Information theoretic analysis of hyperspectral imaging systems with applications to fluorescence microscopy

Sripad Ram

Contact information:

Global Pathology,
Drug Safety Research and Development,
Pfizer Inc.,
10646 Science Center Drive,
San Diego, CA 92121 USA.

Email: sripad.ram@pfizer.com.
Tel: 1-858-622-5904
Fax: 1-858-678-8124

Abstract

We present a general stochastic model for hyperspectral imaging data and derive analytical expressions for the Fisher information matrix for the underlying spectral unmixing problem. We investigate the linear mixing model as a special case and define a linear unmixing performance bound by using the Cramer-Rao inequality. As an application, we consider fluorescence imaging and show how the performance bound provides a spectral resolution limit that predicts how accurately a pair of spectrally similar fluorescent labels can be spectrally unmixed. We also report a novel result that shows how the spectral resolution limit can be overcome by exploiting the phenomenon of anti-Stokes shift fluorescence. In addition, we investigate how photon statistics, channel addition and channel splitting affect the performance bound. Finally by using the performance bound as a benchmark, we compare the performance of the least squares and the maximum likelihood estimators for spectral unmixing. For the imaging conditions tested here, our analysis shows that both estimators are unbiased and that the standard deviation of the maximum likelihood estimator is consistently closer to the performance bound than that of the least squares estimator. The results presented here are based on broad assumptions regarding the underlying data model and are applicable to hyperspectral data acquired with point detectors, sCMOS, CCD and EMCCD imaging detectors.

EDICS: ELI-COL, COI-MCI.

Keywords: Spectral unmixing, Multispectral imaging, Multiplex imaging, Cramer-Rao lower bound.

I. INTRODUCTION

Hyperspectral imaging represents a broad class of techniques that capture spectral and spatial data from the object of interest. Fluorescence microscopy is a powerful tool to study microscopic objects such as biological cells at high spatial and temporal resolution. Fluorescence imaging supports the visualization of multiple targets within the object of interest, which is facilitated by labeling the targets with fluorescent labels that have distinct excitation and emission spectra (Fig. 1A). Consequently, the need for hyperspectral imaging arises in such applications when the emission spectra of the fluorescent labels significantly overlap (Fig. 1B). There exist a variety of approaches to carry out hyperspectral imaging in a fluorescence microscope. For instance, hyperspectral imaging can be carried out either with a confocal ([1], [2], [3]) or a linescanning ([4]) microscope that has spectral detection capability. Alternately, it can also be carried out on a widefield microscope by using either narrowband excitation and emission filters ([5], [6], [7], [8]) or by using an electronically controlled liquid crystal tunable filter ([9], [10]). Common to all these techniques is the underlying hyperspectral data which consists of a sequence of 2D images acquired at different spectral windows. Thus given the hyperspectral data, the goal is then to estimate the relative abundance (typically represented in photon counts) of the different labels at each pixel, and this is referred to as the spectral unmixing problem (Fig. 1C).
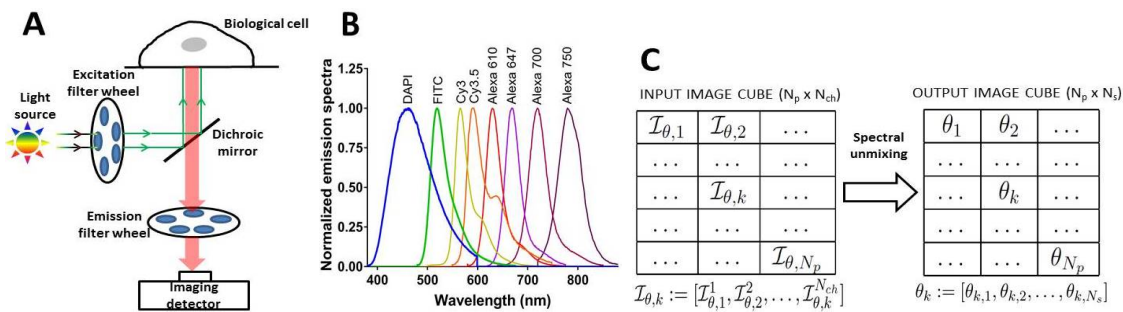

Fig. 1. Hyperspectral imaging in fluorescence microscopy. Panel A shows a typical configuration of a fluorescence microscope for multicolor imaging application. The sample is selectively illuminated at distinct excitation passbands (excitation filter wheel) and the fluorescence signal is then collected at matched emission passbands (emission filter wheel) on an imaging detector. Panel B shows the emission spectra of fluorescent labels exhibiting significant spectral overlap. Panel C illustrates the spectral unmixing problem where given a hyperspectral data set, i.e. the input image cube, which consists of N_{ch} spectral images each with N_p pixels, the goal is to obtain an estimate of the output image cube that represents the relative abundance θ_k of N_s different fluorescent labels at each pixel in the sample. Here $\mathcal{I}_{\theta,k}$ is a random vector with probability density $p_{\theta,k}$ that is a function of $\nu_{\theta,k}$, which describes the signal at the kth pixel in the input image cube.

A common approach to solving the spectral unmixing problem assumes that the number of labels present in the sample is known. Spectral unmixing approaches that make use of this strategy include the least squares estimator ([11], [12], [13]), the maximum likelihood estimator ([14], [15]), and phasor based approaches ([3], [16]). In fluorescence imaging applications, we typically assume a linear mixing model i.e., $\mu_{\theta,k} = A\theta_k$, where $\mu_{\theta,k}$, which describes the signal detected at the kth pixel in the input image cube (see Fig. 1C), is a linear combination of the relative abundance θ_k of the different fluorescent labels present at that pixel in the sample. Here A denotes the mixing matrix that specifies the relative contribution of

each fluorescent label in every spectral channel. The mixing matrix depends on the optical configuration of the hypserspectal microscope and the spectral properties of the fluorescent dye. The mixing matrix can be computed either theoretically or experimentally by performing a calibration experiment with single color control samples.

A fundamental question that arises in hyperspectral imaging applications concerns with its performance limits which deals with the best possible accuracy with which the relative abundance of the different labels can be determined. Knowledge of the performance limit is important as it provides a metric to design and optimize a hyperspectral imaging system. Moreover, it also acts a benchmark to compare the performance of different spectral unmixing algorithms for a given hyperspectral dataset. In this paper, we present results to compute the performance limits of a hyperspectral imaging system. We adopt a stochastic framework to model the acquired data and derive the Fisher information matrix for the spectral unmixing problem. We then consider the linear mixing model as a special case. Using the Cramer-Rao inequality ([17]), we define a linear unmixing performance bound which provides a lower bound to the accuracy with which the relative abundance of the different labels can be estimated.

In the past, several groups have investigated the spectral unmixing problem. In [18], [19], the authors calculated the Fisher information matrix for a deterministic data model that is corrupted by white noise (Gaussian data model). In [15], [20], the authors derived the Fisher information matrix for a data model that is corrupted by shot noise (Poisson data model). The present manuscript provides a broad stochastic framework that is applicable to several different data models that are typically used to describe data from an imaging detector. Specifically, in addition to the above data models our results are also applicable for the Poisson + Gaussian data model and for the stochastic Poisson + Gaussian data model which are typically used to describe imaging data acquired from sCMOS/CCD cameras and EMCCD cameras, respectively.

The paper is organized as follow. In Section II we present the problem formulation, introduce the different data models, and derive the general expression of the Fisher information matrix for the spectral unmixing problem. We then investigate conditions under which the Fisher information matrix is block diagonal as block diagonality has several implications. Next, we introduce the linear mixing model and derive the Fisher information matrix for the same. We also derive analytical expressions of the Fisher information matrix that pertain to the different data models. Further we derive results to address questions regarding the design of hyperspectral imaging systems such as the addition or splitting of spectral channels and its impact on spectral unmixing.

In Section IV, we illustrate the results derived in Section II by considering different hyperspectral imaging configurations and fluorescent label pairs. We show that for a pair of regular fluorescent labels there exists is a spectral resolution limit, which predicts how spectrally close the emission spectra of the two labels can be and still be accurately spectrally resolved. We also show that by exploiting the

phenomenon of anti-Stokes shift fluorescence the spectral resolution limit can be overcome and that the emission spectra of the fluorescent labels can be arbitrarily close to each other. We also illustrate how photon statistics, and the addition or splitting of spectral channels can affect the performance bound for different data models. In Section V, we compare the performance of two spectral unmixing algorithms, namely the least squares and maximum likelihood estimators, on hyperspectral data. Our results show that for the imaging conditions tested here, both estimators are unbiased and their performance come close to the theoretical performance bound with the maximum likelihood estimator being consistently closer to the performance bound than the least squares estimator.

II. Theory

A. Problem formulation

We consider a generic model of a hyperspectral imaging system wherein the sample is illuminated by a light source and the light that is transmitted or emitted by the sample is collected into spectral channels and recorded as a sequence of 2D spectral images, which we refer to as the input image cube. We assume that the size of every spectral image is the same. Let N_p denote the total number of pixels in a spectral image and N_{ch} denote the number of spectral images. We define a pixel block as a sequence of N_{ch} pixels with the same pixel index, say k for $k = 1, \ldots, N_p$, across N_{ch} different spectral images. We assume that the object of interest contains N_s distinct but spectrally overlapping labels. Due to the presence of multiple labels, the signal at the kth pixel block can be described as a superposition of the relative abundance or photons detected from the individual labels that are present at that pixel in the sample. The spectral unmixing problem can then be stated as follows: given an input image cube which is a $N_p \times N_{ch}$ dimensional dataset, the goal is to estimate an output (or spectrally unmixed) image cube θ which is a $N_p \times N_s$ dimensional dataset that contains to the photon counts of N_s labels at each pixel in the sample.

B. Image formation model

Let $\Theta \subseteq \mathbb{R}^{N_p \times N_s}$ be the parameter space and $\theta \in \Theta$ be the unknown parameter vector. The signal detected in the input image cube is modeled as a sequence of random vectors given by

$$\mathcal{I}_\theta = \{\mathcal{I}_{\theta,1}, \mathcal{I}_{\theta,2}, \ldots, \mathcal{I}_{\theta,N_p}\}, \quad \theta \in \Theta, \tag{1}$$

where $\mathcal{I}_{\theta,k} := \{\mathcal{I}_{\theta,k}^j; j = 1, \ldots, N_{ch}\}$ for $k = 1, \ldots, N_p$, and $\mathcal{I}_{\theta,k}^j$ denotes an independent random variable with probability density function $p_{\theta,k}^j$ that models the detected photons at the kth pixel in the jth spectral image. Here we assume that for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$,

A1 $p_{\theta,k}^j$ satisfies the regularity conditions ([17]) required for the calculation of the Fisher information matrix, and

A2 $p_{\theta,k}^j$ depends on θ through a non-negative function $\nu_{\theta,k}^j$ that is differentiable with respect to θ, i.e.,

$$p_{\theta,k}^j \equiv p_{\nu_{\theta,k}^j,k}^j.$$

As we will see, the term $\nu_{\theta,k}^j$ will typically describe the expected number of detected photons at the kth pixel in the jth spectral image for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. We note that $\nu_{\theta,k}^j$ in turn can be expressed as a function of θ which we wish to estimate.

C. Specific data models

We note that the above assumptions regarding the probability density function $p_{\theta,k}^j$ are quite broad. To illustrate how this covers a wide variety of imaging conditions, we consider the following data models.

Gaussian data model: Here we consider the case where the photons detected at the kth pixel in the jth spectral image is modeled as a deterministic signal $\nu_{\theta,k}^j$ that is corrupted by additive Gaussian noise. Hence $\mathcal{I}_{\theta,k}^j$ is given by $\mathcal{I}_{\theta,k}^j = \nu_{\theta,k}^j + W_k^j$ where W_k^j is an independent Gaussian random variable with mean η_k^j and standard deviation σ_k^j for $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. Then the probability density function of $\mathcal{I}_{\theta,k}^j$ is given by

$$p_{\theta,k}^j(z) = \frac{1}{\sqrt{2\pi}\sigma_k^j} \exp\left(-\frac{(z - (\nu_{\theta,k}^j + \eta_k^j))^2}{2(\sigma_k^j)^2}\right), \ z \in \mathbb{R}, \tag{2}$$

$\theta \in \Theta$, $j = 1, \ldots, N_{ch}$, $k = 1, \ldots, N_p$. For the above equation, it is straightforward to verify that assumptions **A1** and **A2** are satisfied.

