NEGATIVE BINOMIAL MATRIX COMPLETION

Yu Lu[★]   *Kevin Bui*[†]   *Roummel F. Marcia*[★]

[★] Department of Applied Mathematics, University of California, Merced, California, USA
[†] Department of Mathematics, University of California, Irvine, California, USA

ABSTRACT

Matrix completion focuses on recovering missing or incomplete information in matrices. This problem arises in various applications, including image processing and network analysis. Previous research proposed Poisson matrix completion for count data with noise that follows a Poisson distribution, which assumes that the mean and variance are equal. Since overdispersed count data, whose variance is greater than the mean, is more likely to occur in realistic settings, we assume that the noise follows the negative binomial (NB) distribution, which can be more general than the Poisson distribution. In this paper, we introduce NB matrix completion by proposing a nuclear-norm regularized model that can be solved by proximal gradient descent. In our experiments, we demonstrate that the NB model outperforms Poisson matrix completion in various noise and missing data settings on real data.

Index Terms— Matrix completion, negative binomial distribution, data recovery, nonconvex optimization

1. INTRODUCTION

Matrix completion aims to recover the low-rank, ground-truth matrix $M \in \mathbb{R}^{m \times n}$ from an incomplete, noisy observation matrix $Y \in \mathbb{R}^{m \times n}$. Such situations occur widely in fields such as medical imaging [1] and social network recovery [2].

Previous research has shown that nuclear norm minimization is an effective method for matrix completion [3, 4], thereby serving as a foundation for many matrix completion algorithms [5–8]. Early works often developed matrix completion models and algorithms based on the noise-free assumption. Consequently, these models and algorithms were extended to more practical settings where the observations are noisy. For example, researchers redesigned them for cases when observations are corrupted with Gaussian noise or similar [9–11]. For count data, typically observed in photon-limited imaging [12] and network traffic analysis [13], the noise is commonly assumed to follow a Poisson distribution, as done in [6, 14–17]. Adapted to this case, Poisson matrix

completion was proposed in [6]. However, the Poisson distribution requires the data's mean and variance to be equal, which is a strong assumption. Hence, we assume a more general distribution, the negative binomial (NB) distribution, for count data whose variance is greater than the mean [18, 19]. The NB noise assumption motivated further modifications to existing statistical models, such as regression [19–21] and matrix factorization [22]. These models have been used in various applications, including traffic accident analysis [23], genomics [24], and signal/image processing [25].

Consisting of a sequence of independent and identically distributed Bernoulli trials, the NB(r, p) distribution has the following probability mass function (p.m.f.):

$$\mathbb{P}(y; r, p) = \binom{r + y - 1}{y}(1 - p)^y p^r,$$

where $y \geq 0$ is the number of successful trials before the rth failure trial and $p \in [0, 1]$ is the failure probability of each trial. The mean of the NB(r, p) distribution is $\mu = r(1-p)/p$, so $p = r/(r + \mu)$. The p.m.f. of the NB(r, p) can be shown to converge to the p.m.f. of the Poisson distribution with mean μ as $r \to \infty$ and $p \to 1$ [26]. As a result, the Poisson distribution can be approximated by the NB distribution.

Let $M \in \mathbb{R}_+^{m \times n}$ be a matrix of non-negative entries, and let $\{Y_{i,j}\}_{(i,j) \in \Omega}$ be a set of noisy observations of the entries of M over an index set $\Omega \subseteq \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$. Under the NB assumption with fixed parameter r, the observed matrix Y has the following entries:

$$Y_{i,j} \sim \text{NB}\left(r, \frac{r}{r + M_{i,j}}\right), \quad \text{for all } (i, j) \in \Omega. \quad (1)$$

Note that each entry $Y_{i,j}$ is independent of each other. Our goal is to recover the low-rank, ground-truth matrix M from the incomplete set $\{Y_{i,j}\}_{(i,j) \in \Omega}$ corrupted with noise modeled as an NB distribution.

