NEGATIVE BINOMIAL OPTIMIZATION FOR LOW-COUNT OVERDISPERSED SPARSE SIGNAL RECONSTRUCTION

Yu Lu and Roummel F. Marcia

Department of Applied Mathematics, University of California, Merced, Merced, CA 95343 USA

ABSTRACT

Low-photon count imaging has been typically modeled by Poisson statistics. This discrete probability distribution model assumes that the mean and variance of a signal are equal. In the presence of greater variability in a dataset than what is expected, the negative binomial distribution is a suitable overdispersed alternative to the Poisson distribution. In this work, we present a framework for reconstructing sparse signals in these low-count overdispersed settings. Specifically, we describe a gradient-based sequential quadratic optimization approach that minimizes the negative log-likelihood corresponding to the negative binomial distribution coupled with a sparsity-promoting regularization term. Numerical experiments on 1D and 2D sparse/compressible signals are presented.

***Index Terms*—** Negative binomial distribution, sparse data recovery, nonconvex optimization, total variation

1. INTRODUCTION

Compressed sensing and sparse signal recovery are active areas of research [1, 2, 3], for which many algorithms have been proposed (e.g., [4, 5, 6]) for efficiently solving these problems, including those in photon-limited settings (e.g., [7, 8, 9]). The discrete low-count data from a detector appears in many settings, such as medical image processing [10], structural variations prediction [11], and network traffic analysis [12]. Discrete low-photon data reconstruction typically use a Poisson process model, whose application to imaging has been well-developed (see e.g., [13, 14, 15]). The Poisson model assumes that the mean and variance are equal. This paper proposes a sparse reconstruction algorithm for low-count settings where there is greater variability in a dataset than what is expected.

The negative binomial distribution is a discrete distribution returning the probability that y successful events happen in a sequence of independent and identically distributed Bernoulli trials until the r^{th} failure event and the probability of failure event is p. The probability mass function is given by

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

Under this negative binomial assumption, we write our observation model as

$$y \sim \text{NB}(r,p).$$

The expectation μ and the variance σ^2 are given by the following formulas:

$$\mu = r\frac{1-p}{p} \quad \text{and} \quad \sigma^2 = r\frac{1-p}{p^2}.$$

These formulas give rise to the expressions

$$\sigma^2 = \mu + \frac{\mu^2}{r} \quad \text{and} \quad p = \frac{\mu}{\sigma^2} = \frac{r}{r+\mu}.$$

Note that as r goes to infinity, the variance σ^2 will tend to the expectation μ, and it can be demonstrated that the Poisson distribution is a special case of the negative binomial distribution.

Let $f^* \in \mathbf{R}_+^n$ be the true signal of interest and $A \in \mathbf{R}_+^{m\times n}$ be the measurement matrix that maps f^* to the vector of observations. Therefore, the expected measurement at the detector is $Af^* \in \mathbf{R}_+^m$. Thus, the mean μ is given by the vector $\mu = Af^*$. The observation vector $y \in \mathbf{R}_+^m$ drawn from a negative binomial distribution model is given by

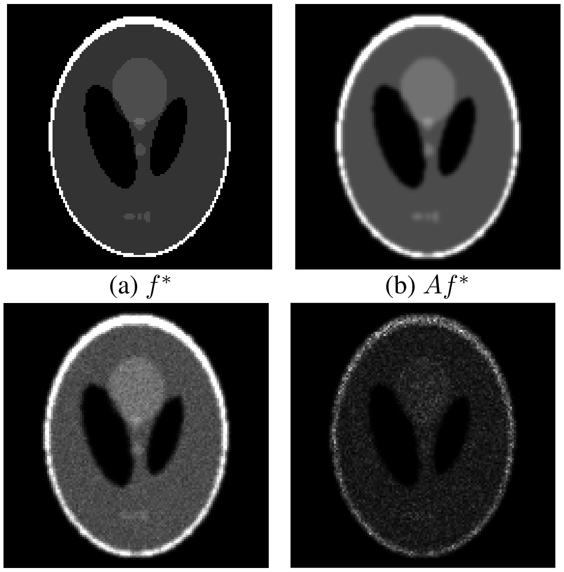
$$y_i \sim \text{NB}(r_i, \frac{\mu_i}{\sigma_i^2}) = \text{NB}\left(r_i, \frac{r_i}{r_i + (Af^*)_i}\right),$$

