Quantitative imaging through a spectrograph – 1: principles & theory

R.A.L. (René) Tolboom

N.J. (Nico) Dam

J.J. (Hans) ter Meulen

University of Nijmegen, Applied Physics,

Toernooiveld 1, NL-6525 ED Nijmegen, the Netherlands

J.M. (Joris) Mooij

J.D.M. (Hans) Maassen

University of Nijmegen, Mathematical Institute,

Toernooiveld 1, NL-6525 ED Nijmegen, the Netherlands

Laser-based optical diagnostics, like planar Laser-Induced Fluorescence and, especially, Raman imaging, often require very selective spectral filtering. We advocate the use of an imaging spectrograph with a broad entrance slit as spectral filter for two-dimensional imaging. A spectrograph in this mode of operation produces output that is a convolution of the spatial and spectral information present in the incident light. We describe an analytical deconvolution procedure, based on Bayesian statistics, that retrieves the spatial information while avoiding excessive noise blow-up. The method allows direct imaging through a spectrograph even under broad-band illumination. In this paper we introduce the formalism and discuss the underlying assumptions. The performance of the procedure is demonstrated on an artificial, but pathological example. In a companion paper (next paper in this issue) the method is applied to the practical case of fuel equivalence ratio Raman imaging in a combustible methane/air mixture.

OCIS codes: 000.2170, 100.1830, 100.3020

Final version, 5th July 2004

1. Introduction

Optical techniques find ever more applications in (reactive or non-reactive) gas flow diagnostics.[1–3] There is good reason for that. Light scattering techniques are as near as one can get to non-intrusive diagnostics, there exist widely available (laser) light sources of unsurpassed spectral brightness that allow specific detection of selected chemical species (of great importance especially in combustion), and a large variety of experimental techniques has been developed for many specific purposes. This paper focusses on two-dimensional (2-D) optical imaging of density distributions of specific chemical species. There are two methods of choice for this purpose,[1,2] Planar Laser-Induced Fluorescence (PLIF) and Planar Raman scattering. Both combine a thin (quasi-monochromatic) light sheet derived from a powerful laser system with (intensified) CCD camera's for detection of the scattered light. Both, also, usually require good spectral filtering of the scattered light, in order to suppress undesired contributions to the measured light intensity. In this paper we discuss the use of an imaging grating spectrograph with a CCD camera on the exit port (an Optical Multichannel Analyser (OMA)) for spectrally selective 2-D imaging. Although we have the specific purpose of Raman imaging in mind, the technique is not restricted to that, and can be used for spectrally selective imaging in general. The use of a spectrograph has two main advantages over the use of spectral band-pass filters: its spectral selectivity is greater, and it provides a spectrum. Thus, undesired spectral contributions will be suppressed more effectively, and even if unexpected spectral interferences occur, a spectrograph will at least show them, whereas they are likely to pass unnoticed when

using band-pass filters.

An ideal imaging grating spectrograph projects a faithful image of its entrance slit onto its exit plane. Where exactly on the exit plane this image ends up also depends on the wavelength of the incident light, as follows directly from the familiar grating equation[4,5] (see Fig. 1),

$$d\,(\sin\alpha \pm \sin\beta) = m\lambda\ , \tag{1}$$

where m is the order of diffraction and d is the groove distance. Both the angle of incidence (α) and of reflection (β) are defined with respect to the grating normal (dash-dotted lines). The '+'-sign applies when α and β are on the same side of the grating normal, and the '$-$'-sign when they are on opposite sides (as in the depicted example). In most cases, a spectrograph is used with a narrow entrance slit to record line (1-D) spectra, *e.g.* Raman spectra in combustion science,[1,6–9] or point (0-D) spectra for increased signal-to-noise ratios.[10,11] In the former case (1-D), one of the axes that span the exit plane carries a spectral scale and the other one still carries purely spatial information (along the height of the entrance slit). One way to visualise planar (2-D) patterns with 1-D imaging is to perform consecutive line measurements for many longitudinal positions. Recently, Sijtsema *et al.*[12] introduced direct 2-D imaging through a spectrograph as a non-intrusive tool for quantitative, planar gas flow visualisation. This technique, further referred to as OMA imaging, is the focus of the present paper. It will be shown that OMA imaging, originally introduced for use with quasi-monochromatic light,[12] can be used in case of extended spectral structure as well.[13]

Below, we concentrate on the theoretical issues related to spectrally selective 2-D imaging using a set-up that incorporates an imaging spectrograph with a diffraction grating operated in first order. The grating effectively produces an output image on the spectrograph exit plane in which the spectral and spatial information present in the incident light are scrambled. It is shown that, in spite of the diffraction, the original (spatial) image can nevertheless be reconstructed for many practical situations. The first section deals with the formalism that describes the scrambling of the spatial and spectral information by the imaging spectrograph. Under not too stringent prerequisites, this scrambling is shown to take the form of a convolution. The second section describes the straightforward analytical deconvolution that, given the spectral composition of the input, should in principle return the spatial distribution. This scheme, however, is shown to suffer in practice from excessive noise blow-up. Therefore, a linear Bayesian deconvolution filter was developed. This filter, that can still be expressed analytically, is described in section 3B, together with a discussion of its performance and the way it is tuned to a particular application. The procedures are illustrated in this paper by means of actual recordings of a square grid that is illuminated by bichromatic light. This is a fairly pathological example, because the original object contains sharp contrast. Any algorithm that provides satisfactory results on this case may therefore be expected to perform well on less demanding cases also.

The mathematics in this paper is not exhaustive; more mathematical detail and properties of the Bayesian filter can be found in the thesis of Tolboom.[14] A companion paper[15] (further referred to as T2) is devoted to the practical application of

the general OMA imaging results derived here to 2-D Raman imaging for fuel equivalence ratio mapping. It turns out that 2-D OMA imaging takes full benefit of an imaging spectrograph and provides a technique that is very time and cost efficient for simultaneous mapping of molecule-specific 2-D density distributions.

2. Convolution by a spectrograph

The images that are the subject of this paper are formed through a spectrograph via a reflection grating operated in first order. The latter diffracts the incident light into its constituent wavelength components in a direction perpendicular to the grooves, as described by Eq. (1). If the spectrograph is operated with a narrow entrance slit (as would normally be the case for spectroscopic measurements), the angle α is well-defined and the light intensity on any position of the spectrograph exit plane is, via the angle β and Eq. (1), directly related to a specific wavelength λ. However, if the entrance slit is *not* narrow, the angle α is not well-defined, and this may give rise to ambiguity in the light intensity distribution in the exit plane. In this broad-slit case, one dimension of the image (say $\hat{x}$) in the exit plane contains both spatial and spectral information. The other dimension (parallel to the grooves of the grating and to the height of the entrance slit; $\hat{y}$) contains spatial information only, and will be omitted in the following discussion. Fig. 2 depicts a cross section of the spectrograph, showing the entrance slit in the focal plane of a collimating lens, a reflective grating and the exit port in the focal plane of a second lens. (In commercial spectrographs[16] mirrors are used instead of lenses, but this does not make a difference to the formalism presented.) The input image on the entrance slit is a real image, formed by an external lens (not

shown in Fig. 2), of some (planar) light source. For clarity, Fig. 2 illustrates the case of bichromatic light, in which the two wavelengths are chosen to be sufficiently distinct to prevent the two resulting images from overlapping. In practice, however, this need not be the case: the incident light may be polychromatic, resulting in overlapping images.

In principle, all rays with a given wavelength λ originating from a specific point x_{in} in the entrance slit plane will be imaged onto exactly one point x_{out} in the exit plane, thus transforming spectral information into spatial information. Conversely, the wavelength λ that can be associated with any position x_{out} in the exit plane also depends on the source position x_{in}. For an extended, polychromatic source, therefore, the relation between λ and x_{out} is ambiguous. (This, in fact, is what usually is loosely referred to as 'loss of spectral resolution' if the entrance slit of a spectrograph is broadened.) In general terms, the relation between the measured light intensity distribution T and the incident intensity distribution S_{in} can be written as[17]

$$T(x_{\mathrm{out}}) = \int_{\lambda} \int_{x_{\mathrm{in}} \in \mathrm{slit}} \mathcal{T}(\lambda, x_{\mathrm{in}};\, x_{\mathrm{out}})\, S_{\mathrm{in}}(\lambda, x_{\mathrm{in}})\, \mathrm{d}\lambda\, \mathrm{d}x_{\mathrm{in}} \quad , \tag{2}$$

with $x_{\mathrm{in,out}}$ coordinates in the spectrograph entrance and exit planes, respectively, T the (spatial) intensity distribution in the exit plane, S_{in} the spectral and spatial intensity distribution in the entrance plane, and $\mathcal{T}$ a transfer function. The integrations run over the whole spectrum and, according to the reasoning above, over the whole range of x_{in} along the width of the entrance slit. As there is input only at the entrance slit, the integration over space can be extended from $-\infty$ to $+\infty$ without affecting the result.[18]

For the present purpose, perfect imaging will be assumed, and only a grating efficiency $\eta(\lambda)$ is accounted for. Under this assumption the transfer function links λ and $x_{\rm in}$ to $x_{\rm out}$ via

$$\mathcal{T}(\lambda, x_{\rm in};\, x_{\rm out}) = \eta(\lambda)\, \delta(x_{\rm out} - f(\lambda, x_{\rm in})) \quad , \tag{3}$$

in which the function $f(\lambda, x_{\rm in})$ depends on grating diffraction and imaging optics. It is interpreted as the function that describes where on the exit plane a monochromatic point source, having wavelength λ and being located at $x_{\rm in}$, would be imaged. The Dirac delta function allows only the signals of those combinations of λ and $x_{\rm in}$ for which a point $x_{\rm out}$ is illuminated to contribute to the signal at that point $x_{\rm out}$.