2. MODEL FORMULATION AND ALGORITHM

In this section, we formulate a model to recover the ground-truth matrix M from the partially observed and NB-noise corrupted matrix $Y \in \mathbb{Z}_+^{m \times n}$ with entries from $\{Y_{i,j}\}_{(i,j) \in \Omega}$. By

(1), the probability of observing Y given the ground-truth matrix M is

$$\mathbb{P}(Y|M) =$$
$$\prod_{(i,j)\in\Omega} \binom{r + Y_{i,j} - 1}{Y_{i,j}} \left(\frac{M_{i,j}}{r + M_{i,j}}\right)^{Y_{i,j}} \left(\frac{r}{r + M_{i,j}}\right)^{r}. \tag{2}$$

Then by Bayes' Theorem, the maximum a posteriori estimate of M is obtained by

$$
\begin{aligned}
\widehat{M} &= \arg\max_{X} \mathbb{P}(X|Y) \\
&= \arg\max_{X} \frac{\mathbb{P}(Y|X)\mathbb{P}(X)}{\mathbb{P}(Y)} \\
&= \arg\min_{X} -\log\mathbb{P}(Y|X) - \log\mathbb{P}(X).
\end{aligned}
\tag{3}
$$

Because

$$
\begin{aligned}
-\log\mathbb{P}(Y|X) = \sum_{(i,j)\in\Omega} &-\log((r + Y_{i,j} - 1)!) + \log(Y_{i,j}!) \\
&+ \log((r-1)!) - r\log(r) - Y_{i,j}\log(X_{i,j}) \\
&+ (r + Y_{i,j})\log(r + X_{i,j}),
\end{aligned}
$$

(3) simplifies to

$$\widehat{M} = \arg\min_{X} F(X) - \log\mathbb{P}(X), \tag{4}$$

where

$$F(X) \equiv \sum_{(i,j)\in\Omega} (r + Y_{i,j})\log(r + X_{i,j}) - Y_{i,j}\log(X_{i,j}).$$

The latter term $-\log\mathbb{P}(X)$ in (4) is the log prior and can be interpreted as a regularizer. Since we desire $\widehat{M}$ to be low rank, we replace the log prior with nuclear norm regularization. Hence, our proposed model to recover the ground-truth matrix M is

$$\widehat{M} = \arg\min_{X} F(X) + \tau\|X\|_{*}, \tag{5}$$

where the nuclear norm $\|\cdot\|_{*}$ is the sum of the singular values of a matrix and $\tau > 0$ is a regularization parameter.

We solve (5) by proximal gradient descent (see e.g., [27, 28]). In each iteration k, we linearize F and solve the following optimization problem:

$$
\begin{aligned}
X^{k+1} = \arg\min_{X} \ & F(X^k) + \langle X - X^k, \nabla F(X^k)\rangle \\
& + \frac{t_k}{2}\|X - X^k\|_F^2 + \tau\|X\|_{*},
\end{aligned}
\tag{6}
$$

where $t_k > 0$ is the reciprocal of the step size (see [6]) and $\|\cdot\|_F$ is the Frobenius norm. Note that each entry of $\nabla F(X^k)$

is given by

$$(\nabla F(X^k))_{i,j} = \begin{cases} \dfrac{r + Y_{i,j}}{r + X_{i,j}^k} - \dfrac{Y_{i,j}}{X_{i,j}^k} & \text{if } (i,j) \in \Omega, \\ 0 & \text{if } (i,j) \notin \Omega. \end{cases}$$

By rearranging (6), we obtain

$$
\begin{aligned}
X^{k+1} &= \arg\min_{X} \|X\|_{*} + \frac{1}{2\frac{\tau}{t_k}}\left\|X - X^k + \frac{1}{t_k}\nabla F(X^k)\right\|_F^2 \\
&= \text{prox}_{\frac{\tau}{t_k}\|\cdot\|_{*}}\left(X^k - \frac{1}{t_k}\nabla F(X^k)\right),
\end{aligned}
\tag{7}
$$

where the proximal operator of $\|\cdot\|_{*}$ is defined by

$$\text{prox}_{\lambda\|\cdot\|_{*}}(Z) = \arg\min_{X} \|X\|_{*} + \frac{1}{2\lambda}\|X - Z\|_F^2, \tag{8}$$

for $\lambda > 0$. The exact solution to (8) is provided by the following lemma:

Lemma 1 (Theorem 2.1, [3]) *Consider the singular value decomposition of $Z \in \mathbb{R}^{m\times n}$ of rank l as follows:*

$$Z = U\Sigma V^\top, \ \Sigma = diag(\{\sigma_i\}_{i=1}^l),$$

where $U \in \mathbb{R}^{m\times l}$ and $V \in \mathbb{R}^{n\times l}$ are matrices with orthonormal columns. For $\lambda > 0$, let

$$\mathcal{D}_\lambda(Z) := U\mathcal{D}_\lambda(\Sigma)V^\top, \ \mathcal{D}_\lambda(\Sigma) = diag\left(\left\{(\sigma_i - \lambda)_+\right\}_{i=1}^l\right),$$

where $(\sigma - \lambda)_+ = \max(0, \sigma - \lambda)$. Then

$$\mathcal{D}_\lambda(Z) = prox_{\lambda\|\cdot\|_{*}}(Z).$$

Therefore, by Lemma 1, (7) simplifies to

$$X^{k+1} = \mathcal{D}_{\frac{\tau}{t_k}}(Z^k), \tag{9}$$

where $Z^k = X^k - \frac{1}{t_k}\nabla F(X^k)$. Because $\frac{1}{t_k}$ is the step size in the gradient descent step in (7), we increase t_k by a factor $\eta > 0$ whenever $F(X^{k+1}) > F(X^k)$ to ensure that $\{F(X^k)\}_{k=0}^\infty$ is monotonically decreasing. The overall algorithm is summarized in Algorithm 1.

3. EXPERIMENT

We conduct experiments on three different datasets corrupted with either Poisson noise or NB noise with dispersion parameters $r = 10$ and $r = 25$. For each noise level and dataset, we consider three different percentages q of known entries in the matrix, with $q = 25\%, 50\%$, and 75%. In each experiment, we compare our proposed NB model (5) with its Poisson counterpart [6]. Each experiment was conducted 20 times with different noisy realizations, and the average of

Algorithm 1 Proximal Gradient Descent for NB Matrix Completion

---
Require: Noisy incomplete matrix Y, upper bound $\alpha > 0$, lower bound $\beta > 0$, regularization parameter τ, penalty multiplier $\eta > 1$, index set Ω, the maximum iteration K, the stopping criteria tol
1: Initialize $X_{i,j}^0 = Y_{i,j}$ if $(i,j) \in \Omega$ else $X_{i,j}^0 = \frac{\alpha+\beta}{2}$
2: $k = 0$
3: **while** $k \leq K$ and $tol < err$ **do**
4:     $Z^k = X^k - \frac{1}{t_k}\nabla F(X^k)$
5:     $X^{k+1} = \mathcal{D}_{\frac{\tau}{t_k}}(Z^k)$
6:     **if** $F(X^{k+1}) > F^k(X^k)$ **then**
7:         $t_k = \eta t_k$
8:         Go back to 4
9:     **end if**
10:     $err = F(X^{k+1}) - F(X^k)$
11:     $k = k + 1$
12: **end while**
13: **return** Recovered matrix $\widehat{M} = X^k$.

---

these trials is presented. For numerical evaluation, we use peak signal-to-noise ratio (PSNR) and normalized root-mean-square error (NRMSE), given as follows:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{M_{\max}^2 \cdot mn}{S}\right)$$

and

$$\text{NRMSE} = \frac{1}{M_{\max} - M_{\min}}\sqrt{\frac{S}{mn}},$$

where $S = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\widehat{M}_{i,j} - M_{i,j}\right)^2$ and $M_{\min}$ and $M_{\max}$ are the minimum and maximum entry values of the ground-truth matrix M, respectively.

The parameters for Algorithm 1 are set as follows: maximum iteration $K = 2000$, the stopping criterion $tol = 1e-6$, and the penalty multiplier $\eta = 1.1$. The bounds α and β are chosen based on the maximum and minimum values of the entries, respectively. The regularization parameter τ is searched within the set $\{10, 20, \ldots, 310\}$.

For the NB models, several methods exist for estimating the parameter r, such as the method-of-moments [29], the maximum quasi-likelihood methods [30], and cross-validation techniques. However, to eliminate potential biases or inaccuracies inherent in the parameter estimation process, our experiments intentionally use the exact value of the parameter r as a prior. In the cases of Poisson noise, we set $r = 1000$ to have our NB model approximate the Poisson model.

One point that needs to be highlighted is that NB noise with $r = 10$ is not always noisier than $r = 25$. By (1), the noise level depends on both the dispersion parameter r and the original matrix entries. For example, if the dispersion parameter r is close to the original data values, the noise level would be minimal at these entries. Conversely, when the dispersion parameter r is distant from the original data values, the corresponding entries can be extremely noisy.

