Fast Bayesian Estimation Using Location-type Variational Representations

Jan Dorazil*, Franz Hlawatsch†, Bernard H. Fleury†, and Radim Burget*

*Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
†Institute of Telecommunications, TU Wien, Vienna, Austria

***Abstract*—A class of iterative Bayesian estimation methods known as Type I methods use a variational representation of the posterior distribution. However, Type I methods can exhibit slow convergence for a family of variational representations referred to as convex/location representation. We analyze the convergence behavior of Type I methods by interpreting them as an iterative maximization of a dual objective function involving a linearization. We then propose a modified method that avoids the linearization and allows a coordinatewise maximization. We demonstrate the advantages of the proposed method in the context of image restoration under Poisson noise.**

***Index Terms*—Bayesian estimation, Type I estimation, half-quadratic minimization, variational representation, convex representation, location parameterization, image restoration.**

I. Introduction

Variational representations provide a powerful framework for Bayesian estimation in a wide class of statistical models [1], [2]. In this framework, certain factors of the posterior distribution are expressed by a variational representation involving a variational parameter. This allows the use of computationally efficient iterative methods that are known as Type I methods [1], [2] (or half-quadratic minimization or gradient linearization [3], [4]). These methods are useful in many applications including sparse signal reconstruction [5], image restoration [6], image deconvolution [7], and optical flow estimation [8].

A specific variational representation, to be referred to as *convex/location representation*, expresses the factors of the posterior distribution as the supremum of a Gaussian probability density function (pdf) weighted by a nonnegative function, with the mean (location) of the Gaussian pdf given by the variational parameter. As shown in [3], the scale η of the Gaussian pdf has a direct effect on the convergence speed of the Type I method. In particular, for a large η, the convergence can be prohibitively slow.

In this paper, we analyze the convergence of Type I methods and present a modified method with significantly faster convergence for large η. We show that Type I methods can be viewed as an iterative maximization of a dual objective function, where a linearization is used in each iteration. We argue that the linearization is responsible for slow convergence in models where η is large. We then propose a new coordinatewise method for maximizing the dual objective function without a linearization. For many practically useful models, each coordinate update amounts to finding the roots of a polynomial of small degree, which results in a low computational complexity. We apply our method to the restoration of images corrupted by Poisson noise and demonstrate that it converges faster than the Type I method, especially when η is large.

The rest of this paper is organized as follows. The probabilistic model and the convex/location variational representation are described in Sections II and III, respectively. In Section IV, we review Type I estimation and present a new interpretation of the Type I method. In Section V, we propose a new method with faster convergence. Finally, simulation results are presented in Section VI.

II. Probabilistic Model

Bayesian estimation of an unknown random vector $\boldsymbol{x}\in\mathbb{R}^m$ from an observed (continuous or discrete) random vector $\boldsymbol{y}$ relies on the posterior pdf $p(\boldsymbol{x}|\boldsymbol{y})$ [9, Sec. 2.4.2]. Following [10], we assume that $p(\boldsymbol{x}|\boldsymbol{y})$ admits a factorization of the form

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{Z(\boldsymbol{y})}\prod_{i=1}^{N}\Psi_i\big(\boldsymbol{a}_i(\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x} - b_i(\boldsymbol{y});\boldsymbol{y}\big), \qquad (1)$$

with partition function $Z(\boldsymbol{y})<\infty$ (which ensures that $p(\boldsymbol{x}|\boldsymbol{y})$ is a proper pdf), "potential functions" (PFs) $\Psi_i(\cdot;\boldsymbol{y})\colon\mathbb{R}\to\mathbb{R}_{>0}$, and $\boldsymbol{a}_i(\boldsymbol{y})\in\mathbb{R}^m$ and $b_i(\boldsymbol{y})\in\mathbb{R}$, for $i=1,\ldots,N$. As we will see in Section III, the desired convex/location representation is obtained by considering PFs of the form

$$\Psi_i(\xi;\boldsymbol{y}) \triangleq \exp\big(-\eta_i\big(\xi^2 - 2f_i(\xi;\boldsymbol{y})\big)\big), \quad \xi\in\mathbb{R}, \qquad (2)$$

with scale parameters $\eta_i>0$ and functions $f_i(\cdot;\boldsymbol{y})\colon\mathbb{R}\to\mathbb{R}$ that are all strictly convex and differentiable. For later use, we define $\boldsymbol{A}(\boldsymbol{y})\triangleq\big(\boldsymbol{a}_1(\boldsymbol{y})\cdots\boldsymbol{a}_N(\boldsymbol{y})\big)^{\mathrm{T}}\in\mathbb{R}^{N\times m}$ and $\boldsymbol{b}(\boldsymbol{y})\triangleq(b_1(\boldsymbol{y})\cdots b_N(\boldsymbol{y}))^{\mathrm{T}}\in\mathbb{R}^N$. The dependence of Ψ_i, f_i, $\boldsymbol{A}$, $\boldsymbol{a}_i$, $\boldsymbol{b}$, and b_i on the observation $\boldsymbol{y}$ will no longer be indicated in what follows.

