Projected Nesterov's Proximal-Gradient Signal Recovery from Compressive Poisson Measurements

Renliang Gu and Aleksandar Dogandžić
ECpE Department, Iowa State University
3119 Coover Hall, Ames, IA 50011
email: {renliang,ald}@iastate.edu

***Abstract*—We develop a projected Nesterov's proximal-gradient (PNPG) scheme for reconstructing sparse signals from compressive Poisson-distributed measurements with the mean signal intensity that follows an affine model with known intercept. The objective function to be minimized is a sum of convex data fidelity (negative log-likelihood (NLL)) and regularization terms. We apply sparse signal regularization where the signal belongs to a closed convex set within the domain of the NLL and signal sparsity is imposed using total-variation (TV) penalty. We present analytical upper bounds on the regularization tuning constant. The proposed PNPG method employs *projected* Nesterov's acceleration step, function restart, and an *adaptive* step-size selection scheme that aims at obtaining a good local majorizing function of the NLL and reducing the time spent backtracking. We establish $O(k^{-2})$ convergence of the PNPG method with step-size backtracking only and no restart. Numerical examples demonstrate the performance of the PNPG method.**

I. Introduction

Signal reconstruction from Poisson-distributed measurements with affine model for the mean signal intensity is important for tomographic, astronomic, optical, microscopic, and hyperspectral imaging [1–4]; it corresponds to the Poisson generalized linear model (GLM) with identity link function [5, Sec. I-A]. In this paper, we propose a projected Nesterov's proximal-gradient (PNPG) method for sparse signal reconstruction under this measurement scenario.

We now illustrate the physics behind the measurement model using a positron emission tomography (PET) example. In PET, antipodal detectors placed on rings count photons that arrive simultaneously and are generated when a positron-electron pair annihilates [1]. The received photon numbers are well modeled using Poisson distribution with the mean equal to the corresponding linear projection of the concentration map of a radioactive material, plus an intercept term corresponding to background radiation, scattering effect, accidental coincidence and various other physical effects. Maximum-likelihood (ML) estimation for this problem was first addressed in [6] using the expectation-maximization (EM) algorithm, see also [4] and references therein for EM-algorithm acceleration. Since then, various regularization terms have been added to this signal reconstruction model, including the nonnegativity and total-variation (TV)-norm regularization on the signal and ℓ_1-norm regularization on the signal coefficients in a transform domain [7]. The sparse Poisson-intensity reconstruction algorithm (SPIRAL) [7, 8] approximates the logarithm function in the underlying negative log-likelihood (NLL) by adding a small positive term to it and descends the objective function (regularized NLL) with proximal steps that employ Barzilai-Borwein (BB) [9] step size in each iteration, followed by backtracking. In this paper, we do not approximate the NLL and propose a *projected* Nesterov acceleration of the proximal iteration, with restart and *adaptive* step size.

This work was supported by the NSF under Grant CCF-1421480.

We introduce the notation: "^{T}", $\mathbf{0}$ and $\mathbf{1}$ denote the transpose, vectors of zeros and ones, respectively; "$\succeq$" and "$\succ$" are the elementwise versions of "$\geq$" and "$>$". For a vector $\boldsymbol{a} = [a_1, \ldots, a_N]^T \in \mathbb{R}^N$, define the indicator function, projection operator, nonnegativity projector:

$$\mathbb{I}_C(\boldsymbol{a}) = \begin{cases} 0, & \boldsymbol{a} \in C \\ +\infty, & \text{otherwise} \end{cases} \tag{1a}$$

$$P_C(\boldsymbol{a}) = \arg\min_{\boldsymbol{x} \in C} \|\boldsymbol{x} - \boldsymbol{a}\|_2^2 \tag{1b}$$

$$\big[(\boldsymbol{a})_+\big]_i = \max(a_i, 0) \tag{1c}$$

and the elementwise logarithm and exponential $[\ln_\circ \boldsymbol{a}]_i = \ln a_i$ and $\big[\exp_\circ \boldsymbol{a}\big]_i = \mathrm{e}^{a_i}$. We define the proximal operator for function $r(\boldsymbol{x})$ scaled by λ:

$$\operatorname{prox}_{\lambda r}\, \boldsymbol{a} = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{a}\|_2^2 + \lambda r(\boldsymbol{x}). \tag{2}$$

