A Parametric Estimation Approach to Instantaneous Spectral Imaging

Figen S. Oktem, *Student Member, IEEE*, Farzad Kamalabadi, *Member, IEEE*, and Joseph M. Davila

Abstract—Spectral imaging, the simultaneous imaging and spectroscopy of a radiating scene, is a fundamental diagnostic technique in the physical sciences with widespread application. Due to the intrinsic limitation of two-dimensional (2D) detectors in capturing inherently three-dimensional (3D) data, spectral imaging techniques conventionally rely on a spatial or spectral scanning process, which renders them unsuitable for dynamic scenes. In this paper, we present a nonscanning (instantaneous) spectral imaging technique that estimates the physical parameters of interest by combining measurements with a parametric model and solving the resultant inverse problem computationally. The associated inverse problem, which can be viewed as a multiframe semiblind deblurring problem (with shift-variant blur), is formulated as a maximum a posteriori (MAP) estimation problem since in many such experiments prior statistical knowledge of the physical parameters can be well estimated. Subsequently, an efficient dynamic programming algorithm is developed to find the global optimum of the nonconvex MAP problem. Finally, the algorithm and the effectiveness of the spectral imaging technique are illustrated for an application in solar spectral imaging. Numerical simulation results indicate that the physical parameters can be estimated with the same order of accuracy as state-of-the-art slit spectroscopy but with the added benefit of an instantaneous, 2D field-of-view. This technique will be particularly useful for studying the spectra of dynamic scenes encountered in space remote sensing.

Index Terms—Computational spectral imaging, imaging spectroscopy, inverse methods, dynamic programming, multiframe image deblurring, parameter estimation of superimposed signals, separable nonlinear least squares problems.

I. INTRODUCTION

OBSERVING the spectrum of a radiating scene, known as *spectral imaging* or *imaging spectroscopy*, is a fundamental diagnostic technique in the physical sciences with application in diverse fields such as physics, chemistry, biology, medicine, astronomy, and remote sensing. The measured spectrum enables sensing the properties of the scene

and is the main source of information about the chemical composition and physical properties of targeted objects. For example, in astrophysical imaging of space plasmas, energy transitions of the constituent matter in the plasma produce spectral emission lines, and measurements of the emitted spectrum provide essential information for inferring the plasma parameters (such as density, temperature, and velocity). Such measurements acquired as a function of time enable the investigation of the complex plasma behavior by revealing how currents and heat flow through the plasma [1], [2].

The objective of spectral imaging is to form images of a scene as a function of wavelength. For a two-dimensional (2D) scene, this requires obtaining a three-dimensional (3D) dataset: one for spectral and two for spatial dimensions. However, obtaining this 3D dataset with inherently 2D detectors pose intrinsic limitations on the spatio-spectral extent of the technique.

To address this limitation, conventional spectral imaging techniques rely on a scanning process to build up the 3D *data cube* from a series of 2D measurements that are acquired sequentially. Typically this data cube is acquired using a spectrometer with a long slit scanning the scene spatially or by using an imager with a series of spectral filters scanning the scene spectrally. In the former case (referred to as rastering or pushbroom) only a thin slice of the scene is observed at a time, whereas in the latter case only one spectral band is observed at a time. Similarly, Fourier and Hadamard transform-based spectrometers perform scanning in a transformed domain (through their movable parts) to build up the 3D data cube [3]. As a result, these conventional methods are best suited for spectral imaging of scenes that remain stationary during the scanning process involved. For dynamic scenes, however, the conventional methods are subject to temporal artifacts.

To effectively operate with non-static scenes, methods that reconstruct the 3D data cube from a single-shot measurement recently have been proposed based on *tomographic* [4]–[8], and *compressive sensing* [9]–[12] approaches. The main idea in the *tomographic* approach is to construct the 3D data cube from its 2D projections, where the projections are obtained through spectrally dispersed images of the scene [4], each dispersed in a different direction and by a different amount (through different diffraction orders of an optical grating). These tomographic techniques, known as *computed tomography imaging spectrometers* (CTIS), require a large set of projections (i.e. dispersed images) to be captured at once, hence demanding a large detector area to be used. The resulting cost generally limits either the field-of-view (FOV) or the resolution. In addition, these techniques suffer
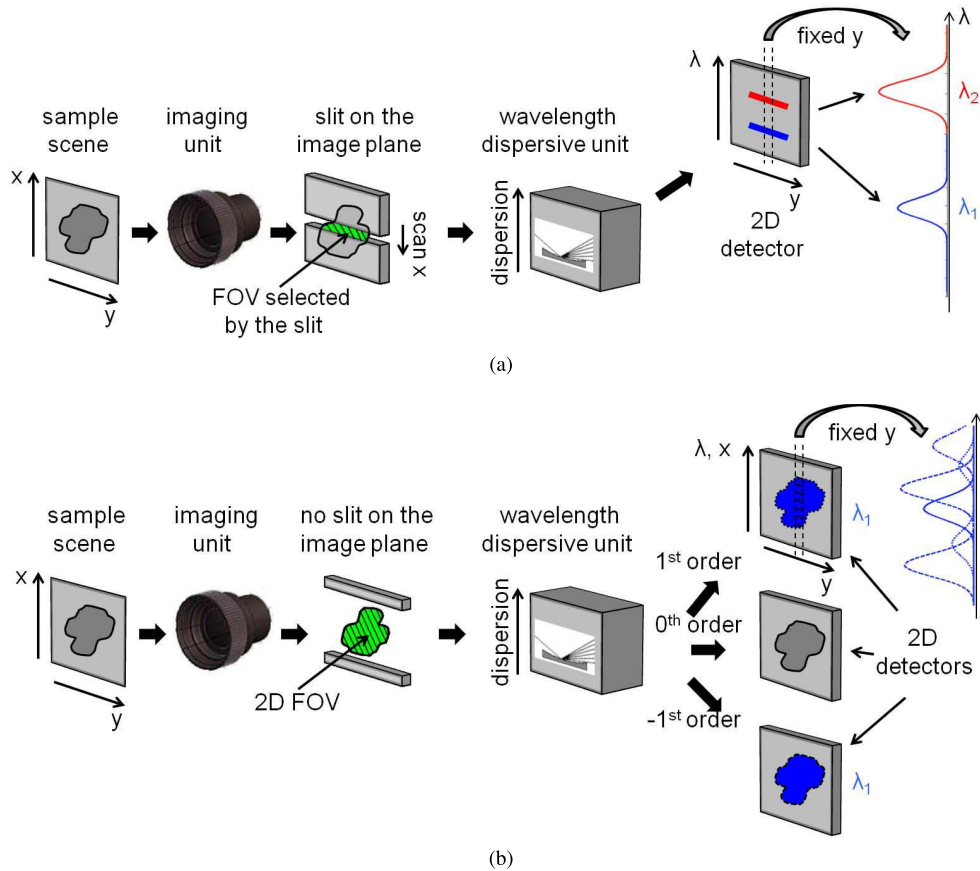
(a)

(b)

Fig. 1. Comparison of (a) the conventional slit spectrometer with (b) instantaneous spectral imager presented in this paper. In both cases, an imaging unit (e.g. lens, mirror) focuses a 2D scene on an image plane. In a slit spectrometer (a), a narrow slit lies on the image plane to limit the FOV to a 1D portion of the scene. The light that passes through the slit is input to a wavelength dispersive unit, which generally consists of collimator optics (e.g., lens), a dispersive element (e.g., diffraction grating), and focusing optics (e.g., lens). Each spectral line in the incoming light is dispersed according to the wavelength and imaged onto a 2D detector. To obtain the spectrum of the entire 2D scene, the slit is moved within the image plane to scan the scene spatially. For the instantaneous spectral imager (b), the slit is widened to achieve an instantaneous 2D FOV. This causes an overlap of spatial and spectral information on the detector: dispersed spectral lines from all spatial positions within the FOV are now superimposed. To decompose this multiplexed data computationally, multiple dispersed images with different diffraction orders are recorded using multiple detectors (in this case three detectors for the orders +1, 0, and −1).

from their incapability of fully sampling the Fourier domain representation of the data cube, which results from the limited angle of projections and the projection-slice theorem [5]. The unsampled Fourier volume, known as the *missing cone*, makes unique reconstruction of the data cube impossible. Consequently, the resulting tomographic reconstruction would suffer from artifacts unless the data deficiency, arising from the missing cone problem, is fully compensated for with additional prior knowledge.

The *compressive sensing* based single-shot approaches, known as *coded aperture snapshot spectral imaging* (CASSI) techniques [9]–[12], exploit the prior knowledge resulting from the assumed sparsity of the spectral data cube. These approaches acquire only a single projection with a coded aperture while simultaneously requiring that the data cube be sparse in some transform domain. Although the imposed sparsity assumption may not always hold, for the cases where such restriction holds, the 3D data cube can be reconstructed from the 2D coded measurement using compressive sensing methods. Several other compressive spectral imagers [13]–[17] also exist for compressive acquisition of the spectral data cube with different sensing architectures. Although these

imagers require multiple successive acquisitions, instead of a single-shot measurement, rendering them not ideal for dynamic scenes, they offer advantages over their conventional counterparts such as improved reconstruction and resolution and reduced acquisition time and storage requirements.

In this paper, we develop a *parametric* single-shot approach to achieve instantaneous (non-scanning) spectral imaging. This approach is specifically designed for scenes that have spectra consisting of discrete lines, which are encountered in various applications [18]–[22] involving astrophysical imaging of space plasmas [1], [2] and atmospheric physics [3], [23]–[25]. Figure 1b depicts a schematic of the observing system involved. This system is quite similar to a conventional slit (pushbroom) spectrometer illustrated in Fig. 1a, but, unlike a slit spectrometer, this system has an instantaneous 2D FOV (rather than a 1D FOV limited by a slit). Also, because spectrally dispersing the 2D input image causes an overlap of spatial and spectral information on the 2D detector, *multiple* dispersed images of the scene are acquired simultaneously in order to gather the needed information for decomposing this multiplexed data.

The system involved in this parametric approach is similar to a CTIS system used in the snapshot tomographic approach [4]–[8], but our technique relies on a parametric approach to reconstruct the 3D data cube from significantly smaller number of dispersed images, i.e., projections, hence allowing a smaller detector area. We note that this task is accomplished by compensating for the resulting missing cone with the prior knowledge that the spectra consist of discrete lines and incorporating this knowledge into the image formation framework via a parametric model.

The system involved in this parametric approach is also similar to a CASSI system [10] with a single disperser, but different in that the CASSI system uses a coded aperture to additionally modulate the imaged 2D scene. While the CASSI system uses the sparsity of the data cube accompanied by the coded measurements for the reconstruction, our method instead uses a parametric reconstruction approach that exploits the specific structure of the spectra and measurements acquired at multiple diffraction orders.

The focus of this paper is the problem of estimating the physical parameters of interest from the measurements of this instantaneous spectral imager by using a parametric model for the measurements. Based on this parametric model, the estimation problem can be viewed as a 1D multiframe semi-blind deblurring problem with shift-variant blur, where multiple blurred images of the same scene are obtained through multiple dispersed images, each with a different diffraction order. We formulate the inverse problem as a maximum *a posteriori* (MAP) estimation problem since in many such experiments prior statistical knowledge of the physical parameters can be well estimated. The resulting nonconvex MAP problem is solved by developing an efficient dynamic programming algorithm, which is an extension of a previously proposed algorithm for maximum likelihood parameter estimation of superimposed signals [26], [27]. The developed algorithm yields parameter estimates that are close to the global optimum of the MAP problem. A local optimization algorithm initialized with these estimates can then be used to obtain the desired global optimum. Through numerical investigations, we verify the results of a Cramer-Rao bound analysis in [28] and [29] demonstrating that the physical parameters can be estimated with the same order of accuracy as the conventional slit spectroscopy, but with the added benefit of an instantaneous 2D FOV.

The paper is organized as follows. In Section II, we introduce the parametric forward model (for the dispersed images). The inverse problem is formulated in Section III as a MAP problem. Section IV presents the dynamic programming algorithm for efficiently solving the MAP problem. Numerical simulation results for an application in solar spectral imaging are presented in Section V. Section VI concludes the paper.