The probabilistic model (1) requires that $Z(\boldsymbol{y})<\infty$ for all $\boldsymbol{y}$. In [2], we show the following condition: *For the pdf* (1), $Z(\boldsymbol{y})<\infty$ *if* $\boldsymbol{A}$ *has full column rank and* $\int_{\mathbb{R}}\Psi_i(\xi)\,\mathrm{d}\xi<\infty$ *for* $i=1,\ldots,N$.

This work was supported in part by the Czech Ministry of Education, Youth and Sports under grant OP JAK CZ.02.01.01/00/23_021/0008829, by the Technology Agency of the Czech Republic under grant CK04000027, and by the Austrian Science Fund (FWF) under grant PAT1538524.

III. Convex/Location Representation

We will next express the PFs Ψ_i and the posterior pdf $p(\boldsymbol{x}|\boldsymbol{y})$ in a way that involves a variational parameter. We first review some basic facts of convex analysis. The *convex conjugate* $f^\star\colon \mathcal{F}^\star \to \mathbb{R}$ of a convex function $f\colon \mathbb{R} \to \mathbb{R}$ is defined as [11, Sec. 3.3]

$$f^\star(\lambda) \triangleq \sup_{\xi\in\mathbb{R}} \big(\lambda\xi - f(\xi)\big),$$

for all $\lambda \in \mathcal{F}^\star \triangleq \{\lambda\in\mathbb{R}\colon \sup_{\xi\in\mathbb{R}} \big(\xi\lambda - f(\xi)\big) < \infty\}$. The convex conjugate $f^\star$ is a convex function. If f is strictly convex, then $f^\star$ is differentiable [12, Th. 26.3]. If, in addition, f is differentiable, then the derivatives of f and $f^\star$ are related as follows [12, Cor. 23.5.1]: for any $\xi \in \mathbb{R}$ and $\lambda \in \mathcal{F}^\star$,

$$f^{\star\prime}(\lambda) = \xi \text{ if and only if } f'(\xi) = \lambda. \tag{3}$$

Let us now consider a PF of the form (2), i.e., $\Psi_i(\xi) = \exp\big(-\eta_i\big(\xi^2 - 2f_i(\xi)\big)\big)$ for $\xi \in \mathbb{R}$, with $\eta_i > 0$ and a strictly convex and differentiable function $f_i\colon \mathbb{R} \to \mathbb{R}$. As we show in [2], $\Psi_i(\xi)$ can be expressed as

$$\Psi_i(\xi) = \sup_{\lambda\in\mathcal{F}_i^\star} G_i(\xi,\lambda)\,\varphi_i(\lambda), \tag{4}$$

where $\mathcal{F}_i^\star \triangleq \{\lambda\in\mathbb{R}\colon \sup_{\xi\in\mathbb{R}} \big(\xi\lambda - f_i(\xi)\big) < \infty\}$ and

$$G_i(\xi,\lambda) \triangleq \exp\big(-\eta_i(\xi-\lambda)^2\big), \tag{5}$$

$$\varphi_i(\lambda) \triangleq \exp\big(\eta_i\big(\lambda^2 - 2f_i^\star(\lambda)\big)\big). \tag{6}$$

According to (5), $G_i(\cdot,\lambda)$ is a Gaussian pdf up to normalization for any $\lambda \in \mathcal{F}_i^\star$; furthermore, the "variational parameter" λ affects the location of $G_i(\cdot,\lambda)$. We will therefore call (4)–(6) the *convex/location representation* of $\Psi_i(\xi)$.

Next, we consider the posterior pdf in (1) and express all PFs $\Psi_i(\xi)$ by their convex/location representation. Inserting (4) into (1) then yields a *convex/location representation of the posterior pdf* $p(\boldsymbol{x}|\boldsymbol{y})$ according to

$$p(\boldsymbol{x}|\boldsymbol{y}) = \sup_{\boldsymbol{\lambda}\in\mathcal{F}^\star} h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y}), \tag{7}$$

where $\boldsymbol{\lambda} \triangleq (\lambda_1 \cdots \lambda_N)^{\mathrm{T}}$, $\mathcal{F}^\star \triangleq \mathcal{F}_1^\star \times \cdots \times \mathcal{F}_N^\star$, and

$$h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y}) \triangleq \frac{1}{Z(\boldsymbol{y})} \prod_{i=1}^{N} G_i(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i, \lambda_i)\,\varphi_i(\lambda_i), \tag{8}$$

with all G_i and φ_i given by (5) and (6), respectively. We will show in the next two sections that the convex/location representation allows for efficient estimation methods.