Unravelling of the spectral and spatial information contained in the input pattern $S_{\rm in}(\lambda, x_{\rm in})$ requires that $S_{\rm in}$ can be factorised, that is,

$$S_{\rm in}(\lambda, x_{\rm in}) = S_\lambda(\lambda) \times S(x_{\rm in}) \quad . \tag{4}$$

This restriction, by the way, is not particular to OMA imaging, but applies to spectrally selective imaging in general. It limits the interpretation of OMAgraphs (that is, a photograph taken through an Optical Multichannel Analyser) to applications in which *either* the light source has a uniform spectral composition over the width of the entrance slit, *or* the contributions from sources with different spatial and spectral distributions do not overlap on the OMAgraph (see below, section 3A). In T2 it will further be discussed when and whether this factorisation is justified for multi-species Raman scattered light. For the moment it suffices to assume that it holds for a single light source that has a spectral profile $S_\lambda(\lambda)$ and a spatial intensity distribution

$S(x_{\mathrm{in}})$. In this case, the factorisation leads to

$$\begin{aligned} T(x_{\mathrm{out}}) &= \int_{\lambda}\int_{x_{\mathrm{in}}} \eta(\lambda)\,\delta(x_{\mathrm{out}} - f(\lambda, x_{\mathrm{in}}))\,S_\lambda(\lambda)\,S(x_{\mathrm{in}})\,\mathrm{d}\lambda\,\mathrm{d}x_{\mathrm{in}} \\ &= \int_{x_{\mathrm{in}}} \eta\Big(\hat{f}(x_{\mathrm{in}};\,x_{\mathrm{out}})\Big)\;S_\lambda\Big(\hat{f}(x_{\mathrm{in}};\,x_{\mathrm{out}})\Big)\;S(x_{\mathrm{in}})\,\mathrm{d}x_{\mathrm{in}} \quad , \end{aligned} \tag{5}$$

with $\hat{f}(x_{\mathrm{in}};\,x_{\mathrm{out}})$ being the inverse of $f(\lambda, x_{\mathrm{in}})$ when solved for λ for a fixed x_{out}. Note that $\hat{f}$ is a function of one variable (x_{in}) only, because x_{out} is a parameter dictated by the point under investigation.[19] In practice, the angles α and β will vary only over small ranges. Linearisation of the sines in the grating equation, Eq. (1) with minus-sign, then provides the relation between the three coordinates (subscripts 0 denote reference positions),

$$M_{\mathrm{s}}(x_{\mathrm{in},0} - x_{\mathrm{in}}) - (x_{\mathrm{out},0} - x_{\mathrm{out}}) = \zeta\lambda\,. \tag{6}$$

This relation implies true imaging (magnification M_{s}, determined by the spectrograph) and linear diffraction (grating constant ζ, characterizing the grating). Both M_{s} and ζ are dimensionless quantities. The Dirac delta function in Eq. (3) selects $f(\lambda, x_{\mathrm{in}}) = x_{\mathrm{out}}$ from Eq. (6). For this simple relation, the 'inverse' function is cast explicitly by

$$\hat{f}(x_{\mathrm{in}};\,x_{\mathrm{out}}) = \lambda = \frac{(x_{\mathrm{out}} - x_{\mathrm{out},0}) - M_{\mathrm{s}}(x_{\mathrm{in}} - x_{\mathrm{in},0})}{\zeta}\,. \tag{7}$$

Substitution of this specific inverse results in $T(x_{\rm out})$ taking the form

$$T(x_{\rm out}) = \int_{x_{\rm in}} \eta\left(\frac{(x_{\rm out}-x_{\rm out,0}) - M_{\rm s}(x_{\rm in}-x_{\rm in,0})}{\zeta}\right) \times S_\lambda\left(\frac{(x_{\rm out}-x_{\rm out,0}) - M_{\rm s}(x_{\rm in}-x_{\rm in,0})}{\zeta}\right)\, S(x_{\rm in})\,{\rm d}x_{\rm in} \tag{8a}$$

$$= \left[(\eta \times S_\lambda) * S\right](x_{\rm out})\ . \tag{8b}$$

This outcome represents a convolution ($*$) of a spectral distribution (that is shifted forward and compressed in its argument) with a spatial distribution.

As it is the spatial pattern S that is of interest, it needs to be reconstructed from the image T via a deconvolution procedure, once the spectral dependency ($\eta \times S_\lambda$) is known. The latter can be obtained by recording $T(x_{\rm out})$ using a narrow entrance slit (located at the reference position $x_{\rm in,0}$ and illuminated with intensity S_0), for which case

$$S_\delta(x_{\rm in}) = S_0\,\delta(x_{\rm in}-x_{\rm in,0})\ , \tag{9}$$

in which the subscript δ indicates the narrow entrance slit. For this particular input the convolution ends up as

$$T_\delta(x_{\rm out}) \;=\; \eta\left(\frac{(x_{\rm out}-x_{\rm out,0})}{\zeta}\right) S_\lambda\left(\frac{(x_{\rm out}-x_{\rm out,0})}{\zeta}\right) \times S_0\ , \tag{10}$$

which, of course, is just a spectrum. Since it is used to deconvolve an OMAgraph for the spectral distribution, we will refer to it as *spectral reference function*, $R(x_{\rm out})$, in the following. In order to make the connection to Eq. (8), Eq. (10) is rewritten as

$$\eta\left(\frac{(x_{\rm out}-x_{\rm out,0}) - M_{\rm s}(x_{\rm in}-x_{\rm in,0})}{\zeta}\right) S_\lambda\left(\frac{(x_{\rm out}-x_{\rm out,0}) - M_{\rm s}(x_{\rm in}-x_{\rm in,0})}{\zeta}\right) = \frac{R(x_{\rm out}-M_{\rm s}(x_{\rm in}-x_{\rm in,0}))}{S_0}\ . \tag{11}$$

Back-substitution of this expression for the spectral reference function R turns the convolution, Eq. (8), into its final form,

$$T(x_{\rm out}) = \int_{x_{\rm in}} R(x_{\rm out} - M_{\rm s}(x_{\rm in} - x_{\rm in,0}))\, \frac{S(x_{\rm in})}{S_0}\, {\rm d}x_{\rm in}\ . \tag{12}$$

Note that the grating efficiency $\eta(\lambda)$ cancelled, because it is present in both the spectral reference image R and the convolved image S. Neither does the grating constant ζ appear explicitly.

An example of the convolving action of the spectrograph is presented in Fig. 3, where three OMAgraphs of a square grid under bichromatic illumination are shown for three different entrance slit widths. Like all subsequent OMAgraphs in this chapter, they were taken through an imaging spectrograph (Acton Research Corporation SpectraPro300i $f/4$ with a 2400 gr/mm grating). The slight curvature of the OMAgraphs is an imaging artifact of the spectrograph; it will be seen below that the deconvolution algorithm can automatically correct for it. Also, the magnification of the spectrograph is 14% larger along the $\hat{x}$-direction than along the $\hat{y}$-direction, which causes the square grid to look rectangular; this is not corrected for in the examples presented here. (But it is corrected for in the companion paper T2.) An intensified CCD camera (Princeton Instruments ICCD-512-T: 512×512 pixels; $\varnothing$25 mm intensifier; 16 bit dynamic range) at the exit port of the spectrograph recorded the OMAgraphs corresponding to regions imaged by a camera objective (Nikon UV-Nikkor 105 mm $f/4.5$) on the entrance slit of the spectrograph. In the images of Fig. 3 the slit width d_s is increased from 40 μm (b) to its maximum value of 3.10 mm (d). The $d_s = 40$ μm OMAgraph is taken as the infinitesimally narrow slit measurement of the spectral reference function

for deconvolution purposes. It contains two lines of the mercury spectrum.[20] Further decrease of the entrance slit width did not result in a narrower line profile, but only decreased the intensity levels. On broadening the entrance slit, the 'spatial' images produced by both lines broaden accordingly (c) and eventually overlap (d). On top of every OMAgraph, the intensity along a single strip is plotted, a horizontal cross section (fixed y) at the height of the white arrows.

The notion of recording a somehow distorted version of the original input image is not unfamiliar. In general imaging experiments, for example, artifacts are introduced by aberrations of the optical detection system and its limited resolution. In image restoration literature[21,22,24] such artifacts are often described in terms of a point-spread function (psf), which describes how a point source is mapped onto the image plane by a convolution procedure that is similar to the one presented here. Although R is not exactly a psf, it can be thought of as acting like one, by linking the input image to the output. Note, however, that R is not spatially invariant.

3. Deconvolution of OMAgraphs

The main result of the previous section is an analytical expression, Eq. (12), for the intensity distribution in the exit plane of a spectrograph as a convolution of the spatial and spectral intensity distributions incident on the entrance slit. In practice, one is often interested in just the spatial intensity distribution. This section focusses on the deconvolution of the spectrograph output with the spectral distribution, which should provide the desired spatial input intensity distribution. Special attention is paid to the role of noise, accumulated on the spectrograph output, in the deconvolution pro-

cedure. First, a straightforward deconvolution algorithm is discussed and evaluated, followed by a more sophisticated approach which yields much better results.

A. Analytical formulation

The general formulation of the convolution as given in Eq. (12) forms the starting point for the deconvolution procedure, and is here restated as

$$T(x_{\mathrm{out}}) = \int_{x_{\mathrm{in}}} R(x_{\mathrm{out}} - M_{\mathrm{s}}(x_{\mathrm{in}} - x_{\mathrm{in},0}))\, G(x_{\mathrm{in}})\, \mathrm{d}x_{\mathrm{in}} + \mathcal{N}(x_{\mathrm{out}}) \;, \tag{13}$$

where the normalised pattern S/S_0 is substituted by G for notational convenience, and a noise term $\mathcal{N}$ is added. The importance of this extra contribution to the formalism will become clear below. The reconstruction of the original spatial distribution S requires a deconvolution of the convolved image T ('broad' entrance slit measurement) with respect to the spectral dependency R ('narrow' slit measurement). This inverse problem can be solved conveniently via Fourier transformation. The Fourier transform (FT) and its inverse are defined by

$$\mathrm{FT}\left[F(x)\right] \equiv \widetilde{F}(k) \stackrel{\mathrm{def}}{=} \int_{x} F(x)\, \exp(-ikx)\, \mathrm{d}x \tag{14a}$$

$$\mathrm{FT}^{-1}\left[\widetilde{F}(k)\right] \equiv F(x) \stackrel{\mathrm{def}}{=} \frac{1}{2\pi} \int_{k} \widetilde{F}(k)\, \exp(ikx)\, \mathrm{d}k \;, \tag{14b}$$

respectively. Application of the Fourier transform to the convolution, Eq. (13), yields

$$\widetilde{T}(k) = \widetilde{R}(k)\, \exp(iM_{\mathrm{s}}kx_{\mathrm{in},0}) \times \widetilde{G}(M_{\mathrm{s}}k) + \widetilde{\mathcal{N}}(k) \;, \tag{15}$$

taking due care of the magnification M_{s} and the shift $x_{\mathrm{in},0}$.[14] The convolution in direct space thus becomes a regular product in reciprocal (k) space, for which the components have decoupled.

