Quantile Multiplicative Updates for Corruption-Robust Nonnegative Matrix Factorization

Zane Collins
Dept. of Mathematics
Harvey Mudd College
Claremont, CA, USA

Jamie Haddock
Dept. of Mathematics
Harvey Mudd College
Claremont, CA, USA

Tyler Headley
Dept. of Mathematics
Harvey Mudd College
Claremont, CA, USA

Luke Wang
Dept. of Mathematics
Harvey Mudd College
Claremont, CA, USA

***Abstract*—Nonnegative matrix factorization (NMF) models have found success in a variety of applications, including document clustering and classification, image processing, and bioinformatics. However, the optimization formulations typically employed for NMF models are often very sensitive to noise and corruption in the data. We introduce a quantile-based variant of the popular multiplicative updates method for training the Frobenius norm-formulation of NMF which avoids the effects of corruption in the data. Our numerical experiments illustrate the promise of this method, and show that in some scenarios, this method applied to the corrupted data recovers factorizations nearly as good as those learned on the uncorrupted data.**

***Index Terms*—Nonnegative matrix factorization, corruption-robust model, quantile-based methods, multiplicative updates**

I. Introduction

As the size of regularly encountered datasets continues to grow, there has been recent emphasis on the development of methods that extract meaningful latent trends from this data. A significant line of research has sought to use dimensionality reduction [30] and topic modeling techniques [31] to identify fundamental archetypes in the data. Many of these approaches are based on *matrix factorization* or the popular model *nonnegative matrix factorization (NMF)* [22].

However, the presence of corruption, outliers, and adversarial noise or perturbations can be entirely disruptive to these techniques or machine learning models in general [3], all while the input data is often too large for end users to inspect for spurious results [21], [26]. The need for methods that are robust to corruption, outliers, and adversarial noise has only expanded in recent years and is increasingly the focus across numerous subfields of numerical linear algebra [8], [16], optimization [1], statistics [18], and machine learning [2].

A. Nonnegative Matrix Factorization

Given a nonnegative matrix $\boldsymbol{D} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$, and a desired rank $r \in \mathbb{N}$, NMF seeks to decompose $\boldsymbol{D}$ into a product of nonnegative dictionary matrix $\boldsymbol{W} \in \mathbb{R}_{\geq 0}^{n_1 \times r}$ and nonnegative representation matrix $\boldsymbol{H} \in \mathbb{R}_{\geq 0}^{r \times n_2}$ so that

$$\boldsymbol{D} \approx \boldsymbol{W}\boldsymbol{H} = \sum_{j=1}^{r} \boldsymbol{w}_j \otimes \boldsymbol{h}_j, \tag{1}$$

JH was partially supported by NSF DMS #2211318 and NSF CAREER #2440040.

where $\boldsymbol{w}_j$ is a column of $\boldsymbol{W}$ and $\boldsymbol{h}_j$ is a row of $\boldsymbol{H}$; see Figure 1 for a visualization of this model. Typically, r is chosen such that $r < \min\{n_1, n_2\}$ to reduce the dimension of the original data matrix or reveal latent themes in the data. Each column of $\boldsymbol{H}$ provides the approximation of the respective column in $\boldsymbol{D}$ in the lower-dimensional space spanned by the columns of $\boldsymbol{W}$. The columns of $\boldsymbol{W}$ are interpreted as *parts* which nonnegatively combine to approximate the columns of $\boldsymbol{D}$. By looking at the row of $\boldsymbol{H}$ corresponding to a column of $\boldsymbol{W}$, one sees to which data (columns of $\boldsymbol{D}$) that part contributes. The nonnegativity of the NMF factor matrices yields clear interpretability; thus, NMF has found application in document clustering [11], [28], [33], and image processing and computer vision [12], [17], [22], amongst others.

