A MAP-based algorithm for spectroscopic semi-blind deconvolution

Article *in* The Analyst · July 2012

Impact Factor: 4.11 · DOI: 10.1039/c2an16213j · Source: PubMed

5 authors, including:

Houzhang Fang
Xidian University
22 PUBLICATIONS **142** CITATIONS
SEE PROFILE

yi Chang
Huazhong University of Science and Techn…
9 PUBLICATIONS **63** CITATIONS
SEE PROFILE

Analyst

Cite this: *Analyst*, 2012, **137**, 3862

www.rsc.org/analyst

PAPER

A MAP-based algorithm for spectroscopic semi-blind deconvolution

Hai Liu, Tianxu Zhang, Luxin Yan,* Houzhang Fang and Yi Chang

Received 6th December 2011, Accepted 18th May 2012
DOI: 10.1039/c2an16213j

Spectroscopic data often suffer from common problems of bands overlapping and random noise. In this paper, we show that the issue of overlapping peaks can be considered as a *maximum a posterior* (MAP) problem and be solved by minimizing an object functional that includes a likelihood term and two *prior* terms. In the MAP framework, the likelihood probability density function (PDF) is constructed based on a spectral observation model, a robust Huber–Markov model is used as spectra *prior* PDF, and the kernel *prior* is described based on a parametric Gaussian function. Moreover, we describe an efficient optimization scheme that alternates between latent spectrum recovery and blur kernel estimation until convergence. The major novelty of the proposed algorithm is that it can estimate the kernel slit width and latent spectrum simultaneously. Comparative results with other deconvolution methods suggest that the proposed method can recover spectral structural details as well as suppress noise effectively.

1 Introduction

Spectroscopic deconvolution and slit width estimation are both classical problems, which are known to be difficult and have attracted major research efforts. The analysis of the precise bandwidth of spectroscopic data is crucial to help understand various inter/intra-molecular processes. However, the spectra recorded by a dispersion spectrophotometer are usually distorted by the response function of the instrument. These distortions generally include the overlapping of adjacent peaks and the aberrancy of the relative intensities among the peaks.

The analysis of overlapped bands in IR spectrum is often assisted by band-narrowing methods,[1–3] curve-fitting procedures,[4] or a combination of both.[5] The former provides more resolved spectra by narrowing the component bands, while the latter potentially allows the estimation of their band parameter, such as positions, widths and area.

Deconvolution has become one of the most useful methods for resolving these problems. For example, Fourier self-deconvolution (FSD) method, developed by Kauppinen *et al.* is the most common method used in infrared spectroscopy.[3] However, deconvolution is an ill-posed inverse problem. Because it needs to recover a high-quality spectrum and instrument response function simultaneously from a degraded spectrum. To settle the problem, *a prior* knowledge is introduced to constrain the solution space of the latent spectrum and instrument response function. In summary, there are two major kinds of *prior*: latent spectral *prior* and kernel *prior*.

The latent spectral *prior*, may include three kinds of constraints such as positivity, smoothness and the same characteristics between the latent and degraded spectra. It is usually assumed that the spectral intensity is positivity. The smoothness constraints include Tikhonov regularization[6] and its variants.[6] In recent years, a high-order statistical method[7] was introduced to estimate latent spectrum and instrument response function simultaneously. It is assumed that the second-order central moment of the degraded and latent spectra are the same. The latent spectrum is obtained by maximizing the high-order even moments of the degraded spectrum. It seems that this approach narrows spectra but enlarges the noise, especially under low signal-to-noise ratio (SNR). Maximum entropy deconvolution[8,9] (MaxEntD) method considers the latent spectrum and the degraded spectrum as different distributions. Then, a unique solution (latent spectrum) is chosen by maximizing the Shannon-entropy[9] or minimizing the cross-entropy[10] (or Kullback–Leibler distance) between the two distributions. While MaxEntD uses the entropic *prior*, massive inference uses a so-called "atomic" *prior*.[8]

As for the kernel *prior*, two aspects of the *a prior* are often used, kernel shape and slit width. First, the kernel shape is usually selected on theoretical or practical grounds, and basically restricted to triangle,[11] Lorentzian,[12] Gaussian and Voigt[13] function. More recently, Lórenz-Fonfría adequately discussed eight different filters, and the conclusion was that the filters, most suited to the FSD method, were the Bessel, BL3 and Gaussian filters.[14] Second, for the kernel slit width, Senga estimated the bandwidth of the degraded spectrum by a complex process.[11] Lórenz-Fonfría discussed highly overlapped bands and

Science and Technology on Multi-spectral Information Processing Laboratory, Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China. E-mail: yanluxin@gmail.com; Fax: +86-027-87543594; Tel: +86-027-87540139

estimated the mean Lorentzian bandwidth of the component bands.[15] For simplicity's sake, slit width usually assumes that the bandwidth of the measured spectra is equal to the width of the narrowest spectral band. It is considered that all the peaks can be recovered if the narrowest peak has been recovered. However, the measured width is often inaccurate, especially when the spectral is seriously degraded.

The major novelty of the proposed algorithm is that it can estimate the kernel slit width and latent spectrum simultaneously. To accomplish these results, our technique benefits from three main contributions. First, our probabilistic model unifies blind and non-blind deconvolution into a single MAP formulation. Second, we introduce Huber–Markov *a prior* knowledge to preserve sharp peaks and suppress noise using the first-order derivative of actual spectrum. Third, to estimate the actual width of the instrumental response, we employ a parametric Gaussian form as kernel shape *prior*. To our knowledge, few publications include the both spectrum and kernel *prior* (or regularization) constraint for spectral deconvolution. In this paper, a MAP-based model is proposed for spectroscopic deconvolution. This algorithm can effectively take advantage of the *a prior* knowledge to estimate the kernel width and latent spectrum simultaneously.

The remainder of this work is organized as follows. In Section 2, the spectrum observation model and spectral characteristics are formulated. The MAP-based recovery model is presented in Section 3. In Section 4, the optimization method and parameter determination are described. Experimental results are provided in Section 5, and Section 6 concludes this paper.

2 Analysis of spectral characteristics

For both coherent and incoherent irradiation, the spectroscopic data is measured by a spectrophotometer. The degraded spectrum can usually be mathematically modeled as a convolution of the actual spectrum with the instrument response function, as well as a random process superposed on the useful signal which describes random errors arising during measurement of the spectrum.[16,17] The relationship between the measured spectrum and actual spectrum can be expressed as

$$g(v) = f(v) \otimes h(v) + n(v), \tag{1}$$

where $g(v)$ is the measured spectroscopic data, and $\otimes$ denotes the convolution operation $f(v) \otimes h(v) = \sum_k h_k f(v-k)$, $f(v)$ stands for actual spectrum and $h(v)$ stands for the point spread function (PSF, also called blur kernel), which collects the intrinsic line-shape function and the instrumental broadening. It usually supports that the length of $h(v)$ is much smaller than $f(v)$.

We classify the IR spectral data into steep and plain regions, according to the first-order derivative of the actual spectrum. The former contains abundant absorption bands, and the latter contains few. Bold lines represent the steep regions, and thin lines represent the plain regions, which are shown in Fig. 1A. Our intriguing finding is that given the same noise over the whole spectrum, the plain and steep regions degrade differently. In other words, the plain and steep regions have different sensitivities to the same noise. For steep regions, the variation of absorption intensity is large, and the noise is not obvious. This can be seen in Fig. 1B. As for the plain regions, whether the absorption intensity is strong or not, the noise is significant. We suggest that the steep regions of the degraded spectrum are more similar to the original spectrum than those of the plain regions. On one hand, if the strong smoothness constraint is imposed on the whole spectrum, the noise will be suppressed but the steep regions will be smoothed. On the other hand, if the mild constraint is incorporated, the steep regions are preserved but much residual noise is left. Based on this idea, we take advantage of the structure information of the spectrum and propose a new adaptive constraint to preserve the details in the steep regions while suppressing the noise in the plain regions.

