \right. \\ \left. - \frac{k-1}{a^4} \sum_{i=1}^{k} d_{i,k+1}^4 - (k-1) \right].$$

From Theorem 5, we immediately derive the following elementary:

Corollary 1 *If w.l.o.g. $a > 1$, then*

$$\begin{aligned} Vol(S)_k^2 &= \frac{a^{2k}}{2^k (k!)^2} \left[\frac{2}{a^4} \sum_{i=1}^{k} \sum_{j=i+1}^{k} d_{i,k+1}^2 d_{j,k+1}^2 + \frac{2}{a^2} \sum_{i=1}^{k} d_{i,k+1}^2 \right. \\ &\quad \left. - \frac{k-1}{a^4} \sum_{i=1}^{k} d_{i,k+1}^4 - (k-1) \right] \\ &> \frac{a^{2k}}{2^k (k!)^2 a^4} \left[2 \sum_{i=1}^{k} \sum_{j=i+1}^{k} d_{i,k+1}^2 d_{j,k+1}^2 \right. \\ &\quad \left. + 2 \sum_{i=1}^{n} d_{i,k+1}^2 - (n-1) \sum_{i=1}^{k} d_{i,k+1}^4 - (k-1) \right]. \end{aligned}$$

7 Simplex volume maximization

Matrix factorization can be cast as an optimization problem where we seek to minimize the Frobenius norm $E = \|\mathbf{V} - \mathbf{WH}\|$ of the difference of a data matrix $\mathbf{V}$ and its low rank approximation $\mathbf{WH}$. Theorem 1 indicates that instead of determining a suitable $\mathbf{W}$ from minimizing the Frobenius norm, we may equivalently determine a solution from fitting a simplex of maximal volume into the data. Note that for other constrained low-rank approximations the concept of maximum-volume is known for quite some time (Goreinov and Tyrtyshnikov 2001). In contrast to Goreinov and Tyrtyshnikov, we optimize the volume of the general simplex and not of the parallelepiped.

Algorithm 1 Simplex Volume Maximization (SiVM)

```
$\mathbf{v}_j \leftarrow \mathbf{v}_{\text{rand}(n)}$     {Select vector $\mathbf{v}_j$ at random from $\mathbf{V}^{d\times n}$}
$\mathbf{w}_1 = \arg\max_k d(\mathbf{v}_k,\ \arg\max_p d(\mathbf{v}_j, \mathbf{v}_p))$     {Find first basis vector}
for $k = 2 \ldots K$ do
    $\phi_{k,p} \leftarrow \phi_{k-1,p} + d(\mathbf{w}_{k-1}, \mathbf{v}_p)$     {Corresponds to: $\sum_{i=1}^{k} d_{i,p}$}
    $\lambda_{k,p} \leftarrow \lambda_{k-1,p} + d(\mathbf{w}_{k-1}, \mathbf{v}_p)^2$     {Corresponds to: $\sum_{i=1}^{k} d_{i,p}^2$}
    $\rho_{k,p} \leftarrow \rho_{k-1,p} + d(\mathbf{w}_{k-1}, \mathbf{v}_p) \times \phi_{k-1}$     {Corresponds to: $\sum_{i=1}^{k} \sum_{j=i+1}^{k} d_{i,p} d_{j,p}$}
    $\mathbf{w}_k = \arg\max_p \left[ d_{\max} \times \phi_{k,p} + \rho_{k,p} - \frac{k-1}{2} \lambda_{k,p} \right]$
end for
```

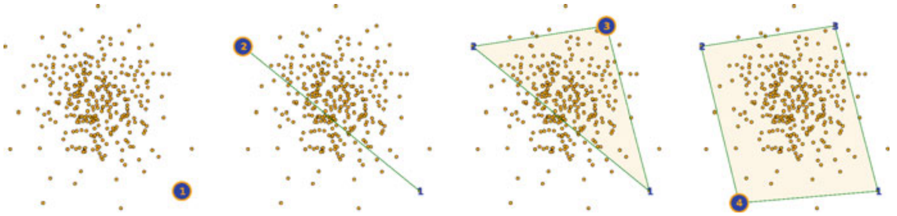

Fig. 5 Example of how the Simplex Volume Maximization (SiVM) algorithm iteratively determines four basis vectors for representation of a data sample by means of convex combinations (best viewed in color). (Color figure online)

Such a simplex could be found by directly optimizing the volume using the Cayley-Menger determinant of the distance matrix of the data. However, for large data sets this approach is ill-advised as it scales with $O(n^2)$ where n is the number of samples. Fortunately, it is possible to iteratively determine a set of k basis vectors in $O(kn)$ that maximize the volume of the simplex. Given a $(k-1)$-simplex S consisting of k vertices, we simply seek to find a new vertex $\mathbf{w}_{k+1} \in \mathbf{V}$ such that

$$\mathbf{w}_{k+1} = \arg\max_{p} \operatorname{Vol}(S \cup \mathbf{v}_p)^2.$$

From Theorem 5 we can now directly derive an iterative algorithm for finding the next best vertex.[9] Due to monotony and since all the $d_{i,j}$ are positive, we can reduce computation efforts by relinquishing to compute the distance squares. This does not significantly alter the algorithm but it is computationally less demanding. We arrive at

$$\mathbf{w}_{k+1} = \arg\max_{p} \left[a \sum_{i=1}^{k} d_{i,p} + \sum_{i=1}^{k} \sum_{j=i+1}^{k} d_{i,p} d_{j,p} - \frac{k-1}{2} \sum_{i=1}^{k} d_{i,p}^2 \right]. \tag{5}$$

For example, for the case where $k = 2$, i.e. for the situation where vertices $\mathbf{w}_1$ and $\mathbf{w}_2$ are already given, the next vertex $\mathbf{w}_3$ will be set to the data point $\mathbf{v}_\pi \in \mathbf{V}$ where

$$\mathbf{v}_\pi = \arg\max_{p} \left[a d_{1,p} + a d_{2,p} + d_{1,p} d_{2,p} - \frac{1}{2} d_{1,p}^2 - \frac{1}{2} d_{2,p}^2 \right].$$

This iterative approach to finding the next vertex translates to the simple and efficient Simplex Volume Maximization (SiVM) approach presented in Algorithm 1. We note that the pairwise distances computed in earlier iterations of the algorithm can be reused in later steps. For retrieving k latent components, we need to compute the distance to all data samples exactly $k+1$ times. The distances are computed with respect to the last selected basis vector. Informally, the algorithm can be formulated as finding the vertex $k+1$ that maximizes the simplex volume given the first k vertices. Figure 5 gives a didactic example on how SiVM iteratively determines basis vectors.

[9] Note that we omit constant values.

7.1 Initialization and parameter selection

The initial vertex for $k = 0$ can not be found from an iterative update. However, it is known that selecting a random vector $\mathbf{v}_1$ from a data sample, determining the data point $\mathbf{v}_2$ that is farthest away and then determining the data point $\mathbf{v}_3$ with the largest distance to $\mathbf{v}_2$, yields a vertex on the convex hull whose overall distance to the data is maximal (Ostrouchov and Samatova 2005). In fact, this initialization step is being used in the well known Fastmap algorithm by Faloutsos and Lin (1995). Therefore, $\mathbf{w}_1 = \arg\max_p d(\mathbf{v}_p, \arg\max_i d(\mathbf{v}_1, \mathbf{v}_i))$.