Several formulations for this nonnegative approximation have been studied [5], [22], [23], [34]; for example,

$$\operatorname*{arg\,min}_{\boldsymbol{W}\geq 0, \boldsymbol{H}\geq 0} \|\boldsymbol{D} - \boldsymbol{W}\boldsymbol{H}\|_F^2 \quad \text{and} \quad \operatorname*{arg\,min}_{\boldsymbol{W}\geq 0, \boldsymbol{H}\geq 0} D(\boldsymbol{D}\|\boldsymbol{W}\boldsymbol{H}),$$

where $D(\cdot\|\cdot)$ is the information divergence. One reason for this popularity is that $\|\cdot\|_F$-NMF and $D(\cdot\|\cdot)$-NMF correspond to the maximum likelihood estimator given an assumed latent generative model and a Gaussian and Poisson model of uncertainty, respectively [4], [10], [32]. Popular training methods include multiplicative updates [22]–[24], projected gradient descent [25], and alternating least squares [19], [20].

B. Notation and Problem Setting

In what follows, we consider the problem of learning a rank-r NMF for the given data matrix $\boldsymbol{D} \in \mathbb{R}^{n_1 \times n_2}$; that is, we seek $\boldsymbol{W} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{H} \in \mathbb{R}^{r \times n_2}$ so that $\boldsymbol{D} \approx \boldsymbol{W}\boldsymbol{H}$. In this paper, we consider the case where the data is additively formed by two pieces, $\boldsymbol{D} = \tilde{\boldsymbol{D}} + \boldsymbol{C}$, where the *uncorrupted data* $\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{W}}\tilde{\boldsymbol{H}}$ is exactly factorizable (but unavailable to us), and the additive corruption $\boldsymbol{C}$ is sparse. We define the fraction of corruptions to be β, that is $\beta := \frac{|\text{supp}(\boldsymbol{C})|}{n_1 n_2}$.

Let $Q_q(\boldsymbol{E})$ denote the empirical q-quantile of the matrix $\boldsymbol{E}$ over all its entries,

$$Q_q(\boldsymbol{E}) = q\text{-quantile}\left\{\boldsymbol{E}_{ij} : i \in [n_1], j \in [n_2]\right\},$$

where the q-quantile of a finite set $S \subset \mathbb{R}$ is defined to be the $\lceil q|S| \rceil$-*th* smallest element of S; that is, $s \in S$ such that $|\{r \in S : r \leq s\}| = \lceil q|S| \rceil$. $|S|$ denotes the cardinality of S.

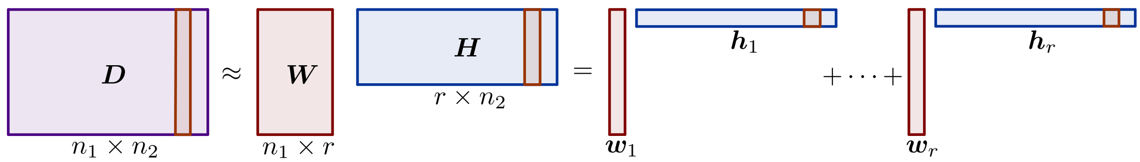

Fig. 1: Given a desired rank r, NMF is formulated as a factorization of a data matrix $\boldsymbol{D} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2}$ into the product of a dictionary matrix $\boldsymbol{W} \in \mathbb{R}_{\geq 0}^{n_1 \times r}$ and representation matrix $\boldsymbol{H} \in \mathbb{R}_{\geq 0}^{r \times n_2}$.

C. Related Work

To address non-Gaussian or Poisson noise in data matrices, on which regular NMF performs poorly, Guo and Zhang [13] propose Sparse Corruption Non-Negative Matrix Factorization (SCNMF). SCNMF uses the parameterized objective function

$$\min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0, \mathbf{S}} \|\mathbf{S}\|_1 + \alpha_1 \|\mathbf{D} - \mathbf{WH} - \mathbf{S}\|_F^2 + \alpha_2 \left\|\tilde{\mathbf{D}} - \mathbf{WH}\right\|_F^2,$$

where $\mathbf{D} \approx \mathbf{WH} + \mathbf{S}$ is the corrupted input matrix, $\mathbf{S}$ is the sparse error matrix, and $\tilde{\mathbf{D}} = \mathbf{WH}$ is the uncorrupted input matrix. In SCNMF, both the corrupted and uncorrupted data are known, which is unrealistic in many applications.

