

Robust Shape-regularized Non-negative Matrix Factorization for Real-time Source Apportionment

Manjyot Singh Nanra*
manjyots21@iitk.ac.in
IIT Kanpur
Kanpur, India

Sharanya Saha*
sharanya21@iitk.ac.in
IIT Kanpur
Kanpur, India

Ashutosh Kumar Shukla
akumars@iitk.ac.in
IIT Kanpur
Kanpur, India

Sachchida Nand Tripathi
snt@iitk.ac.in
IIT Kanpur
Kanpur, India

Purushottam Kar
purushot@cse.iitk.ac.in
IIT Kanpur
Kanpur, India

ABSTRACT

Discovering the anthropogenic sources of pollution is a key step in air quality monitoring and management. Real-time source apportionment (RTSA) is the globally accepted standard for this task and critically uses non-negative matrix factorization (NMF) that is made challenging by the non-convexity of the problem and noisy data. In this work, we develop a technique ROSE-NMF offering improvements over state-of-the-art RTSA and NMF solvers. ROSE-NMF offers provable convergence guarantees, allows diurnal patterns to be specified to guide the solver to an optimal solution, and is robust to row and column outliers. In multiple experiments, ROSE-NMF offered markedly improved performance over standard NMF solvers as well as commercial RTSA solvers. Code for ROSE-NMF is available at <https://github.com/purushottamkar/rose-nmf>.

CCS CONCEPTS

• Computing methodologies → Source separation; • Mathematics of computing → Nonconvex optimization; • Applied computing → Environmental sciences.

KEYWORDS

non-negative matrix factorization, shape regularization, robust learning, source apportionment, air-quality monitoring

ACM Reference Format:

Manjyot Singh Nanra, Sharanya Saha, Ashutosh Kumar Shukla, Sachchida Nand Tripathi, and Purushottam Kar. 2024. Robust Shape-regularized Non-negative Matrix Factorization for Real-time Source Apportionment. In *7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD) (CODS-COMAD 2024)*, January 04–07, 2024, Bangalore, India. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3632410.3632457>

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CODS-COMAD 2024, January 04–07, 2024, Bangalore, India

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1634-8/24/01...\$15.00
<https://doi.org/10.1145/3632410.3632457>

1 INTRODUCTION

Elevated atmospheric Particulate Matter (PM) levels are strongly associated with poor pulmonological and cardiac health outcomes and are a cause for concern for countries such as India that houses 3 of the top 10 most severely polluted cities in the world [1]. Measuring the ambient levels of PM_{2.5} (particulate matter of size 2.5 microns or smaller) is a common first step in detecting poor air quality. However, this alone is insufficient to guide policy action unless the human activities causing elevated PM_{2.5} levels are also identified. The process of identifying the anthropogenic and other sources of PM is known as Source Apportionment (SA).

Several techniques have emerged over the years to conduct SA. *Filter-based source apportionment* (FSA) involves analyzing the chemical composition of PM samples collected using filters that are composed of materials that can efficiently capture PM such as quartz, Teflon, or cellulose [2]. Pollutants present in the samples are identified by subjecting the filters to elemental analysis, X-ray fluorescence, or mass spectrometry [21]. However, FSA has several drawbacks: it takes several months to generate SA results, it cannot resolve short term episodic pollution events or diurnal characteristics since sample collection itself happens over several days to weeks. Un-denuded filter sampling can introduce artifacts leading to overestimated species concentration, and evaporation of semi-volatile compounds can lead to underestimation of concentration for species such as ammonium nitrate [4].

Real-time source apportionment (RTSA) differs from FSA by using instruments such as Aerosol Mass Spectrometers (AMS), Xact, Aethalometer, etc [34] that directly measure the chemical composition of aerosol particles [2, 42]. RTSA offers SA at high temporal resolutions, can detect intermittent pollution episodes and reveal diurnal characteristics of sources [3, 31]. Early works used 24-hourly samples but recent methods collect samples at much higher frequency (hourly or minutely) to offer fine-grained analysis and identify short-lived sources [39]. SA outputs are generated by subjecting the samples to non-negative matrix factorization (NMF).

Although a vast literature exists for NMF, real-life applications such as RTSA can still prove to be challenging owing to the NP-hardness of NMF [38] as well as noise, outliers in data. To address these concerns, this paper develops the ROSE-NMF method that offers a scalable and robust solution for RTSA applications.

Our Contributions. ROSE-NMF offers the following features:

1) It allows the user to guide the solver by specifying known diurnal

patterns in various SA factors, 2) It uses a simple yet effective alternating minimization strategy with provable convergence guarantees, 3) It continuously detects and rectifies outliers or corruptions in the samples to offer improved performance, and 4) In experiments, it offers noticeably superior performance to state-of-the-arts solvers and commercial solvers.

2 RELATED WORKS

The goal in NMF is to decompose a matrix X with non-negative entries into two matrices, say W, H , both with non-negative entries such that $X \approx WH$. The literature on NMF solvers is vast and diverse but comprehensive surveys are available [40].

Solver Strategies. Over the years, several solver strategies have been considered for solving the NMF problem including multiplicative updates, alternating optimization and coordinate descent [12, 15, 22]. Of special mention is the multi-linear engine (ME) line of solvers [27] which are popular for RTSA applications. In particular, the ME-2 algorithm [28] and its commercial implementation called SoFi (Source Finder) [8] is state-of-the-art for RTSA as it facilitates the selection of initial conditions for the ME-2 engine and the analysis of results. We experimentally compare ROSE-NMF to SoFi as well as other leading NMF solvers.

Regularization in NMF. Supervised NMF is an NMF variant wherein each data point is endowed with a class-label [41]. Although this variant does not apply to RTSA settings, solvers for supervised NMF often incorporate graph regularization [17, 41]. It is notable that ROSE-NMF incorporates shape regularization which is distinct from graph regularization. Moreover, the shape priors used by ROSE-NMF are more compact both in terms of space and time complexity. In particular, storing and applying a shape prior has space and time complexity linear in the number of data points whereas the same are quadratic for explicit graph priors.

Multi-factor NMF. Deep NMF [10] performs decomposition into more than two factors e.g., $X \approx W_1 W_2 W_3 H$ or introduces non-linearities e.g., $X \approx W_1 g(W_2 g(W_3 H))$ for some non-linear function g . Note that ROSE-NMF also performs a trivariate decomposition but with the third factor having a restricted diagonal structure. ROSE-NMF exploits this structure to make its updates more scalable which deep NMF methods are unable to do.

Theoretical Analyses. Convergence analyses for NMF solvers typically focus on monotonicity of progress [17, 22]. ROSE-NMF’s alternating minimization approach offers monotonicity by design and offers stronger, local convergence bounds. Several convergence analyses exist for general alternating minimization in literature [18, 24] but present specific deficiencies that make them unsuitable for ROSE-NMF. For example, the *Local Error Bound* in [24] is restrictive and demands that every first-order stationary point be a global optimum, while the analysis in [18] cannot handle constraints nor does it extend to trivariate or multivariate objectives. The analysis in this paper remedies these issues and offers a generic local convergence bound that extends to constrained, multivariate problems while making mild and realistic assumptions.

3 DATA COLLECTION AND PREPARATION

Deployments. High time resolution measurements of elemental components of PM_{2.5} were obtained [35] at two urban sites viz. the

Indian Institute of Technology, Delhi (IITD) (28.54°N, 77.19°E) and the Indian Institute of Tropical Meteorology, New Delhi (IITMD) (28.63°N, 77.17°E). At the IITD site, measurement was done during two seasons; Summer (June-July 2019) and winter (Oct 2019-Jan 2020) resulting in two datasets, namely IITD-S and IITD-W, while at the IITMD, measurements were taken only during winter season (Oct-Dec 2019) resulting in the dataset IITMD-W.

Instrumentation. The Xact 625i ambient continuous multi-metals monitor (Cooper Environmental, Beaverton OR, USA) was used to measure real-time elemental concentrations of aerosols. It operates by sampling ambient aerosols at a flow rate of 16.7 lpm and depositing it on a filter tape [16, 37] which is analyzed using energy dispersive X-ray fluorescence (EDXRF) to determine masses of 30 elemental species (see Table 1 for examples) at half-hourly intervals. The instrument also provides measurement uncertainty for each measurement [6] accounting for uncertainties in spectral deconvolution and sampling. Standard QA/QC steps were taken such as calibrating the flow rate and XRF calibration.