Furthermore, our study shows that the convolution kernel can be described as a convolution of many factors, such as slit function, grating response, and circuit response, *etc.* And the convolution kernel approaches to a Gaussian or Lorentzian-like function. KatraSnik proposed that the PSF of the degraded acousto-optical spectrum can be modeled as a parametric function three side-lobe sinc2.[18] Inspired by this, an attempt was made to use a parametric function to model the PSF. We demonstrate the gains obtained by reducing the blind spectrum deconvolution problem to a semi-blind deconvolution (SBD), in which the kernel is assumed to belong to a class of parametric functions.

3 Our model

Our probabilistic model unifies blind and non-blind deconvolutions into a single MAP formulation. In recent years, the MAP estimation method, which inherently includes *a priori* constraints in the form of *prior* probability density functions (PDFs), has enjoyed increasing popularity. It has been central to the solution of ill-posed inverse problems in a wide range of applications,[8,10,19] such as spectrum denoising, quantitative investigation, and others. We use the MAP framework for spectral deconvolution of an IR spectrum. The purpose is to realize the MAP estimate of a desired spectrum $f(v)$ and the PSF $h(v)$, given the degraded spectrum g. It can be computed by

$$(\hat{f},\hat{h}) = \arg\max p(f,h|g) \tag{2}$$

Applying Bayes's rule, eqn (2) becomes

$$(\hat{f},\hat{h}) = \arg\max \frac{p(g|f,h)p(f)p(h)}{p(g)}. \tag{3}$$

Since $p(f,h|g)$ is independent of g, $p(g)$ can be considered as a constant, and hence, eqn (3) can be rewritten as

$$(\hat{f},\hat{h}) = \arg\max p(g|f,h)p(f)p(h). \tag{4}$$

Using the monotonic logarithm function, eqn (4) can be expressed as

$$(\hat{f},\hat{h}) = \arg\max\{\log p(g|f,h) + \log p(f) + \log p(h)\}. \tag{5}$$

It can be seen that three PDFs need to be constructed. It is very important that what and how much *prior* is introduced to the kernel and spectrum. We now define these terms, and describe our optimization.

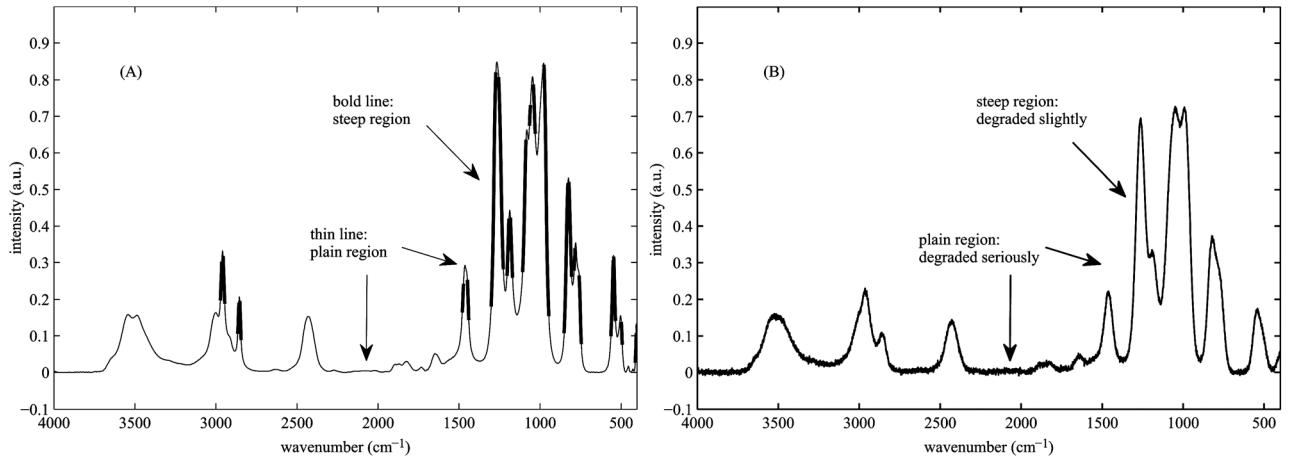

Fig. 1 (A) Original IR spectrum. Bold line represents steep regions; thin line represents plain regions. (B) Degraded (overlapped and noisy) spectrum.

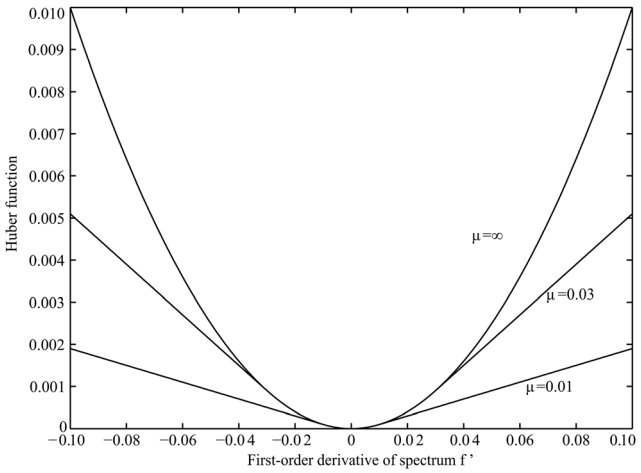

Fig. 2 Huber function with different μ's.

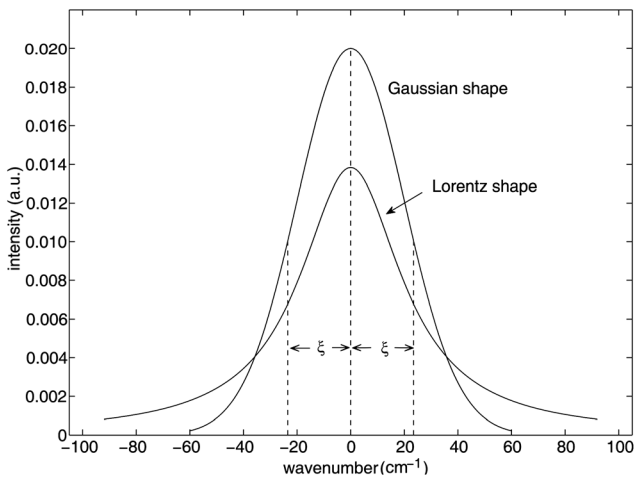

Fig. 3 Different kernels with the same ξ.

3.1 Definition of the probability terms

Likelihood *p*(*g*|*f*,*h*). The first is the likelihood density function, which provides a measure of the conformance of the spectrum to the observed spectrum according to the spectrum observation model. It is determined by the probability density of the noise vector in eqn (1), *i.e.*, $p(g|f,h) = p(n)$. Spectrum noise is modeled as a set of independent and identically distributed (IID) noise random variables for all spectral points, each of which follows a Gaussian distribution. Under these assumptions, the probability density is given by

$$p(n) = \prod_i \mathcal{N}(n_i|0,\sigma_i) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(n_i-0)^2}{2\sigma_i^2}\right)$$
$$= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^N}\exp\left(-\sum_i^N \frac{(f_i\otimes h_i - g_i)^2}{2\sigma^2}\right)$$
$$= \frac{1}{M_1}\exp\left(-\frac{1}{2\sigma^2}\|f\otimes h - g\|^2\right).$$

It can be rewritten as

$$p(g|f,h) \propto \frac{1}{M_1}\exp\left(-\frac{1}{2\sigma^2}\|f\otimes h - g\|^2\right). \quad (6)$$

In this expression, $\|\cdot\|$ denotes the 2-norm operator and M_1 is the constant coefficient.