Poisson data model: Here we assume that the detected photons at each pixel are Poisson distributed. Thus we have $\mathcal{I}_{\theta,k}^j = S_{\theta,k}^j$, where $S_{\theta,k}^j$ is an independent Poisson random variable with mean $\nu_{\theta,k}^j$ for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. he probability density function of $\mathcal{I}_{\theta,k}^j$ is given by

$$p_{\theta,k}^j(z) = \frac{e^{-\nu_{\theta,k}^j}(\nu_{\theta,k}^j)^z}{z!}, \ z = 0, 1, 2, \ldots, \tag{3}$$

$\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ $k = 1, \ldots, N_p$. It immediately follows that assumptions **A1** and **A2** are satisfied. For instance, this specific data model would be applicable for imaging systems such as a confocal or a multiphoton microscope, where the main source of randomness in the data is attributed to shot noise statistics ([21]).

Poisson + Gaussian data model: Here we consider $\mathcal{I}_{\theta,k}^j$ to be given by $\mathcal{I}_{\theta,k}^j = S_{\theta,k}^j + W_k^j$, for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$, and $k = 1, \ldots, N_p$. Here, $S_{\theta,k}^j$ is Poisson random variable with mean $\nu_{\theta,k}^j$ that models the detected photons at the kth pixel in the jth spectral image, and W_k^j is an independent Gaussian random variable with mean η_k^j and standard deviation σ_k^j that models the measurement noise at the kth pixel in the jth spectral image, for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. Here we assume that W_k^j is independent of θ for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. Then the probability density function of $\mathcal{I}_{\theta,k}^j$ is given by ([22])

$$p_{\theta,k}^j(z) = \frac{1}{\sqrt{2\pi}\sigma_k^j} \sum_{l=0}^{\infty} \frac{e^{-\nu_{\theta,k}^j}(\nu_{\theta,k}^j)^l}{l!} e^{-\frac{1}{2}\left(\frac{z-l-\eta_k^j}{\sigma_k^j}\right)^2}, \tag{4}$$

for $z \in \mathbb{R}$, $\theta \in \Theta$, $k = 1, \ldots, N_p$ and $j = 1, \ldots, N_{ch}$. From the above expression it immediately follows that assumption **A2** is satisfied. Moreover, it can also be shown that p_θ satisfies the regularity conditions ([23]) thereby satisfying assumption **A1**. This data model is applicable to a widefield microscope configuration in which the images are acquired with either a CCD or a sCMOS camera. Here, in addition to the shot noise statistics the detected signal is also corrupted by the measurement noise of the detector.

Stochastic-Poisson + Gaussian data model: For completeness, we also consider another data model where in addition to shot noise statistics and measurement noise of the detector, the model takes into account stochastic signal amplification, which, for example, occurs in an electron multiplying CCD camera. Specifically, the detected photons at each pixel that is Poisson distributed with mean $\nu^j_{\theta,k}$ is amplified by a random function **M** that is independent of θ and the amplified signal is further corrupted by additive Gaussian noise ([24]), for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. **M** is typically modeled as a branching process ([25]) with the initial particle count to be Poisson distributed (with mean $\nu^j_{\theta,k}$) and the individual offspring count to be a zero modified geometric distribution. For the above data model, the probability density function can be written as (see [24] for details)

$$
p^j_{\theta,k}(z) = \frac{e^{-\nu^j_{\theta,k}\frac{A}{B}}}{\sqrt{2\pi}\sigma^j_k} \left[e^{-\left(\frac{z-\eta^j_k}{\sqrt{2}\sigma^j_k}\right)^2} + \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta^j_k}{\sqrt{2}\sigma^j_k}\right)^2} \times \right.
$$

$$
\left. \sum_{h=0}^{l-1} \frac{\binom{(l-1)}{h} C^{l-1-h}(D\nu^j_{\theta,k})^{h+1}}{(h+1)! B^{h+l+1}} \right] , \; z \in \mathbb{R}, \tag{5}
$$

where A, B, C and D are constants that are independent of θ for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. From the above equation we see that assumption **A2** is satisfied. Further, it can be shown that the above equation satisfies assumption **A1** ([24]).

D. General expression for the Fisher information matrix

In this section, we derive a general expression of the Fisher information matrix for the output image cube θ using the image formation model described in Section II-B. We also investigate conditions under which the Fisher information matrix $\mathbf{I}(\theta)$ is diagonal, as diagonality of $\mathbf{I}(\theta)$ has several implications. In the present context, since θ is a vector with $N_p \times N_s$ elements, the resulting Fisher information matrix will be a $(N_p \times N_s) \times (N_p \times N_s)$ matrix, which can be very large. Thus a diagonal $\mathbf{I}(\theta)$ can render its calculation to be more tractable. In general for an n-dimensional parameter vector $\theta = (\theta_1, \ldots, \theta_n) \in \Theta$, if $\mathbf{I}(\theta)$ is diagonal, then the limit of the accuracy of θ_i is independent of other components of θ.

Here, the parameter vector is given by

$$
\theta = (\underbrace{\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{1,N_s}}_{\theta_1^T}, \underbrace{\theta_{2,1}, \theta_{2,2} \ldots, \theta_{2,N_s}}_{\theta_2^T}, \ldots,
$$

$$\underbrace{\theta_{N_p,1}, \theta_{N_p,2}, \ldots, \theta_{N_p,N_s}}_{\theta_{N_p}^T}),$$

where $\theta_k = (\theta_{k,1}, \theta_{k,2}, \ldots, \theta_{k,N_s})^T$ denotes the unknown photon counts of the N_s different labels at the kth pixel in the output image cube for $k = 1 \ldots, N_p$. Further we assume that for $j = 1, \ldots, N_{ch}$ and $k \neq m$ and $k, m = 1, \ldots, N_p$, $\frac{\partial \nu_{\theta,k}^j}{\partial \theta_m} = \mathbf{0}$, $\theta \in \Theta$ where $\mathbf{0}$ denotes a $1 \times N_s$ row vector with all elements equal to zero. The validity of this assumption becomes evident in the next section where we will consider an explicit relationship between $\nu_{\theta,k}$ and the components of θ_k for the linear mixing model. As we will see, the above assumption results in the Fisher information matrix being block diagonal, where the diagonal entries pertain to the Fisher information matrix of θ_k for $k = 1, \ldots, N_p$.

In the following theorem, we state two results. The first result is a general expression for the Fisher information matrix pertaining to the image cube $\mathcal{I}_\theta$, which is analogous to a previously published result [24], [26]. The second result investigates conditions for block diagonality of the Fisher information matrix.

Theorem 2.1: Let $\Theta \in \mathbb{R}^{N_p \times N_s}$ denote the parameter space. For $\theta \in \Theta$, let $\mathcal{I}_\theta = \{\mathcal{I}_{\theta,k}^j \mid k = 1, \ldots, N_p, j = 1, \ldots, N_{ch}\}$ denote the input image cube that is defined in eq. 1. Assume that conditions **A1 - A2** are satisfied (see Section II-B).

1. For $\theta \in \Theta$, the Fisher information matrix for the output image cube θ is given by

$$\mathbf{I}(\theta) = \sum_{k=1}^{N_p} \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \left(\frac{\partial \nu_{\theta,k}^j}{\partial \theta} \right)^T \frac{\partial \nu_{\theta,k}^j}{\partial \theta}, \quad \theta \in \Theta,$$

where for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$

$$\alpha_{\theta,k}^j := E \left[\left(\frac{\partial \ln(p_{\theta,k}^j(z_k^j))}{\partial \nu_{\theta,k}^j} \right)^2 \right] \quad (6)$$

and z_k^j denotes a realization of $\mathcal{I}_{\theta,k}^j$.

2. Assume that for $k \neq m$, $k, m = 1, \ldots, N_p$ and $j = 1, \ldots, N_{ch}$, $\frac{\partial \nu_{\theta,k}^j}{\partial \theta_m} = 0$, $\theta \in \Theta$. Then the Fisher information matrix given in result 1 of this Theorem can be written as

$$\mathbf{I}(\theta) = \text{Diag} \left[\mathbf{I}_1(\theta), \mathbf{I}_2(\theta), \ldots, \mathbf{I}_{N_p}(\theta) \right], \quad \theta \in \Theta,$$

where for $\theta \in \Theta$ and $k = 1, \ldots, N_p$

$$\mathbf{I}_k(\theta) := \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \left(\frac{\partial \nu_{\theta,k}^j}{\partial \theta_k} \right)^T \frac{\partial \nu_{\theta,k}^j}{\partial \theta_k}. \quad (7)$$

Proof:

1. See [24], [26] for proof.

8

2. For $\theta \in \Theta$ and $k = 1, \ldots, N_p$, define

$$\mathbf{J}_k(\theta) := \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \left(\frac{\partial \nu_{\theta,k}^j}{\partial \theta} \right)^T \frac{\partial \nu_{\theta,k}^j}{\partial \theta}.$$

Then the Fisher information matrix given in result 1 of this Theorem can be written as

$$\mathbf{I}(\theta) = \sum_{k=1}^{N_p} \mathbf{J}_k(\theta), \quad \theta \in \Theta.$$

For $\theta \in \Theta$, $k = 1, \ldots, N_p$ and $j = 1, \ldots, N_{ch}$, consider the term

$$\frac{\partial \nu_{\theta,k}^j}{\partial \theta} == \left[\mathbf{0}, \mathbf{0}, \ldots, \frac{\partial \nu_{\theta,k}^j}{\partial \theta_k}, \ldots, \mathbf{0} \right],$$

where we have used the fact that for $k \neq m$, $k, m = 1, \ldots, N_p$, $\frac{\partial \nu_{\theta,k}^j}{\partial \theta_m} = \mathbf{0}$ for $\theta \in \Theta$ and $\mathbf{0}$ denotes a $1 \times N_s$ vector with all elements equal to zero. Using the above result, we then have

$$\mathbf{I}(\theta) = \mathbf{J}_1(\theta) + \mathbf{J}_2(\theta) + \cdots + \mathbf{J}_{N_p}(\theta)$$

$$= \sum_{j=1}^{N_{ch}} \alpha_{\theta,1}^j \begin{bmatrix} \left(\frac{\partial \nu_{\theta,1}^j}{\partial \theta_1} \right)^T \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \frac{\partial \nu_{\theta,1}^j}{\partial \theta_1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$$

$$+ \sum_{j=1}^{N_{ch}} \alpha_{\theta,2}^j \begin{bmatrix} \mathbf{0} \\ \left(\frac{\partial \nu_{\theta,2}^j}{\partial \theta_2} \right)^T \\ \vdots \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \frac{\partial \nu_{\theta,2}^j}{\partial \theta_2} & \cdots & \mathbf{0} \end{bmatrix} + \cdots$$

$$+ \sum_{j=1}^{N_{ch}} \alpha_{\theta,2}^j \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \left(\frac{\partial \nu_{\theta,N_p}^j}{\partial \theta_{N_p}} \right)^T \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \frac{\partial \nu_{\theta,N_p}^j}{\partial \theta_{N_p}} \end{bmatrix}$$

$$= \text{Diag} \left[\mathbf{I}_1(\theta), \mathbf{I}_2(\theta), \ldots, \mathbf{I}_{N_p}(\theta) \right], \quad \theta \in \Theta.$$

●

From result 2 of the above theorem we see that the block diagonal representation reduces the computational complexity to calculate the Fisher information matrix. Specifically, for $k = 1, \ldots, N_p$ the kth block matrix $\mathbf{I}_k(\theta)$ is a $N_s \times N_s$ square matrix, which is relatively straightforward to compute.

