POISSON IMAGE RESTORATION WITH LIKELIHOOD CONSTRAINT VIA HYBRID STEEPEST DESCENT METHOD

Shunsuke Ono and Isao Yamada

Department of Communications and Computer Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan

ABSTRACT

This paper proposes a likelihood constrained optimization framework for Poisson image restoration. The likelihood constrained problem considered in this paper is the minimization of convex priors over the level set of the negative-log-likelihood function of the Poisson distribution. It has advantages in parameter selection compared with the minimization of the weighted sum of convex priors and the negative-log-likelihood function, which has been used in conventional methods. The level set is characterized as the fixed point set of a certain quasi-nonexpansive operator, which enables us to apply the hybrid steepest descent method to solve the constrained problem. The proposed framework not only can handle the level set of any convex function whose subgradient is available but also does not require any computationally-expensive procedure such as operator inversion and inner loop. Illustrative numerical examples are also presented.

Index Terms— Poisson image restoration, likelihood constrained optimization, fixed point set characterization, hybrid steepest descent method

1. INTRODUCTION

Image restoration from observations contaminated by Poisson noise is a longstanding problem in various applications from astronomical imaging to medical imaging. State-of-the-art methods for Poisson image restoration, e.g., [1, 2, 3, 4, 5], are based on the minimization of the weighted sum of convex priors and the negative-log-likelihood function of the Poisson distribution, where the convex priors promote some desired property based on a-priori knowledge on the unknown original image, and the negative-log-likelihood function, often called the *generalized Kullback-Leibler (G-KL) divergence* [5], plays a role of the data-fidelity. We refer to this formulation as the *unconstrained problem*. One of the main reasons why these studies adopted the unconstrained problem is that the *proximity operator* of the G-KL divergence is available, resulting in an efficient resolution of the unconstrained problem via convex optimization algorithms of using proximal approaches. Indeed, the *Douglas-Rachford splitting method* [6] and the *alternating direction method of multipliers* [7] have been used for solving the unconstrained problem in the existing methods.

Another possible formulation for Poisson image restoration is the minimization of convex priors over the lower level set of the G-KL divergence, i.e., the upper level set of the likelihood function (possibly with other constraints). We refer to this formulation as the *constrained problem*. The constrained problem has some advantages compared with the unconstrained problem in parameter selection. In the unconstrained problem, we need a careful tuning of the weight that determines the relative importance between the G-KL divergence and the convex priors, in order to obtain a reasonable result. However, this is quite difficult because the connection between their values is totally unclear. On the other hand, in the constrained problem, what we have to select is the level of the G-KL divergence, and it has much clearer meaning than the weight in the unconstrained problem. This is because the level of the G-KL divergence directly represents the likelihood. In other words, solving the constrained problem is equivalent to promoting some desired property based on a-priori knowledge while keeping the likelihood at a certain level. This is quite intuitive, and we can easily utilize some information/criteria to adjust the level of the G-KL divergence independent of the sort of convex priors. However, because of the computational difficulty of the projection onto the level set of the G-KL divergence, convex optimization algorithms of using proximal approaches [6, 7, 8, 9, 10] cannot be directly applied to the constrained problem. It should be remarked that there have been proposed several studies that consider to handle the level set of the G-KL divergence via proximal approaches [11, 12, 13]. We will discuss in detail how our work relates to these prior work in Section 5.

In this paper, we propose an optimization framework for Poisson image restoration based on the constrained problem. We first give an explicit formulation of the constrained problem where multiple convex priors can be treated together with the level set of the G-KL divergence so as to accept various problem design for effective restoration. Second, the level set is exactly characterized as the fixed point set of a newly introduced quasi-nonexpansive operator inspired by the *subgradient projection*. It leads to an equivalent expression of the constrained problem, that is, the minimization of the convex priors over the intersection of the fixed point sets of certain quasi-nonexpansive operators. Third, we reformulate it in a certain product space in order to circumvent the computational difficulty due to the composition of degradation operators (e.g., blur). Finally, we present an efficient algorithmic solution to the reformulated constrained problem via the hybrid steepest descent method [14, 15, 16], which is free from computationally expensive procedures such as operator inversion and inner iteration.