Prior *p*(*f*). The second density function in eqn (5) is the spectrum *prior*, which imposes the spatial constraints on the spectrum. According to the analysis of the spectral, we find that the steep and plain regions obey different local distributions. Some conventional models such as Laplacian *prior* constrain and Gauss–Markov *prior* regularize the corresponding ill-posed problem by forcing spatial smoothness on the spectrum. For example, the general form of the Markov *prior* is denoted by

$$p(f) = \frac{1}{M_2}\exp\left(-\alpha\sum_i \rho(f')\right). \quad (7)$$

In this expression, M_2 is the constant. The quantity f' is a spatial activity measure to intensity f_i, which is often formed by first or second-order differences.[20] In this paper, $f' = (f_{i+1} - f_i)/2$. When $\rho(\cdot)$ is a quadratic potential function as in eqn (7), the corresponding *prior* is regarded as Gauss–Markov

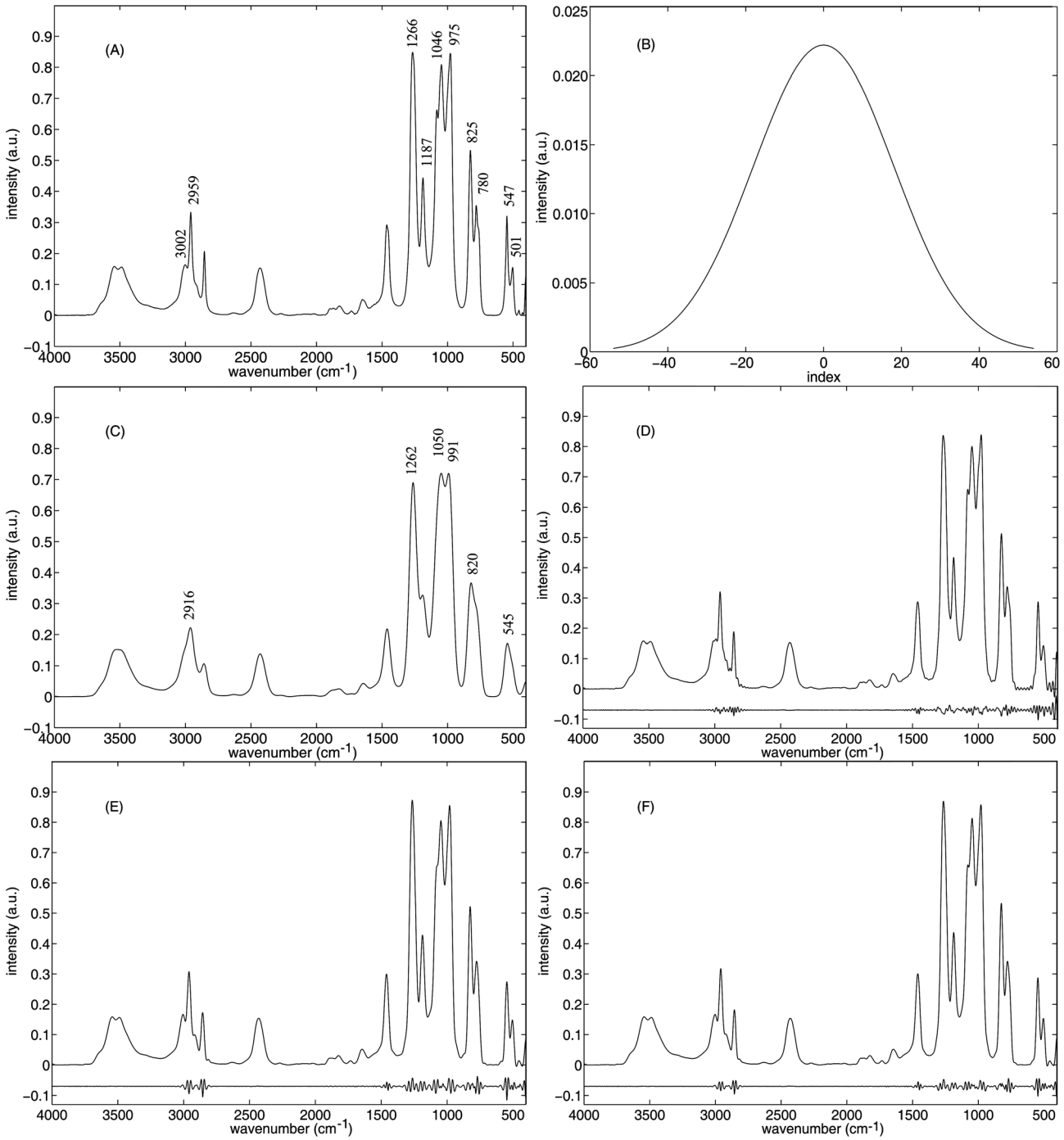

Fig. 4 Deconvolution with noise-free. (A) Original IR spectrum of dimethyl phosphite ($C_2H_7O_3P$) from 4000–400 cm^{-1}. (B) Instrument response function with Gaussian shape and $\sigma = 18$ cm^{-1}. (C) Degraded spectrum convoluted between (A) and (B). (D) Deconvolution by FSD method with a predefined Gaussian kernel $\sigma = 18$ cm^{-1}. (E) Deconvolution using the GMSBD proposed by us. (F) Deconvolution using the suggested HMSBD without predefined kernel. Bottom of the spectrum is the residual noise which equals to original spectrum subtracts the deconvoluted result.

$$\rho(t) = t^2. \tag{8}$$

The criticism of the use of the Gauss–Markov *prior* (and the Laplacian *prior*) is that some high-frequency objects in the spectra tend to be removed. Therefore, a detail-preserving Huber–Markov spectral *prior* is employed in the paper. This *prior* can effectively preserve the detail information in the spectra.[21] The difference between the Huber–Markov *prior* and the Gauss–Markov *prior* is only the potential function $\rho(\cdot)$. The Huber function is defined as

$$\rho(t) = \begin{cases} t^2 & |t| \le \mu \\ 2\mu|t| - \mu^2 & |t| > \mu \end{cases}, \tag{9}$$

where μ is a threshold parameter separating the quadratic and linear regions,[21] as shown in Fig. 2. It is easy to see that the

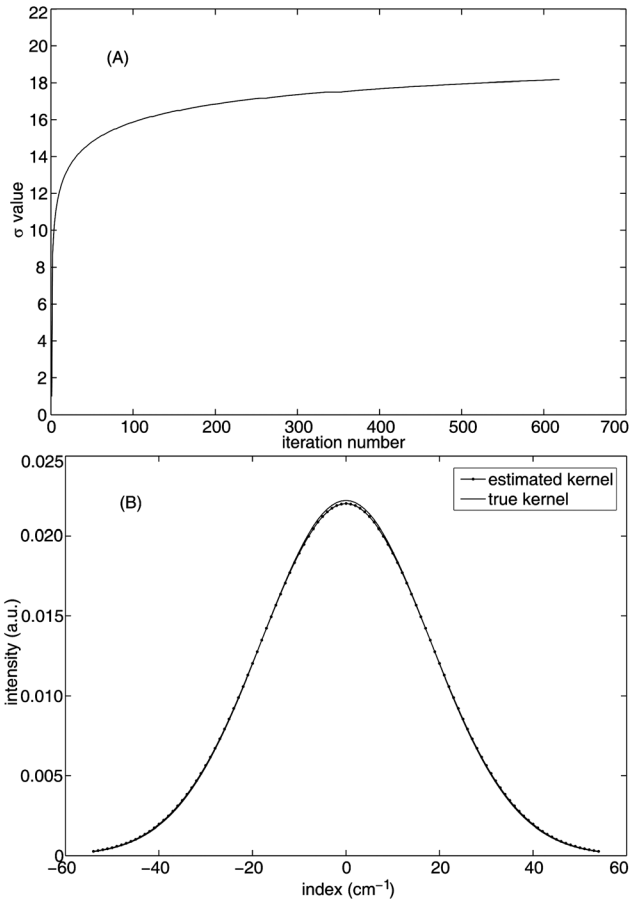

Fig. 5 (A) Convergence curve of the estimated width σ of the blur kernel as a function of the iteration number in the HMSBD of the *noise-free* spectrum (Fig. 4C). (B) Estimated kernel and true kernel.