RTSA via NMF. The SoFi interface was used to obtain RTSA solutions. Standard procedures were followed, such as dynamically down-weighting point outliers [13]. SA Factors were selected based on the highest change in the Q/Q exp value, unexplained variation, factor profile, time series, diurnals, scaled residual, time series correlation with external markers and bootstrapped analysis. Further details of the RTSA execution can be found in [34, 35]

Final Datasets. For each deployment, the post-processed Xact measurements yielded 2 matrices, $X, E \in \mathbb{R}_+^{T \times M}$ where T is the total number of timestamps for which measurements were obtained and M is the number of elemental species measured at each timestamp (see Table 1 for examples of elemental species such as Ca, Fe, Si, etc). $X = [x_{tm}]$ is the matrix of the elemental masses of each of the species at each timestamp and $E = [\epsilon_{mt}]$ is the matrix of measurement uncertainties. On the other hand, the output of the RTSA process yielded, for each deployment, two matrices, namely $G \in \mathbb{R}_+^{T \times K}, F \in \mathbb{R}_+^{K \times M}$ where K is the number of SA factors identified (see Table 1 for examples of factors such as Fireworks, SFC1, Cl-rich, etc). Each column of the matrix G presents a time series showing the temporal variations of the concentration of that factor whereas each row of the matrix F (lovingly called the factor profile matrix) shows the mass spectrum profile of that factor.

4 PROBLEM SETTING

The goal of NMF is straightforward – given measurement and uncertainty matrices $X, E \in \mathbb{R}_+^{T \times M}$ as described above, and a number $K < M$ of the number of factors desired, we seek a decomposition of the form $X \approx GF$ where $G \in \mathbb{R}_+^{T \times K}, F \in \mathbb{R}_+^{K \times M}$. Note that the matrices X, E are guaranteed to have non-negative entries and the matrices G, F must be constrained to have non-negative entries. Most NMF solvers attempt to take into account the measurement uncertainties by solving a variant of the following:

$$\min_{G, F \geq 0} \sum_{t \in [T], m \in [M]} \left(\frac{x_{tm} - \langle g_t, f_m \rangle}{\epsilon_{mt}} \right)^2, \quad (1)$$

$g_t \in \mathbb{R}_+^K$ is the t^{th} row of G and $f_m \in \mathbb{R}_+^K$ is the m^{th} column of F .

Regularized and Constrained NMF. Various regularizers and constraints have been investigated to promote better solutions [12,

	Coal Com- bustion	Non- exhaust	Dust lated	re-	Fireworks	SFC1	SFC2	Cl-rich	Cu-rich	S-rich
IITD-S	✓	✓	✓			✓	✓	✓	✓	✓
IITD-W	✓		✓		✓	✓	✓	✓	✓	✓
IITMD-W	✓		✓		✓	✓	✓	✓	✓	✓
Dominant species	Pb, Se	Ba, Cr	Si, Ca, Fe	Al, Ba, Sr, K	K, As, Rb, Se, Sb	Zn, V, S	Cl, Br	Cu, Br	S	
Anthropo- genic sources	coal com- bustion	brake and tyre wear	road dust resuspen- sion	firework ignition	biomass or crop residue burning	smelting, coal, oil combustion	waste incinera- tion, steel processing	alloy making, electronic waste burn- ing	thermal power plants	
Diurnal character- istics			heavy traf- fic hours	Diwali period, evenings			sharp peak just after sunrise		always high	
References	[9, 19, 23]	[39]	[19, 36]	[14, 20, 30, 32]	[31, 34]	[9, 39]	[31, 34]	[31]	[26]	

Table 1: A summary of the datasets and RTSA results. 3 datasets were collected, namely IITD-S, IITD-W and IITMD-W. The columns correspond to various SA factors. The first 3 rows indicate which factors were resolved in each of the datasets. For example, the “Fireworks” factor was resolved only for IITD-W and IITMD-W since the festival of Diwali occurred during the data collection period of those two datasets. The next four rows indicate, respectively, the metallic species dominant for each factor, the major anthropogenic sources of each factor, any notable diurnal characteristics of that factor and references.

15, 22]. For instance, [8] allow the problem to be constrained using previously obtained solutions, say from literature. For example, if exemplar time series and factor profile matrices F_0, G_0 could be obtained, then the following variant could be solved:

$$\min_{-\alpha \leq \frac{G-G_0}{G_0} \leq \alpha, -\alpha \leq \frac{F-F_0}{F_0} \leq \alpha} \sum_{t \in [T], m \in [M]} \left(\frac{x_{mt} - \langle \mathbf{g}_t, \mathbf{f}_m \rangle}{\epsilon_{mt}} \right)^2, \quad (2)$$

where the quotients $\frac{F-F_0}{F_0}$ and $\frac{G-G_0}{G_0}$ are computed element-wise and $\alpha \in (0, 1)$ is a tolerance hyperparameter.

However, we argue that the form of regularization in (2) is too strict. For example, certain factors have known diurnal variations which can be inferred from past literature. For example, Table 1 indicates that the factor Cl-rich is known to have a peak just after sunrise i.e. the column in G corresponding to Cl-rich is expected to have large values on timestamps corresponding to early morning hours. However this does not mean that the concentration of Cl-rich must exactly match historical values obtained from literature. In fact, the post-sunrise concentration of Cl-rich may vary even day-to-day, making (2) too strict a formulation.

Shape Regularization. Instead of requiring the G matrix to match historical values exactly, we argue that it is much better to impose a *shape* regularization on the G matrix. To explain this, we first develop an alternate indexing. Let the T timestamps in the dataset be divided into D day blocks, each block corresponding to a single day with I timestamps in each block i.e. $T = DI$. For our datasets, $I = 48$ as samples were collected at 30 minute intervals. Thus, each time stamp t can be reindexed as (d, i) where d indicates the day and i indicates the time within that day. We are now ready to define the notion of a shape-scale decomposition.

DEFINITION 1 (SHAPE-SCALE DECOMPOSITION). Given a time series matrix $G \in \mathbb{R}_+^{T \times K}$, for any day $d \in [D]$, consider the sub-matrix $G^d \in \mathbb{R}_+^{I \times K}$ corresponding to rows $\{t : t = (d, i), i \in [I]\}$ of G . Then a shape-scale decomposition of G^d is a pair of matrices $U^d \in [0, 1]^{I \times K}$, $\Sigma^d \in \mathbb{R}_+^{K \times K}$ s.t. Σ^d is diagonal and $G^d = U^d \Sigma^d$.

Note a shape-scale decomposition can be uniquely identified by setting $\Sigma^d = \text{diag}(\sigma_1^d, \dots, \sigma_K^d)$ where $\sigma_k^d = \|\mathbf{g}^{d,k}\|_\infty$ and setting $U^d = G^d (\Sigma^d)^{-1}$ where $\mathbf{g}^{d,k}$ is the k^{th} column of G^d . We are now ready to present the ROSE-NMF algorithm.