IV. Type I Estimation

A. Review of the Type I Estimation Method

An important Bayesian estimator is the maximum a posteriori (MAP) estimator

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \underset{\boldsymbol{x}\in\mathbb{R}^m}{\operatorname{argmax}}\, \log p(\boldsymbol{x}|\boldsymbol{y}). \tag{9}$$

Type I methods offer an efficient approach to MAP estimation that leverages the convex/location variational representation of $p(\boldsymbol{x}|\boldsymbol{y})$ in (7) [1]. Indeed, using (7) and the fact that $\log$ is a strictly increasing function, we can rewrite (9) as

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \underset{\boldsymbol{x}\in\mathbb{R}^m}{\operatorname{argmax}}\, \sup_{\boldsymbol{\lambda}\in\mathcal{F}^\star} \log h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y}). \tag{10}$$

This joint maximization is now performed by repeatedly maximizing $\log h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y})$ alternately with respect to $\boldsymbol{x}$ and $\boldsymbol{\lambda}$. The l-th iteration of the Type I method thus reads

$$\hat{\boldsymbol{x}}^{(l+1)} = \underset{\boldsymbol{x}\in\mathbb{R}^m}{\operatorname{argmax}}\, \log h(\boldsymbol{x},\hat{\boldsymbol{\lambda}}^{(l)};\boldsymbol{y}), \tag{11}$$
$$\hat{\boldsymbol{\lambda}}^{(l+1)} = \underset{\boldsymbol{\lambda}\in\mathcal{F}^\star}{\operatorname{argmax}}\, \log h(\hat{\boldsymbol{x}}^{(l+1)},\boldsymbol{\lambda};\boldsymbol{y}),$$

with $h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y})$ given by (8). As we show in [2], these maximization problems have the following closed-form solutions:

$$\hat{\boldsymbol{x}}^{(l+1)} = \boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{b} + \hat{\boldsymbol{\lambda}}^{(l)}), \tag{12}$$

$$\hat{\lambda}_i^{(l+1)} = f_i'(\boldsymbol{a}_i^{\mathrm{T}}\hat{\boldsymbol{x}}^{(l+1)} - b_i), \quad i = 1,\dots,N, \tag{13}$$

where

$$\boldsymbol{J} \triangleq \boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{A}, \quad \boldsymbol{E} \triangleq \operatorname{diag}\{\eta_1,\dots,\eta_N\}. \tag{14}$$

Note that $\boldsymbol{J}$ is positive-definite since $\boldsymbol{A}$ has full column rank (cf. our condition for $Z(\boldsymbol{y}) < \infty$ in Section II) and $\eta_i > 0$, $i = 1,\dots,N$. The sequence $\big(\hat{\boldsymbol{x}}^{(l)}\big)_{l=1}^{\infty}$ converges to a local maximum or saddle point of $p(\boldsymbol{x}|\boldsymbol{y})$ for arbitrary initialization $\hat{\boldsymbol{\lambda}}^{(1)}$ [13], [14].

B. An Alternative Method

Next, we consider an alternative approximate MAP estimation method, which will be seen in Section IV-C to lead to a new interpretation of the Type I method. The idea is to exchange in (10) the order of the two maximization steps. That is, the method first calculates an estimate of $\boldsymbol{\lambda}$ according to

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}\in\mathcal{F}^\star}{\operatorname{argmax}}\, \sup_{\boldsymbol{x}\in\mathbb{R}^m} \log h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y}), \tag{15}$$

and then obtains an estimate of $\boldsymbol{x}$ by solving $\hat{\boldsymbol{x}} = \operatorname{argmax}_{\boldsymbol{x}\in\mathbb{R}^m} \log h(\boldsymbol{x},\hat{\boldsymbol{\lambda}};\boldsymbol{y})$ (cf. (11)), which results in $\hat{\boldsymbol{x}} = \boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{b} + \hat{\boldsymbol{\lambda}})$ (cf. (12)). We can write (15) as

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}\in\mathcal{F}^\star}{\operatorname{argmax}}\, g(\boldsymbol{\lambda}), \tag{16}$$

with the "dual" objective function [11, Ch. 5]

$$g(\boldsymbol{\lambda}) \triangleq \sup_{\boldsymbol{x}\in\mathbb{R}^m} \log h(\boldsymbol{x},\boldsymbol{\lambda};\boldsymbol{y}). \tag{17}$$

In the Appendix, we derive the closed-form expression

$$g(\boldsymbol{\lambda}) \stackrel{\mathrm{c}}{=} \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda} + 2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{P} - \boldsymbol{E})\boldsymbol{b} - 2\sum_{i=1}^{N}\eta_i f_i^\star(\lambda_i), \tag{18}$$

where $\stackrel{\mathrm{c}}{=}$ denotes equality up to an additive constant and

$$\boldsymbol{P} \triangleq \boldsymbol{E}\boldsymbol{A}\boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}. \tag{19}$$