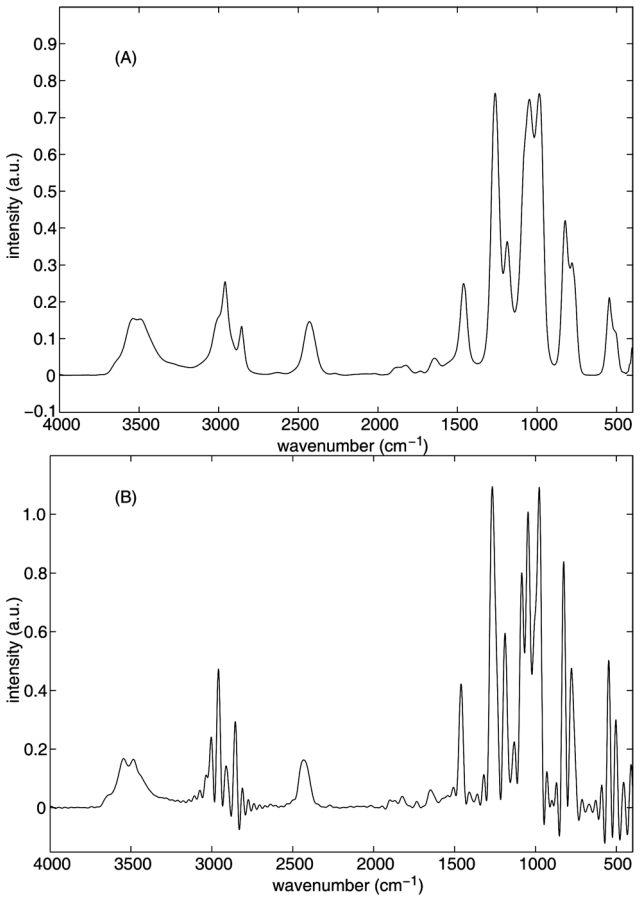

Fig. 6 Deconvolution results by FSD[14] method with incorrect kernels. (A) Infra-deconvolution with Gaussian kernel with $\sigma = 10$ cm^{-1}. (B) Over-deconvolution with Gaussian kernel with $\sigma = 26$ cm^{-1}.

Gauss–Markov *prior* can be regarded as a special case of the Huber–Markov *prior* when μ approaches $+\infty$.

Prior *p*(*h*). The third density function in eqn (5) is the recovery kernel *prior*. To get the kernel PDF, we discuss the kernel prior from two aspects. On one hand, the PSF can be modeled with a parametric function according to *a prior* knowledge of the kernel shape. Symbol h_ξ stands for the kernel shape function by the bandwidth ξ. In case of Lorentzian function kernel, ξ stands for the half width at half maximum γ,

$$h_\gamma(v) = \frac{1}{\pi}\frac{\gamma}{(v^2+\gamma^2)}, \tag{10}$$

and for Gaussian function, $\xi = \sqrt{2\ln 2}\sigma$ also can be referred to the width of the kernel, shown in Fig. 3,

$$h_\sigma(v) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{v^2}{2\sigma^2}\right). \tag{11}$$

Other kinds of kernel shape can be similarly expressed, such as triangle, Bessel, Vogit and so on. Without any *a prior* knowledge about the kernel, the kernel recovered from measurement spectrum often approximates a Gaussian-like line shape,[7] called blind deconvolution (BD). In this paper, we just choose the Gaussian line shape function to elaborate the semi-blind deconvolution (SBD).

On the other hand, the kernel width of PSF is represented by the kernel smoothness *prior*: the width of the Gaussian corresponds to its smoothness. The larger the width of kernel is, the smoother the kernel is. The kernel smoothness is measured by its first-order derivative. Senga choose the triangle function[11] as kernel function, and Laplacian-*prior* should be used. As for the Lorentzian and Gauss kernel (shown in Fig. 3), Gauss–Markov *prior* was chosen to constrain the smoothness of the kernel.

$$p(h) \propto \frac{1}{M_3}\exp\left(-\beta\sum_i |h'_\sigma|^2\right). \tag{12}$$

In this expression, $h' = (h_{i+1} - h_i)/2$.

Substituting eqn (6), (7) and (12) in eqn (5), our MAP problem is transformed to an object minimization problem that minimizes the negative logarithm of the probability we have defined, *i.e.*, the object $E(f,h) = -\sigma^2\log(p(f,h|g))$. After some manipulation, M_1, M_2 and M_3 can be safely dropped, we get the objective functional:

$$E(f,h) = \tfrac{1}{2}\|f \otimes h_\sigma - g\|^2 + \alpha\sum\rho(f') + \beta\sum|h'_\sigma|^2, \tag{13}$$

in which the first term is the data fidelity term, $\sum\rho(f')$ and $\sum|h'_\sigma|^2$ acts as the regularization terms. These are balanced by α and β, which can now be called the regularization parameters. Our configuration of α and β will be described in next section.

Table 1 Band distortions in deconvoluted spectra by FSD, GMSBD and HMSBD in Fig. 4

Band position[a]		3002	2959	1266	1187	1046	975	825	780	547	501	RMSE[b]
Position	FSD	−9	0	+1	−1	+2	+4	0	+1	+1	+6	3.755
	GMSBD	+3	0	+2	0	−1	+6	0	−2	−1	+4	2.664
	HMSBD	0	0	−1	0	1	+6	0	−2	−1	+4	2.429

[a] In cm^{-1}, obtained from the band maximum. [b] Root mean square error in the determination of the position.

According to the selection of potential function, we call the proposed methods Huber–Markov semi-blind deconvolution (HMSBD), Gauss–Markov semi-blind deconvolution (GMSBD), respectively.

4 Optimization

An efficient alternative approach, based on the lagged diffusivity fixed pointed scheme and gradient descent, was presented by You and Kaveh,[22] followed by Chan and Wong.[23] We employ the algorithm from image restoration to the spectral deconvolution. The gradient descent optimization method is used for the minimum problem in eqn (13). We solve for f and h by applying alternating minimization. The first-order necessary conditions associated with the alternating minimization of the object function can be expressed as follows.

4.1 Optimizing *f*

In this step, we fix h and optimize f. The functional $E(f,h)$ can be simplified to $E(f)$ by removing constant-value terms:

$$E(f) = \tfrac{1}{2}\|f \otimes h_\sigma - g\|^2 + \alpha\sum\rho(f'), \tag{14}$$

which is convex on IR and the solution, when it exists, is unique and can be obtained by gradient descent method. Minimization of the objective functional (14) with respect to f is carried out using the Euler–Lagrange (E–L) eqn (15), with Neumann boundary conditions. The E–L equation is a linear partial differential equation, the derivation is presented as follows:

$$\frac{\delta E}{\delta f} = \sum_{i\in L}\left(\frac{\partial F}{\partial f} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial f'}\right)\right). \tag{15}$$

The corresponding gradient element of the *prior* term is given by

$$\frac{\partial F}{\partial f} = (f \otimes h_\sigma - g)\otimes h_\sigma(-v)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial f'}\right) = \begin{cases} 2f'' & |f'| \le \mu \\ 0 & |f'| > \mu \end{cases}.$$

Therefore, eqn (15) is rewritten as

$$\frac{\delta E}{\delta f} = \begin{cases} (f \otimes h_\sigma - g)\otimes h_\sigma(-v) - 2\alpha f'' & |f'| \le \mu \\ (f \otimes h_\sigma - g)\otimes h_\sigma(-v) & |f'| > \mu \end{cases}. \tag{16}$$

For the Gauss–Markov *prior*, its E–L equation is eqn (17) (top),

$$\frac{\delta E}{\delta f} = (f \otimes h_\sigma - g)\otimes h_\sigma(-v) - 2\alpha f''. \tag{17}$$

Then, the latent spectrum is solved by employing a successive approximations iteration

$$\hat{f}_{n+1} = \hat{f}_n + t_n\left(-\frac{\delta E}{\delta f_n}\right), \tag{18}$$

where n is the iteration number, and t_n is the time step. If t_n is too small, the convergence will be very slow. On the contrary, if it is too large, the algorithm will be unstable or divergent. By making a second-order Taylor series approximation to the object function at the current state $\hat{f}_n$, a quadratic step size approximation becomes[24]

$$t_n = \frac{(\nabla E(f_n))^T \nabla E(f_n)}{(\nabla E(f_n))(\nabla^2 E(f_n))\nabla E(f_n)}, \tag{19}$$

where $\nabla E(f_n)$ and $\nabla^2 E(f_n)$ are the first-order and Hessian matrix of the objective functional $E(f)$, respectively.