5 ROBUST SHAPE-REGULARIZED NMF

The first key idea behind ROSE-NMF is to note that (2) regularizes G^d which is too restrictive and that we should regularize only the shape matrix i.e. U^d . To do so, we use a *shape prior* which is a matrix of the form $U_0 \in [0, 1]^{I \times M}$ and perform a *trivariate* decomposition. Note that this is unlike traditional NMF and (1), (2) all of which perform a bivariate decomposition. The scale matrix Σ^d is not regularized which allows the absolute concentrations of various factors to differ from historical values as well as day to day, so long as these variations have similar diurnal shapes. In the following, we have $U^d \in [0, 1]^{I \times K}$ and Σ^d is a diagonal matrix and $F \in \mathbb{R}_+^{K \times M}$ is the factor matrix as before, $\mathbf{u}_{(d,i)} \in \mathbb{R}_+^K$ is the i^{th} row of U^d and $\mathbf{f}_m \in \mathbb{R}_+^K$ is the m^{th} column of F . Note that we learn a separate U^d, Σ^d per day, but F is shared among all days.

$$\min_{\substack{U^d \in [0,1]^{I \times K}, \\ \Sigma^d \geq 0, F \geq 0}} \sum_{\substack{d \in [D], \\ i \in [I], m \in [M]}} \left(\frac{x_{m(d,i)} - \langle \mathbf{u}_{(d,i)}, \Sigma^d \mathbf{f}_m \rangle}{\epsilon_{m(d,i)}} \right)^2 + \frac{\lambda}{2} \cdot \sum_{d \in [D]} \|U^d - U_0\|_{\text{fro}}^2, \quad (3)$$

where $\|\cdot\|_{\text{fro}}$ is the Frobenius norm but we can use any strongly convex regularizer. We note that minor regularizers may also be introduced to U^d, F to condition the problem (see Section 6.1). We comment that ROSE-NMF can readily handle formulations such as (2) with constraints of the form $-\alpha \leq \frac{F-F_0}{F_0} \leq \alpha$ with minor modifications but we do not explore such variants in this paper. We now move on to developing a solver for (3). To avoid notational clutter, we will show the calculations using $\epsilon_{m(d,i)} = 1$ and note that general values of $\epsilon_{m(d,i)}$ will merely change certain least squares problems derived below into weighted least squares problems. Before we conclude this discussion to present the algorithm, we note that the trivariate problem in (3) cannot be trivially reduced to a bivariate problem by absorbing the scale matrix Σ^d into the factor profile matrix F since Σ^d varies from day to day while F is shared.

Alternating Minimization. We rewrite (3) more compactly:

$$\min_{\substack{U^d \in [0,1]^{I \times K}, \\ \Sigma^d \geq 0, F \geq 0}} \sum_{\substack{d \in [D], \\ i \in [I]}} \left(\mathbf{x}_{(d,i)} - \langle \mathbf{u}_{(d,i)}, \Sigma^d F \rangle \right)^2 + \frac{\lambda}{2} \sum_{d \in [D]} \|U^d - U_0\|_{\text{fro}}^2,$$

where $\mathbf{x}_{(d,i)} \in \mathbb{R}_+^M$ is the vector of measurements on day d , timestamp $i \in [I]$. We do not optimize this directly as it is an NP-hard problem in general [38]. Instead, we notice that if we create 3 sets of variables $\{U^d\}_{d \in [D]}, \{\Sigma^d\}_{d \in [D]}, \{F\}$, then, upon fixing any two sets, finding the optimal value of variables in the remaining set is simple. We justify this using derivations below. This gives us the following simple recipe for optimizing objective (3):

- (1) Fix $\{\Sigma^d\}_{d \in [D]}, \{F\}$ and update $\{U^d\}_{d \in [D]}$.
- (2) Fix $\{U^d\}_{d \in [D]}, \{F\}$ and update $\{\Sigma^d\}_{d \in [D]}$.
- (3) Fix $\{U^d\}_{d \in [D]}, \{\Sigma^d\}_{d \in [D]}$ and update $\{F\}$.
- (4) Repeat epochs (steps 1 \rightarrow 2 \rightarrow 3) until convergence

Update U^d . Upon fixing $\{\Sigma^d\}_{d \in [D]}, \{F\}$, for any $d \in [D]$, the optimal value of $\mathbf{u}_{(d,i)}$, i.e. the i^{th} row of U^d , can be found as:

$$\min_{\mathbf{u} \in [0,1]^K} \left(\mathbf{x}_{(d,i)} - \langle \mathbf{u}, \Sigma^d F \rangle \right)^2 + \frac{\lambda}{2} \cdot \|\mathbf{u} - \mathbf{u}_0^i\|_{\text{fro}}^2,$$

where \mathbf{u}_0^i is the i^{th} row of U_0 . This is a ridge regression problem with box constraints which is a convex problem for which standard solvers exist. The update to F upon fixing $\{U^d, \Sigma^d\}_{d \in [D]}$ is similar.

Update Σ^d . Upon fixing $\{U^d\}_{d \in [D]}, \{F\}$, Σ^d can be found as:

$$\min_{\Sigma^d \geq 0} \sum_{i \in [I], m \in [M]} \left(x_{m(d,i)} - \langle \mathbf{u}_{(d,i)}, \Sigma^d \mathbf{f}_m \rangle \right)^2$$

As Σ^d is diagonal hence symmetric, the properties of the trace operator give $\langle \mathbf{u}_{(d,i)}, \Sigma^d \mathbf{f}_m \rangle = \text{trace}(\Sigma^d \mathbf{f}_m \mathbf{u}_{(d,i)}^\top) = \langle \boldsymbol{\sigma}^d, \mathbf{z}^{(m,di)} \rangle$ where $\mathbf{z}^{(m,di)} \stackrel{\text{def}}{=} \text{diag}(\mathbf{f}_m \mathbf{u}_{(d,i)}^\top) \in \mathbb{R}_+^K$ and $\boldsymbol{\sigma}^d \stackrel{\text{def}}{=} \text{diag}(\Sigma^d) \in \mathbb{R}_+^K$. This allows us to rewrite the problem as a simple least squares problem with box constraints

$$\min_{\boldsymbol{\sigma}^d \geq 0} \sum_{i \in [I], m \in [M]} \left(x_{m(d,i)} - \langle \boldsymbol{\sigma}^d, \mathbf{z}^{(m,di)} \rangle \right)^2.$$

5.1 Corrective Updates for Robust Learning

Upon convergence of the above alternating minimization process, we notice several timestamps (d, i) with poor reconstruction i.e.

$\|\mathbf{x}_{(d,i)} - \langle \mathbf{u}_{(d,i)}, \Sigma^d F \rangle\|_2 \gg 0$ or else some row or column of some U^d deviating significantly from that in U^0 . Given the brittleness of the least squares estimator [11, 25], it is likely that in order to fit these outlier rows, a compromise value of the factor matrix F (as F is shared) was learnt which in turn gave us suboptimal values of U^d, Σ^d for non-outlier days too. We address this by executing Phase II and III of ROSE-NMF where these outlier rows are either corrected or excluded from the training process.

Handling row corruptions in X . Rows of X were sorted in decreasing order of $\|\mathbf{x}_{(d,i)} - \langle \mathbf{u}_{(d,i)}, \Sigma^d F \rangle\|_2$ and the Kneedle algorithm [33] was used to select an appropriate number of rows of X which to consider corrupted. For each such row, say (d, i) , a temporary correction was made by setting $\mathbf{x}_{(d,i)} = \langle \mathbf{u}_{(d,i)}, \Sigma^d F \rangle$.

Handling row corruptions in U . Rows of U were sorted in decreasing order of $\|\mathbf{u}_{(d,i)} - \mathbf{u}_0^i\|_2$ and the Kneedle algorithm was used to select an appropriate number of rows of U which to consider corrupted. For each such row, say (d, i) , a temporary correction was made by setting $\mathbf{u}_{(d,i)} = \mathbf{u}_0^i$ and $\mathbf{x}_{(d,i)} = \langle \mathbf{u}_0^i, \Sigma^d F \rangle$.

Handling column corruptions in U . Corrupted columns of U were similarly identified by looking at their distance from the corresponding column in U_0 . Days on which such column corruptions occurred were excluded from training.

After handling corruptions, regular alternations updating the parameters U^d, Σ^d, F in turn were carried out till convergence – this constituted one phase. Multiple such phases were executed. However in all but the last phase, only row corruptions in X/U were considered. In the last phase, column corruptions in U were considered in addition to row corruptions in X/U . Note that the set of rows considered corrupted was recomputed after each phase since the change in the factor profile and time series matrices could have caused some other rows to get revealed as corrupted. Final evaluation was done after undoing all such corrections to X .

6 THEORETICAL ANALYSIS

In this section, we present a convergence bound for ROSE-NMF. We present a generic bound which is instantiated to ROSE-NMF in Section 6.1. We first introduce the concept of marginally strongly convex/smooth functions and marginal optimality. These are adapted from [18] but modified to handle multivariate and constrained optimization problems. Note that Defns 2 and 3 are stated for trivariate functions but readily extend to arbitrary multivariate functions.