In [9], Díaz and Steele evaluated three different NMF models based on the L_1, L_2, and $\mathrm{L}_{2,1}$ norms on the basis of robustness to Gaussian and salt-and-pepper noise. They found that the algorithms based on the L_1 norm and $\mathrm{L}_{2,1}$ norm were more robust, performing similarly on feature selection and robustness to both types of noise. The algorithm based on the L_1 norm was particularly interesting as it allowed for the identification and removal of noise; however, it suffered from slower computation than algorithms for other models.

Recent works have proposed methods that introduce *quantile statistics* of the in-iteration residual information to detect and avoid corrupted data in large-scale *unconstrained* linear systems and linear regression [6], [14], [15]. These methods have been based upon the randomized Kaczmarz method [29]. Quantile information has previously been introduced into linear least-squares models to provide robustness to outlier or corrupted measurements [27].

D. Contributions and Organization

Our primary contribution is introducing the quantile-statistic into the popular multiplicative updates method for NMF, producing a method which is robust to additive, sparse corruption in the data. We provide an initial theoretical result which illustrates a simple case under which the NMF objective on the uncorrupted data is known to be non-increasing; that is *the factorization learned by our method on the corrupted data obeys the same guarantees as the factorization learned by regular multiplicative updates applied to the uncorrupted data.* We additionally illustrate the promise of our method with a suite of numerical experiments on real and synthetic data.

II. Quantile Multiplicative Updates

We propose a variant of the classical multiplicative updates (MU) method of Lee and Seung [23]. Our algorithm, *quantile multiplicative updates (QMU)*, seeks to avoid the effect of corrupted entries in the input data matrix $\boldsymbol{D}$ by masking the effects of matrix entries in $\boldsymbol{D}$ corresponding to unusually large magnitude entries in the residual matrix $\boldsymbol{E} = |\boldsymbol{D} - \boldsymbol{WH}|$ given approximate factors $\boldsymbol{W}$ and $\boldsymbol{H}$. To avoid updating according to the effect of elements suspected of being corrupted, we adapt the weighted NMF (WNMF) multiplicative update rules [35] to mask those suspected corrupted entries. We identify suspected corruptions by identifying those entries of $\boldsymbol{E}$ which are larger than a significant fraction, q, of the entries of $\boldsymbol{E}$; that is, $\boldsymbol{E}_{ij} > Q_q(\boldsymbol{E})$. Each iteration of our algorithm consists of two parts; first, we calculate the masking matrix which masks any entries corresponding to entries of $\boldsymbol{E}$ larger than a quantile of all entries of $\boldsymbol{E}$ in Algorithm 2; second, in Algorithm 1 we apply the weighted multiplicative update using the masking matrix computed in Algorithm 2.

Algorithm 1 Quantile Multiplicative Updates (QMU)

Input: data matrix $\boldsymbol{D} \in \mathbb{R}^{n_1 \times n_2}$, integer N, quantile q
Initialize $\boldsymbol{W} \in \mathbb{R}^{n_1 \times r}$, $\boldsymbol{H} \in \mathbb{R}^{r \times n_2}$, $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$
for $i = 1, \cdots, N$ **do**
  $\boldsymbol{M} \leftarrow \text{Quantile Mask}(\boldsymbol{D}, \boldsymbol{W}, \boldsymbol{H}, q)$
  $\boldsymbol{W} \leftarrow \boldsymbol{W} \circ \frac{(\boldsymbol{M} \circ \boldsymbol{D})\boldsymbol{H}^\top}{(\boldsymbol{M} \circ \boldsymbol{WH})\boldsymbol{H}^\top}$
  $\boldsymbol{H} \leftarrow \boldsymbol{H} \circ \frac{\boldsymbol{W}^\top(\boldsymbol{M} \circ \boldsymbol{D})}{\boldsymbol{W}^\top(\boldsymbol{M} \circ \boldsymbol{WH})}$
return $\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{M}$