2. POISSON OBSERVATION MODEL

Throughout the paper, let $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{N}$ denote the sets of all real, nonnegative real numbers, and nonnegative integers, respectively. Consider the following Poisson observation model:

$$\mathbf{v} = \mathcal{D}_P(\mathbf{\Phi}\hat{\mathbf{u}}), \tag{1}$$

We thank the reviewers for their careful reading and valuable comments. This work is supported in part by JSPS Grants-in-Aid for JSPS fellows (24·2522) and (B-21300091).

where $\hat{\mathbf{u}} \in \mathbb{R}_+^N$ ($N \in \mathbb{N}$ is the number of pixels) is the unknown original image, $\mathbf{\Phi} : \mathbb{R}^N \to \mathbb{R}^M$ a linear degradation operator such that $\mathbf{\Phi}\hat{\mathbf{u}} \in \mathbb{R}_+^M$, $\mathcal{D}_P$ a Poisson noise contamination process, and $\mathbf{v} = [v_1, \ldots, v_M]^t \in \mathbb{N}^M$ an observation ($(\cdot)^t$ stands for the transposition). In this model, $\mathbf{v}$ is assumed to be a sample of $M (\leq N)$ dimensional independent Poisson random vector $\mathbf{V}$ with the following probability distribution:

$$P(\mathbf{V} = \mathbf{v}|\boldsymbol{\lambda}) = \prod_{i=1}^{M} \left[\frac{(\alpha\lambda_i)^{v_i}}{v_i!} \exp(-\alpha\lambda_i) \right], \tag{2}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_M]^t = \mathbf{\Phi}\hat{\mathbf{u}} \in \mathbb{R}_+^M$, and $\alpha \in (0, \infty)$ is a scaling parameter which determines the noise intensity. The negative log-likelihood function is derived as

$$-\ln P(\mathbf{V} = \mathbf{v}|\boldsymbol{\lambda}) = \sum_{i=1}^{M} \left[-v_i \ln \alpha\lambda_i + \ln v_i! + \alpha\lambda_i \right]. \tag{3}$$

Taking into account the case that $v_i = 0$ and the constant $\ln v_i!$, a data-fidelity function $D_{\mathbf{v},\alpha} \in \Gamma_0(\mathbb{R}^M)$[1] for the Poisson noise contamination (2) is obtained as follows:

$$\begin{aligned} &D_{\mathbf{v},\alpha} : \mathbb{R}^M \to (-\infty, \infty] \\ &: \mathbf{x} \mapsto \sum_{i=1}^{M} \begin{cases} \alpha x_i - v_i \ln \alpha x_i, & \text{if } v_i > 0 \text{ and } x_i > 0, \\ \alpha x_i, & \text{if } v_i = 0 \text{ and } x_i \geq 0, \\ \infty, & \text{otherwise,} \end{cases} \end{aligned} \tag{4}$$

which is the so-called *generalized Kullback-Leibler (G-KL) divergence* [5]. We will use the *subdifferential*[2] of the G-KL divergence $\partial D_{\mathbf{v},\alpha}$ given by

$$\begin{aligned} &\partial D_{\mathbf{v},\alpha} : \mathbb{R}^M \to 2^{\mathbb{R}^M} \\ &: x_i \mapsto \begin{cases} \alpha - \frac{v_i}{x_i}, & \text{if } v_i > 0 \text{ and } x_i > 0, \\ \alpha, & \text{if } v_i = 0 \text{ and } x_i > 0, \\ (-\infty, \alpha], & \text{if } v_i = 0 \text{ and } x_i = 0, \\ \varnothing, & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

A selection of the *subgradient* of $D_{\mathbf{v},\alpha}$ at $\mathbf{x}$ is denoted by $D'_{\mathbf{v},\alpha}(\mathbf{x}) \in \partial D_{\mathbf{v},\alpha}(\mathbf{x})$.

3. PROPOSED METHOD

3.1. Problem Formulation

Define the level set of the G-KL divergence for the level $\rho \in \mathbb{R}$ as

$$\text{lev}_{\leq\rho} D_{\mathbf{v},\alpha} := \{\mathbf{x} \in \mathbb{R}^M |\ D_{\mathbf{v},\alpha}(\mathbf{x}) \leq \rho\}, \tag{6}$$

which plays a role of the data-fidelity (i.e., likelihood) constraint. Then the constrained problem for Poisson image restoration based on the observation model (1) is formulated as follows.

Problem 3.1 (Likelihood constrained problem)**.**

$$\text{Find } \mathbf{u}^\star \in \arg \min_{\substack{\mathbf{u} \in C_{255} \\ \mathbf{\Phi}\mathbf{u} \in \text{lev}_{\leq\rho} D_{\mathbf{v},\alpha}}} \sum_{k=1}^{K} {}^{\gamma} f_k(\mathbf{M}_k \mathbf{u}), \tag{7}$$

where

$$C_{255} := \{\mathbf{x} \in \mathbb{R}^N |\ x_i \in [0, 255] \text{ for } i = 1, \ldots, N\}, \tag{8}$$

is a numerical constraint for eight-bit grayscale images, $\mathbf{M}_k : \mathbb{R}^N \to \mathcal{M}_k$ ($k = 1, \ldots, K$) are bounded linear operators ($\mathcal{M}_k$ ($k = 1, \ldots, K$) are finite dimensional real Hilbert spaces), $f_k \in \Gamma_0(\mathcal{M}_k)$ ($k = 1, \ldots, K$) (possibly nonsmooth) convex functions whose proximity operators[3] are available, and ${}^{\gamma} f_k$ ($k = 1, \ldots, K$) their Moreau envelope[4].