DEFINITION 2 (MARGINAL STRONG CONVEXITY/SMOOTHNESS). *A continuously differentiable function $f : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \rightarrow \mathbb{R}$ in three variables over the constraint set $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is considered α -marginally strongly convex (MSC) and β -marginally strongly smooth (MSS) in its first variable if for every value of $\mathbf{y}, \mathbf{z} \in \mathcal{Y} \times \mathcal{Z}$, the (univariate) function $f(\cdot, \mathbf{y}, \mathbf{z}) : \mathcal{X} \rightarrow \mathbb{R}$ is α -strongly convex and β -strongly smooth, i.e., for every $\mathbf{x}^1, \mathbf{x}^2 \in \mathcal{X}$, we have*

$$\frac{\alpha}{2} \|\mathbf{x}^2 - \mathbf{x}^1\|_2^2 \leq f(\mathbf{x}^2, \mathbf{y}, \mathbf{z}) - f(\mathbf{x}^1, \mathbf{y}, \mathbf{z}) - \langle \mathbf{g}, \mathbf{x}^2 - \mathbf{x}^1 \rangle \leq \frac{\beta}{2} \|\mathbf{x}^2 - \mathbf{x}^1\|_2^2,$$

where $\mathbf{g} \stackrel{\text{def}}{=} \nabla_{\mathbf{x}} f(\mathbf{x}^1, \mathbf{y}, \mathbf{z})$ is the partial derivative of f w.r.t its first variable at $(\mathbf{x}^1, \mathbf{y}, \mathbf{z})$. A similar condition is imposed for f to be considered MSC/MSS in its second and third variables.

DEFINITION 3 (MARGINAL OPTIMALITY). Let f be a function of three variables over the constraint set $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For any $\mathbf{y}, \mathbf{z} \in \mathcal{Y} \times \mathcal{Z}$, we say that $\tilde{\mathbf{x}} \in \mathcal{X}$ is a marginally optimal coordinate with respect to \mathbf{y}, \mathbf{z} and use the shorthand $\tilde{\mathbf{x}} \in \text{mOPT}_f(\mathbf{y}, \mathbf{z})$, if $f(\tilde{\mathbf{x}}, \mathbf{y}, \mathbf{z}) \leq f(\mathbf{x}, \mathbf{y}, \mathbf{z})$ for all $\mathbf{x} \in \mathcal{X}$. We similarly extend the notion of marginal optimality to the second and third variables.

It is important to note that marginally strongly convex functions need not even be convex and thus standard convex optimization analyses [5] do not apply to such functions. For example, the bivariate function $f(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} x\mathbf{y} + x^2/4 + y^2/4$ satisfies MSC but is non-convex. Instead, for such optimization problems, it is popular to adopt alternating minimization as indeed ROSE-NMF does. To start the analysis, we first specify our assumptions.

Simplified Notation. We will use $\Theta \stackrel{\text{def}}{=} \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ as a shorthand for the constraint set and $\theta \stackrel{\text{def}}{=} (\mathbf{x}, \mathbf{y}, \mathbf{z})$ to denote a point.

ASSUMPTION 1 (STATIONARITY AT OPTIMA). For a constrained optimization problem with objective f and constraint set Θ , let $\theta^* \stackrel{\text{def}}{=} \arg \min_{\theta \in \Theta} f(\theta)$ be the global optimum. Then $\nabla f(\theta^*) = \mathbf{0}$.

We note that this is a relatively mild assumption and is always satisfied if θ^* lies in the relative interior of the feasible set Θ (using elementary results, say [7, Proposition 1.2]).

ASSUMPTION 2 (ROBUST STABILITY). For an objective function f and constraint set Θ , let $\theta^* \stackrel{\text{def}}{=} (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \Theta$ be the global optimum. Then, f satisfies the C -robust stability property in the region $\hat{\Theta} \subseteq \Theta$ if for some $C > 0$, we have for any $\theta \stackrel{\text{def}}{=} (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \hat{\Theta}$,

$$f(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}) - 3f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq C \cdot (f(\mathbf{x}, \mathbf{y}, \mathbf{z}) - f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}) - \text{err}),$$

where $\tilde{\mathbf{x}} \in \text{mOPT}_f(\mathbf{y}, \mathbf{z})$, $\tilde{\mathbf{y}} \in \text{mOPT}_f(\tilde{\mathbf{x}}, \mathbf{z})$ and $\tilde{\mathbf{z}} \in \text{mOPT}_f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ and $\text{err} \stackrel{\text{def}}{=} \langle \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}, \mathbf{y}, \mathbf{z}), \mathbf{x} - \tilde{\mathbf{x}} \rangle + \langle \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \mathbf{z}), \mathbf{y} - \tilde{\mathbf{y}} \rangle + \langle \nabla_{\mathbf{z}} f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}), \mathbf{z} - \tilde{\mathbf{z}} \rangle$.

The dense notation aside, Assumption 2 posits a simple proposition. The right hand side estimates the amount by which a single set of alternating optimization steps can decrease the objective value when starting at $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Lemma 2 below shows that the left hand side estimates how far is $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ from the optimum $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$. Thus, Assumption 2 merely states that if it gets challenging to decrease the objective value by a large amount using alternating optimization, we must be close to the optimum.

Ensuring Assumption 2 globally can be challenging but Lemma 1 shows that it is possible to ensure it in a region Θ close to the optimum. Assumption 2 is an extension of a similar assumption in [18] but can handle constrained optimization problems (since arbitrary constraint sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are assumed) unlike [18] which assumes the constraint set to be like $\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^r$ i.e., no constraints.

LEMMA 1. Suppose f takes bounded values i.e. $|f(\mathbf{x}, \mathbf{y}, \mathbf{z})| \leq B$ for all $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \Theta$ and satisfies α -MSC and β -MSS with respect to all its variables, and satisfies Assumption 1. Then, for any $r > 0$, f satisfies C -robust stability in a ball of radius r around the optima i.e., $\hat{\Theta} \stackrel{\text{def}}{=} \{\theta : \|\theta - \theta^*\|_2^2 \leq r\}$, with $C = \frac{\beta r}{4B}$.

PROOF. Applying MSS and Assumption 1 at the optimum for the first coordinate gives $f(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) - f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq \frac{\beta}{2} \|\mathbf{x} - \mathbf{x}^*\|_2^2$.

Doing the same for other coordinates and adding the inequalities gives $f(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}) - 3f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq \frac{\beta}{2} \|\theta - \theta^*\|_2^2$, where $\theta \stackrel{\text{def}}{=} (\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $\theta^* \stackrel{\text{def}}{=} (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$. Now, since we defined $\tilde{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \mathbf{y}, \mathbf{z})$, standard results on constrained optimization (say [7, Proposition 1.3]) tell us that we must have $\langle \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}, \mathbf{y}, \mathbf{z}), \mathbf{x} - \tilde{\mathbf{x}} \rangle \geq 0$. Applying this to the other coordinates tells us that $\text{err} > 0$. Thus, C -robust stability is satisfied if $\frac{\beta}{2} \|\theta - \theta^*\|_2^2 \leq C \cdot (f(\mathbf{x}, \mathbf{y}, \mathbf{z}) - f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})) \leq 2CB$ by using the bound on f . This tells us that for any radius $r > 0$, C robust stability is satisfied in the region $\hat{\Theta} \stackrel{\text{def}}{=} \{\theta : \|\theta - \theta^*\|_2^2 \leq r\}$ for any $C > \frac{\beta r}{4B}$ which finishes the proof. \square

LEMMA 2. For a function f satisfying α -MSC, β -MSS and Assumptions 1 and 2, we have $\|\theta - \theta^*\|_2^2 \leq \frac{C\beta}{\alpha} \|\theta - \tilde{\theta}\|_2^2$ where $\tilde{\theta} \stackrel{\text{def}}{=} (\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$.