3.1. Experiment 1: Bike-Sharing Count Data

Table 1: Experiment 1 results. PSNRs and NRMSEs of the Poisson model and our proposed NB model on the Capital bike-sharing data with three different known-data level and three different noise level. **Bold** indicates best value.

Mask	Model	PSNR (↑) / NRMSE (↓)		
		$r = 10$	$r = 25$	Poisson
$q = 25\%$	Poisson	16.47 / 15.13%	20.37 / 9.59%	**22.29 / 7.71%**
	NB	**18.05 / 12.52%**	**21.06 / 8.87%**	21.92 / 8.04%
$q = 50\%$	Poisson	17.72 / 13.03%	21.69 / 8.26%	**27.12 / 4.41%**
	NB	**18.41 / 12.08%**	**22.58 / 7.63%**	26.75 / 4.60%
$q = 75\%$	Poisson	19.12 / 11.09%	22.74 / 7.30%	**30.12 / 3.12%**
	NB	**21.60 / 8.32%**	**24.63 / 5.88%**	29.65 / 3.30%

We examine the bike-sharing data aggregated on an hourly basis on Saturdays between 2011 and 2012 within the Capital Bike Share system[1], resulting in an 24×105 matrix. The average count is 194 and the median count is 163, so NB noise with $r = 10$ will be stronger than with $r = 25$. The results of the models applied to this dataset is provided in Table 1.

Across all three masking levels, the NB model demonstrates more accurate results than the Poisson model when the data is corrupted by NB noise with $r = 10, 25$. However, for Poisson noise, the Poisson model slightly outperforms the NB model as expected: the difference in performance is only at most 0.50 for PSNR and 0.33% for NRMSE. The most significant difference between the proposed NB model and the Poisson model occurs at the highest noise level, $r = 10$, indicating that the NB model yields more accurate results in complex noisy situations than the Poisson model. As the level of missing data increases, we observe a decrease in accuracy for both the Poisson and the proposed NB model.

3.2. Experiment 2: Vehicle Count Data

We examine the hourly inbound traffic data at Robert F. Kennedy Bridge Queens/Bronx Plaza between January 2021 and March 2021[2]. The traffic data records the number of vehicles that passed by without the E-ZPass tag or with dysfunctional tags. Note that the the mean and median of the dataset are 12 and 4, respectively. Like in Experiment 1, the data is formatted as a 24×90 matrix, shown in Fig. 1.

Table 2 presents the PSNRs and NRMSEs of the Poisson and NB models. As expected, we observe that the NB model is more accurate than the Poisson model for the NB noise cases, while the Poisson model achieves slightly better results for the Poisson noise case. The difference in performance for the Poisson noise case is less than 0.1 in PSNR and less than 0.04% in NRMSE. Across all noise levels, the results for $q = 75\%$ are more accurate than those for $q = 25\%$ because the

---
[1]Accessed at `https://doi.org/10.24432/C5W894`
[2]Accessed at `https://data.ny.gov/Transportation/Hourly-Traffic-on-Metropolitan-Transportation-Auth/qzve-kjga/about_data`

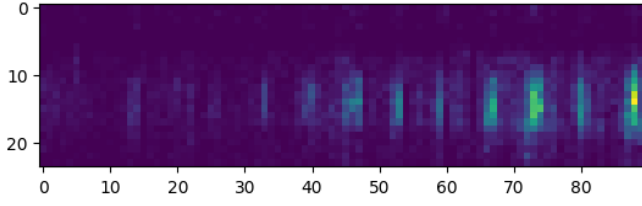**Fig. 1**: Hourly inbound traffic data at Robert F. Kennedy Bridge Queens/Bronx Plaza from January 2021 to March 2021.
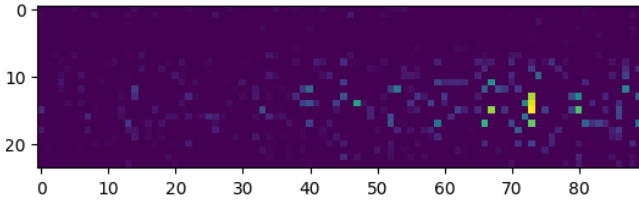
Table 2: Experiment 2 results. PSNRs and NRMSEs of the Poisson model and our proposed NB model on the hourly inbound traffic data at Robert F. Kennedy Bridge Queens/Bronx Plaza with three different known-data level and three different noise level. **Bold** indicates best value.