4.2 Optimizing *h*

In this step, we fix f and compute the optimal h. Eqn (13) is simplified to

$$E(h) = \tfrac{1}{2}\|h_\sigma \otimes f - g\|^2 + \beta\sum|h'_\sigma|^2. \tag{20}$$

Minimization with respect to the scalar parameter σ is determined by differentiation of the objective functional eqn (20)

$$\frac{\partial E}{\partial \sigma} = \sum_{i\in L}(h_\sigma \otimes f - g)\left(\frac{\partial h_\sigma}{\partial \sigma}\otimes f\right)_i + \beta\sum_{i\in L}\left(\frac{\partial}{\partial \sigma}|h'_\sigma|^2\right)_i = 0, \tag{21}$$

where in eqn (21)

$$\frac{\partial h_\sigma}{\partial \sigma} = \frac{1}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{v^2}{2\sigma^2}\right)\cdot\left(\frac{v^2}{\sigma^2} - 1\right), \tag{22}$$

and

$$h'_\sigma = \frac{\partial h_\sigma}{\partial v} = -\frac{v}{\sqrt{2\pi}\sigma^3}\exp\left(-\frac{v^2}{2\sigma^2}\right), \tag{23}$$

$$\frac{\partial}{\partial \sigma}|h'_\sigma|^2 = \frac{v^2}{\pi\sigma^7}\exp\left(-\frac{v^2}{\sigma^2}\right)\left(\frac{v^2}{\sigma^2} - 3\right). \tag{24}$$

Eqn (21) is solved for σ using the bisection method. The discrete support of the Gaussian kernel is limited to $6\sigma + 1$, which in our experiments is much smaller than the spectrum size. The numerical integral of h_σ is normalized to 1, and the σ_0 width is initialized to 1.

We declare convergence when for more than two consecutive iterations both the kernel width and latent spectrum change less

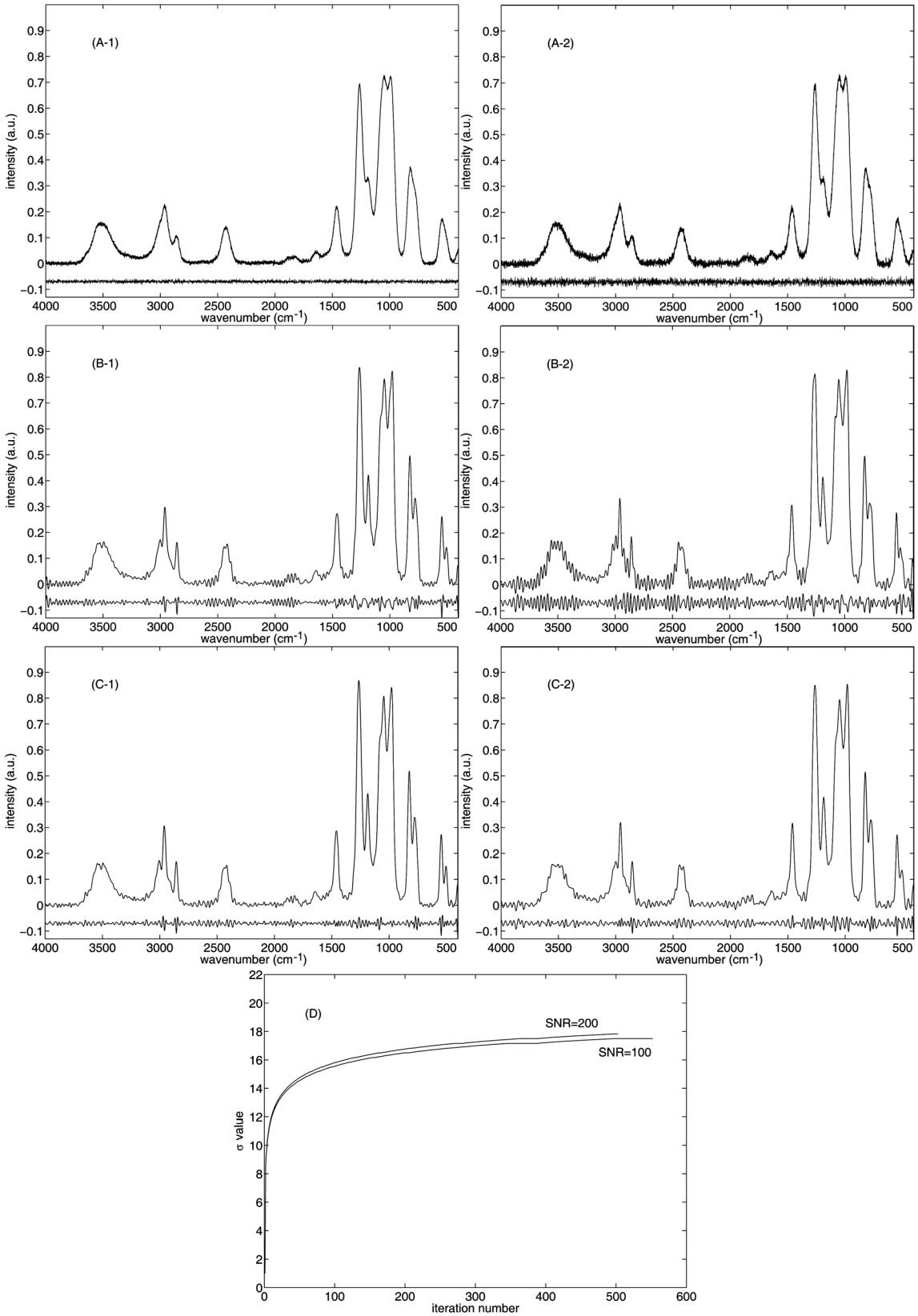

Fig. 7 Comparison between HMSBD and FSD method with different levels of noise, SNR = 200 (left column) and SNR = 100 (right column). (A) Degraded spectra, equal to *noise-free* spectrum (Fig. 4C) plus Gaussian white noise. (B) Deconvolution using the FSD[14] method (narrow factor of 1.5 and 1.4). (C) Deconvolution using the suggested HMSBD method. (D) Convergence curves of the estimated kernel width σ under different SNRs by HMSBD.

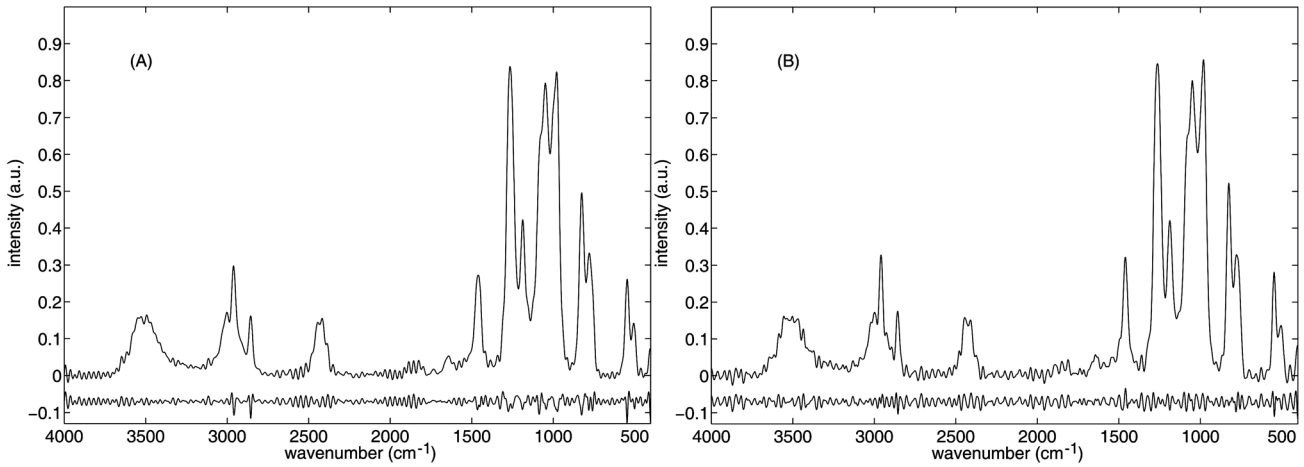

Fig. 8 Deconvoluted results by the suggested GMSBD method using the same parameters setting with HMSBD. (A) SNR = 200. (B) SNR = 100.