PROOF. Repeated applications of MSC and Assumption 1 yield $f(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}, \mathbf{z}^*) + f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}) - 3f(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \geq \frac{\alpha}{2} \|\theta - \theta^*\|_2^2$. Repeated applications of MSS and a telescoping sum yield $f(\mathbf{x}, \mathbf{y}, \mathbf{z}) - f(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}) - \text{err} \leq \frac{\beta}{2} \|\theta - \tilde{\theta}\|_2^2$, finishing the proof using Assm 2. \square

Theorem 3 below presents the convergence analysis for alternating optimization that offers an exponential rate of convergence. This was observed in experiments as well when ROSE-NMF was found to converge within a very few number of iterations.

THEOREM 3. For a constrained problem with objective f , constraint set Θ and optimum θ^* that satisfies α -MSC, β -MSS, and Assumptions 1 and 2, if alternating minimization is initialized at a point $\theta^0 \in \hat{\Theta}$, then for any $\epsilon > 0$, within $T = O\left(\log \frac{1}{\epsilon}\right)$ iterations, it offers an iterate θ^T such that $f(\theta^T) \leq f(\theta^*) + \epsilon$.

PROOF. Let $\theta \stackrel{\text{def}}{=} (\mathbf{x}, \mathbf{y}, \mathbf{z})$ be an iterate of the algorithm. Alternating minimization performs successive marginal optimizations $\mathbf{x}^+ \in \text{mOPT}_f(\mathbf{x}, \mathbf{z})$, $\mathbf{y}^+ \in \text{mOPT}_f(\mathbf{x}^+, \mathbf{z})$, $\mathbf{z}^+ \in \text{mOPT}_f(\mathbf{x}^+, \mathbf{y}^+)$ and constructs the next iterate as $\theta^+ \stackrel{\text{def}}{=} (\mathbf{x}^+, \mathbf{y}^+, \mathbf{z}^+)$. Let $\phi \stackrel{\text{def}}{=} f(\theta) - f(\theta^*)$ and $\phi^+ \stackrel{\text{def}}{=} f(\theta^+) - f(\theta^*)$. Applying MSS and Assumption 1 gives

$$\phi^+ \leq \frac{\beta}{2} \|\theta^+ - \theta^*\|_2^2 \leq \frac{\beta}{2} \|\theta^+ - \theta\|_2^2$$

Applying MSC on the various iterations of the algorithm tells us

$$f(\mathbf{x}, \mathbf{y}, \mathbf{z}) - f(\mathbf{x}^+, \mathbf{y}, \mathbf{z}) \geq \langle \nabla_{\mathbf{x}} f(\mathbf{x}^+, \mathbf{y}, \mathbf{z}), \mathbf{x} - \mathbf{x}^+ \rangle + \frac{\alpha}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2$$

However, since $\mathbf{x}^+ \in \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \mathbf{y}, \mathbf{z})$, standard results (say [7, Proposition 1.3]) tell us that $\langle \nabla_{\mathbf{x}} f(\mathbf{x}^+, \mathbf{y}, \mathbf{z}), \mathbf{x} - \mathbf{x}^+ \rangle \geq 0$ which gives $f(\mathbf{x}, \mathbf{y}, \mathbf{z}) - f(\mathbf{x}^+, \mathbf{y}, \mathbf{z}) \geq \frac{\alpha}{2} \|\mathbf{x}^+ - \mathbf{x}\|_2^2$. Similarly, we get $f(\mathbf{x}^+, \mathbf{y}, \mathbf{z}) - f(\mathbf{x}^+, \mathbf{y}^+, \mathbf{z}) \geq \frac{\alpha}{2} \|\mathbf{y}^+ - \mathbf{y}\|_2^2$ as well as $f(\mathbf{x}^+, \mathbf{y}^+, \mathbf{z}) - f(\mathbf{x}^+, \mathbf{y}^+, \mathbf{z}^+) \geq \frac{\alpha}{2} \|\mathbf{z}^+ - \mathbf{z}\|_2^2$. A telescoping sum of these inequalities gives $\phi - \phi^+ = f(\theta) - f(\theta^+) \geq \frac{\alpha}{2} \|\theta^+ - \theta\|_2^2$. Hereon, the analysis is similar to [18, Thm 4.3]. We get $\phi^+ \leq \frac{\beta}{2} \|\theta^+ - \theta^*\|_2^2 \leq \beta \left(\|\theta^+ - \theta\|_2^2 + \|\theta - \theta^*\|_2^2 \right)$ i.e. $\phi^+ \leq \beta(1 + C\kappa) \|\theta^+ - \theta\|_2^2 \leq 2\kappa(1 + C\kappa)(\phi - \phi^+)$, where $\kappa \stackrel{\text{def}}{=} \beta/\alpha$ and we used $(a + b)^2 \leq 2(a^2 + b^2)$ and Lemma 2. Simple manipulations yield $\phi^+ \leq \eta \cdot \phi$, where $\eta \stackrel{\text{def}}{=} \frac{2\kappa(1+C\kappa)}{1+2\kappa(1+C\kappa)} < 1$, finishing the proof. \square

6.1 Instantiating Theorem 3 to ROSE-NMF

We notice that the alternating updates derived in Section 5 show that the objective function in (3) is marginally convex. For U^d , the objective is marginally strongly convex due to the shape regularization. Marginal strong convexity is more challenging to guarantee with respect to the variables Σ^d and F but can be immediately ensured by adding a tiny regularization of the form $\lambda_0 \cdot (\|F\|_F^2 + \sum_{d \in [D]} \|\Sigma^d\|_F^2)$ for some small value of λ_0 that does not distort the objective too much. Marginal strong smoothness is easier to ensure since the objective function is doubly differentiable with all quantities bounded. Assumption 1 can be ensured if there are no degenerate solutions i.e., the optimal values for U^d, Σ^d, F do not have any 0 coordinates. Assumption 2 can be ensured by initializing close-enough to the optimum and using Lemma 1.

7 EXPERIMENTAL RESULTS

Experimental Setup. Steps outlined in Section 3 were carried out to obtain the 30-species measurement from the Xact instrument. Species for which more than 80% measurements were below minimum detection limit (MDL) were removed altogether. For various datasets (see Table 4 for dataset details) this left 22-23 species for RTSA analysis. All 3 datasets offered on an average of 40–44 timestamps per day. All experiments were carried out on a 64-bit machine with Intel® Core™ i7-6500U CPU @ 2.50GHz, 4 cores, 16 GB RAM and Ubuntu 16.04 OS. Code for ROSE-NMF is available at <https://github.com/purushottamkar/rose-nmf>.

Shape Prior. The shape prior matrix U_0 was obtained by taking all the G_d matrices offered by the ME-2 solution, taking coordinate-wise median and then doing a shape-scale decomposition. Exploring alternate sources of shape, such as mining published literature, is a topic of ongoing and future work.

Initialization. Two separate threads of experiments were carried out, one where ROSE-NMF was initialized using an NMF solution obtained using the popular scikit-learn library [29], and another one in which ROSE-NMF was initialized using a solution by the ME-2 algorithm obtained via the SoFi package. The `max_iter` parameter for sklearn was set to a large value 100,000 to ensure that it achieved convergence. Note that this did not necessarily result in a prohibitive runtime for NMF since the coordinate descent solver was chosen that offers fairly inexpensive iterations. In either case, the initialization algorithm offered G, F matrices. The factor matrix F was taken directly as initialization, while the day-wise time series matrices G_d were subjected to shape-scale decomposition to get an initialization for Σ^d .

Evaluation. All competitor methods were evaluated on their reconstruction error as well as value of the overall objective i.e., including the regularizer. As Figure 1 shows, ROSE-NMF offers good convergence in terms of both reconstruction and overall error. Note that evaluation for ROSE-NMF was done on original ground truth, i.e., after undoing all corrections to X .