II. PARAMETRIC FORWARD PROBLEM

In a slit spectrometer (see Fig. 1a), a narrow slit lies on the image plane of a lens or mirror, hence limiting the field-of-view to a 1D portion of the scene. As a result, only a thin slice of the scene is observed at a time. The light

that passes through the slit then enters into a wavelength dispersive unit where each spectral line in the incoming light is dispersed according to wavelength and imaged onto a 2D detector [3]. Because the dispersion plane is aligned to be perpendicular to the long side of the slit, one dimension in the detector corresponds only to the wavelength (λ) whereas the other dimension corresponds to the spatial dimension (y) admitted through the slit. Hence the spatial and spectral information do not overlap on the detector, and each dispersed spectral line observed at the detector is associated with a single position on the scene. To obtain the spectral information of an entire 2D scene, the scene is scanned spatially using the slit, i.e., the 1D instrument FOV is pointed to a series of adjacent spatial positions (x) on the scene with a narrow slit exposure taken at each pointing location. This approach is not suitable for dynamic scenes which evolve on timescales faster than the scanning process involved. For example, in solar spectroscopy, the scanning takes on the order of minutes (to cover an active/dynamic region of interest) whereas the physical processes occurring in solar plasma change on the order of seconds [6].

To overcome this limitation with an instantaneous spectral imager, the width of the entrance slit is increased to obtain an instantaneous *2D* FOV (see Fig. 1b), such that a 2D image of the scene is now allowed at the image plane of the imaging unit. Then the light from the 2D scene enters into the dispersive unit where each spectral line in the incoming beam is dispersed and imaged onto a 2D detector. Because the input to the dispersive system is now a 2D image, dispersion causes an overlap of spatial and spectral information on the detector. More specifically, dispersed spectral lines from all positions along the spatial dimension parallel to the dispersion plane are now superimposed at the output. To overcome the difficulty of decomposing this multiplexed data, *multiple* spectrally dispersed images of the 2D scene are recorded simultaneously using multiple detectors. These dispersed images differ by the amount and direction of dispersion as determined by different diffraction (spectral) orders. In particular, a negative diffraction order indicates the reversal of the dispersion direction, and higher diffraction orders indicate larger amounts of dispersion.

A. Model Assumptions

First we list the assumptions that are used in the development of the parametric forward model.

- The spectrum is composed of discrete spectral lines, and each spectral line has a Gaussian shape due to thermal broadening [2], [21], [30].
- The medium of interest is optically thin [2]; hence scattering and absorption effects are negligible, and do not contribute to the shape of the spectrum.
- Continuum background is negligible.
- Each dispersed image is monochromatic, hence resulting from a single spectral line with known central wavelength λ_0.

The application domains where these assumptions commonly hold are various [18]–[22] including astrophysical imaging of
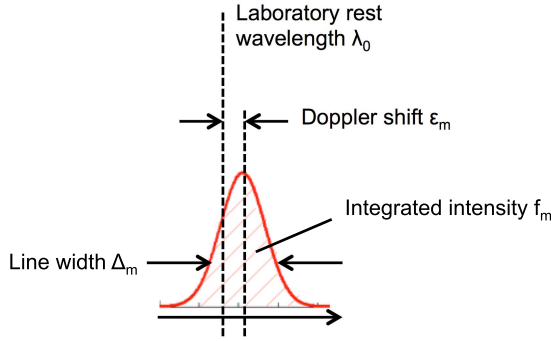
Fig. 2. Spectrum at a single spatial location. In the parametric model, the spectrum at the mth spatial location is modeled by a Gaussian spectral line, and is fully characterized by the three parameters: integrated line intensity f_m, a line width Δ_m, and a Doppler shift ϵ_m.
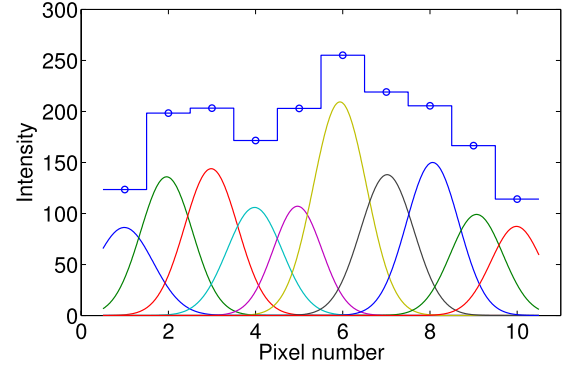

Fig. 3. Illustration of one of the columns of the simulated dispersed image (in a ten-pixel detector column). Bars correspond to the total intensity observed at each pixel, and colored spectral lines show the contributions from different pixels.

space plasmas [1], [2] and atmospheric physics [3], [23]–[25].

Here the third assumption requires that either the continuum background is relatively small compared to the sum of the contributions of the dispersed spectral lines, or it can be subtracted from the measurements through pre-processing. Moreover, the last assumption will be approximately true if one spectral line is dominant (i.e. strongest) in the passband of the instrument (i.e. the spectrum filtered by a spectral filter is dominated by a single spectral line), and if that does not hold, the FOV can be made small enough such that monochromatic dispersed images of the scene, resulting from adjacent spectral lines in the spectrum, do not overlap on the detector. This will allow us to use each monochromatic image separately for the reconstruction as before. Hence here we restrict our formulation to the cases where the spectrum at each spatial position consists of only one Gaussian spectral line. Our goal in the inversion is then to estimate the parameters of the Gaussian spectral lines at all spatial positions within the 2D FOV.

B. Parametric Image Formation Model

In the parametric model, the spectrum at each spatial location is modeled by a Gaussian spectral line. With this assumption, the dispersed spectral line from pixel m' can be fully characterized by three parameters of interest: an integrated line intensity (strength) $f_{m'}$, a line width (broadening) $\Delta_{m'}$, and a line center shift (Doppler shift) $\epsilon_{m'}$ (see Fig. 2). Here the Gaussian assumption is due to thermal broadening [2], [21], [30]. The line width is the result of the thermal motions of the emitting particles along the line of sight in the radiating scene such that the width is proportional to the temperature of the emitters. Doppler shifts in wavelength (variations in the line center position with respect to the central wavelength λ_0) are associated with coherent flows along the line of sight.

Dispersed images are then expressed as superposition of dispersed spectral lines from different spatial positions on the scene. If the dispersion plane of the dispersive unit is aligned to be parallel to the columns of pixels on the detector (as illustrated in Fig. 1b), then spectra from neighboring columns are not mixed; that is, in the dispersed image, spectra from neighboring

spectral lines only from positions along a single column are superimposed. This allows us to treat the 2D problem as a 1D problem where each column of the dispersed image is modeled independently.

Then considering a column of pixels of length M, the dispersed spectral lines from all of these pixels are superimposed; hence the observed intensity at any detector pixel is the sum of contributions from all of these spectral lines. Fig. 3 illustrates this superposition on a single column of pixels.

Consequently, the observed intensity at any detector pixel is given by the sum of contributions from all of these parametric spectral lines. In particular, the contribution of each spectral line (to the total intensity observed at a pixel) is given by the integrated intensity of the spectral line over that pixel. Let x' be a continuous variable in pixel units that denotes the vertical location on the detector, and let the mth detector pixel correspond to the range $m - 1/2 \leq x' < m + 1/2$, where $m = 1, \ldots, M$. Then the contribution of the spectral line at pixel m' to the total intensity at pixel m can be found by integrating the Gaussian spectral line over the mth pixel:

$$c_{m'} = \int_{m-1/2}^{m+1/2} \frac{f_{m'}}{\sqrt{2\pi}|a|\Delta_{m'}} e^{-\frac{(x'-m'-a\,\epsilon_{m'})^2}{2(a\Delta_{m'})^2}} dx', \quad (1)$$

$$= \frac{f_{m'}}{\sqrt{2\pi}|a|\Delta_{m'}} \int_{-1/2}^{1/2} e^{-\frac{(x'+m-m'-a\,\epsilon_{m'})^2}{2(a\Delta_{m'})^2}} dx', \quad (2)$$

$$= f_{m'} \frac{\mathrm{erf}(t_2) - \mathrm{erf}(t_1)}{2}, \quad (3)$$

where the error function $\mathrm{erf}(t)$ is

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x'^2) dx', \quad (4)$$

$$t_{1,2} = \frac{m - m' \mp 1/2 - a\epsilon_{m'}}{\sqrt{2}\,|\,a\,|\,\Delta_{m'}}, \quad (5)$$

and both line width $\Delta_{m'}$ and Doppler shift $\epsilon_{m'}$ are measured in pixel units and in the first diffraction order ($a = +1$). At higher orders, these are scaled by the spectral order a because the physical spread of any spectral interval on the detector is enlarged by the order a (since the dispersion amount varies linearly with order) [31]. Also a negative sign in the order indicates reversal of the direction of dispersion, which affects the relative direction of line center shift (Doppler shift).

Then the total intensity at the detector pixel m, denoted by y_m^a, is given by the sum of contributions from all spectral lines at pixels $m' = 1, \ldots, M$:

$$y_m^a = \sum_{m'=1}^{M} f_{m'} \frac{\text{erf}(t_2) - \text{erf}(t_1)}{2} \quad \text{if } a \neq 0, \tag{6}$$

for all $m = 1, \ldots, M$. This is true for any dispersed image with order $a \neq 0$. For the zero order image, which is the result of direct imaging without any dispersion, the total intensity at each pixel is simply the integrated line intensity of the spectral line at that pixel:

$$y_m^0 = f_m. \tag{7}$$

Subsequently, the complete parametric model for the intensities of the ath order image can be expressed as

$$y_m^a = \sum_{m'=1}^{M} f_{m'} \phi_{m-m'}^a(\boldsymbol{\Theta}_{m'}), \tag{8}$$

where $\boldsymbol{\Theta}_{m'} = [\epsilon_{m'}, \Delta_{m'}]$, and the contribution amount $\phi_{m-m'}^a(\boldsymbol{\Theta}_{m'})$ is

$$\phi_{m-m'}^a(\boldsymbol{\Theta}_{m'}) = \begin{cases} \dfrac{\text{erf}(t_2) - \text{erf}(t_1)}{2} & \text{if } a \neq 0; \\ \delta_{m-m'} & \text{if } a = 0, \end{cases} \tag{9}$$

with δ_m denoting the Kronecker delta function. The amount of contributions from superimposed signals are known up to the parameters $\boldsymbol{\Theta}_{m'}$ and contain the nonlinearity in the model.

The unknowns in this model are the spectral line parameters $\boldsymbol{\Psi}_{m'} = [\boldsymbol{\Theta}_{m'}, f_{m'}]$ satisfying the constraints $\boldsymbol{\Psi}_{m'} \in \Omega$. The constraint set Ω equals to $\Lambda \times \Pi$, where Λ and Π denote the constraint sets for $\boldsymbol{\Theta}_{m'}$ and $f_{m'}$, respectively: $\Lambda = \{(\epsilon_{m'}, \Delta_{m'}) \in \mathbb{R} \times \mathbb{R}^+ : |\epsilon_{m'}| \leq \epsilon_{\max} \text{ and } \Delta_{\min} \leq \Delta_{m'} \leq \Delta_{\max}\}$, and $\Pi = \{f_{m'} \in \mathbb{R}^+ : f_{\min} \leq f_{m'} \leq f_{\max}\}$. (The number of superimposed signals is known, and equal to the number of pixels M.)

If we define the vectors $\mathbf{y}^a = [y_1^a \ldots y_M^a]^\top$, $\mathbf{f} = [f_1 \ldots f_M]^\top$, $\boldsymbol{\epsilon} = [\epsilon_1 \ldots \epsilon_M]^\top$, and $\boldsymbol{\Delta} = [\Delta_1 \ldots \Delta_M]^\top$ (with the superscript $^\top$ denoting the transpose of a vector), each of these vectors has the size M. Then, based on the parametric model, each dispersed image $\mathbf{y}^a$ can be viewed as a blurred version of the same input image $\mathbf{f}$ with a different spatially-varying filter of unknown parameters $\boldsymbol{\epsilon}$ and $\boldsymbol{\Delta}$. On the other hand, when the order is zero, there is no dispersion and hence no blur on the input image; that is $\mathbf{y}^0 = \mathbf{f}$.