Furthermore, we did not yet set a value for a. Because it is known that the volume of a simplex is maximal if its vertices are equidistant, we assumed a constant edge length of a when formulating Theorem 5. Obviously, in a practical application of SiVM, the constant a has to be chosen appropriately. From 5 we see that the maximal possible volume is bounded from above by setting the free parameter a to the maximum of all observed distances $d_{\max}$. Assuming all edge lengths to be $d_{\max}$ must result in the largest possible volume. At the same time, we have a lower bound for each iterative update step which corresponds to choosing the minimal distance $d_{\min} = \min d(\mathbf{w}_j, \mathbf{w}_i)$ where $i \neq j$ and $\mathbf{w}_j, \mathbf{w}_i \in \mathbf{W}$. Assuming equidistant vertices, i.e., the edges have the same lengths and setting a to $d_{\min}$ will therefore result in the smallest possible volume under the current configuration. Consequently, setting a to $d_{\min}$ maximizes the lower bound and as such optimizes the volume as desired. As the minimal distance changes with each iteration step, we have to adapt a accordingly. However, experimental validation did not reveal a significant empirical difference to setting a to $d_{\max}$. Actually, for larger data sets, $d_{\max}$ and $d_{\min}$ were found to be similar in practice. Since setting a to $d_{\max}$ offers slight advantages in terms of computational efficiency, we adhere to this strategy.

Finally, we did not address the computation of the coefficient matrix $\mathbf{H}$ so far. This, however, is straightforward once a suitable set of basis vectors $\mathbf{W}$ has been determined. More precisely, $\mathbf{H}$ can be found from solving the following constrained quadratic optimization problem

$$\begin{aligned} \min \quad & \|\mathbf{v}_i - \mathbf{W}\mathbf{h}_i\| \\ \text{s.t.} \quad & \mathbf{1}^T\mathbf{h}_i = 1 \\ & \mathbf{h}_i \succeq \mathbf{0}\,. \end{aligned}$$

Moreover, this process can be fully parallelized as the coefficients of data vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ are independent

7.2 Computational complexity

Based on SiVM, we propose the following 2-steps matrix factorization approach called SiVM-NMF: (1) compute the basis vectors using SiVM and (2) compute the coefficients $\mathbf{H}$ as shown above.

Step 1 mainly consists of an iterative computation of distances. For each basis vector $\mathbf{w}_i$ we have to compute the distance $d(\mathbf{w}_i, \mathbf{v}_i)$, $\mathbf{v}_i \in \mathbf{V}$ once. If we assume k

basis vectors, this translates to $O(k \cdot n)$. The three simple addition operations in lines 4-6 of Algorithm 1 do not increase computation times substantially for large n and any conventional distance metric (e.g. in our experiments, they typically require less than 0.1% of the overall computation time). Usually, we can assume that $k \ll n$, which leads to linear complexity $O(n)$ for step 1.

Step 2 is highly dependent on the constraints imposed on the factorization method. It requires to compute the coefficients $\mathbf{H}$ that minimize the Frobenius norm given the basis elements $\mathbf{W}$ that result from step 1. In the experiments presented below, we apply quadratic programming approaches for computing $\mathbf{H}$. As this has to be done once for each data sample in $\mathbf{V}$, the complexity is of $O(n \cdot QP(k)))$. Thus for $k \ll n$, the complexity is linear in the second step as well. Moreover, once suitable basis vectors $\mathbf{W}$ have been determined, computing the coefficients for each data sample $\mathbf{v}_i \in \mathbf{V}$ could be done in parallel to further reduce computation time.

Thus, the overall running time complexity of SiVM-NMF is $O(n)$ assuming $k \ll n$ for some fix k.

8 Experiments

In the following, we present our experimental evaluation of SiVM-NMF. Our intention here is to investigate the following questions:

- **(Q1)** Does SiVM-NMF perform comparably to related methods in terms of run-time and accuracy?
- **(Q2)** Does SiVM-NMF find basis factors that are more easily interpretable than the ones computed by related methods?
- **(Q3)** Does SiVM-NMF scale well to large-scale data sets?

To this aim, we have implemented[10] SiVM-NMF and related factorization methods in scientific Python using the h5py library[11] and the cvxopt library by Dahl and Vandenberghe.[12] For reading hyperspectral images, we used the SPy library.[13] Although SiVM-NMF can be easily parallelized, we used a serial version running on a standard Intel-Quadcore 2. GHz computer.

Overall, we decided for three experimental setups. Our first experimental setup evaluates and compares the run-time and accuracy performance using synthetically generated data in order to address **Q1**. The second setup investigates **Q2** by applying SiVM-NMF to several small, real-world sustainability data sets already discussed in Sect. 3: U.S. States temperatures and precipitations values, global energy consumption of the world's countries, and the environmental performance index. Finally, we investigated SiVM-NMF's performance to analyze several medium-scale and large-scale, real-world hyperspectral images: AVIRIS and Natural Scene Spectral Images.

[10] http://pymf.googlecode.com.

[11] http://h5py.googlecode.com.

[12] http://abel.ee.ucla.edu/cvxopt/.

[13] http://spectralpython.sourceforge.net/.

Table 1 Description of the real world data sets used in the experimental evaluation

Name	Size		# Entries
U.S. temperature	50 × 480 (State × month)		24,000
U.S. precipitation	50 × 480 (State × month)		24,000
Electricity consumption	28 × 211 (Year × country)		5,908
Name of AVIRIS spectral image	Size		# Entries
	Band × (X × Y)	Band × size	
Indian Pines	220 × (145 × 145)	220 × 21,025	4,625,500
Cuprite, Nevada	50 × (400 × 350)	50 × 140,000	7,000,000
Moffett Field, California	56 × (500 × 350)	56 × 175,000	9,800,000
Jasper Ridge area, California	60 × (600 × 512)	60 × 307,200	18,432,000
Name of natural scene spectral image	Size		# Entries
	Band × (X × Y)	Band × size	
Sameiro area (Braga)	33 × (1,018 × 1,339)	33 × 1,363,102	44,982,366
Ruivães (Vieira do Minho)	33 × (1,017 × 1,338)	33 × 1,360,746	44,904,618
Museum of the Monastery (Mire de Tibães)	33 × (1,018 × 1,267)	33 × 1,289,806	42,563,598
Gualtar campus (University of Minho)	33 × (1,019 × 1,337)	33 × 1,362,403	44,959,299
Terras de Bouro (Minho region)	32 × (1,020 × 1,339)	33 × 1,365,780	45,070,740
Picoto area (Braga)	33 × (1,021 × 1,338)	33 × 1,366,098	45,081,234
Ribeira area (Porto)	33 × (1,017 × 1,340)	33 × 1,362,780	44,971,740
Souto (Minho region)	33 × (1,018 × 1,340)	33 × 1,364,120	45,015,960

Specifically, we investigated **Q2** on the AVIRIS Indian Pines hyperspectral image and **Q3** on all 12 hyperspectral images. The used data sets and their statistics are summarized in Table 1 and further explained in the following subsections. The used matrix factorization methods were:

NMF: Standard non-negative matrix factorization. Although it does not fulfill**R1**, we compare to it for the sake of completeness. Of course, sub-sampling strategies and other advanced version of NMF could be employed to scale to massive data sets.