An efficient iterative method for solving the maximization problem (16) can be obtained as follows. Let $\hat{\boldsymbol{\lambda}}^{(l+1)}$ denote the iterate of $\boldsymbol{\lambda}$ calculated at the l-th iteration. We first linearize the quadratic term $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ occurring in (18) around $\hat{\boldsymbol{\lambda}}^{(l)}$, i.e.,

$$\begin{aligned}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda} &\approx \hat{\boldsymbol{\lambda}}^{(l)\mathrm{T}}\boldsymbol{P}\hat{\boldsymbol{\lambda}}^{(l)} + 2\hat{\boldsymbol{\lambda}}^{(l)\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}}^{(l)}) \\ &\stackrel{\mathrm{c}}{=} 2\hat{\boldsymbol{\lambda}}^{(l)\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}.\end{aligned} \tag{20}$$

Inserting this approximation into (18) yields the following approximation of the dual objective function $g(\boldsymbol{\lambda})$:

$$\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)}) \stackrel{\mathrm{c}}{=} 2\hat{\boldsymbol{\lambda}}^{(l)\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda} + 2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{E})\boldsymbol{b} - 2\sum_{i=1}^{N}\eta_i f_i^{\star}(\lambda_i). \tag{21}$$

Substituting $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)})$ for $g(\boldsymbol{\lambda})$ in (16), we obtain the new iterate of $\boldsymbol{\lambda}$ as

$$\hat{\boldsymbol{\lambda}}^{(l+1)} = \underset{\boldsymbol{\lambda}\in\mathcal{F}^{\star}}{\operatorname{argmax}}\, \tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)}). \tag{22}$$

This maximization problem is simpler than (16). Indeed, $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)})$ is concave in $\boldsymbol{\lambda}$, whereas $g(\boldsymbol{\lambda})$ contains both concave and convex terms. Thus, we can find the maximum of $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)})$ by setting the gradient of $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)})$ to zero. Writing $\boldsymbol{P}=\boldsymbol{E}\boldsymbol{Q}$ (with $\boldsymbol{Q}\triangleq\boldsymbol{A}\boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}$, cf. (19)) and recalling that $\boldsymbol{E}=\mathrm{diag}\{\eta_1,\ldots,\eta_N\}$, Eq. (21) can be rewritten as

$$\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)}) \stackrel{\mathrm{c}}{=} 2\sum_{i=1}^{N}\eta_i\,\tilde{g}_i(\lambda_i;\hat{\boldsymbol{\lambda}}^{(l)}), \tag{23}$$

with

$$\tilde{g}_i(\lambda_i;\hat{\boldsymbol{\lambda}}^{(l)}) \triangleq \hat{\boldsymbol{\lambda}}^{(l)\mathrm{T}}\boldsymbol{q}_i\lambda_i + \lambda_i(\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{b}-b_i) - f_i^{\star}(\lambda_i), \tag{24}$$

where $\boldsymbol{q}_i^{\mathrm{T}}$ denotes the i-th row of $\boldsymbol{Q}$. According to (23) and (24), setting the gradient of $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}^{(l)})$ to zero is equivalent to the N scalar equations $\frac{\mathrm{d}}{\mathrm{d}\lambda_i}\tilde{g}_i(\lambda_i;\hat{\boldsymbol{\lambda}}^{(l)}) = 0$ and, in turn, $\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{b}+\hat{\boldsymbol{\lambda}}^{(l)}) - b_i - f_i^{\star\prime}(\lambda_i) = 0$, for $i=1,\ldots,N$. Thus, the solutions $\hat{\lambda}_i^{(l+1)}$ satisfy $f_i^{\star\prime}(\hat{\lambda}_i^{(l+1)}) = \boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{b}+\hat{\boldsymbol{\lambda}}^{(l)}) - b_i$. Invoking (3), we finally obtain the closed-form expressions

$$\hat{\lambda}_i^{(l+1)} = f_i'\big(\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{b}+\hat{\boldsymbol{\lambda}}^{(l)}) - b_i\big), \quad i=1,\ldots,N. \tag{25}$$

We conclude that, as a result of our linearization of $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ in (20), the maximization problem (22) admits an efficient coordinatewise solution. As shown in [15], the sequence $\big(\hat{\boldsymbol{\lambda}}^{(l)}\big)_{l=1}^{\infty}$ obtained by (22) converges to a local maximum or saddle point of $g(\boldsymbol{\lambda})$, at thus it attempts to solve the original problem (16), (17).