where r_i is often referred to as the "dispersion parameter" associated with the ith observation. In this work, we assume that r is constant for each component of the measurement vector y. Our goal is to reconstruct the true sparse/compressible signal of interest f^* from the observed data y. For illustration, see Fig. 1. We note that the negative binomial distribution has been used previously in applications such as matrix factorization [16] and data regression [17].

2. PROBLEM FORMULATION

Under the assumption that the measurements in y are independent and are identically drawn from a negative binomial

This work is partly supported by National Science Foundation grants DMS 1840265 and IIS 1741490.



(a) f^* (b) Af^*

(c) y_{Poisson} (d) y_{NB}

Fig. 1. Example of observation model. (a) True image f^*. (b) The true detector intensity $\mu = Af^*$. Here, A is a blurring operator. (c) Observed measurement y_{Poisson} drawn from a Poisson distribution. (d) Observed measurement y_{NB} drawn from a negative binomial distribution with $r = 9$.

distribution, the probability of observing the data vector y is given by

$$P(y|f) = \prod_{i=1}^{m} \binom{r+y_i-1}{r-1} \left(\frac{r}{r+(Af)_i}\right)^r \left(\frac{(Af)_i}{r+(Af)_i}\right)^{y_i}.$$

Applying a maximum likelihood approach to estimate f^* yields the following constrained optimization problem:

$$\begin{aligned} &\underset{f \in \mathbb{R}^n}{\text{maximize}} \quad P(y|f) \\ &\text{subject to} \quad f \geq 0, \end{aligned}$$

which can be formulated equivalently using a negative log-likelihood objective function as

$$\begin{aligned} &\underset{f \in \mathbb{R}^n}{\text{minimize}} \quad F(f) \equiv \sum_{i=1}^{m} (r+y_i)\log(r+(Af)_i) - y_i \log((Af)_i) \\ &\text{subject to} \quad f \geq 0. \qquad (1) \end{aligned}$$

We impose the an additional constraint on f, namely, that f is structured, e.g., sparse or compressible (see e.g., [1]), which we accomplish by adding a regularization term $\text{pen}(f)$ in the objective function. Our negative binomial sparse reconstruction algorithm takes on the form

$$\begin{aligned} &\underset{f \in \mathbb{R}^n}{\text{minimize}} \quad F(f) + \tau \text{pen}(f) \\ &\text{subject to} \quad f \geq 0, \qquad (2) \end{aligned}$$

where $\tau > 0$ is a regularization parameter that balances the data fitting term $F(f)$ with the sparsity-promoting regularization term $\text{pen}(f)$. In our experiment, we use three types of penalty terms: (1) the ℓ_1-norm $\|f\|_1$, (2) the ℓ_1-norm $\|W^\top f\|_1$ for an orthonormal basis W, and the total variation (TV) norm $\|f\|_{\text{TV}}$.

3. ALGORITHM

Gradient-based approaches are computationally efficient methods for solving optimization problems. Here, the derivatives of $F(f)$ in (2) are given by

$$\nabla F(f) = \sum_{i=1}^{m} \left(\frac{r+y_i}{r+e_i^\top Af} - \frac{y_i}{e_i^\top Af} \right) A^\top e_i$$

$$\nabla^2 F(f) = A^\top \sum_{i=1}^{m} \left[\left(\frac{r+y_i}{(r+e_i^\top Af)^2} - \frac{y_i}{(e_i^\top Af)^2} \right) e_i e_i^\top \right] A$$

where e_i is the ith canonical basis unit vector. To solve (2), we generate and solve a sequence of subproblems of the form