E. Linear mixing model and the mixing matrix

In the previous theorem we derived a general expression of the Fisher information matrix for the output image cube θ pertaining to a general hyperspectral imaging system. Here we next consider the case when the non negative function $\nu_{\theta,k}^j$ is a linear superposition of the components of the unknown parameter vector θ for $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. For $\theta \in \Theta$, define $\mu_{\theta,k} := [\nu_{\theta,k}^1, \nu_{\theta,k}^2, \ldots, \nu_{\theta,k}^{N_{ch}}]^T$, $k = 1, \ldots, N_p$. Then for the linear mixing model we have

$$\nu_{\theta,k} := A\theta_k, \quad k = 1, \ldots, N_p, \quad \theta \in \Theta, \tag{8}$$

where A is a $N_{ch} \times N_s$ dimensional matrix with elements $\{a_{ij}; i = 1, \ldots, N_{ch}; j = 1, \ldots, N_s\}$ known as the *mixing matrix* and $\theta_k = [\theta_{k,1}, \theta_{k,2}, \ldots, \theta_{k,N_s}]^T$, $k = 1, \ldots, N_p$. The columns of A pertain to the different labels, the rows of A pertain to the different spectral channels, and the element a_{ij} denotes the contribution of the jth label in the ith channel. As we will see, the mixing matrix will play an important role is governng the behavior of the Fisher information matrix for the linear mixing model. We note that the linear mixing model is used to model hyperspectral data in many applications including fluorescence microscopy.

In the following Theorem we derive the Fisher information matrix for the linear mixing model. We also consider a special case where $\nu_{\theta,k} = \theta_k$ for $\theta \in \Theta$, $k = 1, \ldots, N_p$ and $N_{ch} = N_s$. This special case pertains to an ideal imaging configuration where the signal from a given label can be collected without being corrupted by signal from other labels at every pixel block in the input mage cube.

Theorem 2.2: Let $\mathbf{I}(\theta)$ denote the Fisher information matrix of the output image cube θ that is given in Theorem 2.1.

1. For $\theta \in \Theta$ and $k = 1, \ldots, N_p$ let $\nu_{\theta,k}$ be given by eq. 8. Then the Fisher information matrix of the output image cube θ for the linear mixing model is given by

$$\mathbf{I}(\theta) = \text{Diag} \left[A^T \mathcal{G}_1(\theta)A, A^T \mathcal{G}_2(\theta)A, \ldots, A^T \mathcal{G}_{N_p}(\theta)A \right],$$

where $\theta \in \Theta$, A denotes the mixing matrix,

$$\mathcal{G}_k(\theta) := \begin{bmatrix} \alpha_{\theta,k}^1 & 0 & \cdots & 0 \\ 0 & \alpha_{\theta,k}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha_{\theta,k}^{N_{ch}} \end{bmatrix}, \tag{9}$$

and $\alpha_{\theta,k}^j$ is given by eq. 6 for $j = 1, \ldots, N_{ch}$, $k = 1, \ldots, N_p$ and $\theta \in \Theta$.

2. Let $N_{ch} = N_s$. For $\theta \in \Theta$ and $k = 1, \ldots, N_p$, let $\nu_{\theta,k} = \theta_k$. Then the Fisher information matrix for the output image cube θ is given by

$$\mathbf{I}_{bc}(\theta) = \text{Diag} \left[\mathcal{G}_1(\theta), \mathcal{G}_2(\theta), \ldots, \mathcal{G}_{N_p}(\theta) \right], \quad \theta \in \Theta,$$

where $\mathcal{G}_k(\theta)$ is given by eq. 9 for $\theta \in \Theta$ and $k = 1, \ldots, N_p$.

Proof:

1. By definition of ν_{θ_k}, we have for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$,

$$\nu_{\theta,k}^j = \sum_{l=1}^{N_s} a_{jl}\theta_{k,l} = a_{j1}\theta_{k,1} + a_{j2}\theta_{k,2} + \cdots + a_{jN_s}\theta_{k,N_s}. \tag{10}$$

Hence for $k \neq m$ and $k, m = 1, \ldots, N_p$, we have

$$\frac{\partial \nu_{\theta,k}^j}{\partial \theta_m} = \begin{bmatrix} \frac{\partial \nu_{\theta,k}^j}{\partial \theta_{m,1}} & \frac{\partial \nu_{\theta,k}^j}{\partial \theta_{m,2}} & \cdots & \frac{\partial \nu_{\theta,k}^j}{\partial \theta_{m,N_s}} \end{bmatrix} = \mathbf{0},$$

for $\theta \in \Theta$ and $j = 1, \ldots, N_{ch}$, and $\mathbf{0}$ denotes a $1 \times N_s$ vector with all elements equal to zero. Hence the Fisher information matrix will be given by result 2 of Theorem 2.1. To prove the desired result, it is sufficient if we show that $\mathbf{I}_k(\theta) = A^T \mathcal{G}_k(\theta) A$, where $\mathbf{I}_k(\theta)$ is defined by eq. 7. Substituting eq. 10 in eq. 7, we have

$$\mathbf{I}_k(\theta) = \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \left(\frac{\partial \nu_{\theta,k}^j}{\partial \theta_k} \right)^T \frac{\partial \nu_{\theta,k}^j}{\partial \theta_k}$$

$$= \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jN_s} \end{pmatrix} \begin{pmatrix} a_{j1} & a_{j2} & \cdots & a_{jN_s} \end{pmatrix}$$

$$= \sum_{j=1}^{N_{ch}} \alpha_{\theta,k}^j \begin{pmatrix} a_{j1}^2 & a_{j1}a_{j2} & \cdots & a_{j1}a_{jN_s} \\ a_{j2}a_{j1} & a_{j2}^2 & \cdots & a_{j2}a_{jN_s} \\ \vdots & \vdots & \cdots & \vdots \\ a_{jN_s}a_{j1} & a_{jN_s}a_{j2} & \cdots & a_{jN_s}^2 \end{pmatrix}$$

$$= A^T \mathcal{G}_k(\theta) A, \quad \theta \in \Theta, \ k = 1, \ldots, N_p.$$

2. Substituting $\nu_{\theta,k} = \theta_k$ for $\theta \in \Theta$ and $k = 1, \ldots, N_p$ in eq. 7 and simplifying the result follows. ●

From result 1 of the above theorem we see that the Fisher information matrix for the linear mixing model depends on the mixing matrix A and the term $\alpha_{\theta,k}^j$ which is given by eq. 6. Note that the analytical expression of $\alpha_{\theta,k}^j$ depends on the probability density function of the data model that describes the input image cube. In the next Corollary we derive analytical expression for three of the data models that we discussed in section II-C.

From result 2 of the above Theorem, the Fisher information matrix $\mathbf{I}_{bc}(\theta)$ can be considered as a special case of result 1 with the mixing matrix being equal to a $N_s \times N_s$ identity matrix. As we will see, this result will pertain to the best case imaging scenario in that it provides the best possible limit to the accuracy with which photons from a label can be estimated in the output image cube. We note that a hyperspectral imaging system with $A = \mathbf{1}$ may not be practically realizable, for example, if there is significant spectral overlap between the labels used in the sample. Nevertheless, this provides an important benchmark to assess the effect of spectral overlap on spectral unmixing accuracy.

Corollary 2.1: For $\theta \in \Theta$, let $\mathbf{I}(\theta)$ denote the Fisher information matrix given by result 1 of Theorem 2.2, and for $k = 1, \ldots, N_p$ let $\nu_{\theta,k}$ be given by eq. 8.

1. **Gaussian data model**. For $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$, let $p^j_{\theta,k}$ be given by eq. 2. Then for $\theta \in \Theta$

$$\mathbf{I}(\theta) = \text{Diag}\left[A^T \mathcal{G}_1^g A, A^T \mathcal{G}_2^g A, \ldots, A^T \mathcal{G}_{N_p}^g A \right],$$

where for $k = 1, \ldots, N_p$ and $\theta \in \Theta$

$$\mathcal{G}_k^g(\theta) := \left(\text{Dia} \frac{1}{(\sigma_k^1)^2}, \frac{1}{(\sigma_k^2)^2}, \ldots, \frac{1}{(\sigma_k^{N_{ch}})^2} \right)$$

and σ_k^j denotes the standard deviation of Gaussian noise for $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$.

2. **Poisson data model**. For $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$, let $p^j_{\theta,k}$ be given by eq. 3. Then for $\theta \in \Theta$

$$\mathbf{I}(\theta) = \text{Diag}\left[A^T \mathcal{G}_1^p(\theta) A, A^T \mathcal{G}_2^p(\theta) A, \ldots, A^T \mathcal{G}_{N_p}^p(\theta) A \right],$$

where for $k = 1, \ldots, N_p$ and $\theta \in \Theta$

$$\mathcal{G}_k^p(\theta) := \text{Diag}\left(\frac{1}{\nu_{\theta,k}^1}, \frac{1}{\nu_{\theta,k}^2}, \ldots, \frac{1}{\nu_{\theta,k}^{N_{ch}}} \right).$$

3. **Poisson + Gaussian data model**. For $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$, let $p^j_{\theta,k}$ be given by eq. 4. Then

$$\mathbf{I}(\theta) = \text{Diag}\left[A^T \mathcal{G}_1^{pg}(\theta) A, A^T \mathcal{G}_2^{pg}(\theta) A, \ldots, A^T \mathcal{G}_{N_p}^{pg}(\theta) A \right] \quad \theta \in \Theta,$$

where for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$

$$\mathcal{G}_k^{pg}(\theta) := \text{Diag}\left(\Gamma_{\theta,k}^1, \Gamma_{\theta,k}^2, \ldots, \Gamma_{\theta,k}^{N_{ch}} \right),$$

$$\Gamma_{\theta,k}^j := \int_{\mathbb{R}} \frac{(\zeta_{\theta,k}^j(z))^2}{p_{\theta,k}^j(z)} dz - 1,$$

$p^j_{\theta,k}$ is given by eq. 4 and

$$\zeta_{\theta,k}^j(z) := \sum_{l=1}^{\infty} \frac{[\nu_{\theta,k}^j]^{l-1} e^{-\nu_{\theta,k}^j}}{(l-1)!} \frac{1}{\sqrt{2\pi}\sigma_k^j} e^{-\frac{1}{2}\left(\frac{z-l-\eta_k^j}{\sigma_k^j} \right)^2}, \quad z \in \mathbb{R}.$$

Proof: 1. Substituting for $p^j_{\theta,k}$ (eq. 2) in eq. 6, we have

$$\alpha_{\theta,k}^j = E\left[\left(\frac{\partial \ln(p_{\theta,k}^j(z_k^j))}{\partial \nu_{\theta,k}^j} \right)^2 \right]$$

$$= \frac{1}{(\sigma_k^j)^4} E[(z_k^j)^2 - 2z_k^j(\eta_k^j + \nu_{\theta,k}^j) + (\eta_k^j + \nu_{\theta,k}^j)^2] = \frac{1}{(\sigma_k^j)^2},$$

where $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. In deriving the above result we have made use of the fact that the random variables $\mathcal{I}_{\theta,k}^j$ are mutually independent of each other, $E[(z_k^j)^2] = \text{Var}(z_k^j) + (E[z_k^j])^2$ and $E[z_k^j] = \eta_k^j + \nu_{\theta,k}^j$ for $\theta \in \Theta$, $k = 1, \ldots, N_p$ and $j = 1, \ldots, N_{ch}$. Substituting the above result in eq. 9 the result follows.

2. For an independent Poisson random variable with mean $\nu_{\theta,k}^j$, where $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$, we have

$$\alpha_{\theta,k}^j = E\left[\left(\frac{\partial \ln(p_{\theta,k}^j(z_k^j))}{\partial \nu_{\theta,k}^j}\right)^2\right] = E\left[\left(\frac{z_k^j}{\nu_{\theta,k}^j} - 1\right)^2\right]$$

$$= \frac{E[(z_k^j)^2]}{(\nu_{\theta,k}^j)^2} - 1 = \frac{1}{\nu_{\theta,k}^j},$$

where $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. Substituting the above expression in eq. 9 the result follows.