Table 2 Figures of merit for the performance of FSD, GMSBD and HMSBD methods under three SNR conditions

		Method		
Merits	SNR	FSD	GMSBD	HMSBD
CC	Noise-free	0.9993	0.9990	0.9996
	200	0.9984	0.9980	0.9989
	100	0.9973	0.9975	0.9981
WCC	Noise-free	0.9994	0.9992	0.9996
	200	0.9989	0.9985	0.9991
	100	0.9977	0.9978	0.9984

than threshold values: $|\sigma_{n+1} - \sigma_n| < d_1$ and $\|f_{n+1} - f_n\|/\|f_n\| < d_2$, where d_1 and d_2 are predetermined coefficients. The deconvolution procedure described in this section is summarized as follows:

Algorithm 1 Spectral deconvolution

Input degraded spectrum g

1: Select α,β and μ

2: Initialize $f_0 = g$, $\sigma_0 = 1$, $\sigma_k >> 1$

3: WHILE $\{ |\sigma_{n+1} - \sigma_n| > d_1 , \| f_{n+1} - f_n \|/\| f_n | > d_2 \}$

(i) Fix $\sigma_{n+1} = \sigma_n$, solve $f_{n+1} = \arg\min_f E(f, \sigma_n)$ using gradient descent flow method.

(ii) Fix $f_{n+1} = f_n$, solve $\sigma_{n+1} = \arg\min_\sigma E(f_{n+1}, \sigma)$ using bisection method.

4: END

Output deconvoluted spectrum and kernel width.

In this paper d_1 and d_2 are small positive constants between 10^{-9} and 10^{-7}.

4.3 Optimization details and parameters

Algorithm 1 shows the skeleton of our algorithm, where the iterative optimization steps and the termination criteria are given. Two parameters in our algorithm, *i.e.*, α and β given in eqn (13), are adjustable. α and β correspond to the probability parameters in eqn (7) and (12) for the spectrum and kernel *priors*, and their values are adapted from their initial values over iterations of the optimization. We set $\alpha = \alpha/\lambda_1$, $\beta = \beta/\lambda_2$. Then, after each iteration of optimization, the values of α and β are divided by λ_1 and λ_2, respectively, where we usually set $\lambda_1 = 1.01$, $\lambda_2 = 1.02$ to reduce the influence of the spectrum *prior* and increase that of the spectrum likelihood.

As noted above, the multiscale statistics require a noise estimate for evaluation. We have evaluated several similar techniques in common use for estimating Gaussian noise based on various differencing schemes that, in general, give comparable results in simulations but are all upwardly biased. We use median absolute differences to estimate δ defined in the following equations.[25,26]

$$\delta = \frac{1.4826}{\sqrt{2}} \text{median}\left\{ |g_i - g_{i-1}|, \quad i = 2, \cdots, n \right\}. \qquad (25)$$

It is worth noting that the method proposed by Turner[27] using second-order difference spectra followed by filtering to remove spike and/or signal artifacts showed the minimum bias from a brief survey of available methods with synthetic spectra. Initializing the egularization parameters, we set $\alpha_0 = 20\delta$, $\beta_0 \in$ [250,350]. In the sequel, spectrum intensities are normalized to the range [0,1].

5 Experimental results and discussion

To show the performance of the proposed algorithm, a series of experiments were carried out on various simulated and real spectra. In the following experiments, we evaluated three kinds of performance of the proposed method, namely, estimation of the spectral slit width, spectral narrowing and noise suppression with different SNRs.

In order to show that the proposed algorithm was not very sensitive to the parameter set, the same parameters were used in the following deconvoluted experiments, *i.e.* $\alpha_0 = 20\delta$, $\beta_0 = 300$, $\mu = 0.02$. The Fourier self-deconvolution[14] (FSD) and GMSBD methods were also tested to make a comparative analysis with the proposed algorithm. For the SBD method, we tested the

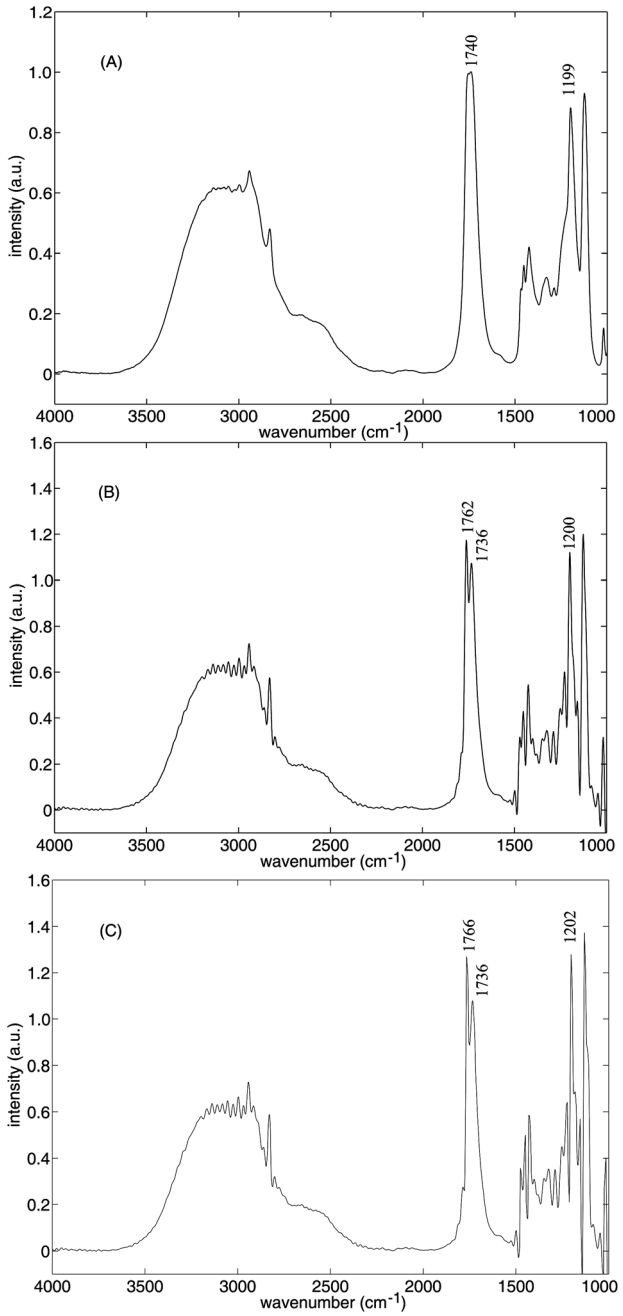

Fig. 9 Real IR spectrum experiment. (A) IR spectrum of aliphatic, carboxylic acid, ether from 4000 to 1000 cm^{-1}. (B) Deconvolution using the FSD[14] method with the kernel width 15.8 cm^{-1} estimated by HMSBD. (C) Deconvolution using the suggested HMSBD method.