$$\begin{aligned} f^{j+1} = &\underset{f \in \mathbb{R}^n}{\arg\min} \quad F^j(f) + \tau \text{pen}(f) \\ &\text{subject to} \quad f \geq 0, \qquad (3) \end{aligned}$$

where $F^j(f)$ is a quadratic approximation to $F(f)$. Specifically, we use a second-order Taylor series approximation to $F(f)$ around the current iterate f^j given by

$$F^j(f) = F(f^j) + (f - f^j)^\top \nabla F(f^j) + \frac{\alpha_j}{2} \|f - f^j\|_2^2,$$

where the parameter α_j is chosen via Barzilai-Borwein criteria [18] with $\alpha_j I \approx \nabla^2 F(f^j)$. This approach is similar to the framework in [4, 9, 19]. Setting $q^j = f^j - \frac{1}{\alpha_j} \nabla F(f^j)$, then (3) can be written equivalently as

$$\begin{aligned} f^{j+1} = &\underset{f \in \mathbb{R}^n}{\arg\min} \quad \frac{1}{2}\|f - q^j\|_2^2 + \frac{\tau}{\alpha_j} \text{pen}(f) \\ &\text{subject to} \quad f \geq 0. \qquad (4) \end{aligned}$$

The algorithm terminates when consecutive iterates or the corresponding objective function $F(f)$ values do not change significantly.

Here we present three types of penalty terms, $\text{pen}(f)$, for which the solution to (4) can be computed efficiently using the methods in [9].

1. Canonical basis. If f is sparse, then $\text{pen}(f) = \|f\|_1$ is commonly used to promote sparsity [20]. Even though this penalty is not differentiable, the minimizer in (4) is given by the closed-form expression $f^{j+1} = \left[q^j - \frac{\tau}{\alpha_j} \mathbb{1} \right]_+$, where $[\,\cdot\,]_+ \equiv \max\{0, \,\cdot\,\}$ (component-wise) and $\mathbb{1}$ is the vector of ones of appropriate length.

2. Orthonormal basis. If $f = W\theta$, where W is an orthogonal matrix and θ is sparse, then we use the penalty term $\text{pen}(f) = \|W^\top f\|_1 = \|\theta\|_1$. Then, writing $\theta = u - v$, where $u, v \geq 0$, we reformulate (4) as

$$\begin{aligned} f^{j+1} \;=\; \underset{u,v\in\mathbb{R}^n}{\arg\min} \quad & \frac{1}{2}\|W(u-v) - q^j\|_2^2 + \frac{\tau}{\alpha_j}\mathbb{1}^\top(u+v) \\ \text{subject to} \quad & u \geq 0, v \geq 0, W(u-v) \geq 0, \end{aligned}$$

which has a differentiable objective function and can be solved using alternating minimization [21]. In our numerical experiments, we use the wavelet basis transform.

3. Total variation. The total variation norm measures first-order vertical and horizontal differences in magnitude of an image. In our experiments, we use the anisotropic total variation (TV) norm of an $n \times n$ image f, which is given by

$$\|f\|_{\text{TV}} \triangleq \sum_{s=1}^{n-1}\sum_{t=1}^{n} |f_{s,t} - f_{s+1,t}| + \sum_{s=1}^{n}\sum_{t=1}^{n-1} |f_{s,t} - f_{s,t+1}|.$$

The subproblem corresponding to (4) with $\text{pen}(f) = \|f\|_{\text{TV}}$ is solved using the approach in [22].

4. EXPERIMENTS

We conduct three experiments to demonstrate the effectiveness of our proposed approach.

- **Experiment I** consists of reconstructing a one-dimensional signal of length 15,000, where the number of non-zeros is 1500. The measurement y is drawn from a negative binomial distribution with $r = 1$.
- **Experiment II** consists of reconstructing a 2-dimensional image (the modified Shepp-Logan phantom image in MATLAB) with size 128×128, where the observation data y is drawn from a Poisson distribution.
- **Experiment III** is similar to Experiment II, but the observation data y is drawn from a negative binomial distribution with dispersion parameter $r = 9$.

