J Math Imaging Vis (2014) 49:467–480
DOI 10.1007/s10851-013-0477-9

Characterization of Maximum Likelihood Solutions to Image Reconstruction in Photon Emission Tomography

François Chapeau-Blondeau · Christian Jeanguillaume

Published online: 3 December 2013

Abstract For photon emission tomography, the maximum likelihood (ML) estimator for image reconstruction is generally solution to a nonlinear equation involving the vector of measured data. No explicit closed-form solution is known in general for such a nonlinear ML equation, and numerical resolution is usually implemented, with a very popular iterative method formed by the expectation-maximization algorithm. The numerical character of such resolutions usually makes it difficult to obtain a general characterization of the performance of the ML solution. We show that the nonlinear ML equation can be replaced by an equivalent system of two dual linear equations nonlinearly coupled. This formulation allows us to exhibit explicit (to some extent) forms for the solutions to the ML equation, in general conditions corresponding to the various possible configurations of the imaging system, and to characterize their performance with expressions for the mean-squared error, bias and Cramér-Rao bound. The approach especially applies to characterize the ML solutions obtained numerically, and offers a theoretical framework to contribute to better appreciation of the capabilities of ML reconstruction in photon emission tomography.

Keywords Photon emission tomography · Image reconstruction · Maximum likelihood

F. Chapeau-Blondeau (✉)
Laboratoire d'Ingénierie des Systèmes Automatisés (LISA),
Université d'Angers, 62 avenue Notre Dame du Lac,
49000 Angers, France
e-mail: chapeau@univ-angers.fr

C. Jeanguillaume
LISA & Service de Médecine Nucléaire, Hôpital Larrey,
CHU Angers, Angers, France

1 Introduction

Photon emission tomography is an important imaging modality for noninvasive observation of the internal structure of three-dimensional objects, especially of biological interest [1, 5]. It relies on a radioisotope tracer selectively tagging the internal constituents and emitting high-energy (gamma) photons, either directly as in single photon emission tomography, or indirectly through local annihilation of positrons as in positron emission tomography. The high-energy photons reach the outside of the object, and are collected by an external array of detectors. These measured photon data are then processed to reconstruct an image of the spatial distribution of the tracer inside the object. For such a photon-limited imaging modality, a finite-dimensional statistical modeling is often found appropriate, to describe both the physics of the process and the fluctuations inherent to photon detection. In such a statistical framework, a reference approach to image reconstruction is through the general method of maximum likelihood (ML) [41, 42]. Statistical ML image reconstruction has been shown superior to deterministic linear-filtering based reconstruction methods like the filtered backprojection method for instance (although improvements are still proposed [6]) in that it usually exhibits improved signal-to-noise ratio, improved image resolution, even at low counts of photons [8].

In this context of photon emission tomography, the ML estimator for image reconstruction is generally solution to a nonlinear equation involving the set of measured data. No explicit closed-form solution is known in general for such a nonlinear ML equation. In practice, numerical resolution is usually implemented, with a very popular iterative method formed by the expectation-maximization (EM or EM-ML) algorithm [22, 42, 44]. ML image reconstruction is important both as a theoretical reference and as a practical methodology, and accordingly it has been the subject of numerous

studies and refinements, especially its practical implementation through EM-ML [4, 23]. An efficient method proposed to accelerate the reconstruction is the ordered subset method, which implements the successive EM-ML iterations sequentially on subsets of the data which are evenly distributed around the scene to be imaged [11, 21]. To reduce the noise inherent to ML estimation from a single realization of a statistical process [24], it has been found useful to stop the EM-ML algorithm before convergence, according to various rules, with the effect of biasing the reconstructed image toward its initial estimate [11, 45, 46]. Also, to regularize the reconstructed image, a penalty criterion based on a priori knowledge can be added to the likelihood and interpreted in a Bayesian or maximum a posteriori framework [7, 14, 15, 17, 19, 26, 29, 40].

Depending on the configuration of the imaging system, the nonlinear ML equation can have nonunique solutions, and in this case the outcome of the EM-ML algorithm generally varies with the initial condition [42, 44]. Also, the numerical character of the estimation realized by the EM-ML algorithm or any other numerical resolution, usually makes it difficult to obtain a general characterization of the performance of the ML estimator, for instance in terms of mean-squared error, bias, variance or Cramér-Rao bound. In this report, we show that the nonlinear ML equation for image reconstruction in photon emission tomography, can be replaced by an equivalent system of two linear equations nonlinearly coupled. This formulation allows us to exhibit explicit (to some extent) forms for the solutions to the ML equation, in general conditions corresponding to the various possible configurations of the imaging system. With this approach, we also provide a characterization of the performance of the ML estimator, by means of explicit expressions for its mean-squared error, bias and Cramér-Rao bound. The approach especially applies to characterize the ML solutions numerically obtained by the EM-ML algorithm or by any other numerical resolution. The results offer a theoretical framework to contribute to better appreciation of the capabilities of ML reconstruction in photon emission tomography. In the paper, we develop the theoretical characterization of the solutions to the direct unconstrained ML estimation, which forms an important reference with a well-defined status. We then discuss its connections to the positively-constrained ML problem often considered in practice, and to the practically important EM-ML algorithm and its initialization and convergence conditions.

2 Model for Photon Emission Tomography

We consider a standard model of photon emission tomography as for instance described in [42, 44]. We have N independent sources of photons realized by the N voxels defining the object to be imaged. Each such source is modeled as a Poisson process with intensity θ_i, for $i = 1$ to N. The N parameters θ_i are the unknowns to be estimated for image reconstruction, that we organize in the $N \times 1$ column vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_N]^\top$.

Measurements are collected by an array of M photon detectors, delivering M mutually independent data counts y_j having Poisson distribution of intensity λ_j, associated with the probability

$$\Pr(y_j) = \exp(-\lambda_j)\frac{\lambda_j^{y_j}}{y_j!}, \quad \text{for } j = 1, 2, \ldots, M. \tag{1}$$

The intensities λ_j of the M observable Poisson processes, are related to the intensities θ_i of the N unobservable Poisson processes, through the M affine functions

$$\lambda_j(\boldsymbol{\theta}) = r_j + \sum_{i=1}^{N} p_{ji}\theta_i, \quad \text{for } j = 1, 2, \ldots, M, \tag{2}$$

or under matrix form

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = \boldsymbol{r} + \mathbf{P}\boldsymbol{\theta}, \tag{3}$$

with the $M \times 1$ column vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_M]^\top$. The $M \times 1$ vector $\boldsymbol{r} = [r_1, r_2, \ldots, r_M]^\top$ denotes the (assumed known) intensities of M independent Poisson processes representing the background events on each of the M detectors (background radiation, random coincidences, scatter, etc.) [16, 38, 43]. The $M \times N$ matrix $\mathbf{P}$ in Eq. (3) is formed with nonnegative coefficients $p_{ji} \in [0, 1]$ interpretable as the probability that a photon emitted by source i is measured by detector j. These coefficients p_{ji} are usually fixed by the tomographic imaging system (chiefly its geometry) and attached to it [42, 44]. There is a normalization condition for each source i reading

$$\sum_{j=1}^{M} p_{ji} = s_i, \tag{4}$$

with often $s_i = 1$ for all $i = 1$ to N, as a result of the probabilistic interpretation of the coefficients p_{ji}, but this is not a necessary restriction [33, 42].

3 ML Estimation

For estimation of the unknown source parameters $\boldsymbol{\theta}$ from the vector of data $\boldsymbol{y} = [y_1, y_2, \ldots, y_M]^\top$, a useful approach is provided by the maximum likelihood (ML) method. The likelihood of the vector parameter $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}, \boldsymbol{y}) = \prod_{j=1}^{M} \exp(-\lambda_j)\frac{\lambda_j^{y_j}}{y_j!}, \tag{5}$$

and the loglikelihood is

$$\ln L(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{j=1}^{M} \big\{ -\lambda_j(\boldsymbol{\theta}) + y_j \ln\big[\lambda_j(\boldsymbol{\theta})\big] - \ln(y_j!) \big\}. \tag{6}$$

ML estimation seeks that value of the parameter $\boldsymbol{\theta}$ maximizing the likelihood or equivalently the loglikelihood.

In order to maximize the loglikelihood, we consider the kth component of its gradient

$$\frac{\partial}{\partial \theta_k} \ln L(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{j=1}^{M} \left[-p_{jk} + \frac{y_j p_{jk}}{\lambda_j(\boldsymbol{\theta})} \right]$$

$$= -s_k + \sum_{j=1}^{M} \frac{y_j p_{jk}}{\lambda_j(\boldsymbol{\theta})}, \tag{7}$$

where the second equality follows from the condition of Eq. (4). Equating to zero the N derivatives similar to Eq. (7) provides a system of N equations

$$\sum_{j=1}^{M} \frac{y_j p_{jk}}{\lambda_j(\boldsymbol{\theta})} = s_k, \quad \text{for } k = 1, 2, \ldots, N, \tag{8}$$

or equivalently, thanks to Eq. (2),

$$\sum_{j=1}^{M} \frac{y_j p_{jk}}{r_j + \sum_{i=1}^{N} p_{ji}\theta_i} = s_k, \quad \text{for } k = 1, 2, \ldots, N. \tag{9}$$

The system of Eqs. (8) or (9) is a system of N nonlinear equations in the N unknowns $[\theta_1, \theta_2, \ldots, \theta_N]^\top = \boldsymbol{\theta}$. It is commonly solved numerically by the expectation-maximization (EM-ML) algorithm [22, 42, 44], through the iteration

$$\theta_k^{\text{new}} = \frac{\theta_k^{\text{old}}}{s_k} \sum_{j=1}^{M} \frac{y_j p_{jk}}{\lambda_j(\boldsymbol{\theta}^{\text{old}})}, \quad \text{for } k = 1, 2, \ldots, N. \tag{10}$$

The second derivative of the loglikelihood from Eq. (7) yields

$$\frac{\partial^2}{\partial \theta_k \partial \theta_\ell} \ln L(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{j=1}^{M} \frac{-y_j p_{jk} p_{j\ell}}{[\lambda_j(\boldsymbol{\theta})]^2} \tag{11}$$

$$= \sum_{j=1}^{M} \frac{-y_j p_{jk} p_{j\ell}}{(r_j + \sum_{i=1}^{N} p_{ji}\theta_i)^2}. \tag{12}$$

Since the p_{ji}'s are nonnegative, the derivatives of Eqs. (11)–(12) constitute the coefficients of a negative semidefinite quadratic form of $\mathbb{R}^{N\times 1}$ [42, 44] ensuring that the loglikelihood $\ln L(\boldsymbol{\theta}, \boldsymbol{y})$ is a concave ($\cap$) function in the variable $\boldsymbol{\theta}$. Therefore, any extremum of $\ln L(\boldsymbol{\theta}, \boldsymbol{y})$ is a global maximum. Iteration of the EM-ML algorithm of Eq. (10) converges to such a global maximum of $\ln L(\boldsymbol{\theta}, \boldsymbol{y})$ providing an ML estimate of $\boldsymbol{\theta}$ [22, 42, 44].

4 Characterization of the ML Solution Set

The EM-ML algorithm of Eq. (10) provides a way to one solution to the ML estimation, which especially may depend on the initialization of the iteration process [42, 44]. We shall now try to obtain a more general characterization of the solution set to the ML estimation.