II. Measurement Model

Consider N independent measurements $\boldsymbol{y} = (y_n)_{n=1}^N$ that follow the Poisson distribution with means $[\Phi\boldsymbol{x} + \boldsymbol{b}]_n$, where $\boldsymbol{x} = (x_i)_{i=1}^p \in C$ is the *unknown* $p \times 1$ signal vector that we wish to reconstruct and $\Phi = [\boldsymbol{\phi}_1\, \boldsymbol{\phi}_2 \cdots \boldsymbol{\phi}_N]^T \in \mathbb{R}_+^{N\times p}$, $\boldsymbol{b}$, and C are the *known* sensing matrix, intercept term, and nonempty closed convex set that $\boldsymbol{x}$ belongs to, respectively; the intercept $\boldsymbol{b}$ models background radiation and scattering determined, e.g., by calibration before the measurements $\boldsymbol{y}$ have been collected. Upon ignoring constant terms, we obtain the NLL function

$$\mathcal{L}(\boldsymbol{x}) = \mathbf{1}^T(\Phi\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{y}) + \sum_{n=1}^N y_n \ln\left(\frac{y_n}{\boldsymbol{\phi}_n^T\boldsymbol{x} + b_n}\right) \tag{3}$$

in the form of generalized Kullback-Leibler (KL) divergence [10]. The NLL $\mathcal{L}(\boldsymbol{x}) : \mathbb{R}^p \mapsto \mathbb{R}_+$ is a convex function of the signal $\boldsymbol{x}$. Assume

$$C \subseteq \operatorname{dom}\mathcal{L}(\boldsymbol{x}) \tag{4}$$

which ensures that $\mathcal{L}(\boldsymbol{x})$ is computable for all $\boldsymbol{x} \in C$; here,

$$\operatorname{dom}\mathcal{L}(\boldsymbol{x}) = \left\{\boldsymbol{x} \in \mathbb{R}^p \;\middle|\; \begin{matrix} \boldsymbol{\phi}_n^T\boldsymbol{x} + b_n \geq 0 \text{ for } y_n = 0 \\ \boldsymbol{\phi}_n^T\boldsymbol{x} + b_n > 0 \text{ for } y_n > 0 \end{matrix}, \forall n \right\}.$$

Nonnegativity signal constraint. In practice, $\boldsymbol{x}$ represents signal density or concentration, which are nonnegative quantities; hence $\boldsymbol{x} \in \mathbb{R}^p_+$. The selection

$$C = \mathbb{R}^p_+ \tag{5a}$$

satisfies (4) when $\boldsymbol{b} \succ \boldsymbol{0}$; here,

$$P_C(\boldsymbol{a}) = P_{\mathbb{R}_+}(\boldsymbol{a}) = (\boldsymbol{a})_+. \tag{5b}$$

We adopt the unconstrained regularization framework and minimize

$$f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + u r(\boldsymbol{x}) \tag{6a}$$

where $u > 0$ is a scalar tuning constant and

$$r(\boldsymbol{x}) = \mathrm{TV}(\boldsymbol{x}) + \mathbb{I}_C(\boldsymbol{x}) \tag{6b}$$

is the regularization term that imposes the TV-domain sparsity and convex-set constraints [11], $\mathrm{TV}(\boldsymbol{x}) \in \{\mathrm{TV}_{\mathrm{iso}}(\boldsymbol{x}), \mathrm{TV}_{\mathrm{ani}}(\boldsymbol{x})\}$,

$$\mathrm{TV}_{\mathrm{iso}}(\boldsymbol{x}) = \sum_{i=1}^{p} \sqrt{\sum_{j \in \mathcal{N}_i} (x_i - x_j)^2} \tag{7a}$$

$$\mathrm{TV}_{\mathrm{ani}}(\boldsymbol{x}) = \sum_{i=1}^{p} \sum_{j \in \mathcal{N}_i} |x_i - x_j| \tag{7b}$$

are isotropic and anisotropic TV penalties, and $\mathcal{N}_i$ is an index set of appropriately chosen neighbors of x_i. For images, each set $\mathcal{N}_i$ consists of two pixels at most, with one below and the other on the right of the ith pixel, if possible [11].

For sufficiently large regularization tuning parameter u, the minimum point

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) \tag{8}$$

reduces to the constant signal $\boldsymbol{1}x_0^\star$ because $\mathrm{TV}(\boldsymbol{1}x_0) = 0$ for any x_0; here, the optimal constant signal level

$$x_0^\star = \arg\min_{x_0} f(\boldsymbol{1}x_0) = \arg \min_{\substack{x_0 \\ \boldsymbol{1}x_0 \in C}} \mathcal{L}(\boldsymbol{1}x_0) \tag{9}$$

is not a function of u. For nonnegative C in (5a), $x_0^\star = \big(\arg\min_{x_0} \mathcal{L}(\boldsymbol{1}x_0)\big)_+$. In the following, we find bounds U on u beyond which $\boldsymbol{x}^\star = \boldsymbol{1}x_0^\star$; these bounds apply to general convex differentiable NLLs.