Algorithm 2 Quantile Mask

Input: data matrix $\boldsymbol{D}$, factors $\boldsymbol{W}$, $\boldsymbol{H}$, quantile q
Initialize $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$
$\boldsymbol{E} \leftarrow |\boldsymbol{D} - \boldsymbol{WH}|$
for $i = 1, \cdots, n_1$, $j = 1, \cdots, n_2$ **do**
  $\boldsymbol{M}_{ij} \leftarrow \begin{cases} 1 & \text{if } \boldsymbol{E}_{ij} \leq Q_q(\boldsymbol{E}) \\ 0 & \text{otherwise.} \end{cases}$
return M

In our initial theoretical result, we illustrate (using prior guarantees for standard MU [23]) that if our masking matrix masks exactly those entries which are corrupted in $\boldsymbol{D}$, then the objective function on the uncorrupted data $\tilde{\boldsymbol{D}}$ is non-increasing under QMU applied to $\boldsymbol{D}$. We do not include the proof here due to space constraints.

Proposition 1: Let $\tilde{\boldsymbol{D}}$ be the uncorrupted data (which is generally unavailable) and $\boldsymbol{D} = \tilde{\boldsymbol{D}} + \boldsymbol{C}$ be the corrupted input data. Define $\Omega_C = \{(i,j) : i \in [n_1], j \in [n_2], C_{ij} \neq 0\}$

and $\Omega_C^c = ([n_1] \times [n_2]) \setminus \Omega_C$. Suppose the masking matrix $\boldsymbol{M}$ learned in line 4 of Algorithm 1 satisfies $\text{supp}(\boldsymbol{M}) = \Omega_C^c$. Then the model error on the uncorrupted data,

$$\|(\tilde{\boldsymbol{D}} - \boldsymbol{W}\boldsymbol{H})_{\Omega_C^c}\|_F^2,$$

is non-increasing under the QMU updates applied to corrupted input data $\boldsymbol{D}$, lines 5 and 6 of Algorithm 1.

III. Numerical Experiments

In this section, we provide numerical experiments that illustrate the promise of the QMU method and explore its behavior for various data and hyperparameter scenarios. Code for all experiments can be found at our Github repository. In each experiment, we print the average time required for a single model training trial and report these timings on the associated figures. The experiments presented here were performed in Python 3.11.8 with Numpy 1.26.4 on a 2024 MacBook Pro with 16 GB of RAM and Apple M4 processor.

A. Synthetic Data

In the following experiments, we generate synthetic data matrices by generating an exactly factorizable matrix $\tilde{\boldsymbol{D}} = \hat{\boldsymbol{W}}\hat{\boldsymbol{H}} \in \mathbb{R}^{120\times 100}$ where $\hat{\boldsymbol{W}} \in \mathbb{R}^{120\times \hat{r}}$ and $\hat{\boldsymbol{H}} \in \mathbb{R}^{\hat{r}\times 100}$ have integer entries sampled uniformly at random from $\{0, \cdots, 100\}$. We set $\text{rank}(\tilde{\boldsymbol{D}}) = \hat{r} = 40$. We then form the corrupted data $\boldsymbol{D}$ by sampling entries uniformly at random from $\tilde{\boldsymbol{D}}$ and adding values sampled independently from $|\mathcal{N}(0, 10^{12})|$. In all experiments, unless otherwise specified, we set the model rank equal to the rank of the uncorrupted data, $r = \hat{r}$, and the predicted uncorrupted proportion (quantile) q equal to the true uncorrupted proportion $1 - \beta$. In each experiment that follows, we train each model for 400 iterations and run 10 trials with different factor matrix initializations. We plot the mean error as a line and shade in the region between the minimum and maximum errors over all trials.