K-Means: Standard k-means implementation. Note that is also does not fulfill **R1** so we compare to it only for sake of completeness.

C-NMF/AA: Convex-NMF resp. Archetypal Analysis, see Sect. 2 for references.

CH-NMF: Convex-hull NMF as introduced in(Thurau et al. 2009).To accommodate for large-scale, we used FastMap (Faloutsos and Lin 1995) for efficiently computing candidates of convex-hull vertices.

CH-NMF R: From a robust statistics viewpoint, the distribution of distances may contain information on potential outliers. We remove the outliers by trimming the most extreme data points as described by Ostrouchov and Samatova (2005). That is we take a constant number r of large distances, cluster the corresponding data points, and choose a central point of the largest cluster as candidate. In Sect. 3, we called this approach “Robust”.

SiVM: Simplex Volume Maximization as proposed in the present paper but it does not compute the reconstruction weights of the data points.

SiVM-NMF: Also computes the reconstruction weights of the data points.

SiVM+NMF: SiVM-NMF used as initialization for NMF.

8.1 Q1: Synthetic data

In order to evaluate the performance of SiVM-NMF and to compare it to that of previously proposed factorization methods, we generated 5 different data sets randomly sampled from a 5-dimensional uniform distribution. Each data set was then processed using k-means clustering, Convex-hull NMF (CH-NMF), and Convex NMF (C-NMF) and Archetypal Analysis (AA). We measured the Frobenius norm and running times for each randomly generated set.

Figure 6 shows the Frobenius norm for up to 50,000 randomly sampled data points (note the logarithmic scale on the y axis) averaged over the five datasets. For all the methods in this test, we computed 10 basis vectors and used the same number of iterations for computing them. It can be seen that NMF yields the lowest Frobenius norm on average. We attribute this to its less restrictive constraint (non-negativity). CH-NMF and SiVM-NMF impose the same additional convexity constraint and are found to yield very similar results.

Figure 6 shows the average computation time in seconds for the tested methods. We decided to exclude the running times for computing the coefficient vectors for CH-NMF and SiVM-NMF. This allows for a better comparison of the two most related methods. As the computation of coefficients is an inherent part of k-means and NMF, we could not exclude it in these cases. We find the computation of SiVM-NMF basis vectors to require only small fractions of a second even for data sets consisting of several thousand samples.

To further investigate the relation between the minimization of the Frobenius norm and the maximization of the Simplex Volume, we decided for an in-depth comparison of Convex-NMF (or Archetypal Analysis) and Simplex Volume Maximization. Convex-NMF attempts to directly minimize the Frobenius under the convexity constraint discussed in Sect. 5. SiVM-NMF, in contrast, maximizes the simplex volume for detecting suitable basis vectors to represent the data by means of convex combinations. Figure 6 shows the results for iid data samples from a 3-dimensional cube averaged over three runs. Note that we had to reduce the number of sampled data points as both C-NMF and AA do not scale to more than a few thousand data samples. It can be seen that both approaches result in almost identical average reconstruction errors. However, Fig. 6 shows that with respect to computation time (this time including computation times for the coefficient matrices **H**), SiVM-NMF is orders of magnitude faster. Thus, it offers a viable alternative to AA and C-NMF as it yields similar or even better reconstruction errors but runs fast enough to allow for large-scale data analysis.

To summarize, the results clearly show that **Q1** can be answered affirmatively.

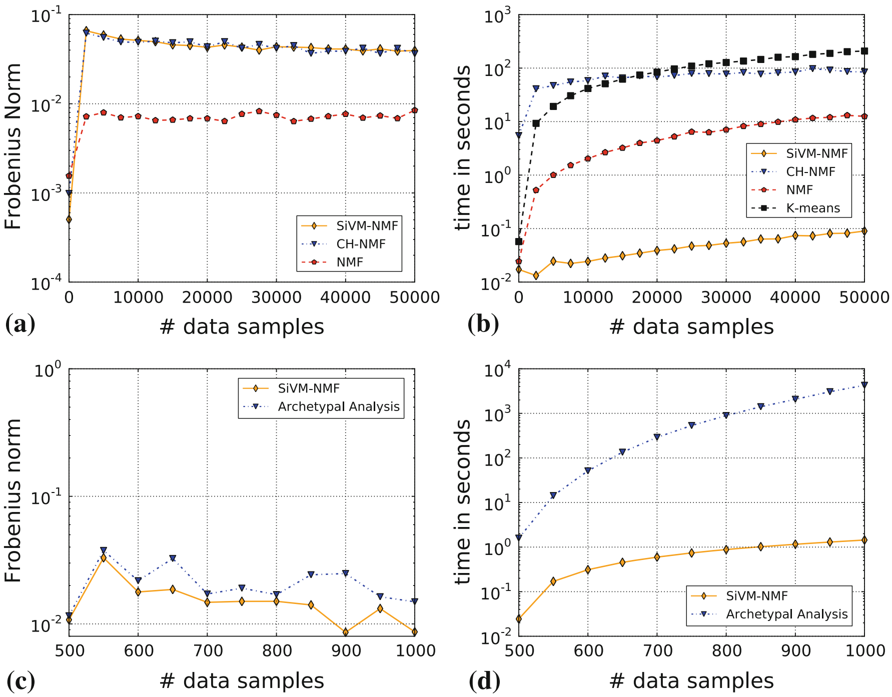

Fig. 6 Accuracy and running time comparison on synthetic data. The Frobenius norm is divided by the number of matrix elements (best viewed in color). (**a**) Resulting Frobenius norms averaged over 5 runs for various latent component detection techniques, (**b**) Runtimes averaged over 5 runs, (**c**) Resulting Frobenius norms of C-NMF/AA and SiVM-NMF averaged over 3 runs, (**d**) Runtimes C-NMF/AA and SiVM-NMF averaged over 3 runs. (Color figure online)

8.2 Q2: Temperature, precipitation, and energy consumption

The results presented already in Sect. 3 clearly answer **Q2** affirmatively. Figure 7 summarize the quantitative results for these data sets. Specifically, they show the Frobenious norms and the running times of several non-negative matrix factorizations for the HCS temperature and precipitation normals as well as for the EIA's global electricity consumptions. As one can see, SiVM-NMF performs very well compared to other convexity constrained NMF approaches as well as NMF and K-Means in terms of Frobenius norm and runtime. Additionally, it can also be used as initialization method for NMF (SiVM+NMF) to get similar or even slightly better reconstruction errors as NMF. This is additional evidence that **Q1** can be answered affirmatively.