3. Substituting eq. 4 in eq. 6 and simplifying (see [22] for details), we have

$$\alpha_{\theta,k}^j = \Gamma_{\theta,k}^j = \int_{\mathbb{R}} \frac{(\zeta_{\theta,k}^j(z))^2}{p_{\theta,k}^j(z)} dz - 1,$$

for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. Substituting the above expression in eq. 9 the result follows.

•

From the above Corollary we see that the Fisher information matrix for the Gaussian data model is independent of the term $\nu_{\theta,k}^j$ and depends only on the mixing matrix A and the variance $(\sigma_k^j)^2$ of the Gaussian noise for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. This is in contrast to the Poisson and the Poisson + Gaussian data models where the Fisher information matrix depends on $\nu_{\theta,k}^j$ which describes the expected number of detected photons at the kth pixel in the jth channel for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. An implication of this result is that if the underlying data model is indeed Gaussian then the Fisher information matrix of the output image cube is solely governed by the optical configuration of the imaging system and is independent of the photon budget. We note that the above result for the Gaussian data model is consistent with a previously published result ([19]).

From result 2 of Corollary 2.1, we see that the Fisher information matrix for the Poisson data model shows inverse dependence on the expected photon count. Consequently, the inverse Fisher information matrix will show linear dependence on the expected photon count. According to the Cramer-Rao inequality, for an unbiased estimator $\hat{\theta}$ of an n-dimensional vector parameter θ, we have $Var(\hat{\theta}_i) \geq [\mathbf{I}^{-1}(\theta)]_{ii}$, where $[\mathbf{I}^{-1}(\theta)]_{ii}$ denotes the Cramer-Rao lower bound of $\hat{\theta}_i$ for, $i = 1, \ldots, n$. An implication of this result is that the Cramer-Rao lower bound of the photon count estimate for the labels will increase with increasing value of the expected photon count. Since the (square root of) Cramer-Rao lower bound provides a lower bound to the variance (or standard deviation) of the photon count estimate, we will use the ratio of the

square root of the Cramer-Rao lower bound of photon count to the expected photon count as a performance measure of the spectral unmixing problem (See Definition 4.2 in Results section).

F. Channel addition

An important question that arises in the design of hyperspectral imaging systems concerns the number of spectral channels that is required to capture the signal from the label of interest in order achieve optimal spectral unmixing. Specifically, given an input image cube with N_{ch} spectral channels the question arises as to whether adding an extra spectral channel will provide any benefit for spectral unmixing. In the next Corollary we show that for the linear mixing model with N_s labels, the Fisher information matrix for the output image cube pertaining to a $N_{ch} + 1$ channel hyperspectral imaging system is greater than that of the Fisher information matrix pertaining to a N_{ch} channel hyperspectral imaging system.

Corollary 2.2: For $\theta \in \Theta$, let $\mathbf{I}(\theta \mid N_{ch})$ denote the Fisher information matrix given by result 1 of Theorem 2.2 pertaining to an input image cube with N_{ch} spectral channels. Then for $\theta \in \Theta$, $\mathbf{I}(\theta \mid N_{ch} + 1) \geq \mathbf{I}(\theta \mid N_{ch})$.

Proof: By definition of $\mathbf{I}(\theta \mid N_{ch})$, it is sufficient to show that $\mathbf{I}_k(\theta \mid N_{ch} + 1) \geq \mathbf{I}_k(\theta \mid N_{ch})$ for $\theta \in \Theta$ and $k = 1, \ldots, N_p$, where $\mathbf{I}_k(\theta \mid N_{ch}) = A^T \mathcal{G}_k(\theta) A$ for $\theta \in \Theta$, A is a $N_{ch} \times N_s$ matrix and $\mathcal{G}_k(\theta)$ is a $N_{ch} \times N_{ch}$ diagonal matrix, for $k = 1, \ldots, N_p$ that is given by eq. 9. In the case of a $N_{ch} + 1$ channel hyperspectral imaging system, we have for $\theta \in \Theta$ and $k = 1, \ldots, N_p$

$$\mathbf{I}_k(\theta \mid N_{ch} + 1) = \tilde{A}^T \tilde{\mathcal{G}}_k(\theta) \tilde{A}, \tag{11}$$

where $\tilde{A}$ is $(N_{ch} + 1) \times N_s$ matrix and $\tilde{\mathcal{G}}_k(\theta)$ is $(N_{ch} + 1) \times (N_{ch} + 1)$ diagonal matrix, for $k = 1, \ldots, N_p$. Since the number of labels is the same, we rearrange the mixing matrix $\tilde{A}$ such that the addition of an extra spectral channel results in the addition of a row at the bottom of this matrix. Hence $\tilde{A}$ can be written as $\tilde{A} = \begin{pmatrix} A \\ \mathbf{R} \end{pmatrix}$ where $\mathbf{R}$ is a $1 \times N_s$ vector that pertains to the $(N_{ch} + 1)$th channel. Similarly, the matrix $\tilde{\mathcal{G}}_k(\theta)$ can be written as

$$\tilde{\mathcal{G}}_k(\theta) = \begin{pmatrix} \mathcal{G}_k(\theta) & \mathbf{0} \\ \mathbf{0}^T & \alpha_{\theta,k}^{N_{ch}+1} \end{pmatrix}, \quad \theta \in \Theta, \quad k = 1, \ldots, N_p,$$

where $\mathbf{0}$ denotes a $N_{ch} \times 1$ zero vector and $\alpha_{\theta,k}^{N_{ch}+1}$ is a scalar term given by eq. 6 and pertains to the $(N_{ch} + 1)$th diagonal entry of $\tilde{\mathcal{G}}_k(\theta)$. Substituting the above expressions of $\tilde{\mathcal{G}}_k(\theta)$ and $\tilde{A}$ in eq. 11, we have

$$\mathbf{I}_k(\theta \mid N_{ch} + 1) = (A^T \ \mathbf{R}^T) \begin{pmatrix} \mathcal{G}_k(\theta) & \mathbf{0} \\ \mathbf{0}^T & \alpha_{\theta,k}^{N_{ch}+1} \end{pmatrix} \begin{pmatrix} A \\ \mathbf{R} \end{pmatrix}$$

$$= (A^T \ \mathbf{R}^T) \begin{pmatrix} \mathcal{G}_k(\theta) A \\ \alpha_{\theta,k}^{N_{ch}+1} \mathbf{R} \end{pmatrix} = \mathbf{I}_k(\theta \mid N_{ch}) + \alpha_{\theta,k}^{N_{ch}+1} \mathbf{R}^T \mathbf{R},$$

where $\theta \in \Theta$ and $k = 1, \ldots, N_p$. By definition, $\alpha_{\theta,k}^{N_{ch}+1} \geq 0$ for $\theta, \in \Theta$ and $k = 1, \ldots, N_p$ and $\mathbf{R}^T \mathbf{R} \geq 0$. From this the result follows. ●

In deriving the above result, we made no specific assumptions about the underlying data model that describes the input image cube. Therefore the above result is applicable to a wide variety of imaging configurations for the linear mixing model. An immediate implication of the above result is that the limit of the accuracy of photon count for a given label improves by adding extra spectral channels to detect the signal from that label.

G. Channel splitting and photon partitioning

In the previous section, we saw how adding an extra channel to a hyperspectral imaging system can improve the Fisher information matrix. In many practical situations, channel addition may not be feasible since the overall spectral bandwidth to collect the signal from a given label cannot be increased. In such cases, a question then arises as to whether spectral subsampling within the passband would be beneficial. Specifically, if we have a hyperspectral imaging system with N_{ch} spectral channels and if one of the spectral channels are split into two channels of smaller spectral bandwidth, then under what conditions is channel splitting beneficial? In the next corollary we show that channel splitting is beneficial for the Poisson data model.

Corollary 2.3: For $\theta \in \Theta$, let $\mathbf{I}(\theta \mid N_{ch})$ denote that Fisher information matrix given by result 2 of Theorem 2.1 where N_{ch} denotes the number of spectral channels. Assume that for $j = 1, \ldots, N_{ch}$ there exists a partition function $0 < \gamma_\theta^j < 1$ with $\frac{\partial \gamma_\theta^j}{\partial \theta} \neq 0$ for $\theta \in \Theta$ and $j = 1, \ldots, N_{ch}$ such that the addition of an extra channel splits the expected photon counts into two fractions given by $\gamma_\theta^j \nu_{\theta,k}^j$ and $(1 - \gamma_\theta^j)\nu_{\theta,k}^j$, where $\nu_{\theta,k}^j$ denotes the expected photon count of the kth pixel in the jth spectral image, for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$. If we consider a Poisson data model, then

$$\mathbf{I}(\theta \mid N_{ch} + 1) > \mathbf{I}(\theta \mid N_{ch}), \quad \theta \in \Theta.$$

Proof: By definition of $\mathbf{I}(\theta \mid N_{ch})$, it is sufficient to show that $\mathbf{I}_k(\theta \mid N_{ch} + 1) \geq \mathbf{I}_k(\theta \mid N_{ch})$ for $\theta \in \Theta$ and $k = 1, \ldots, N_p$, where $\mathbf{I}_k(\theta \mid N_{ch})$ is given by eq. 7. Without loss of generality, we set $N_{ch} = 1$. For a Poisson data model, it can be shown that $\alpha_{\theta,k}^j = \nu_{\theta,k}^j$ for $\theta \in \Theta$, $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$ where $\alpha_{\theta,k}^j$ is defined in eq. 6. Substituting this in eq. 7, we get

$$\mathbf{I}_k(\theta \mid 1) = \frac{1}{\nu_{\theta,k}^1}\left(\frac{\partial \nu_{\theta,k}^1}{\partial \theta_k}\right)^T \frac{\partial \nu_{\theta,k}^1}{\partial \theta_k}, \quad k = 1, \ldots, N_p, \quad \theta \in \Theta.$$

If a second spectral channel is introduced ($N_{ch} = 2$) and a fraction of the photons detected in the first channel is now captured in the second channel, then we have for $\theta \in \Theta$ and $k = 1, \ldots, N_p$,

$$\mathbf{I}_k(\theta \mid 2) = \frac{1}{\gamma_\theta^1 \nu_{\theta,k}^1}\left(\frac{\partial(\gamma_\theta^1 \nu_{\theta,k})}{\partial \theta_k}\right)^T \frac{\partial(\gamma_\theta^j \nu_{\theta,k})}{\partial \theta_k}$$

$$+ \frac{1}{(1 - \gamma_\theta^1)\nu_{\theta,k}^1}\left(\frac{\partial((1 - \gamma_\theta^1)\nu_{\theta,k})}{\partial \theta_k}\right)^T \frac{\partial((1 - \gamma_\theta^j)\nu_{\theta,k})}{\partial \theta_k}$$

$$= \frac{1}{\nu_{\theta,k}^1} \left(\frac{\partial \nu_{\theta,k}^1}{\partial \theta_k} \right)^T \frac{\partial \nu_{\theta,k}^1}{\partial \theta_k} + \frac{\nu_{\theta,k}^1}{\gamma_\theta^1(1-\gamma_\theta^1)} \left(\frac{\partial \gamma_\theta^1}{\partial \theta_k} \right)^2$$

$$= \mathbf{I}(\theta \mid 1) + \frac{\nu_{\theta,k}^1}{\gamma_\theta^1(1-\gamma_\theta^1)} \left(\frac{\partial \gamma_\theta^1}{\partial \theta_k} \right)^2.$$

From this the result follows. ●

Note that unlike Corollary 2.2 which used the analytical expression for the Fisher information matrix given in Theorem 2.2, the above corollary used a more general expression for the Fisher information matrix given by result 2 of Theorem 2.1. The reason for this is due to the dependence of θ by the partition function γ_θ^j for $j = 1, \ldots, N_{ch}$ and $\theta \in \Theta$. As we will show in Section IV-E, the above result is only true for the Poisson data model. In fact for the Poisson + Gaussian data model we show that the limit of the accuracy of the photon count exhibit complex behavior with channel splitting and eventually deteriorates when increasing the number of partitioned channels. We note that the result given in Corollary 2.3 was also reported in [20] for the Poisson data model.