C. Observation Model With Noise

In vector-matrix form, the observation model with noise is given by

$$\tilde{\mathbf{y}}^a = \sum_{m'=1}^{M} f_{m'} \mathbf{h}_{m'}^a(\boldsymbol{\Theta}_{m'}) + \mathbf{n}^a, \tag{10}$$

where

$$\mathbf{h}_{m'}^a(\boldsymbol{\Theta}_{m'}) = [\phi_{1-m'}^a(\boldsymbol{\Theta}_{m'}), \ldots, \phi_{M-m'}^a(\boldsymbol{\Theta}_{m'})]^\top, \tag{11}$$

and $\mathbf{n}^a = [n_1^a \ldots n_M^a]^\top$ is the noise vector with $n_m^a \sim N(0, \sigma_a^2)$ representing white Gaussian noise that is uncorrelated across

both different pixels m and orders a. In practice this noise model is valid when the following conditions are satisfied: first, photon noise is the dominant source of noise in the measurements rather than the thermal and readout noise (which can be ensured by sufficiently long integration time); second, a strong spectral line is measured through all pixels (so that the values of y_m^a are large enough to well approximate the poisson noise as Gaussian noise); third, signal-to-noise ratio (SNR) is sufficiently high (such that noise standard deviation can be approximated as constant over each dispersed image).

Here we are interested in the case that multiple dispersed images at different diffraction orders are simultaneously available. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ be the set of all orders that are measured with N being the number of different orders, and $\{\sigma_{a_1}, \sigma_{a_2}, \ldots, \sigma_{a_N}\}$ be the corresponding noise standard deviations for these measurements. Defining the $M \times M$ matrix

$$\mathbf{H}^a(\boldsymbol{\Theta}) = [\mathbf{h}_1^a(\boldsymbol{\Theta}_1), \ldots, \mathbf{h}_M^a(\boldsymbol{\Theta}_M))], \tag{12}$$

with $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1^\top, \ldots, \boldsymbol{\Theta}_M^\top]^\top$ (i.e. $\boldsymbol{\Theta} = [\boldsymbol{\epsilon}, \boldsymbol{\Delta}]$), the observation model for each order a can be rewritten more compactly as

$$\tilde{\mathbf{y}}^a = \mathbf{H}^a(\boldsymbol{\Theta}) \, \mathbf{f} + \mathbf{n}^a. \tag{13}$$

Then by stacking all measured dispersed images into a single vector, $\tilde{\mathbf{y}}$, the complete model becomes

$$\tilde{\mathbf{y}} = \mathbf{H}(\boldsymbol{\Theta})\mathbf{f} + \mathbf{n}, \tag{14}$$

where

$$\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{\mathbf{y}}^{a_1} \\ \tilde{\mathbf{y}}^{a_2} \\ \vdots \\ \tilde{\mathbf{y}}^{a_N} \end{bmatrix}, \quad \mathbf{H}(\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{H}^{a_1}(\boldsymbol{\Theta}) \\ \mathbf{H}^{a_2}(\boldsymbol{\Theta}) \\ \vdots \\ \mathbf{H}^{a_N}(\boldsymbol{\Theta}) \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}^{a_1} \\ \mathbf{n}^{a_2} \\ \vdots \\ \mathbf{n}^{a_N} \end{bmatrix}.$$

III. INVERSE PROBLEM

In the inverse problem, the goal is to estimate the unknown spectral line parameters $\mathbf{f}$ and $\boldsymbol{\Theta}$ from the measurements $\tilde{\mathbf{y}}$ based on the model (14). This problem can be viewed as a *multiframe*, *semi-blind* deblurring problem with *shift-variant* blur. Here *semi-blind* refers to the fact that although the parametric form of each blur is known, the blur parameters $\boldsymbol{\Delta}$ and $\boldsymbol{\epsilon}$ are unknown and must therefore be estimated jointly with the original image $\mathbf{f}$. The term *multiframe* refers to the availability of multiple blurred images of the same object through different diffraction orders.

We formulate the inverse problem as a maximum *a posteriori* (MAP) estimation problem that incorporates prior knowledge of the statistics of the spectral line parameters. Such priors can be obtained from the measurements of existing conventional pushbroom spectrometers, as will be illustrated in Section V-A. Incorporation of prior information helps to regularize the inverse problem, hence prevents noise amplification that would result from over fitting to the noisy data.

Treating the parameter vectors $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ as *independent* random vectors, the MAP estimates of $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ from the measurements $\tilde{\mathbf{y}}$ are given by

$$\arg \max_{\substack{\mathbf{f} \in \Pi^M \\ [\boldsymbol{\epsilon}, \boldsymbol{\Delta}] \in \Lambda^M}} p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon}) \, p(\mathbf{f}) p(\boldsymbol{\Delta}) p(\boldsymbol{\epsilon}), \tag{15}$$

where $p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ represents the conditional probability density function (pdf) of $\tilde{\mathbf{y}}$ given $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ (equivalently, the likelihood function of the unknown parameters), and $p(\mathbf{f})$, $p(\boldsymbol{\Delta})$, and $p(\boldsymbol{\epsilon})$ denote the prior distributions. These prior distributions specify the probability of each parameter independently from the observed data; hence they describe the information we have on each parameter prior to observing the data.

On the other hand, the conditional pdf $p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ comes from the noisy observation model in (10) and (14), and has the following form:

$$p(\tilde{\mathbf{y}} \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon}) = \prod_{i=1}^{N} \frac{1}{(\sqrt{2\pi} \, \sigma_{a_i})^M} e^{-\frac{1}{2\sigma_{a_i}^2} \|\tilde{\mathbf{y}}^{a_i} - \mathbf{H}^{a_i}(\boldsymbol{\Theta})\mathbf{f}\|_2^2}. \quad (16)$$

The role of this pdf in the estimation problem is to force the estimates of $\mathbf{f}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\epsilon}$ to match the observation model closely.

Note that a noisy observation of $\mathbf{f}$ is available through the zeroth order image, $\tilde{\mathbf{y}}^0$. This gives an immediate statistical model for $\mathbf{f}$. Hence if $\tilde{\mathbf{y}}^0$ is observed at a sufficiently high SNR such that the conditional pdf $p(\tilde{\mathbf{y}}^0 \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ is more concentrated around the true value of $\mathbf{f}$ as compared to the prior of $\mathbf{f}$ (yielding the prior $p(\mathbf{f})$ to be effectively constant where the conditional pdf $p(\tilde{\mathbf{y}}^0 \mid \mathbf{f}, \boldsymbol{\Delta}, \boldsymbol{\epsilon})$ is nonzero), then the prior of $\mathbf{f}$ is not necessary to yield a useful estimate. For this reason, we ignore $p(\mathbf{f})$ in the MAP formulation, which yields a simpler form of a separable nonlinear least squares problem [32], and to a more efficient estimation algorithm (as will be discussed in the next section). However, we note that if it were desired to keep the prior $p(\mathbf{f})$ in the MAP formulation, a dynamic programming algorithm could still be derived to solve the MAP problem, but owing to less efficiency.

For the priors $p(\boldsymbol{\Delta})$ and $p(\boldsymbol{\epsilon})$, we assume that parameters at different pixels are independently distributed. After combining all of these together, and taking the logarithm of (15), the MAP estimation problem becomes

$$\min_{\substack{\mathbf{f} \in \Pi^M \\ [\boldsymbol{\epsilon}, \boldsymbol{\Delta}] \in \Lambda^M}} \left\{ \sum_{i=1}^{N} \frac{1}{2\sigma_{a_i}^2} \|\tilde{\mathbf{y}}^{a_i} - \mathbf{H}^{a_i}(\boldsymbol{\Theta})\mathbf{f}\|_2^2 \right.$$
$$\left. - \sum_{m=1}^{M} (\log p(\Delta_m) + \log p(\epsilon_m)) \right\}. \quad (17)$$

We can express this as

$$\min_{\substack{\mathbf{f} \in \Pi^M \\ \boldsymbol{\Theta} \in \Lambda^M}} \left\{ \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta})\mathbf{f}\|_W^2 + \sum_{m=1}^{M} \Gamma(\boldsymbol{\Theta}_m) \right\}, \quad (18)$$

where $\Gamma(\boldsymbol{\Theta}_m) = -2\log p(\Delta_m) - 2\log p(\epsilon_m)$ is the regularization functional arising from priors, and W denotes the weighted norm in (17) where for each diffraction order the sum of squared residuals is weighted with the reciprocal of the noise variance at that order. That is, in vector-matrix form, $\|.\|_W^2 = (.)^* \mathbf{W}(.)$ with $\mathbf{W}$ being the inverse covariance matrix for the data, and superscript $*$ denoting the conjugate transpose of a vector.

This inverse problem belongs to the class of *separable nonlinear least squares problems* [32] with regularization. In these problems the observation model is given by a linear combination of parametrically prespecified nonlinear functions. The parameters of interest can be grouped into two categories: parameters that affect the observations in a linear fashion ($\mathbf{f}$ in our case), and parameters that affect the observations in a nonlinear fashion ($\boldsymbol{\Theta}$ in our case). We note that implicit in the formulation of the estimation problem is that (18) has a unique global minimum.

IV. DYNAMIC PROGRAMMING (DP) ALGORITHM

We now focus on developing an efficient and globally converging algorithm for solving our MAP problem. Note that this MAP estimation problem requires solving a nonlinear and nonconvex optimization problem with possibly local minima. (Nonconvexity of the problem is apparent since the Hessian matrix of the objective function is not positive semidefinite for all feasible points [33].) This can create difficulty in efficiently finding the global minimum of the problem using local optimization methods such as gradient-descent type methods [34] and expectation-maximization type algorithms [35]. This is because such methods can converge only to one of the local minimums depending on the initialization. Many of the methods proposed for the general separable nonlinear least squares problems [32] and for problems involving superimposed signals (see [36]–[38]), are such local optimization methods.

While efficient local optimization methods may suffer from convergence to local minima, global optimization methods can guarantee convergence to the global solution. There are two types of *general-purpose* global optimization methods: deterministic versus stochastic. The deterministic global optimization methods (such as exhaustive search, branch and bound method [39]) can guarantee convergence to a global solution within a certain tolerance value; but it is not practical to employ these methods in most applications because of their high computational complexity [40]. On the other hand, stochastic methods (such as simulated annealing) lower the computational cost in return for weaker guarantees for global convergence (in probabilistic sense) [41].

Here we develop an efficient global optimization method, *specialized* for our problem, which combines the strengths of deterministic and stochastic (global optimization) methods: global convergence guarantee within a certain tolerance value, as the deterministic approaches, and lower computational complexity, as the stochastic approaches. The key idea in this method is to perform a computationally efficient search that is equivalent to exhaustive search by exploiting the special form of the objective function to optimize, which arises from the limited interaction of superimposed signals (i.e. Gaussian line profiles) with few of their closest neighbors. This special form, so-called as *local interaction* in [26], yields to a Markovian-like property of the globally optimal solutions (in the sense of deterministic dependence, rather than statistical dependence) allowing us to break the optimization problem into smaller subproblems. The resulting dynamic programming (DP) algorithm has a computational cost that is linear in the number

of superimposed signals as opposed to the exponential cost in exhaustive search.

A. Local Interaction Signal Model

The development of the dynamic programming algorithm relies on one major assumption: local interaction, implying that each superimposed signal interacts (overlaps) with only a few of its closest neighbors. Let $r \geq 1$ be the number of closest neighbors with which each superimposed signal overlaps on both sides. Then, mathematically the *local interaction model* [26] is expressed as

$$\mathbf{h}_i(\mathbf{\Theta}_i)^* \mathbf{h}_j(\mathbf{\Theta}_j) \approx 0 \quad \text{for } |i - j| > r. \tag{19}$$

This requires that the ith and jth columns of $\mathbf{H}(\mathbf{\Theta})$, denoted by $\mathbf{h}_i(\mathbf{\Theta}_i)$ and $\mathbf{h}_j(\mathbf{\Theta}_j)$ (associated with the ith and jth superimposed signals), are approximately orthogonal if they are separated by more than r columns. (Note that the ith column of $\mathbf{H}(\mathbf{\Theta})$, $\mathbf{h}_i(\mathbf{\Theta}_i)$, contains the contributions of the ith superimposed signal to measurements at all detector pixels.)