4.1 A Linear System and Its Dual for ML

The ML equation of Eqs. (8) or (9) forms a nonlinear system in the N unknowns $[\theta_1, \theta_2, \ldots, \theta_N]^\top = \boldsymbol{\theta}$. To characterize the general solution to this nonlinear system, we introduce the set of M auxiliary variables $z_j = y_j/\lambda_j$, for $j = 1$ to M. In the variable $\boldsymbol{z} = [z_1, z_2, \ldots, z_M]^\top$, the nonlinear ML equation of Eq. (8) is equivalent to

$$\mathbf{P}^\top \boldsymbol{z} = \boldsymbol{s}, \tag{13}$$

with the $N \times 1$ vector $\boldsymbol{s} = [s_1, s_2, \ldots, s_N]^\top$. Because of Eq. (4), the linear system of Eq. (13) always has for solution the $M \times 1$ vector $\boldsymbol{z} = [1, 1, \ldots, 1]^\top = \mathbf{1}_M$. Thus, the linear system of Eq. (13) has at least one solution $\boldsymbol{z} = \mathbf{1}_M$, and it may have others more. The general solution to Eq. (13) is expressible as $\boldsymbol{z} = \mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)$, with $\text{Ker}(\mathbf{P}^\top)$ the null space of $\mathbf{P}^\top$.

When solving Eq. (13), each component $z_j = y_j/\lambda_j$ of the unknown $\boldsymbol{z}$ has to be considered as a free independent variable. The reason is that, when we are at the stage of solving the ML equation of Eq. (8) and the subsequent Eq. (13), then y_j is known from measurement, but on the contrary λ_j is an unknown scalar, transferring this status to the auxiliary variable z_j. Then every z_j composing a solution $\boldsymbol{z}$ to Eq. (13), determines a value $\widehat{\lambda}_j = y_j/z_j$ for the unknown λ_j. In vector form, $\widehat{\boldsymbol{\lambda}} = \boldsymbol{y}./\boldsymbol{z}$, with $\boldsymbol{y}./\boldsymbol{z}$ indicating component-wise division, i.e. the $M \times 1$ vector with component y_j/z_j for $j = 1$ to M. And this $\widehat{\boldsymbol{\lambda}}$ satisfies the ML equation of (8), because $\boldsymbol{z}$ satisfies Eq. (13). Next, because of the model of Eq. (3), this $\widehat{\boldsymbol{\lambda}}$ satisfying Eq. (8), may have been produced by any source $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ satisfying

$$\mathbf{P}\widehat{\boldsymbol{\theta}} = \boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r}, \tag{14}$$

yet with possibly no $\widehat{\boldsymbol{\theta}}$ solving Eq. (14) and allowing to reach $\widehat{\boldsymbol{\lambda}}$ solving Eq. (8). In this way, not all $\boldsymbol{z}$ solving Eq. (13) can be associated with a $\boldsymbol{\theta}$ solving Eq. (8). But on the contrary, every solution $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ to Eq. (8) is necessarily associated with a $\boldsymbol{z} = \boldsymbol{y}./\boldsymbol{\lambda}(\widehat{\boldsymbol{\theta}})$ solving Eq. (13). Therefore, the solutions to the linear Eq. (14) with $\boldsymbol{z}$ solving Eq. (13) define the only possible solutions to the nonlinear ML equation of (8). It is now tractable to explicitly characterize the solutions to Eq. (14), as a linear system parameterized by $\boldsymbol{z}$ solution to Eq. (13), by means of standard linear algebra considerations [1, 37].

The linear system of Eq. (14) has a solution (at least one) if and only if $\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r}$ is an $M\times 1$ vector belonging to range($\mathbf{P}$), the range of $\mathbf{P}$, which is a subspace of $\mathbb{R}^{M\times 1}$ also known as the consistency space of $\mathbf{P}$ which is identical to $\mathrm{Ker}(\mathbf{P}^\top)^\perp$, the orthogonal complement of the null space $\mathrm{Ker}(\mathbf{P}^\top)$. Equivalently, the linear system of Eq. (14) has a solution (at least one) if and only if

$$\mathbf{P}\mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r})=\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r}, \tag{15}$$

where $\mathbf{P}^+$ is the Moore-Penrose pseudoinverse matrix of $\mathbf{P}$. And when the condition of Eq. (15) is satisfied, any solution to Eq. (14) is given by

$$\widehat{\boldsymbol{\theta}}=\mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r})+\big(\mathbf{I_N}-\mathbf{P}^+\mathbf{P}\big)\boldsymbol{u} \tag{16}$$

$$=\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}+\widehat{\boldsymbol{\theta}}_{\mathbf{0}}, \tag{17}$$

with

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}=\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}(\boldsymbol{y})=\mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r}), \tag{18}$$

and

$$\widehat{\boldsymbol{\theta}}_{\mathbf{0}}=\big(\mathbf{I_N}-\mathbf{P}^+\mathbf{P}\big)\boldsymbol{u}, \tag{19}$$

where $\mathbf{I_N}$ is the $N\times N$ identity matrix and $\boldsymbol{u}$ is an arbitrary vector from $\mathbb{R}^{N\times 1}$.

The pseudoinverse $\mathbf{P}^+$ always satisfies $\mathbf{P}\mathbf{P}^+\mathbf{P}=\mathbf{P}$ and as a consequence, the $N\times 1$ vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ of Eq. (19) verifies $\mathbf{P}\widehat{\boldsymbol{\theta}}_{\mathbf{0}}=\mathbf{0}_{\boldsymbol{M}}$ and stands as an arbitrary vector from the null space of $\mathbf{P}$. This occurs for any $N\times 1$ vector $\boldsymbol{u}$ and therefore establishes the matrix $(\mathbf{I_N}-\mathbf{P}^+\mathbf{P})$ as the projector of $\mathbb{R}^{N\times 1}$ on this null space $\mathrm{Ker}(\mathbf{P})$.

Also, the pseudoinverse $\mathbf{P}^+$ always satisfies $\mathbf{P}^+\mathbf{P}\mathbf{P}^+=\mathbf{P}^+$. As a consequence, the $N\times 1$ vector $\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}$ of Eq. (18), when projected on $\mathrm{Ker}(\mathbf{P})$ yields $(\mathbf{I_N}-\mathbf{P}^+\mathbf{P})\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}=(\mathbf{P}^+-\mathbf{P}^+\mathbf{P}\mathbf{P}^+)(\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r})=\mathbf{0}_N$. Therefore $\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}$ of Eq. (18) is always a $N\times 1$ vector from the subspace $\mathrm{Ker}(\mathbf{P})^\perp$ orthogonal to the null space $\mathrm{Ker}(\mathbf{P})$.

The null space $\mathrm{Ker}(\mathbf{P})$ and its orthogonal complement $\mathrm{Ker}(\mathbf{P})^\perp$ in $\mathbb{R}^{N\times 1}$ form the direct sum $\mathrm{Ker}(\mathbf{P})\oplus\mathrm{Ker}(\mathbf{P})^\perp=\mathbb{R}^{N\times 1}$, ensuring that any vector of $\mathbb{R}^{N\times 1}$ uniquely decomposes as a vector from $\mathrm{Ker}(\mathbf{P})$ plus a vector from $\mathrm{Ker}(\mathbf{P})^\perp$. This is especially the case for the true (unknown) value $\boldsymbol{\theta}=\boldsymbol{\theta}^{\mathrm{true}}$ of the parameter, which uniquely decomposes as

$$\boldsymbol{\theta}^{\mathrm{true}}=\big[\boldsymbol{\theta}_{\mathbf{1}}\in\mathrm{Ker}(\mathbf{P})^\perp\big]+\big[\boldsymbol{\theta}_{\mathbf{0}}\in\mathrm{Ker}(\mathbf{P})\big]. \tag{20}$$

The subspace $\mathrm{Ker}(\mathbf{P})^\perp$ is called the measurement space, and the part $\boldsymbol{\theta}_{\mathbf{1}}$ of $\boldsymbol{\theta}^{\mathrm{true}}$ or the part $\widehat{\boldsymbol{\theta}}_{\boldsymbol{z}}$ of $\widehat{\boldsymbol{\theta}}$, is called the measurement part. Meanwhile, the part of $\boldsymbol{\theta}^{\mathrm{true}}$ or of $\widehat{\boldsymbol{\theta}}$ formed by the null vector $\boldsymbol{\theta}_{\mathbf{0}}$ or $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ of the null space $\mathrm{Ker}(\mathbf{P})$, is called the null part. For consistency with the model of Eq. (3), one has

$$\boldsymbol{\theta}_{\mathbf{1}}=\mathbf{P}^+(\boldsymbol{\lambda}-\boldsymbol{r}), \tag{21}$$

or also $\boldsymbol{\lambda}=\boldsymbol{r}+\mathbf{P}\boldsymbol{\theta}_{\mathbf{1}}$ showing that the measured data are influenced only by the measurement part $\boldsymbol{\theta}_{\mathbf{1}}$ of $\boldsymbol{\theta}^{\mathrm{true}}$ and not by the null part $\boldsymbol{\theta}_{\mathbf{0}}$.

We therefore obtain the following characterization of the general solution to the ML equation of Eq. (8). We consider any $\boldsymbol{z}$ from the subset $\{\mathbf{1}_{\boldsymbol{M}}+\mathrm{Ker}(\mathbf{P}^\top)\}$ and realizing a vector $\boldsymbol{y}./\boldsymbol{z}-\boldsymbol{r}$ belonging to the consistency space $\mathrm{range}(\mathbf{P})=\mathrm{Ker}(\mathbf{P}^\top)^\perp$ defined by Eq. (15). Each such $\boldsymbol{z}$ determines by Eqs. (16)–(17) the solutions to the ML equation of Eq. (8), and there are no other solutions to Eq. (8).

Equations (14) based on $\mathbf{P}$ and (13) based on $\mathbf{P}^\top$, can be seen as two dual linear systems; one is underdetermined when the other is overdetermined, and conversely. By combining the solutions to Eqs. (13) and (14), the general solution to the ML equation of (8) can be further specified in the following way, in the various conditions which can arise when N unknown sources $\boldsymbol{\theta}$ have to be estimated from M measured data $\boldsymbol{y}$.

4.2 Underdetermined System

When $\mathbf{P}$ as rank $\mathrm{rank}(\mathbf{P})\leq M<N$, then the dual system from Eq. (13) with N equations in $M<N$ unknowns, is an overdetermined system with the unique solution $\boldsymbol{z}=\mathbf{1}_{\boldsymbol{M}}$ and $\mathrm{Ker}(\mathbf{P}^\top)$ is the trivial subspace $\{\mathbf{0}_{\boldsymbol{M}}\}$. In this condition, the nonlinear ML equation (8) is equivalent to the linear equation (14) at $\boldsymbol{z}=\mathbf{1}_{\boldsymbol{M}}$, reading

$$\mathbf{P}\widehat{\boldsymbol{\theta}}=\boldsymbol{y}-\boldsymbol{r}. \tag{22}$$

The consistency space is $\mathrm{Ker}(\mathbf{P}^\top)^\perp=\mathbb{R}^{M\times 1}$. The primal system from Eq. (22) with M equations in $N>M$ unknowns, is an underdetermined system, with an infinite number of solutions under the form $\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}_{\mathbf{1}}+\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ according to Eqs. (16)–(17) at $\boldsymbol{z}=\mathbf{1}_{\boldsymbol{M}}$, i.e. with

$$\widehat{\boldsymbol{\theta}}_{\mathbf{1}}=\mathbf{P}^+(\boldsymbol{y}-\boldsymbol{r}). \tag{23}$$

There is thus an infinite number of distinct ML estimators $\widehat{\boldsymbol{\theta}}$, each for every arbitrary vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ from the null space $\mathrm{Ker}(\mathbf{P})$ in Eq. (19). However, the solution in the measurement space $\mathrm{Ker}(\mathbf{P})^\perp$, i.e. $\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}_{\mathbf{1}}$ from Eq. (23), always forms the solution of minimum Euclidean norm to Eq. (22). Also, there is no other ML estimator than these $\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}_{\mathbf{1}}+\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ solutions to Eq. (22).