Remark 3.1 (Note on Problem 3.1)**.**

- Each function $f_k(\mathbf{M}_k \cdot)$ is designed based on a-priori knowledge on the unknown original image $\hat{\mathbf{u}}$. Useful examples of $f_k(\mathbf{M}_k \cdot)$ are the ℓ^1 norm of some tight frame and the total variation [18]. In the former case, f_k corresponds the ℓ^1 norm and $\mathbf{M}_k$ the tight frame. In the latter case, $f_k : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+ : (\mathbf{x}, \mathbf{y}) \mapsto \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} \sqrt{x_{i,j} + y_{i,j}}$ and $\mathbf{M}_k : \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N : \mathbf{x} \mapsto (\mathbf{D}_v \mathbf{x}, \mathbf{D}_h \mathbf{x})$, where $\mathbf{D}_v, \mathbf{D}_h$ are vertical and horizontal discrete gradient operators, and $n_v \times n_h$ the image size ($n_v n_h = N$). Note that Problem 3.1 can incorporate multiple convex priors together, so that it accepts various design of the objective function.
- If we choose a sufficiently small γ (e.g., $\gamma \leq 0.001$), the efficacy of ${}^{\gamma} f_k$ is almost same as the original nonsmooth function f_k, as stated in [16] (this is also confirmed by the experimental results in Section 4).

3.2. Fixed Point Characterization

To make Problem 3.1 tractable, as the first step, we characterize $\text{lev}_{\leq\rho} D_{\mathbf{v},\alpha}$ as the fixed point set[5] of the following operator:

$$T_{\text{G-KL}} := T_{\text{sp}D_{\mathbf{v},\alpha}} \circ P_{C_B^{\mathbf{v}}}, \tag{11}$$

where

$$P_{C_B^{\mathbf{v}}} : \mathbb{R}^M \to C_B^{\mathbf{v}} : x_i \mapsto \begin{cases} L, & \text{if } x_i < L \text{ and } v_i > 0, \\ 0, & \text{if } x_i < 0 \text{ and } v_i = 0, \\ U, & \text{if } x_i > U, \\ x_i, & \text{otherwise} \end{cases} \tag{12}$$

[1] Let $\mathcal{H}$ be a real Hilbert space equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$. A function $f : \mathcal{H} \to (-\infty, \infty]$ is called *proper lower semicontinuous convex* if $\text{dom}(f) := \{\mathbf{x} \in \mathcal{H} |\ f(\mathbf{x}) < \infty\} \neq \emptyset$, $\text{lev}_{\leq\alpha}(f) := \{\mathbf{x} \in \mathcal{H} |\ f(\mathbf{x}) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$, and $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ and $\lambda \in (0, 1)$, respectively. The set of all proper lower semicontinuous convex functions on $\mathcal{H}$ is denoted by $\Gamma_0(\mathcal{H})$.

[2] Let $f : \mathcal{H} \to (-\infty, \infty]$ be proper. The subdifferential of f is $\partial f : \mathcal{H} \to 2^{\mathcal{H}} : \mathbf{x} \to \{\mathbf{g} \in \mathcal{H} |\ (\forall \mathbf{y} \in \mathcal{H})\ \langle \mathbf{y} - \mathbf{x}, \mathbf{g} \rangle + f(\mathbf{x}) \leq f(\mathbf{y})\}$.

[3] For any $\gamma \in (0, \infty)$, the *proximity operator* of index γ of $f \in \Gamma_0(\mathcal{H})$ is defined by

$$\text{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H} : \mathbf{x} \mapsto \arg \min_{\mathbf{y} \in \mathcal{H}} \left\{ f(\mathbf{y}) + \frac{1}{2\gamma} \|\mathbf{x} - \mathbf{y}\|^2 \right\}, \tag{9}$$

[4] Every $f \in \Gamma_0(\mathcal{H})$ can be approximated with any accuracy by a differentiable convex function:

$${}^{\gamma} f(\mathbf{x}) := \min_{\mathbf{y} \in \mathcal{H}} \left\{ f(\mathbf{y}) + \frac{1}{2\gamma} \|\mathbf{x} - \mathbf{y}\|_2^2 \right\}, \tag{10}$$

which is called the *Moreau envelope* [17, 16] of index $\gamma \in (0, \infty)$ of f, and its gradient $\nabla^{\gamma} f(\mathbf{x}) = \frac{\mathbf{x} - \text{prox}_{\gamma f}(\mathbf{x})}{\gamma}$ is $\frac{1}{\gamma}$-Lipschitzian, i.e., $\|\nabla^{\gamma} f(\mathbf{x}) - \nabla^{\gamma} f(\mathbf{y})\| \leq \frac{1}{\gamma} \|\mathbf{x} - \mathbf{y}\|$. In addition, the chain rule yields $\nabla({}^{\gamma} f \circ \mathbf{M}) = \mathbf{M}^* \circ (\nabla^{\gamma} f) \circ \mathbf{M}$, where $(\cdot)^*$ stands for the adjoint.