1) Comparing standard multiplicative updates and QMU: In our first experiment, we compare the behavior of standard MU and QMU applied to the uncorrupted data $\tilde{\boldsymbol{D}}$, and corrupted data $\boldsymbol{D}$ with corruption rate $\beta = 0.2$. We run standard MU on corrupted data $\boldsymbol{D}$ and uncorrupted data $\tilde{\boldsymbol{D}}$, as well as QMU on corrupted data $\boldsymbol{D}$. All errors are measured against the uncorrupted data $\tilde{\boldsymbol{D}}$. Figure 2 demonstrates that although standard MU with uncorrupted data reaches the lowest overall error, in the presence of corrupted entries, QMU vastly outperforms standard MU.

2) Relative error against corrupted and uncorrupted data: In our second experiment, we compare the errors of the matrix factorization produced by QMU on $\boldsymbol{D}$ with model rank $r = 40$ measured against uncorrupted data $\tilde{\boldsymbol{D}}$ and corrupted data $\boldsymbol{D}$ with corruption rate $\beta = 0.1$. In Figure 3, QMU fits the uncorrupted data $\tilde{\boldsymbol{D}}$ rather than the corrupted data $\boldsymbol{D}$, despite not having access to $\tilde{\boldsymbol{D}}$, indicating that QMU effectively identifies and ignores corrupted entries.

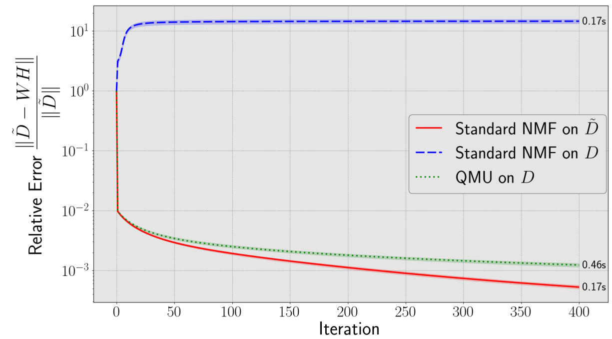

Fig. 2: Standard multiplicative updates on uncorrupted data $\tilde{\boldsymbol{D}} \in \mathbb{R}^{120\times 100}$ with rank $\hat{r} = 40$ and corrupted data $\boldsymbol{D}$ with corruption rate $\beta = 0.2$ versus QMU on corrupted data $\boldsymbol{D}$. All errors are measured against uncorrupted data $\tilde{\boldsymbol{D}}$.

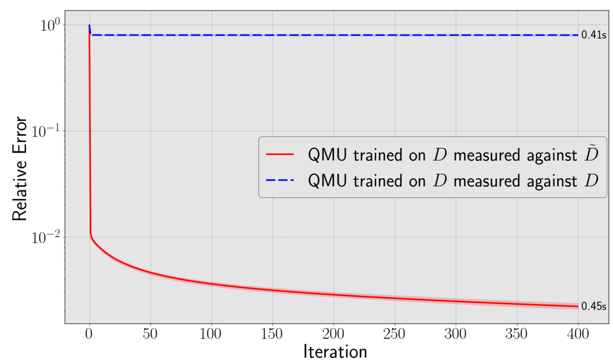

Fig. 3: Relative error of matrix factorization produced by QMU on $\boldsymbol{D}$ with model rank $r = 40$ measured against uncorrupted data $\tilde{\boldsymbol{D}} \in \mathbb{R}^{120\times 100}$ with rank $\hat{r} = 40$ and corrupted data $\boldsymbol{D}$ with $\beta = 0.1$.

3) Varying quantile q with fixed corruption rate β: Here, we vary the quantile q while holding the data corruption rate β constant. We again generate uncorrupted data $\tilde{\boldsymbol{D}}$ and corrupted data $\boldsymbol{D}$ with fixed corruption rate $\beta = 0.2$, and use model rank $r = 40$. Figure 4 illustrates that quantile $q = 0.8 = 1 - \beta$ performs best, while when too many entries are masked ($1 - q > \beta$) QMU performs slightly worse, and when we fail to mask the corruptions ($1 - q < \beta$) QMU diverges. This highlights the need to select a conservative quantile q; however, in real-world use, this may be difficult to guarantee.