Mask	Model	PSNR (↑)/NRMSE (↓)		
		$r = 10$	$r = 25$	Poisson
$q = 25\%$	Poisson	21.44/8.47%	21.22/8.70%	**22.99/7.09%**
	NB	**23.97/6.37%**	**22.44/7.59%**	22.97/7.11%
$q = 50\%$	Poisson	27.72/4.14%	22.17/7.80%	**25.62/5.24%**
	NB	**28.36/3.82%**	**23.75/6.50%**	25.58/5.27%
$q = 75\%$	Poisson	28.69/3.68%	23.59/6.62%	**27.16/4.39%**
	NB	**29.79/3.24%**	**26.47/4.80%**	27.10/4.42%

more available information by the observed entries improves the inference of the missing entries. Since the dataset values are closer to 10 than to 25, $r = 25$ results in a noisier matrix. Consequently, the results for $r = 10$ are more accurate than those for $r = 25$.

For the case when $q = 25\%$ of the data entries are known, Fig. 2 visualizes the reconstruction results of the data matrix corrupted with Poisson noise, while Fig. 3 visualizes the reconstruction results of the data matrix corrupted with NB
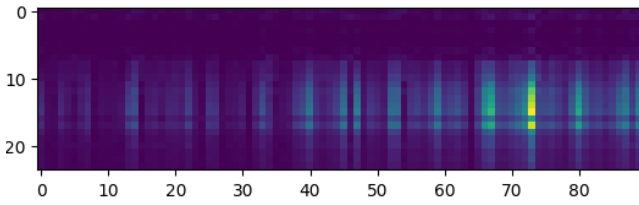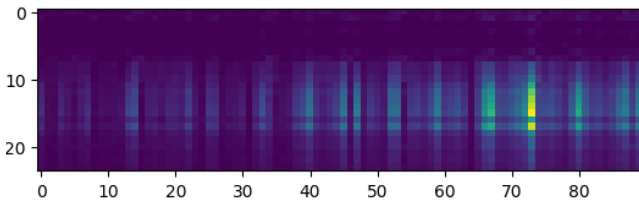
$(r = 10)$ noise. In Fig. 2, the reconstruction results between the Poisson and NB model appear very similar, as validated by their nearly equal PSNRs and RMSEs. On the other hand, in Fig. 3, the difference between the two reconstruction results is apparent. The NB reconstruction significantly resembles the original data matrix more than the Poisson reconstruction. Moreover, the NB reconstruction recovers stronger signals in the traffic count data than the Poisson reconstruction around the 73th and 89th days of 2021. Such signals are not clear in
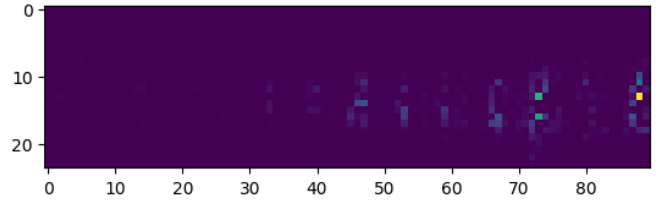

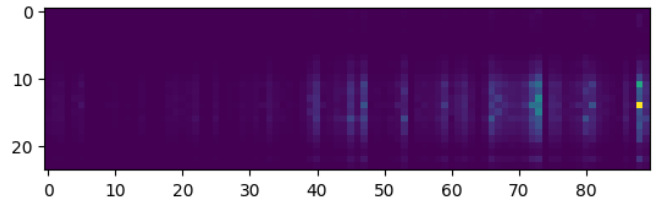(a) Data with Poisson noise and 25% known entries

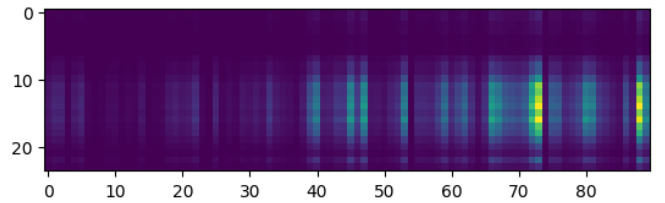

(b) Poisson Reconstruction

(c) NB Reconstruction

Fig. 2: Experiment 2 (Poisson) results. (a) Traffic data (Fig. 1) corrupted with Poisson noise and only 25% of its entries are known. (b) Poisson reconstruction with PSNR = 22.99 and RMSE = 7.09%. (c) NB reconstruction with PSNR = 22.97 and RMSE = 7.11%.