[5] The fixed point set of an operator $T : \mathcal{H} \to \mathcal{H}$ is defined by $\text{Fix}(T) := \{\mathbf{x} \in \mathcal{H} |\ T(\mathbf{x}) = \mathbf{x}\}$.

is the projection onto

$$C_B^{\mathbf{v}} := \left\{ \mathbf{x} \in \mathbb{R}^M \,\middle|\, \begin{smallmatrix} x_i \in [L,U], \text{ if } v_i > 0 \\ x_i \in [0,U], \text{ if } v_i = 0 \end{smallmatrix} \right\}, \tag{13}$$

$$\begin{aligned} &T\mathrm{sp}_{D_{\mathbf{v},\alpha}} : C_B^{\mathbf{v}} \to \mathbb{R}^M : \\ &\mathbf{x} \mapsto \begin{cases} \mathbf{x} - \frac{D_{\mathbf{v},\alpha}(\mathbf{x}) - \rho}{\|D'_{\mathbf{v},\alpha}(\mathbf{x})\|_2^2} D'_{\mathbf{v},\alpha}(\mathbf{x}), & \text{if } D_{\mathbf{v},\alpha}(\mathbf{x}) > \rho, \\ \mathbf{x}, & \text{otherwise,} \end{cases} \end{aligned} \tag{14}$$

a variant of the so-called *subgradient projection*, $\|\cdot\|_2$ the ℓ^2 norm, and $L, U \in (0, \infty)$. The operator $T\mathrm{sp}_{D_{\mathbf{v},\alpha}}$ is well-defined on $C_B^{\mathbf{v}} \subset \mathrm{dom} D_{\mathbf{v},\alpha} := \{\mathbf{x} \in \mathbb{R}^M | \ D_{\mathbf{v},\alpha} < \infty\}$. Concatenating $T\mathrm{sp}_{D_{\mathbf{v},\alpha}}$ and $P_{C_B^{\mathbf{v}}}$ as (11), the obtained operator $T_{\text{G-KL}}$ is well-defined over $\mathbb{R}^M$. Moreover, by choosing sufficiently small L and large U, the set $C_B^{\mathbf{v}}$ includes $\mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha}$, i.e., $\mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha} \cap C_B^{\mathbf{v}} = \mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha}$. Indeed, in the case that the observation $\mathbf{v}$ is an eight-bit image, we see that $L = 1$ and $U = 255$ satisfy the condition. This is because, in such a case, if $v_i > 0$ then the corresponding x_i is same/lager than 1 due to that the distribution (2) ensures $x_i > 0$, and each pixel of an eight-bit image takes a nonnegative integer.

Proposition 3.1. *Assume that* $\mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha} \cap P_{C_B^{\mathbf{v}}} \neq \emptyset$. *Then the operator* $T_{\text{G-KL}}$ *is attracting quasi-nonexpansive with* $\mathrm{Fix}(T_{\text{G-KL}}) = \mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha} \cap P_{C_B^{\mathbf{v}}}$, *i.e., for* $\mathbf{x} \in \mathbb{R}^M \setminus \mathrm{Fix}(T_{\text{G-KL}})$ *and* $\mathbf{y} \in \mathrm{Fix}(T_{\text{G-KL}})$,

$$\|T_{\text{G-KL}}(\mathbf{x}) - \mathbf{y}\|_2 < \|\mathbf{x} - \mathbf{y}\|_2.$$

By Proposition 3.1, we can rewrite Problem 3.1 into the following problem:

$$\text{Find } \mathbf{u}^\star \in \arg \min_{\substack{\mathbf{u} \in \mathrm{Fix}(P_{C_{255}}) \\ \mathbf{\Phi}\mathbf{u} \in \mathrm{Fix}(T_{\text{G-KL}})}} \sum_{k=1}^{K} {}^\gamma f_k(\mathbf{M}_k \mathbf{u}), \tag{15}$$

where $P_{C_{255}}$ is the projection onto C_{255} given by

$$P_{C_{255}} : \mathbb{R}^N \to \mathbb{R}^N : x_i \mapsto \begin{cases} 0, & \text{if } x_i < 0, \\ x_i, & \text{if } 0 \leq x_i \leq 255, \\ 255, & \text{if } x_i > 255, \end{cases} \tag{16}$$

and $\mathrm{Fix}(P_{C_{255}}) = C_{255}$.