4) Varying corruption rate β and quantile q together: In this experiment, we vary the quantile q and the corruption rate β together such that $q + \beta = 1$. We choose the model rank $r = 40$. Figure 5 displays the training error of the different learned models for each combination tested of q and β. We observe that error increases as corruption rate β increases, however, this error increase is gradual as only slightly less information is available in each iteration.

5) Varying model rank r: In this experiment, we measure the effectiveness of QMU on corrupted synthetic data when varying the model rank r. We again generate uncorrupted data

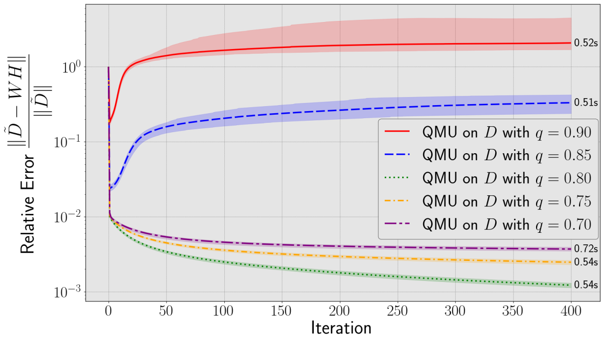

Fig. 4: Relative error per iteration for QMU on $\boldsymbol{D}$ with corruption rate $\beta = 0.2$ and model rank $r = 40$ measured against uncorrupted data $\tilde{\boldsymbol{D}} \in \mathbb{R}^{120\times 100}$ with rank $\hat{r} = 40$.

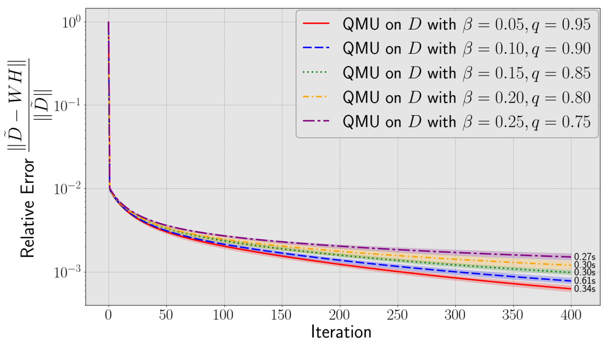

Fig. 5: Varying β and q together such that $q+\beta = 1$. We apply QMU to $\boldsymbol{D}$ and use model rank $r = 40$. Error is measured against uncorrupted data $\tilde{\boldsymbol{D}} \in \mathbb{R}^{120\times 100}$ with rank $\hat{r} = 40$.

$\tilde{\boldsymbol{D}}$ and corrupted data $\boldsymbol{D}$ with corruption rate $\beta = 0.1$. We run QMU on $\boldsymbol{D}$ with model ranks $r \in \{10, 20, 40, 60\}$ and measure error against $\tilde{\boldsymbol{D}}$. Figure 6 illustrates model performance suffers if underestimating the data rank, but overestimating the rank does not significantly affect performance.

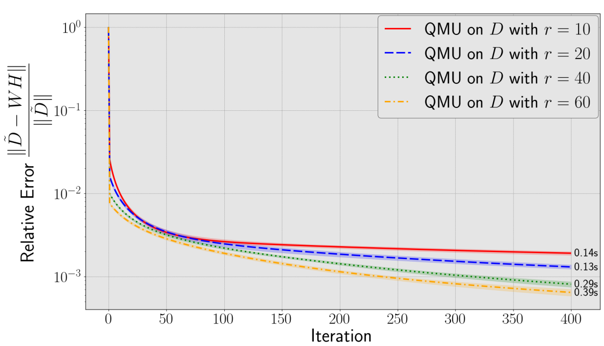

Fig. 6: Relative error of QMU applied to $\boldsymbol{D}$ with corruption rate $\beta = 0.1$ and model ranks $r \in \{10, 20, 40, 60\}$. All errors are measured against $\tilde{\boldsymbol{D}} \in \mathbb{R}^{120\times 100}$ with rank $\hat{r} = 40$.