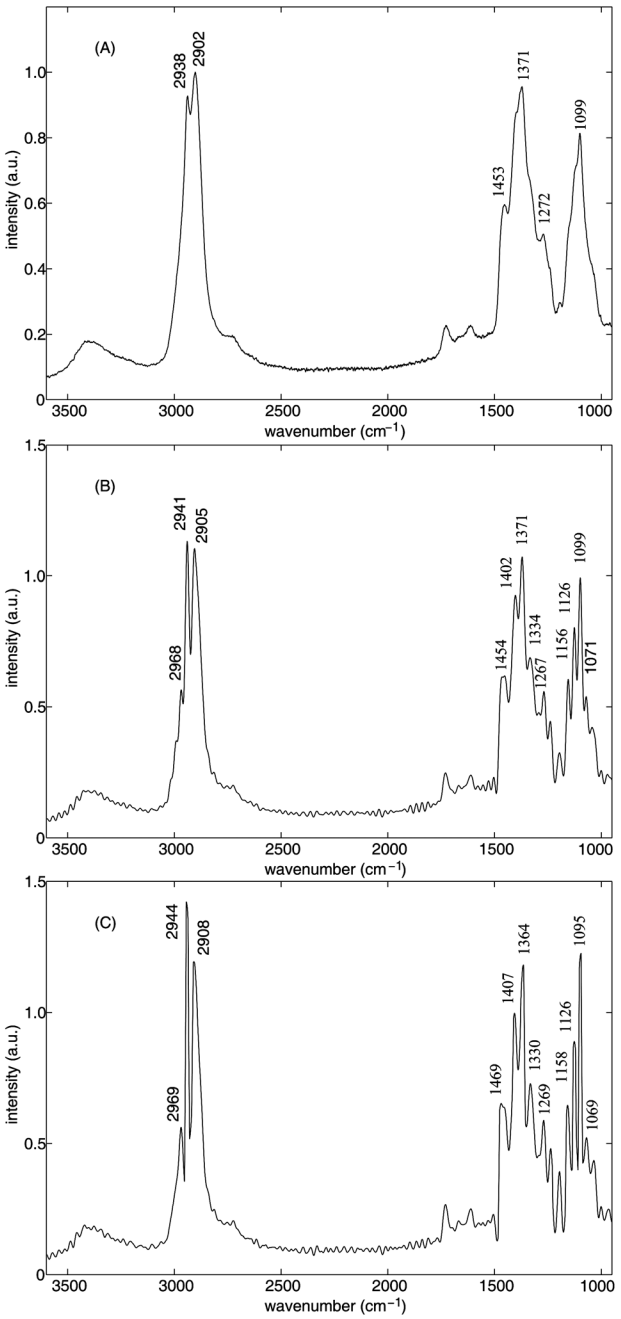

Fig. 10 Real Raman spectrum experiment. (A) Raman spectrum of xanthan from 3600 to 950 cm^{-1}. (B) Deconvolution using the FSD[14] method with kernel width 16.0 cm^{-1}. (C) Deconvolution using the suggested HMSBD method.

performance Gauss–Markov and Huber–Markov *prior*. Because the FSD method needed a known PSF, we applied a predefined kernel.

5.1 Simulation and experimental spectral data

For the simulated data, following the eqn (1), we simulated degraded spectra data by computer on the basis of the experimental IR spectra. Fig. 4A presents the IR spectrum of dimethyl phosphite ($C_2H_7O_3P$) from 4000 to 400 cm^{-1} at 1 cm^{-1} resolution. σ and SNR represent the overlapping degree of adjacent bands and the noise level, respectively. In general, we convolved Fig. 4A with a Gaussian kernel width σ of 18 cm^{-1}, as illustrated in Fig. 4B, to generate the degraded (overlapped and noisy) spectrum (Fig. 4C). The degraded spectrum becomes much smoother and less resolved, and the bands become wider and lower. To investigate the robustness to noise of the deconvolution methods, white Gaussian noise was added to the degraded

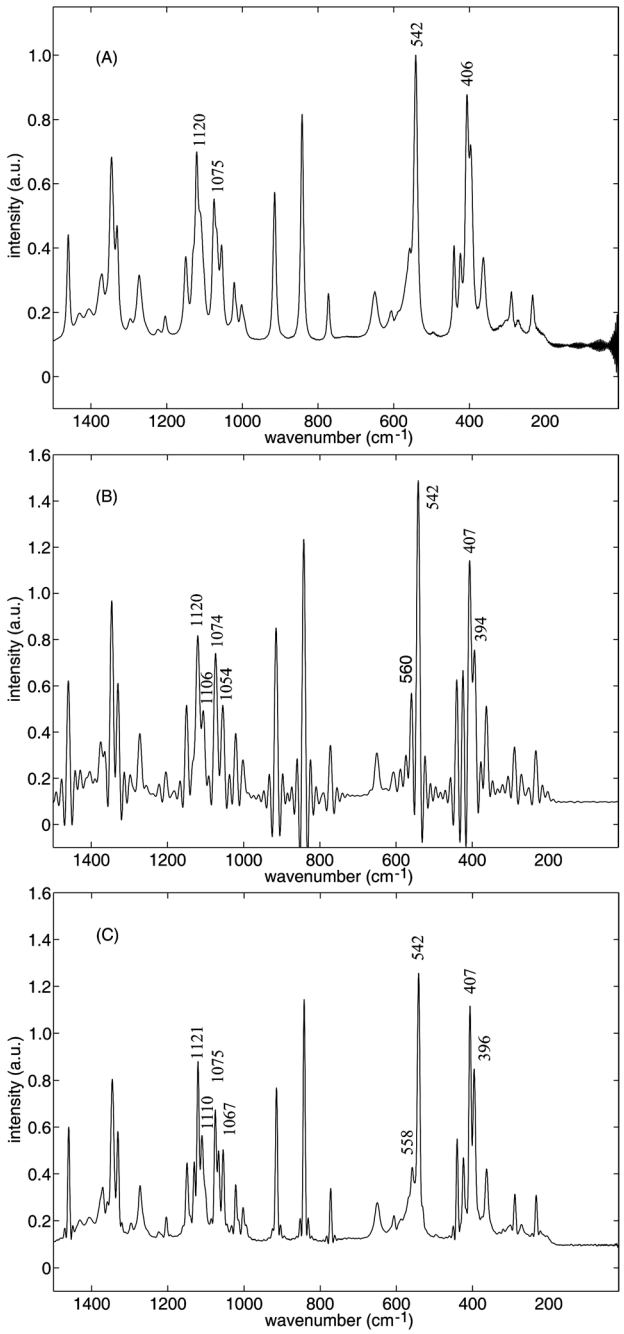

Fig. 11 Real Raman spectrum experiment. (A) Raman spectrum of D(+)-glucopyranose from 1500 to 10 cm^{-1}. (B) Deconvolution using the FSD[14] method with kernel width 9.5 cm^{-1}. (C) Deconvolution using the suggested HMSBD method.

spectrum (Fig. 4C) with two different SNRs of 200 and 100, as shown in Fig. 7A1 and A2, respectively.

For the real data, the spectra are sampled on a Perkin Elmer System 2000 interferometer using a Nd:YAG laser. Raman spectra will be recorded using the following parameters: excitation (1064 nm), laser power (400 mW), Raman shift (3600–0 cm^{-1}), accumulations (256), resolution (1 cm^{-1}). We downloaded the measurement from the online supply network.

5.2 Estimation of spectral mean Gaussian bandwidth

In Fig. 4C, it is difficult to distinguish the peak at 825 cm^{-1} from that 780 cm^{-1}, and at 547 from 501 cm^{-1}. There are overlapping bands between the peaks such as 1266 and 1187 cm^{-1}, 1046 and 975 cm^{-1}. First, we applied three methods on the noise-free degraded spectrum. For FSD, we used the accurate Gaussian kernel in the same way as simulation (Fig. 4B). For GMSBD and HMSBD, we started from a Gaussian kernel with an initial width $\sigma_0 = 1$. Using the eqn (25), we estimate the $\delta = 0.00081$, and we set $\alpha_0 = 0.02$.

Fig. 4D–F show deconvolution results by FSD (with a known kernel, shown in Fig. 4B), GMSBD, and the proposed algorithm. All the deconvoluted spectra are clearly resolved. Visual analysis of deconvoluted spectrum clearly shows that the HMSBD method splits the overlapped peaks very well. In Fig. 4F (HMSBD method), the peak at 820 cm^{-1} is split into two peaks 825 and 780 cm^{-1}, respectively. The peak at 545 cm^{-1} is split into peaks at 574 and 501 cm^{-1}. While in Fig. 4D (FSD method) and Fig. 4E (GMSBD method), the overlapping bands are separated slightly. Therefore, for the noise-free spectrum, it seems clear that the HMSBD not only estimates the blur kernel reliably, but also produces a much narrower spectrum with more details than the FSD method.

The width σ converges at about 18.2 cm^{-1} after 602 iterations (see Fig. 5A), and the estimated kernel was plotted in Fig. 5B. The estimated kernel by the HMSBD method is a close match to the actual kernel.

