Analysis of p-norm Regularized Subproblem Minimization for Sparse Photon-Limited Image Recovery

Aramayis Orkusyan[†], Lasith Adhikari[⋆], Joanna Valenzuela[⋆], and Roummel F. Marcia[⋆]

[†]Department of Mathematics, California State University, Fresno, Fresno, CA 93740, USA

[⋆]Applied Mathematics, University of California, Merced, Merced, CA 95343, USA

Problem Formulation

Setting: Critical to accurate reconstruction of sparse signals from low-dimensional low-photon count observations is the solution of nonlinear optimization problems that promote sparse solutions.

Goal: Analyze zero-finding methods for solving the p-norm regularized minimization subproblems arising from a sequential quadratic approach.

Observation Model

Photon-limited data observations generally follow a Poisson distribution with a certain mean detector photon intensity [1], i.e.,

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{f}^*),$$

where

$\mathbf{y} \in \mathbb{Z}_+^m$ = a vector of observed photon counts,
$\mathbf{f}^* \in \mathbb{R}_+^n$ = the vector of true signal intensity,
$\mathbf{A} \in \mathbb{R}_+^{m\times n}$ = the system matrix.

Approach: The unknown signal $\mathbf{f}^*$ is estimated using the *maximum likelihood principle*.

Nonconvex Optimization Problem

The Poisson reconstruction problem with a non-convex p-norm penalty term has the following constrained optimization form:

$$\begin{aligned} &\underset{\mathbf{f}\in\mathbb{R}^n}{\text{minimize}} \quad \Phi(\mathbf{f}) \equiv F(\mathbf{f}) + \tau \,\|\mathbf{f}\|_p^p \\ &\text{subject to} \quad \mathbf{f} \succeq 0, \end{aligned} \tag{1}$$

where $\tau > 0$ is a regularization parameter, $F(\mathbf{f})$ is the negative Poisson log-likelihood function

$$F(\mathbf{f}) = \mathbf{1}^T\mathbf{A}\mathbf{f} - \sum_{i=1}^{m} y_i \log(\mathbf{e}_i^T\mathbf{A}\mathbf{f} + \beta),$$

where $\mathbf{1}$ is an m-vector of ones, $\mathbf{e}_i$ is the i-th column of the $m \times m$ identity matrix, $\beta > 0$ (typically $\beta \ll 1$), and $\|\mathbf{f}\|_p^p = \sum_{i=1}^n |\mathbf{f}_i|^p$ $(0 \le p < 1)$, which bridges the convex ℓ_1 norm to the ℓ_0 counting seminorm.

Seperable Quadratic Subproblems

To solve the minimization problem in (1), $F(\mathbf{f})$ is approximated by second-order Taylor series expansion, where the Hessian in the Taylor series is replaced by a scaled identity matrix $\alpha_k\mathbf{I}$, where $\alpha_k > 0$ [2, 3, 4]. A simple manipulation to this quadratic approximation will lead into a sequence of subproblems of the form

$$\begin{aligned} \mathbf{f}^{k+1} = \underset{\mathbf{f}\in\mathbb{R}^n}{\arg\min} \quad & \frac{1}{2}\,\|\,\mathbf{f} - \mathbf{s}^k\,\|_2^2 \; + \frac{\tau}{\alpha_k}\|\mathbf{f}\|_p^p \\ \text{subject to} \quad & \mathbf{f} \succeq 0, \end{aligned} \tag{2}$$

where

$$\mathbf{s}^k = \mathbf{f}^k - \frac{1}{\alpha_k}\nabla F(\mathbf{f}^k).$$

Note that the subproblem (2) can be separated into scalar minimization problems of the form

$$\begin{aligned} f_s^* = \underset{f\in\mathbb{R}}{\arg\min} \quad & \Omega_s(f) = \frac{1}{2}(f-s)^2 + \lambda|f|^p, \\ \text{subject to} \quad & f \ge 0. \end{aligned} \tag{3}$$

where f and s denote elements of the vectors $\mathbf{f}$ and $\mathbf{s}^k$ respectively and $\lambda = \tau/\alpha_k$ [5].

Minimizing the Scalar Subproblem

Given a regularization parameter $\lambda > 0$ and p-norm for $\Omega_s(f)$ in (3), there exists a threshold value $\gamma_p(\lambda)$ (that explicitly depends on p and λ) such that if $s \le \gamma_p(\lambda)$, the global minimum of (3) is $f_s^* = 0$; otherwise, the global minimum will be a non-zero value (see Fig. 1).

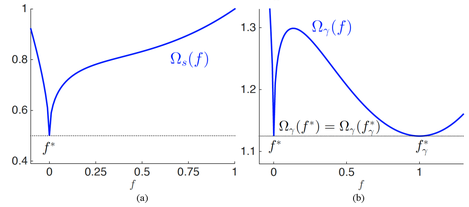

Figure 1: The plot of the scalar quadratic function $\Omega_s(f)$, where $p = 0.5$ and $\lambda = 1.0$. (a) When $s < \gamma_p(\lambda)$, then $f_s^* = 0$ is the unique global minimum. (b) When $s = \gamma_p(\lambda)$, there are global minima at $f^* = 0$ and f_γ^*. If $s > \gamma_p(\lambda)$, then the global minimum is uniquely at some $f_s^* > 0$.