8.3 Q2+Q3: Hyperspectral image analysis

Hyperspectral images are often generated from airborne sensors like the NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The AVIRIS first began operations aboard a NASA research craft in 1987. It has since become a standard

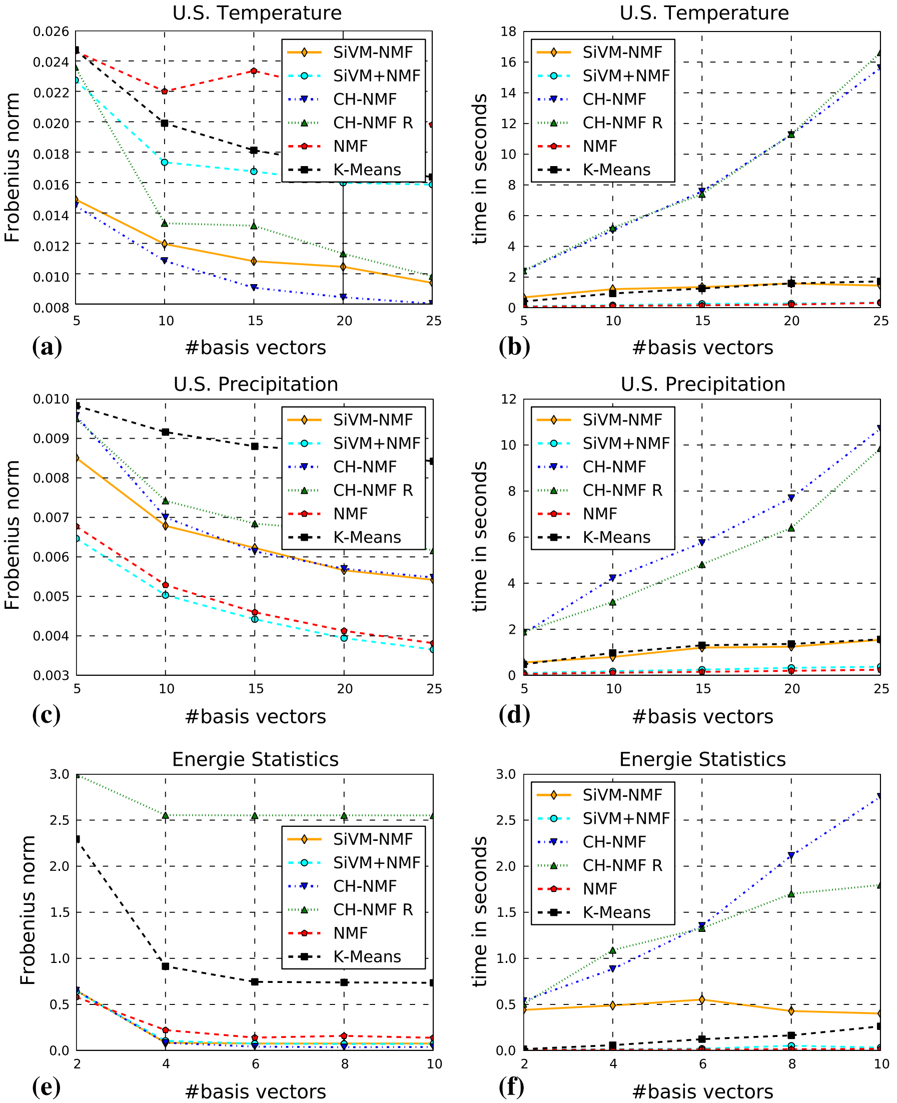

Fig. 7 Accuracy and running time comparison on the data sets from Sect. 3. The Frobenius norm is divided by the number of matrix elements (best viewed in color). (**a**) Resulting Frobenius norm, (**b**) Runtime performance, (**c**) Resulting Frobenius norm, (**d**) Runtime performance, (**e**) Resulting Frobenius norm, (**f**) Runtime performance. (Color figure online)

instrument used for remote sensing of the Earth, collecting spectral radiance data for characterization of the Earth's surface and atmosphere. AVIRIS and hyperspectral data is often used in the fields such as oceanography, environmental science, snow hydrology, geology, volcanology, soil and land management, atmospheric and aerosol studies, agriculture, and limnology. For each location flown over, reflectance data (used for quantitative characterization of surface features) for several contiguous spec-

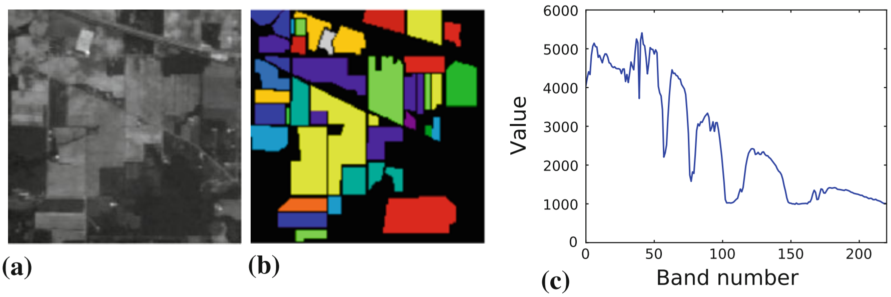

Fig. 8 AVIRIS data sets: (**a**) Band 10 (0.5μ m) of the AVIRIS Indian Pines data set. (**b**) Its ground truth. The classes are (1) alfalfa, (2) corn-notill, (3) corn- min, (4) corn, (5) grass/pasture, (6) grass/trees, (7) grass/pasture-mowed, (8) hay-windrowed, (9) oats, (10) soybeans-notill, (11) soybeans- min, (l2) soybeans-clean, (13) wheat, (14) woods, (15) building-grass-tree-drive, and (16) stone-steel towers. (**c**) An example spectral profile taken from the AVIRIS Indian Pines data set (best viewed in color). (**a**) AVIRIS Indian Pines, (**b**) Ground Truth, (**c**) Spectral Band. (Color figure online)

tral channels (bands) is recorded. Although the reflectance is a continuous function of wavelength, each feature (band number) corresponds to a discrete sampling of a particular location's spectral profile. Consequently, a 3-dimensional cube of data is generated for each area analyzed by AVIRIS. Consider for example Fig. 8a–c showing the June 1992 AVIRIS data set collected over the Indian Pines test site in an agricultural area of northern Indiana. The image has 145 × 145 pixels with 220 spectral bands and contains approximately two-thirds agricultural land and on-third forest or other elements. This results in a $220 \times 21,050$ band by pixel location matrix A containing non-negative reflectance data. In general, the cube of data is even larger. One of our hyperspectral image consists of $33 \times 1,018 \times 1,339 = 33 \times 1,363,102$ band by pixel matrix, i.e., in total about 45 million entries. In practice, one often finds images with even hundreds of millions of entries. In other words, scaling is important for analyzing hyperspectral images. Based on a location's spectral profile, one is typically interested in determining what primary physical components exists within an area flown over. To determine these components, one commonly applies some form of non-negative matrix factorization to the band by pixel matrix. Using the many locations, one aims at obtaining the k components, i.e., basis vectors that could best be added together to reconstruct each location's spectral profile as closely as possible.