Once the Fourier components $\widetilde{T}(k)$ and $\widetilde{R}(k)$ have been assessed, the Fourier components of the original image at the entrance slit can formally be obtained analytically by solving Eq. (15) for $\widetilde{G}$,

$$\widetilde{G}(M_{\mathrm{s}}k) = \frac{\widetilde{T}(k) - \widetilde{\mathcal{N}}(k)}{\widetilde{R}(k)\,\exp(iM_{\mathrm{s}}kx_{\mathrm{in},0})}\,, \tag{16}$$

and after an inverse Fourier transform the entrance slit image in direct space, $G(x_{\mathrm{in}})$, results.

In practical OMA imaging, the signal at the exit plane is recorded by a CCD camera, requiring a reformulation (discretization) of the analytical problem. Details on this procedure are given in the Appendix.

Fig. 4 demonstrates the effect of the analytical deconvolution procedure (Appendix, Eq. (A6)) to the data of Fig. 3. In this measurement, there was noise contained in both the spectral reference signal (Fig. 3(b)) and the convolved distribution (Fig. 3(d)). The straightforward deconvolution result is extremely noisy (image 4(a)) and contains virtually no useful quantitative information, as can be seen clearly in the graph of the single-strip cross section above the image. Note that this occurs in spite of the fact that both input images (Figs. 3(b) and (d)) are not particularly noisy. A closer look at the reciprocal space components (image 4(b)) indicates that the resulting noise can be attributed to large high-k-components[25] that do not cancel each other on inverse Fourier transformation. The Fourier transform diagonalizes the convolution problem and hence the deconvolution. However, since many of its high k-components are small, the inverse problem is (numerically) ill-conditioned and sensitive to small variations (noise!) in the recorded data. In a more physical explana-

tion, the deconvolution algorithm attempts to increase the contrast in the image, *i.e.* it attempts to restore structure in the input that was smeared out by the convolution: any structure that is present in the measured (convolved) image must have been more pronounced in the actual spatial distribution on the entrance slit (before convolution). There is, however, no way for the reconstruction algorithm to distinguish real data from noise, that is added during detection and superposes artificial structure (on a pixel-to-pixel basis, that is, high-frequency) onto the convolved signal. Since the analytical algorithm does not treat the noise any different than the 'real' data, it generates high k-components to account for the rapid pixel-to-pixel fluctuations in the measured (signal + noise) data. Unfortunately, it is not trivial to discriminate between 'real' data and noise.

Zero-padding the data (to array sizes up to 2^{13} pixels) to increase the resolution in k-space, did not alter the deconvolved result in direct space significantly. Additional windowing of the raw data to rigidly impose periodic boundary conditions as required for the Fourier transform is also insufficient to restore the original image. This means that the noise in the straightforward reconstruction is not a consequence of ill-satisfied periodic boundary conditions. At the cost of loosing spatial resolution, binning two neighbouring pixels is a trivial low-pass filter for the spectral and the convolved data. A subsequent deconvolution yields Fig. 4(c), which looks less noisy than (a). Indeed, its corresponding Fourier power spectrum (Fig. 4(d)) shows less contributions from high k-components. However, although the structure of the object can now be recognized, the S/N ratio in Fig. 4(c) is still poor, as compared to the input (Fig. 3(d); compare especially the two single-strip cross sections above each image).

Clearly, additional filtering is required to separate data from noise by suppressing 'exploded' $\widetilde{G}(k)$-contributions, especially at higher reciprocal space components. There exist filters that globally do this job,[22] but often they are *ad hoc*, and do not provide dramatic improvements. A more general algorithm dedicated to filtered deconvolution is presented below.

B. Linear Bayesian deconvolution filter

The goal is to reconstruct the original spatial pattern $\{G_n\}$ (a strip of the input) that gave rise to a measurement $\{T_m\}$ (a strip of the output) for a given spectral convolution function $\{R_m\}$. As the actual input $\{G_n\}$ is in principle unknown, many candidates might have given the same measured output (including noise), and one of these has to be selected. This boils down to finding the *most likely reconstruction of* $\{G_n\}$*, given the incomplete knowledge provided by the experiment.* 'Incomplete' means that the actual noise contribution is not known, for example. What we do know, however, is (*i*) a measurement of the output and (*ii*) the convolution model of Eq. (12), which links the output to the input. A Bayesian statistical analysis[23] prescribes a procedure for quantifying the term 'most likely' and for assigning a reliability to the reconstruction. For this purpose, the intensities of the input pattern, the measured data, and the noise are thought of as stochastic variables to which probability density functions (pdf) are associated, describing the probability that a stochastic variable takes a certain value. The spectral reference function is assumed to be well assessed as the response to a peak input. The principles of the data filter are presented in this paper; details can be found in the thesis of Tolboom.[14]

In this section, upper-case letters denote stochastic variables, the corresponding lower-case letters represent their actual values (data). Bold-face vector notation indicates that we deal with the entire set of N stochastic variables or their specific values, respectively; subscripts indicate a single component.

Given the pdf's and the experimental fact that the measurement of $\boldsymbol{T}$ resulted in the particular outcome $\boldsymbol{t}$ (denoted by $\boldsymbol{T} = \boldsymbol{t}$), the reconstruction problem is equivalent to calculating the conditional expectation values

$$\mathbb{E}\big[\boldsymbol{G}\,\big|\,\boldsymbol{T} = \boldsymbol{t}\big] \equiv \boldsymbol{\mathcal{F}}(\boldsymbol{t}) \tag{17}$$

as a function $\boldsymbol{\mathcal{F}}$ of the measured output $\boldsymbol{t}$. These conditional expectation values are the best *a posteriori* estimates of the *a priori* input $\boldsymbol{g}$, given the measured outcome $\boldsymbol{T} = \boldsymbol{t}$.[26] The criterion 'best' is defined as that particular $\boldsymbol{\mathcal{F}}(\boldsymbol{t})$ that minimizes the mean square errors

$$\mathbb{E}\big[|G_n - \mathcal{F}_n(\boldsymbol{t})|^2\big] \quad \forall_n \tag{18}$$

as based on the measured $\boldsymbol{T} = \boldsymbol{t}$. This section provides the best reconstruction by a direct calculation of the conditional expectation value. It will be Bayes' theorem which is used to link *a priori* knowledge of both input and output to *a posteriori* knowledge. Different ways exist to arrive at the final result, for example by a direct minimization of the mean square error, either analytically or graphically.[14] However, we think that the formulation presented below gives the clearest insight in the mathematics involved.

In the statistical approach, both the input image $\boldsymbol{G}$ and the noise $\boldsymbol{\mathcal{N}}$ are modeled

by stochastic variables, that is,

$$G_n = c + \sigma X_n \quad \forall_n \qquad \text{and} \qquad \mathcal{N}_m = b + \tau Y_m \quad \forall_m \ , \tag{19}$$

respectively, where the X_n's and Y_m's are taken *a priori* from mutually independent standard normal distributions, a choice that will be commented upon below. It is assumed in the model that all pixels in a strip are characterized by one set of parameters $\{c, \sigma, b, \tau\}$. The mutual independence (denoted by '$\perp\!\!\!\perp$') of all distributions $\boldsymbol{X}$ and $\boldsymbol{Y}$ implies that

$$\begin{cases} X_m \perp\!\!\!\perp X_n (m \neq n) & \text{the pdf of the optics before the spectrograph} \\ & \text{is neglected;} \\ Y_m \perp\!\!\!\perp Y_n \, (m \neq n) & \text{the noise is accounted for per individual pixel;} \\ X_m \perp\!\!\!\perp Y_n \quad \forall_{m,n} & \text{the noise is not correlated to the signal at all.} \end{cases} \tag{20}$$

Since G (the normalized input signal) is a dimensionless quantity, the parameters c and σ are dimensionless, too. On the other hand, b and τ have the same dimension as $\mathcal{N}$, which is [count]. It can be shown[14] that the pdf (ϱ) of the original image G_n taking on the value g_n can be derived from the standard normal distribution as

$$\varrho_{G_n}(g_n) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2} \left(\frac{g_n - c}{\sigma} \right)^2 \right] \quad \forall_n \tag{21}$$

and similarly for the noise term

$$\varrho_{\mathcal{N}_m}(\nu_m) = \frac{1}{\sqrt{2\pi}\,\tau} \exp\left[-\frac{1}{2} \left(\frac{\nu_m - b}{\tau} \right)^2 \right] \quad \forall_m \ . \tag{22}$$

The parameters b and c are the averages of the noise and the normalised input image, respectively, and τ^2 and σ^2 are the variances in the corresponding signals. The

distribution ϱ_{G_n} (Eq. (21)) illustrates the dilemma in choosing σ. On the one hand, σ has to be sufficiently large to allow all the reasonable data, but on the other hand, it should be small enough to exclude negative values (the input g_m is necessarily positive), effectively by assigning small probabilities to them. The exclusion of negative numbers is not an issue for the noise model. The addition of 'mean' c to the input is an extension to existing analytical models (like documented by MacKay[27]), since it shifts the *a priori* pdf to the (positive) intensities expected on physical principles. A physical approach to determine σ and τ is discussed at the end of this section.

The discretized form of the convolution model, Eq. (A2), provides the link between the assumed stochastic variables Eq. (19) and the output, which we can measure. These output values,

$$T_m = \sum_n R_{m-n+n_0}\, G_n + \mathcal{N}_m \quad \forall_m \ , \tag{23}$$

are also treated as stochastic parameters in the current approach. Once the *values* $\boldsymbol{g} = (g_0, \ldots, g_{N-1})$ of the input pattern $\boldsymbol{G} = (G_0, \ldots, G_{N-1})$ are specified, the pdf for the output follows from this equation as the conditional pdf (or: likelihood)

$$\varrho_{T_m}\big(t_m \,\big|\, \boldsymbol{G} = \boldsymbol{g}\big) = \frac{1}{\sqrt{2\pi}\,\tau} \exp\left[-\frac{1}{2}\left(\frac{t_m - [\sum_n R_{m-n+n_0}\, g_n + b]}{\tau}\right)^2\right] \quad \forall_m \ . \tag{24}$$

This is an equation for a normal distribution once more, but now centered around the value $(\sum_n R_{m-n+n_0}\, g_n + b)$ with variance τ.