In these experiments, we compare our proposed method to SPIRAL-TAP [9]. We use the percentage root-mean-square error ($\text{RMSE}(\%) = 100\|\hat{f} - f^*\|_2/\|f^*\|_2$) to measure the distance between the computed solution $\hat{f}$ to the true solution f^*. At most 100 iterations were used because no significant improvements are observed past this limit. We used the default parameters in [9]. We used the given disperson parameter r, but this value can be estimated using cross-validation [23], method-of-moment [24], and maximum quasi-likelihood methods [25].

Experiment I: 1D data with canonical basis sparsity

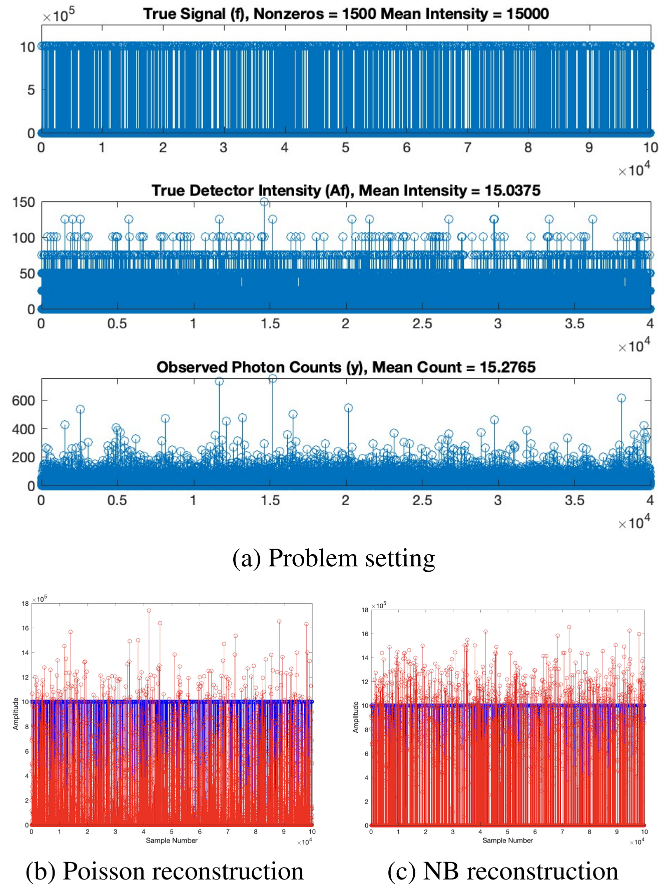


Fig. 2. Results for Experiment I: 1D data with canonical basis sparsity. (a) The true signal f^*, the true detector intensity Af^*, and the observed photon counts y drawn from a negative binomial (NB) distribution with $r = 1$. (b) Reconstruction using a Poisson model with 5015 non-zero components and RMSE = 47.0% (c) Reconstruction using a NB model with 1527 non-zero components and RMSE = 15.2%.

For this experiment, the one-dimensional signal f^* is sparse in the canonical basis. Consequently, we choose the ℓ_1-norm as the penalty term. The measurement data is drawn from a negative binomial distribution with dispersion parameter $r = 1$, which represents a highly overdispersed dataset. The true signal f^*, the true detector intensity Af^*, and the observed photon counts y are shown in Fig. 2(a). The reconstruction using SPIRAL-TAP is shown in Fig. 2(b), which has 5015 non-zero components and an RMSE of 47.0%. The reconstruction using our proposed method is shown in Fig. 2(c), which, in contrast, has 1527 non-zero components and a lower RMSE of 27.2%.