III. METHODS

A. Computing the Fisher information matrix and the mixing matrix A

The Fisher information matrix was calculated using the FandPLimitTool ([27]) which is a MATLAB based software package that contains an extensive suite of tools to calculate the Fisher information matrix for a wide variety of parameter estimation problems in fluorescence microscopy. For the linear spectral unmixing problem, the Fisher information matrix depends on the mixing matrix A which needs to be computed for for illustrating the results. We note that in a practical situation, A will be experimentally determined by imaging single color control samples under identical imaging conditions to that of the sample that contains N_s different labels. Here we present a simplified approach to theoretically calculate A by using relevant reference data such as the the excitation and emission spectra of fluorescent labels, the spectral emissivity of the fluorescent light source and passbands of the excitation and emission filters. The elements of the mixing matrix A can be written as

$$a_{ij} = \xi_{ex}(i,j)\xi_{em}(i,j), \qquad i = 1, \ldots, N_{ch}, \quad j = 1, \ldots, N_s,$$

where for $i = 1, \ldots, N_{ch}$ and $j = 1, \ldots, N_s$

$$\xi_{ex}(i,j) := \frac{b(i,j)}{\max(b(1,j), b(2,j), \ldots, b(N_{ch}, j))},$$

$$b(i,j) := \int_{\lambda_{i,min}^{ex}}^{\lambda_{i,max}^{ex}} t_{j,ex}(\lambda) E_{LS}(\lambda) d\lambda,$$

$$\xi_{em}(i,j) := \int_{\lambda_{i,min}^{em}}^{\lambda_{i,max}^{em}} t_{j,em}(\lambda) t_{opt}(\lambda) t_{cam}(\lambda) d\lambda.$$

In the above equations, $[\lambda_{i,min}^{ex}, \lambda_{i.max}^{ex}]$ and $[\lambda_{i,min}^{em}, \lambda_{i.max}^{em}]$ denote the excitation and emission passbands corresponding to the ith channel, $t_{j,ex}$ and $t_{j,em}$ denote the fluorescence excitation and emission spectra of the jth dye, $j = 1, \ldots, N_s$, E_{LS} denotes the spectral emissivity of the fluorescent light source, t_{opt} denotes the spectral transmitivity of the optical components (e.g., objective and tube lens) in the emission light path, and t_{cam} denotes the spectral sensitivity of the detector. For simplicity, we set $t_{opt}(\lambda) = 1$ and $t_{cam}(\lambda) = 1$, and also assume that the transmission efficiency of the excitation and emission filters is equal to 1 in the passbands and equal to zero outside the passbands. The excitation and emission spectra of the fluorescent labels and the spectral emissivity of the fluorescent light source were obtained from the Pubspectra database [28] and were modified such that their sum equals 1. The above integrals were numerically evaluated using the Trapezoidal rule.

B. Hyperspectral data simulation

All simulations and estimations were carried out in the MATLAB programming language. To simulate hyperspectral data, we use the above approach to compute the mixing matrix A and use eq. 8 to compute the expected photon counts of the fluorescent labels $\nu_{\theta,k}$ at the kth pixel in the input image cube. For the Poisson + Gaussian data model, we first create a Poisson realization of $\nu_{\theta,k}$ and then add Gaussian noise with mean $\eta_k^j = 0$ e$^-$/pixel and standard deviation $\sigma_{w,k}^j = 8$ e$^-$/pixel to the Poisson data.

IV. RESULTS

In this section, we illustrate the results obtained in the previous section concerning the linear mixing model. We first define a few entities that we will use throughout this section.

Definition 4.1: For $\theta \in \Theta$, $i = 1, \ldots, N_s$ and $k = 1, \ldots, N_p$, the **Linear Unmixing Performance (LUP) bound** for the ith label at the kth pixel in the output image cube is defined as $\delta_{k,i} = \sqrt{[\mathbf{I}_k^{-1}(\theta)]_{ii}}$, $\theta \in \Theta$, and the **LUP bound for the best case scenario** is defined as $\delta_{k,i}^{bc} = \sqrt{[\mathcal{G}_k^{-1}(\theta)]_{ii}}$, where $\mathbf{I}_k(\theta) = A^T \mathcal{G}_k(\theta) A$, A denotes the mixing matrix and $\mathcal{G}_k(\theta)$ is given by 9, for $\theta \in \Theta$ and $k = 1, \ldots, N_p$.

Definition 4.2: For $\theta \in \Theta$, $i = 1, \ldots, N_s$ and $k = 1, \ldots, N_p$, the **normalized LUP (nLUP) bound** for the ith label at the kth pixel in the output image cube is defined as $\frac{\delta_{k,i}}{\theta_{k,i}}$, and the **nLUP bound for the best case scenario** is defined as $\frac{\delta_{k,i}^{bc}}{\theta_{k,i}}$, where $\delta_{k,i}$ and $\delta_{k,i}^{bc}$ are defined in Definition 4.1.

A. Spectral resolution for a pair of fluorophores

An important question that arises in hyperspectral imaging applications concerns with how accurately two spectrally overlapping fluorescent labels can be discerned from the acquired data. Here we address this problem by considering a pair of fluorescent labels whose emission maxima are separated by a distance d_s (see Fig. 2A). We assume an imaging configuration where the sample is illuminated (sequentially or simultaneously) to excite the labels such that the excitation passband for a given label overlaps with its

corresponding excitation maxima. The fluorescence signal is then detected (sequentially or simultaneously) in two distinct spectral channels, where the emission passband of each spectral channel covers the corresponding emission maxima of a particular label. We then ask the question how decreasing the spectral distance d_s affects the spectral resolution of the two fluorophores. For the current discussion, we consider the fluorescent labels with traditional Stokes shift fluorescence signal ([29]) where the peak of the emission spectra for the label occurs at a longer wavelength (lower energy) than the peak of its corresponding excitation spectra.

Figs. 2B and 2C show the nLUP bounds for labels 1 and 2, respectively, for different data models as a function of the spectral distance d_s. The figures also show the nLUP bounds for labels 1 and 2 for the best case scenario, which pertains to the ideal imaging configuration wherein the signal collected from a given label is not corrupted by photons from other labels present in the sample. Here, to simulate the decrease in d_s, we shift the excitation and emission spectra of label 2 and also the corresponding passbands of the excitation and emission filters towards label 1 while leaving the excitation and emission spectra for label 1 unchanged. By definition, the nLUP bound is a ratio of the LUP bound to the expected photon count for a given label (Definition 4.2). A large numerical value of the nLUP bound predicts a relatively high level of error in estimating the photon count at a given pixel while a small numerical value predicts a relatively low level of error in estimating the photon count at a given pixel.

From the figures we see that when the labels are spectrally well separated the nLUP bound is numerically close to the nLUP bound for the best case scenario. This is expected since there is no spectral crosstalk between the excitation or emission spectra of the two labels. Consistent with this, we see that the mixing matrix A is diagonal (see Fig. 2C). As the spectral distance decreases, the excitation/emission spectra of the labels start to overlap and the mixing matrix is no longer diagonal. Consequently, the numerical value of the nLUP bound becomes bigger and starts to deviate from the nLUP bound for the best case scenario. When the spectral distance further decreases, the nLUP bound continues to deteriorate and we see that the numerical values of the off-diagonal terms of the A matrix, which accounts for the spectral cross talk between the labels, is comparable to the diagonal terms. Note that the nLUP bound for the Poisson + Gaussian data model is consistently higher than that for the Poisson data model, since in the former model the presence of additive Gaussian noise corrupts the detected photon counts at each pixel in the input image cube which leads to higher uncertainty in estimating the photon counts at a given pixel.

We note that the difference between the nLUP bound and the nLUP bound for the best case scenario is referred to as the multiplexing disadvantage ([15]) which accounts for the deterioration in the limit of the accuracy of photon count from the best case scenario when we consider the effect of detecting unwanted photons from other labels present in the sample. As we will see in rest of this section, the nLUP bound for the best case scenario will provide an important benchmark for studying the performance limits of hyperspectral imaging systems. In our present example the shape of the emission spectra for labels 1 and

2 are dissimilar, and therefore even for a very small spectral distance of 1 nm, the nLUP bound remains finite for both labels.
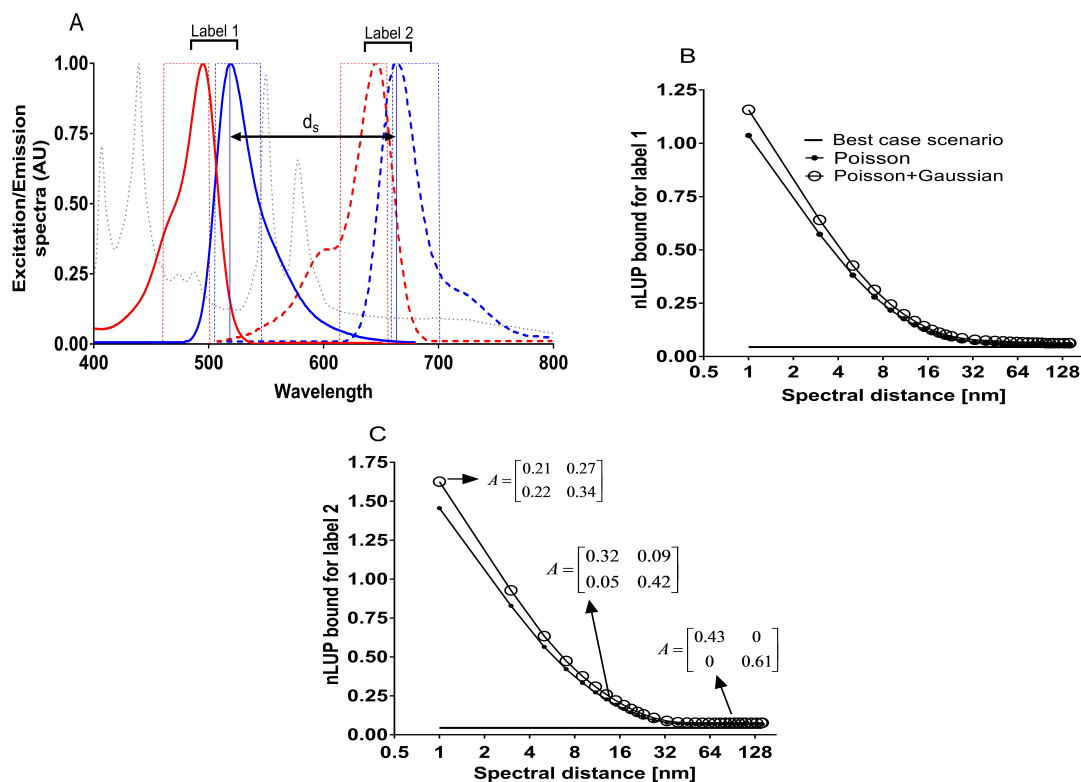

Fig. 2. The effect of changing the spectral distance on spectral resolution. Panel A shows the normalized excitation (red lines) and emission spectra (blue lines) of two fluorescent labels that are spectrally separated by a distance d_s. The panel also shows the spectral emissivity of a metal hallide broadband light source that is used to excite the two labels. Here, for both labels we consider the traditional Stoke's shift for their fluorescence emission spectra. Panels B and C show the nLUP bound for labels 1 and 2 at a pixel, respectively, for the Poisson and the Poisson + Gaussian data models and also for the best case scenario pertaining to the Poisson data model. Panel C also shows the mixing matrix pertaining to different values of d_s. In panels B and C, the expected photon count for both labels is set to be 500 photons per pixel and for the Poisson + Gaussian data model we consider the mean and standard deviation of the Gaussian noise to be 0 e$^-$/pixel and 8 e$^-$/pixel, respectively.