Suitability of this local interaction model to our problem follows from the Gaussian shape of superimposed signals and a Cramer-Rao bound analysis. Note that in our observation model, superimposed signals are Gaussian line profiles, each centered around a different pixel on the detector (hence most of its energy is concentrated around that pixel). Therefore, for each Gaussian, the interaction (overlap) is limited to the closest neighboring Gaussians and is determined by the width of the Gaussian, which itself is determined by the amount of dispersion in the instrument (i.e. higher dispersion results in wider width, hence larger interaction). Therefore, the extent of interaction, r, depends on the amount of dispersion, which is a design choice for the instrument.

Moreover, a study of the Cramer-Rao bound, a lower bound on estimation uncertainties [42], [43], reveals that large dispersion (such that more than a few Gaussians overlap with each other) is not an optimal design choice because it results in significantly larger errors in the parameter estimates [28], [29]. That is, useful instrument models can be restricted without loss of generality to Gaussian line profiles which interact only with few of their closest neighbors (typically, $r \leq 4$).

A more general discussion of suitability of the local interaction model to a wide class of separable least squares problems has been given in [26]. It was shown, based on Cramer-Rao bound analysis, that in many instances superimposed signals interacting with more than a few neighbors cannot be separated to any meaningful accuracy; that is, useful models can be restricted to those with local interaction.

Other than the amount of dispersion in the instrument, there are other factors that affect the choice of r from the algorithmic point of view. As will be shown in the next two sections, the choice of r, which impacts the approximation in (19), provides a mechanism for making a tradeoff between the accuracy of DP estimates (in terms of closeness to the desired MAP estimates) and computational complexity.

B. Dynamic Programming Algorithm

The dynamic programming algorithm presented in this paper is an extension of a previously proposed method for

parameter estimation of superimposed signals [26], [27]. This algorithm was presented for the maximum likelihood problem and when each superimposed signal interacts with only one neighbor on both sides. Here we extend the algorithm to the MAP framework (that involves priors) and to superimposed signals interacting with arbitrary number of neighbors. A preliminary version of the extended MAP algorithm was presented in [44].

For simplicity and without loss of generality, we ignore the weights in the least squares term of the MAP functional in (18), and treat the problem with identical weights of unity. The more general case can be simply handled within this framework after scaling each measurement vector ($\tilde{\mathbf{y}}^{a_i}$) and measurement matrix ($\mathbf{H}^{a_i}(\mathbf{\Theta})$) by the standard deviation of the corresponding measurement noise (σ_{a_i}).

The main idea in the dynamic programming algorithm is to break the MAP optimization problem into smaller subproblems that are related to each other recursively. This recursive multistage optimization process is enabled by the local interaction model as follows: for any parameter set $(\mathbf{f}, \mathbf{\Theta})$, any extent of interaction r with $1 \leq r \leq M - 1$, and any pixel k in the range $1 \leq k < M - r$, the objective function in the MAP formulation can be decomposed as follows:

$$\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta})\mathbf{f}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m)$$

$$= \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2$$
$$+ \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]})\mathbf{f}_{[k+r+1:M]}\|^2$$
$$- \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m)$$
$$+ 2\text{Re}\{\mathbf{f}_{[1:k]}^* \mathbf{H}^*(\mathbf{\Theta}_{[1:k]}) \mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]}) \mathbf{f}_{[k+r+1:M]}\}$$
$$\approx \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2$$
$$+ \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\mathbf{\Theta}_{[k+r+1:M]})\mathbf{f}_{[k+r+1:M]}\|^2$$
$$- \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 + \sum_{m=1}^{M} \Gamma(\mathbf{\Theta}_m)$$
$$\triangleq \tilde{J}(\mathbf{\Psi}),$$

where $\mathbf{\Theta}_{[i:j]}$ denotes $[\mathbf{\Theta}_i \ \mathbf{\Theta}_{i+1} \ \ldots \ \mathbf{\Theta}_j]$, $\mathbf{f}_{[i:j]}$ denotes the corresponding similar representation, and $\mathbf{H}(\mathbf{\Theta}_{[i:j]})$ is a submatrix composed of ith to jth columns of $\mathbf{H}(\mathbf{\Theta})$. The approximate equality holds from the local interaction assumption in (19); hence, when the local interaction model holds, solving the MAP problem is equivalent to minimizing $\tilde{J}(\mathbf{\Psi})$.

This decomposed objective function has the generic functional form of

$$\tilde{J}(\mathbf{\Psi}) = \tilde{J}_1(\mathbf{\Psi}_{[1:k]}, \mathbf{\Psi}_{[k+1:k+r]}) + \tilde{J}_2(\mathbf{\Psi}_{[k+1:k+r]}, \mathbf{\Psi}_{[k+r+1:M]}) \tag{20}$$

with $\mathbf{\Psi}_i = (\mathbf{\Theta}_i, f_i)$, where the function $\tilde{J}_1(.)$ contains the first term of $\tilde{J}(\mathbf{\Psi})$, and $\tilde{J}_2(.)$ contains the next two terms, in addition to the prior terms. This form enables us to efficiently find the global optimum of $\tilde{J}(\mathbf{\Psi})$ via dynamic programming [45]–[47]. This is because given $\mathbf{\Psi}_{[k+1:k+r]}$ for any k, the variables $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_k$ and $\mathbf{\Psi}_{k+r+1}, \ldots, \mathbf{\Psi}_M$

are decoupled. As a result, if $\tilde{J}(\boldsymbol{\Psi})$ is optimized for a given $\boldsymbol{\Psi}_{[k+1:k+r]}$, then the optimal values of $\boldsymbol{\Psi}_1, \dots, \boldsymbol{\Psi}_k$ are a function of only $\boldsymbol{\Psi}_{[k+1:k+r]}$, and hence can be denoted as $\boldsymbol{\Psi}_{[1:k]}^*(\boldsymbol{\Psi}_{[k+1:k+r]})$. These optimal values can further be obtained by optimizing only $\tilde{J}_1(.)$. This property of the globally optimal solutions is similar to the Markov property of random processes (where in our case deterministic dependence replaces the role of statistical dependence).

This shows that our problem satisfies the *principle of optimality* of the theory of dynamic programming [45]: subsets of an optimal solution of the original problem are themselves optimal solutions to its subproblems. This allows us to efficiently solve the high-dimensional problem by solving smaller subproblems that are related to each other recursively. More specifically, if we define the kth subproblem as finding $\boldsymbol{\Psi}_{[1:k]}^*(\boldsymbol{\Psi}_{[k+1:k+r]})$ for any given $\boldsymbol{\Psi}_{[k+1:k+r]}$, then it can be solved by using the solution of the $(k-1)$th subproblem:

$$\boldsymbol{\Psi}_{[1:k]}^*(\boldsymbol{\Psi}_{[k+1:k+r]})$$
$$= \arg \min_{\substack{\boldsymbol{\Psi}_k \in \Omega \\ \boldsymbol{\Psi}_{[1:k-1]} \in \boldsymbol{\Psi}_{[1:k-1]}^*(\boldsymbol{\Psi}_{[k:k+r-1]})}} \tilde{J}_1(\boldsymbol{\Psi}_{[1:k]}, \boldsymbol{\Psi}_{[k+1:k+r]}). \quad (21)$$

This limits the search for $\boldsymbol{\Psi}_{[1:k-1]}$ to a reduced set given by the solution of the $(k-1)$th subproblem, and hence yields a significant computational gain over the exhaustive search of the original problem. (Indeed, if the $(k-1)$th subproblem has a unique solution, this reduced set contains only one solution.) The global minimum of $\tilde{J}(\boldsymbol{\Psi})$ can then be computed recursively through $M-r$ stages, where at the kth stage the kth subproblem is solved through recursion, while k increases from 1 to $M-r$.

As a final observation, we note that each subproblem can also be simplified. Explicitly, the kth subproblem is

$$\min_{\substack{\boldsymbol{\Theta}_{[1:k]} \in \Lambda^k \\ \mathbf{f}_{[1:k]} \in \Pi^k}} \left\{ \|\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 \right.$$
$$\left. + \sum_{m=1}^{k} \Gamma(\boldsymbol{\Theta}_m) \right\}, \quad (22)$$

Here, the minimization over $\boldsymbol{\Theta}_{[1:k]}$ can be solved separately by eliminating $\mathbf{f}_{[1:k]}$ from (22) based on the *variable projection* technique of separable nonlinear least squares problems [32]. This results in the following equivalent problem:

$$\min_{\boldsymbol{\Theta}_{[1:k]} \in \Lambda^k} \left\{ \|\mathbf{P}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:k]})}^{\perp}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}]\|^2 \right.$$
$$\left. + \sum_{m=1}^{k} \Gamma(\boldsymbol{\Theta}_m) \right\}, \quad (23)$$

where $P_{\mathbf{A}}^{\perp} = I - \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ is the projection matrix onto the orthogonal complement of the column space of $\mathbf{A}$.

With all these observations, the steps in the dynamic programming algorithm are summarized below.

1) *Initialization Stage ($k = 1$):*

a) For each $(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]}) \in \Omega^r$, solve the following problem

$$\hat{\boldsymbol{\Theta}}_{[1:1]}(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]}) = \arg \min_{\boldsymbol{\Theta}_1 \in \Lambda}$$
$$\{\|\mathbf{P}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:1]})}^{\perp}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[2:1+r]})\mathbf{f}_{[2:1+r]}]\|^2 + \Gamma(\boldsymbol{\Theta}_1)\},$$

through exhaustive search over $\boldsymbol{\Theta}_1 \in \Lambda$.
b) Record the optimal values as a function of $\boldsymbol{\Theta}_{[2:1+r]}$:

$$\boldsymbol{\Theta}_{[1:1]}^*(\boldsymbol{\Theta}_{[2:1+r]}) = \{\boldsymbol{\Theta}_{[1:1]} \in \Lambda : \boldsymbol{\Theta}_{[1:1]}$$
$$= \hat{\boldsymbol{\Theta}}_{[1:1]}(\boldsymbol{\Theta}_{[2:1+r]}, \mathbf{f}_{[2:1+r]})$$
$$\text{for some } \mathbf{f}_{[2:1+r]} \in \Pi^r\}.$$

2) *Update Stages ($k = 2, \dots, M - r$):*

a) For each $(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) \in \Omega^r$, solve the following problem

$$\hat{\boldsymbol{\Theta}}_{[1:k]}(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}) = \arg \min_{\substack{\boldsymbol{\Theta}_k \in \Lambda \\ \boldsymbol{\Theta}_{[1:k-1]} \in \boldsymbol{\Theta}_{[1:k-1]}^*(\boldsymbol{\Theta}_{[k:k+r-1]})}}$$
$$\left\{ \|\mathbf{P}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:k]})}^{\perp}[\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}\|^2 \right.$$
$$\left. + \sum_{m=1}^{k} \Gamma(\boldsymbol{\Theta}_m) \right\},$$

through exhaustive search over $\boldsymbol{\Theta}_k \in \Lambda$ and $\boldsymbol{\Theta}_{[1:k-1]} \in \boldsymbol{\Theta}_{[1:k-1]}^*(\boldsymbol{\Theta}_{[k:k+r-1]})$.
b) Record the optimal values as a function of $\boldsymbol{\Theta}_{[k+1:k+r]}$:

$$\boldsymbol{\Theta}_{[1:k]}^*(\boldsymbol{\Theta}_{[k+1:k+r]})$$
$$= \big\{ \boldsymbol{\Theta}_{[1:k]} \in \Lambda^k :$$
$$\boldsymbol{\Theta}_{[1:k]} = \hat{\boldsymbol{\Theta}}_{[1:k]}(\boldsymbol{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})$$
$$\text{for some } \mathbf{f}_{[k+1:k+r]} \in \Pi^r \big\}.$$

3) *Final Stage:*

a) To obtain the final estimate of $\boldsymbol{\Theta}$, solve the following problem

$$\hat{\boldsymbol{\Theta}} = \arg \min_{\substack{\boldsymbol{\Theta}_{[M-r+1:M]} \in \Lambda^r \\ \boldsymbol{\Theta}_{[1:M-r]} \in \boldsymbol{\Theta}_{[1:M-r]}^*(\boldsymbol{\Theta}_{[M-r+1:M]})}}$$
$$\left\{ \|\mathbf{P}_{\mathbf{H}(\boldsymbol{\Theta}_{[1:M]})}^{\perp}\tilde{\mathbf{y}}\|^2 + \sum_{m=1}^{M} \Gamma(\boldsymbol{\Theta}_m) \right\}, \quad (24)$$

through exhaustive search over $\boldsymbol{\Theta}_{[M-r+1:M]} \in \Lambda^r$ and $\boldsymbol{\Theta}_{[1:M-r]} \in \boldsymbol{\Theta}_{[1:M-r]}^*(\boldsymbol{\Theta}_{[M-r+1:M]})$.

b) Estimate of $\mathbf{f}$ is then given by

$$\hat{\mathbf{f}} = [\mathbf{H}^*(\hat{\boldsymbol{\Theta}})\mathbf{H}(\hat{\boldsymbol{\Theta}})]^{-1}\mathbf{H}^*(\hat{\boldsymbol{\Theta}})\tilde{\mathbf{y}}. \quad (25)$$

The relation of the dynamic programming algorithm to the MAP problem is stated in the following theorem. This theorem is a generalization of [26, Th. 5].