Outlier Removal Handicap. Since ROSE-NMF benefits significantly from correcting certain outlier rows, it may be argued that removing a select few rows would improve performance of other algorithms as well. Table 2 tests this by allowing all algorithms to set aside a certain number of rows from evaluation that they

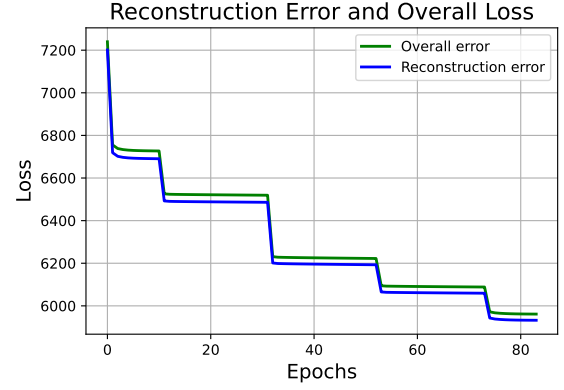


Figure 1: A plot showing the convergence of ROSE-NMF across 4 phases with each phase making corrections/removals for outliers followed by alternating updates for U^d, Σ^d, F . Notice the sharp decline in both reconstruction and overall loss after the completion of each phase. Although the curves may look flat within each phase, there is a gentle downward slope that is not appreciable clearly in this zoomed-out plot.

consider corrupted. The number of such rows was kept the same for all algorithms and is mentioned in Table 3. Unsurprisingly, the performance of all algorithms improved but as Table 2 shows, ROSE-NMF continues to offer upto 15% smaller reconstruction error no matter how it is initialized or whether outlier removal is done or not. This indicates that the retraining that ROSE-NMF does in each phase after correcting/removing outliers is crucial to learning better factor matrices F which in turn, allow better time series matrices U^d to be learnt. It is also interesting that despite the NMF solution being so poor initially, ROSE-NMF with NMF initialization often beats the commercial solver ME-2 by a healthy margin.

Comparison of Reconstruction Error. Figures 3 and 4 show the difference in reconstruction errors for a certain row of the matrix X . ROSE-NMF offers reconstruction errors 3 orders of magnitude smaller than NMF and comparable or smaller than ME-2.

Detection of Outliers. Figure 2 shows several instances where ROSE-NMF was able to detect departures from the prior shape matrix. For instance, it detected a departure from the usual pattern of the fireworks factor which usually peaks in the later hours of the day but for one day, the solver seemed to suggest significant firework contribution in the mid-morning hours, indicating some outlier behavior. Rectifying/removing such instances allows ROSE-NMF to learn better factor profiles and better time series matrices.

Better agreement with Shape Prior. Figures 5 and 6 show that ROSE-NMF offers improved agreement with the shape prior than both competitor models. In Figure 5, for the factors Cl-rich, SFC2 and Fireworks, ROSE-NMF offered total squared error of (0.001, 0.001, 0.007) respectively whereas NMF offered (0.013, 0.009, 0.03). In Figure 6, for the factors Fireworks, Coal Combustion and SFC2, ROSE-NMF offered total squared error of (0.01, 0.01, 0.01) respectively whereas ME-2 offered (0.02, 0.02, 0.03). Note that for Figure 5, ROSE-NMF was initialized using NMF while for Figure 6, ROSE-NMF was initialized using ME-2 itself. The plots show that

	Dataset	NMF Solution	ME-2 Solution	ROSE-NMF		% Improvement
				NMF Initialized	ME-2 Initialized	
IITD-W	Entire Dataset	9318059	7122	6729	6682	5.52%
	Post Outlier Removal	7776364	5896	5667	5858	3.88%
IITD-S	Entire Dataset	24065922	5711	4850	4877	15.1%
	Post Outlier Removal	20084291	3978	3588	3581	9.8%
IITMD-W	Entire Dataset	46510837	6980	6588	6626	5.62%
	Post Outlier Removal	8543806	5689	5573	5652	2.03%

Table 2: Comparison of reconstruction errors for ROSE-NMF vs NMF (from sklearn) and ME-2 (from SoFi) across the 3 datasets. The last column reports the relative improvement ROSE-NMF offered over its nearest competitor. ROSE-NMF was initialized using both NMF and ME-2. The NMF solver was offered sufficiently large number of iterations to ensure that it converged. Results are reported on the entire dataset as well as after outlier removal. The competitors were offered a handicap where they could choose to not be evaluated on a certain number of rows on which they considered outliers. ROSE-NMF continues to dominate irrespective of initialization or outlier removal, offering 2–15% improvement in reconstruction error.

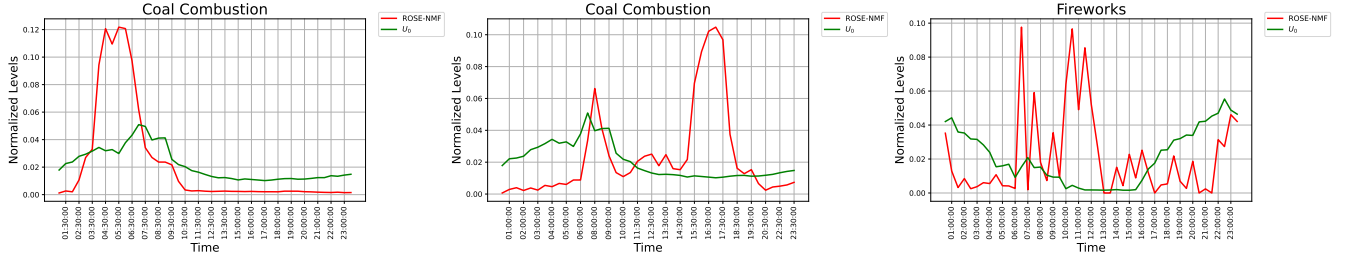


Figure 2: Illustrations of column corruptions detected by ROSE-NMF. In each case, a certain factor exhibited a marked departure from the shape/trend specified by U_0 . Removal of such outlier days from training allows learning better factor profile matrices.

Dataset	NMF-Intialized		ME2-Intialized	
	Rows in U	Rows in X	Rows in U	Rows in X
IITD-W	160	154	169	113
IITD-S	109	108	72	111
IITMD-W	105	106	83	146

Table 3: The number of rows identified for correction by ROSE-NMF due to either U corruption or X corruption. Results are shown for ROSE-NMF initialized both ways. Recall that final evaluation is done on the original ground truth matrix X and not the corrected one.

Dataset	Count			
	# Days (D)	# Times-tamps (T)	# Factors (K)	# Species (M)
IITD-S	46	1861	8	22
IITD-W	103	4453	8	23
IITMD-W	93	4137	8	22

Table 4: Dataset Statistics

ROSE-NMF is able to take an initial solution and offer improved agreement in terms of shape and diurnal variation of factors.

8 CONCLUSION

This paper presented ROSE-NMF, a scalable routine for performing NMF decomposition in the presence of outliers, and using helpful advice in the form of shape priors for one of the factors. ROSE-NMF enjoys provable convergence guarantees and in experiments, outperformed not only popular solvers such as sklearn, but also

commercial ones like ME-2. ROSE-NMF leaves several avenues of future work open, such as better ways of handling column corruptions in U^d as ROSE-NMF does not attempt to correct these, instead choosing to omit those rows from training. Further experimentation with RTSA using data from other instruments such as AMS and Black Carbon Aethalometer (BCA) is also an interesting prospect.

ACKNOWLEDGMENTS

The authors acknowledge financial support received under the Centre of Excellence in Advanced Technologies for Monitoring Air-quality iNdicators (ATMAN) approved by the office of the PSA,

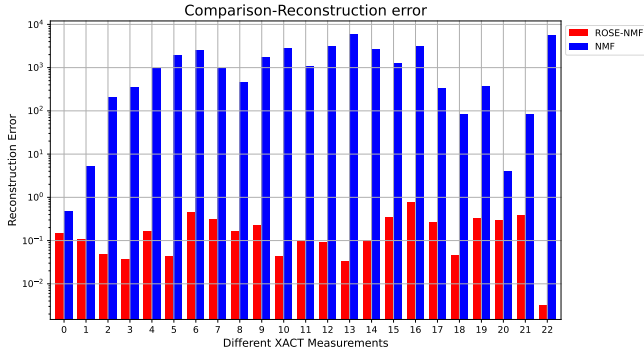


Figure 3: Comparison of reconstruction errors for a specific row of the X matrix on November 28, 2019, at 17:30 using NMF and ROSE-NMF. ROSE-NMF offers reconstruction errors upto 3 orders of magnitude smaller than NMF across species.