3.3. Reformulation on Product Space

The problem (15) is still not manageable because of the composition of $\mathbf{\Phi}$. To resolve this difficulty, we further reformulate it in the product space $\mathcal{X} := \mathbb{R}^N \times \mathbb{R}^M$ equipped with the inner product $\langle (\mathbf{u}, \boldsymbol{\vartheta}), (\mathbf{u}', \boldsymbol{\vartheta}') \rangle_{\mathcal{X}} := \mathbf{u}^t \mathbf{u}' + \boldsymbol{\vartheta}^t \boldsymbol{\vartheta}'$ for $(\mathbf{u}, \boldsymbol{\vartheta}), (\mathbf{u}', \boldsymbol{\vartheta}') \in \mathcal{X}$ and its induced norm $\|\cdot\|_{\mathcal{X}}$. Define

$$T_{\mathcal{X}} : \mathcal{X} \to \mathcal{X} : (\mathbf{u}, \boldsymbol{\vartheta}) \mapsto (P_{C_{255}}(\mathbf{u}), T_{\text{G-KL}}(\boldsymbol{\vartheta})). \tag{17}$$

Then, by Proposition 3.1 and the firm nonexpansivity[6] of $P_{C_{255}}$, we see that $T_{\mathcal{X}}$ is attracting quasi-nonexpansive with

$$\mathrm{Fix}(T_{\mathcal{X}}) = \mathrm{Fix}(P_{C_{255}}) \times \mathrm{Fix}(T_{\text{G-KL}}). \tag{18}$$

[6] A operator T is called *nonexpansive* if, for every $\mathbf{x}, \mathbf{y} \in \mathcal{H}$,

$$\|T(\mathbf{x}) - T(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

In addition, T is called *firmly nonexpansive* if $2T - I$ is nonexpansive (I denotes the identity operator). Note that the firm nonexpansivity implies the attracting quasi-nonexpansivity.

Incidentally, by letting $\Psi(\mathbf{u}, \boldsymbol{\vartheta}) := \frac{1}{2}\|\mathbf{\Phi}\mathbf{u} - \boldsymbol{\vartheta}\|_2^2$, we have

$$\begin{aligned} \{(\mathbf{u}, \mathbf{\Phi}\mathbf{u}) \in \mathcal{X} | \ \mathbf{u} \in \mathbb{R}^N\} &= \arg \min_{(\mathbf{u}, \boldsymbol{\vartheta}) \in \mathcal{X}} \Psi(\mathbf{u}, \boldsymbol{\vartheta}) \\ &= \mathrm{Fix}(I - \mu \nabla \Psi), \end{aligned} \tag{19}$$

and Fact 17.5 in [16] generates

$$\begin{aligned} &(I - \mu \nabla \Psi) : \mathcal{X} \to \mathcal{X} : \\ &(\mathbf{u}, \boldsymbol{\vartheta}) \mapsto (\mathbf{u}, \boldsymbol{\vartheta}) - \mu(\mathbf{\Phi}^t \mathbf{\Phi}\mathbf{u} - \mathbf{\Phi}^t \boldsymbol{\vartheta}, \boldsymbol{\vartheta} - \mathbf{\Phi}\mathbf{u}), \end{aligned} \tag{20}$$

which is nonexpansive if we use $\mu \in (0, \frac{2}{\kappa})$ for a Lipschitz constant $\kappa \in (0, \infty)$, e.g., $\|\mathbf{A}\|_{op}^2$, where $\mathbf{A} := [\mathbf{\Phi} \ \ -\mathbf{I}_M] \in \mathbb{R}^{M \times (M+N)}$, $\mathbf{I}_M$ the $M \times M$ identity matrix, and $\|\cdot\|_{op}$ the operator norm. By (18), (19), and Proposition 4.35 of [19], we see that

$$\begin{aligned} &\mathbf{u} \in \mathrm{Fix}(P_{C_{255}}) \text{ and } \mathbf{\Phi}\mathbf{u} \in \mathrm{Fix}(T_{\text{G-KL}}) \\ &\Leftrightarrow (\mathbf{u}, \boldsymbol{\vartheta}) \in \mathrm{Fix}((I - \mu \nabla \Psi) \circ T_{\mathcal{X}}). \end{aligned} \tag{21}$$