Theorem 1: If the local interaction model holds exactly for some $r \geq 1$, i.e.

$$\mathbf{h}_i(\boldsymbol{\Theta}_i)^*\mathbf{h}_j(\boldsymbol{\Theta}_j) = 0 \quad \text{for } |i - j| > r, \quad (26)$$

for all i, $j = 1, \ldots, M$, then the estimates obtained with the dynamic programming algorithm (with this value of r) are same as the MAP estimates obtained by solving (18).

Proof: See Appendix. □

This theorem shows that the exact MAP estimates can be obtained with the DP algorithm when the local interaction model holds exactly (i.e. with exact orthogonality). However, for our spectral imaging problem, exact orthogonality as in (26) is not possible because of the Gaussian nature of the overlapping signals, hence there is some unavoidable deviation from exact orthogonality. Fortunately, it has been shown, under some regularity conditions, that the dynamic programming algorithm is robust in the sense that deviations from exact orthogonality continuously perturb the DP estimates from the exact MAP estimate, and moreover, the resulting deviation from the exact MAP estimate is upper-bounded by a constant that is proportional to the deviation from exact orthogonality [27]. Therefore, for any well-conditioned problem, if the deviation from orthogonality is small enough, then the DP estimates are close to the desired MAP estimate. As a result, the accuracy of the DP estimates is controlled by the amount of deviation from exact orthogonality, which is indeed controlled by the choice of r. Therefore, as the value of r is increased, the accuracy of DP estimates will be improved. In practice, the DP estimates can be used as initialization for a local optimization method to obtain the exact global MAP estimate.

As a final remark, we note that this parameter estimation algorithm is quite general, and it can be applied to other problems involving different superimposed signals and priors. Superimposed signal models and the resulting separable nonlinear least squares problems are of wide interest in various applications such as sensor array processing, communications, imaging, robotics, and vision [32]. Two commonly encountered problems are estimation of frequency and amplitude of superimposed sinusoids, and estimation of position, width, and amplitude of overlapping pulses of given shape (as our problem) [26]. The dynamic programming algorithm is applicable to any such separable nonlinear least squares problem for which the local interaction signal model is suited.

C. Computational Aspects

We now consider the computational requirements of the generalized dynamic programming algorithm. Note that nonconvex minimization problem at each stage is solved through exhaustive search over the parameter space restricted by the constraint sets. This requires discretization of the search space to a finite number of parameter values. Let q be the number of quantization levels used in exhaustive search for each scalar parameter, and n and p be the number of scalar parameters in each Θ_m and f_m, respectively. (In our problem, we have $n = 2$ and $p = 1$ with the parameters in Θ_m being Δ_m and ϵ_m.)

For the minimization problem at each stage, we need to evaluate objective functions of the form

$$\|\mathbf{P}^{\perp}_{\mathbf{H}(\tilde{\Theta}_1)}(\tilde{\mathbf{y}} - \mathbf{H}(\tilde{\Theta}_2)\mathbf{f})\|^2 + \Gamma((\tilde{\Theta}_1)). \tag{27}$$

We will simply state the computational requirement in terms of the number of function evaluations of this form (although these function evaluations have different costs at different stages).

At the kth stage, the objective function needs to be evaluated for all possible values of Θ_k and $\Theta_{[1:k-1]}$. Assuming that each subproblem has a unique solution, the recorded set $\Theta^*_{[1:k-1]}(\Theta_{[k:k+r-1]})$ has at most the size of the vector $\mathbf{f}_{[k:k+r-1]}$; therefore, there are at most $(q^p)^r$ different values for $\Theta_{[1:k-1]}$. With q^n possible values for Θ_k, the objective function at the kth stage needs to be evaluated q^{rp+n} times. Moreover, this exhaustive search is repeated for every possible value of $(\Theta_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})$, hence $q^{(p+n)r}$ times. Because there are a total of $M - r - 1$ update stages, the total number of objective function evaluations at these stages is $(M - r - 1) \times q^{(p+n)r} \times q^{rp+n} = (M - r - 1)q^{r(2p+n)+n}$. Using a similar argument, the initialization and final stages require $q^{r(p+n)+n}$ and $q^{r(p+n)}$ function evaluations, respectively. Therefore, it follows that the total computational effort of the dynamic programming algorithm is of $\mathcal{O}(q^{r(2p+n)+n})$, while the exhaustive search of the original problem over the entire parameter space is of $\mathcal{O}(q^{M(p+n)})$. Hence the computational cost is exponential only in the number of interacting signals, r, while linear in the total number of superimposed signals, M, as opposed to the exponential cost in M in exhaustive search of the original problem. With typically $M \gg r$, this shows the computational efficiency of the dynamic programming algorithm compared to the exhaustive search of the original problem over the entire parameter space.

As mentioned before, there exists a bounded discrepancy between DP estimates and the global MAP estimates because of the approximation in the local interaction model. A second source of discrepancy will arise from performing the optimization (exhaustive search) at each stage over a discretized parameter space (rather than over continuous values of the parameters). Clearly, the discretization needs to be fine enough to remain close to the desired global MAP estimates. In practice, a local optimization method initialized with DP estimates will be used subsequently in order to refine these estimates and obtain the exact global MAP estimate.

As a final remark, we note that the computational complexity of the DP algorithm can be further reduced through parallel implementations of the dynamic programming algorithm [48], or through the approximate version of the algorithm [49] which has significantly lower computational cost.

V. SAMPLE APPLICATION

In this section, we illustrate the performance of the instantaneous spectral imaging technique and the MAP estimation framework for an application in solar spectral imaging [1]. For this, we consider a prominent solar emission line in the extreme ultraviolet (EUV) regime, with a central wavelength of $\lambda_0 = 19.512$ nm. Our goal is to estimate the parameters of this emission line (consisting of integrated intensity, line width and Doppler shift parameters) within a 2D FOV from the observations of the instantaneous spectral imager. These emission line parameters yield estimates of the key physical parameters of the Sun's extended atmosphere such as temperature, plasma density and velocity, and hence enable the investigation of the dynamic plasma behavior [50], [51].
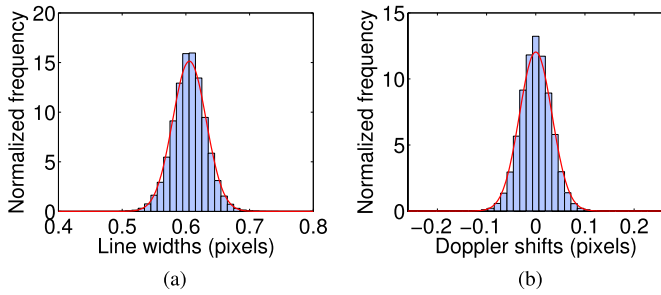
Fig. 4. Normalized histograms of line widths and Doppler shifts for the solar spectral imaging application. Fitted distributions are shown (in red) on top of the histograms.

A. Estimation of the Prior Distributions

To apply the MAP approach, we need to specify the prior distributions of line widths, Δ_m, and Doppler shifts, ϵ_m. The choice of these priors is application-dependent. Formerly, parameters at different pixels are assumed to be independently distributed. Here we further treat them as having the same distribution, hence as independent and identically distributed random variables from pixel-to-pixel (which is approximately true for quite-Sun regions). Therefore two density distributions, one for line widths Δ_m and one for Doppler shifts ϵ_m, need to be estimated.

For this, we use quite-Sun observations made with a conventional pushbroom (slit) spectrometer [51] during December of 2006. Each observation with a slit is associated with a 1D portion of the scene admitted through the slit, and arrives at a different time instance. Here we treat the line widths and Doppler shifts obtained from slit data, i.e. $\Delta_m(t)$ and $\epsilon_m(t)$, as independent and identically distributed (i.i.d.) over space m and stationary over time t. With this, each slit data obtained at a different time instance and slit position provides a different 1D realization of the line widths and Doppler shifts. Figure 4 shows the resulting histograms obtained from a large set of slit data. For density estimation, the histograms of line widths and Doppler shifts are normalized by the number of total observations (so that bin counts sum to one). Here we note that the parameters in the histograms are shown in pixel units, rather than in physical units, in order to match the units in the parametric model. The implicit step in this conversion is discussed in [28] and [29].

Because parameter values are clustered around one value in the histograms, Gaussian distribution is used to model their density distributions:

$$\Delta_m \sim N(\mu_\Delta, \sigma_\Delta^2), \tag{28}$$
$$\epsilon_m \sim N(\mu_\epsilon, \sigma_\epsilon^2), \tag{29}$$

with each parameter i.i.d. over pixels, $(\mu_\Delta, \sigma_\Delta^2)$ denoting the mean and variance of the distribution of line widths, and $(\mu_\epsilon, \sigma_\epsilon^2)$ denoting the mean and variance of the distribution of Doppler shifts. These mean and variance parameters are respectively estimated from the data using the maximum likelihood approach, which are then given by the sample mean and sample variance [52]. Gaussian distributions with these estimated parameters are shown (in red) on top of the histograms. The resulting prior terms to be used in the MAP

estimation are then given by

$$\Gamma(\boldsymbol{\Theta}_m) = \frac{(\Delta_m - \mu_\Delta)^2}{\sigma_\Delta^2} + \frac{(\epsilon_m - \mu_\epsilon)^2}{\sigma_\epsilon^2}, \tag{30}$$

for all $m = 1, \ldots, M$.

B. Numerical Results

Computer simulation results are presented to demonstrate the effectiveness of the parametric MAP approach for estimating the spectral line parameters from the measurements of instantaneous spectral imager. For this, we work with a portion of the slit data used in Section V-A. The spectral line parameters obtained from this solar data are used to simulate the measurements of the instantaneous spectral imager. More explicitly, for a detector of size 100×100 pixels (i.e. $M = 100$), the measurement $\tilde{\mathbf{y}}^a$ along each detector column is simulated based on the parametric model in equation (13) (i.e., as the superposition of the Gaussian line profiles characterized with these spectral line parameters). Such simulated measurements are obtained in three different orders $a \in \{0, +1, -1\}$. Also for each order, the additive noise term, $\mathbf{n}^a$, is randomly and independently generated according to Gaussian distribution, where each component has zero mean and variance of σ^2. The first row of Fig. 5 shows an example of the simulated noisy measurements with a noise standard deviation of $\sigma = 2$.

In order to estimate the spectral line parameters from these noisy measurements, the dynamic programming algorithm is used with the extent of interaction $r = 2$, hence with the model that each Gaussian profile, at most, interacts with its two closest neighbors on both sides. To define the constraint sets involved, line width and Doppler shift parameters are restricted to the ranges observed in the histograms (see Fig. 4). Constraining all integrated intensity parameters to a single range likewise would require a large amount of quantization levels in the algorithm and hence computational load, because the dynamic range of intensities over all pixels is large (see Fig. 6a). Instead, a different range is assigned to each integrated intensity parameter at a different pixel. In particular, each intensity parameter is constrained to lie around the value of the zeroth order measurement at that pixel, since the zeroth order measurements are Gaussian distributed around the true integrated intensities.