Computing the Threshold Value $\gamma_p(\lambda)$

When $s = \gamma_p(\lambda)$, there exists f_γ^* such that

$$\Omega_\gamma(f_\gamma^*) = \Omega_\gamma(0) \quad \text{and} \quad \Omega_\gamma'(f_\gamma^*) = 0. \tag{4}$$

By solving (4) simultaneously, we can explicitly find the threshold value $\gamma_p(\lambda)$ for given p and λ values [6]. Specifically,

$$\gamma_p(\lambda) = (2\lambda(1-p))^{\frac{1}{2-p}} + \lambda p(2\lambda(1-p))^{\frac{p-1}{2-p}} \quad \text{and} \quad f_\gamma^* = (2\lambda(1-p))^{\frac{1}{2-p}}.$$

For any $s > \gamma_p(\lambda)$, the unique minimum f_s^* of $\Omega_s(f)$ is greater than 0 and is obtained by setting Ω_s' to 0:

$$\Omega_s'(f_s^*) = f_s^* - s + \lambda p(f_s^*)^{p-1} = 0. \tag{5}$$

We now describe zero-finding algorithms to compute the root f_s^*.

Zero-Finding Algorithms

Fixed-point iteration method: The fixed-point iteration method defines a sequence of points $\{f_n\}$ given by $f_{n+1} = G(f_n)$. Previous methods (see, e.g., [5, 6]) for finding the root of $\Omega_s'(f)$ use the fixed point iteration:

$$f_{n+1} = G(f_n) = s - \lambda p f_n^{p-1}.$$

Newton's method: The iterations for Newton's method are given by

$$f_{n+1} = f_n - \frac{\Omega_s'(f_n)}{\Omega_s''(f_n)} = \frac{s f_n^{2-p} + \lambda p(p-2) f_n}{f_n^{2-p} + \lambda p(p-1)}.$$

Initialization

If $s = \gamma_p(\lambda) + \varepsilon$ for some $\varepsilon > 0$, we now analyze how to estimate f_s^* to initialize the above zero-finding methods.

First-order Taylor series approximation: To define the initial point, we can linearize $\Omega_s'(f)$ around f_γ^* and find the zero of the linearization. This leads to the initialization

$$f_s^0 = f_\gamma^* + \delta, \quad \text{where } \delta = \frac{\varepsilon}{1 + \lambda p(p-1)(f_\gamma^*)^{p-2}}.$$

Second-order Taylor series approximation: Similarly, we can use a second-order Taylor approximation to Ω_s' around f_γ^*, which yields the following approximation:

$$f_s^0 = f_\gamma^* + \delta, \quad \text{where } \delta = \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

where $a = \frac{\lambda p(p-1)(p-2)}{2}(f_\gamma^*)^{p-3}$, $b = 1 + \lambda p(p-1)(f_\gamma^*)^{p-2}$, and $c = -\varepsilon$.

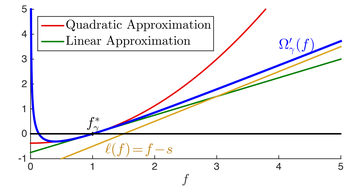

Figure 2: Approximations to $\Omega_\gamma'(f)$ centered at f_γ^*. As f increases, both the linear and quadratic Taylor approximation diverge from $\Omega_\gamma'(f)$. In contrast, the approximation $\ell(f) = f - s$, which are the first two terms in $\Omega_\gamma'(f)$, is more accurate for large values of f.

Bounds on f_s^*

The following lemma allows us to describe the asymptotic behavior of f_s^*.

Lemma 1. *Let $\lambda > 0$ and $0 \le p < 1$. Then for $s \ge \gamma_p(\lambda)$, $\lambda p(1-p)(f_s^*)^{p-2} \le \frac{p}{2}$.*

Theorem 1. *For $\lambda > 0$ and $0 \le p < 1$, the minimizer, f_s^*, of Ω_s is bounded by $f_s^* \le s$. If $0 \le p \le \frac{1}{2}$, then the minimizer is further bounded by $\frac{2}{3}s \le f_s^* \le s$.*

Note that Theorem 1 implies that as s increases, so does f_s^*. Moreover, as $s \to \infty$, $(f_s^*)^{p-1} \to 0$, and therefore, by (5), $f_s^* \to s$. Thus, a sensible initial estimate for f_s^* is s.

Guarantee of Convergence

Fixed point iteration method: Let $e_n = f_n - f^*$ and $e_{n+1} = f_{n+1} - f^*$ represent the errors on the n-th and n+1-th iterations respectively. For fixed point iteration, we have

$$e_{n+1} = f_{n+1} - f^* = G(f_n) - f^* = e_n G'(f^*) + e_n^2 G''(\xi).$$

For small e_n, $e_{n+1} \approx e_n G'(f^*)$. In our context, $G(f) = s - \lambda p f^{p-1}$ and $G'(f) = \lambda p(1-p) f^{p-2}$. By Lemma 1, $G'(f) < 1$. Therefore, the error is decreasing and the fixed point iteration method is guaranteed to converge.