The problem in OMA imaging, however, is that we measure the *output* $\boldsymbol{T} = \boldsymbol{t}$, from which we want to retrieve the pdf for the *input* pattern. This (*a posteriori*) pdf is also a conditional probability density function, $\varrho_{G_q}\big(g_q \,\big|\, \boldsymbol{T} = \boldsymbol{t}\big)$, but it cannot be

calculated readily. Fortunately, it suffices for our problem to calculate the conditional expectation value of G_q, which is defined for discrete stochastic variables as

$$\begin{aligned}\mathbb{E}\big[G_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] &= \sum_{g_q} g_q \,\mathbb{P}\big[G_q = g_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] \\ &= \sum_{\boldsymbol{g}} g_q \,\mathbb{P}\big[\boldsymbol{G}=\boldsymbol{g} \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] \quad \forall_q \,. \end{aligned} \tag{25}$$

(Note that these relations contain a summation over *values* g_q and $\boldsymbol{g}$ rather than a summation over *components* q. Since the individual components are independent, this does not introduce additional components.[14]). The conditional probability $\mathbb{P}$ (on an event; for discrete stochastic variables) that was introduced in Eq. (25) is related to the conditional pdf ϱ (of a value; for continuous stochastic variables) via

$$\mathbb{P}\big[G_q \in (a,b) \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] = \int_a^b \varrho_{G_q}\big(g_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big)\,\mathrm{d}g_q \,. \tag{26}$$

It is at this point that Bayes' theorem

$$\mathbb{P}\big[\boldsymbol{G}=\boldsymbol{g} \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] = \frac{\mathbb{P}\big[\boldsymbol{T}=\boldsymbol{t} \,\big|\, \boldsymbol{G}=\boldsymbol{g}\big] \times \mathbb{P}[\boldsymbol{G}=\boldsymbol{g}]}{\mathbb{P}[\boldsymbol{T}=\boldsymbol{t}]} \tag{27}$$

enters the statistical approach for the deconvolution by reversing the 'information' ($\boldsymbol{T}=\boldsymbol{t}$) and 'unknown' ($\boldsymbol{G}=\boldsymbol{g}$) arguments of the likelihood, turning Eq. (25) into

$$\mathbb{E}\big[G_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] = \sum_{\boldsymbol{g}} g_q \,\frac{\mathbb{P}\big[\boldsymbol{T}=\boldsymbol{t} \,\big|\, \boldsymbol{G}=\boldsymbol{g}\big] \times \mathbb{P}[\boldsymbol{G}=\boldsymbol{g}]}{\mathbb{P}[\boldsymbol{T}=\boldsymbol{t}]} \quad \forall_q \,. \tag{28}$$

The denominator does not depend on $\boldsymbol{g}$, so it can be taken out of the sum. Additionally, the decomposition relation of probabilities is applied to the denominator

$$\mathbb{P}[\boldsymbol{T}=\boldsymbol{t}] = \sum_{\boldsymbol{g}} \mathbb{P}\big[\boldsymbol{T}=\boldsymbol{t} \,\big|\, \boldsymbol{G}=\boldsymbol{g}\big] \times \mathbb{P}[\boldsymbol{G}=\boldsymbol{g}] \,, \tag{29}$$

and the conditional expectation value becomes

$$\mathbb{E}\big[G_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] = \frac{\sum\limits_{\boldsymbol{g}} g_q \,\mathbb{P}\big[\boldsymbol{T}=\boldsymbol{t}\,\big|\,\boldsymbol{G}=\boldsymbol{g}\big] \times \mathbb{P}[\boldsymbol{G}=\boldsymbol{g}]}{\sum\limits_{\boldsymbol{g}} \mathbb{P}\big[\boldsymbol{T}=\boldsymbol{t}\,\big|\,\boldsymbol{G}=\boldsymbol{g}\big] \times \mathbb{P}[\boldsymbol{G}=\boldsymbol{g}]} \quad \forall_q \,, \tag{30}$$

showing that the numerator and the denominator contain the same probabilities, and that the numerator carries the additional value g_q that is averaged. This is a relation familiar from statistical physics, for example, where the population of a grand canonical ensemble is normalised by the partition function.[28]

In the continuum limit of the stochastic variables, the last equation contains Riemann summations that go over into integrals over pdf's for every pixel q,

$$\begin{aligned}
\mathbb{E}\big[G_q \,\big|\, \boldsymbol{T}=\boldsymbol{t}\big] &= \frac{\int\limits_{\mathbb{R}^N} g_q \,\varrho_{\boldsymbol{T}}\big(\boldsymbol{t}\,\big|\,\boldsymbol{G}=\boldsymbol{g}\big) \times \varrho_{\boldsymbol{G}}(\boldsymbol{g})\,\mathrm{d}\boldsymbol{g}}{\int\limits_{\mathbb{R}^N} \varrho_{\boldsymbol{T}}\big(\boldsymbol{t}\,\big|\,\boldsymbol{G}=\boldsymbol{g}\big) \times \varrho_{\boldsymbol{G}}(\boldsymbol{g})\,\mathrm{d}\boldsymbol{g}} \\
&= \frac{\int\limits_{\mathbb{R}^N} g_q \,\exp\!\Big[-\tfrac{1}{2}\,\big|\tfrac{\boldsymbol{t}-\boldsymbol{R}*\boldsymbol{g}-b}{\tau}\big|^2\Big] \times \exp\!\Big[-\tfrac{1}{2}\,\big|\tfrac{\boldsymbol{g}-c}{\sigma}\big|^2\Big]\,\mathrm{d}\boldsymbol{g}}{\int\limits_{\mathbb{R}^N} \exp\!\Big[-\tfrac{1}{2}\,\big|\tfrac{\boldsymbol{t}-\boldsymbol{R}*\boldsymbol{g}-b}{\tau}\big|^2\Big] \times \exp\!\Big[-\tfrac{1}{2}\,\big|\tfrac{\boldsymbol{g}-c}{\sigma}\big|^2\Big]\,\mathrm{d}\boldsymbol{g}} \\
&= \frac{\int\limits_{\mathbb{R}^N} g_q \,\exp\!\Big[-\tfrac{1}{2}\sum\limits_m \Big(\big|\tfrac{t_m-\sum_n R_{m-n+n_0}\,g_n-b}{\tau}\big|^2 + \big|\tfrac{g_m-c}{\sigma}\big|^2\Big)\Big]\,\mathrm{d}\boldsymbol{g}}{\int\limits_{\mathbb{R}^N} \exp\!\Big[-\tfrac{1}{2}\sum\limits_m \Big(\big|\tfrac{t_m-\sum_n R_{m-n+n_0}\,g_n-b}{\tau}\big|^2 + \big|\tfrac{g_m-c}{\sigma}\big|^2\Big)\Big]\,\mathrm{d}\boldsymbol{g}} \,,
\end{aligned} \tag{31}$$

where it is used that all stochastic variables are independent and that the pre-factors in the numerator and the denominator cancel. Because the $\boldsymbol{g}$ and $\boldsymbol{R}$ values in the first part of the exponentials are coupled via the convolution, these integrals cannot be performed analytically in direct space. Substitution of the Fourier transforms of the $\boldsymbol{g}$, $\boldsymbol{R}$, $\boldsymbol{t}$, b, and c values provides a convenient change of the coordinates of integration that decouples the integrand once more. However, this change of coordinates is rather intricate and introduces some complications with respect to the coordinates and domain of integration (see appendix 6.A in the thesis of Tolboom[14]). Analytical

evaluation via reciprocal space of the integrals in Eq. (31) eventually leads to the key formula

$$\mathbb{E}\big[G_q \,\big|\, \boldsymbol{T} = \boldsymbol{t}\big] = \frac{c\left(\frac{\tau}{\sigma}\right)^2 - b\widetilde{R}_{k=0}}{\widetilde{R}_{k=0}^2 + \left(\frac{\tau}{\sigma}\right)^2} + \mathrm{FT}^{-1}\left[\frac{\widetilde{R}^*_{k,n_0}\,\widetilde{t}_k}{\left|\widetilde{R}_{k,n_0}\right|^2 + \left(\frac{\tau}{\sigma}\right)^2}\right] \tag{32}$$

for the restoration of the unknown original input image. (The tilde denotes a Fourier transform; see Appendix.)

Equation (32) is the main result of this paper, describing the reconstruction of the most probable input pattern by means of filtered deconvolution of the measured output. Since the expression is analytically closed, it enables a straightforward implementation in computer code without iterative loops. In Eq. (32) the first term denotes an offset, around which data is scattered by the FT^{-1} in the second term. The parameter c is always taken as the average over the convolved image $\boldsymbol{t}$.

The filter has only one free model parameter, the fraction σ/τ, instead of two (σ and τ separately) as might perhaps be expected from the supposition of the stochastic variables of Eq. (19). This effective filter parameter can be interpreted as a measure for the *contrast* between the 'real' structure in the signal (*e.g.* the grid lines and the numbers) and the 'noise' accumulated in the measurement.

Equation (32) is the linear Bayesian-filtered version of the straightforward deconvolution, Eq. (A6). To compare the two results we rearrange the argument of the FT^{-1} in Eq. (32) according to

$$\frac{\widetilde{R}^*_{k,n_0}\,\widetilde{t}_k}{\left|\widetilde{R}_{k,n_0}\right|^2 + \left(\frac{\tau}{\sigma}\right)^2} \;=\; \frac{\left|\widetilde{R}_{k,n_0}\right|^2}{\left|\widetilde{R}_{k,n_0}\right|^2 + \left(\frac{\tau}{\sigma}\right)^2}\,\frac{\widetilde{t}_k}{\widetilde{R}_{k,n_0}} \;=\; \frac{1}{1 + \left(\frac{\tau}{\sigma}\right)^2\left|\widetilde{R}_{k,n_0}\right|^{-2}}\,\frac{\widetilde{t}_k}{\widetilde{R}_{k,n_0}}\,. \tag{33}$$

This shows that the filter effectively suppresses all k-components of the straightforward deconvolution by a factor $[1 + (\tau/\sigma)^2|\widetilde{R}_{k,n_0}|^{-2}] \geq 1$. Alternatively, the power of

the Fourier transformed components within the inverse FT can be written as

$$\underbrace{\left|\frac{\widetilde{R}^{*}_{k,n_0}}{\left|\widetilde{R}_{k,n_0}\right|^2+\left(\frac{\tau}{\sigma}\right)^2}\right|^2}_{\substack{\text{filtered deconvolution}\\ \text{function}}} = \underbrace{\left[\frac{\left|\widetilde{R}_{k,n_0}\right|^2}{\left|\widetilde{R}_{k,n_0}\right|^2+\left(\frac{\tau}{\sigma}\right)^2}\right]^2}_{\text{filter}} \times \underbrace{\left|\frac{1}{\widetilde{R}_{k,n_0}}\right|^2}_{\substack{\text{deconvolution}\\ \text{function}}} . \qquad (34)$$

Written this way, the Bayesian filter with the choice of normal distributions for the stochastic variables is seen to yield similar results as the Wiener deconvolution filter.[22]

The pre-factor 'filter' is a measure for the power attributed by the filter to the 'real' data, depending only on the spectral reference function R and the regularization parameter σ/τ. It is clear that if τ/σ approaches zero (*i.e.* the measured structures are completely due to structure in the input image), the non-filtered deconvolution, Eq. (A6), will remain. In that limit, the offset in Eq. (32) approaches $-b/\widetilde{R}_{k=0}$, corresponding to subtraction of the (constant) noise level b from the deconvolved data. The results in this paper are obtained for $b = 0$, so there is no such additional offset correction for accumulated noise. The limit of $\tau/\sigma \to \infty$, on the other hand, would physically correspond to a measured output that is dominated by noise. In this case, the filter effectively suppresses *all* Fourier components, resulting in the prediction of a flat input distribution.