In total, we used 12 spectral images. The first data set, the AVIRIS Indian Pines data set, has extensive ground-truth information available, thus allowing us to qualitatively compare the performance of SiVM-NMF to NMF. For the other 11 data sets, we do not have the ground truth. Therefore, we used them only for run-time and reconstruction accuracy comparison. Overall, there were three other AVIRIS images and eight hyperspectral images of natural scenes. The eight hyperspectral images of natural scenes are due to (Foster et al. 2004)[14] and are a mixture of rural scenes from the Minho region of Portugal, containing, rocks, trees, leaves, grass, and earth and of urban scenes from the cities of Porto and Braga, Portugal. Images were obtained during the summers of

[14] Available from http://personalpages.manchester.ac.uk/staff/david.foster/Hyperspectral_images_of_natural_scenes_04.html.

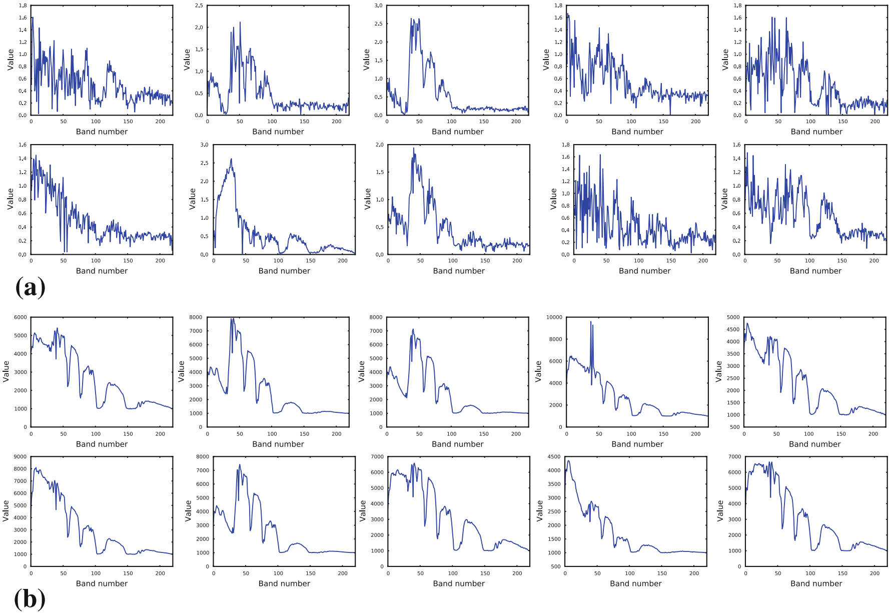

Fig. 9 The first 10 out of 14 basis vectors (band number vs. value) found on the Indian Pines data set. (**a**) for NMF, they retain the non-negativity of the reflectance data but they no longer maintain any of the spectral continuity. (**b**) For SiVM-NMF, they retain the non-negativity of the reflectance data and maintain the spectral continuity. (**a**) NMF Basis Vectors, (**b**) SiVM-NMF Basis Vectors. (Color figure online)

2002 and 2003, almost always under a clear sky. Particular care was taken to avoid scenes containing movement. Scenes were illuminated by direct sunlight in clear or almost clear sky.

8.3.1 Q2: Qualitative analysis: AVIRIS Indian pines

We applied NMF with random initialization to the Indian Pines data set and obtain the 14 spectral basis shown in Fig. 9a. While the 14 basis profiles shown here retain the non-negativity of the reflectance data, they no longer maintain any of the spectral continuity. This is not to say that the basis profiles found will not sum up to reconstruct the original profiles. The basis profiles found are just significantly corrupted with noise, do not correspond to actual profiles and, hence, do not maintain the clear structure that would allow one to determine if they correspond to other known surfaces (i.e. sand, vegetation, etc.). In contrast, SiVM-NMF found the basis vectors shown in Fig. 9b. These basis profiles should be compared to those obtained by NMF shown in Fig. 9a. They are much smoother (preserve continuity better) than those obtained using NMF: they are actual spectral profiles.

To further investigate the difference in performance, we computed abundance maps for the entire image. That is, we assigned every data point to the basis vector with the largest proportion value. Figure 10 shows examples of resulting maps. Both maps should be compared to the ground truth in Fig. 8b. As one can see, the SiVM-NMF basis vector separate the classes better than the NMF basis vector. To quantify this,

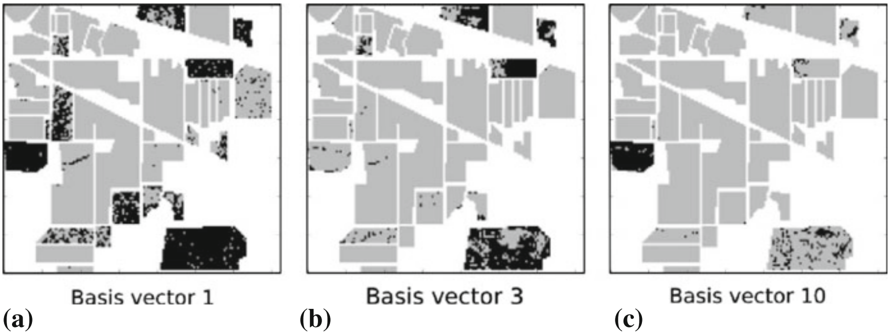

Fig. 10 Example abundance maps founds on the labeled AVIRIS Indian Pines data set for (**a**) NMF and (**b, c**) SiVM-NMF. Pixels in white are unlabeled. Pixels in *gray* indicate pixels from other classes. Remaining pixels have abundance levels > 0.2. This should be compared to the ground truth in Fig. 8b. As one can see, NMF mixes the *light blue* class (grass/pasture) and the *red* classes (wheat/woods/building-grass-tree-drive/stone-steel towers). SiVM-NMF separates the red classes better from the other classes than NMF. (**a**) NMF BV 1, (**b**) SiVM-NMF BV 3, (**c**) SiVM-NMF BV 10

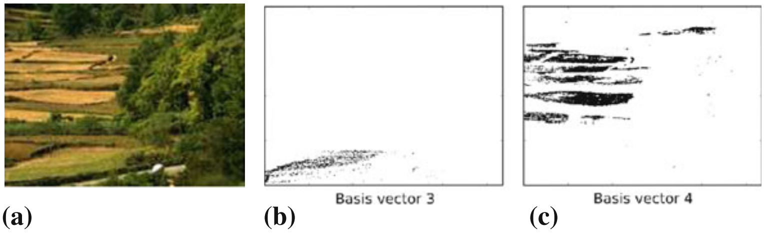

Fig. 11 SiVM-NMF abundance maps for large-scale "Terras de Bouro (Minho region)" hyperspectral image. We do not have ground truth for this data set but one can see that the different types of greens are well separated (best viewed in color). (**a**) Colore picture of "Terras de Bouro **Minho region**. (Color figure online)

we also computed the sums of Shannon entropies of the abundance distributions. For SiVM-NMF, the entropy was 21.33 whereas NMF's entropy was 32.77. This indicates that SiVM-NMF actually uses fewer number of basis vectors to describe each ground truth class, and that its basis vectors are better representatives of the ground truth classes.