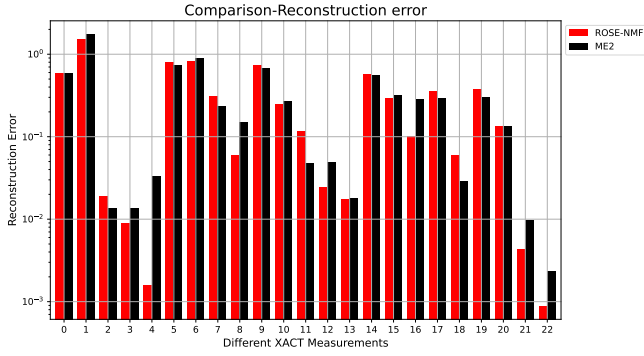


Figure 4: Comparison of reconstruction errors for a specific row of the X matrix on October 30, 2019, at 13:00 using ME-2 and ROSE-NMF solutions. ROSE-NMF offers comparable or better reconstruction than ME-2 for most species.

Government of India from Bloomberg Philanthropies, the Open Philanthropy and the Clean Air Fund. P.K. thanks Microsoft Research India and Tower Research for research grants.

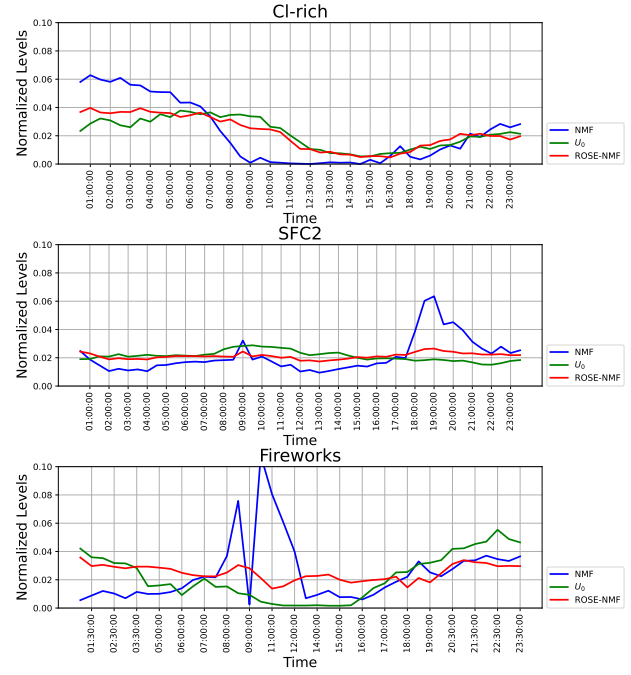


Figure 5: A comparison between NMF (sklearn) and ROSE-NMF in terms of agreement between U^d and U_0 for various factors. ROSE-NMF offers better agreement with U_0 than NMF as measured by total squared error (values in text).

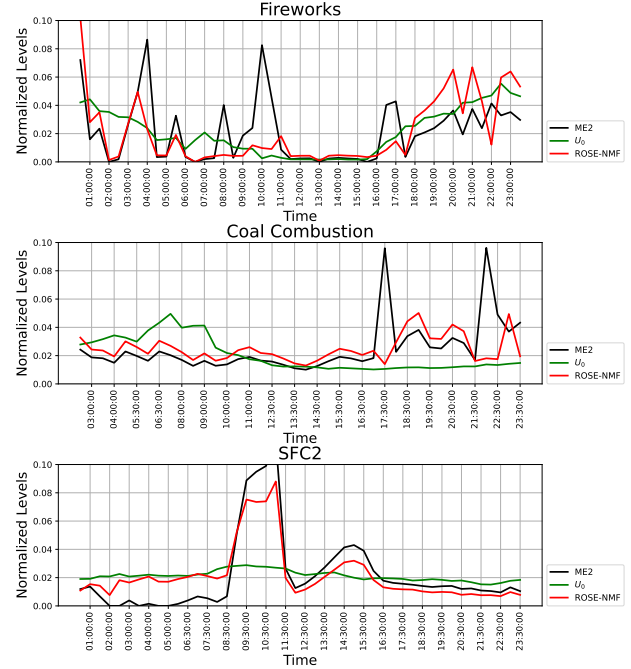


Figure 6: A comparison between ME-2 (SoFi) and ROSE-NMF in terms of agreement between U^d and U_0 for various factors. ROSE-NMF offers better agreement with U_0 compared to ME-2 as measured by total squared error (values in text).