B. Improving the spectral resolution by using anti-Stokes shift fluorescence

In the previous section we saw that with traditional fluorescent labels which exhibit Stokes shift fluorescence, there is a limit to how spectrally close the two labels can be and still be accurately spectrally unmixed. In this section, we show that by replacing one of the labels with a fluorophore that exhibits anti-Stokes shift fluorescence we can overcome the spectral resolution limit. Fig. 3 shows the excitation and emission spectra for such a pair of labels, where label 1 is a traditional fluorescent label as in Fig. 2 and label 2 exhibits anti-Stokes shift fluorescence ([29]) wherein the peak of its emission spectra occurs at a shorter wavelength (higher energy) relative to the peak of its excitation spectra. It should be pointed out that at present there is significant interest in developing such probes especially for biological applications ([30], [31], [32]). For example, upconverting nanoparticles are one such a class of fluorescent labels

that typically exhibit strong absorption in the near infrared wavelength range (700 - 1000 nm) and emit fluorescence signal in the visible range (400 - 650 nm) ([33], [34]).

In fig. 3 we have simulated the excitation and emission spectra for a hypothetical label that exhibits anti-Stokes shift fluorescence. We consider an imaging configuration where the labels are sequentially illuminated over two excitation passbands that cover their corresponding excitation maximas, and the fluorescence signal is sequentially collected over emission passbands that cover their emission maximas. To simulate the decrease in spectral distance, we shift the excitation and emission spectra of label 2 and their corresponding excitation and emission passbands towards label 1, and keep the excitation and emission spectra of label 1 fixed.

Figs. 3B and 3C show the nLUP bounds for labels 1 and 2, respectively, pertaining to the Poisson and Poisson+Gaussian data models. The figures also show the nLUP bound for the best case scenario pertaining to the Poisson data model. From the figures we see that the nLUP is almost constant for all values of the spectral distance. Specifically, the mixing matrix A remains diagonal irrespective of the spectral distance of separation between the two labels. This can attributed to anti-Stokes shift fluorescence which significantly eliminates the overlap of the excitation spectra between labels 1 and 2 when the spectral distance decreases. Therefore for above the imaging configuration there is very limited spectral cross talk between the two labels and consequently the nLUP bound remains constant and is consistently close to the nLUP bound for the best case scenario. Note that similar to Fig. 2, the nLUP bound for the Poisson+Gaussian data model is consistently higher than the nLUP bound for the Poisson data model.

C. Effect of photon count

We next investigate how changing the expected photon count impacts the linear unmixing performance bound. Here, we consider two fluorescent labels Cy3 and Cy3.5 which have significant overlapping excitation and emission spectra (Figs. 4A and 4B). We assume an imaging configuration in which the fluorophores are sequentially excited at distinct excitation passbands and the corresponding fluorescence signal is then sequentially collected at the indicated emission passbands.

Figs. 4C and 4D show the nLUP bound for the Poisson and the Poisson + Gaussian data models along with the nLUP bound for the best case scenario pertaining to the Poisson data model. We see that as the expected photon count /pixel increases, the numerical value of the nLUP bound becomes smaller and approaches the nLUP bound for the best case scenario. Note that the nLUP bound for the Poisson + Gaussian data model is consistently higher than that of the Poisson data model, and the difference starts to diminish as the expected photon count / pixel increases. Taken together, these results imply that increasing the photon/light budget would result in an overall improvement in the performance bound.
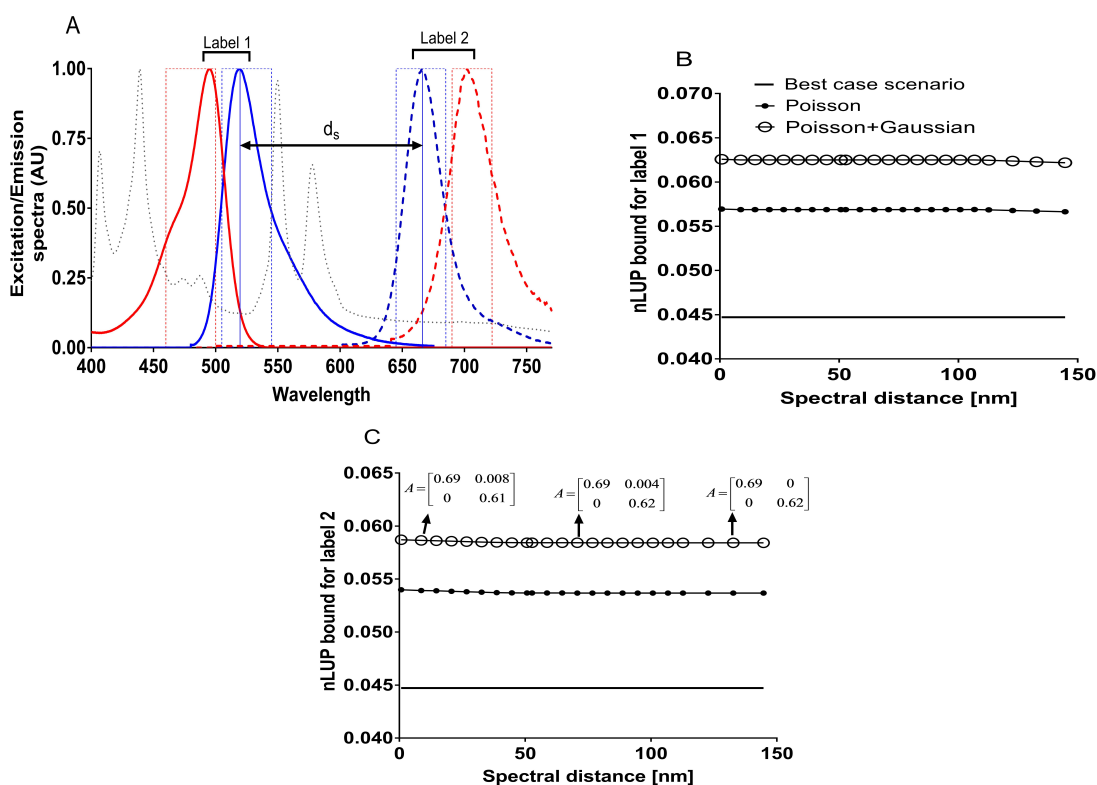
Fig. 3. Improving the spectral resolution by using anti-Stokes shift fluorescences. Panel A shows the normalized excitation (red lines) and emission spectra (blue lines) of two fluorescent labels that are spectrally separated by a distance d_s. The panel also shows the spectral emissivity of a metal hallide broadband light source that is used to excite the two labels. Here, label 1 is the same as that shown in Fig. 2 while label 2 is a fluorophore with anti-Stokes shift fluorescence emission, where the peak of its emission spectra is at a shorter wavelength (higher energy) than the peak of its excitation spectra. Panels B and C show the nLUP bounds for labels 1 and 2, respectively, at a pixel for the Poisson and the Poisson + Gaussian data models and also for the best case scenario pertaining to the Poisson data model. Panel C also shows the mixing matrix pertaining to different values of d_s. The numerical values used to generate the above plots are identical to those used in Fig. 2.

D. Channel addition

We next investigate the effect of adding extra spectral channels to collect the signal from a label on the spectral unmixing accuracy. Here again we consider Cy3 and Cy3.5 labels and calculate the LUP bound that is defined in Definition 4.1. Unlike the results given in Figs. 3 and 4, here we compute the LUP bound since the expected photon count is fixed for the fluorescent labels. We assume an imaging configuration such that the fluorescence signal from each label is sequentially collected in one or more spectral channels (i.e. different emission passbands) by repetitively exciting the sample at the same excitation passband. For example, for the Cy3.5 fluorophore, consider a 6 channel configuration with emission passbands in the range of 590 - 610 nm, 610 - 630 nm, 630 - 650 nm, 650 - 670 nm, 670 - 690 nm and 690 - 710 nm (see Fig. 5A). Here the sample will be sequentially excited 6 times at the excitation passband of 542 - 582 nm and each time the fluorescence signal will be collected in a different emission passband pertaining to Cy3.5. Note that in this configuration the number of detected photons from the fluorescent label will increase with increasing number of channels assuming that there is negligible photobleaching effect which
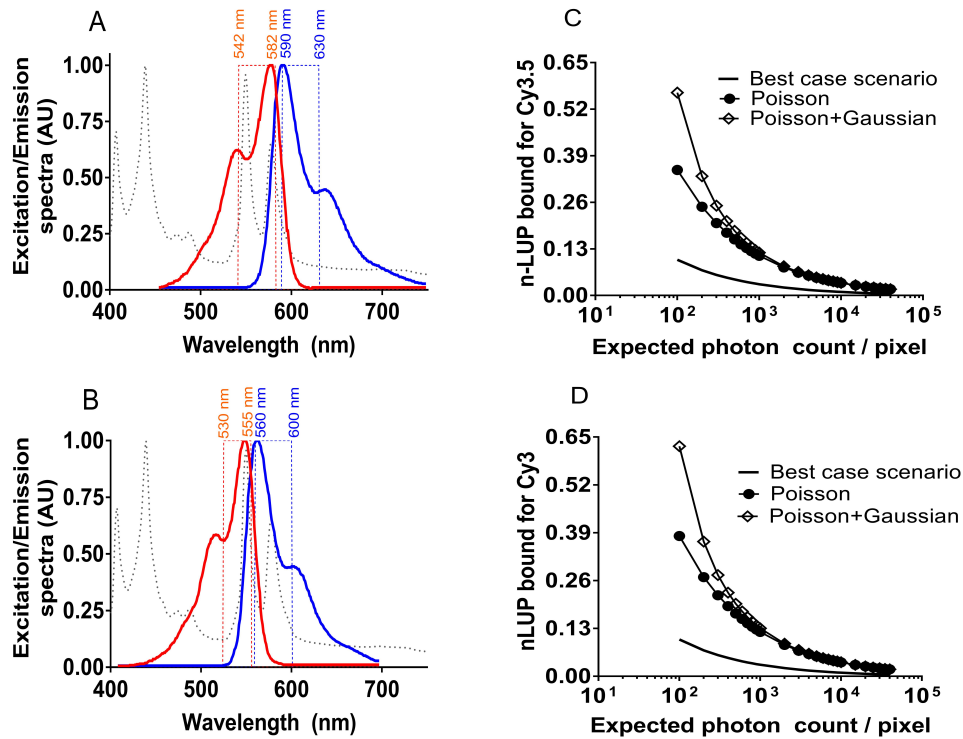
Fig. 4. Effect of photon count on the nLUP bound. Panels A and B show the normalized excitation (red lines) and emission (blue lines) spectra for Cy3.5 and Cy3 fluorescent labels, respectively, along with the spectral emissivity of a metal hallide light source (dotted line). The panels also show the corresponding excitation and emission passbands for each fluorescent label. Panels C and D show the behavior of the nLUP bound for Cy3.5 and Cy3, respectively, at a single pixel as a function of the expected photon count for the Poisson and the Poisson + Gaussian data models. As reference, the panel also shows the nLUP bound for the best case scenario pertaining to the Poisson data model. Here, we assume the expected photon count to be the same for both fluorophores and for the Poisson + Gaussian data model, the mean and standard deviation of the readout noise is set to be 0 e^-/pixel and $8e^-$/pixel, respectively.

typically diminishes the fluorescence intensity of the label with repeated excitation. We note that such an imaging configuration has been implemented, for example, using specialized emission filters that are placed before the imaging detector in the light path in which the passband of the filter can be changed either electronically ([9]) or optomechanically by changing the relative angle of incidence of light on the emission filter ([7])