Proposition 3.2. *Let* $\mathcal{U}$ *be a finite dimensional real Hilbert space equipped with the standard inner product* $\langle \cdot, \cdot \rangle$ *and its induced norm* $\|\cdot\|$, $R : \mathcal{U} \to \mathcal{U}$ *a nonexpansive operator,* $T : \mathcal{U} \to \mathcal{U}$ *an attracting quasi-nonexpansive operator, and* $\triangleright(\mathrm{Fix}(R \circ T), r) := \{\mathbf{x} \in \mathcal{H} | \ \inf_{\mathbf{y} \in \mathrm{Fix}(R \circ T)} \|\mathbf{x} - \mathbf{y}\| \geq r\}$. *Assume that* $\mathrm{Fix}(R \circ T)$ *is nonempty, bounded, and closed. Then* $R \circ T$ *is quasi-shrinking with* $\mathrm{Fix}(R \circ T)$ *on any bounded closed convex set* E *such that* $\mathrm{Fix}(R \circ T) \subset E$, *i.e.,*

$$\begin{aligned} &K : r \in [0, \infty) \mapsto \\ &\begin{cases} \inf_{\mathbf{x} \in \triangleright(\mathrm{Fix}(R \circ T), r) \cap E} \Big\{ \inf_{\mathbf{y} \in \mathrm{Fix}(R \circ T)} \|\mathbf{x} - \mathbf{y}\| \\ \qquad - \inf_{\mathbf{y} \in \mathrm{Fix}(R \circ T)} \|R \circ T(\mathbf{x}) - \mathbf{y}\| \Big\}, \\ \qquad \qquad \textit{if } \triangleright(\mathrm{Fix}(R \circ T), r) \cap E \neq \emptyset, \\ \infty, \qquad \qquad \textit{otherwise}. \end{cases} \end{aligned}$$

satisfies $K(r) = 0 \Leftrightarrow r = 0$.

By the attracting quasi-nonexpansivity of $T_{\mathcal{X}}$ with $\mathrm{Fix}(T_{\mathcal{X}})$, the nonexpansivity of $(I - \mu \nabla \Psi)$, and Proposition 3.2, we see that the operator $(I - \mu \nabla \Psi) \circ T_{\mathcal{X}}$ is *quasi-shrinking* with $\mathrm{Fix}((I - \mu \nabla \Psi) \circ T_{\mathcal{X}})$ on any bounded closed convex set including it. Note that the boundedness and closedness of $\mathrm{Fix}((I - \mu \nabla \Psi) \circ T_{\mathcal{X}})$ immediately follow from those of C_{255} and $\mathrm{lev}_{\leq\rho} D_{\mathbf{v},\alpha} \cap P_{C_B^{\mathbf{v}}}$. Finally, the problem (15) (i.e., Problem 3.1) is reformulated as follows:

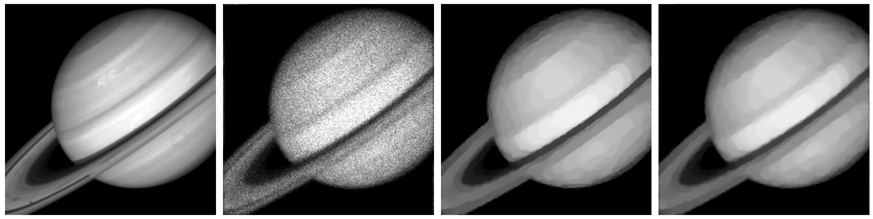
$$\begin{aligned} &\text{Find } (\mathbf{u}^\star, \boldsymbol{\vartheta}^\star) \in \Omega \\ &\qquad := \arg \min_{(\mathbf{u}, \boldsymbol{\vartheta}) \in \mathrm{Fix}((I - \mu \nabla \Psi) \circ T_{\mathcal{X}})} F(\mathbf{u}, \boldsymbol{\vartheta}), \end{aligned} \tag{22}$$

where $F(\mathbf{u}, \boldsymbol{\vartheta}) := \sum_{k=1}^{K} {}^\gamma f_k(\mathbf{M}_k \mathbf{u})$.

3.4. Algorithm

We solve the problem (22) via the *hybrid steepest descent method* [14, 16]. The hybrid steepest descent method can minimize a differentiable convex function over the fixed point set of a quasi-shrinking operator. Obviously, the problem (22) is a special case of such a formulation, so that an algorithmic solution to (22) via the hybrid steepest descent method is obtained as follows:

$$\begin{aligned} &(\mathbf{u}^{(k+1)}, \boldsymbol{\vartheta}^{(k+1)}) := (I - \mu \nabla \Psi) \circ T_{\mathcal{X}}(\mathbf{u}^{(k)}, \boldsymbol{\vartheta}^{(k)}) \\ &\quad - t^{(k+1)} \nabla F \left((I - \mu \nabla \Psi) \circ T_{\mathcal{X}}(\mathbf{u}^{(k)}, \boldsymbol{\vartheta}^{(k)}) \right), \end{aligned} \tag{23}$$