Newton's method: Let f_c be a critical point of $\Omega_s'(f)$ i.e. $\Omega_s''(f_c) = 0$. In particular, $f_c = (\lambda p(1-p))^{\frac{1}{2-p}}$ and for any $f > f_c$, $\Omega_s''(f) = 1 + \lambda p(p-1) f^{p-2} > 0$ i.e. $\Omega_s'(f)$ is increasing in the interval (f_c, ∞). Then, $\Omega_s'''(f) = \lambda p(p-1)(p-2) f^{p-3} > 0$ for all $f \in (0, \infty)$, which implies $\Omega_s'(f)$ is convex. Finally, we note that $f_c < (2\lambda p(p-1))^{\frac{1}{2-p}} = f_\gamma^* \le f^*$, i.e $\Omega_s'(f)$ has a root in (f_c, ∞). Therefore, $\Omega_s'(f)$ is increasing, convex, and has a zero in (f_c, ∞), and Newton's method is guaranteed to converge from any starting point in the interval (f_c, ∞) (see [7]).

Rate of Convergence

For fixed point iteration, the number of iterations required to converge:

$$n_{\text{Fixed Point}} \ge \frac{\ln\varepsilon - \ln|e_0|}{\ln C_1}, \text{where } C_1 = \lambda p(1-p)(f_s^*)^{p-2}.$$

For Newton's method, the number of iterations required to converge:

$$n_{\text{Newton}} \ge \frac{1}{\ln 2}\ln\left(\frac{\ln C_2 + \ln\varepsilon}{\ln C_2 + \ln e_0}\right), \text{where } C_2 = \frac{1}{2}\frac{\lambda p(1-p)(2-p)(f_s^*)^{p-3}}{1 - \lambda p(1-p)(f_s^*)^{p-2}}.$$

Example: When $p = 0.5$, $\lambda = 1$, $\varepsilon = 10^{-8}$, $e_0 = s - f^*$, and $\gamma_p(\lambda) \le s \le 11$, the theoretical number of iterations required to converge is given in Fig. 3.

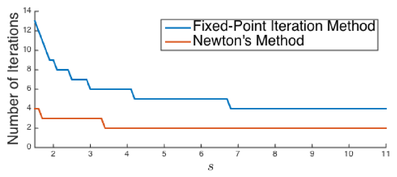

Figure 3: Theoretical number of iterations required to converge as a function of s.

Note that when s is near $\gamma_p(\lambda)$ (e.g., $s \approx 1.5$), the fixed-point iteration method takes many more iterations than Newton's method. For large s, fixed-point iterations only require four iterations. However the number of floating point operations for fixed-point iterations is much smaller than that for Newton's method. Since s can take on any real value, we expect the average performance of fixed-point iteration and Newton's method will be comparable.

Numerical Experiments

We reconstructed a 3D cubic phantom in fluorescence molecular tomography with two embedded fluorescence capillary rods inside it. Fig. 4 shows the true signal ($\mathbf{f}^*$) and our reconstruction.

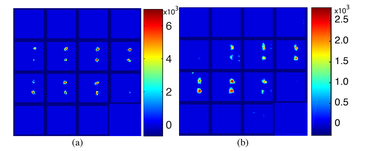

Figure 4: (a) Horizontal slices of a simulated fluorescence capillary rod targets. (b) Reconstruction using p-norm ($p = 0.74$) regularized subproblem minimization.

Method	Time (sec)	Iterations
Fixed-point iteration	21.2829	1,281,974
Newton's method	21.0128	476,585

Table 1: Time and iteration average over 10 trials for fixed-point iteration and Newton's method to reconstruct the fluorescence molecular tomography data.

While Newton's method in theory should converge to the solution faster than fixed-point iterations, the number of floating-point operations needed to perform each iteration offsets the computational advantage of using derivative information.

Acknowledgements

This work was supported by U. S. National Science Foundation Grants CMMI 1333326 and DMS-1359484. We thank Dr. Dianwen Zhu and Prof. Changqing Li for providing the numerical codes for generating the fluorescence molecular tomography data.

References

[1] D. L. Snyder and M. I. Miller, "Random point processes in space and time," *Springer-Verlag, New York, NY*, 1991.

[2] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, 1988.

[3] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms; theory and practice," *Image Processing, IEEE Trans. on*, vol. 21, no. 3, pp. 1084–1096, March 2012.

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

[5] L. Adhikari and R. F. Marcia, "Nonconvex relaxation for Poisson intensity reconstruction," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brisbane Australia, April 2015.

[6] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *2013 IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 217–224.

[7] D. R. Kincaid and E. W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, American Mathematical Society, 3rd edition, 2002.