To summarize, the results clearly indicate that **Q2** can be answered affirmatively.

8.3.2 Q3: Quantitative analysis: large-scale hyperspectral images

To see how well SiVM-NMF scales to large data sets, i.e., to investigate **Q3**, we considered the other 11 hyperspectral images. Figure 11 shows some abundance maps found for the "Terras de Bouro (Minho region)" hyperspectral image consisting of $32 \times 1,365,780 = 43,704,960$ entries. We compared SiVM-NMF to other convexity constrained NMF methods in terms of running time. For the three smaller AVIRIS spectral images, we also compared to NMF and report the Frobenius norm error, cf. Fig. 12. The runtime results on the other eight large-scale images are summarized in Fig. 13.

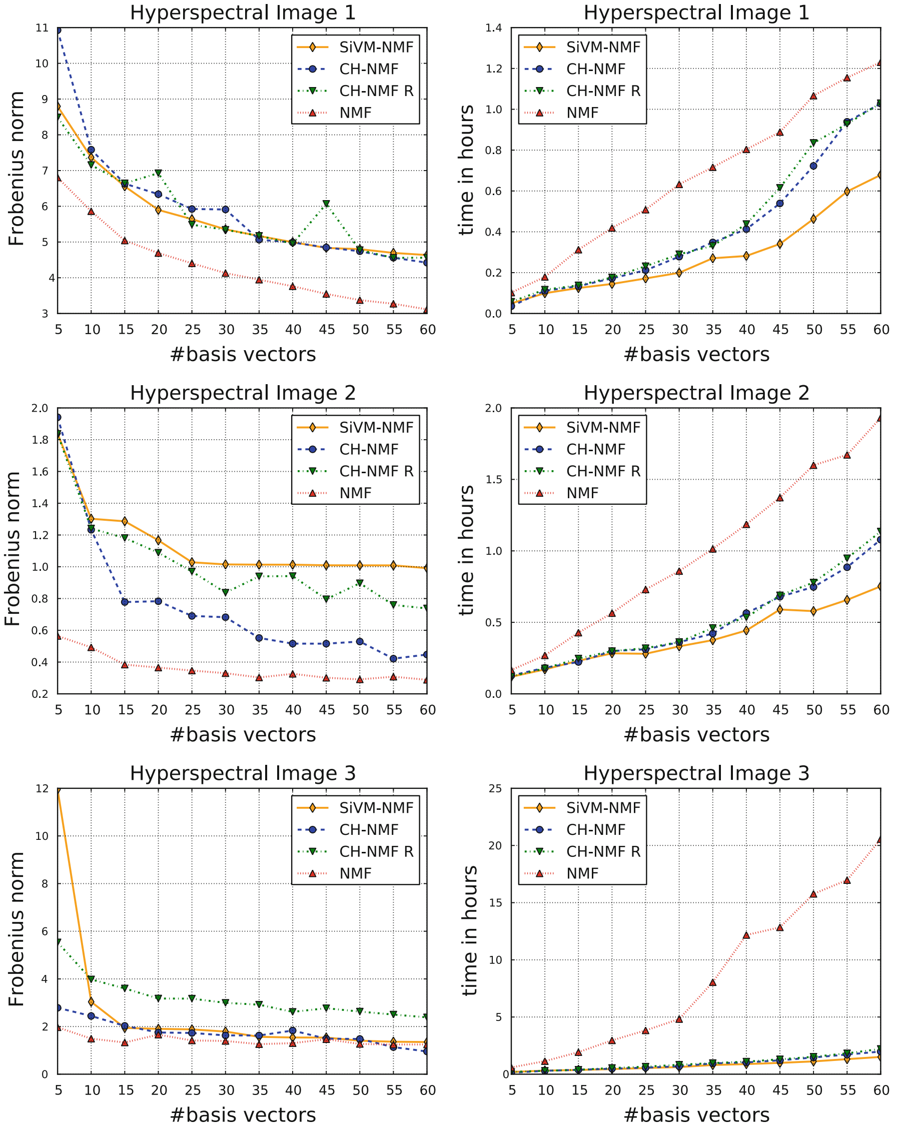

Fig. 12 Accuracy and running time comparison on the AVIRIS data sets. The Frobenius norm is divided by the number of matrix elements (best viewed in color). (Color figure online)

As one can see, SiVM-NMF compares favorable to all baselines and scales well to massive data sets. Furthermore, all CH-NMF baselines can only compute at most 60 basis vectors, 2-times as many basis vectors as there are dimensions. This is because they only identify basis vectors essentially lying on the convex hull. In contrast, SiVM-NMF can compute so called overcomplete representations, i.e., the number of basis vectors is greater than the dimensionality of the input. Overcomplete representations have been advocated because they have greater robustness in the presence of noise, can

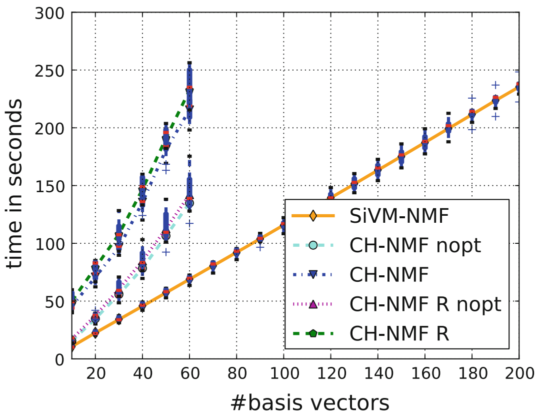

Fig. 13 Box-plot of the running times over all the large-scale, natural scene hyperspectral images. The "nopt" label for CH-NMF (R) denotes versions of the corresponding methods that do not compute the *k* best candidates from the found *l* convex-hull vertex candidates but simply use directly the *l* candidates as basis vectors. This avoids the computationally demanding optimization step but yields reconstructions of lower quality. (Color figure online)

be sparser, and can have greater flexibility in matching structure in the data. Exploring this feature is left for future work.

Finally, in an experiment not reported here, we used our SiVM-NMF Python implementation running on a standard desktop computer to factorize 80 million tiny Google images, a matrix with 30 billion entries. It took about 4 h to compute a single basis vector on a single core machine.These results clearly show that **Q3** can be answered affirmatively.

To summarize, the experimental results show that all questions **Q1–Q3** can be answered affirmatively.

9 Conclusions

As Gomes (2009) highlights, the "development of policies for a sustainable future presents unique computational problems in scale, impact, and richness". Motivated by this, we have presented a novel method for finding latent components in massive data sets, called SiVM-NMF, that is fast and scales well. We formulated the problem as a classical constrained matrix factorization problem that should minimize the Frobenius norm. The solution we presented, however, instead maximizes the volume of the simplex spanned by the latent components. This novel formulation is not only equivalent when imposing convexity constraints, it also allows for rapidly finding the latent components. More importantly, the latent components are actual data points so that they are easy to interpret even by non-experts. That is, in contrast to most existing matrix factorization techniques, SiVM-NMF is not focussing on the model but is driven by the question "how much can the data itself help us to factorize matrices?". This direction is particularly appealing given that intuitively understandable results are, in our opinion, another key issue for many computational sustainability problems.