A. Upper-bounding the regularization constant u

Theorem 1: Consider reconstructing an image $X \in \mathbb{R}^{I \times J}$ by minimizing the penalized NLL (6a) for convex and differentiable NLL $\mathcal{L}(\boldsymbol{x})$ and penalty $r(\boldsymbol{x})$ in (6b), where $\boldsymbol{x} = \mathrm{vec}\, X$. Then, the minimum point $\boldsymbol{x}^\star$ in (8) is a constant image $\boldsymbol{x}^\star = \boldsymbol{1}x_0^\star$ if

$$u \geq U \tag{10}$$

where

$$U = \begin{cases} U_0, & \mathrm{TV}(\boldsymbol{x}) = \mathrm{TV}_{\mathrm{ani}}(\boldsymbol{x}) \\ \sqrt{2}U_0, & \mathrm{TV}(\boldsymbol{x}) = \mathrm{TV}_{\mathrm{iso}}(\boldsymbol{x}) \end{cases} \tag{11a}$$

$$U_0 = \frac{1}{2(I+J-2)} \max\Big[\max_{i<I} \big|w_{i,j}(G)\big|, \max_{j<J} \big|w_{j,i}\big(G^T\big)\big|\Big] \tag{11b}$$

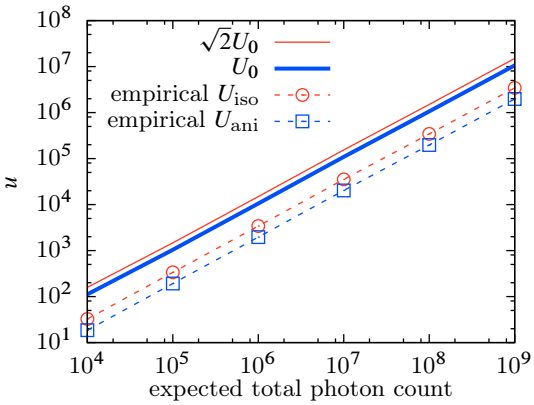


Figure 1: Analytical and empirical upper bounds on u as functions of the total expected photon count $\boldsymbol{1}^T \Phi \boldsymbol{x}$.

and $G = \mathrm{vec}^{-1}\big[\nabla \mathcal{L}(\boldsymbol{1}x_0^\star)\big]$ is an $I \times J$ matrix. Here,

$$w_{k,\ell}(Z) = (2k-1)h_{I_Z,\ell}(Z) - 2(I_Z - 1)h_{k,\ell}(Z) - \big(\mathbb{1}_{\ell>1} + \mathbb{1}_{\ell<J_Z}\big) \sum_{j=1}^{J_Z} h_{k,j}(Z) \tag{12a}$$

$$h_{k,\ell}(Z) = \sum_{i=1}^{k} z_{i,\ell} = \sum_{i=1}^{I_Z} \sum_{j=1}^{J_Z} z_{i,j}\, \mathbb{1}_{k-i\geq 0, \ell-j=0} \tag{12b}$$

where $Z = (z_{i,j})$ is an $I_Z \times J_Z$ matrix, $h_{k,\ell}(\cdot)$ is a partial sum of elements of the ℓth column of Z from the top to the kth entry, and $\mathbb{1}_{\mathcal{A}} = \begin{cases} 1, & \mathcal{A} \text{ true} \\ 0, & \mathcal{A} \text{ false} \end{cases}$.

Proof: See [12]. □

For 1D signal where (7a) and (7b) coincide and the number of columns in X is $J = 1$, U in (10) becomes $U = \max_k \Big(\Big|\sum_{i=1}^{k} \big[\nabla \mathcal{L}\big(\boldsymbol{1}x_0^\star\big)\big]_i\Big|\Big)$.

Fig. 1 shows the analytical bounds U in (11a) and the empirical upper bounds on u (beyond which $\boldsymbol{x}^\star$ is a constant image) as functions of signal-to-noise ratio (SNR) for one noise realization under the measurement scenario in Section IV. Our analytical bounds follow the trends of the empirical bounds well, where the analytical bounds are 4.5 to 6.1 larger than the empirical bounds.

III. Reconstruction

We propose a PNPG approach for minimizing (6a) that combines convex-set projection with Nesterov acceleration [13, 14] and applies function restart and adaptive step size. The pseudo code in Algorithm 1 summarizes our PNPG method.

A. PNPG method

Iteration i of the *PNPG method* proceeds as follows:

$$\theta^{(i)} = \frac{1}{2}\left[1 + \sqrt{1 + 4\big(\theta^{(i-1)}\big)^2}\right] \tag{13a}$$

$$\overline{\boldsymbol{x}}^{(i)} = P_C\Big(\boldsymbol{x}^{(i-1)} + \tfrac{\theta^{(i-1)}-1}{\theta^{(i)}}\big(\boldsymbol{x}^{(i-1)} - \boldsymbol{x}^{(i-2)}\big)\Big) \tag{13b}$$

$$\boldsymbol{x}^{(i)} = \mathrm{prox}_{\beta^{(i)} u r}\Big(\overline{\boldsymbol{x}}^{(i)} - \beta^{(i)} \nabla \mathcal{L}\big(\overline{\boldsymbol{x}}^{(i)}\big)\Big) \tag{13c}$$