B. Swimmer Data

In this section, we apply the proposed method to a toy image dataset called the `Swimmer` dataset [7], which is composed of 11×20-pixel images such as that of Figure 7a. We let $\tilde{\boldsymbol{D}}$ be the matrix with columns that are vectorized images from the `Swimmer` dataset, and note that this matrix has rank $\hat{r} = 16$.

In this experiment, we test QMU on the `Swimmer` dataset to better understand its effectiveness on real data. We use a model rank of $r = 17$ and set $q = 0.95$, and train our models for 400 iterations. We build $\boldsymbol{D}$ by corrupting $\beta = 0.05$ of the entries with values drawn from $|\mathcal{N}(0, 25)|$; Figure 7c presents one of the `Swimmer` figures after corruption. In Figure 7, we reconstruct this image using standard MU and QMU. We see that while the standard MU algorithm reconstructs uncorrupted data well (Figure 7b), when applied to corrupted data, the reconstruction (Figure 7d) is poor. In contrast, QMU masks these corruptions for a better reconstruction (Figure 7e).

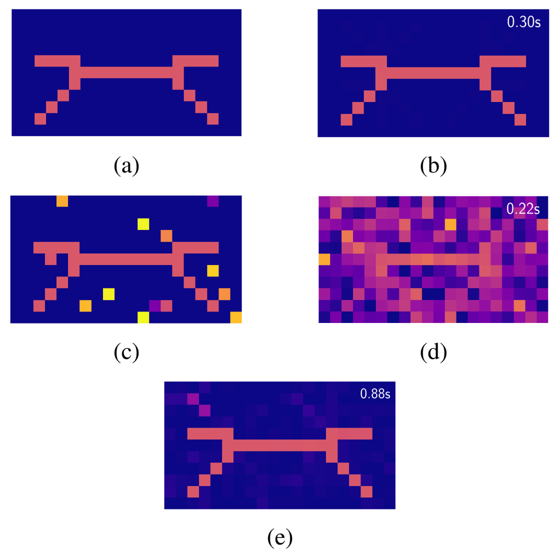

Fig. 7: Comparison of standard multiplicative updates and QMU in reconstructing `Swimmer` image 17. (a) `Swimmer` 17, no corruptions. (b) Reconstruction of `Swimmer` 17 by standard MU with rank 17. (c) `Swimmer` 17 with $\beta = 0.05$ of entries corrupted by noise drawn from $|\mathcal{N}(0, 25)|$. (d) Reconstruction of `Swimmer` 17 by standard MU with model rank $r = 17$ trained on $\boldsymbol{D}$. (e) Reconstruction of `Swimmer` 17 by QMU with rank $r = 17$ and $q = 0.95$ trained on $\boldsymbol{D}$.

IV. Conclusion

In this paper, we introduce a quantile-based variant of the popular multiplicative updates method for training the Frobenius norm-formulation of NMF which avoids the effects of additive, sparse corruption in the data. Our numerical experiments illustrate the promise of this method on synthetic and real data and show that in some scenarios, applying this method to corrupted data recovers factorizations nearly as good as those learned on uncorrupted data. We additionally provide an initial theoretical result which guarantees that, under a simple assumption on the support of the quantile masks, QMU applied to corrupted data obeys nearly the same guarantees as multiplicative updates on uncorrupted data.

References

[1] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*, volume 28. Princeton University Press, 2009.

[2] B. Biggio and F. Roli. Wild patterns: Ten years after the rise of adversarial machine learning. In *Proc. ACM SIGSAC Conf. Comput. Commun. Sec.*, pages 2154–2156, 2018.

[3] W. Brendel, J. Rauber, and M. Bethge. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. In *Int. Conf. Learn. Rep.*, 2018.

[4] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Comput. Intel. Neurosc.*, 2009, 2008.

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

[6] N. Coria, J. Haddock, and J. Pacheco. On quantile randomized Kaczmarz for linear systems with time-varying noise and corruption. 2024. Submitted.

[7] David Donoho and Victoria Stodden. When does non-negative matrix factorization give a correct decomposition into parts? *Lect. Notes Comput. Sc.*, 16, 2003.