The graphs in Fig. 5 illustrate the action of the (filtered) deconvolution in reciprocal space for the single-strip data $\widetilde{\boldsymbol{R}}$ of Fig. 3(b). At the right in Fig. 5 are the powers of the deconvolution function (b) and of the filtered deconvolution function (d) on a linear scale. The spectrum (b) is dominated by high-k components, and these cause the extreme noise in the straightforward reconstruction (*e.g.* Fig. 4(a)). As can be seen from Fig. 5(c), it is exactly these high-k components that are strongly sup-

pressed by the 'filter'-factor, Eq. (34). As a result, the power spectrum of the filtered deconvolution (Fig. 5(d), which should be compared to Fig. 5(b)) is dominated by (real) structure in the low-k components, a contribution that is present as only a minor wrinkle in the unfiltered spectrum (b). The exact shape of the filter (c) strongly depends on the shape of $|\widetilde{R}_k|^2$ (Fig. 5(a)). In this particular case, the presence of two spectral lines causes the oscillations in $|\widetilde{R}_k|^2$, and in all plots derived from it. For a single-line spectrum, for example, all curves would have been smooth.

Fig. 6 illustrates the dependence of the reconstruction on the filter parameter σ/τ in some more detail. The 'turning point' for the filter behaviour, as can be seen from Eq. (34), lies at about $\sigma/\tau \approx |R|^{-1}$. For $\sigma/\tau \lesssim |R|^{-1}$, the structure in the measured image is taken as dominated by noise, whereas for $\sigma/\tau \gtrsim |R|^{-1}$ it is attributed to the input image itself. The former situation results in a relatively uniform input image (as reflected in the smooth reconstruction of Fig. 6(a)) onto which read-out noise has imposed structure. The latter situation corresponds to an input image with rich structure onto which relatively little noise has accumulated. The deconvolution therefore produces a wildly fluctuating image (c), approaching the non-filtered deconvolution result. In the best result (here $\sigma/\tau = 6$ count^{-1}; see below), the original object is very well reproduced, and the other parts of the image are nearly empty. (Ideally, they would be completely dark.) The horizontal cross sections (white arrows, similar to those in Figs. 3 and 4) emphasize the truthful recovery of the input image; three distinct minima represent the grid lines in the single-strip data, and an intermediate structure is present that comes from cutting the '68'. The 'grid' minima are separated by approximately 43 pixels and they are somewhat broader than 4 pixels (full width

at half maximum), in perfect agreement with the printed grid of 1:10 for the line thickness. Still, there are a few small undershoots (oscillations and negative numbers), also in the best result. As discussed after the equations Eq. (21) and Eq. (22) for the pdf's, this does not come as a surprise.

Although rigorous mathematical procedures exist for estimating the regularization parameter σ/τ (see e.g. Engl *et al.*[29]), we have opted for a more physical approach. The physical situation is that light is transmitted through the entrance slit only. Thus, after deconvolution the signal should ideally be confined to a finite area on the CCD chip corresponding to the entrance slit image, while the rest of the chip is empty. Define the *image contrast* as $\langle P(\text{in})\rangle/\langle P(\text{out})\rangle$, with $\langle P(\text{in})\rangle$ the average power 'inside' the reconstructed entrance slit image (pixel numbers 176 through 335), and $\langle P(\text{out})\rangle$ the average power 'outside' this part of the image (pixel numbers 1 through 165). The 'best' σ/τ should then maximize the image contrast. From a plot of the image contrast against σ/τ (Fig. 7), this is found to be the case for $\sigma/\tau = 6$ count^{-1}. Note that the maximum in Fig. 7 is rather broad, so that the filter performance is not very critical to the exact value of σ/τ that is used.

A final check on the quality of the reconstructed image is to re-convolve it with the spectral reference function as done in Fig. 8. The difference of this result with the measured data is also indicated, being one to two orders of magnitude smaller than the actual data. The average power in the re-convolved image is 0.946 times the average power contained in the original data, so the filter attributes approximately 5% of the power to noise for $\sigma/\tau = 6$ count^{-1}.

4. Some properties of OMA imaging & the reconstruction of OMAgraphs

A. Linearity

The linearity of the deconvolution filter can be demonstrated by considering the deconvolution of any linear combination of two measurements $\boldsymbol{T} = (a_1\boldsymbol{t}_1 + a_2\boldsymbol{t}_2)$, given by

$$\begin{aligned}\mathbb{E}\big[G_q\,\big|\,\boldsymbol{T} = a_1\boldsymbol{t}_1 + a_2\boldsymbol{t}_2\big] &= (a_1 + a_2)\,\frac{c\left(\frac{\tau}{\sigma}\right)^2 - b\widetilde{R}_{k=0}}{\widetilde{R}_{k=0}^2 + \left(\frac{\tau}{\sigma}\right)^2} + \\ &\qquad + \mathrm{FT}^{-1}\left[\frac{\widetilde{R}^*_{k,n_0}\,\{a_1 t_{1k} + a_2 t_{2k}\}}{\left|\widetilde{R}_{k,n_0}\right|^2 + \left(\frac{\tau}{\sigma}\right)^2}\right] \\ &= a_1\mathbb{E}\big[G_q\,\big|\,\boldsymbol{T} = \boldsymbol{t}_1\big] + a_2\mathbb{E}\big[G_q\,\big|\,\boldsymbol{T} = \boldsymbol{t}_2\big]\ . \end{aligned} \tag{35}$$

The factor $(a_1 + a_2)$ before the offset term is maybe not obvious. Since the total signal now consists of two separate measurements (t_1 and t_2) multiplied by the factors a_1 and a_2, the average value c and the average noise level b of the two images need to be multiplied by the same factor. Note that the linearity is a consequence of our specific choice of standard normal distributions in Eq. (19), rather than that it has been pre-supposed in the derivation of Eq. (32). The physical implication for OMA imaging is that (spectral) structures that do not overlap in an OMAgraph can be treated separately. This allows various sources with different spectral distributions to be recorded in a single OMAgraph, as long as their contributions do not overlap in the exit plane. (See Fig. 3(c), the deconvolved image can be found in Tolboom's thesis.[14]) Thus, the factorization requirement stated in Eq. (4) is seen not to be any more severe for OMA imaging than for any other kind of spectrally selective imaging.

B. 2-D image reconstruction

The spectrograph, like any other optical device, introduces imaging defects. Since the impact of these effects increases with the distance from the optical axis, especially the non-diffracting ($\hat{y}$) dimension will suffer from defects. The imaging artifacts of the particular system used in our experiments can be judged from Fig. 3(a). The two spectral lines (images of a straight entrance slit) are curved, and they become a bit less sharp towards the upper and lower ends. Both effects do not noticeably vary over the width of the image. Since horizontal cross sections are treated individually, this justifies the perfect-imaging assumption of Eq. (3). Furthermore, the strip-by-strip deconvolution provides two additional advantages related to inherent properties of deconvolution. If the spectral reference function is recorded with the same spectrograph settings as the image (but with a narrow entrance slit, of course; compare Figs. 3(a) and (c)), the reference image will contain a spectral reference function on each image line (pixel row). The deconvolution will (*i*) consider the (1-D) psf as part of the spectral reference function and it will (*ii*) center the deconvolved image around the origin.[30] Thus, if individual strips of an OMAgraph are deconvolved with a spectral reference function of the corresponding strip in a spectral reference image (like that of Fig. 4(a)), both the additional blurring and the curvature are corrected for. In some practically complicated cases, the spectral reference function cannot be determined for every strip, or needs to be averaged for sufficient signal-to-noise ratio. In such cases, the reconstruction will still contain residual blurring and/or the curvature will persist. Nevertheless, these images can often be used in quantitative studies, as

shown in the companion paper T2.

C. Improvements and restrictions

A main advantage of the Bayesian deconvolution filter presented here is that it is expressed in closed form, Eq. (32). Thus, the expectation values can be calculated directly, which obviates the need for CPU-intensive optimization schemes. This, however, has its price. We had to assume linear imaging (Eq. (6)), and found that the linearity of the filter in fact arises from our choice of normally distributed stochastic variables (Eq. (19)). The linear imaging assumption was required for the spectrograph output to take on the form of a convolution (Eq. (12)). It may break down for extended spectral structures, depending also on the equipment. Probably, the best check of whether or not this assumption is justified will be empirical.

The choice of normal distributions for the stochastic variables is in itself not expected to be a serious restriction, but it does allow (unphysical) negative values in the final result (see the discussion of Fig. 6). More elaborate distributions could remedy this, but probably at the expense of not yielding a closed expression for the filter any more. This, for example, is the case with the Maximum Entropy method,[31] where positivity of the reconstruction is assured, but a numerical global optimization algorithm is required to find it.

Finally, there is the assumption of (piecewise) factorizability of the input (Eq. (4)). We do not consider this a restriction that is particular to OMA imaging, but rather one that holds for all spectrally selective quantitative imaging schemes. In fact, we consider it one of the advantages of using a spectrograph over, for instance, band-pass

filters: *if* the factorisability assumption breaks down, the spectrum will at least tell you that it does.

5. Summary and conclusion

Under the hypothesis of factorizability of the input to a linearized spectrograph into a purely spectral and a purely spatial part, the entangling of spatial and spectral information by an OMA set-up is described effectively by a convolution. Reconstruction is achieved by a dedicated linear Bayesian deconvolution filter, depending on one free model parameter only. The data filter is based on Gaussian probability density functions for the unknown spatial input and the accumulated noise, allowing for a closed analytical filter expression (Eq. 32). The resulting reconstruction shows clear contrast and a good reproduction of the factual input. Moreover, the algorithm prescribes a recipe for generating quantitatively interpretable data, thus satisfying an essential criterion for quantitative 2-D imaging.