Algorithm 1: PNPG method

Input: $\boldsymbol{x}^{(-1)}$, u, $\mathtt{m}$, ξ, η, and threshold ϵ
Output: $\arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$
Initialization: $\theta^{(0)} \leftarrow 0$, $\boldsymbol{x}^{(0)} \leftarrow \mathbf{0}$, $i \leftarrow 0$, $\kappa \leftarrow 0$ and $\beta^{(1)}$ by the BB method
repeat
  $i \leftarrow i+1$ and $\kappa \leftarrow \kappa + 1$
  evaluate (13a) and (13b)
  while *true* **do** `// backtracking search`
    solve the proximal-mapping step (13c)
    if *condition* (14) *holds* **then**
      break
    else
      $\beta^{(i)} \leftarrow \xi\beta^{(i)}$ and $\kappa \leftarrow 0$
  if $i > 1$ *and* $f\left(\boldsymbol{x}^{(i)}\right) > f\left(\boldsymbol{x}^{(i-1)}\right)$ **then** `// restart`
    $\theta^{(i-1)} \leftarrow 1$, $i \leftarrow i-1$, and continue
  if *condition* (17a) *holds with threshold* ϵ **then**
    declare convergence
  if $\kappa \geq \mathtt{m}$ **then** `// adapt step size`
    $\kappa \leftarrow 0$ and $\beta^{(i+1)} \leftarrow \beta^{(i)}/\xi$
  else
    $\beta^{(i+1)} \leftarrow \beta^{(i)}$
until *convergence declared or maximum number of iterations exceeded*

where $\beta^{(i)} > 0$ is an *adaptive step size* chosen to satisfy the *majorization condition*

$$\mathcal{L}\left(\boldsymbol{x}^{(i)}\right) \leq \mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right) + \left(\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}^{(i)}\right)^T \nabla\mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right) + \frac{1}{2\beta^{(i)}}\left\|\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}^{(i)}\right\|_2^2 \tag{14}$$

using the following simple adaptation scheme that aims at keeping $\beta^{(i)}$ as large as possible:

i) • if there have been no step-size backtracking events or increase attempts for $\mathtt{m}$ consecutive iterations ($i-\mathtt{m}$ to $i-1$), start with a larger step size

$$\bar{\beta}^{(i)} = \frac{\beta^{(i-1)}}{\xi} \qquad \text{(increase attempt)} \tag{15a}$$

where $\xi \in (0,1)$ is a *step-size adaptation parameter*;
• otherwise start with

$$\bar{\beta}^{(i)} = \beta^{(i-1)}; \tag{15b}$$

ii) (backtracking search) select

$$\beta^{(i)} = \xi^{t_i}\bar{\beta}^{(i)} \tag{15c}$$

where $t_i \geq 0$ is the smallest integer such that (15c) satisfies the majorization condition (14); *backtracking event* corresponds to $t_i > 0$.

We select the initial step size $\bar{\beta}^{(0)}$ using the BB method [9].

If there has been an attempt to change the step size in any of the previous $\mathtt{m}$ consecutive iterations, we start the backtracking search ii) with the step size from the latest completed iteration. Consequently, the step size will be approximately piecewise-constant as a function of the iteration index i; see Fig. 2,

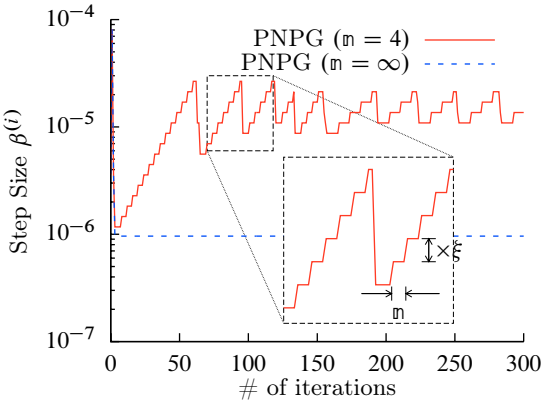


Figure 2: PNPG step size as a function of the iteration index.

which shows the evolution of the adaptive step size $\beta^{(i)}$ for one realization of the noisy measurements in Section IV. Here, $\mathtt{m}$ controls the size of the neighborhood around $\bar{\boldsymbol{x}}^{(i)}$ over which $\mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right) + \left(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(i)}\right)^T \nabla\mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right) + \frac{1}{2\beta^{(i)}}\left\|\boldsymbol{x} - \bar{\boldsymbol{x}}^{(i)}\right\|_2^2$ majorizes $\mathcal{L}(\boldsymbol{x})$ and also the time spent backtracking: larger $\mathtt{m}$ yields a larger neighborhood and leads to less backtracking. A step-size adaptation scheme with $\mathtt{m}=0$ has been used in [14] for basis pursuit denoising (BPDN), i.e., Gaussian linear model NLL and ℓ_1-norm penalty; $\mathtt{m}=0$ means that we always initialize the step-size search aggresively with an increase attempt (15a).