The parameter space restricted by these ranges must also be discretized (to a finite number of values) for the exhaustive search in the DP algorithm. A straightforward option is uniform discretization [26] of each range where the number of quantization levels is chosen based on the Cramer-Rao error bounds of the parameters [28], [29]. Instead, here we choose a nonuniform quantization grid to take into account the normal distribution of the parameters. More specifically, the grid is designed by dividing each parameter range to regions of equal probability, rather than of equal length. Hence the resulting grid is more dense around the mean of the parameter, where most of its realizations will lie.

The estimates obtained with the dynamic programming algorithm are refined by a gradient-based interior-point
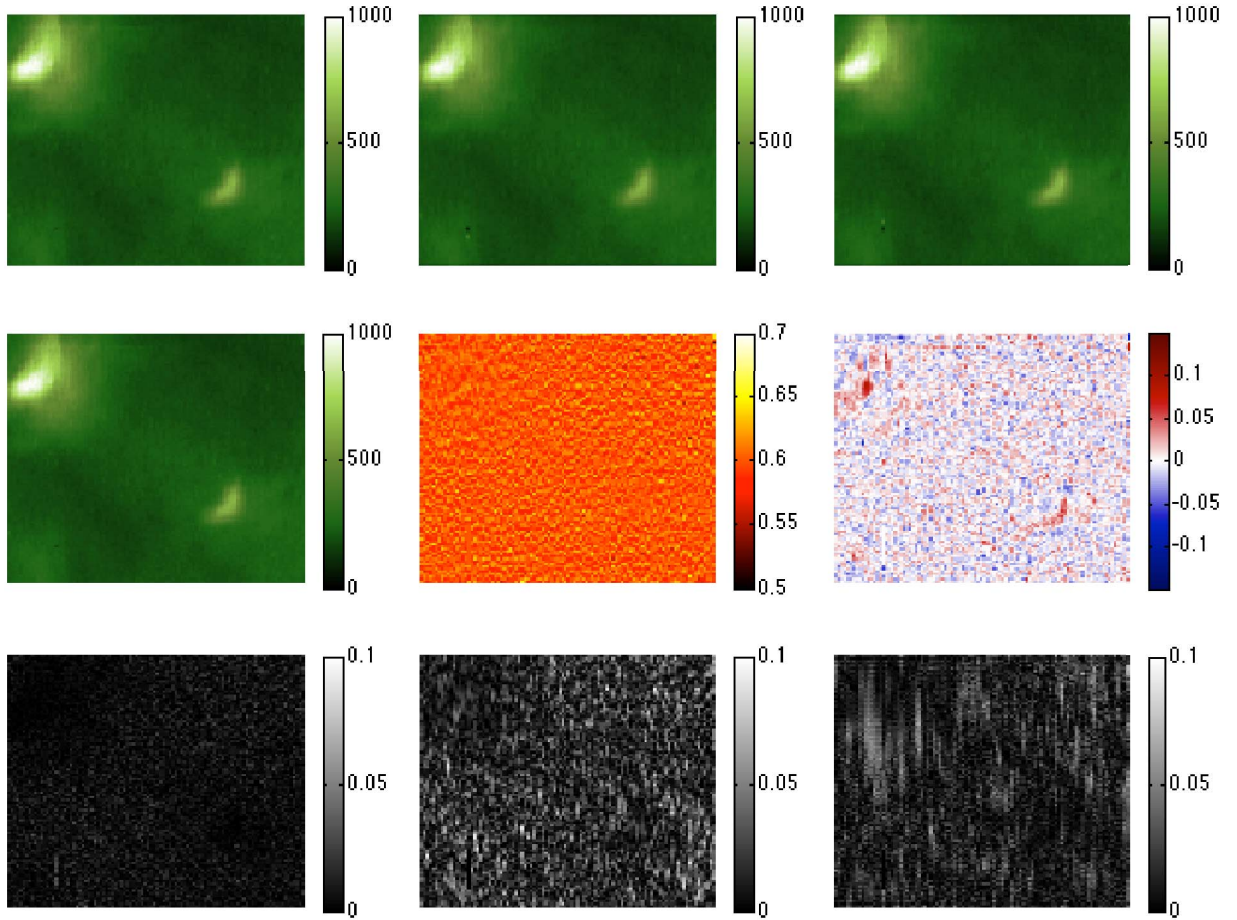
Fig. 5. Noisy observations in a solar imaging application and the estimated parameters of the Gaussian spectral lines from these observations. First row (from left to right): Simulated measurements of the instantaneous spectral imager at three diffraction orders 0, +1, and −1 on a detector of size 100×100 pixels. Second row (from left to right): Estimates of integrated line intensities, line widths, and Doppler shifts, which fully characterize the spectra at all spatial locations. (The true spectral line parameters are obtained from real solar data, and line width and Doppler shift parameters are shown in pixel units.) Third row (from left to right): The estimation errors for integrated intensities, line widths, and Doppler shifts as represented by the absolute difference between the estimated and true parameters. (The absolute differences for integrated intensities and line widths are normalized by their corresponding true values).

algorithm (a local optimization method). For the overall evaluation of the estimation, estimated parameters in each column are compared with their true values by using (i) the normalized absolute difference error $E_{diff}(m) = |f_m - \hat{f}_m|/f_m$, (ii) average absolute difference $E_{aad} = \sum_m |f_m - \hat{f}_m|/M$, and (iii) the root-mean-square (RMS) error $E_{rms} = (\sum_m (f_m - \hat{f}_m)^2/M)^{1/2}$ (similarly for Δ_m's and ϵ_m's.)

The middle row of Figure 5 shows the estimates of the integrated intensities, line widths, and Doppler shifts obtained from the given noisy observations. The last row of Fig. 5 also shows the absolute differences between the estimated and true parameters at each pixel. Here the absolute differences for integrated intensities and line widths are also normalized by their corresponding true values. The average absolute differences are typically less than 1% for intensities, 2.5% for line widths and 0.02 (pixels) for Doppler shifts; whereas the RMS errors are typically less than 2 for intensities, and 0.02 (pixels) for line widths and Doppler shifts. When converted to the physical units, this estimation with as low as three measured orders has the same order of accuracy as state-of-the-art slit spectroscopy used for this application [51], but with the added benefit of an instantaneous 2D FOV. Note that measuring more than three orders can help further to reduce the errors in the parameter

estimates. The quantification of the amount of improvement with additional orders is a topic of future study.

Figure 6 additionally illustrates the estimated spectra at two spatial locations and the estimates of the spectral line parameters over a column of the 2D scene. Figure 6a shows these two spatial locations and the column considered. The spectrum at each spatial location is a Gaussian function characterized by the spectral line parameters at that location and is shown in Fig. 6b both for the estimated and true parameters. Figure 6c also shows the estimates of integrated intensities, line widths, and Doppler shifts over all spatial locations along the marked column, and illustrates good agreement with the true values.

To evaluate the performance of the parametric MAP approach further, we investigate the effect of the noise standard deviation (hence SNR) on the estimation accuracy of the spectral line parameters. For this, Monte Carlo simulations are performed for a total of 40 random parameter sets (generated according to the prior distributions), and the numerical averages of RMS errors from these runs are computed for cases with varying noise standard deviation.

Fig. 7 shows the average RMS errors of the parameter estimates as a function of the noise standard deviation. To indicate
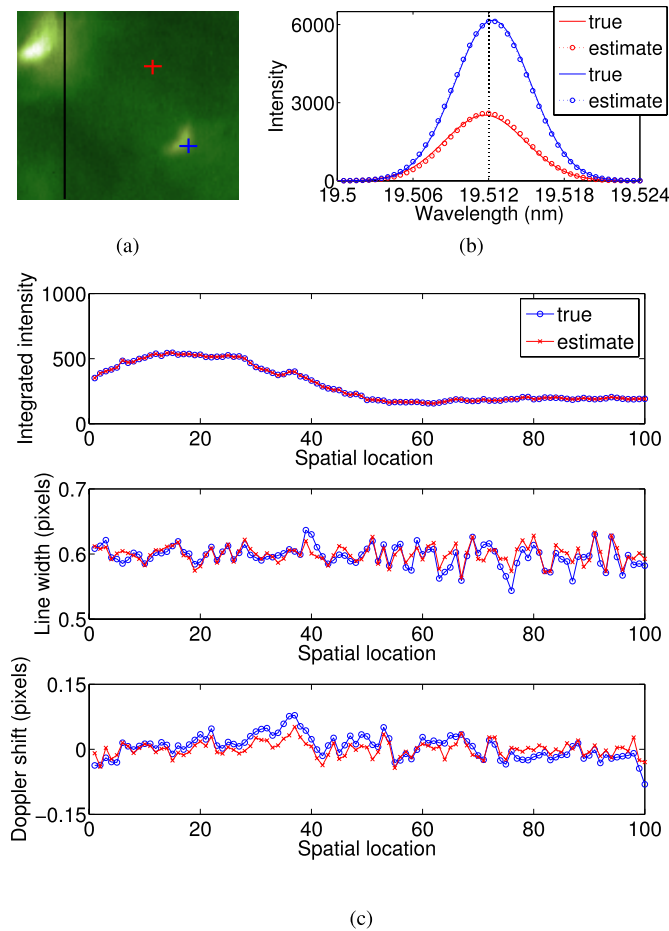
Fig. 6. Illustrations of the estimated spectra at two spatial locations and the estimates of the spectral line parameters over a column of the 2D scene. The two spatial locations and the column are marked in the spatial intensity image in (a). The estimated and true spectra associated with the marked spatial locations are shown in (b). (Here each spectrum is a Gaussian function characterized by the spectral line parameters at that spatial location.) Estimates of integrated intensities, line widths, and Doppler shifts associated with all spatial locations along the marked column are shown in (c).

the improvement in the accuracy of estimates as compared to a trivial estimate where all parameters are set to their known mean values without any estimation (more specifically, line width and Doppler shift estimates are set to mean values in their prior distributions and integrated intensity estimates are set to the zeroth order measurements), the RMS error of this trivial estimate is also shown in the figures with a dashed line. As seen, for the noise-free case with $\sigma = 0$, the true parameter values are always obtained. Moreover, the errors for integrated intensities strongly depend on the noise standard deviation since the zeroth order measurement directly provides a noisy observation of integrated intensities. In fact, for the low noise regime the intensity estimates obtained with the DP algorithm do not show significant improvement over the zeroth order measurement.

For the line widths and Doppler shifts, the dependence on the noise standard deviation σ is weaker at the high noise regime. This is because, in this regime, the estimation is highly dominated by the priors (rather than the measurements). Also we note that the estimation accuracy is comparable to the slit spectroscopy when σ is smaller than 4 (corresponding to an SNR of ~ 50 when SNR is defined as the ratio of the signal
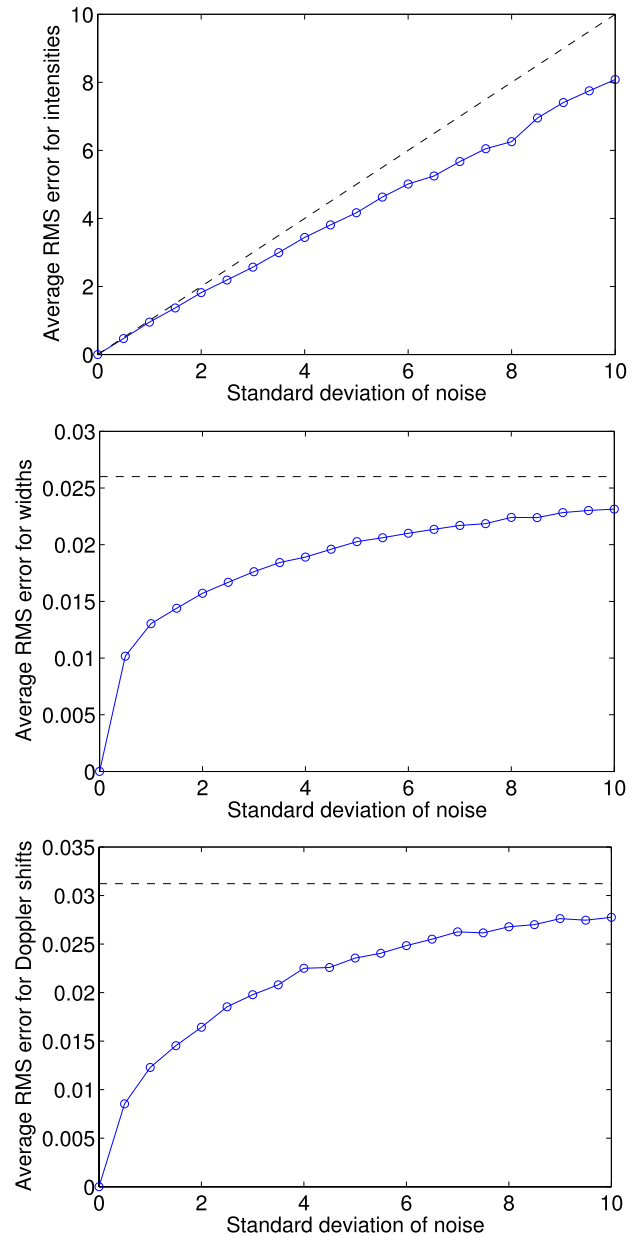

Fig. 7. RMS errors for the estimates of the intensity, width, and Doppler shift parameters as a function of the noise standard deviation when $\{0, +1, -1\}$ orders are measured along a 50-pixel detector column (i.e. $M = 50$). For comparison, dashed lines show the RMS errors of the trivial estimates where all parameters are set to their known mean values, without any estimation.

mean to the standard deviation of the noise). To achieve similar accuracy at higher noise levels, more spectral orders (than three) will be needed.