C. A New Interpretation of the Type I Method

The i-th row of $\boldsymbol{Q}=\boldsymbol{A}\boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}$ is $\boldsymbol{q}_i^{\mathrm{T}} = \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}$. Inserting this expression into (25) yields an expression of $\hat{\lambda}_i^{(l)}$ that is equivalent to the expression obtained by inserting (12) into (13). This equivalence provides the following new interpretation of the Type I method: each iteration maximizes a concave approximation $\tilde{g}$ of the dual objective function g that is obtained by a linearization of the quadratic term $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ occurring in g.

Next, we present an analysis of the convergence speed of the Type I method that is based on this interpretation. Because $\tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}})$ was obtained from $g(\boldsymbol{\lambda})$ by linearizing $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ in (18), the approximation error $\varepsilon(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}) \triangleq g(\boldsymbol{\lambda}) - \tilde{g}(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}})$ is equal to the difference between $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ and its linearized version on the right-hand side of (20), i.e.,

$$\varepsilon(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}) = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda} - \big(\hat{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}\hat{\boldsymbol{\lambda}} + 2\hat{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})\big), \tag{26}$$

where we replaced $\hat{\boldsymbol{\lambda}}^{(l)}$ by $\hat{\boldsymbol{\lambda}}$ to simplify the notation. Since $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\lambda}$ is a quadratic function, it equals its second-order Taylor series expansion, $\hat{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}\hat{\boldsymbol{\lambda}} + 2\hat{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}}) + (\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\mathrm{T}}\boldsymbol{P} \times (\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})$. Inserting this expression into (26) yields $\varepsilon(\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}) = (\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\mathrm{T}}\boldsymbol{P}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})$, or, with a slight abuse of notation,

$$\varepsilon(\tilde{\boldsymbol{\lambda}}) = \tilde{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}\tilde{\boldsymbol{\lambda}}, \tag{27}$$

where $\tilde{\boldsymbol{\lambda}} \triangleq \boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}}$. For fixed $\tilde{\boldsymbol{\lambda}}$, this approximation error grows with the scale parameters η_i. To see this, let us scale all η_i, or equivalently the matrix $\boldsymbol{E}=\mathrm{diag}\{\eta_1,\ldots,\eta_N\}$, by a factor $\sigma>0$. Substituting $\sigma\boldsymbol{E}$ for $\boldsymbol{E}$ in the expression for $\boldsymbol{P}$ in (19) while recalling (14) yields $\boldsymbol{P}_\sigma \triangleq \sigma\boldsymbol{E}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\sigma\boldsymbol{E}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\sigma\boldsymbol{E} = \sigma\boldsymbol{P}$, and hence the approximation error in (27) is changed to $\varepsilon_\sigma(\tilde{\boldsymbol{\lambda}}) \triangleq \tilde{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}_\sigma\tilde{\boldsymbol{\lambda}} = \sigma\tilde{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{P}\tilde{\boldsymbol{\lambda}} = \sigma\epsilon(\tilde{\boldsymbol{\lambda}})$. Thus, larger values of η_i can be expected to result in a larger approximation error. We conjecture that the larger approximation error, in turn, results in a slower convergence of the Type I method. This conjecture is supported by the theoretical convergence rates of [3] as well as by our experimental results in Section VI, both of which demonstrate that larger values of η_i lead to a slower convergence of the Type I method.

V. The Proposed Method

The above analysis suggests a relation between the linearization underlying the Type I method and the method's slow convergence for large η_i. We now propose an iterative method that avoids the linearization but is still able to solve the maximization in (16) coordinatewise. In the l-th iteration, all λ_i are sequentially updated according to

$$\hat{\lambda}_i^{(l+1)} = \underset{\lambda_i\in\mathcal{F}_i^{\star}}{\operatorname{argmax}}\, g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big), \quad i=1,\ldots,N, \tag{28}$$

where $g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big)$ is the dual objective function $g(\boldsymbol{\lambda})$ in (16) restricted such that $\lambda_j=\hat{\lambda}_j$ for all $j\neq i$; here, $\hat{\lambda}_j$ is the last available iterate of λ_j (calculated at iteration $l-1$ or l).

Let p_{ij} denote the (i,j)-th entry of $\boldsymbol{P}$ and θ_i the i-th entry of $\boldsymbol{\theta}\triangleq(\boldsymbol{P}-\boldsymbol{E})\boldsymbol{b}$. From (18), we obtain

$$g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big) \stackrel{\mathrm{c}}{=} p_{ii}\lambda_i^2 + 2\alpha_i\lambda_i - 2\eta_i f_i^{\star}(\lambda_i),$$

with $\alpha_i \triangleq \theta_i + \sum_{j\neq i}\hat{\lambda}_j p_{ji}$. To solve (28), we set $\frac{\mathrm{d}}{\mathrm{d}\lambda_i}g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big)$ to zero, which gives $f_i^{\star\prime}(\lambda_i) = \frac{1}{\eta_i}(p_{ii}\lambda_i + \alpha_i)$. By (3), this is equivalent to

$$f_i'\bigg(\frac{p_{ii}\lambda_i+\alpha_i}{\eta_i}\bigg) = \lambda_i. \tag{29}$$