(a) Original. (b) Degraded. PSNR=21.09 dB SSIM=0.4702 (c) Existing method. PSNR=30.80 dB SSIM=0.8849 (d) Proposed method. PSNR=31.10 dB SSIM=0.8990

Fig. 1. The resulting images: Our estimate (d) is as well-restored as (c) obtained by solving (25) which uses the hand-optimized w.

where ∇F is the gradient of F given by

$$\nabla F : \mathcal{X} \to \mathcal{X} : (\mathbf{u}, \boldsymbol{\vartheta}) \mapsto \left(\sum_{k=1}^{K} \mathbf{M}_k^* \circ (\nabla^\gamma f_k) \circ \mathbf{M}_k(\mathbf{u}), \mathbf{0} \right),$$

which is Lipschitzian, and $(t^{(k)})_{k \geq 1} \subset \mathbb{R}_+$ a slowly decreasing nonnegative sequence satisfying $\lim_{k\to\infty} t^{(k)} = 0$ and $\sum_{k\geq 1} t^{(k)} = \infty$. It is confirmed from (12), (14), and (20) that $T_{\mathcal{X}}$ and $\nabla \Psi$ can be easily computed. Also, ∇F is available as long as the proximity operator of f_k ($k = 1, \ldots, K$) are. Theorem 5 of [15] guarantees that the sequence $(\mathbf{u}^{(k)}, \boldsymbol{\vartheta}^{(k)})_{k\geq 0}$ generated by (23) reaches the solution set of the problem (22), i.e.,

$$\lim_{k\to\infty} \min_{(\mathbf{u}^\star, \boldsymbol{\vartheta}^\star)\in\Omega} \|(\mathbf{u}^{(k)}, \boldsymbol{\vartheta}^{(k)}) - (\mathbf{u}^\star, \boldsymbol{\vartheta}^\star)\|_{\mathcal{X}} = 0. \quad (24)$$

4. NUMERICAL EXAMPLES

We examined the proposed method on Poisson image deblurring. This kind of situation is often appeared in the context of photon-limited applications, for example, astronomical image restoration. In our experiment, an astronomical image 'Saturn' whose size is $n_v \times n_h = 256 \times 256$ ($N = 65,536$) was blurred by 5×5 uniform-blur and then contaminated by a Poisson noise with scaling parameter $\alpha = 0.2$. For the convex priors, we simply used the total variation, i.e., $K = 1$ (see also Remark 3.1). In this case, the proximity operator of f_1 is given by a generalized soft-thresholding (see, e.g., Example III.3 of [5]). The parameter γ of the Moreau envelope was set to 0.001. The level of the G-KL divergence ρ was adjusted to $D_{\mathbf{v},\alpha}(\boldsymbol{\Phi}\hat{\mathbf{u}})$. We compare our results with the results obtained by solving the unconstrained problem [1, 2, 3, 4], i.e., the minimization of the weighted sum of the total variation and the G-KL divergence:

$$\text{find } \mathbf{u}^\star \in \arg\min_{\mathbf{u}\in C_{255}} \{ w f_1(\mathbf{M}_1 \mathbf{u}) + D_{\mathbf{v},\alpha}(\boldsymbol{\Phi}\mathbf{u}) \}, \quad (25)$$

where $w \in (0, \infty)$. Note that, even in the experimental setting where the original image $\hat{\mathbf{u}}$ is known, a suitable value of the weight cannot be determined easily. The parameters (μ, $t^{(k)}$) are set to (0.2, $k^{-\frac{2}{3}}$). The unconstrained problem is solved by the *primal-dual splitting algorithm* [10][7].

The resulting images are shown in Fig. 1. We see that our estimate (d) is close (even a little better in the sense of PSNR and SSIM [20]) to the estimate (c) obtained by solving the unconstrained problem that uses a hand-optimized weight. This implies that the proposed likelihood constrained optimization is more favorable in parameter selection than the unconstrained one; the Moreau envelope of the total variation used in the proposed method is as effective as the original nonsmooth one.

[7] Basically, what algorithm we use for solving (25) does not significantly affect the recovery performance.

5. CONCLUDING REMARKS

We have presented a likelihood constrained optimization framework for Poisson image restoration. The level set of the G-KL divergence is characterized as the fixed point set of an attracting quasi-nonexpansive operator inspired by the subgradient projection, and reformulated the constrained problem in a product space to circumvent the difficulty of the computation caused by matrix composition. The reformulated constrained problem is efficiently solved by the hybrid steepest descent method. Numerical examples have shown that the proposed method is effective for the deblurring of images contaminated by Poisson noise.