The majorization condition (14) ensures that the iterate $\boldsymbol{x}^{(i)}$ attains lower (or equal) objective function than the intermediate signal $\bar{\boldsymbol{x}}^{(i)}$ [15, Lemma 2.3]:

$$f\left(\boldsymbol{x}^{(i)}\right) \leq f\left(\bar{\boldsymbol{x}}^{(i)}\right) - \frac{1}{2\beta^{(i)}}\left\|\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}^{(i)}\right\|_2^2. \tag{16}$$

However, (16) *does not* guarantee monotonicity of the PNPG iteration: we apply the function restart [16] to restore the monotonicity and improve convergence of the PNPG iteration, see Algorithm 1.

For the proximal mapping in (13c), we use an inner iteration with the TV-based denoising method in [11].

Convergence criteria. The outer- and inner-iteration convergence criteria are

$$\delta^{(i)} \triangleq \left\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i-1)}\right\|_2 < \epsilon\left\|\boldsymbol{x}^{(i)}\right\|_2 \tag{17a}$$

$$\left\|\boldsymbol{x}^{(i,j)} - \boldsymbol{x}^{(i,j-1)}\right\|_2 < \eta\delta^{(i-1)} \tag{17b}$$

where $\epsilon > 0$ is the convergence threshold, j is the inner-iteration index, $\boldsymbol{x}^{(i,j)}$ is the iterate of $\boldsymbol{x}$ in the jth inner iteration step within the ith step of the (outer) PNPG iteration (13), and the convergence tuning constant $\eta \in (0,1)$ is chosen to trade off the accuracy and speed of the inner iterations and provide sufficiently accurate solutions to (13c).

The above iteration can be viewed as an accelerated sequence of iterative denoising steps (13c), where the proximal step (2) performs denoising of the noisy signal $\boldsymbol{a}$ by imposing signal constraints through the regularization term $\lambda r(\boldsymbol{x})$ and the convex-set projection in (13b) ensures that $\mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right)$ and $\nabla\mathcal{L}\left(\bar{\boldsymbol{x}}^{(i)}\right)$ exist, see also (4).

Nesterov's acceleration can be removed by replacing (13a) with the assignment $\theta^{(i-1)} = 1$, which leads to the proximal-gradient (PG) iteration with adaptive step size. In this case,

larger step size $\beta^{(i)}$ renders larger $\delta^{(i)}$, which follows easily from [14, Lemma 2], see also (17a). Consequently, it will be *harder* for the PG method with our adaptive step selection to converge compared with the standard PG method that employs step-size backtracking only (i.e., $\mathbb{m} = +\infty$). This trend is confirmed empirically for our PNPG method as well, compare PNPG with finite and infinite $\mathbb{m}$ in Fig. 4a in Section IV.

We now bound the convergence speed of the PNPG method with step-size backtracking only ($\mathbb{m} = +\infty$) and no restart, which can be thought of as a Poisson compressed-sensing adaptation of fast iterative shrinkage-thresholding algorithm (FISTA) [15]. Although PNPG with finite $\mathbb{m}$ converges faster than PNPG with infinite $\mathbb{m}$ empirically (see Section IV), it is not easy to analyze its convergence rate.

Theorem 2 (Convergence of PNPG with $\mathbb{m} = +\infty$ and no restart): Assuming $\mathbb{m} = +\infty$ and no restart, the convergence of PNPG iterates $\boldsymbol{x}^{(k)}$ to the minimum point $\boldsymbol{x}^\star$ in (8) is bounded as follows:

$$f\left(\boldsymbol{x}^{(k)}\right) - f\left(\boldsymbol{x}^\star\right) \leq \frac{2\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star\right\|_2^2}{\beta^{(k)} k^2} \tag{18}$$

for all $k \geq 1$.

Proof: Along the lines of [15, Th. 4.4] with adjustment made to accommodate the convex-set projection in (13b), based on the nonexpansiveness of the convex projection [17, Th. 9.9]. □

Theorem 2 applies to general scenarios with continuous and convex $\mathcal{L}(\boldsymbol{x})$, as well as convex penalties $r(\boldsymbol{x})$.

Here, in contrast with [15], we leave the right-hand side of (18) as a function of the step size $\beta^{(k)}$ instead of further bounding $\beta^{(k)}$ with the Lipschitz constant of $\mathcal{L}(\boldsymbol{x})$. The reason is infinite Lipschitz constant of (3) when $\boldsymbol{b} = \boldsymbol{0}$.