(a) Data with NB ($r = 10$) noise and 25% known entries

(b) Poisson Reconstruction

(c) NB Reconstruction

Fig. 3: Experiment 2 (NB) results. (a) Traffic data (Fig. 1) corrupted with NB ($r = 10$) noise and only 25% of its entries are known. (b) Poisson reconstruction with PSNR = 21.44 and RMSE = 8.47%. (c) NB reconstruction with PSNR = 23.97 and RMSE = 6.37%.

the corrupted matrix shown in Fig. 3a. Overall, our proposed NB model is a more general and robust model for recovering realistic count data.

3.3. Experiment 3: Microscopy Image

We examine a 512×512 microscopy image of mouse brain tissues [31]. The maximum pixel intensity value of this data is 242, the minimum value is 2, and the mean and median pixel intensity values are 14 and 10, respectively. Here, we break the image into 8×8 patches, vectorize them, and concatenate them into a 64×4096 matrix, and then approximate it by low-rank matrix with rank $l = 32$. This low-rank matrix is the ground-truth matrix M, shown in Fig. 4a, for this experiment. Subsequently, we corrupt Fig. 4a by NB noise with $r = 10$ and observe $q = 75\%$ of the entries. The Poisson and NB reconstructed images are shown in Fig. 4c-4d, where the Poisson reconstruction has PSNR $= 22.36$ and NRMSE $= 7.62\%$ while the NB reconstruction has PSNR $= 29.08$ and NRMSE $= 3.51\%$, demonstrating the improved performance of the proposed NB model over the Poisson model for this experiment.

(a) Low-Rank Image   (b) Noisy Image

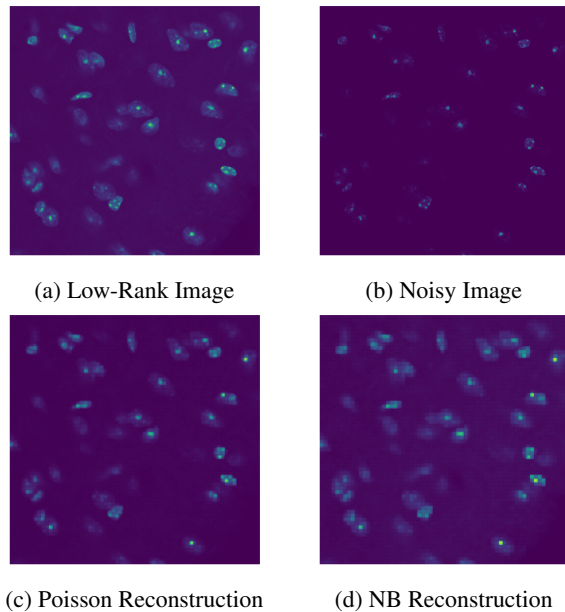(c) Poisson Reconstruction   (d) NB Reconstruction

Fig. 4: Results on the microscopy image of mouse brain tissues from [31]. (a) Low-rank image of the original. (b) Incomplete, noisy image with NB noise $r = 10$ NB noise level and $q = 75\%$ data-known level. (c) Poisson reconstruction with PSNR $= 22.36$ and NRMSE $= 7.62\%$. (d) NB model reconstruction with PSNR $= 29.08$ and RMSE $= 3.51\%$.

4. CONCLUSION

We proposed NB matrix completion to accommodate overdispersed count data often encountered in real-world scenarios. To perform NB matrix completion, we introduced a model formulated using maximum a posteriori and nuclear norm regularization and solved it by proximal gradient descent. Through three experiments, we compared our proposed model with its Poisson counterpart. Our experiments demonstrated that the NB model significantly outperformed the Poisson model when the data is corrupted with NB noise. On the other hand, when the data is corrupted with Poisson noise, the NB model with large dispersion parameter r maintained comparable performance as the Poisson model. In extreme cases of information loss, both the NB and Poisson models faced challenges in achieving high accuracy results.