VI. CONCLUSION

We have presented a new spectral imaging modality with a slitless configuration that admits 2D instantaneous FOV. In this instantaneous spectral imaging technique, spectrally dispersed images of a 2D scene are simultaneously measured in several diffraction orders. The parameters of the spectral lines (within the scene) are then estimated by using these measurements with a parametric model and solving the resultant inverse problem computationally. The associated inverse problem can be viewed as a multiframe semi-blind

deblurring problem with shift-variant blur, and is tackled here by using a MAP estimation framework where the prior distributions of the spectral line parameters are estimated from the measurements of existing conventional slit spectrometers. An efficient dynamic programming algorithm is developed to find the global optimum of the resulting nonconvex MAP problem. This algorithm yields parameter estimates that are close to the global optimum of the MAP problem, which can then be refined by using a local optimization method.

We have investigated the application of the technique in solar spectral imaging. Computer simulation results suggest that spectral line parameters can be estimated with the same order of accuracy as the conventional slit spectroscopy, but with the added benefit of providing an instantaneous 2D field-of-view. Moreover, this estimation accuracy is achievable with as low as three dispersed images. This illustrates the advantage of the parametric approach over the tomographic approaches [4]–[8] which generally require significantly larger number of dispersed images.

It is important to note that this parametric approach is specifically designed for objects that have spectra consisting of Gaussian-shaped spectral lines. Here we provide a means of estimating the parameters of each spectral line over a 2D FOV, provided that dispersed images of the scene resulting from different spectral lines are non-overlapping at the detector. The extension of the approach to the cases with overlapping dispersed images will require extension of our formulation and treatment to the cases where the spectrum for any spatial location consists of multiple Gaussians, rather than a single Gaussian, and will be a topic for future study.

To conclude, this parametric approach to spectral imaging offers the means for effective estimation of spectral line parameters over an instantaneous 2D FOV. The estimated spectral line parameters can be used to infer the key physical parameters of a radiating medium (such as the temperature, density, and flow speed of the particles involved in the radiation). These inferred parameters enable the investigation of the dynamic behavior by revealing how currents and heat flow through the radiating medium. Such a capability resulting from the presented technique in this paper is particularly useful for studying the spectra of dynamic scenes in a wide variety of space remote sensing applications.

APPENDIX
PROOF OF THEOREM 1

This proof is a generalization of [31, Proof of Th. 5] to $r \geq 1$ case, and presented here for completeness.

Let $\hat{\mathbf{\Theta}}$ denote the estimate obtained with the DP algorithm, and $\{\mathbf{\Theta}^o, \mathbf{f}^o\}$ denote the exact MAP estimate given by solving (18), hence

$$\{\mathbf{\Theta}^o, \mathbf{f}^o\} = \arg \min_{\substack{\mathbf{f} \in \Pi^M \\ \mathbf{\Theta} \in \Lambda^M}} \left\{ ||\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta})\mathbf{f}||_W^2 + \sum_{m=1}^M \Gamma(\mathbf{\Theta}_m) \right\}. \quad (31)$$

Our goal is to show that $\hat{\mathbf{\Theta}} = \mathbf{\Theta}^o$. Note that this will imply that $\hat{\mathbf{f}} = \mathbf{f}^o$ since $\hat{\mathbf{f}}$ is obtained from $\hat{\mathbf{\Theta}}$ using (25).

First, we use induction to prove that

$$\hat{\mathbf{\Theta}}_{[1:k]}(\mathbf{\Theta}^o_{[k+1:k+r]}, \mathbf{f}^o_{[k+1:k+r]}) = \mathbf{\Theta}^o_{[1:k]} \quad \text{for } k=1, \ldots, M-r. \quad (32)$$

For this, similar to (25), we set

$$\hat{\mathbf{f}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})$$
$$= [\mathbf{H}^*(\hat{\mathbf{\Theta}}_{[1:k]})\mathbf{H}(\hat{\mathbf{\Theta}}_{[1:k]})]^{-1}\mathbf{H}^*(\hat{\mathbf{\Theta}}_{[1:k]})$$
$$\times (\tilde{\mathbf{y}} - \mathbf{\Theta}_{[k+1:k+r]}\mathbf{f}_{[k+1:k+r]}) \quad (33)$$

to the optimal values of the problem (22).

Then, from the update equation of $\mathbf{\Theta}$ in the algorithm, we have

$$\{\hat{\mathbf{\Theta}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]}), \hat{\mathbf{f}}_{[1:k]}(\mathbf{\Theta}_{[k+1:k+r]}, \mathbf{f}_{[k+1:k+r]})\}$$
$$= \arg \min_{\substack{\mathbf{\Theta}_k \in \Lambda \\ \mathbf{\Theta}_{[1:k-1]} \in \mathbf{\Theta}^*_{[1:k-1]}(\mathbf{\Theta}_{[k:k+r-1]}) \\ \mathbf{f}_{[1:k]} \in \Pi^k}} \Big\{ \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}_{[k+1:k+r]})\mathbf{f}_{[k+1:k+r]}$$
$$- \mathbf{H}(\mathbf{\Theta}_{[1:k]})\mathbf{f}_{[1:k]}\|^2 + \sum_{m=1}^k \Gamma(\mathbf{\Theta}_m) \Big\}. \quad (34)$$

We first prove the base case of induction for $k = 1$:

$$\{\hat{\mathbf{\Theta}}_{[1:1]}(\mathbf{\Theta}^o_{[2:1+r]}, \mathbf{f}^o_{[2:1+r]}), \hat{\mathbf{f}}_{[1:1]}(\mathbf{\Theta}^o_{[2:1+r]}, \mathbf{f}^o_{[2:1+r]})\}$$
$$= \arg \min_{\substack{\mathbf{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \{\|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:1]})f_1\|^2 + \Gamma(\mathbf{\Theta}_1)\}$$
$$= \arg \min_{\substack{\mathbf{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \Big\{ \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:1]})f_1\|^2$$
$$+ \Gamma(\mathbf{\Theta}_1) + \|\mathbf{H}(\mathbf{\Theta}^o_{[2+r:M]})\mathbf{f}^o_{[2+r:M]}\|^2$$
$$- 2\text{Re}\{(\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}^o_{[2:1+r]})\mathbf{f}^o_{[2:1+r]} - \mathbf{H}(\mathbf{\Theta}_{[1:1]})f_1)^*$$
$$\times \mathbf{H}(\mathbf{\Theta}^o_{[2+r:M]})\mathbf{f}^o_{[2+r:M]}\} + \sum_{m=2}^M \Gamma(\mathbf{\Theta}^o_m)\Big\}$$
$$= \arg \min_{\substack{\mathbf{\Theta}_1 \in \Lambda \\ f_1 \in \Pi}} \Big\{ \|\tilde{\mathbf{y}} - \mathbf{H}(\mathbf{\Theta}^o_{[2:M]})\mathbf{f}^o_{[2:M]} - \mathbf{H}(\mathbf{\Theta}_{[1:1]})f_1\|^2$$
$$+ \Gamma(\mathbf{\Theta}_1) + \sum_{m=2}^M \Gamma(\mathbf{\Theta}^o_m)\Big\}$$
$$= \{\mathbf{\Theta}^o_1, f^o_1\}. \quad (35)$$

The second equality follows from (26) which implies that $\mathbf{H}(\mathbf{\Theta}_{[1:1]})^*\mathbf{H}(\mathbf{\Theta}^o_{[2+r:M]}) = 0$, and from the fact that adding terms that are independent of $\mathbf{\Theta}_1$ and f_1 do not affect the minimization. The fourth equality follows from the optimality of $\{\mathbf{\Theta}^o, \mathbf{f}^o\}$.

Now, for the inductive step, suppose that

$$\hat{\mathbf{\Theta}}_{[1:i]}(\mathbf{\Theta}^o_{[i+1:i+r]}, \mathbf{f}^o_{[i+1:i+r]}) = \mathbf{\Theta}^o_{[1:i]} \quad (36)$$

for some $i \in \{1, 2, \ldots, M - r\}$. We want to prove that the statement holds for $i + 1$. Similar to the base step, we have

$$\{\hat{\boldsymbol{\Theta}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]}),$$
$$\hat{\mathbf{f}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]})\}$$

$$= \arg \min_{\substack{\boldsymbol{\Theta}_{i+1} \in \Lambda \\ \boldsymbol{\Theta}_{[1:i]} \in \boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]}) \\ \mathbf{f}_{[1:i+1]} \in \Pi^{i+1}}} \left\{ \| \tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}) \mathbf{f}^o_{[i+2:i+r+1]} \right.$$
$$- \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]}) \mathbf{f}_{[1:i+1]} \|^2 + \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m)$$
$$+ \| \mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]}) \mathbf{f}^o_{[i+r+2:M]} \|^2$$
$$- 2\mathrm{Re}\{ (\tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}) \mathbf{f}^o_{[i+2:i+r+1]}$$
$$- \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]}) \mathbf{f}_{[1:i+1]})^* \mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]})$$
$$\left. \times \mathbf{f}^o_{[i+r+2:M]} \} + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \right\}$$

$$= \arg \min_{\substack{\boldsymbol{\Theta}_{i+1} \in \Lambda \\ \boldsymbol{\Theta}_{[1:i]} \in \boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]}) \\ \mathbf{f}_{[1:i+1]} \in \Pi^{i+1}}} \left\{ \| \tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:M]}) \mathbf{f}^o_{[i+2:M]} \right.$$
$$- \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]}) \mathbf{f}_{[1:i+1]} \|^2 + \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m)$$
$$\left. + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \right\} \tag{37}$$

where the first equality follows from (34), and from (26) which implies that $\mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]})^* \mathbf{H}(\boldsymbol{\Theta}^o_{[i+r+2:M]}) = 0$. Also, by the optimality of $\{\boldsymbol{\Theta}^o, \mathbf{f}^o\}$,

$$\{\boldsymbol{\Theta}^o_{[1:i+1]}, \mathbf{f}^o_{[1:i+1]}\}$$
$$= \arg \min_{\substack{\boldsymbol{\Theta}_{[1:i+1]} \in \Lambda^{i+1} \\ \mathbf{f}_{[1:i+1]} \in \Pi^{i+1}}} \left\{ \| \tilde{\mathbf{y}} - \mathbf{H}(\boldsymbol{\Theta}^o_{[i+2:M]}) \mathbf{f}^o_{[i+2:M]} - \mathbf{H}(\boldsymbol{\Theta}_{[1:i+1]}) \right.$$
$$\left. \times \mathbf{f}_{[1:i+1]} \|^2 + \sum_{m=1}^{i+1} \Gamma(\boldsymbol{\Theta}_m) + \sum_{m=i+2}^{M} \Gamma(\boldsymbol{\Theta}^o_m) \right\} \tag{38}$$