As already envisioned by such eighteenth-century philosophers as Jean Jacques Rousseau, John Locke, and John Stuart Mill, government requires that everyone have the right to influence political and environmental decisions that affect them. A basic assumption is that everybody is — or should be—essentially equal, in both their concern for environmental issues and their competency to make decisions about them. However, in order to make these decisions, (at least informed) individuals need accurate and understandable models. One of the benefits of SiVM-NMF for latent component analysis is that the resulting basis vectors are often readily interpretable even to non-expert users. As shown by our extensive empirical results, by extracting the most extreme instances of a data set, SiVM-NMF yields basis elements that are well distinguishable data points. Since this accommodates the principle of opposites in human cognition, the resulting basis elements are quite easy to interpret across several sustainability problems. In the case of hyperspectral images, the spectral profiles found by SiVM-NMF may not make complete sense to the reader. However, with only a little training in the field of remote sensing, one can immediately identify the features corresponding to the profile of, say, water and could consequently make some skilled interpretation of the basis obtained.

There are several interesting avenues for future work. First of all, SiVM-NMF relies on distance computations only. Although, we have here focussed on the Euclidean distance, it can directly be applied to other distances. For example, using the cosine distance will lead to latent components that maximize the angular difference. From a robust statistics viewpoint, the most extreme instances of a data set might actually be outliers that should be removed for instance by trimming them in a similar fashion as done in robust CH-NMF. Climate change, energy efficiency, etc., data naturally evolves over time so that SiVM-NMF should be extended to deal with temporal data. Data streams and tensors are further interesting data types. Another interesting avenue is parallelization. In a preliminary implementation using the map-reduce framework (Dean and Ghemawat 2008) we observed a linear scaling with the number of cores.

Overall our contribution and results are an encouraging sign that applying matrix factorizations in the wild, that is on gigantic matrices with billions of entries may not be insurmountable.

Acknowledgements The authors would like to thank Katharina Morik for her encouragement and the anonymous reviewers for their comments. Kristian Kersting and Mirwaes Wahabzada were supported by the Fraunhofer ATTRACT Fellowship STREAM.

Appendix

Proof (of Theorem 1) We first note that

$$\left\| \mathbf{V} - \mathbf{W}^{d\times(k+1)}\mathbf{H}^{(k+1)\times n} \right\|^2 = \sum_i \left\| \mathbf{v}_i - \mathbf{W}^{d\times(k+1)}\mathbf{h}_i^{k+1} \right\|^2 \tag{6}$$

and that

$$\left\| \mathbf{V} - \mathbf{W}^{d\times k}\mathbf{H}^{k\times n} \right\|^2 = \sum_i \left\| \mathbf{v}_i - \mathbf{W}^{d\times k}\mathbf{h}_i^{k} \right\|^2 . \tag{7}$$

Consider the simplex $\mathbf{W}^{d\times k}$. If $\mathbf{v}_i$ is a point inside the simplex, then there exists an optimal k-dimensional coefficient vector $\mathbf{h}_i^k$ where $\mathbf{1}^T\mathbf{h}_i^k = 1$ and $\mathbf{h}_i^k \succeq \mathbf{0}$ such that

$\left\| \mathbf{v}_i - \mathbf{W}^{d\times k}\mathbf{h}_i^k \right\|^2 = 0$. If $\mathbf{v}_i$ is a point outside of the simplex, $\mathbf{W}^{d\times k}\mathbf{h}_i^k$ will be its projection onto the nearest facet of the simplex and $\left\| \mathbf{v}_i - \mathbf{W}^{d\times k}\mathbf{h}_i^k \right\|^2 > 0$.

Adding a randomly chosen vertex $\mathbf{w}_{k+1} \in \mathbf{V}$, $\mathbf{w}_{k+1} \notin \mathbf{W}^{d\times k}$ introduces new facets to the simplex and increases its volume. For any $\mathbf{v}_i$, we can determine an optimal coefficient vector $\mathbf{h}_i^{k+1}$ from solving the constraint quadratic problem

$$\begin{aligned} \min \quad & \left\| \mathbf{v}_i - \mathbf{W}^{d\times(k+1)}\mathbf{h}_i^{k+1} \right\| \\ \text{s.t.} \quad & \mathbf{1}^T\mathbf{h}_i^{k+1} = 1, \ \mathbf{h}_i^{k+1} \succeq \mathbf{0} \end{aligned}$$

and have to distinguish four cases:

(i) If $\mathbf{v}_i$ was inside $\mathbf{W}^{d\times k}$ it will also be inside $\mathbf{W}^{d\times(k+1)}$ and there exists an optimal $k+1$-dimensional coefficient vector $\mathbf{h}_i^{k+1}$ such that the reconstruction error vanishes.

(ii) If $\mathbf{v}_i$ was outside $\mathbf{W}^{d\times k}$ but is inside the extended simplex $\mathbf{W}^{d\times(k+1)}$, the reconstruction error decreases $\left\| \mathbf{v}_i - \mathbf{W}^{d\times k}\mathbf{h}_i^k \right\|^2 > \left\| \mathbf{v}_i - \mathbf{W}^{d\times(k+1)}\mathbf{h}_i^{k+1} \right\|^2$.

(iii) and (iv) If $\mathbf{v}_i$ was outside $\mathbf{W}^{d\times k}$ and remains outside of $\mathbf{W}^{d\times(k+1)}$, it may or may not be projected to a new, possibly closer facet. Either way $\left\| \mathbf{v}_i - \mathbf{W}^{d\times k}\mathbf{h}_i^k \right\|^2 \geq \left\| \mathbf{v}_i - \mathbf{W}^{d\times(k+1)}\mathbf{h}_i^{k+1} \right\|^2$.

Therefore, after extending the simplex $\mathbf{W}^{d\times k}$ by a vertex $\mathbf{w}_{k+1}$, none of the terms on the right hand side of (6) is larger than the corresponding term on the right hand side of (7). □

Proof (of Theorem 2) We consider the Cayley-Menger determinant as introduced in (4). To simplify notation, we set $\delta_{i,j} = d_{i,j}^2$ and $\alpha = a^2$. If we assume equidistant edge lengths a for the first n vertices, the CMD reads

$$\det(\mathbf{A}) = \begin{vmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & \alpha & \alpha & \dots & \delta_{1,k+1} \\ 1 & \alpha & 0 & \alpha & \dots & \delta_{2,k+1} \\ 1 & \alpha & a & 0 & \dots & \delta_{3,k+1} \\ \vdots & & & & \ddots & \vdots \\ 1 & \delta_{1,k+1} & \delta_{2,k+1} & \delta_{3,k+1} & \dots & 0 \end{vmatrix}.$$