REFERENCES

- [1] Costanza Acciai, Zhenyi Zhang, Fenjuan Wang, Zhangxiong Zhong, and Giovanni Lonati. 2017. Characteristics and source Analysis of trace Elements in PM_{2.5} in the Urban Atmosphere of Wuhan in Spring. *Aerosol and Air Quality Research* 17, 9 (2017), 2224–2234. <https://doi.org/10.4209/aaqr.2017.06.0207>
- [2] CA Belis, Michail Pikridas, Franco Lucarelli, E Petralia, Fabrizia Cavalli, Giulia Calzolari, M Berico, and J Sciare. 2019. Source apportionment of fine PM by combining high time resolution organic and inorganic chemical composition datasets. *Atmospheric Environment: X* 3 (2019), 100046.
- [3] Sahil Bhandari, Shahzad Gani, Kanan Patel, Dongyu S Wang, Prashant Soni, Zainab Arub, Gazala Habib, Joshua S Apte, and Lea Hildebrandt Ruiz. 2020. Sources and atmospheric dynamics of organic aerosol in New Delhi, India: insights from receptor modeling. *Atmospheric Chemistry and Physics* 20, 2 (2020), 735–752.
- [4] H. S. Bhowmik, A. Shukla, V. Lalchandani, J. Dave, N. Rastogi, M. Kumar, V. Singh, and S. N. Tripathi. 2022. Inter-comparison of online and offline methods for measuring ambient heavy and trace elements and water-soluble inorganic ions (NO_3^- , SO_4^{2-} , NH_4^+ , and Cl^-) in PM_{2.5} over a heavily polluted megacity, Delhi. *Atmospheric Measurement Techniques* 15, 9 (2022), 2667–2684. <https://doi.org/10.5194/amt-15-2667-2022>
- [5] Stephen Boyd and Lieven Vandenbergh. 2004. *Convex Optimization*. Cambridge University Press.
- [6] Steven G Brown, Shelly Eberly, Pentti Paatero, and Gary A Norris. 2015. Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and guidance on reporting PMF results. *Science of the Total Environment* 518 (2015), 626–635.
- [7] Sebastian Bubeck. 2015. Convex Optimization: Algorithms and Complexity. *Foundations and Trends® in Machine Learning* 8, 34 (2015), 231–357.
- [8] Francesco Canonaco, Monica Crippa, Jay Gates Slowik, Urs Baltensperger, and André SH Prévôt. 2013. SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data. *Atmospheric Measurement Techniques* 6, 12 (2013), 3649–3661.
- [9] Yunhua Chang, Kan Huang, Mingjie Xie, Congrui Deng, Zhong Zou, Shoudong Liu, and Yanlin Zhang. 2018. First long-term and near real-time measurement of trace elements in China's urban atmosphere: temporal variability, source apportionment and precipitation effect. *Atmospheric Chemistry and Physics* 18, 16 (2018), 11793–11812.
- [10] Wen-Sheng Chen, Qianwen Zeng, and Binbin Pan. 2022. A survey of deep nonnegative matrix factorization. *Neurocomputing* 491 (2022), 305–320. <https://doi.org/10.1016/j.neucom.2021.08.152>
- [11] Yudong Chen, Constantine Caramanis, and Shie Mannor. 2013. Robust Sparse Regression under Adversarial Corruption. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*.
- [12] Andrzej Cichocki and Anh-Huy Phan. 2009. Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* E92.A, 3 (2009), 708–721. <https://doi.org/10.1587/transfun.E92.A.708>
- [13] M. Crippa, F. Canonaco, J. G. Slowik, I. El Haddad, P. F. DeCarlo, C. Mohr, M. F. Heringa, R. Chirico, N. Marchand, B. Temime-Roussel, E. Abidi, L. Poulain, A. Wiedensohler, U. Baltensperger, and A. S. H. Prévôt. 2013. Primary and secondary organic aerosol origin by combined gas-particle phase source apportionment. *Atmospheric Chemistry and Physics* 13, 16 (2013), 8411–8426.
- [14] Dabrina D Dutcher, Kevin D Perry, Thomas A Cahill, and Scott A Copeland. 1999. Effects of indoor pyrotechnic displays on the air quality in the Houston Astrodome. *Journal of the Air & Waste Management Association* 49, 2 (1999), 156–160.
- [15] Cédric Févotte and Jérôme Idier. 2011. Algorithms for Nonnegative Matrix Factorization with the β -Divergence. *Neural Computation* 23, 9 (09 2011), 2421–2456. https://doi.org/10.1162/NECO_a.00168
- [16] Markus Furger, Maria Cruz Minguillón, Varun Yadav, Jay G Slowik, Christoph Hüglin, Roman Fröhlich, Krag Pettersson, Urs Baltensperger, and André SH Prévôt. 2017. Elemental composition of ambient aerosols measured with high temporal resolution using an online XRF spectrometer. *Atmospheric Measurement Techniques* 10, 6 (2017), 2061–2076.
- [17] Qi Huang, Xuesong Yin, Songcan Chen, Yigang Wang, and Bowen Chen. 2020. Robust nonnegative matrix factorization with structure regularization. *Neuro-computing* 412 (2020), 72–90. <https://doi.org/10.1016/j.neucom.2020.06.049>
- [18] Prateek Jain and Purushottam Kar. 2017. Non-convex Optimization for Machine Learning. *Foundations and Trends® in Machine Learning* 10, 3–4 (2017), 142–363.
- [19] Cheol-Heon Jeong, Jon M Wang, and Greg J Evans. 2016. Source apportionment of urban particulate matter using hourly resolved trace metals, organics, and inorganic aerosol components. *Atmospheric Chemistry and Physics Discussions* (2016), 1–32.
- [20] UC Kulshrestha, T Nageswara Rao, S Azhaguvel, and MJ Kulshrestha. 2004. Emissions and accumulation of metals in the atmosphere due to crackers and sparkles during Diwali festival in India. *Atmospheric Environment* 38, 27 (2004), 4421–4425.
- [21] Matthew S Landis, J Patrick Pancras, Joseph R Graney, Emily M White, Eric S Edgerton, Allan Legge, and Kevin E Percy. 2017. Source apportionment of ambient fine and coarse particulate matter at the Fort McKay community site, in the Athabasca Oil Sands Region, Alberta, Canada. *Science of the Total Environment* 584 (2017), 105–117.
- [22] Daniel Lee and H. Sebastian Seung. 2000. Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp (Eds.), Vol. 13. MIT Press.
- [23] Qian Li, Hongguang Cheng, Tan Zhou, Chunye Lin, and Shu Guo. 2012. The estimated atmospheric lead emissions in China, 1990–2009. *Atmospheric Environment* 60 (2012), 1–8.
- [24] Zhi-Quan Luo and Paul Tseng. 1993. Error bounds and convergence analysis of feasible descent methods: A general approach. *Annals of Operations Research* 46, 1 (1993), 157–178.
- [25] Bhaskar Mukhoty, Govind Gopakumar, Prateek Jain, and Purushottam Kar. 2019. Globally-convergent Iteratively Reweighted Least Squares for Robust Regression Problems. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- [26] Hermine Nalbandian. 2012. Trace element emissions from coal. *IEA Clean Coal Centre* 601 (2012).
- [27] Pentti Paatero. 1999. The Multilinear Engine: A Table-Driven, Least Squares Program for Solving Multilinear Problems, including the n-Way Parallel Factor Analysis Model. *Journal of Computational and Graphical Statistics* 8, 4 (1999), 854–888. <http://www.jstor.org/stable/1390831>
- [28] Pentti Paatero and Philip K. Hopke. 2009. Rotational tools for factor analytic models. *Journal of Chemometrics* 23, 2 (2009), 91–100.
- [29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [30] Siwatt Pongpiachan, Akihiro Iijima, and Junji Cao. 2018. Hazard quotients, hazard indexes, and cancer risks of toxic metals in PM₁₀ during firework displays. *Atmosphere* 9, 4 (2018), 144.
- [31] Pragati Rai, Markus Furger, Imad El Haddad, Varun Kumar, Liwei Wang, Atinderpal Singh, Kuldeep Dixit, Deepika Bhattu, Jean-Eudes Petit, Dilip Ganguly, et al. 2020. Real-time measurement and source apportionment of elements in Delhi's atmosphere. *Science of the Total Environment* 742 (2020), 140332.
- [32] Pragati Rai, Markus Furger, Jay G Slowik, Francesco Canonaco, Roman Fröhlich, Christoph Hüglin, María Cruz Minguillón, Krag Pettersson, Urs Baltensperger, and André SH Prévôt. 2020. Source apportionment of highly time-resolved elements during a firework episode from a rural freeway site in Switzerland. *Atmospheric Chemistry and Physics* 20, 3 (2020), 1657–1674.
- [33] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*. 166–171. <https://doi.org/10.1109/ICDCSW.2011.20>
- [34] Ashutosh K Shukla, Vipul Lalchandani, Deepika Bhattu, Jay S Dave, Pragati Rai, Navaneeth M Thamban, Suneeti Mishra, Sreenivas Gaddamidi, Nidhi Tripathi, Pawan Vats, et al. 2021. Real-time quantification and source apportionment of fine particulate matter including organics and elements in Delhi during summertime. *Atmospheric Environment* 261 (2021), 118598.
- [35] Ashutosh Kumar Shukla, Sachchida Nand Tripathi, Francesco Canonaco, Vipul Lalchandani, Ravi Sahu, Deepchandra Srivastava, Jay Dave, Navaneeth Meena Thamban, Sreenivas Gaddamidi, Lokesh Sahu, et al. 2023. Spatio-temporal variation of C-PM_{2.5} (composition based PM_{2.5}) sources using PMF*PMF (double-PMF) and single-combined PMF technique on real-time non-refractory, BC and elemental measurements during post-monsoon and winter at two sites in Delhi, India. *Atmospheric Environment* 293 (2023), 119456.
- [36] Anna Tobler, Deepika Bhattu, Francesco Canonaco, Vipul Lalchandani, Ashutosh Shukla, Navaneeth M Thamban, Suneeti Mishra, Atul K Srivastava, Deewan S Bisht, Suresh Tiwari, et al. 2020. Chemical characterization of PM_{2.5} and source apportionment of organic aerosol in New Delhi, India. *Science of The Total Environment* 745 (2020), 140924.
- [37] Anja H Tremper, Anna Font, Max Priestman, Samera H Hamad, Tsai-Chia Chung, Ari Pribadi, Richard JC Brown, Sharon L Goddard, Nathalie Grassineau, Krag Pettersson, et al. 2018. Field and laboratory evaluation of a high time resolution x-ray fluorescence instrument for determining the elemental composition of ambient aerosols. *Atmospheric Measurement Techniques* 11, 6 (2018), 3541–3557.
- [38] Stephen A. Vavasis. 2010. On the Complexity of Nonnegative Matrix Factorization. *SIAM Journal on Optimization* 20, 3 (2010), 1364–1377. <https://doi.org/10.1137/070709967>
- [39] S Visser, Jay G Slowik, M Furger, P Zotter, N Bukowiecki, F Canonaco, U Flechsig, K Appel, DC Green, AH Tremper, et al. 2015. Advanced source apportionment of size-resolved trace elements at multiple sites in London during winter. *Atmospheric Chemistry and Physics* 15, 19 (2015), 11291–11309.
- [40] Yu-Xiong Wang and Yu-Jin Zhang. 2013. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering*

- 25, 6 (2013), 1336–1353. <https://doi.org/10.1109/TKDE.2012.51>
- [41] Jianchao Yang, Shuicheng Yang, Yun Fu, Xuelong Li, and Thomas Huang. 2008. Non-negative graph embedding. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1–8. <https://doi.org/10.1109/CVPR.2008.4587665>
- [42] Qi Zhang, Jose L Jimenez, Manjula R Canagaratna, Ingrid M Ulbrich, Nga L Ng, Douglas R Worsnop, and Yele Sun. 2011. Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. *Analytical and bioanalytical chemistry* 401 (2011), 3045–3067.