For the underdetermined system, the solutions of the nonlinear ML equation (8) are thus equivalent to the solutions of the linear equation (22) which has the same form as the model equation (3). The linear equation (22), when it admits solutions, defines the moment estimates, as called in [44], which are equivalent to the ML estimates. This is no longer true, as we shall see, in the case of the overdetermined system, when Eq. (22) typically admits no solution and an ML estimator solution to Eq. (8) still exists in general.

4.3 Uniquely Determined System

When $\text{rank}(\mathbf{P}) = M = N$, then the dual system from Eq. (13) is a uniquely determined system with the unique solution $z = \mathbf{1}_M$ and the trivial null space $\text{Ker}(\mathbf{P}^\top) = \{\mathbf{0}_M\}$. Also, the primal system formed by Eq. (14) at $z = \mathbf{1}_M$, is a uniquely determined system with the unique solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ realizing the unique ML estimator, since the pseudoinverse $\mathbf{P}^+$ coincides with the true inverse $\mathbf{P}^{-1}$ and Eq. (19) gives the null vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} \in \{\mathbf{0}_N\} = \text{Ker}(\mathbf{P})$.

4.4 Overdetermined System

When $\text{rank}(\mathbf{P}) \leq N < M$ or $\text{rank}(\mathbf{P}) < N = M$, then the dual system from Eq. (13) is in general an underdetermined system with an infinite number of solutions under the form $z = \mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)$ and $\text{Ker}(\mathbf{P}^\top) \supset \{\mathbf{0}_M\}$. Equivalently, using the projector $\mathbf{I_M} - \mathbf{P}\mathbf{P}^+$ onto $\text{Ker}(\mathbf{P}^\top)$, any z solution to Eq. (13) is expressible as $z = \mathbf{1}_M + (\mathbf{I_M} - \mathbf{P}\mathbf{P}^+)\boldsymbol{v}$ for any vector $\boldsymbol{v}$ from $\mathbb{R}^{M\times 1}$. For each z solution to Eq. (13), the associated primal system from Eq. (14) is in general an overdetermined system. The consistency space is $\text{Ker}(\mathbf{P}^\top)^\perp \subset \mathbb{R}^{M\times 1}$, and in such case, Eq. (14) has the unique solution

$$\widehat{\boldsymbol{\theta}} = \mathbf{P}^+(\boldsymbol{y}./z - \boldsymbol{r}) \tag{24}$$

or no solution. The solution to Eq. (14) exists as Eq. (24) if and only if Eq. (15) is satisfied. For each z solving Eq. (13), the solution in Eq. (24), when it exists, defines an ML estimator. One such a solution to the ML equation of Eq. (8) can be expected to always exist, in nondegenerate tomographic systems where the intensities $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are not all simultaneously zero. In such nondegenerate conditions, for any valid data $\boldsymbol{y} \in [0, \infty[^M$, the likelihood $L(\boldsymbol{\theta}, \boldsymbol{y})$ of Eqs. (5)–(6) is > 0 for any finite $\boldsymbol{\theta}$, with a vanishing asymptotic behavior $L(\boldsymbol{\theta}, \boldsymbol{y}) \to 0$ for large $\boldsymbol{\theta}$ of norm $\|\boldsymbol{\theta}\|_2 \to \infty$. Since $L(\boldsymbol{\theta}, \boldsymbol{y})$ is a concave ($\cap$) function in the variable $\boldsymbol{\theta}$, there necessarily exists at least one finite $\boldsymbol{\theta}$ maximizing $L(\boldsymbol{\theta}, \boldsymbol{y})$ and therefore solving Eq. (8); and such $\boldsymbol{\theta}$ is bound to follow as a solution to Eq. (14) under the form of Eq. (24) for some z in the subset $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)\}$.

It can also be argued that such an ML solution can generally be expected to be unique. A z in $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)\}$ verifies Eq. (13) which represents a system of R independent scalar equations, to be satisfied by the $M > R$ scalar components of z, with $R = \text{rank}(\mathbf{P}^\top) = \text{rank}(\mathbf{P}) \leq N < M$. Equivalently, since the dimension of $\text{Ker}(\mathbf{P}^\top)$ is $M - R$, the $M \times 1$ vector z solving Eq. (13) is forced to live in the subset $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)\}$ by R scalar constraints. Next, for a fixed data vector $\boldsymbol{y}$, for an ML solution to exist, z solving Eq. (13) has also to satisfy Eq. (15). Since also $\text{rank}(\mathbf{P}\mathbf{P}^+) = \text{rank}(\mathbf{P}) = R$, the system of M scalar equations formed by Eq. (15) contains only $M - R$ independent scalar constraints while the R others are redundant and automatically satisfied. Equivalent to Eq. (15) is that the $M \times 1$ vector $(\boldsymbol{y}./z - \boldsymbol{r})$ in Eq. (14) has to live in the consistency space $\text{range}(\mathbf{P}) = \text{Ker}(\mathbf{P}^\top)^\perp$ having dimension $R = \text{rank}(\mathbf{P})$. This is obtained by $M - R$ independent scalar constraints expressed by Eq. (15), reducing the degrees of freedom of $(\boldsymbol{y}./z - \boldsymbol{r})$ from M to R. Since $\boldsymbol{y}$ and $\boldsymbol{r}$ are fixed given $M \times 1$ vectors, this is in fact onto z that the $M - R$ independent scalar constraints are imposed by Eq. (15). Summarizing, on the $M \times 1$ vector z, R scalar constraints are imposed by Eq. (13) and $M - R$ by Eq. (15). This is a total of M constraints imposed together by Eqs. (13) and (15) which generally determine one unique z, which in Eq. (24) fixes one unique ML estimator. This is true except maybe in some peculiar conditions where more than one solution for z might exist.

It should be noted that, because of the division by z in Eq. (15), the combination of Eqs. (13) and (15) which in general uniquely determines z, forms a system of *nonlinear* equations in z. This is a resurgence of the nonlinearity initially present in the ML equation (8). Once a solution z to the system of Eqs. (13) and (15) is known, an explicit form follows for the ML estimator $\widehat{\boldsymbol{\theta}}$ with Eq. (24). Beyond the theoretical characterization of the ML estimator it enables, the nonlinear system formed by Eqs. (13) and (15), from a practical standpoint may not be straightforward to solve. It offers nevertheless another route to the ML estimator, complementing for instance the EM-ML algorithm of Eq. (10), since both approaches are bound to yield the same ML estimator when it is unique.

We note that in general the simple choice $z = \mathbf{1}_M$ does not solve Eq. (15). This would impose $(\boldsymbol{y} - \boldsymbol{r})$ to permanently remain an eigenvector of $\mathbf{P}\mathbf{P}^+$ for any data $\boldsymbol{y}$. More degrees of freedom with $z \neq \mathbf{1}_M$ are usually needed to satisfy Eq. (15) with an arbitrary $\boldsymbol{y}$. The eigendirections of $\mathbf{P}\mathbf{P}^+$ are fixed, and $z \neq \mathbf{1}_M$ is needed to place $(\boldsymbol{y}./z - \boldsymbol{r})$ in an eigendirection of $\mathbf{P}\mathbf{P}^+$ for any $\boldsymbol{y}$. A simple interpretation of Eq. (24) is that the ML estimator $\widehat{\boldsymbol{\theta}}$ performs a nonuniform attenuation or amplification of the data $\boldsymbol{y}$, selectively for each detector j, controlled by z from $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)\}$.

An Approximation to the ML Solution It can be envisaged, as an approximation to the ML solution, to select a z which does not solve Eq. (15), then Eq. (14) has no solution for this z. Nevertheless, $\widehat{\boldsymbol{\theta}}$ from Eq. (24) still achieves the minimum of the Euclidean norm $\|\mathbf{P}\widehat{\boldsymbol{\theta}} - (\boldsymbol{y}./z - \boldsymbol{r})\|_2$ associated with Eq. (14), and as such this $\widehat{\boldsymbol{\theta}}$ may constitute a useful estimator, especially via the following path. If the z which does not solve Eq. (15) is nevertheless selected from the subset $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^\top)\}$, then this z solves Eq. (13) and therefore determines $\widehat{\boldsymbol{\lambda}} = \boldsymbol{y}./z$ satisfying the ML equation (8). If the norm is some small scalar $\varepsilon > 0$, i.e. $\|\mathbf{P}\widehat{\boldsymbol{\theta}} - (\boldsymbol{y}./z - \boldsymbol{r})\|_2 = \varepsilon$, then we have the vector $\mathbf{P}\widehat{\boldsymbol{\theta}} + \boldsymbol{r}$ which differs from the vector $\boldsymbol{y}./z$ by a small vector $\boldsymbol{\epsilon}$ with the small norm $\|\boldsymbol{\epsilon}\|_2 = \varepsilon$, i.e.

$\mathbf{P}\widehat{\boldsymbol{\theta}} + \boldsymbol{r} = \boldsymbol{y}./z + \boldsymbol{\epsilon} \approx \boldsymbol{y}./z = \widehat{\boldsymbol{\lambda}}$. In this respect, the estimator $\widehat{\boldsymbol{\theta}}$ from Eq. (24) performs close, by the small amount $\boldsymbol{\epsilon}$, to a genuine ML estimator exactly solving Eq. (8).

The characterization of the ML estimator under the form of Eq. (24) is thus also interesting as a form to suggest useful alternative estimators. The procedure will be, employing Eq. (24), to select a z from the subset $\{\mathbf{1}_M + \mathrm{Ker}(\mathbf{P}^\top)\}$ achieving a small norm $\|\mathbf{P}\widehat{\boldsymbol{\theta}} - (\boldsymbol{y}./z - \boldsymbol{r})\|_2$ relative to some appropriate level of precision. Then, such a z will settle in Eq. (24) an estimator behaving close to a genuine ML estimator. Various heuristics can be devised to select such an interesting z, possibly iteratively, among the form $z = \mathbf{1}_M + (\mathbf{I_M} - \mathbf{P}\mathbf{P}^+)\boldsymbol{v}$ for any vector $\boldsymbol{v}$ from $\mathbb{R}^{M\times 1}$ which uses the projector $\mathbf{I_M} - \mathbf{P}\mathbf{P}^+$ onto $\mathrm{Ker}(\mathbf{P}^\top)$.

4.5 Expressing the Pseudoinverse

With N unknown sources $\boldsymbol{\theta}$ to be estimated from M measured data $\boldsymbol{y}$ several configurations can arise, which control the explicit expression ascribable to the Moore-Penrose pseudoinverse matrix $\mathbf{P}^+$ introduced in Eq. (15) and which served to us to express the solutions to the nonlinear ML equation of (8) for tomographic reconstruction. From standard linear algebra, these possible expressions for $\mathbf{P}^+$ come as follows [1, 37].

A possible configuration in tomography [42, 44] is when the $M \times N$ system matrix $\mathbf{P}$ as rank $\mathrm{rank}(\mathbf{P}) = M < N$, i.e. $\mathbf{P}$ has M independent rows. In this case, the expression of the pseudoinverse $\mathbf{P}^+$ is

$$\mathbf{P}^+ = \mathbf{P}^\top \left(\mathbf{P}\mathbf{P}^\top\right)^{-1}, \tag{25}$$

and as a consequence $\mathbf{P}^+$ is the right inverse of $\mathbf{P}$ verifying $\mathbf{P}\mathbf{P}^+ = \mathbf{I_M}$. This ensures that Eq. (15) is automatically satisfied, and the subsequent solution exists.

A special configuration is when $\mathrm{rank}(\mathbf{P}) = M = N$, yielding $\mathbf{P}^+ = \mathbf{P}^{-1}$ the true inverse. Here $\mathbf{P}$ is perfectly invertible, and $\mathbf{P}^+\mathbf{P} = \mathbf{I_N}$ cancels the vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in Eq. (19). There is then a unique ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ from Eq. (16).