If f_i' is a polynomial of degree at most three, this equation can be solved in closed form; otherwise, efficient numerical methods are available. A solution $\hat{\lambda}_i$ of (29) is a (possibly local) maximizer of $g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big)$ if and only if

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda_i^2}\,g\big(\lambda_i;\{\hat{\lambda}_j\}_{j\neq i}\big)\bigg|_{\hat{\lambda}_i} = 2\big(p_{ii} - \eta_i f_i^{\star\prime\prime}(\hat{\lambda}_i)\big) < 0,$$

where we assumed that $f_i^\star$ is twice differentiable. Here, $f_i^{\star\prime\prime}$ can be evaluated as follows. Differentiating the left-hand and right-hand parts of (3) and combining the resulting equations yields $1 = f_i''(\hat{\xi}_i) f_i^{\star\prime\prime}(\hat{\lambda}_i)$ or equivalently $f_i^{\star\prime\prime}(\hat{\lambda}_i) = \frac{1}{f_i''(\hat{\xi}_i)}$, for any $\hat{\xi}_i$ satisfying $\hat{\lambda}_i = f_i'(\hat{\xi}_i)$. Thus, evaluation of $f_i^{\star\prime\prime}(\hat{\lambda}_i)$ amounts to solving the equation $f_i'(\hat{\xi}_i) = \hat{\lambda}_i$ for $\hat{\xi}_i$. Note that there is a unique solution since f_i' is strictly increasing.

VI. Simulation Study

We compare the convergence speed of the Type I method and the proposed method for synthetic data whose generation is motivated by the problem of restoring images from a (potentially incomplete) set of photon-limited measurements [16].

A. Simulation Setup

Let $\boldsymbol{x} = (x_1 \cdots x_m)^{\mathrm{T}}$ with $x_k > 0$ represent (via columnwise or rowwise stacking) the true two-dimensional image, and let $\boldsymbol{y} = (y_1 \cdots y_m)^{\mathrm{T}}$ with $y_k \in \{0, 1, 2, \dots\}$ represent a measured discrete-valued image. Each entry y_k of $\boldsymbol{y}$ is Poisson distributed [16] with intensity x_k, i.e.,

$$p(y_k|x_k) = \text{Poisson}(y_k; x_k). \quad (30)$$

Furthermore, y_k is conditionally independent, given x_k, of y_l and x_l for all $l \neq k$. The prior pdf is a truncated version of the Gaussian pdf

$$p_{\mathrm{G}}(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\left(\sum_{j=1}^{2m} (\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{x})^2 + (\mathbb{1}^{\mathrm{T}}\boldsymbol{x} - m\gamma)^2\right)\right), \quad (31)$$

where $\boldsymbol{g}_j$ is defined such that $\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{x}$ yields the vertical or horizontal (depending on j) difference between adjacent pixels of the image represented by $\boldsymbol{x}$, $\mathbb{1}$ is the all-ones vector of length m, and $\gamma > 0$ is a parameter. This prior promotes smoothness of the image represented by $\boldsymbol{x}$ (via $\sum_{j=1}^{2m} (\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{x})^2$) and closeness of $\sum_{k=1}^{m} x_m$ to $m\gamma$ (via $(\mathbb{1}^{\mathrm{T}}\boldsymbol{x} - m\gamma)^2$). We can rewrite (31) as

$$p_{\mathrm{G}}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; m\gamma\boldsymbol{\Sigma}\mathbb{1}, \boldsymbol{\Sigma}), \text{ with } \boldsymbol{\Sigma} \triangleq (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} + \mathbb{1}\mathbb{1}^{\mathrm{T}})^{-1}, \quad (32)$$

where $\boldsymbol{G}$ is the matrix with rows $\boldsymbol{g}_j^{\mathrm{T}}$. Note that $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} + \mathbb{1}\mathbb{1}^{\mathrm{T}}$ is nonsingular because $\mathbb{1}$ spans the nullspace of $\boldsymbol{G}$. When sampling from $p_{\mathrm{G}}(\boldsymbol{x})$, we reject nonpositive samples x_k in order to satisfy the condition $x_k > 0$; this is equivalent to replacing (32) by a truncated Gaussian pdf.