If we now compare (37) and (38), the objective functions are the same, while the constraint sets are different with the latter one containing the former one. But because the minimizer $\boldsymbol{\Theta}^o_{[1:i+1]}$ of (38) is yet an element of the smaller constraint set in (37), it must also be the minimizer of (37). (This follows because $\boldsymbol{\Theta}^o_{[1:i]} = \hat{\boldsymbol{\Theta}}_{[1:i]}(\boldsymbol{\Theta}^o_{[i+1:i+r]}, \mathbf{f}^o_{[i+1:i+r]})$ by the assumption (36), and hence $\boldsymbol{\Theta}^o_{[1:i]} \in \boldsymbol{\Theta}^*_{[1:i]}(\boldsymbol{\Theta}_{[i+1:i+r]})$). Therefore,

$$\{\hat{\boldsymbol{\Theta}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]}),$$
$$\hat{\mathbf{f}}_{[1:i+1]}(\boldsymbol{\Theta}^o_{[i+2:i+r+1]}, \mathbf{f}^o_{[i+2:i+r+1]})\} = \{\boldsymbol{\Theta}^o_{[1:i+1]}, \mathbf{f}^o_{[1:i+1]}\}. \tag{39}$$

This completes the induction, hence the statement in (32) is proved. With $k = M - r$, this gives that

$$\hat{\boldsymbol{\Theta}}_{[1:M-r]}(\boldsymbol{\Theta}^o_{[M-r+1:M]}, \mathbf{f}^o_{[M-r+1:M]}) = \boldsymbol{\Theta}^o_{[1:M-r]}, \tag{40}$$

as a result, $\boldsymbol{\Theta}^o_{[1:M-r]} \in \boldsymbol{\Theta}^*_{[1:M-r]}(\boldsymbol{\Theta}_{[M-r+1:M]})$. Using this in (24) together with the similar arguments used above, it is easily seen that $\hat{\boldsymbol{\Theta}} = \boldsymbol{\Theta}^o$.

REFERENCES

[1] K. J. H. Phillips, U. Feldman, and E. Landi, *Ultraviolet and X-Ray Spectroscopy of the Solar Atmosphere*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[2] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*. New York, NY, USA: Wiley, 2008.

[3] G. G. Shepherd, *Spectral Imaging of the Atmosphere*, vol. 82. San Diego, CA, USA: Academic, 2002.

[4] T. Okamoto and I. Yamaguchi, "Simultaneous acquisition of spectral image information," *Opt. Lett.*, vol. 16, no. 16, pp. 1277–1279, 1991.

[5] M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: Experimental calibration and reconstruction results," *Appl. Opt.*, vol. 34, no. 22, pp. 4817–4826, 1995.

[6] C. C. Kankelborg and R. J. Thomas, "Simultaneous imaging and spectroscopy of the solar atmosphere: Advantages and challenges of a 3-order slitless spectrograph," *Proc. SPIE*, vol. 4498, pp. 16–26, Dec. 2001.

[7] W. R. Johnson, D. W. Wilson, W. Fink, M. Humayun, and G. Bearman, "Snapshot hyperspectral imaging in ophthalmology," *J. Biomed. Opt.*, vol. 12, no. 1, p. 014036, 2007.

[8] N. Hagen and E. L. Dereniak, "Analysis of computed tomographic imaging spectrometers. I. Spatial and spectral resolution," *Appl. Opt.*, vol. 47, no. 28, pp. F85–F95, 2008.

[9] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14013–14027, 2007.

[10] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[11] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[12] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," *Opt. Lett.*, vol. 36, no. 14, pp. 2692–2694, 2011.

[13] T. Sun and K. Kelly, "Compressive sensing hyperspectral imager," in *Proc. Comput. Opt. Sensing Imag. (COSI)*, San Jose, CA, USA, Oct. 2009, paper CTuA5.

[14] Y. August, C. Vachman, Y. Rivenson, and A. Stern, "Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains," *Appl. Opt.*, vol. 52, no. 10, pp. D46–D54, 2013.

[15] J. E. Fowler, "Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 2014, pp. 684–688.

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

[17] F. Soldevila *et al.*, "Single-pixel polarimetric imaging spectrometer by compressive sensing," *Appl. Phys. B*, vol. 113, no. 4, pp. 551–558, 2013.

[18] S. Minin and F. Kamalabadi, "Uncertainties in extracted parameters of a Gaussian emission line profile with continuum background," *Appl. Opt.*, vol. 48, no. 36, pp. 6913–6922, 2009.

[19] D. D. Lenz and T. R. Ayres, "Errors associated with fitting Gaussian profiles to noisy emission-line spectra," *Publ. Astronomical Soc. Pacific*, vol. 140, pp. 1104–1106, Nov. 1992.

[20] D. A. Landman, R. Roussel-Dupre, and G. Tanigawa, "On the statistical uncertainties associated with line profile fitting," *Astrophys. J.*, vol. 261, pp. 732–735, Oct. 1982.

[21] J. Ireland, "Precision limits to emission-line profile measuring experiments," *Astrophys. J.*, vol. 620, no. 2 p. 1132, 2005.

[22] L. Lading and A. Jensen, "Estimating the spectral width of a narrowband optical signal," *Appl. Opt.*, vol. 19, no. 16, pp. 2750–2756, 1980.

[23] P. B. Hays and R. G. Roble, "A technique for recovering Doppler line profiles from Fabry–Perot interferometer fringes of very low intensity," *Appl. Opt.*, vol. 10, no. 1, pp. 193–200, 1971.

[24] T. L. Killeen and P. B. Hays, "Doppler line profile analysis for a multichannel Fabry–Perot interferometer," *Appl. Opt.*, vol. 23, no. 4, pp. 612–620, 1984.

[25] E. J. Mierkiewicz, F. L. Roesler, S. M. Nossal, and R. J. Reynolds, "Geocoronal hydrogen studies using Fabry–Perot interferometers, part 1: Instrumentation, observations, and analysis," *J. Atmos. Solar-Terrestrial Phys.*, vol. 68, no. 13, pp. 1520–1552, 2006.

[26] S. F. Yau and Y. Bresler, "Maximum likelihood parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 804–820, Feb. 1993.

[27] S. F. Yau and Y. Bresler, "On the robustness of parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2825–2836, Nov. 1996.

[28] F. S. Oktem, F. Kamalabadi, and J. M. Davila, "Cramer–Rao bounds and instrument optimization for slitless spectroscopy," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 2169–2173.

[29] F. S. Oktem, "Computational imaging and inverse techniques for high-resolution and instantaneous spectral imaging," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Illinois at Urbana-Champaign, Champaign, IL, USA, 2014.

[30] J. M. Hollas, *Modern Spectroscopy*. New York, NY, USA: Wiley, 2004.

[31] J. M. Lerner, "Imaging spectrometer fundamentals for researchers in the biosciences—A tutorial," *Cytometry A*, vol. 69A, no. 8, pp. 712–734, 2006.

[32] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[34] I. Griva, S. G. Nash, and A. Sofer, *Linear and Nonlinear Optimization*. Philadelphia, PA, USA: SIAM, 2009.

[35] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[36] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[37] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. R1–R26, Mar. 1989.

[38] H. Kwakernaak, "Estimation of pulse heights and arrival times," *Automatica*, vol. 16, no. 4, pp. 367–377, 1980.

[39] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Oper. Res.*, vol. 14, no. 4, pp. 699–719, 1966.

[40] L. Liberti and N. Maculan, *Global Optimization: From Theory to Implementation*, vol. 84. New York, NY, USA: Springer-Verlag, 2006.

[41] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, vol. 65. New York, NY, USA: Wiley, 2005.

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[43] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*. New York, NY, USA: Wiley, 1971.

[44] F. S. Oktem, F. Kamalabadi, and J. M. Davila, "Parameter estimation for instantaneous spectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 8148–8152.

[45] R. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[46] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1. Belmont, MA, USA: Athena Scientific, 1995.

[47] M. Sniedovich, *Dynamic Programming: Foundations and Principles*. Boca Raton, FL, USA: CRC Press, 2010.

[48] W. Rytter, "On efficient parallel computations for some dynamic programming problems," *Theoretical Comput. Sci.*, vol. 59, no. 3, pp. 297–307, 1988.

[49] S. Yau and Y. Bresler, "A compact Cramer–Rao bound expression for parametric estimation of superimposed signals," *IEEE Trans. Signal Process.*, vol. 40, no. 5, pp. 1226–1230, May 1992.

[50] F. Kamalabadi, "Multidimensional image reconstruction in astronomy," *IEEE Signal Process. Mag.*, vol. 27, no. 1, pp. 86–96, Jan. 2010.

[51] P. R. Young *et al.*, "EUV emission lines and diagnostics observed with Hinode/EIS," *Publ. Astron. Soc. Jpn.*, vol. 59, pp. S857–S864, Nov. 2007.

[52] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition*. New York, NY, USA: Wiley, 2011.

Figen S. Oktem is currently a Post-Doctoral Research Associate with the NASA Goddard Space Flight Center, Greenbelt, MD, USA, where she is conducting research in the Heliophysics Science Division. She received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2007 and 2009, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA, in 2014.

Her research spans the areas of statistical signal processing, optical information processing, computational imaging, and inverse problems with applications to biomedical imaging and remote sensing. Her main research interest is the development of novel computational imaging and sensing technologies, and their applications in physical and life sciences.

Dr. Oktem was a recipient of the Professor Kung Chie Yeh Endowed Fellowship from the Department of Electrical and Computer Engineering, UIUC, in 2012, and a NASA Graduate Research Fellowship from 2012 to 2014.

Farzad Kamalabadi (M'02) received the B.S. degree in computer systems engineering from the University of Massachusetts, Amherst, MA, USA, in 1991, and the M.S. and Ph.D. degrees in electrical engineering from Boston University, Boston, MA, USA, in 1994 and 2000, respectively.

Since 2000, he has been on the faculty of the University of Illinois at Urbana-Champaign, Champaign, IL, USA, where he is currently a Professor in the Department of Electrical and Computer Engineering, an Affiliate Professor in the Department of Statistics, and a Research Professor in the Coordinated Science Laboratory. From 2010 to 2012, he was also a Program Director with the U.S. National Science Foundation, Arlington, VA, USA. He was a Visiting Fellow with SRI International, Menlo Park, CA, USA, in Fall 2002, a National Aeronautics and Space Administration (NASA) Faculty Fellow with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, in Summer 2003, and a Visiting Professor with the French National Research Institute for Computer Science and Control, Sophia Antipolis, France, in 2008 and 2009. His research interests are in the areas of multidimensional and statistical signal processing, inverse problems in image formation, sensor array processing, tomography, and the development of optical and radio sensing and imaging techniques and their applications to space physics and astronomy.

Dr. Kamalabadi has served as a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE, and currently serves as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Joseph M. Davila is currently an Astrophysicist with the Solar Physics Branch, NASA Goddard Space Flight Center, Greenbelt, MD, USA. He was born in 1948. He received the B.S. degree in mechanical engineering from Lamar University, Beaumont, TX, USA, in 1972, the B.S. degree in physics from the University of California at Irvine, Irvine, CA, USA, in 1978, and the Ph.D. degree in astronomy from the University of Arizona, Tucson, AZ, USA, in 1982.

He is a member of the American Astronomical Society, the American Geophysical Union, and the International Astronomical Union. His research interests have included the linear and nonlinear theory of hydromagnetic waves, hydromagnetic instabilities due to energetic particle beams, resonance absorption in inhomogeneous plasmas, the acceleration of high-speed wind streams in solar and stellar coronal holes, and plasma heating in closed magnetic structures. He has also published research on the acceleration of cosmic rays, the transport of energetic particles within the Galaxy, the modulation of Galactic cosmic rays by the solar wind, and the propagation of solar cosmic rays in the interplanetary medium.

Dr. Davila was the Principal Investigator for the Solar Extreme-Ultraviolet Research Telescope and Spectrograph, and the Project Scientist for Solar Terrestrial Relations Observatory (STEREO). He is the lead Scientist for COR1 on the STEREO mission, and Co-I on the Hinode and Solar Orbiter missions.