[8] P. Du, A. Bouteiller, G. Bosilca, T. Herault, and J. Dongarra. Algorithm-based fault tolerance for dense matrix factorizations. In *Proc. ACM SIGPLAN Symp. Princ. Pract. Parallel Program.*, page 225–234, 2012.

[9] Alex Díaz and Damian Steele. Analysis of the robustness of nmf algorithms, 2021.

[10] P. Favaro and S. Soatto. *3-d shape estimation and image restoration: Exploiting defocus and motion-blur*. Springer Science & Business Media, 2007.

[11] E. Gaussier and C. Goutte. Relation between PLSA and NMF and implications. In *Proc. ACM SIGIR Conf. on Research and Development in Inform. Retrieval*, pages 601–602, 2005.

[12] D. Guillamet and J. Vitria. Non-negative matrix factorization for face recognition. In *Proc. Catalonian Conf. on Artif. Intel.*, pages 336–344. Springer, 2002.

[13] Zhibo Guo and Ying Zhang. A sparse corruption non-negative matrix factorization method and application in face image processing and recognition. *Measurement*, 136:429–437, 2019.

[14] J. Haddock, A. Ma, and E. Rebrova. On subsampled quantile randomized Kaczmarz. In *Proc. Allerton Conf. Comm. Cont. Comput.*, 2023.

[15] J. Haddock, D. Needell, E. Rebrova, and W. Swartworth. Quantile-based iterative methods for corrupted systems of linear equations. *SIAM Journal on Matrix Analysis and Applications*, 43(2):605–637, 2022.

[16] M. F. Hoemmen and M. A. Heroux. Fault-tolerant iterative methods. Technical report, Sandia National Lab (SNL-NM), Albuquerque, NM (United States), 2011.

[17] P. O. Hoyer. Non-negative sparse coding. In *Proc. IEEE Workshop on Neural Networks for Sig. Process.*, pages 557–565. IEEE, 2002.

[18] P. J. Huber. Robust statistics. In *Int. Encyc. Statist. Sci.*, pages 1248–1251. Springer, 2011.

[19] D. Kim, S. Sra, and I. S. Dhillon. Fast projection-based methods for the least squares nonnegative matrix approximation problem. *Stat. Anal. Data Min.*, 1(1):38–51, 2008.

[20] H. Kim and H. Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. A.*, 30(2):713–730, 2008.

[21] A. Kurakin, I. J. Goodfellow, and S. Bengio. Adversarial machine learning at scale. In *Int. Conf. Learn. Rep.*, 2017.

[22] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[23] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Proc. Adv. Neur. In.*, pages 556–562, 2001.

[24] H. Lee, J. Yoo, and S. Choi. Semi-supervised nonnegative matrix factorization. *IEEE Signal Proc. Let.*, 17(1):4–7, 2009.

[25] C. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural Comput.*, 19(10):2756–2779, 2007.

[26] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. In *Int. Conf. Learn. Rep.*, 2018.

[27] Peter J Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984.

[28] F. Shahnaz, M. Berry, V. Pauca, and R. Plemmons. Document clustering using nonnegative matrix factorization. *Inform. Process. Manag.*, 42(2):373–386, 2006.

[29] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.

[30] Laurens Van Der Maaten, Eric O Postma, H Jaap Van Den Herik, et al. Dimensionality reduction: A comparative review. *Journal of machine learning research*, 10(66-71):13, 2009.

[31] Ike Vayansky and Sathish AP Kumar. A review of topic modeling methods. *Information Systems*, 94:101582, 2020.

[32] T. Virtanen, A. T. Cemgil, and S. Godsill. Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In *Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Process.*, pages 1825–1828. IEEE, 2008.

[33] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273. ACM, 2003.

[34] Z. Yang, H. Zhang, Z. Yuan, and E. Oja. Kullback-Leibler divergence for nonnegative matrix factorization. In *Proc. Int. Conf. on Artif. Neural Networks*, pages 250–257. Springer, 2011.

[35] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. *Learning from Incomplete Ratings Using Non-negative Matrix Factorization*, pages 549–553.