Another configuration is when $\mathrm{rank}(\mathbf{P}) = N < M$, i.e. $\mathbf{P}$ has N independent columns. In this case, the expression of the pseudoinverse is

$$\mathbf{P}^+ = \left(\mathbf{P}^\top\mathbf{P}\right)^{-1}\mathbf{P}^\top, \tag{26}$$

and as a consequence $\mathbf{P}^+$ is the left inverse of $\mathbf{P}$ verifying $\mathbf{P}^+\mathbf{P} = \mathbf{I_N}$. This is enough to cancel the null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ of Eq. (19), in the solution to Eq. (14). Its solutions exist in the measurement space $\mathrm{Ker}(\mathbf{P})^\perp$ according to Eq. (24) or do not exist.

Other more peculiar configurations, when $\mathbf{P}$ has a rank strictly smaller than the minimum of M and N, can still be handled, from standard linear algebra, to obtain an explicit expression for the pseudoinverse $\mathbf{P}^+$.

5 Performance of the ML Estimator

Equation (16) parameterized by z solution to Eq. (13) offers an analytical form for the solution to the ML reconstruction in emission tomography. This can serve to a characterization of the performance of the estimators deduced from Eq. (16), by determining their bias and mean-squared error as we shall now do. Yet before, we briefly recall a general limit of performance in estimation established via the Fisher information.

5.1 Fisher Information

In general, for estimation of an $N \times 1$ vector parameter $\boldsymbol{\theta}$ from the $M \times 1$ data $\boldsymbol{y}$, the Fisher information matrix $\mathbf{F} = [F_{k\ell}]$ is the symmetric $N \times N$ matrix with the $F_{k\ell}$ element defined from the likelihood $L(\boldsymbol{\theta}, \boldsymbol{y})$ as the statistical average

$$F_{k\ell} = -\left\langle \frac{\partial^2}{\partial\theta_k \partial\theta_\ell} \ln L(\boldsymbol{\theta}, \boldsymbol{y}) \right\rangle. \tag{27}$$

The Fisher information allows for a general characterization of the performance in estimation. Any estimator $\widetilde{\boldsymbol{\theta}}(\boldsymbol{y})$ for the parameter $\boldsymbol{\theta}$, achieves an estimation error $\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$, associated with a bias $\boldsymbol{b}(\widetilde{\boldsymbol{\theta}}) = \langle \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \rangle$, and a correlation matrix of the error $\mathbf{C}$ satisfying the Cramér-Rao inequality [1]

$$\mathbf{C} = \left\langle (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \right\rangle \geq \mathbf{B}_{\mathrm{CR}}, \tag{28}$$

where the Cramér-Rao lower bound is

$$\mathbf{B}_{\mathrm{CR}} = \boldsymbol{b}\boldsymbol{b}^\top + \left(\mathbf{I_N} + \boldsymbol{\nabla}\boldsymbol{b}^\top\right)\mathbf{F}^+\left(\mathbf{I_N} + \boldsymbol{\nabla}\boldsymbol{b}^\top\right)^\top, \tag{29}$$

with the $N \times 1$ nabla derivation operator $\boldsymbol{\nabla}\cdot = [\partial\cdot/\partial\theta_1, \ldots, \partial\cdot/\partial\theta_N]^\top$. The pseudoinverse $\mathbf{F}^+$ in Eq. (28) reduces to the true inverse $\mathbf{F}^{-1}$ when the square matrix $\mathbf{F}$ is nonsingular.

In the present model for emission tomography, for the Poisson processes y_j we have from Eq. (1) the statistical average $\langle y_j \rangle = \lambda_j$, for $j = 1$ to M. Then, based on Eq. (2), we have Eq. (11) leading to

$$F_{k\ell} = \sum_{j=1}^{M} \frac{p_{jk}\, p_{j\ell}}{\lambda_j(\boldsymbol{\theta})}. \tag{30}$$

We therefore have in matrix form

$$\mathbf{F} = \mathbf{P}^\top\, \mathbf{diag}\left[\frac{1}{\lambda_j(\boldsymbol{\theta})}\right]\mathbf{P}, \tag{31}$$

where $\mathbf{diag}[\cdot]$ denotes the diagonal matrix with the specified elements, and with the proviso that the Fisher information is evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^{\mathrm{true}}$, the true value of the vector parameter [28]. The characterization of Eq. (31) of the Fisher information matrix in emission tomography is well known and

can be found in [20, 28] for instance. From Eq. (31), we can push the characterization a little further by noting

$$\mathbf{F} = \mathbf{A}^\top \mathbf{A}, \tag{32}$$

with the $M \times N$ matrix

$$\mathbf{A} = \mathbf{diag}\left[\frac{1}{\sqrt{\lambda_j}}\right]\mathbf{P}. \tag{33}$$

And based on a property of the pseudoinverse [1],

$$\mathbf{F}^+ = \left(\mathbf{A}^\top \mathbf{A}\right)^+ = \mathbf{A}^+ \mathbf{A}^{+\top}, \tag{34}$$

since $\mathbf{A}^{\top +} = \mathbf{A}^{+\top}$ always.

From matrix theory [1], if a matrix $\boldsymbol{X_1}$ with M columns has (full column) rank M, and a matrix $\boldsymbol{X_2}$ with M rows has (full row) rank M, then $(\boldsymbol{X_1 X_2})^+ = \boldsymbol{X_2}^+ \boldsymbol{X_1}^+$. In $\mathbf{A}$ of Eq. (33), the $M \times M$ diagonal matrix $\mathbf{diag}[1/\sqrt{\lambda_j}]$ has always (full column) rank M. Then, only when the system is not overdetermined, i.e. when $M \leq N$ and $\mathbf{P}$ has rank M, can we conclude that

$$\mathbf{A}^+ = \mathbf{P}^+ \, \mathbf{diag}[\sqrt{\lambda_j}], \quad M \leq N, \tag{35}$$

since $(\mathbf{diag}[1/\sqrt{\lambda_j}])^+ = (\mathbf{diag}[1/\sqrt{\lambda_j}])^{-1} = \mathbf{diag}[\sqrt{\lambda_j}]$. And in this case, the pseudoinverse Fisher information matrix of Eq. (34) can receive a more explicit expression as

$$\mathbf{F}^+ = \mathbf{P}^+ \, \mathbf{diag}[\lambda_j] \mathbf{P}^{+\top}, \quad M \leq N. \tag{36}$$

Otherwise, when $M > N$ for the overdetermined system, Eq. (35) does not hold, nor Eq. (36).

One can also use the variance-covariance matrix of the error

$$\mathbf{V} = \left\langle \left(\widetilde{\boldsymbol{\theta}} - \langle \widetilde{\boldsymbol{\theta}} \rangle\right)\left(\widetilde{\boldsymbol{\theta}} - \langle \widetilde{\boldsymbol{\theta}} \rangle\right)^\top \right\rangle, \tag{37}$$

related to the correlation matrix $\mathbf{C} = \mathbf{V} + \boldsymbol{b}\boldsymbol{b}^\top$, and, from Eqs. (28)–(29), satisfying the inequality

$$\mathbf{V} \geq \nabla \langle \widetilde{\boldsymbol{\theta}} \rangle^\top \, \mathbf{F}^+ \left(\nabla \langle \widetilde{\boldsymbol{\theta}} \rangle^\top\right)^\top. \tag{38}$$

5.2 Bias

5.2.1 *Underdetermined System*

We start with the condition of an underdetermined system as in Sect. 4.2 when the ML estimator is $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$. The measurement part $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}}(\boldsymbol{y}) = \mathbf{P}^+(\boldsymbol{y} - \boldsymbol{r})$ is dependent on the data $\boldsymbol{y}$, according to Eq. (23). However, the null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} = \widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$ of Eq. (19) can, in general, also be considered dependent on the data $\boldsymbol{y}$, if the estimation algorithm constructs it in this way. We then have the statistical average

$$\langle \widehat{\boldsymbol{\theta}} \rangle = \langle \widehat{\boldsymbol{\theta}}_{\mathbf{1}} \rangle + \langle \widehat{\boldsymbol{\theta}}_{\mathbf{0}} \rangle. \tag{39}$$

With the statistical average $\langle y_j \rangle = \lambda_j$, for $j = 1$ to M, in vector form, from Eq. (3), we have $\langle \boldsymbol{y} \rangle = \boldsymbol{\lambda} = \boldsymbol{r} + \mathbf{P}\boldsymbol{\theta}$. From Eqs. (23) and (21) we then obtain

$$\langle \widehat{\boldsymbol{\theta}}_{\mathbf{1}} \rangle = \mathbf{P}^+\left(\langle \boldsymbol{y} \rangle - \boldsymbol{r}\right) = \mathbf{P}^+(\boldsymbol{\lambda} - \boldsymbol{r}) = \boldsymbol{\theta}_{\mathbf{1}}. \tag{40}$$

In this way, Eq. (39) gives

$$\langle \widehat{\boldsymbol{\theta}} \rangle = \boldsymbol{\theta}_{\mathbf{1}} + \langle \widehat{\boldsymbol{\theta}}_{\mathbf{0}} \rangle. \tag{41}$$

Confronting Eq. (41) with $\boldsymbol{\theta} = \boldsymbol{\theta}^{\text{true}}$ from Eq. (20), we conclude that the ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ has a bias $\boldsymbol{b}(\widehat{\boldsymbol{\theta}}) = \langle \widehat{\boldsymbol{\theta}} \rangle - \boldsymbol{\theta}$ which reduces to the $N \times 1$ vector

$$\boldsymbol{b}(\widehat{\boldsymbol{\theta}}) = \langle \widehat{\boldsymbol{\theta}}_{\mathbf{0}} \rangle - \boldsymbol{\theta}_{\mathbf{0}}. \tag{42}$$

The ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ for the underdetermined system of Sect. 4.2 is therefore unbiased if and only if $\langle \widehat{\boldsymbol{\theta}}_{\mathbf{0}} \rangle = \boldsymbol{\theta}_{\mathbf{0}}$. This means that the estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ has to construct its estimate by incorporating a null vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ whose average is the null vector $\boldsymbol{\theta}_{\mathbf{0}}$ of $\text{Ker}(\mathbf{P})$ that makes up the true parameter $\boldsymbol{\theta}^{\text{true}} = \mathbf{P}^+(\boldsymbol{\lambda} - \boldsymbol{r}) + \boldsymbol{\theta}_{\mathbf{0}}$, if ever it exists and can be known.

5.2.2 *Uniquely Determined System*

We note that a zero bias is automatically achieved in the condition of a uniquely determined system as in Sect. 4.3, when the ML estimator reduces to $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$.

5.2.3 *Overdetermined System*

We now turn to the condition of an overdetermined system as in Sect. 4.4 when the estimator is $\widehat{\boldsymbol{\theta}} = \mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ of Eq. (24). If we want this $\widehat{\boldsymbol{\theta}}$ of Eq. (24) to be an exact ML estimator solving Eq. (8), then in general the value of $\boldsymbol{z}$ in Eq. (24) has to be precisely matched to the data $\boldsymbol{y}$, by solving the nonlinear system of Eqs. (13) and (15) as explained in Sect. 4.4. Therefore, in general, the exact ML estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ of Eq. (24) involves a data-dependent $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{y})$. Then this estimator has the statistical average

$$\left\langle \widehat{\boldsymbol{\theta}}(\boldsymbol{y}) \right\rangle = \mathbf{P}^+\left(\langle \boldsymbol{y}./\boldsymbol{z} \rangle - \boldsymbol{r}\right). \tag{43}$$

The overdetermined system is associated with the trivial null space $\text{Ker}(\mathbf{P}) = \{\mathbf{0}_N\}$, and from Eq. (20) the unknown true parameter reduces to the form $\boldsymbol{\theta} = \boldsymbol{\theta}^{\text{true}} = \mathbf{P}^+(\boldsymbol{\lambda} - \boldsymbol{r})$ of Eq. (21). The bias $\boldsymbol{b}(\widehat{\boldsymbol{\theta}}) = \langle \widehat{\boldsymbol{\theta}} \rangle - \boldsymbol{\theta}$ therefore follows as

$$\boldsymbol{b}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^+\left(\langle \boldsymbol{y}./\boldsymbol{z} \rangle - \boldsymbol{\lambda}\right). \tag{44}$$

We always have $\langle \boldsymbol{y} \rangle = \boldsymbol{\lambda}$. But we have seen in Sect. 4.4 that in general $\boldsymbol{z}(\boldsymbol{y}) \neq \mathbf{1}_M$ for the ML estimator. The bias of Eq. (44) is thus unlikely to vanish in general. It is however difficult to further specify the statistical evaluation of the ML estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^+(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ in the absence of an explicit expression for $\boldsymbol{z}(\boldsymbol{y})$ pending to analytical resolution of the nonlinear system of Eqs. (13) and (15).