IV. Numerical Examples

Relative square error (RSE) is adopted as the main metric to assess the performance of the compared algorithms:

$$\mathrm{RSE} = \frac{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_{\mathrm{true}}\|_2^2}{\|\boldsymbol{x}_{\mathrm{true}}\|_2^2} \tag{19}$$

where $\boldsymbol{x}_{\mathrm{true}}$ and $\widehat{\boldsymbol{x}}$ are the true and reconstructed signal, respectively.

Consider PET reconstruction of the 128×128 concentration map $\boldsymbol{x}$ in Fig. 3a, which represents simulated radiotracer activity in human chest. Assume that the corresponding 128×128 attenuation map $\boldsymbol{\kappa}$ is known, given in Fig. 3b, which is needed to model the attenuation of the gamma rays [1] and compute the sensing matrix Φ in this application, see (20) and the following discussion. We collect the photons from 90 equally spaced directions over 180°, with 128 radial samples at each direction. Here, we adopt the parallel strip-integral matrix Γ [18, Ch. 25.2] and use its implementation in the Image Reconstruction Toolbox (IRT) [19], which leads to the sensing matrix Φ:

$$\Phi = w\,\mathrm{diag}\big[\exp_\circ(-\Gamma\boldsymbol{\kappa} + \boldsymbol{c})\big]\Gamma \tag{20}$$

where $\boldsymbol{c}$ is a known vector generated using a zero-mean independent, identically distributed (i.i.d.) Gaussian sequence with variance 0.3 to model the detector efficiency variation, and w is a known scaling constant, which we use to control the expected total number of detected photons, $\mathbf{1}^T\,\mathrm{E}(\boldsymbol{y}) = \mathbf{1}^T\Phi\boldsymbol{x}$, which is an SNR measure. Assume that the background radiation, scattering effect, and accidental coincidence combined together lead to a known (generally nonzero) intercept term $\boldsymbol{b}$. The elements of the intercept term have been set to a constant equal to 10 % of the sample mean of $\Phi\boldsymbol{x}$: $\boldsymbol{b} = \frac{\mathbf{1}^T\Phi\boldsymbol{x}}{10N}\mathbf{1}$.

The above model, choices of parameters in the PET system setup, and concentration map have been adopted from IRT [19, `emission/em_test_setup.m`].

We compare filtered backprojection (FBP) [1] and proximal-gradient methods that aim at minimizing (6a) with isotropic TV norm (7a) and nonnegative $\boldsymbol{C}$ in (5a):

(i) PNPG with adaptive step-size parameters $\mathbb{m} \in \{0, 4, +\infty\}$, $\xi = 0.8$, and inner-iteration convergence constant $\eta = 10^{-3}$,

(ii) SPIRAL with NLL term (3), which improves its numerical stability compared with the original code in [8].

The iterative methods use the convergence criterion (17a) with

$$\epsilon = 10^{-6} \tag{21}$$

have the maximum number of iterations limited to 10^4, and have been initialized by FBP reconstructions implemented in the IRT [19], see also [1].

The NLL centering via generalized KL divergence in (3) is *critical* for numerical stability and avoiding cancellation errors. Accordingly, we have also incorporated this NLL into the SPIRAL code. In particular, we have replaced the Poisson NLL in the SPIRAL paper and code [7, 8] with the generalized KL form (3), which is also slightly more general than the NLL in [7] because it allows for the nonzero intercept term $\boldsymbol{b}$; we keep the small additive constant in logarithm calculations. We observed that the original SPIRAL fails to converge at larger SNRs whereas our modified SPIRAL code converges successfully in these cases. The Matlab implementation of the PNPG and SPIRAL methods and numerical examples is available at https://github.com/isucsp/imgRecSrc.

We adopt the following form of the regularization constant u:

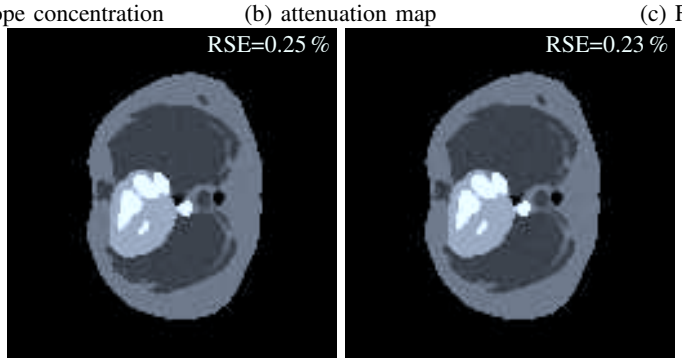
$$u = 10^a \tag{22}$$

vary a in the range $[-6, 3]$ over a grid with spacing 0.5, and seek reconstructions with the smallest RSEs.

Figs. 3c and 3d show reconstructions for one random realization of the noise and detector variation $\boldsymbol{c}$, with the expected total annihilation photon count equal to 10^8 and regularization constant a optimized for RSE performance. The optimal a is 0.5 for both PNPG and SPIRAL. PNPG and SPIRAL perform similarly and clearly outperform FBP; here, PNPG is 3 times faster than SPIRAL.