Since OMA imaging requires just an imaging spectrograph and a single camera, this is a reliable, relatively cheap, and efficient technique for quantitative imaging experiments. A practical application is discussed in the companion paper on the following pages.

Acknowledgements

We appreciate the critical and constructive discussions we had with Marianna Sijtsema. This work was made possible by financial support of the Technology Foundation (STW), the applied science division of the Netherlands Organisation for Scientific Re-

search (NWO) and the technology program of the Ministry of Economic Affairs.

6. Appendix: discretization

The photodetection chip in a CCD camera[34] has a finite (say N) number of pixels of finite (non-zero) dimension. These properties require a discretization of the analytical formulation, Eq. (13). Since the pixels have a finite dimension, they already integrate the signal over a finite region of x_{out}. All integrals over space therefore become a sum over the sub-integrals, *i.e.* the pixel values. The discretized version of Eq. (13) thus reads

$$T_{n_{\mathrm{out}}} = \sum_{n_{\mathrm{in}}} R_{n_{\mathrm{out}} - M_{\mathrm{s}}(n_{\mathrm{in}} + n_{\mathrm{in},0})} \times G_{n_{\mathrm{in}}} + \mathcal{N}_{n_{\mathrm{out}}} \ , \tag{A1}$$

where the indices n_{out} and n_{in} denote discrete positions (or, equivalently, pixel numbers) and replace the continuous parameters x_{out} and x_{in}, respectively. This discretization implies that not only the output signal at the exit port is discretized, but also the input signal at the entrance slit. Thus, the aim will be to estimate the signal G as a function of *pixel number* (discretized position). The index $[n_{\mathrm{out}} - M_{\mathrm{s}}(n_{\mathrm{in}} + n_{\mathrm{in},0})]$ of the spectral reference function R labels the (discretized) wavelength. It is expressed in terms of both the discrete positions n_{in} and n_{out}, and need not be an integer because it contains the magnification $M_{\mathrm{s}} \in \mathbb{R}$. From Eq. (6), however, we know that for every wavelength λ, any specific $M_{\mathrm{s}} n_{\mathrm{in}}$ (or $M_{\mathrm{s}} x_{\mathrm{in}}$) may be converted into a specific n_{out} (or x_{out}), where the indices n_{in} and n_{out} both denote pixel numbers. Therefore, the reference to 'in' and 'out' can be omitted and the indices can simply be any integer counter without reference to the specific planes (input or output).

Another consequence of the CCD chip having only a finite number of pixels, is that T and R can only be recorded for a finite number of points. The experimental

settings must, of course, be such that this finite segment is representative for the complete signal, so that this representative part can be put into the convolution. When assuming periodic boundary conditions (period N), the discretized convolution in direct space (Eq. (A1)) becomes

$$T_m = \sum_{n=0}^{N-1} R_{m-n+n_0} \times G_n + \mathcal{N}_m \ . \tag{A2}$$

All subsequent summations and products over indices will be from 0 through $N-1$.

The discretized form of the Fourier transform and its inverse are[32]

$$\widetilde{F}_k = \sum_{n=0}^{N-1} F_n \ \exp(-\frac{2\pi ikn}{N}) \qquad \text{and} \qquad F_n = \frac{1}{N} \sum_{k=0}^{N-1} \widetilde{F}_k \ \exp(\frac{2\pi ikn}{N}) \ , \tag{A3}$$

respectively. Like m and n, the index k is merely a counter, which denotes the reciprocal space component of a Fourier transformed signal. The discrete Fourier transforms are implemented in a computer program as Fast Fourier Transforms (FFT's) on 512 ($= 2^9$) data points. The discretized analogue of Eq. (15) is

$$\widetilde{T}_k = \widetilde{R}_{k,n_0} \times \widetilde{G}_k + \widetilde{\mathcal{N}}_k \qquad \text{with} \qquad \widetilde{R}_{k,n_0} \stackrel{\text{def}}{=} \widetilde{R}_k \ \exp(\frac{2\pi ik\, n_0}{N}) \ , \tag{A4}$$

and the relation for the discretized deconvolved signal in reciprocal space (Eq. (16)) becomes

$$\widetilde{G}_k = \frac{\widetilde{T}_k - \widetilde{\mathcal{N}}_k}{\widetilde{R}_{k,n_0}} \ . \tag{A5}$$

Since the phase factor $\exp(2\pi ikn_0/N)$ will always appear in combination with $\widetilde{R}_k$, it is absorbed in the latter for notational brevity; this is indicated by the additional subscript n_0 to $\widetilde{R}_{k,n_0}$. The pixel values in direct space are

$$G_n \ = \ \mathrm{FT}^{-1}\Big[\widetilde{G}_k\Big] \ = \ \mathrm{FT}^{-1}\left[\frac{\widetilde{T}_k - \widetilde{\mathcal{N}}_k}{\widetilde{R}_{k,n_0}}\right] \ = \ \frac{1}{N}\sum_{k=0}^{N-1}\left[\frac{\widetilde{T}_k - \widetilde{\mathcal{N}}_k}{\widetilde{R}_{k,n_0}}\right] \exp(\frac{2\pi ikn}{N}) \ , \tag{A6}$$

where the index n numbers the pixels (formally in the entrance slit plane). Since the Fourier algorithm is essentially just a mathematical trick, the physical information should be contained only in the *real* part of the FT^{-1}. The *imaginary* part ought to vanish. The program that performed the deconvolution, Eq. (A6), was tested successfully on computer generated data (not shown). The average power[33] in the imaginary part was always found to be about 25 orders of magnitude smaller than the average power in the real part, being zero indeed within the limit of computational accuracy.

References

1. A.C. Eckbreth. *Laser Diagnostics for Combustion Temperature and Species*. (Gordon & Breach, Amsterdam, NL, 1996), 2nd ed.
2. K. Kohse-Höinghaus and J.B. Jeffries. *Applied Combustion Diagnostics*. (Taylor & Francis, New York (NY), USA, 2002).
3. W. Merzkirch. *Flow visualisation*. (Academic Press, Orlando (NY), USA, 1987).
4. W. Demtröder. *Laser Spectroscopy – Basic Concepts and Instrumentation*. Vol. 5 of *Springer Series in Chemical Physics* (Springer Verlag, Berlin, D, 1996), 2nd enlarged ed.
5. E. Hecht. *Optics*. (Addison Wesley, San Francisco (CA), USA, 2002), 4th ed.
6. S. P. Nandula, T. M. Brown, P. A. Skaggs, R. W. Pitz, and P. A. DeBarber. *Multi-species line Raman measurements in H_2-air turbulent flames*. AIAA-paper 1994-0227 (1994).
7. M. Mansour and Y. Chen. *Line Raman, Rayleigh, and laser-induced predissociation fluorescence technique for combustion with a tunable KrF excimer laser.*

Appl. Opt. **35**(21), 4252–4260 (1996).

8. F. Rabenstein and A. Leipertz. *One-dimensional, time-resolved Raman measurements in a sooting flame made with* 355 *nm excitation*. Appl. Opt. **37**(21), 4937–4943 (1998).

9. G. Grünefeld, H. Schlüter, and P. Andresen. *Simultaneous multiple-line Rayleigh-/Raman/LIF measurements in combustion*. Appl. Phys. B **70**(3), 309–313 (2000).

10. J. O. Gilmore, S. Sharma, D. Fletcher, and D. Bershader. *Single-pulse spontaneous Raman scattering measurements in an expanding nitrogen/oxygen admixture*. AIAA-paper 1995-2125 (1995).

11. G. Tejeda, J. M. Fernández-Sánchez, and S. Montero. *High-performance dual Raman spectrometer*. Appl. Spectrosc. **51**(2), 265–276 (1997).

12. N.M. Sijtsema, R.A.L. Tolboom, N.J. Dam, and J.J. ter Meulen. *Two-dimenisonal multispecies imaging of a supersonic nozzle flow*. Opt. Lett. **24**, 664-666 (1999).

13. R.A.L. Tolboom, N.J. Dam, N.M. Sijtsema, and J.J. ter Meulen. *Quantitative spectrally resolved imaging through a spectrograph*. Opt. Lett. **28**, 2046-2048 (2003).

14. R. Tolboom. *Expanding Laser Diagnostics in non-seeded Compressible Flow Research*. Ph.D. thesis, University of Nijmegen, Nijmegen (NL) (2002), ISBN 90-9016193-7. Available from

    `http://webdoc.ubn.kun.nl/mono/t/tolboom_r/expaladii.pdf`.

15. R.A.L. Tolboom, N.J. Dam, and J.J. ter Meulen. *Quantitative imaging through a spectrograph – 2: stoichiometry mapping by Raman scattering*. Appl. Opt. (2004).

Companion paper: see next paper in this issue.

16. A. Thorne, U. Litzén, and S. Johansson. *Spectrophysics: principles and applications*. (Springer Verlag, Berlin, D, 1999.)

17. R.A.L. Tolboom, N.M. Sijtsema, N.J. Dam, and J.J. ter Meulen. *Raman imaging for combustion diagnostics*. AIAA paper 2000-0956 (2000).

18. Unless explicitly stated otherwise, the integration limits will be from $-\infty$ to $+\infty$ for (reciprocal) space coordinates and from 0 to $+\infty$ for wavelengths and frequencies.

19. Compare this to rolling two dice, λ and $x_{\rm in}$, the sum of their results being $x_{\rm out}$. There are various combinations of $(\lambda, x_{\rm in})$ leading to, for example, $x_{\rm out} = 7$. These combinations are described by $f(\lambda, x_{\rm in})$. On the other hand, once the 'outcome' $x_{\rm out}$ is chosen to be 7, there is only one corresponding λ for every $x_{\rm in}$, namely $\hat{f}(x_{\rm in};\, x_{\rm out})$.

20. The line at $\lambda = 579.40$ nm is an unresolved doublet. In fact, the exact wavelengths do not matter for the processing of OMAgraphs.

21. R.C. Gonzalez and R.E. Woods. *Digital image processing*. (Prentice Hall, Upper Saddle River (NJ), USA, 2002).

22. K.C. Castleman. *Digital image processing*. (Prentice Hall, Englewood Cliffs, NJ (USA), 1996).

23. See e.g. P.M. Lee. *Bayesian statistics: an introduction*. (Arnold, London, UK, 1997.) 2$^{\rm nd}$ ed.

24. H.C. Andrews and B.R. Hunt. *Digital image restoration*. (Prentice Hall, Engle-

wood Cliffs, NJ (USA), 1977).