Bias in an Approximation to the ML Solution As an alternative, a tractable reference is the situation where $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ of Eq. (24) is constructed with a data-independent z to implement an approximation to the ML estimator in the sense described at the end of Sect. 4.4. For the estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ from Eq. (24) with a fixed data-independent z, the average of Eq. (43) reduces to

$$\langle\widehat{\boldsymbol{\theta}}\rangle = \mathbf{P}^{+}(\boldsymbol{\lambda}./z - \boldsymbol{r}), \tag{45}$$

and the bias to

$$\boldsymbol{b}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^{+}(\boldsymbol{\lambda}./z - \boldsymbol{\lambda}). \tag{46}$$

For a fixed $z \neq \mathbf{1}_M$ the bias of Eq. (46) is also unlikely to vanish in general; this would require $\boldsymbol{\lambda}./z - \boldsymbol{\lambda}$ to always remain in the null space $\mathrm{Ker}(\mathbf{P}^{+})$, and this is usually not possible with $\mathbf{P}$ fixed by the tomographic system and $\boldsymbol{\lambda}$ varying with the imaged scene. The bias of Eq. (46) vanishes in the special case $z = \mathbf{1}_M$, but this choice removes the flexibility of a more interesting choice $z \neq \mathbf{1}_M$ from the subset $\{\mathbf{1}_M + \mathrm{Ker}(\mathbf{P}^{\top})\}$ as explained in Sect. 4.4.

From Eq. (46) one also has the derivative

$$\boldsymbol{\nabla}\boldsymbol{b}^{\top} = \mathbf{P}^{\top}\,\mathbf{diag}[1/z_j]\mathbf{P}^{+\top} - \mathbf{I_N}. \tag{47}$$

5.3 Correlation of the Error

The estimator $\widehat{\boldsymbol{\theta}}$ achieves an estimation error $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ with a correlation given by the $N \times N$ symmetric matrix

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \big\langle(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\top}\big\rangle. \tag{48}$$

5.3.1 Underdetermined System

In the condition of the underdetermined system of Sect. 4.2 when the ML estimator is $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \widehat{\boldsymbol{\theta}}_{\mathbf{1}}(\boldsymbol{y}) + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$, by developing and rearranging Eq. (48), and using $\langle\widehat{\boldsymbol{\theta}}_{\mathbf{1}}\rangle = \boldsymbol{\theta}_{\mathbf{1}}$ from Eq. (40), one obtains

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{C_1} + \mathbf{C_0} + \mathbf{C_{01}} + \mathbf{C_{01}}^{\top}, \tag{49}$$

with the $N \times N$ matrices

$$\mathbf{C_1} = \big\langle(\widehat{\boldsymbol{\theta}}_{\mathbf{1}} - \boldsymbol{\theta}_{\mathbf{1}})(\widehat{\boldsymbol{\theta}}_{\mathbf{1}} - \boldsymbol{\theta}_{\mathbf{1}})^{\top}\big\rangle = \big\langle\widehat{\boldsymbol{\theta}}_{\mathbf{1}}\widehat{\boldsymbol{\theta}}_{\mathbf{1}}^{\top}\big\rangle - \boldsymbol{\theta}_{\mathbf{1}}\boldsymbol{\theta}_{\mathbf{1}}^{\top}, \tag{50}$$

$$\mathbf{C_0} = \big\langle(\widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}})(\widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}})^{\top}\big\rangle, \tag{51}$$

and

$$\mathbf{C_{01}} = \big\langle\widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\widehat{\boldsymbol{\theta}}_{\mathbf{1}} - \boldsymbol{\theta}_{\mathbf{1}})^{\top}\big\rangle = \big\langle\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\widehat{\boldsymbol{\theta}}_{\mathbf{1}}^{\top}\big\rangle - \langle\widehat{\boldsymbol{\theta}}_{\mathbf{0}}\rangle\boldsymbol{\theta}_{\mathbf{1}}^{\top}. \tag{52}$$

From Eq. (23) it comes

$$\big\langle\widehat{\boldsymbol{\theta}}_{\mathbf{1}}\widehat{\boldsymbol{\theta}}_{\mathbf{1}}^{\top}\big\rangle = \mathbf{P}^{+}\big\langle(\boldsymbol{y} - \boldsymbol{r})(\boldsymbol{y} - \boldsymbol{r})^{\top}\big\rangle\mathbf{P}^{+\top}. \tag{53}$$

For the M independent Poisson processes y_j of Eq. (1) one has the statistical average $\langle y_j^2\rangle = \lambda_j^2 + \lambda_j$, for $j = 1$ to M, yielding

$$\big\langle\boldsymbol{y}\boldsymbol{y}^{\top}\big\rangle = \boldsymbol{\lambda}\boldsymbol{\lambda}^{\top} + \mathbf{diag}[\lambda_j], \tag{54}$$

whence

$$\big\langle(\boldsymbol{y} - \boldsymbol{r})(\boldsymbol{y} - \boldsymbol{r})^{\top}\big\rangle = (\boldsymbol{\lambda} - \boldsymbol{r})(\boldsymbol{\lambda} - \boldsymbol{r})^{\top} + \mathbf{diag}[\lambda_j]. \tag{55}$$

Then Eq. (53) becomes

$$\big\langle\widehat{\boldsymbol{\theta}}_{\mathbf{1}}\widehat{\boldsymbol{\theta}}_{\mathbf{1}}^{\top}\big\rangle = \boldsymbol{\theta}_{\mathbf{1}}\boldsymbol{\theta}_{\mathbf{1}}^{\top} + \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{+\top}. \tag{56}$$

As a result, the correlation matrix of the error in Eq. (49) is expressible as

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{+\top} + \big\langle(\widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}})(\widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}})^{\top}\big\rangle + \mathbf{C_{01}} + \mathbf{C_{01}}^{\top}. \tag{57}$$

Further characterization of the error correlation matrix $\mathbf{C}(\widehat{\boldsymbol{\theta}})$ requires to explicitly specify the null vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$ implemented by the estimation algorithm to construct its ML estimate.

For instance, a special case is the situation of the ML estimator $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ with a constant $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ independent of the data $\boldsymbol{y}$. In this case, the error correlation matrix of Eq. (57) becomes

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{b}\boldsymbol{b}^{\top} + \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{+\top}, \tag{58}$$

with the constant bias $\boldsymbol{b} = \widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}}$ from Eq. (42). Moreover, in this special case of a constant bias, the right-hand side of Eq. (58) precisely matches the Cramér-Rao bound $\mathbf{B}_{\mathrm{CR}}$ of Eq. (29), thanks to Eq. (36) for the underdetermined system. Therefore, an ML estimator under the form $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ with a constant $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$, is always efficient, since its error has a correlation matrix $\mathbf{C}(\widehat{\boldsymbol{\theta}})$ reaching the Cramér-Rao lower bound $\mathbf{B}_{\mathrm{CR}}$. However, we are dealing here with biased ML estimators, and so the Cramér-Rao bound is not intrinsic but is dependent on the estimator itself, via its constant bias $\boldsymbol{b} = \widehat{\boldsymbol{\theta}}_{\mathbf{0}} - \boldsymbol{\theta}_{\mathbf{0}}$. It is therefore reasonable to ask for the ML estimator $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ associated with the lowest (in some sense) Cramér-Rao bound $\mathbf{B}_{\mathrm{CR}}$. A sensible approach is to concentrate on the trace $\mathrm{tr}[\mathbf{C}(\widehat{\boldsymbol{\theta}})]$ of the correlation matrix of the error, which matches the sum of the mean-squared error of each component of the estimator, i.e. $\mathrm{tr}[\mathbf{C}(\widehat{\boldsymbol{\theta}})] = \sum_{i=1}^{N}\langle(\widehat{\theta}_i - \theta_i)^2\rangle$, and as such offers a convenient scalar assessment of the performance in estimation. Then, converting Eq. (58) into an equality on traces, the estimator-dependent part of $\mathrm{tr}[\mathbf{C}(\widehat{\boldsymbol{\theta}})]$ is $\mathrm{tr}[\boldsymbol{b}\boldsymbol{b}^{\top}] = \boldsymbol{b}^{\top}\boldsymbol{b}$. Since the inner product $\boldsymbol{b}^{\top}\boldsymbol{b}$, being the squared norm of the bias $\boldsymbol{b}$, is nonnegative, it results that the efficient ML estimator that minimizes $\mathrm{tr}[\mathbf{C}(\widehat{\boldsymbol{\theta}})]$ is the unbiased one with the constant null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} = \boldsymbol{\theta}_{\mathbf{0}}$. However, since

the null part $\boldsymbol{\theta_0}$ of $\boldsymbol{\theta}^{\text{true}}$ is usually unknown, such estimator is usually inaccessible.

One can turn to an ML estimator $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$ with a nonconstant data-dependent null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$, to possibly expect a better performance from $\mathbf{C}(\widehat{\boldsymbol{\theta}})$ in Eq. (57), possibly associated with a lower Cramér-Rao bound $\mathbf{B}_{\text{CR}}$ in Eq. (29) especially afforded by the extra degree of freedom from $\nabla \boldsymbol{b}$. The feasibility of this option rests on the ability of devising an appropriate data-dependent null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$ for the estimator.

5.3.2 Uniquely Determined System

In the special case of a uniquely determined system as in Sect. 4.3, there is a unique ML estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ which is unbiased, $\mathbf{P}^{+} = \mathbf{P}^{-1}$ exists as a true inverse, and Eq. (58) is still valid but reduces to

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^{-1}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{-1\top}. \tag{59}$$

The right-hand side of Eq. (59) matches here the Cramér-Rao bound $\mathbf{B}_{\text{CR}}$ of Eq. (29), manifesting that the ML estimator is efficient.

5.3.3 Overdetermined System

In the condition of the overdetermined system of Sect. 4.4, the estimator is $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ given by Eq. (24) with in general a data-dependent $z = z(\boldsymbol{y})$. The estimation error is $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{\lambda})$ having a correlation matrix, which after some straightforward algebra, follows as

$$\begin{aligned}\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{b}\boldsymbol{b}^{\top} + \mathbf{P}^{+}\big[&\big\langle(\boldsymbol{y}./z)(\boldsymbol{y}./z)^{\top}\big\rangle \\ &- \langle \boldsymbol{y}./z\rangle\langle \boldsymbol{y}./z\rangle^{\top}\big]\mathbf{P}^{+\top},\end{aligned} \tag{60}$$

which is lower-bounded by the Cramér-Rao bound $\mathbf{B}_{\text{CR}}$ of Eq. (29). Further specification of $\mathbf{C}(\widehat{\boldsymbol{\theta}})$ from Eq. (60) would require an explicit expression for $z(\boldsymbol{y})$ solving the nonlinear system of Eqs. (13) and (15).