5. REFERENCES

[1] Yue Hu, Xiaohan Liu, and Mathews Jacob, "A generalized structured low-rank matrix completion algorithm for MR image recovery," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1841–1851, 2018.

[2] Gunjan Mahindre, Anura P Jayasumana, Kelum Gajamannage, and Randy Paffenroth, "On sampling and recovery of topology of directed social networks–a low-rank matrix completion based approach," in *2019 IEEE 44th Conference on Local Computer Networks (LCN)*. IEEE, 2019, pp. 324–331.

[3] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[4] Emmanuel Candes and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, pp. 717–772, 2009.

[5] Dimitris Bertsimas and Michael Lingzhi Li, "Fast exact matrix completion: A unified optimization framework for matrix completion," *Journal of Machine Learning Research*, vol. 21, no. 231, pp. 1–43, 2020.

[6] Yang Cao and Yao Xie, "Poisson matrix recovery and completion," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1609–1620, 2015.

[7] Emmanuel J Candes and Yaniv Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[8] Jared Tanner and Ke Wei, "Normalized iterative hard thresholding for matrix completion," *SIAM Journal on Scientific Computing*, vol. 35, no. 5, pp. S104–S125, 2013.

[9] Suriya Gunasekar, Ayan Acharya, Neeraj Gaur, and Joydeep Ghosh, "Noisy matrix completion using alternating minimization," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2013, pp. 194–209.

[10] Raghunandan Keshavan, Andrea Montanari, and Sewoong Oh, "Matrix completion from noisy entries," *Advances in Neural Information Processing Systems*, vol. 22, 2009.

[11] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh, "Low-rank matrix completion with noisy observations: a quantitative comparison," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2009, pp. 1216–1222.

[12] Zachary T Harmany, Roummel F Marcia, and Rebecca M Willett, "Sparsity-regularized photon-limited imaging," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 772–775.

[13] Juan Andrés Bazerque, Gonzalo Mateos, and Georgios B Giannakis, "Inference of Poisson count processes using low-rank tensor data," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5989–5993.

[14] Lasith Adhikari and Roummel F Marcia, "Nonconvex relaxation for Poisson intensity reconstruction," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1483–1487.

[15] Kevin Bui, Yifei Lou, Fredrick Park, and Jack Xin, "Weighted anisotropic–isotropic total variation for Poisson denoising," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 1020–1024.

[16] Zachary T Harmany, Roummel F Marcia, and Rebecca M Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2011.

[17] Triet Le, Rick Chartrand, and Thomas J Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, 2007.

[18] Francis J Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[19] William Gardner, Edward P Mulvey, and Esther C Shaw, "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models.," *Psychological Bulletin*, vol. 118, no. 3, pp. 392, 1995.

[20] P. D. Allison and R. P. Waterman, "Fixed-effects negative binomial regression models," *Sociological Methodology*, vol. 32, no. 1, pp. 247–265, 2002.

[21] Huiming Zhang and Jinzhu Jia, "Elastic-net regularized high-dimensional negative binomial regression," *Statistica Sinica*, vol. 32, no. 1, pp. 181–207, 2022.

[22] O. Gouvert, T. Oberlin, and C. Févotte, "Negative binomial matrix factorization," *IEEE Signal Processing Letters*, vol. 27, pp. 815–819, 2020.

[23] Mark Poch and Fred Mannering, "Negative binomial analysis of intersection-accident frequencies," *Journal of Transportation Engineering*, vol. 122, no. 2, pp. 105–113, 1996.

[24] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth, "edgeR: A bioconductor package for differential expression analysis of digital gene expression data," *bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.

[25] Yu Lu and Roummel F Marcia, "Negative binomial optimization for low-count overdispersed sparse signal reconstruction," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 1948–1952.

[26] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, Pearson Education, 2012.

[27] Jean Jacques Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, vol. 255, pp. 2897–2899, 1962.

[28] Patrick L Combettes and Valérie R Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale modeling & simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[29] S. J. Clark and J. N. Perry, "Estimation of the negative binomial parameter κ by maximum quasi-likelihood," *Biometrics*, pp. 309–316, 1989.

[30] W. W. Piegorsch, "Maximum likelihood estimation for the negative binomial dispersion parameter," *Biometrics*, pp. 863–867, 1990.

[31] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard, "A Poisson-Gaussian denoising dataset with real fluorescence microscopy images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11710–11718.