25. As the input signal $S(x)$ of the ft is purely real, the Fourier transformed data $\widetilde{S}(k)$ is even under complex conjugation, *i.e.* $\widetilde{S}(-k) = \widetilde{S}^*(k)$. Therefore, only the positive k-components need to be plotted to represent all power information.

26. R. Durrett. *Probability: Theory and Examples* (Duxbury Press, Belmont, CA (USA), 1991).

27. D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms* (2002). (Draft 2.4.1)
`http://www.inference.phy.cam.ac.uk/itprnn/book.pdf`.

28. M. Plischke and B. Bergersen. *Equilibrium Statistical Physics*. (World Scientific Publishing, Singapore, 1994), 2nd ed.

29. H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. (Kluwer, Dordrecht, NL, 1996.)

30. This is a general property of Fourier transformation. For visual clarity, the reconstructions presented in this paper (Figs. 4 and 6) have been recentered in the image frames.

31. B. Buck and V.A. Macaulay. *Maximum Entropy in Action*. (Clarendon Press, Oxford, UK, 1991.)

32. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C – The Art of Scientific Computation*. (Cambridge University Press, Cambridge (MA), USA, 1992), 2nd ed.

33. In this paper, the convention is that the *power* or *absolute square* of a signal S

is defined as the signal times its complex conjugate (*i.e.* $|S|^2 = S \times S^*$), and the *norm* of the signal $\|S\|$ is the integral over S.

34. See e.g. ref. 1, section 3.6.

List of Figures

4 Strip-by-strip deconvolution of Fig. 3(d) with the spectrum of Fig. 3(b) via the unfiltered Fourier transformation algorithm, Eq. (A6). The purely spatial images (a) and (c) show data in direct space, the corresponding power spectra are shown in (b) and (d), respectively (first half of the k-components only). (a,b): direct deconvolution; note the large high k-components in (b). (c,d): ditto, but with two-pixel binning; the high k-components partly cancel (d). On top: single-strip cross sections at the position of the arrows. (Zero baseline indicated in (a,c); left ordinates omitted in (b,d) to emphasize the low k-components.)

5 Power spectra (single strips at the location of the arrows in Fig. 3). (a) Fourier transformed spectral reference function $\widetilde{R}_k$ of Fig. 3(b); (b) non-filtered deconvolution function; (c) filtering of (b) (note logarithmic scale); and (d) linear Bayesian filtered deconvolution function. The power of the filtering function (c) is the pre-factor 'filter' of Eq. (34). (c) and (d) are calculated for $\sigma/\tau = 6$ count^{-1}.

6 Strip-by-strip linear Bayesian deconvolution of Fig. 3(d) with the spectrum of Fig. 3(b). From right to left are the results for three ratios σ/τ, ranging from (a) 'too low' ($\sigma/\tau = 1$ count^{-1}) via (b) 'best' ($\sigma/\tau = 6$ count^{-1}) to (c) 'too high' ($\sigma/\tau = 800$ count^{-1}). Above the images are the single-strip cross sections (similar to those in Figs. 3 and 4), and on top are their power spectra for the first half of k-components. The images are scaled individually, but the graphs are all on the same linear grey scale. The (left) ordinates of the power spectra are omitted to show the similarity of the hardly filtered, low k-components.

`Tolboom_ApplOpt2_F6.eps` 46

7 Contrast (defined in the text) as a function of σ/τ. The maximum in the curve is taken as the best σ/τ for the deconvolution as it minimises the *relative* power in the physically dark region. The corresponding value for the non-filtered results (Fig. 4(a)) is: $\langle P(\text{in})\rangle/\langle P(\text{out})\rangle = 1.35$.

`Tolboom_ApplOpt2_F7.eps` 47

8 The re-convolved image of the data that were obtained with a deconvolution for $\sigma/\tau = 6$ count^{-1} (solid curve) and its difference with the original, measured data (residual; grey curve). The difference is 0.04 count on average and has a standard deviation of over 7 kcount.

`Tolboom_ApplOpt2_F8.eps` 47

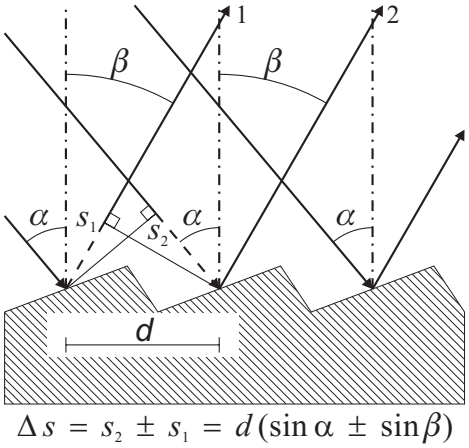

Figure 1. Schematic representation of the diffraction of light by a grating.

`Tolboom_ApplOpt2_F1.eps`

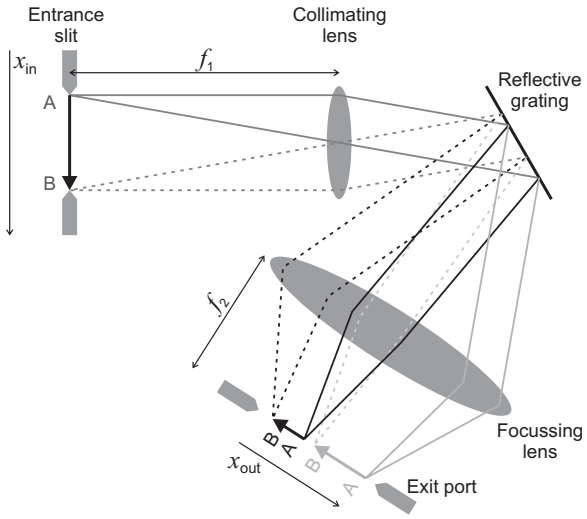

Figure 2. Cross section of a spectrograph including a first order ray trace for bichromatic light. The heights of entrance slit and exit port as well as the grooves of the grating are perpendicular to the plane of the picture. Also indicated in the figure are the coordinates x_{in} and x_{out} that appear in the formalism of section 2.

`Tolboom_ApplOpt2_F2.eps`

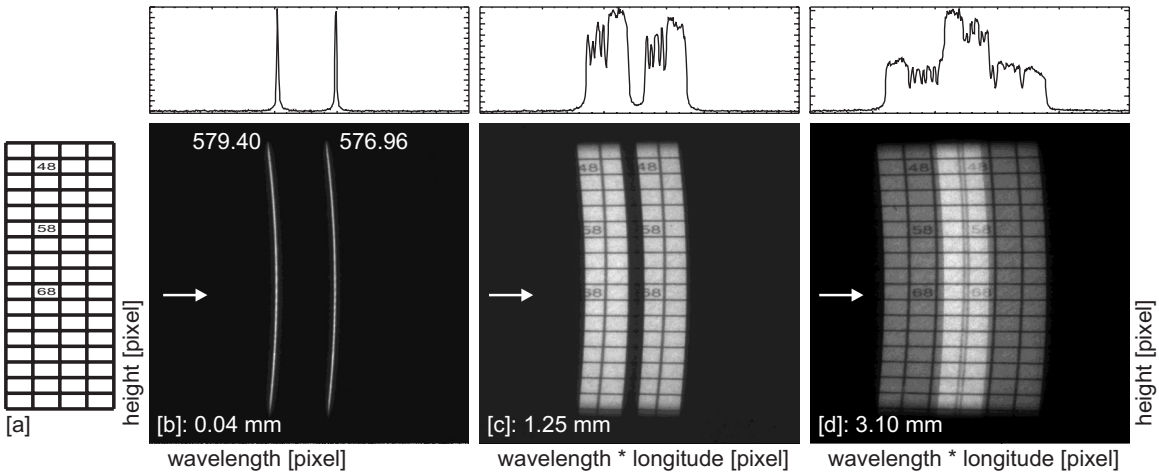

Figure 3. Three OMAgraphs recorded with different widths of the entrance slit of the spectrograph, as indicated in the images. The horizontal axes contain both spectral and spatial information, and the vertical axes are purely spatial. (a) Imaged object, a 5.0×5.0 mm^2 grid printed on white paper with a line width of 0.5 mm; (b) spectrum of the light source (a Hg(Ar) calibration lamp; λ (nm) indicated), recorded by reflection off white paper; (c) and (d) OMAgraphs of the object shown in (a) under illumination with the same source as in (b). The graphs on top of the images are single-strip cross sections of the images at the position of the arrows, cutting the lower circles of the '68' on the grid. All images are scaled individually. The curvature of the images and the horizontal extrusion of the grid are artifacts of the spectrograph.

`Tolboom_ApplOpt2_F3.eps`

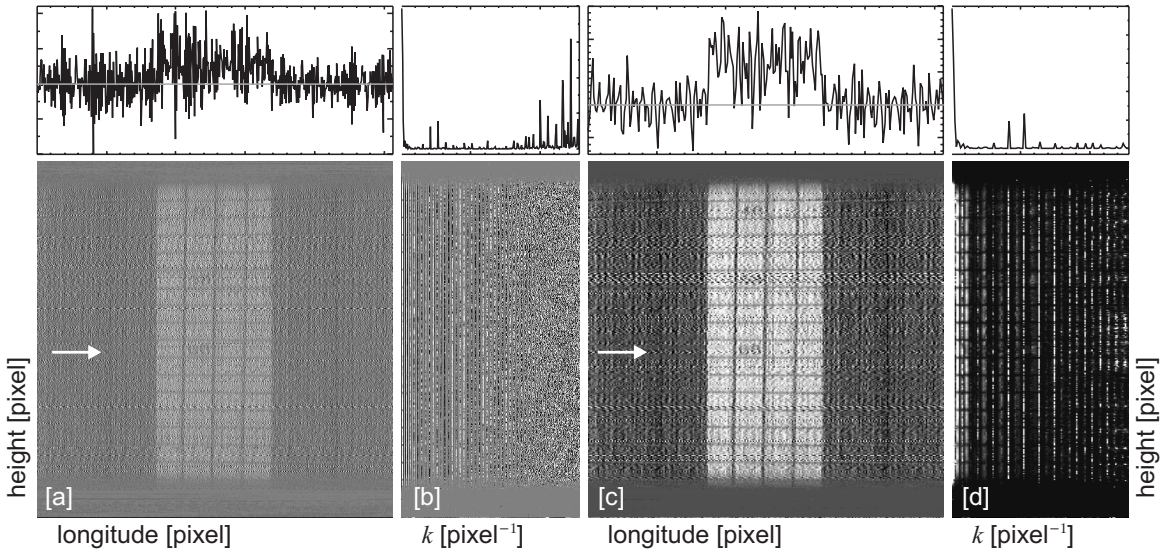

Figure 4. Strip-by-strip deconvolution of Fig. 3(d) with the spectrum of Fig. 3(b) via the unfiltered Fourier transformation algorithm, Eq. (A6). The purely spatial images (a) and (c) show data in direct space, the corresponding power spectra are shown in (b) and (d), respectively (first half of the k-components only). (a,b): direct deconvolution; note the large high k-components in (b). (c,d): ditto, but with two-pixel binning; the high k-components partly cancel (d). On top: single-strip cross sections at the position of the arrows. (Zero baseline indicated in (a,c); left ordinates omitted in (b,d) to emphasize the low k-components.)