Experiment II: 2D data drawn from a Poisson distribution

For this experiment, we use the 128×128 modified Shepp-Logan phantom image in MATLAB. The observation data y is drawn from a Poisson distribution. We consider both spar-

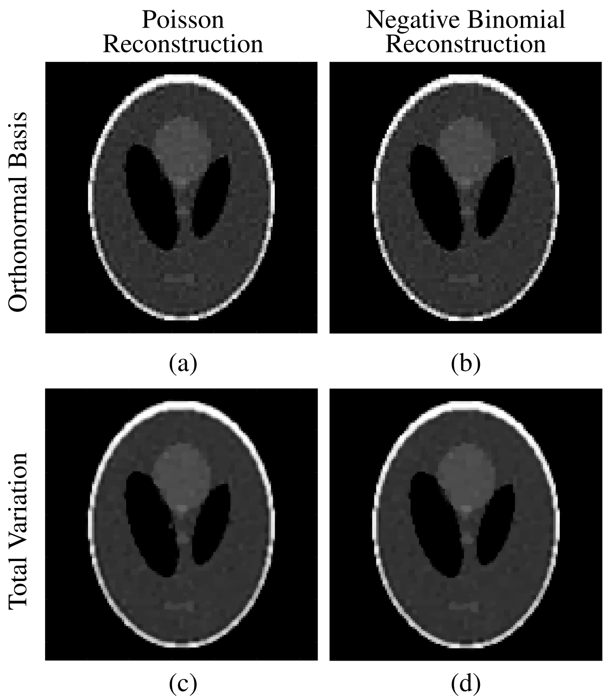


Fig. 3. Results for Experiment II: 2D data drawn from a **Poisson** distribution. (a) The Poisson reconstruction using the orthonormal basis (ONB) penalty with RMSE = 24.7%. (b) The negative binomial (NB) reconstruction using the ONB penalty with RMSE = 24.6%. (c) The Poisson reconstruction using a total variation (TV) penalty with RMSE = 24.6%. (d) The NB reconstruction using a TV penalty with RMSE = 22.9%.

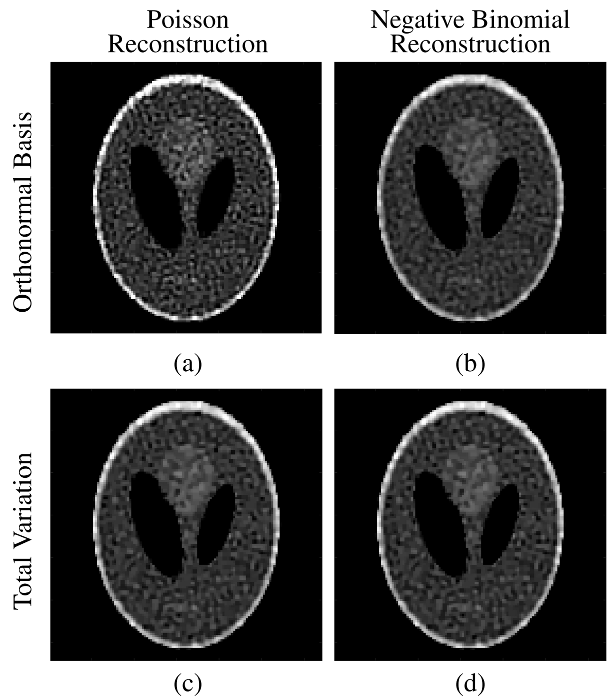


Fig. 4. Results for Experiment III: 2D data drawn from a **negative binomial** distribution with dispersion parameter $r = 9$. (a) The Poisson reconstruction using the ONB penalty with RMSE = 37.7%. (b) The NB reconstruction using the ONB penalty with RMSE = 34.9%. (c) The Poisson reconstruction using a TV penalty with RMSE = 34.8%. (d) The NB reconstruction using a TV penalty with RMSE = 32.7%.

sifying orthonormal basis (ONB) and the total variation (TV) to promote sparsity in the reconstruction. The reconstructions using the Poisson and negative binomial models are presented in Fig. 3. The RMSE for the Poisson reconstructions using the ONB and TV penalties are similar (24.7% and 24.6%, respectively). These values are comparable to the RMSE for the NB reconstruction using the ONB penalty (24.6%). However, the RMSE for the NB reconstruction using the TV penalty is significantly lower (22.9%). For the NB reconstructions, we used a dispersion parameter value of $r = 1,000$ to mimic the Poisson model used to generate the data.