Figs. 5C and 5D show the LUP bound for Cy3.5 and Cy3 labels, respectively, as a function of the number of spectral channels for two different data models. From the figure we see that for both the Poisson and the Poisson + Gaussian data models the LUP bound improves by increasing the number of channels. This is consistent with Corollary 2.2, where we showed that the Fisher information matrix for an output image cube pertaining to an $N_{ch} + 1$ channel hyperspectral system is greater than or equal to that of the Fisher information matrix for an N_{ch} channel hyperspectral imaging system. Note that for both labels, the LUP bound for the Poisson + Gaussian data model is consistently higher than that of the Poisson data model, which is consistent with the behavior that we observed in Fig. 4. We also observe that the addition of spectral channels beyond a certain wavelength range is not beneficial as we see that the LUP bound starts to plateau out.
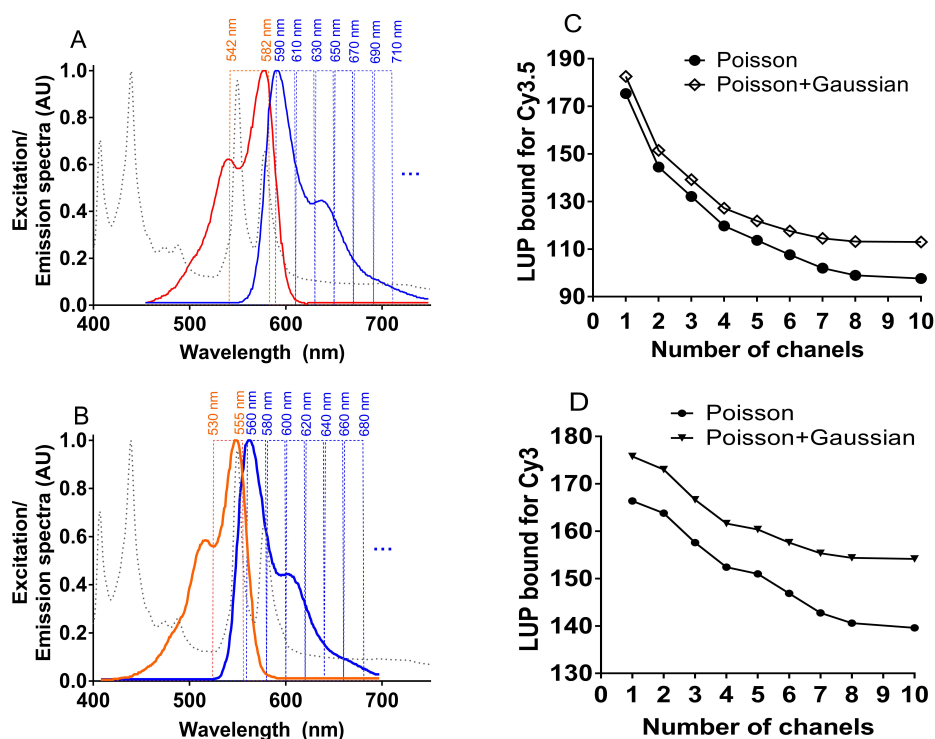
Fig. 5. Effect of channel addition on the LUP bound. Panels A and B show the normalized excitation (red lines) and emission (blue lines) spectra for Cy3.5 and Cy3 fluorescent labels, respectively, along with the spectral emissivity of a metal hallide light source (dotted line). The panels also show the excitation passband and the emission passbands pertaining to the different spectral channels. Panel C shows the behavior of the LUP bound for Cy3.5 at a single pixel as a function of the number of channels after channel addition for different data models. Panel D shows the same for Cy3. In Panels C and D, the expected photon count is set to be 3000 for both labels, and the numerical values of the mean and standard deviation of the Gaussian noise component are identical to those used in Fig. 4.

E. Channel splitting

In the previous section we investigated how the addition of spectral channels impacts the LUP bound for a given label. In many situations, the addition of spectral channels is not feasible. More specifically the emission passband for a given fluorescent label is typically fixed. In such cases a question arises as to whether splitting the available passband into narrower spectral channels will render any benefit. Here we investigate this question for Cy3 and Cy3.5 labels. For each fluorescent label, we consider a partitioning scheme wherein we begin with a single emission passband and then partition the passband into 2, 4, 8, 13, 20 and 40 smaller bands. For example, for Cy3.5 we start with the emission passband of 590 nm - 630 nm (see Fig. 4), which we split into two narrower channels, i.e. 590 nm - 610 nm and 610 nm - 630 nm, then into 4 narrower channels, i.e. 590 nm - 600 nm, 600 nm - 610 nm, 610 nm - 620 nm and 620 nm - 630 nm, and so on. A similar partitioning scheme is also adopted for Cy3 where we start with the emission passband of 560 - 600 nm.

We consider an imaging configuration where the labels are sequentially excited at a specific excitation passband followed by simultaneous detection of the emitted photons across the different spectral channels for that label. Unlike the imaging configuration that we considered in Section IV-D, here the total number

of photons collected from a given label remains the same as we increase the number of partitioned channels. We note that the above imaging configuration can be realized in a spectral confocal microscope or in a hyperspectral line scanning microscope that is equipped with a dispersive optical element such as a prism or a monochromator, which spectrally disperses the incident light along a line and the dispersed signal is then captured on a linear imaging detector.
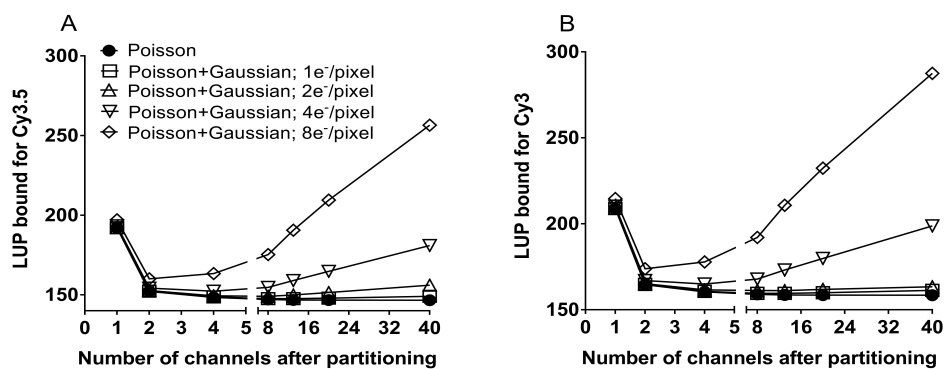

Fig. 6. The effect of channel splitting on the LUP bound. Panel A shows the behavior of the LUP bound for Cy3.5 at a single pixel as a function of the number of channels after channel splitting for the Poisson and the Poisson + Gaussian data models. Panel B shows the same for Cy3. In Panels A and B, the expected photon count is set to be 3000 for both labels, and for the Poisson + Gaussian data model we assume the mean of the Gaussian noise to be 0 e$^-$/pixel and different values of standard deviation as indicated in the legend in panel A.

Figs. 6A and 6B, show the LUP bound for Cy3.5 and Cy3 labels at a given pixel as a function of the number of partitioned channels for different data models. For the Poisson data model we see that channel splitting improves the LUP bound for both labels. Specifically, the LUP bound decreases with increasing number of channels after partitioning. This is consistent with Corollary 2.3, where we showed that for the Poisson data model the Fisher information matrix for a hyperspectral imaging system with $N_{ch} + 1$ partitioned channels is greater than that with N_{ch} partitioned channels.

Figs. 6A and 6B also show the LUP bound for the Poisson + Gaussian data model for different readout noise levels. Here we assume that the readout noise remains the same as the number of partitioned channels increases. Unlike the Poisson data model, we see that the LUP bound first improves but then progressively deteriorates with increasing number of partitioned channels. Specifically, when the number of channels initially increases, this provides additional spectral information which in turn results in an improvement of the LUP bound. However, as the number of partitioned channels increases, the passband for each spectral channel becomes narrower and in turn the expected photon count per spectral channel decreases while the readout noise remains the same.

Note that the extent of this deterioration is proportional to the standard deviation of the readout noise. For example for Cy3.5 label, if the standard deviation of the readout noise is 8 e$^-$ /pixel, then the LUP bound varies from 160 photons to 210 photons when the number of channels after partitioning increases from 2 to 20. On the other hand, if the standard deviation of the readout noise is 2 e$^-$/pixel, then the LUP bound varies from 149 photons to 152 photons when the number of channels after partitioning increases

from 2 to 20. We note that a similar result has been reported for the localization accuracy problem in fluorescence microscopy, where the limit of the accuracy of the X/Y location coordinate also shows an analogous behavior as a function of pixel size of the imaging detector for the Poisson and the Poisson + Gaussian data models ([35]).

V. ASSESSMENT OF SPECTRAL UNMIXING ALGORITHMS

In the previous sections we investigated the behavior of the linear unmixing performance bound for different experimental configurations and imaging conditions. A fundamental question remains whether there exists an unbiased estimator than will attain the performance bound. To address this question, we consider two widely used algorithms for spectral unmixing namely the least squares (LS) estimator and the maximum likelihood (ML) estimator, and evaluate their performance on simulated hyperspectral imaging data. Recall from Theorem 2.2 that the block diagonality of the Fisher information matrix implies that the accuracy with which the photon counts are estimated at the kth pixel in the output image cube is independent of the accuracy of photon count estimates at the mth pixel when $k \neq m$. Hence for the current discussion it is sufficient to evaluate the performance of the spectral unmixing algorithms for hyperspectral data simulated at a single pixel. For $k = 1, \ldots, N_p$, let $z_k := [z_k^1 \ z_k^2 \ \ldots \ z_k^{N_{ch}}]^T$, denote the detected photon counts at the kth pixel block in the input image cube. The LS estimator of θ_k can be written as

$$\hat{\theta}_k(LS) := (A^T A)^{-1} A^T z_k, \quad k = 1, \ldots, N_p,$$

where A denotes the mixing matrix. The ML estimator can be written as

$$\hat{\theta}_k(ML) = \max_{\theta \in \Theta} \left(\sum_{j=1}^{N_{ch}} \ln(p_{\theta,k}^j(z_k^j)) \right), \quad k = 1, \ldots, N_p,$$

where $p_{\theta,k}^j$ denotes the probability density function of the photon counts detected at the kth pixel in the jth spectral channel, for $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$.

From the definition it is straightforward to see that for the Poisson data model (eq. 3) the LS estimator is unbiased, since

$$E[\hat{\theta}_k(LS)] = (A^T A)^{-1} A^T E[z_k] = (A^T A)^{-1} A^T \nu_{\theta,k}$$

$$= (A^T A)^{-1} (A^T A) \theta_k = \theta_k, \ \theta \in \Theta, \ k = 1, \ldots, N_p,$$

where we have used eq. 8. However, if the hyperspectral data follows either the Gaussian data model (eq. 2) or the Poisson + Gaussian data model (eq. 4), then we have

$$E[\hat{\theta}_k(LS)] = (A^T A)^{-1} A^T E[z_k] = (A^T A)^{-1} A^T (\nu_{\theta,k} + \eta_k)$$

$$= (A^T A)^{-1} (A^T (A\theta_k + \eta_k) = \theta_k + (A^T A)^{-1} A^T \eta_k,$$

where $\eta_k := [\eta_k^1 \ \eta_k^2 \ldots \ \eta_k^{N_{ch}}]^T$ and $k = 1, \ldots, N_p$. From the above equation we see that the LS estimator will be unbiased only when the mean of the Gaussian noise component η_k^j is equal to zero for $j = 1, \ldots, N_{ch}$ and $k = 1, \ldots, N_p$.
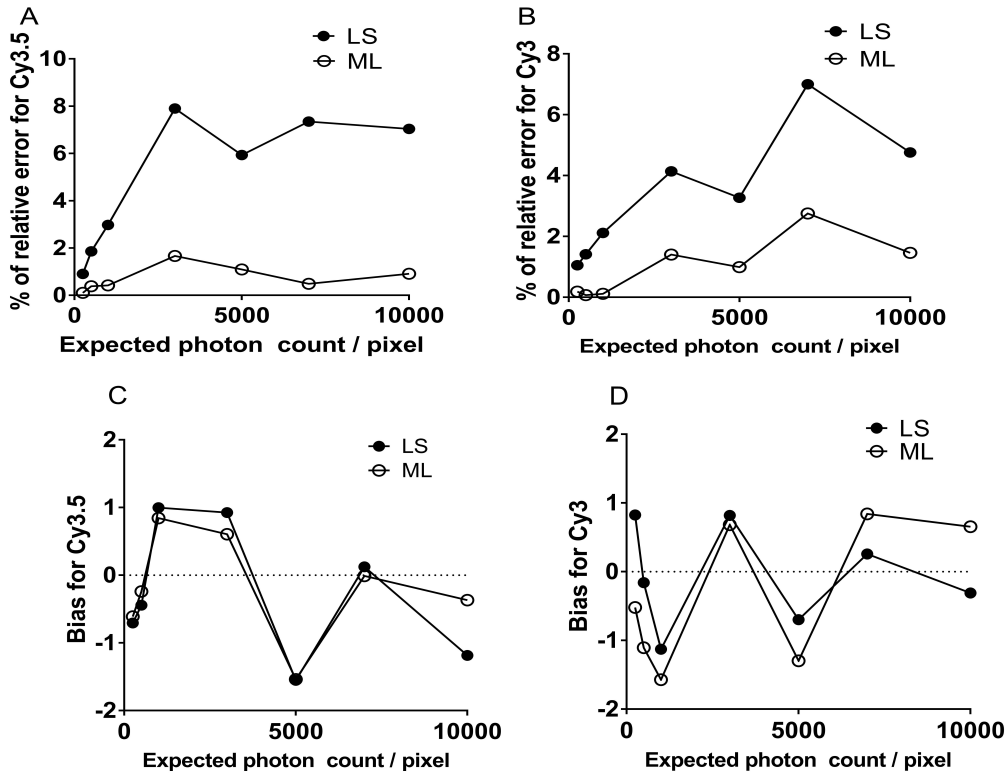