While we use this model to generate the data, we cannot use it in the estimation methods because the Poisson likelihood function $p(y_k|x_k)$ does not admit a convex/location representation. Instead, following [6], we use the Anscombe likelihood function with a quadratic extension, which provides a good approximation of the Poisson likelihood function, especially for large intensity x_k. Thus, $p(y_k|x_k)$ is defined by

$$-\log p(y_k|x_k) \stackrel{c}{=} \begin{cases} \beta(x_k, y_k), & x_k \geq 0, \\ \beta(0, y_k) + \beta'(0, y_k)x_k + \frac{1}{2}\beta''(0, y_k)x_k^2, & x_k < 0, \end{cases} \quad (33)$$

with $\beta(x, y) \triangleq 2\left(\sqrt{y + \frac{3}{8}} - \sqrt{x + \frac{3}{8}}\right)^2$. Furthermore, we use the improper prior

$$\tilde{p}(\boldsymbol{x}) = \prod_{j=1}^{2m} \tilde{p}_j(\boldsymbol{x}), \quad \text{with } \tilde{p}_j(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{g}_j^{\mathrm{T}}\boldsymbol{x})^2\right). \quad (34)$$

In contrast to (32), this prior promotes only smoothness, while the sum of the x_k does not enter. This is a common practice followed in many image processing applications [3], [14].

The posterior pdf $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\tilde{p}(\boldsymbol{x})$ induced by (33) and (34) can be written in the form (1) by equating the factors $\Psi_i(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i)$ for $i = 1, \dots, m$ with $p(y_i|x_i)$ in (33) and for $i = m+1, \dots, 3m$ with $\tilde{p}_{i-m}(\boldsymbol{x})$ in (34). This is achieved by suitably defining $\boldsymbol{a}_i$, b_i, and Ψ_i. The PFs Ψ_i corresponding to $p(y_i|x_i)$ and $\tilde{p}_j(\boldsymbol{x})$ admit the convex/location representation (4)–(6) with $\eta_i > \frac{\sqrt{y_i + 3/8}}{2(3/8)^{3/2}}$ and $\eta_j > 1$, respectively [2].

We implemented the two methods in C on an Intel(R) Core(TM) i5-7500 CPU with clock rate 3.40 GHz. For numerical and linear algebra operations, we used the GNU Scientific Library (www.gnu.org/software/gsl) and the OpenBLAS library (www.openmathlib.org/OpenBLAS). The inverse $\boldsymbol{J}^{-1}$ was precomputed before running the iterations; thus, one iteration of the Type I method and the proposed method involves, respectively, one matrix-vector product and m vector-vector products. We note that when $\boldsymbol{J}$ has a special structure, precomputing $\boldsymbol{J}^{-1}$ may be disadvantageous because it may be possible to efficiently compute the updates of the Type I and/or the proposed method by other means.

B. Simulation Results

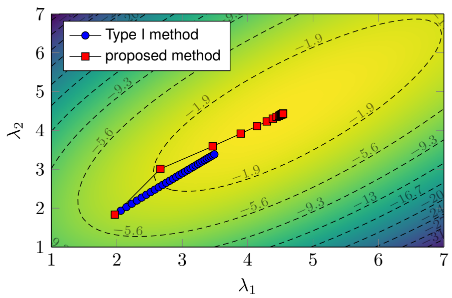

Fig. 1: Iterated estimates of $\boldsymbol{\lambda}$ for 30 iterations of the Type I method (blue bullets) and the proposed method (red squares). The contour plot shows the dual objective function $g(\boldsymbol{\lambda})$ to be maximized (see (16)–(18)).

Fig. 1 visualizes the convergence of the two methods for images of size 2×1 (thus, $m = 2$) and $\gamma = 5$. One can see that the proposed method converged in around seven iterations whereas the Type I method did not converge in 30 iterations.

Next, we consider images of size 50×50 (thus, $m = 2500$) and γ in the range $[1, 10]$. According to (32), increasing γ results in a larger prior mean of $\boldsymbol{x}$, which, because of the condition $\eta_i > \frac{\sqrt{y_i + 3/8}}{2(3/8)^{3/2}}$ combined with (30) and the fact that the mean of Poisson$(y_i; x_i)$ is x_i, necessitates larger values of η_i. Fig. 2 shows the runtime and number of iterations, averaged over 20 realizations of $\boldsymbol{y}$, that are required by the two

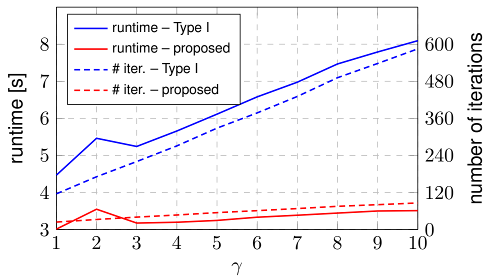

Fig. 2: Mean runtime and number of iterations required by the Type I method and the proposed method to converge, for different values of γ.

methods to converge in the sense that $\frac{\|\hat{\boldsymbol{\lambda}}^{(l+1)}-\hat{\boldsymbol{\lambda}}^{(l)}\|_2}{\|\hat{\boldsymbol{\lambda}}^{(l)}\|_2} < 10^{-5}$. (For a threshold less than 10^{-5}, we observed that the Type I method sometimes does not converge.) One can see that the runtime and number of iterations of the proposed method are significantly smaller than those of the Type I method, and they increase significantly less fast with growing γ. In further experiments with smaller and larger synthetic images, we observed that for $\gamma = 10$ the proposed method always converged at least twice as fast as the Type I method. We finally remark that in our experiments, we always observed the two methods to converge to the same point.