Error in an Approximation to the ML Solution In the simpler approach of a constant data-independent z when $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ approximates the ML estimator, Eq. (60) reduces to

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{b}\boldsymbol{b}^{\top} + \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j/z_j]\mathbf{P}^{+\top}, \tag{61}$$

with the bias $\boldsymbol{b}$ from Eq. (46). Moreover, through Eq. (47), the Cramér-Rao bound of Eq. (29) here gives

$$\mathbf{B}_{\text{CR}} = \boldsymbol{b}\boldsymbol{b}^{\top} + \mathbf{P}^{\top}\,\mathbf{diag}[1/z_j]\mathbf{P}^{+\top}\mathbf{F}^{+}\mathbf{P}^{+}\,\mathbf{diag}[1/z_j]\mathbf{P}, \tag{62}$$

or equivalently

$$\begin{aligned}\mathbf{B}_{\text{CR}} = \boldsymbol{b}\boldsymbol{b}^{\top} &+ \mathbf{P}^{\top}\,\mathbf{diag}[1/z_j]\mathbf{P}^{+\top}\big(\mathbf{P}^{\top}\,\mathbf{diag}[1/\lambda_j]\mathbf{P}\big)^{+} \\ &\times \mathbf{P}^{+}\,\mathbf{diag}[1/z_j]\mathbf{P}.\end{aligned} \tag{63}$$

The right-hand side of Eq. (61) generally differs from the Cramér-Rao bound $\mathbf{B}_{\text{CR}}$ in Eqs. (62)–(63), and the estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ is not efficient since its does not saturate the Cramér-Rao bound.

In the special case $z = \mathbf{1}_M$, the bias $\boldsymbol{b}$ vanishes, the correlation of Eq. (61) reduces to

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{+\top}, \tag{64}$$

while the Cramér-Rao bound of Eq. (63) becomes

$$\mathbf{B}_{\text{CR}} = \big(\mathbf{P}^{\top}\,\mathbf{diag}[1/\lambda_j]\mathbf{P}\big)^{+} = \mathbf{F}^{+}. \tag{65}$$

Since Eq. (36) no longer holds in general for the overdetermined system with $M > N$, it results that the right-hand side of Eq. (64) generally differs from the Cramér-Rao bound $\mathbf{B}_{\text{CR}}$ in Eq. (65), and the unbiased estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ is not efficient either, since its does not saturate the Cramér-Rao bound.

For any fixed z, replacing $\boldsymbol{b}$ from Eq. (46) in Eq. (61) leads for the estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ to

$$\mathbf{C}(\widehat{\boldsymbol{\theta}}) = \mathbf{P}^{+}\mathbf{A_1}\mathbf{P}^{+\top}, \tag{66}$$

with the $M \times M$ symmetric matrix

$$\mathbf{A_1} = (\boldsymbol{\lambda}./z - \boldsymbol{\lambda})(\boldsymbol{\lambda}./z - \boldsymbol{\lambda})^{\top} + \mathbf{diag}[\lambda_j/z_j]. \tag{67}$$

At $z = \mathbf{1}_M$, Eq. (66) reduces to Eq. (64) when the estimator is $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$. An interesting issue is to examine the feasibility of exploiting $z \neq \mathbf{1}_M$ solving Eq. (13) in order to obtain a biased estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./z - \boldsymbol{r})$ that would improve over the unbiased estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ at $z = \mathbf{1}_M$. This will be characterized by a correlation matrix in Eq. (66) at $z \neq \mathbf{1}_M$ smaller than that of Eq. (64) at $z = \mathbf{1}_M$, i.e.

$$\mathbf{P}^{+}\mathbf{A_1}\mathbf{P}^{+\top} < \mathbf{P}^{+}\,\mathbf{diag}[\lambda_j]\mathbf{P}^{+\top}, \tag{68}$$

or equivalently,

$$\mathbf{P}^{+}\mathbf{A_2}\mathbf{P}^{+\top} < \mathbf{0}_N, \tag{69}$$

with the $M \times M$ symmetric matrix

$$\begin{aligned}\mathbf{A_2} &= \mathbf{A_1} - \mathbf{diag}[\lambda_j] \\ &= (\boldsymbol{\lambda}./z - \boldsymbol{\lambda})(\boldsymbol{\lambda}./z - \boldsymbol{\lambda})^{\top} + \mathbf{diag}[\lambda_j/z_j - \lambda_j].\end{aligned} \tag{70}$$

A requirement less stringent than Eq. (69) is a condition on the trace

$$\text{tr}\big(\mathbf{P}^{+}\mathbf{A_2}\mathbf{P}^{+\top}\big) < 0, \tag{71}$$

which ensures a global improvement of the mean square errors by the estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ at $\boldsymbol{z} \neq \mathbf{1}_{\boldsymbol{M}}$. The condition of Eq. (71) can be satisfied with a matrix $\mathbf{A_2}$ with $\mathrm{tr}(\mathbf{A_2}) < 0$ and furthermore with all its diagonal elements negative, i.e.

$$\left(\frac{1}{z_j} - 1\right)^2 \lambda_j^2 + \left(\frac{1}{z_j} - 1\right)\lambda_j < 0 \quad \text{for } j = 1, 2, \ldots, M. \tag{72}$$

There are usually enough degrees of freedom in the $M \times 1$ vector $\boldsymbol{z}$ to satisfy Eq. (72) for all $j = 1$ to M, and thus ensuring Eq. (71). For $\lambda_j > 1$, Eq. (72) is satisfied when $z_j \in]1, \lambda_j/(\lambda_j - 1)[$, while for $0 < \lambda_j \leq 1$, Eq. (72) is satisfied when $z_j \in]-\infty, \lambda_j/(\lambda_j - 1)[\cup]1, \infty[$. Moreover, for z_j in the allowed range, the left-hand size of Eq. (72) is maximally negative when $z_j = 2\lambda_j/(2\lambda_j - 1)$.

In this way, in the condition of not too small average counts $\lambda_j > 1$ for all $j = 1$ to M, any fixed $z_j \in]1, \lambda_j/(\lambda_j - 1)[$ leads to an improved estimator with a reduced mean-squared error $\langle(\widehat{\theta_i} - \theta_i)^2\rangle$ for each component i of the vector parameter $\boldsymbol{\theta}$. This is a sufficient condition for a negative trace in Eq. (71). Such an estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ at fixed $\boldsymbol{z} \neq \mathbf{1}_{\boldsymbol{M}}$ will usually be biased, but the division of the data y_j by $z_j > 1$ reduces the fluctuations in the estimate, and the overall effect is to reduce the mean-squared error of each component i. Yet, since λ_j is usually unknown, the upper bound $\lambda_j/(\lambda_j - 1)$ for z_j as to be set approximately, as $y_j/(y_j - 1)$ for instance. A fixed z_j picked in such an approximate interval may fail to satisfy Eq. (72); nevertheless, if such failing z_j are not too many, a global improvement may still persist with a negative trace in Eq. (71). This property of an improved estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ is obtained even if the fixed $\boldsymbol{z} \neq \mathbf{1}_{\boldsymbol{M}}$ is not in the subset $\{\mathbf{1}_{\boldsymbol{M}} + \mathrm{Ker}(\mathbf{P}^{\top})\}$. However, this may not provide a substantial practical benefit in general; for large intensity λ_j the interval $]1, \lambda_j/(\lambda_j - 1)[$ of beneficial z_j shrinks to 1; the improved biased estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ goes to the simpler unbiased $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ with a benefit gradually vanishing. One can expect a more uniform benefit with the exact ML estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z}(\boldsymbol{y}) - \boldsymbol{r})$ with $\boldsymbol{z}(\boldsymbol{y})$ solving Eqs. (13) and (15), but whose performance resists analytical characterization because of the nonlinearity introduced by $\boldsymbol{z}(\boldsymbol{y})$, which again may be seen as a manifestation of the intrinsic nonlinear character of the ML equation (8).

6 Summary and Discussion

6.1 Summary

For image reconstruction in photon emission tomography, we have shown that the nonlinear ML equation of (8) can be replaced by the two dual linear equations (13) and (14) nonlinearly coupled through the auxiliary variable $\boldsymbol{z}$. Linear algebra then applied to these two dual linear equations allowed us to obtain an analytical characterization of the forms and conditions of existence of the ML solutions, parameterized by the auxiliary variable $\boldsymbol{z}$, and of their performance in terms of bias, correlation matrix of the error and Cramér-Rao bound. For estimating N unknown sources $\boldsymbol{\theta}$ from M measured data $\boldsymbol{y}$, the properties of the resulting ML estimators can be summarized as follows.

For an underdetermined system with $M < N$, there is usually an infinite number of ML estimators, under the form $\widehat{\boldsymbol{\theta}} = [\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})] + [\widehat{\boldsymbol{\theta}}_{\mathbf{0}} \in \mathrm{Ker}(\mathbf{P})]$. Such an estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ usually has a non-vanishing bias $\boldsymbol{b}(\widetilde{\boldsymbol{\theta}})$ in Eq. (42) and a correlation matrix of the error $\mathbf{C}(\widetilde{\boldsymbol{\theta}})$ in Eq. (57). The measurement part $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) \in \mathrm{Ker}(\mathbf{P})^{\perp}$ from Eq. (23) is the minimum Euclidean-norm ML estimator, but it misses the null part $\boldsymbol{\theta}_{\mathbf{0}}$ if one exists in the true parameter $\boldsymbol{\theta}^{\mathrm{true}}$ from Eq. (20) to be estimated. It is however not straightforward to characterize the null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} \in \mathrm{Ker}(\mathbf{P})$ to usefully contribute to the ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ since the data $\boldsymbol{y}$ are not affected by $\boldsymbol{\theta}_{\mathbf{0}}$ to be estimated. If the EM-ML algorithm as in Eq. (10) is employed in this configuration, it will usually converge to an ML solution that will pick a given $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$, yet with no explicit control upon it and possibly varying with the initialization [42, 44]. A more sensible construction of the null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ may be based on a priori information about the image $\boldsymbol{\theta}^{\mathrm{true}}$ to be estimated. This would benefit from a separate characterization of the null space $\mathrm{Ker}(\mathbf{P})$ attached to the tomographic system, to appreciate under what form the true parameter image $\boldsymbol{\theta}^{\mathrm{true}}$ can contain some component in $\mathrm{Ker}(\mathbf{P})$, for instance in the spirit of the statistical analyses of [2, 9, 31, 47]. From another viewpoint, this relates to the ability of appreciating if a minimum Euclidean-norm property which is conferred to the ML estimator by the choice $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ is a plausible approximation to approach the true image $\boldsymbol{\theta}^{\mathrm{true}}$. A minimum Euclidean-norm solution out of the subset $\{\mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \mathrm{Ker}(\mathbf{P})\}$ may not necessarily match the true image $\boldsymbol{\theta}^{\mathrm{true}}$ which is a priori not bound to be of minimum ℓ_2 norm in this subset. A possibly more suitable condition could be to characterize the true image $\boldsymbol{\theta}^{\mathrm{true}}$ as verifying a condition of minimum ℓ_0 pseudonorm relative to some basis of $\mathbb{R}^{N\times 1}$. This represents a condition of sparsity on the true image $\boldsymbol{\theta}^{\mathrm{true}}$ which often suits natural images in some common basis like Fourier or cosine or wavelet bases. The approach for estimation would then be to construct the null part $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ so as to obtain an ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ as the minimum ℓ_0-pseudonorm solution out of the subset $\{\mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r}) + \mathrm{Ker}(\mathbf{P})\}$ or sparsest solution relative to some appropriate basis of $\mathbb{R}^{N\times 1}$. This is likely to produce a data-dependent $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} = \widehat{\boldsymbol{\theta}}_{\mathbf{0}}(\boldsymbol{y})$ to match the data-dependent part $\widehat{\boldsymbol{\theta}}_{\mathbf{1}}(\boldsymbol{y}) = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ and yield an ML estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$

sparse in the specified basis. Such approach can benefit from the techniques of minimum-norm reconstruction of sparse signals recently developed for Poisson data especially for compressive sensing [13, 18, 30, 32, 39]. Strategies to reconstruct images with a number N of voxels exceeding the number M of detectors are especially important for high-resolution imaging [25, 36]. This requires to combine the M measurements, with prior knowledge on the N-voxel reconstructed image to control its null part. With a given tomographic scanner, the reconstruction size N can even in principle be progressively increased above M, provided the corresponding $M \times N$ system matrix $\mathbf{P}$ is accessible at each step, while controlling with some relevant metric the quality of the reconstructed image. In this respect, the Fisher information as described in Sect. 5.1, can provide a basis for such metrics of image quality [3, 10]. In this way, maximum benefit can be sought from a given scanner with M detectors for high-resolution imaging at $N > M$.