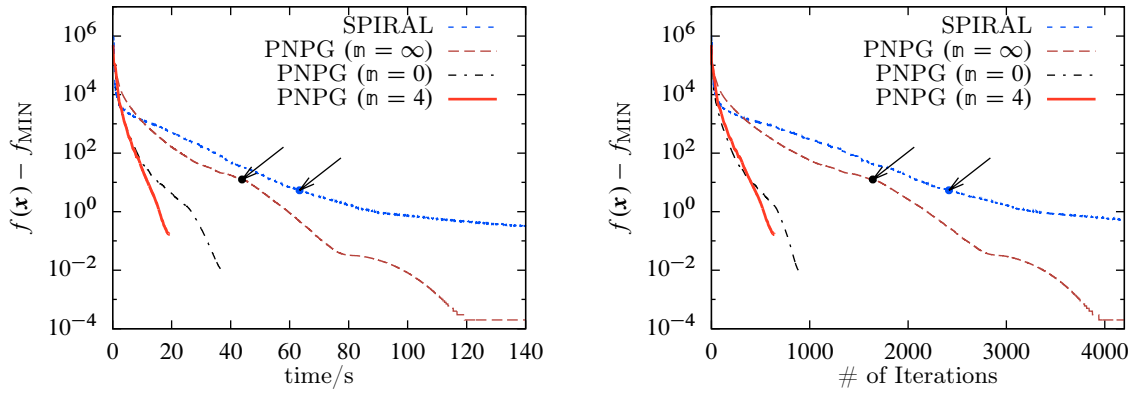
Fig. 4 shows the evolutions of the centered objectives $f(\boldsymbol{x}) - f_{\mathrm{MIN}}$ as functions of CPU time and iteration index for the reconstructions in Fig. 3; here $f_{\mathrm{MIN}} = \min_{\boldsymbol{x}} f(\boldsymbol{x})$. We run PNPG ($\mathbb{m} = \infty$) and SPIRAL past convergence points to illustrate their convergence rates; these convergence points for the threshold in (21) are labeled using arrows. PNPG ($\mathbb{m} = 4$) meets the convergence criterion (21) twice faster than PNPG ($\mathbb{m} = \infty$) and 3 times faster than SPIRAL, and reaches lower objective function than these competitors. The advantage of PNPG ($\mathbb{m} = 4$) over PNPG ($\mathbb{m} = \infty$) is due to its use of step-size adaptation, see also Fig. 2 where the step size of PNPG ($\mathbb{m} = 4$) is consistently larger than

(a) radio-isotope concentration (b) attenuation map (c) FBP



(d) SPIRAL (e) PNPG ($\mathbb{m} = 4$)

Figure 3: (a) True emission concentration map, (b) density map, and (c)–(e) reconstructions of the emission concentration map.



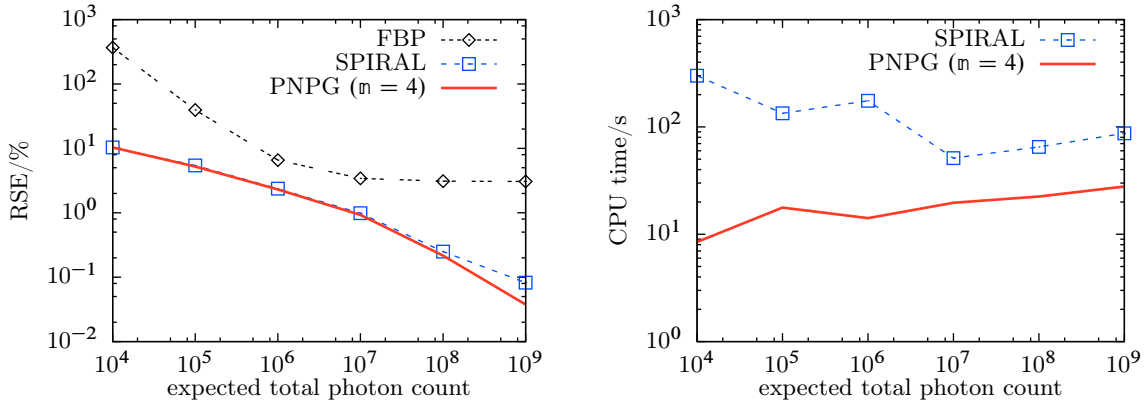
(a) (b)

Figure 4: Centered objective functions as functions of (a) the CPU time and (b) number of iterations.

that of PNPG ($\mathbb{m} = \infty$). Indeed, PNPG ($\mathbb{m} = \infty$) does not adapt to the local curvature of the NLL. The advantage of PNPG ($\mathbb{m} = 4$) over PNPG ($\mathbb{m} = 0$) is due to the *patient* nature of its step-size adaptation, which leads to better local majorizing function of the NLL and reduced time spent backtracking compared with the aggressive PNPG ($\mathbb{m} = 0$) scheme. Fig. 4b presents the centered objectives as functions of the iteration index and hence does not account for time spent backtracking: PNPG ($\mathbb{m} = 4$) remains superior and converges in fewer iterations than PNPG ($\mathbb{m} = 0$).