VII. Conclusion

The Type I estimation method for Bayesian models using the convex/location representation converges slowly in the case of large scale parameters. Building on a new interpretation of the Type I method that provides insight into this issue, we proposed a new estimation method that maximizes a dual objective function coordinatewise. For many practical models, this method converges faster and is less complex than the Type I method. These advantages were verified for the problem of image restoration under Poisson noise.

Possible directions for future research include modifications of the proposed method that are able to leverage a special structure of the matrix $\boldsymbol{J}$, such as tridiagonal, block-tridiagonal, Toeplitz, or block-Toeplitz. Our preliminary results indicate good results for tridiagonal $\boldsymbol{J}$. It would also be interesting to study the convergence of the proposed method for other relevant data distributions such as Rayleigh, Rice, or approximations thereof, and to evaluate the method on real data.

Appendix

To derive (18), we first insert (8) into (17), which yields

$$g(\boldsymbol{\lambda}) \overset{c}{=} \sum_{i=1}^{N} \log \varphi_i(\lambda_i) + \sup_{\boldsymbol{x}\in\mathbb{R}^m} \sum_{i=1}^{N} \log G_i(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i, \lambda_i). \quad (35)$$

Using (6), the first term becomes

$$\sum_{i=1}^{N} \log \varphi_i(\lambda_i) = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{\lambda} - 2\sum_{i=1}^{N} \eta_i f_i^{\star}(\lambda_i), \quad (36)$$

and using (5), the second term becomes

$$\begin{aligned} &\sup_{\boldsymbol{x}\in\mathbb{R}^m} \sum_{i=1}^{N} \log G_i(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i, \lambda_i) \\ &= -\inf_{\boldsymbol{x}\in\mathbb{R}^m} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{\lambda}) \\ &= -\inf_{\boldsymbol{x}\in\mathbb{R}^m} (\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{\mu} + (\boldsymbol{b}+\boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{b}+\boldsymbol{\lambda}), \end{aligned}$$

where

$$\boldsymbol{\mu} \triangleq \boldsymbol{J}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{b}+\boldsymbol{\lambda}). \quad (37)$$

Using the fact that $\inf_{\boldsymbol{x}\in\mathbb{R}^m}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}-\boldsymbol{\mu}) = 0$ because $\boldsymbol{J}$ is positive-definite, we obtain further

$$\begin{aligned} &\sup_{\boldsymbol{x}\in\mathbb{R}^m} \sum_{i=1}^{N} \log G_i(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i, \lambda_i) \\ &\quad = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{\mu} - (\boldsymbol{b}+\boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{E}(\boldsymbol{b}+\boldsymbol{\lambda}) \\ &\quad = \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{E})\boldsymbol{\lambda} + 2\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{E})\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{P}-\boldsymbol{E})\boldsymbol{b}, \quad (38) \end{aligned}$$

where the final expression was obtained by inserting (37) and (14) and using (19). Finally, inserting (36) and (38) into (35) yields (18).

References

[1] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. NIPS*, Vancouver, Canada, Dec. 2005, pp. 1559–1566.

[2] J. Dorazil, B. H. Fleury, and F. Hlawatsch, "Variational representation methods for Bayesian estimation within a class of Gibbs-Boltzmann distributions: A unified view," 2025, in preparation.

[3] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, Dec. 2005.

[4] M. Nikolova and R. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, May 2007.

[5] D. P. Wipf and B. D. Rao, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, Sep. 2011.

[6] Y. Marnissi, Y. Zheng, E. Chouzenoux, and J.-C. Pesquet, "A variational Bayesian approach for image restoration—Application to image deblurring with Poisson-Gaussian noise," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 722–737, May 2017.

[7] D. Wipf and H. Zhang, "Revisiting Bayesian blind deconvolution," *J. Mach. Learn. Res.*, vol. 15, pp. 3775–3814, Nov. 2014.

[8] J. Dorazil, B. H. Fleury, and F. Hlawatsch, "Bayesian methods for optical flow estimation using a variational approximation, with applications to ultrasound," in *Proc. IEEE ICASSP*, Rhodes, Greece, Jun. 2023.

[9] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. Hoboken, NJ, USA: Wiley, 1994.

[10] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, Sep. 2004.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[12] T. R. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton University Press, 1997.

[13] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 204–221, Feb. 1998.

[14] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.

[15] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, Apr. 2003.

[16] M. Makitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, Jul. 2010.