`Tolboom_ApplOpt2_F4.eps`

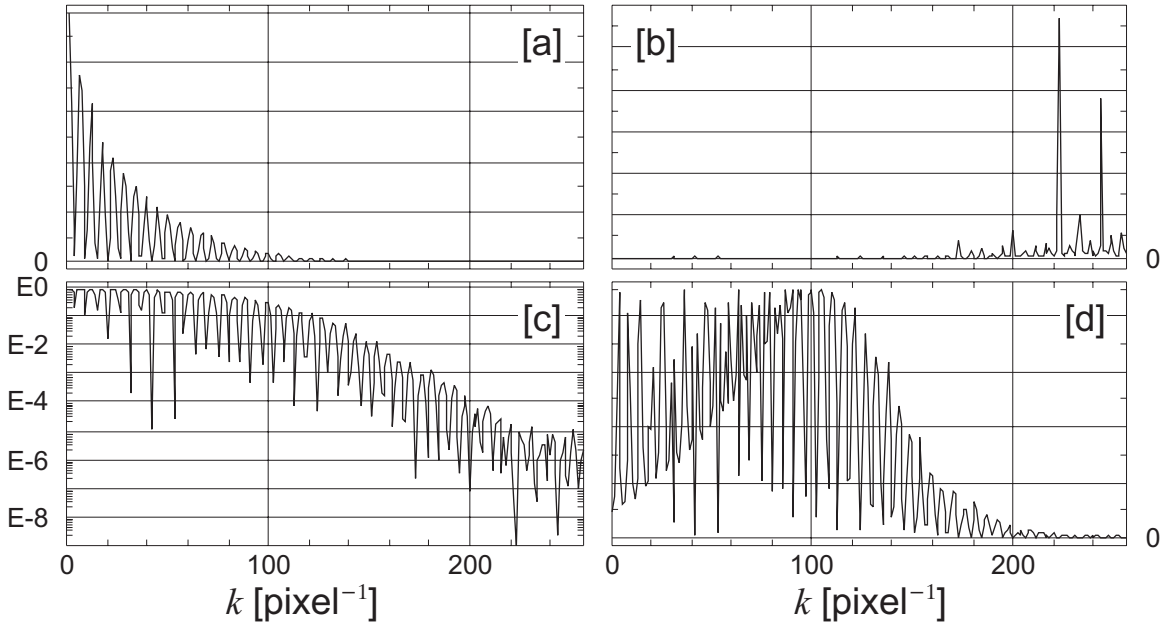

Figure 5. Power spectra (single strips at the location of the arrows in Fig. 3). (a) Fourier transformed spectral reference function $\widetilde{R}_k$ of Fig. 3(b); (b) non-filtered deconvolution function; (c) filtering of (b) (note logarithmic scale); and (d) linear Bayesian filtered deconvolution function. The power of the filtering function (c) is the pre-factor 'filter' of Eq. (34). (c) and (d) are calculated for $\sigma/\tau = 6$ count^{-1}.

`Tolboom_ApplOpt2_F5.eps`

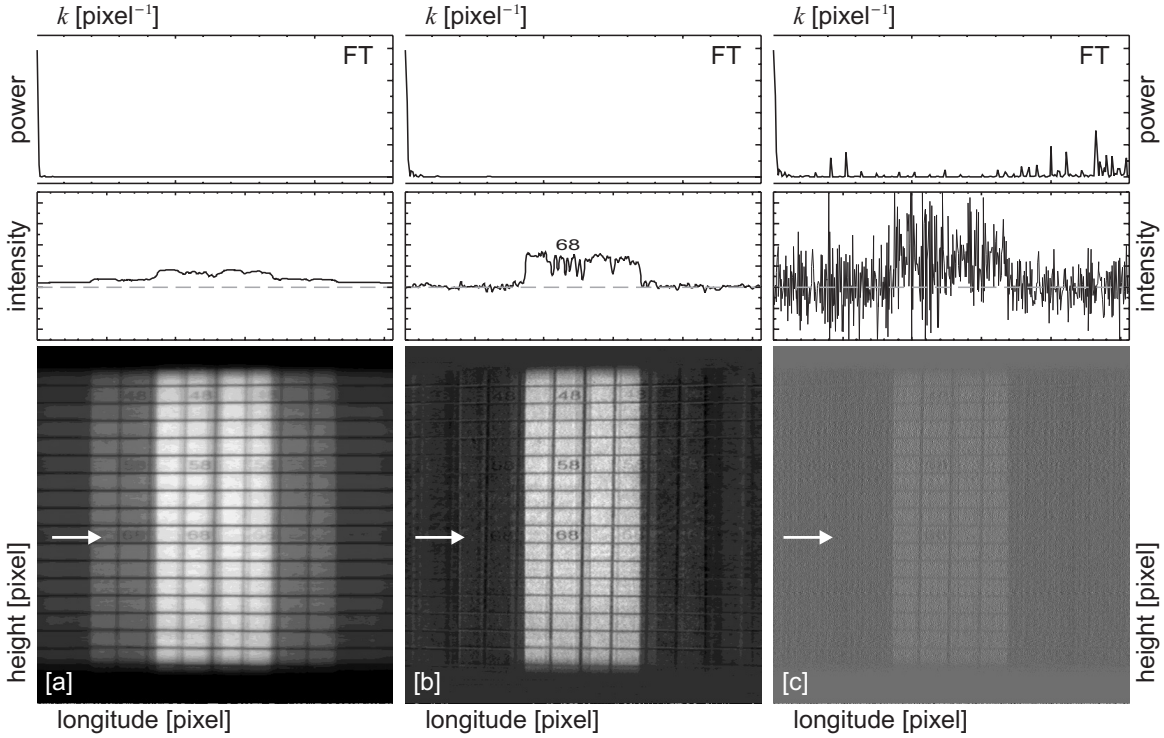

Figure 6. Strip-by-strip linear Bayesian deconvolution of Fig. 3(d) with the spectrum of Fig. 3(b). From right to left are the results for three ratios σ/τ, ranging from (a) 'too low' ($\sigma/\tau = 1$ count^{-1}) via (b) 'best' ($\sigma/\tau = 6$ count^{-1}) to (c) 'too high' ($\sigma/\tau = 800$ count^{-1}). Above the images are the single-strip cross sections (similar to those in Figs. 3 and 4), and on top are their power spectra for the first half of k-components. The images are scaled individually, but the graphs are all on the same linear grey scale. The (left) ordinates of the power spectra are omitted to show the similarity of the hardly filtered, low k-components.

`Tolboom_ApplOpt2_F6.eps`

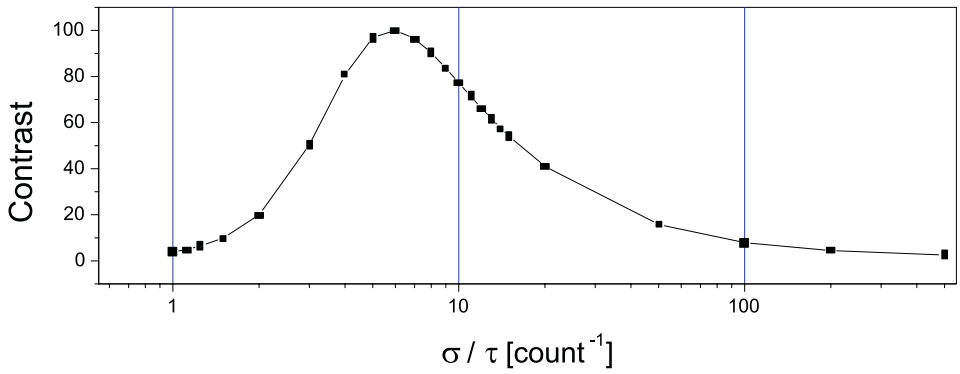

Figure 7. Contrast (defined in the text) as a function of σ/τ. The maximum in the curve is taken as the best σ/τ for the deconvolution as it minimises the *relative* power in the physically dark region. The corresponding value for the non-filtered results (Fig. 4(a)) is: $\langle P(\text{in})\rangle/\langle P(\text{out})\rangle = 1.35$.

`Tolboom_ApplOpt2_F7.eps`

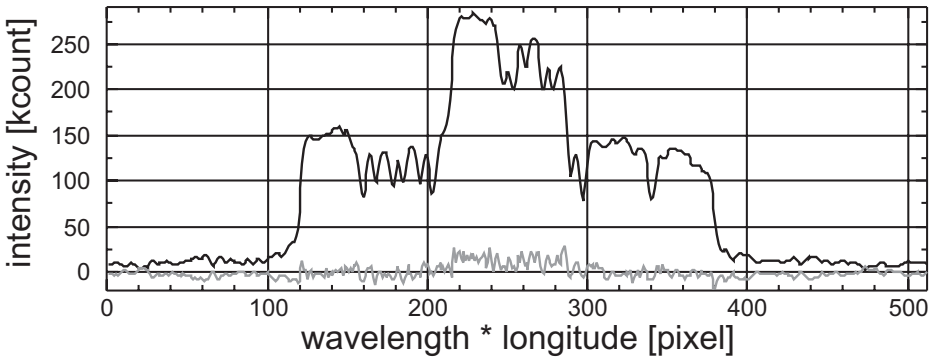

Figure 8. The re-convolved image of the data that were obtained with a deconvolution for $\sigma/\tau = 6$ count^{-1} (solid curve) and its difference with the original, measured data (residual; grey curve). The difference is 0.04 count on average and has a standard deviation of over 7 kcount.

`Tolboom_ApplOpt2_F8.eps`