Fig. 7. Performance of spectral unmixing algorithms on hyperspectral data. The figure shows the behavior of the LS and ML estimators on simulated hyperspectral data. Panels A and B show the % of relative error of the estimators for Cy3.5 and Cy3 labels, respectively, for different expected photon counts. Here we assume that the expected photon count is the same for both fluorescence labels. Panels C and D show the bias of the estimators for Cy3.5 amd Cy3, respectively, for different expected photon counts.

We consider the hyperspectral imaging configuration illustrated in Fig. 5 where the fluorescent labels Cy3 and Cy3.5 are sequentially excited and their fluorescence signal is detected across 6 distinct spectral channels. We assume that the expected photon count for each label is the same and simulate the hyperspectral data as described in Section III-B. For each value of expected photon count, we create 10,000 realizations of the data and estimate the photon counts associated with each fluorescent label by using the different spectral unmixing algorithms. To assess the performance of the unmixing algorithms, we calculate the bias and the relative error which are given by

$$\text{Bias for } j^{th} \text{ label} = \text{mean}(\hat{\theta}_j) - \theta_j,$$

$$\% \text{ of relative error for } j^{th} \text{ label} = 100 \times \frac{\frac{\text{mean}(\hat{\theta}_j)}{\text{std}(\hat{\theta}_j)} - \frac{\delta_j}{\theta_j}}{\frac{\delta_j}{\theta_j}},$$

where $j = 1, 2$, δ_j/θ_j denotes the nLUP bound for θ_j (see Definition 4.2).

Fig. V shows the performance of the LS and ML estimators for the hyperspectral data generated for different values of the expected photon counts. From the figure, we see that the performance of LS and ML estimators are relative close to the nLUP bound in that the % of relative error for both labels are within 8% of their corresponding nLUP bounds. Note that the performance of the ML estimator is consistently better than that of the LS estimator for a range of expected photon count values, since the % of relative error for the ML estimator is lower than that for the LS estimator. This is expected since we used the Poisson + Gaussian data model to simulate the hyperspectral data and used the appropriate probability density function to estimate the photon counts in the ML estimator. The figure also shows the bias for the LS and ML estimators. By definition the LS estimator is unbiased and consistent with this we see that the bias estimates of the LS estimator for both labels are equally distributed between positive and negative values. In particular, the magnitude of the bias estimate is within ± 1 photon from the true value of the photon count. An almost identical behavior is observed for the ML estimator which suggests that the ML estimator is also unbiased for the imaging conditions tested here.

VI. CONCLUSION

A general stochastic model for hyperspectral imaging data was presented and analytical expressions for the Fisher information matrix was derived. The model is based on relatively broad assumptions about the underlying probability density function that describes the hyperspectral data and allows for a wide variety of imaging conditions. As an application, we considered the linear mixing model and showed that the Fisher information matrix becomes block diagonal. Using the Cramer-Rao inequality, we introduced a linear unmixing performance bound and showed how this can be used to predict the spectral resolution limit for two spectrally overlapping fluorescent labels. We also showed how the spectral resolution limit can be surpassed in a standard microscope configuration by using the phenomenon of anti-Stokes shift fluorescence. Further, we investigated the effects of channel addition and channel splitting on the the behavior of the performance bound and illustrated them through concrete examples. Finally, we evaluated the performance of the least squares and maximum likelihood estimators for spectral unmixing, and studied their bias and variance behaviors at different photon/light budgets. In conclusion we note that the results and analysis presented here extend prior studies by providing a comprehensive framework to analyze the performance limits of wide variety of hyperspectral imaging systems and spectral unmixing algorithms.

REFERENCES

[1] M. E. Dickinson, G. Bearman, S. Tille, R. Lansford, and S. E. Fraser, "Multispectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy," *Biotechniques*, vol. 31, pp. 1272–1278, 2001.

[2] T. J. Fountaine, S. M. Wincovitch, D. H. Geho, S. H. Garfield, and S. Pittaluga, "Multispectral imaging of clinically relevant cellular targets in tonsil and lymphoid tissue using semiconductor quantum dots," *Modern Pathology*, vol. 19, pp. 1181–1191, 2006.

[3] F. Cutrale, V. Trivedi, L. A. Trinh, C. Chiu, J. M. Choi, M. S. Artiga, and S. E. Fraser, "Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging," *Nature Methods*, vol. 14, pp. 149–152, 2017.

[4] P. J. Cutler, M. D. Malik, S. Liu, J. M. Byars, D. S. Lidke, and K. A. Lidke, "Multicolor quantum dot tracking using a high-speed hyperspectral line-scanning microscope," *PLoS ONE*, vol. 8, pp. e64 320, doi:10.1371/journal.pone.0 064 320, 2013.

[5] Y. Garini, I. T. Young, and G. McNamara, "Spectral imaging: principles and applications," *Cytometry Part A*, vol. 69A, pp. 735–747, 2006.

[6] T. Zimmermann, J. Marrison, K. Hogg, and P. O'Toole, "Clearing up the signal: spectral imaging and linear unmixing in fluorescence microscopy," *Methods in Molecular Biology*, vol. 1075, pp. 129–148, 2014.

[7] P. Favreau, C. Hernandez, A. S. Lindsey, D. F. Alvarez, T. Rich, P. Prabhat, and S. J. Leavesley, "Thin-film tunable filters for hyperspectral fluorescence microscopy," *Journal of Biomedical Optics*, vol. 19, pp. 011 017, 1–11, 2014.

[8] P. Favreau, C. Hernandez, T. Heaster, D. F. Alvarez, T. Rich, P. Prabhat, and S. J. Leavesley, "Excitation scanning hyperspectral-imaging microscope," *Journal of Biomedical Optics*, vol. 19, pp. 046 010, 1–11, 2014.

[9] N. Gat, "Imaging spectroscopy using tunable filters: a review," *Proceedings of the SPIE*, vol. 4056, pp. 50–64, 2000.

[10] E. Stack, C. Wang, K. A. Roman, and C. C. Hoyt, "Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis," *Methods*, vol. 70, pp. 46–58, 2014.

[11] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 529–545, 2001.

[12] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 59–69, 2013.

[13] J. R. Mansfield, C. Hoyt, and R. M. Levenson, "Visualization of microscopy-based spectral imaging data from multi-label tissue sections," *Current Protocols in Molecular Biology*, vol. 84, pp. 14.19.1–14.19.15, 2008.

[14] S. E. Bialkowski, "Overcoming the multiplex disadvantage by using maximum-likelihood inversion," *Applied Spectroscopy*, vol. 52, pp. 591–598, 1998.

[15] D. R. Fuhrmann, C. Preza, J. A. O'Sullivan, and D. L. Snyder, "Spectrum estimation from quantum limited interferograms," *IEEE Transactions on signal processing*, vol. 52, pp. 950–961, 2004.

[16] F. Fereidouni, A. N. Bader, and H. C. Gerritsen, "Spectral phasor analysis allows rapid and reliable unmixing of fluorescence microscopy spectral images," *Optics Express*, vol. 20, 2012.

[17] C. R. Rao, *Linear statistical inference and its applications.* Wiley, New York, USA., 1965.

[18] R. Neher and E. Neher, "Optimizing imaging parameters for the separation of multiple labels in a fluorescence image," *Journal of Microscopy*, vol. 213, pp. 46–62, 2004.

[19] M. Cubero-Castan, J. Chanussot, X. Briottet, M. Shimoni, and V. Achard, "An unmixing-based method for the analysis of thermal hyperspectral images," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 7859–7863.

[20] A. Esposito, M. Popleteeva, and A. R. venkitaraman, "Maximizing the biochemical resolving power of fluorescence microscopy," *PLoS ONE*, vol. 8, p. e77392, 2013.

[21] J. B. Pawley, *Handbook of biological confocal microscopy*, 3rd ed. Springer, 2006.

[22] S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidimensional Systems and Signal Processing*, vol. 17, pp. 27–58, 2006.

[23] S. Ram, "Resolution and localization in single molecule microscopy," Ph.D. dissertation, University of Texas at Arlington/University of Texas Southwestern Medical Center at Dallas, 2007.

[24] J. Chao, E. S. Ward, and R. J. Ober, "Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices," *Multidimensional Systems and Signal Processing*, vol. 23, pp. 349–379, 2012.

[25] M. Kimmel and D. E. Axelrod, *Branching processes in biology.* Springer, New York, USA, 2002.

[26] J. Chao, S. Ram, E. S. Ward, and R. J. Ober, "Ultrahigh accuracy imaging modality for super-localization microscopy," *Nature Methods*, vol. 10, pp. 335–338, 2013.

[27] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober, "Quantitative study of single molecule location estimation techniques," *Optics Express*, vol. 17, pp. 23 352–23 373, 2009.

[28] G. McNamara, A. Gupta, J. Reynaert, T. D. Coates, and C. Boswell, "Spectral imaging microscopy web sites and data," *Cytometry Part A*, vol. 5, pp. 121–132, 2006.

[29] J. Lakowicz, *Principles of fluorescence spectroscopy*. Springer USA, 2006.

[30] Y. Li, G. Dong, M. Peng, L. Wondraczek, and J. Qiu, "Anti-stokes fluorescent probe with incoherent excitation," *Scientific Reports*, vol. 4:4059, pp. 1–6, 2014.

[31] P. S. K., S. Sreedharan, H. Singh, N. H. Green, C. Smythe, J. A. Thomas, and A. Das, "Imaging cellular trafficking processes in real time using lysosome targeted up-conversion nanoparticles," *Chemical Communications*, vol. 53, pp. 12 672–12 675, 2017.

[32] L. Zhou, R. Wang, C. Yao, X. Li, C. Wang, X. Zhang, C. Xu, A. Zeng, D. Zhao, and F. Zhang, "Single-band upconversion nanoprobes for multiplexed simultaneous in situ molecular mapping of cancer biomarkers," *Scientific Reports*, vol. 6:6938, pp. 1–12, 2014.

[33] A. Gnach and A. Bednarkiewicz, "Lanthanide-doped up-converting nanoparticles: merits and challenges," *Nano Today*, vol. 47, pp. 532–563, 2012.

[34] S. Wilhelm, "Perspectives for upconverting nanoparticles," *ACS Nano*, no. 11, pp. 10 644–10 653, 2017.

[35] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single molecule microscopy," *Biophysical Journal*, vol. 86, pp. 1185–1200, 2004.