For an overdetermined system with $M > N$, there is usually one ML estimator of the form $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ from Eq. (24), for every $\boldsymbol{z}$ solving the nonlinear system of Eqs. (13) and (15), or equivalently for any $\boldsymbol{z}$ from the subset $\{\mathbf{1}_{\boldsymbol{M}} + \mathrm{Ker}(\mathbf{P}^{\top})\}$ and solving Eq. (15). Such an estimator has a bias $\boldsymbol{b}(\widetilde{\boldsymbol{\theta}})$ given in Eq. (44) and a correlation matrix of the error $\mathbf{C}(\widetilde{\boldsymbol{\theta}})$ in Eq. (60). We argued in Sect. 4.4 that it can be commonly expected one unique solution $\boldsymbol{z} \neq \mathbf{1}_{\boldsymbol{M}}$ to the system of Eqs. (13) and (15), and therefore one unique ML estimator from Eq. (24). The resolution of the system of Eqs. (13) and (15) is the difficult part of the estimation as it involves a nonlinearity. If the EM-ML algorithm as in Eq. (10) is employed in this configuration, it will in principle converge to an ML estimator $\widehat{\boldsymbol{\theta}}$ implicitly incorporating a solution to Eqs. (13) and (15) for $\boldsymbol{z}$, yet without knowing or relying on the general analytical form of Eq. (24) which is generic for the ML estimator. Our approach makes this analytical form explicit, which is useful to a characterization of the ML estimator. Specific methods can be devised for the effective resolution of the nonlinear system of Eqs. (13) and (15), possibly iterative like the EM-ML algorithm which is an indirect resolution thereof.

Finally, for a uniquely determined system with $M = N$, there is a unique ML estimator $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ which is unbiased and efficient, with a correlation matrix of the error $\mathbf{C}(\widetilde{\boldsymbol{\theta}})$ in Eq. (59) matching the Cramér-Rao bound $\mathbf{B}_{\mathrm{CR}}$ of Eq. (29).

6.2 Positivity Constraint

The present study analyzes the solutions to the plain ML approach which is an essential reference for estimation, with a well-defined status. Accordingly, we did not impose here any positivity constraint in the mathematical derivation of the ML solutions. However, physical reasons dictate positivity to the image estimator $\widehat{\boldsymbol{\theta}}$, i.e. $\widehat{\theta}_k \geq 0, \forall k = 1, \ldots, N$. With the positivity constraint, the ML problem is more difficult to solve analytically, and the general theoretical analysis developed here for the unconstrained ML problem no longer applies as such. The direct path to maximize the likelihood of Eq. (5) under the constraint of a positive $\boldsymbol{\theta}$, is to use the derivatives as in Eq. (7) and involve them into the Kuhn-Tucker conditions [42, 44, 48]. This constrained maximization is difficult to realize analytically and precludes general theoretical analyzability of the solution. However, as an alternative, some indirect paths are possible to obtain a positive solution to the unconstrained maximization of the ML equation (8). In practice, the EM-ML algorithm which is very often used for unconstrained maximization, when initialized on a positive estimate in Eq. (10), is guaranteed to converge to a positive ML solution, provided one exists [42, 44]. The existence of positive solutions to the ML equation (8) is conditioned both by the system matrix $\mathbf{P}$ and by the data $\boldsymbol{y}$, and is difficult to a priori prove in general conditions. When running EM-ML, the existence of a positive solution is usually empirically assumed and well verified in practice. Theoretically, the solutions to the unconstrained ML equation we have described here, can further be involved in a post-processing argument to a posteriori reconsider the positivity constraint on the ML solution, in the following way.

For an underdetermined system, the unconstrained ML solution is of the form $\widehat{\boldsymbol{\theta}} = [\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})] + [\widehat{\boldsymbol{\theta}}_{\mathbf{0}} \in \mathrm{Ker}(\mathbf{P})]$. If $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$, the unconstrained ML solution in the measurement space, is positive, then any $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in $\mathrm{Ker}(\mathbf{P})$ that preserves the positivity of $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ leads to a valid ML solution satisfying the positivity constraint and reaching the same (absolute) maximum of the likelihood. These solutions are completely equivalent to the solutions to the positively-constrained ML problem that would have resulted from the direct resolution through the Kuhn-Tucker conditions. In this way, from the set of solutions to the unconstrained ML problem of Eq. (8), we obtain as a subset all the solutions to the positively-constrained ML problem. From a positive unconstrained ML solution $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$, any vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in $\mathrm{Ker}(\mathbf{P})$ leading to a positive $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ is, in finite N dimension, the analogue of a null function of the infinite-dimensional object space as considered in [9]. Such null functions of [9] form a subset of the null space of the imaging system. Bounds have been determined in [9] for such null functions, which can possibly be transposed to vectors in finite dimension, and serve to assess the extent of the solution set of the positively constrained ML problem, when the measurement part $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ is positive. However, there is a priori no necessity for the measurement part $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ to be positive; in the same way there is no necessity for the measurement part $\boldsymbol{\theta}_{\mathbf{1}}$ of the true parameter $\boldsymbol{\theta}^{\mathrm{true}}$ in Eq. (20) to be positive; the positivity constraint applies to the complete vectors $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{\mathrm{true}}$. Therefore, in the case where $\widehat{\boldsymbol{\theta}}_{\mathbf{1}} = \mathbf{P}^{+}(\boldsymbol{y} - \boldsymbol{r})$ is not positive, then

some $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in Ker($\mathbf{P}$) may be able to realize a positive solution $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$. A priori, such a $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in Ker($\mathbf{P}$) is not the analogue of a null function of [9], and the bounds of [9] do not apply to it. The resulting $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ realizes a valid ML solution satisfying the positivity constraint and reaching the same (absolute) maximum of the likelihood as the unconstrained ML solution. In the same way, the set of positive solutions so obtained is completely equivalent to the set of solutions to the positively-constrained ML problem that would have resulted from the direct resolution through the Kuhn-Tucker conditions. Here also, from the set of solutions to the unconstrained ML problem of Eq. (8), we obtain as a subset all the solutions to the positively-constrained ML problem. On the contrary, with a non positive $\widehat{\boldsymbol{\theta}}_{\mathbf{1}}$, if there is no $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ in Ker($\mathbf{P}$) able to realize a positive $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathbf{1}} + \widehat{\boldsymbol{\theta}}_{\mathbf{0}}$, then the set of solutions to the unconstrained ML problem contains no solution to the positively-constrained ML problem. The solutions to the positively-constrained ML problem will follow from the direct approach via the Kuhn-Tucker conditions. These constrained solutions are not solutions to the unconstrained ML problems, and they achieve a smaller value for the maximum likelihood compared to the unconstrained ML solutions that still achieve the absolute maximum of the likelihood.

The regime of the underdetermined system where the positively-constrained ML solutions form a subset of the unconstrained ML solutions especially holds for the data realization $\boldsymbol{y} = \boldsymbol{\lambda}$ with the positive ML solution $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\text{true}}$ satisfying Eq. (8), and this regime can be expected to persist for weakly noisy data when $\boldsymbol{y}$ does not depart much from $\boldsymbol{\lambda}$. Also in such a regime, for a positive solution $\widehat{\boldsymbol{\theta}}$ satisfying the ML equation (8), a vector $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ can be taken from the null space Ker($\mathbf{P}$), and since Ker($\mathbf{P}$) is a vector space, for any scalar ε the vector $\widehat{\boldsymbol{\theta}} + \varepsilon\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ satisfies the ML equation (8). For small enough nonzero ε one would expect $\widehat{\boldsymbol{\theta}} + \varepsilon\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$ to remain positive, i.e. to form another positive ML solution. It can then be reasonably expected rather commonly, non-unique positive ML solutions in the underdetermined case when Ker($\mathbf{P}$) $\supset \{\mathbf{0}_N\}$. The EM-ML algorithm, when initialized on a positive estimate, will converge to one of the positive ML solutions, possibly dependent on the initialization [42, 44]. It is an interesting issue to investigate which solution will be picked by EM-ML, yet this demands an analysis specific to this algorithm and its initialization.

For an overdetermined system, the unconstrained ML problem is expected to have at least one solution, displaying the form $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$, defined by one $\boldsymbol{z}$ existing from the subset $\{\mathbf{1}_M + \text{Ker}(\mathbf{P}^{\top})\}$ and solving Eq. (15). The positivity constraint on $\widehat{\boldsymbol{\theta}}$ does not translate into a simpler constraint decidable on $\boldsymbol{z}$. If the unconstrained ML solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ turns out to be positive, then it also stands as a solution to the positively-constrained ML problem. On the contrary, if this solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ is not positive, then the solutions to the positively-constrained ML problem are distinct, they follow through the direct approach via the Kuhn-Tucker conditions, and achieve a smaller value of the maximum likelihood compared to the unconstrained ML solution. These constrained ML solutions in particular are not bound to the form $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{+}(\boldsymbol{y}./\boldsymbol{z} - \boldsymbol{r})$ since they generally do not solve Eq. (8).

Finally, for a uniquely determined system, there is a unique solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ to the unconstrained ML problem. If this solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ is positive, then it also stands as the unique solution to the positively-constrained ML problem. On the contrary, if this solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ is not positive, then the solutions to the positively-constrained ML problem are distinct, they follow through the direct approach via the Kuhn-Tucker conditions, and achieve a smaller value of the maximum likelihood compared to the unconstrained ML solution. As we already mentioned, for a given system matrix $\mathbf{P}$, it is not a priori easy to theoretically characterize the data $\boldsymbol{y}$ leading to a positive $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$; and one is relegated to direct calculation and checking of $\widehat{\boldsymbol{\theta}}$ for each given $\boldsymbol{y}$. However, it can be affirmed that, on average, $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$ is indeed positive, since from unbiasedness, $\langle\widehat{\boldsymbol{\theta}}\rangle = \langle\mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})\rangle = \boldsymbol{\theta}^{\text{true}} \geq \mathbf{0}_N$. This ensures at least a reasonable amount of realizations of the data $\boldsymbol{y}$ associated with a positive solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$. Based on the Poisson statistics of the data $\boldsymbol{y}$, it may be possible to develop further, at least in probabilistic terms, the characterization of the set of data $\boldsymbol{y}$ leading to a positive solution $\widehat{\boldsymbol{\theta}} = \mathbf{P}^{-1}(\boldsymbol{y} - \boldsymbol{r})$, for a given invertible matrix $\mathbf{P}$, and even when $\mathbf{P}$ gradually departs from invertibility toward an underdetermined or an overdetermined system. This could provide ground to the empirical observation that in practice a positive solution to the unconstrained ML problem often exists and can be found by EM-ML started from a positive initial estimate.

6.3 Convergence of EM-ML

The essential contribution of the present study is, for the ML method which is an essential reference for estimation, to develop an original approach for analyzing the ML solutions and their theoretical properties. Accordingly, we did not put forward any specific algorithm for practical calculation of the ML solution. Any practical algorithm, iterative algorithms like EM-ML, direct matrix inversion or computation of the pseudoinverse (when practically tractable), are equally concerned by our results. As we mentioned in the Introduction, practical iterative algorithms like EM-ML are often stopped before convergence to a true ML solution characterized here [11, 45, 46]. This is usually intended as a regularization (smoothing) procedure, with the effect of biasing the reconstructed image toward its initial estimate, often controlled by empirical stopping rules. The resulting solutions however, do not own a clear theoretical status or optimality property, which hinders their mathematical analysis.