We briefly discuss how our main contributions are related to prior work. The proposed Poisson image restoration based on the constrained problem (7) is an advanced method in the sense of the parameter setting compared with unconstrained approaches [1, 2, 3, 4, 5]. Use of the likelihood constraint in Poisson noise contamination scenario was first considered by Chierchia *et al.* [11], where they employed an outer approximation of the constraint via certain multiple hyper planes. Carlavan *et al.* [12] and Teuber *et al.* [13] also adopted the likelihood constraint where the projection onto it is substituted by the proximity operator of the G-KL divergence with appropriate γ obtained by solving a certain discrepancy equation. Such γ varies based on the input, so that their algorithms have to calculate γ using Newton's method in each iteration, resulting in inner loop. Compared with them, our proposed optimization scheme can solve the likelihood constrained problem (7) without using any approximation related to the constraint, as well as it does not require computationally-expensive procedures such as operator inversion and inner loop. We should also mention the generalized Haugazeau's algorithm [21], which was developed for minimizing strictly convex functions over the fixed point set of a certain quasi-nonexpansive operator and applied to a constrained total-variation-based image restoration [22]. This algorithm can also be applied to the likelihood constrained problem if the objective function is strictly convex. On the other hand, out proposed optimization scheme admits non-strict convex objective function, which implies its wider applicability than the generalized Haugazeu's algorithm. In addition, the proposed formulation (7) allows us to adopt a variety of prior design with Moreau envelope relaxation, such as multiple priors [23, 5] and more elaborated priors [24, 25, 26, 27].

We finally remark that the proposed method can handle the level set of any function whose subgradient is available. This is because the key principle of the proposed method in Proposition 3.1 is not based on any specific property of the G-KL divergence. Hence, there are many other possible applications, for example, restoration from Poisson-Gaussian mixture [28] and multiplicative noise contamination [29, 30] based on suitable likelihood constraints.

6. REFERENCES

[1] P. L. Combettes and J. C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Topics in Signal Process.*, vol. 1, pp. 564–574, 2007.

[2] F. X. Dupé, J. M. Fadili, and J. L. Starck, "A proximal iteration for deconvolving Poisson noisy images using sparse representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310–321, 2009.

[3] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Vis. Commun. Image Repres.*, vol. 21, pp. 193–199, 2010.

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

[5] N. Pustelnik, C. Chaux, and J. C. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, 2011.

[6] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[7] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite elements approximations," *Comput. Math. Appl.*, vol. 2, pp. 17–40, 1976.

[8] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2010.

[9] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[10] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optimization Theory and Applications*, 2012, DOI 10.1007/s10957-012-0245-9.

[11] G. Chierchia, N. Pustelnik, J. C. Pesquet, and B. Pesquet-Popescu, "A proximal approach for constrained cosparse modelling," in *Proc. IEEE ICASSP*, 2012, pp. 3433–3436.

[12] M. Carlavan and L. Blanc-Féraud, "Sparse Poisson noisy image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1834–1846, 2012.

[13] T. Teuber, G. Steidl, and R. H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Inverse Problems*, vol. 29, no. 3, 2013.

[14] I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," in *Inherently Parralel Algorithm for Feasibility and Opt. and Their App.*, D. Butnariu *et al*, Ed., pp. 473–504. Elsevier, 2001.

[15] I. Yamada and N. Ogura, "Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings," *Numer. Funct. Anal. Optim.*, vol. 25, pp. 619–655, 2004.

[16] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings," in *Fixed-Point Algorithm for Inverse Problems in Science and Engineering*, H. H. Bauschke *et al*, Ed., pp. 345–390. Springer-Verlag, 2011.

[17] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Ser. A Math.*, vol. 255, pp. 2897–2899, 1962.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[19] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, 2011.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[21] P. L. Combettes, "Strong convergence of block-iterative outer approximation methods for convex optimization," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 538–564, 2000.

[22] P. L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1213–1222, 2004.

[23] S. Ono, T. Miyata, and K. Yamaoka, "Total variation-wavelet-curvelet regularized optimization for image restoration," in *Proc. of IEEE ICIP*, 2011, pp. 2721–2724.

[24] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. and Simul.*, vol. 7, no. 3, pp. 1005–1028, 2008.

[25] S. Ono, T. Miyata, I. Yamada, and K. Yamaoka, "Missing region recovery by promoting blockwise low-rankness," in *Proc. of IEEE ICASSP*, 2012, pp. 1281–1284.

[26] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, 2012.

[27] S. Ono and I. Yamada, "A convex regularizer for reducing color artifact in color image recovery," in *Proc. of CVPR*, 2013, (accepted).

[28] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, 2011.

[29] Y. Huang, M. Ng, and Y. Yen, "A new total variation method for multiplicative noise removal," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 20–40, 2009.

[30] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1720–1730, 2010.