Experiment III: 2D data drawn from an NB distribution

This experiment is similar to Experiment II, but here, the observations are drawn from a negative binomial distribution (with dispersion parameter $r = 9$) instead of a Poisson distribution. The reconstructions using the Poisson and negative binomial models are presented in Fig. 4. Because the measurement data are much noisier (see Fig. 1), we expect the reconstruction RMSEs for both models to be higher than those in Experiment II. However, unlike the results for Experiment II, the RMSE for the Poisson reconstructions using the ONB and TV penalties are not similar (37.7% and 34.8%, respectively). The TV penalty RMSE value is comparable to the RMSE for the NB reconstruction using the ONB penalty (34.9%). However, the RMSE for the NB reconstruction using the TV penalty is significantly lower (32.7%).

5. CONCLUSION

Low-count observations arise in many practical applications of interest, including medical imaging and night imaging. In this work, we presented an alternative to Poisson statistics for modeling low-photon count measurements in the presence of greater variability in a dataset than what is expected. For sparse signal recovery, we proposed a sequential quadratic optimization approach that uses the negative binomial log-likelihood with a sparsifying penalty term to promote sparsity in the reconstruction. We considered three types of penalties (the widely used ℓ_1-norm, the ℓ_1-norm in conjunction with a sparsifying basis, and the total variation norm), whose corresponding quadratic subproblems are efficiently solved. We conducted numerical experiments in 1D and 2D where the measurements are drawn from both Poisson and negative binomial distributions, and in all cases, using a negative binomial model improve the accuracy of the reconstructed signals.

6. REFERENCES

[1] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[2] R. M. Willett, R. F. Marcia, and J. M. Nichols, "Compressed sensing for practical optical imaging systems: a tutorial," *Optical Engineering*, vol. 50, no. 7, pp. 072601–072601, 2011.

[3] L. Stanković, E. Sejdić, S. Stanković, M. Daković, and I. Orović, "A tutorial on sparse signal reconstruction and its applications in signal processing," *Circuits, Systems, and Signal Processing*, vol. 38, pp. 1206–1263, 2019.

[4] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.

[5] D. Ito, S. Takabe, and T. Wadayama, "Trainable ISTA for sparse signal recovery," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3113–3125, 2019.

[6] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6202–6216, 2012.

[7] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, 2008.

[8] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[9] Z. T. Harmany, R. F. Marcia, and R. 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, 2012.

[10] X. Ye, Y. Chen, and F. Huang, "Computational acceleration for MR image reconstruction in partially parallel imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1055–1063, 2010.

[11] M. Banuelos, S. Sindi, and R. F. Marcia, "Negative binomial optimization for biomedical structural variant signal reconstruction," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 906–910.

[12] S. B. Slimane and T. Le-Ngoc, "A doubly stochastic Poisson model for self-similar traffic," in *Proceedings IEEE International Conference on Communications ICC '95*, 1995, vol. 1, pp. 456–460 vol.1.

[13] J.-L. Starck, F. Murtagii, and A. Bijaoui, "Multiresolution support applied to image filtering and restoration," *Graphical models and image processing*, vol. 57, no. 5, pp. 420–431, 1995.

[14] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized em algorithms," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[15] M. Raginsky, R. M. Willett, Z. T. Harmany, and R. F. Marcia, "Compressed sensing performance bounds under poisson noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 3990–4002, 2010.

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

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

[18] J. Barzilai and J. M. Borwein, "Two-Point Step Size Gradient Methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 01 1988.

[19] E. G. Birgin, J. M. Martínez, and M. Raydan, "Nonmonotone spectral projected gradient methods on convex sets," *SIAM Journal on Optimization*, vol. 10, no. 4, pp. 1196–1211, 2000.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[21] Z. Harmany, D. Thompson, R. Willett, and R. F. Marcia, "Gradient projection for linearly constrained convex optimization in sparse signal recovery," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 3361–3364.

[22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[23] C. Gu, "Cross-validating non-Gaussian data," *Journal of Computational and Graphical Statistics*, vol. 1, no. 2, pp. 169–179, 1992.

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

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