In this respect, our results, which provide a characterization of the exact ML solutions, can serve to appreciate the situation of an empirical solution stopped before convergence, in relation to the two extreme references formed by the chosen initial estimate and the optimal ML solution analyzed here. Some works however can iterate EM-ML up to convergence [11, 42, 44]. There also exist specific situations where EM-ML is iterated up to convergence. This is the case when specially designed smooth basis functions are used instead of voxels for the reconstruction [12, 27]. This is also the case when post-processing is implemented after convergence on the ML solution for noise smoothing or for incorporating extra available information [34, 35]. Such post-processing of the ML solution offers potentialities for improved signal-to-noise ratio and spatial resolution and flexibility in the filtering operations [34], yet usually at increased computational cost, but can become more competitive as computer power continues to grow. Such approaches are directly concerned by the characterization of the ML solutions accomplished here.

6.4 Final Remarks

As discussed, the theoretical characterization of the ML solutions can be useful to the practically extensively used EM-ML algorithm. Knowing the precise structure of the ML solution set as characterized here, can help to better understand the behavior of the EM-ML algorithm which is often empirically exploited, as a function of its initialization, in the presence of positivity constraint, the solution it picks, and where it stands when stopped before convergence. Also the expressions of the ML estimators derived here, strongly rely on the pseudoinverse matrix $\mathbf{P}^+$. In practice, this is usually a very large matrix. However, such a matrix $\mathbf{P}^+$ is a fixed matrix characterizing a given tomographic imaging system, and it can then be computed once and for all and attached to the imaging system. Meanwhile, the EM-ML algorithm as in Eq. (10) as to be iterated anew for each new data $\boldsymbol{y}$. ML approaches, post-processed and regularized ML approaches, therefore still hold potential of improvement for photon emission tomography and image reconstruction.

References

1. Barrett, H.H., Myers, K.J.: Foundations of Image Science. Wiley, New York (2004)
2. Barrett, H.H., Wilson, D.W., Tsui, B.M.W.: Noise properties of the EM algorithm: I. Theory. Phys. Med. Biol. **39**, 833–846 (1994)
3. Barrett, H.H., Denny, J.L., Wagner, R.F., Myers, K.J.: Objective assessment of image quality. II. Fisher information. Fourier crosstalk, and figures of merit for task performance. J. Opt. Soc. Am. A **12**, 834–852 (1995)
4. Barrett, H.H., Hunter, W.C.J., Miller, B.W., Moore, S.K., Chen, Y., Furenlid, L.R.: Maximum-likelihood methods for processing signals from gamma-ray detectors. IEEE Trans. Nucl. Sci. **56**, 725–735 (2009)
5. Bovik, A.C.: Handbook of Image and Video Processing. Academic Press, San Diego (2000)
6. Carroll, R.B., Mair, B.A.: A new model and reconstruction method for 2D PET based on transforming detector tube data into detector arc data. J. Math. Imaging Vis. **14**, 165–185 (2001)
7. Chen, Y., Ma, J., Feng, Q., Luo, L., Shi, P., Chen, W.: Nonlocal prior Bayesian tomographic reconstruction. J. Math. Imaging Vis. **30**, 133–146 (2008)
8. Chornoboy, E.S., Chen, C.J., Miller, M.I., Miller, T.R., Snyder, D.L.: An evaluation of maximum likelihood reconstruction for SPECT. IEEE Trans. Med. Imaging **9**, 99–110 (1990)
9. Clarkson, E., Barrettt, H.: Bounds on null functions of linear digital imaging systems. J. Opt. Soc. Am. A **15**, 1355–1360 (2007)
10. Clarkson, E., Shen, F.: Fisher information and surrogate figures of merit for the task-based assessment of image quality. J. Opt. Soc. Am. A **27**, 2313–2326 (2010)
11. Cloquet, C., Defrise, M.: MLEM and OSEM deviate from the Cramer-Rao bound at low counts. IEEE Trans. Nucl. Sci. **60**, 134–143 (2013)
12. Daube-Witherspoon, M.E., Matej, S., Karp, J.S., Lewitt, R.M.: Application of the row action maximum likelihood algorithm with spherical basis functions to clinical PET imaging. IEEE Trans. Nucl. Sci. **48**, 24–30 (2001)
13. Dupé, F.X., Fadili, J.M., Starck, J.L.: A proximal iteration for deconvolving Poisson noisy images using sparse representations. IEEE Trans. Image Process. **18**, 310–321 (2009)
14. Fessler, J.A.: Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. IEEE Trans. Image Process. **5**, 493–506 (1996)
15. Fessler, J.A., Hero, A.O.: Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. IEEE Trans. Image Process. **4**, 1417–1429 (1995)
16. Fessler, J.A., Rogers, W.L.: Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. IEEE Trans. Image Process. **5**, 1346–1358 (1996)
17. Green, P.J.: Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Trans. Med. Imaging **9**, 84–93 (1990)
18. Harmany, Z.T., Marcia, R.F., Willett, R.M.: Sparsity-regularized photon-limited imaging. In: Proceedings 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, The Netherlands, 14–17 April 2010, pp. 772–775 (2010)
19. Hebert, T.I., Leahy, R.: A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. IEEE Trans. Med. Imaging **8**, 194–202 (1989)
20. Hero, A.O., Fessler, J.A.: A recursive algorithm for computing Cramer-Rao-type bounds on estimator covariance. IEEE Trans. Inf. Theory **40**, 1205–1210 (1994)
21. Hudson, H.M., Larkin, R.S.: Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imaging **13**, 601–609 (1994)
22. Lange, K., Carson, R.: EM reconstruction algorithms for emission and transmission tomography. J. Comput. Assist. Tomogr. **8**, 306–316 (1984)
23. Lange, K., Bahn, M., Little, R.: A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. IEEE Trans. Med. Imaging **6**, 106–114 (1987)
24. Le, T., Chartrand, R., Asaki, T.J.: A variational approach to reconstructing images corrupted by Poisson noise. J. Math. Imaging Vis. **27**, 257–263 (2007)
25. Le Meunier, L., Slomka, P.J., Dey, D., Ramesh, A., Thomson, L.E., Hayes, S.W., Friedman, J.D., Cheng, V., Germano, G., Berman, D.S.: Enhanced definition PET for cardiac imaging. J. Nucl. Cardiol. **17**, 414–426 (2010)
26. Levitan, E., Herman, C.: A maximum a posteriori expectation maximization algorithm for image reconstruction in emission tomography. IEEE Trans. Med. Imaging **6**, 185–192 (1987)

27. Lewitt, R.M.: Alternatives to voxels for image representation in iterative reconstruction algorithms. Phys. Med. Biol. **37**, 705–716 (1992)
28. Li, Q., Asma, E., Qi, J., Bading, J.R., Leahy, R.M.: Accurate estimation of the Fisher information matrix for the PET image reconstruction problem. IEEE Trans. Med. Imaging **23**, 1057–1064 (2004)
29. Liang, Z., Hart, H.: Bayesian reconstruction in emission computerized tomography. IEEE Trans. Nucl. Sci. **35**, 788–792 (1988)
30. Lingenfelter, D.J., Fessler, J.A., He, Z.: Sparsity regularization for image reconstruction with Poisson data. In: Proceedings SPIE, Computational Imaging VII, San Jose, CA, vol. 7246, p. 72460F (2009)
31. Llacer, J., Veklerov, E., Coakley, K.J., Hoffman, E.J., Nunez, J.: Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies. IEEE Trans. Med. Imaging **12**, 215–231 (1993)
32. Luo, J., Liu, J., Li, W., Zhu, Y., Jiang, R.: Image reconstruction from sparse projections using S-transform. J. Math. Imaging Vis. **43**, 227–239 (2012)
33. Novak, R.D., Kolaczyk, E.D.: A statistical multiscale framework for Poisson inverse problems. IEEE Trans. Inf. Theory **46**, 1811–1825 (2000)
34. Nuyts, J., Fessler, J.A.: A penalized-likelihood image reconstruction method for emission tomography, compared to post-smoothed maximum-likelihood with matched spatial resolution. IEEE Trans. Med. Imaging **22**, 1042–1052 (2003)
35. Nuyts, J., Baete, K., Bequé, D., Dupont, P.: Comparison between MAP and post-processed ML for image reconstruction in emission tomography when anatomical knowledge is available. IEEE Trans. Med. Imaging **24**, 667–675 (2005)
36. Panin, V.Y., Kehren, F., Michel, C., Casey, M.: Fully 3-D PET reconstruction with system matrix derived from point source measurements. IEEE Trans. Med. Imaging **25**, 907–921 (2006)
37. Pearson, C.E.: Handbook of Applied Mathematics. Van Nostrand, New York (1974)
38. Qi, J., Leahy, R.M.: A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. IEEE Trans. Med. Imaging **18**, 293–305 (1999)
39. Raginsky, M., Willett, R.M., Harmany, Z.T., Marcia, R.F.: Compressed sensing performance bounds under Poisson noise. IEEE Trans. Signal Process. **58**, 3990–4002 (2010)
40. Randarajan, A., Hsiao, I.T., Gindi, G.: A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. J. Math. Imaging Vis. **12**, 199–217 (2000)
41. Rockmore, A.J., Macovski, A.: A maximum likelihood approach to emission image reconstruction from projections. IEEE Trans. Nucl. Sci. **23**, 1428–1432 (1976)
42. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. IEEE Trans. Med. Imaging **1**, 113–122 (1982)
43. Stayman, J.W., Fessler, J.A.: Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. IEEE Trans. Med. Imaging **19**, 601–615 (2000)
44. Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. J. Am. Stat. Assoc. **80**, 8–20 (1985)
45. Veklerov, E., Llacer, J.: Stopping rule for the EM algorithm based on statistical hypothesis testing. IEEE Trans. Med. Imaging **6**, 313–319 (1988)
46. Veklerov, E., Llacer, J., Hoffman, E.J.: MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule. IEEE Trans. Nucl. Sci. **35**, 603–607 (1988)
47. Wilson, D.W., Tsui, B.M.W., Barrett, H.H.: Noise properties of the EM algorithm: II. Monte Carlo. Phys. Med. Biol. **39**, 847–871 (1994)
48. Zangwill, W.I.: Nonlinear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs (1969)

François Chapeau-Blondeau was born in France in 1959. He received the Engineer Diploma from ESEO, Angers, France, in 1982, the Ph.D. degree in electrical engineering from University Pierre et Marie Curie, Paris 6, France, in 1987, and the *Habilitation* degree from the University of Angers, France, in 1994. In 1988, he was a research associate in the Department of Biophysics at the Mayo Clinic, Rochester, Minnesota, USA, working on biomedical ultrasonics. Since 1990, he has been with the University of Angers, France, where he is currently a professor of electronic and information sciences. His research interests include signal processing and imaging, and the interactions between physics and information sciences.

Christian Jeanguillaume was born in France in 1951. He graduated in Medicine in 1979 (M.D.) specialized in nuclear medicine in 1982 and he received the doctorat d'État es Sciences Physiques in 1989 (Ph.D.). He first worked in the nuclear medicine department in CHU of Creteil France, and was a researcher in the Laboratoire de physique des solides (CNRS, UMR8502, Orsay, France). He is currently Maître de Conférences at the University of Angers, and Praticien Hospitalier at the CHU of Angers. His research interests are: emission tomography, new methods of gamma ray imaging with large hole collimator, inverse problems, parathyroid and brain radionuclide imaging, electron energy loss spectroscopy in the electron microscope (EELS, STEM).