Interchanging two rows or two columns leaves the determinant unchanged as does adding a multiple of one row to another row or adding a multiple of one column to another column. We exchange the first and the last row as well as the first and the last column. To set most elements of the upper left submatrix to zero, we subtract the last row multiplied by α from each row. All entries will now be zero, except for the diagonal elements and the elements in the last and last but one row and column. Next, we add the first n rows times α^{-1} to the last row. The first elements in the last but one row can be set to zero by adding the first i rows times $\delta_{i,k}\alpha^{-1}$ to the last row. Finally, we subtract the last column from the last but one column. These manipulations yield

$$\det(\mathbf{A}) = \begin{vmatrix} -\alpha & 0 & \dots & \delta_{1,k+1} - \alpha & 1 \\ 0 & -\alpha & \dots & \delta_{2,k+1} - \alpha & 1 \\ 0 & 0 & \dots & \delta_{3,k+1} - \alpha & 1 \\ \vdots & & \ddots & & \\ 0 & 0 & \dots & \left[\sum_{i=1}^{k} \delta_{i,k} \frac{\delta_{i,k}-\alpha}{\alpha}\right] & \left[1 + \sum_{i=1}^{k} \frac{\delta_{i,k}}{\alpha}\right] \\ 0 & 0 & \dots & \left[1 + \sum_{i=1}^{k} \frac{\delta_{i,k}-\alpha}{\alpha}\right] & \left[\frac{n}{\alpha}\right] \end{vmatrix}.$$

To eliminate the sums in the last but one column and the last row, we multiply the last column with its reciprocal value. The determinant is now in upper triangular form and since the determinant of a triangular matrix equals the product of the diagonal entries, we can now write

$$\det(\mathbf{A}) = -\alpha^k \left[\sum_{i=1}^{k} \frac{\delta_{i,k}(\delta_{i,k} - \alpha)}{\alpha} \cdot \frac{k}{\alpha} - \left(1 + \sum_{i=1}^{k} \frac{\delta_{i,k}}{\alpha} \cdot \left(1 + \sum_{i=1}^{k} \frac{\delta_{i,k} - \alpha}{\alpha} \right) \right) \right].$$

After some tedious but straightforward algebra, this further simplifies to

$$\det(\mathbf{A}) = -\alpha^k \left[\frac{n-1}{\alpha^2} \sum_{i=1}^{k} \delta_{i,k+1}^2 + (k-1) \right.$$
$$\left. - \frac{2}{\alpha} \sum_{i=1}^{k} \delta_{i,k+1} - \frac{2}{\alpha^2} \sum_{i=1}^{k} \sum_{j=i+1}^{k} \delta_{i,k+1} \delta_{j,k+1} \right].$$

Plugging $\det(\mathbf{A})$ back into (3) yields the following more accessible expression for the simplex Volume

$$\frac{(-1^{k+1})(-\alpha^k)}{2^k (k!)^2} \left[\frac{k-1}{\alpha^2} \sum_{i=1}^{k} \delta_{i,k+1}^2 + (k-1) \right.$$
$$\left. - \frac{2}{\alpha} \sum_{i=1}^{n} \delta_{i,k+1} - \frac{2}{\alpha^2} \sum_{i=1}^{k} \sum_{j=i+1}^{k} \delta_{i,k+1} \delta_{j,k+1} \right].$$

For any k, either (-1^{k+1}) or $(-\alpha^k)$ will be negative, which leads to the form required by the theorem

$$\mathrm{Vol}(S)_k^2 = \frac{\alpha^k}{2^k(k!)^2}\left[\frac{2}{\alpha^2}\sum_{i=1}^{k}\sum_{j=i+1}^{k}\delta_{i,k+1}\delta_{j,k+1} + \frac{2}{\alpha}\sum_{i=1}^{k}\delta_{i,k+1} - \frac{k-1}{\alpha^2}\sum_{i=1}^{k}\delta_{i,k+1}^2 - (k-1)\right].$$

□

References

Achlioptas D, McSherry F (2007) Fast computation of low-rank matrix approximations. J ACM 54(9):1–19

Aguilar O, Huerta G, Prado R, West M (1998) Bayesian inference on latent structure in time series. In: Bernardo J, Bergen J, Dawid A, Smith A (eds) Bayesian statistics. Oxford University Press, Oxford

Blumenthal LM (1953) Theory and applications of distance geometry. Oxford University Press, Oxford

Chan B, Mitchell D, Cram L (2003) Archetypal analysis of galaxy spectra. Mon Not R Astron Soc 338(3):790–795

Chang CI, Wu CC, Liu WM, Ouyang YC (2006) A new growing method for simplex-based endmember extraction algorithm. IEEE T Geosci Remote 44(10):2804–2819

Crippen G (1988) Distance geometry and molecular conformation. Wiley, New York

Cutler A, Breiman L (1994) Archetypal analysis. Technometrics 36(4):338–347

Dean J, Ghemawat S (2008) Mapreduce: simplified data processing on large clusters. Commun ACM 51(1):107–113

Ding C, Li T, Jordan M (2010) Convex and semi-nonnegative matrix factorizations. IEEE T Pattern Anal 32(1):45–55

Drineas P, Kannan R, Mahoney M (2006) Fast Monte Carlo algorithms III: computing a compressed approixmate matrix decomposition. SIAM J Comput 36(1):184–206

Faloutsos C, Lin KI (1995) FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In: Proceedings of the ACM SIGMOD international conference on management of data, San Diego

Foster D, Nascimento S, Amano K (2004) Information limits on neural identification of coloured surfaces in natural scenes. Visual Neurosci 21:331–336

Gomes C (2009) Computational sustainability. The Bridge, National Academy of Engineering 39(4):6–11

Goreinov SA, Tyrtyshnikov EE (2001) The maximum-volume concept in approximation by low-rank matrices. Contemp Math 280:47–51

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24(7):498–520

Kersting K, Wahabzada M, Thurau C, Bauckhage C (2010) Hierarchical convex NMF for clustering massive data. In: Proceedings of the 2nd Asian Conference on Machine Learning (ACML-10)

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–799

Lucas A, Klaassen P, Spreij P, Straetmans S (2003) Tail behaviour of credit loss distributions for general latent factor models. Appl Math Finance 10(4):337–357

MacKay D (2009) Sustainable energy—without the hot air. UIT Cambridge Ltd, Cambridge

Miao L, Qi H (2007) Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. IEEE T Geosci Remote 45(3):765–777

Nascimento JMP, Dias JMB (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE T Geosci Remote 43(4):898–910

Ostrouchov G, Samatova N (2005) On fastmap and the convex hull of multivariate data: toward fast and robust dimension reduction. IEEE T Pattern Anal 27(8):1340–1434

Sippl M, Sheraga H (1986) Cayley-Menger coordinates. Proc Natl Acad Sci 83(8):2283–2287

Spearman C (1904) General intelligence objectively determined and measured. Am J Psychol 15:201–293

Thurau C, Kersting K, Bauckhage C (2009) Convex non-negative matrix factorization in the wild. In: Proceedings of the IEEE International Conference on Data Mining, Miami

Thurau C, Kersting K, Wahabzada M, Bauckhage C (2010) Convex non-negative matrix factorization for massive datasets. Knowl Inf Syst (KAIS). doi:10.1007/s10115-010-0352-6

Winter ME (1999) N-FINDR: an algorithm for fast and autonomous spectral endmember determination in hyperspectral data. In: Proceedings of the International Conference on Applied Geologic Remote Sensing, Vancouver