Fig. 5 shows the minimum average (over 15 random realizations of the noise and detector variation $\boldsymbol{c}$) RSEs and CPU times as functions of the expected total photon counts $\mathbf{1}^T \Phi \boldsymbol{x} \in \{10^4, 10^5, \ldots, 10^9\}$, where a has been selected to minimize the average RSE for each method at each expected total photon count. For both PNPG and SPIRAL the optimal a increases with the SNR: -0.5 for $\mathbf{1}^T \mathrm{E}(\boldsymbol{y}) \in \{10^4, 10^5\}$, 0 for $\mathbf{1}^T \mathrm{E}(\boldsymbol{y}) \in \{10^6, 10^7\}$, 0.5 for $\mathbf{1}^T \mathrm{E}(\boldsymbol{y}) = 10^8$, and 1 for $\mathbf{1}^T \mathrm{E}(\boldsymbol{y}) = 10^9$. As the expected total annihilation photon count increases in Fig. 5a, FBP reaches a performance floor whereas PNPG and SPIRAL continue to improve thanks to the signal sparsity and nonnegativity constraints that they employ. As the SNR increases, the convergence points of SPIRAL and PNPG diverge, which explains the difference between RSEs of the two methods at large SNRs in Fig. 5a. This trend is observed already in Fig. 4a where $\mathbf{1}^T \mathrm{E}(\boldsymbol{y}) = 10^8$.

Fig. 5b shows average CPU times as functions of SNR. For the same convergence threshold (21), PNPG is 3 to 30



(a) (b)

Figure 5: (a) Minimum average RSEs and (b) corresponding average CPU times as functions of the expected total photon counts.

times faster than SPIRAL, where the gap between the two methods is particularly large in the practically important low-SNR scenario.

V. Conclusion

We developed a framework for penalized NLL minimization where the signal that we wish to reconstruct belongs to a nonempty closed convex set within the domain of the NLL: we employ a proximal-gradient scheme with *projected* Nesterov's acceleration, function restart, and *patient adaptive step size selection*. We applied the proposed framework to reconstruct sparse signals from measurements that follow Poisson GLM with identity link and known intercept term. Our PNPG approach is computationally efficient compared with the state-of-the-art SPIRAL method.

Future work includes automating the selection of the regularization parameter u.

References

[1] J. M. Ollinger and J. A. Fessler, "Positron-emission tomography", *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43–55, 1997.

[2] M. Bertero *et al.*, "Image deblurring with Poisson data: From cells to galaxies", *Inverse Prob.*, vol. 25, no. 12, pp. 123006-1–123006-26, 2009.

[3] R. M. Willett *et al.*, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection", *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.

[4] H. Wang and P. Miller, "Scaled heavy-ball acceleration of the Richardson-Lucy algorithm for 3D microscopy image restoration", *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 848–854, 2014.

[5] R. Gu and A. Dogandžić. (Mar. 2015). Reconstruction of nonnegative sparse signals using accelerated proximal-gradient algorithms. version 3. arXiv: 1502.02613 [stat.CO].

[6] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, 1982.

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

[8] *The sparse Poisson intensity reconstruction algorithms (SPIRAL) toolbox*, accessed 27-November-2013. [Online]. Available: http://drz.ac/code/spiraltap.

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

[10] L. Zanni *et al.*, "Numerical methods for parameter estimation in Poisson data inversion", *J. Math. Imaging Vis.*, vol. 52, no. 3, pp. 397–413, 2015.

[11] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems", *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[12] R. Gu and A. Dogandžić, "Upper bounds on total-variation regularization constants", 2015, in preparation.

[13] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$", in *Sov. Math. Dokl.*, vol. 27, 1983, pp. 372–376.

[14] Y. Nesterov, "Gradient methods for minimizing composite functions", *Math. Program.*, vol. 140, no. 1, pp. 125–161, 2013.

[15] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems", *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[16] B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes", *Found. Comput. Math.*, vol. 15, no. 3, pp. 715–732, 2013.

[17] A. Beck, *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. Philadelphia, PA: SIAM, 2014.

[18] J. A. Fessler, *Image reconstruction*, accessed 30-January-2015, 2009. [Online]. Available: http://web.eecs.umich.edu/~fessler/book/a-geom.pdf.

[19] ——, *Image reconstruction toolbox*, accessed 30-October-2014. [Online]. Available: http://www.eecs.umich.edu/~fessler/code.