We also tested the FSD method with inaccurate kernels. It is usually the case that the bandwidth of the measured spectra is equal to the width of the narrowest spectral band. However, the complexity of the natural spectrum makes it easy to obtain smaller or larger values than the actual value. To obtain deeper understanding of these phenomena, we selected the narrower and wider kernels to recover the degraded spectrum, respectively. The outcome is shown in Fig6A and B, where the Gaussian kernel width is 10 and 26 cm^{-1}. Compared to the actual spectrum (Fig. 4D), the recovery spectrum in Fig. 6A is severely infra-deconvoluted, and in Fig. 6B shows obvious over-deconvolution. This demonstrates that the non-blind deconvolution such as FSD, is sensitive to the width of the kernel. At the source of this problem is the aspiration for general kernel width which needs to measure manually. On the contrary, the proposed method HMSBD method does not suffer from these problems.

Furthermore, for the deconvoluted spectra by FSD, GMSBD and HMSBD method, the band distortions were investigated. In the original spectrum (Fig. 4A), the ten absorbance bands at 3002, 2959, 1266, 1187, 1046, 975, 825, 780, 547, 501 cm^{-1} are taken as references. Table 1 lists these bands distortions in position, width, height and area between the deconvoluted and actual spectrum (Fig. 4A). Roots of mean square error (RMSE) of these distortions are calculated. In terms of RMSE, the distortion of the positions is smaller than those by the FSD method.

5.3 Huber–Markov spectral semi-blind deconvolution

Effect of the noise. For the results of FSD, we execute the source code from the authors and hand-tune parameters to

produce the best possible results. And the kernel used in FSD is the Gaussian line-shape. For SNR = 200, we compute the $\delta = 0.0041$, and set $\alpha_0 = 0.05$; for SNR = 100, $\delta = 0.0078$ and set $\alpha_0 = 0.08$. Fig. 7B1 and B2 and C1 and C2 present the deconvolution results for noisy spectra with FSD and HMSBD respectively. The residual noise is shown at the bottom of the result. Therefore, we conclude that the proposed HMSBD method recovers more spectral details and suppresses more noise than FSD.

Fig. 7D presents the estimated kernel width from noisy spectra by the HMSBD method. It can be seen that, as SNR is decreased, the estimated width of the kernel becomes smaller; as a result, the recovered spectrum becomes less deconvoluted and with more residual noise. On the other hand, for the better noise suppression, a larger α should be chosen to favor a smoother resolution, which will also result in a narrow kernel, and consequently lead to infra-deconvolution. This is why the spectrum in Fig. 7D (SNR = 100) is much less deconvoluted than that in Fig. 7D (SNR = 200). Therefore, for the proposed HMSBD method, regularization parameter α should be chosen as a balance between narrowing capability and noise suppression, especially at low SNRs.

Effect of Huber function. To demonstrate the effectiveness of the Huber function, we compare the performances of the HMSBD (Huber–Markov *prior*) method with GMSBD (Gauss–Markov *prior*). With the same parameter setting, Fig. 8A and B show the results of the GMSBD method with SNR = 200 and 100. It is clear that the whole spectrum consists of plain regions and steep regions. On one hand, the ability to suppress noise of HMSBD is the same as GMSBD in the plain region. On the other hand, the percentage of steep regions is far less than those of the plain regions. Therefore, the total difference between two methods is not significant. However, for the GMSBD method, the noise will be suppressed but the steep regions will be also smoothed. The HMSBD can suppress noise adaptively and preserve the detail of the steep regions. We suggest that the HMSBD method suppresses noise in the plain regions and preserves detail in the steep regions.

5.4 Quantitative assessment

To go further and provide a quantitative assessment of the new spectroscopic deconvolution method, two quality indexes are employed. These are the correlation coefficient (CC) and weighted correlation coefficient (WCC). The CC is defined as

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}\left(f_i - \overline{f}\right)\left(\hat{f}_i - \overline{\hat{f}}\right)}{\sqrt{\sum_{i-1}^{N}\left(f_i - \overline{f}\right)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{f}_i - \overline{\hat{f}}\right)^2}}$$

the horizontal bar denotes the average of all the spectroscopic data. Pearson's CC represents the average similarity between the trends of the actual and recovered spectra, and the larger value denotes the better match; for example, CC = 1 means a perfect match.

The WCC method[28] is used to evaluate the similarity between the original spectrum and the object spectrum. It is defined by

$$\mathrm{WCC} = \frac{\sum_{i=1}^{N} w_i\left(f_i - \overline{f}\right)\left(\hat{f}_i - \overline{\hat{f}}\right)}{\sqrt{\sum_{i-1}^{N} w_i\left(f_i - \overline{f}\right)^2}\sqrt{\sum_{i=1}^{N} w_i\left(\hat{f}_i - \overline{\hat{f}}\right)^2}},$$

where w is the weight array, and the horizontal bar denotes the weighted means. The weight array w was constructed in terms of the intensities of actual spectrum and the difference between the actual and recovered spectra.

Table 2 shows the resulting CC and WCC for each method based on simulated data with different SNRs. It is clear that HMSBD is superior to FSD and GMSBD under all SNR conditions.

Therefore, we conclude that the proposed HMSBD method recovers more spectral details and suppresses more noise than the FSD and GMSBD methods.

5.5 Deconvolution of experimental spectra

Three more challenging real examples are shown in Fig. 9–11, all contain overlapping peaks and random noise. The top, middle and bottom rows correspond to the original spectra, FSD and HMSBD results. There is no actual bandwidth to the experimental spectra, the bandwidth used in FSD method is estimated from HMSBD.

Fig. 9A shows the infrared absorbance spectrum of aliphatic, carboxylic acid, ether from 4000 to 400 cm^{-1} at 1 cm^{-1}. The spectrum is deconvoluted by FSD using 15.8 cm^{-1} Gaussian width (equal to the estimated width by HMSBD). The peak at 1740 cm^{-1} is split into two peaks at 1766 and 1736 cm^{-1}, respectively. Here, we computed $\delta = 0.0015$ and set $\alpha_0 = 0.03$. The Gaussian width estimates by HMSBD method equals 15.8 cm^{-1} (not shown). The peak at 1740 cm^{-1} is split into two peaks at 1762 and 1736 cm^{-1}, respectively. It seems that the deconvolution result in Fig. 9C is more resolved than that in Fig. 9B.

The HMSBD method is also applied in Raman spectra. Fig. 10A is the Raman spectrum of xanthan[29] ($C_{35}H_{49}O_{29}$) from 3600 to 1050 cm^{-1} at 1 cm^{-1}. The deconvolution spectrum shown in Fig. 10B is far more resolved. Here, we computed $\delta = 0.0033$ and set $\alpha_0 = 0.07$. The HMSBD method estimates the Gaussian width to be 15.8 cm^{-1}. The peaks 2938 and 2902 cm^{-1} are split far from each other. Five bands are resolved between 1500 and 1160 cm^{-1}, and the peak 1099 is split into five peaks, 1158, 1126, 1095, 1069 and 1050 cm^{-1}. The FSD method also works well in the accurate kernel achieved from HMSBD. The result is shown in Fig. 10B.

Fig. 11A shows the 1495 length Raman spectral data of D(+)-glucopyranose[30] ($C_6H_{12}O_6$) from 1500 to 5 cm^{-1} at 1 cm^{-1}. Fig. 11C is the deconvoluted data. Here, we compute $\delta = 0.0006$ and set $\alpha_0 = 0.01$. The kernel width σ estimated by HMSBD approximates 9.5 cm^{-1}. The peak at 406 cm^{-1} is split into two peaks at 396 and 407 cm^{-1}. The peaks at 542, 1075 and 1120 cm^{-1} are also split into two peaks. Noise is suppressed very well from 200 to 5 cm^{-1}, and closes nearly to a line. The result from the FSD method (Fig. 11B) shows more noise.

Computing times in the above examples are between 35 and 55 seconds. The proposed algorithm is implemented by MATLAB R2010a on a computer with Intel Core 2 Duo 2-GHz CPU, 2-G RAM, and Microsoft Windows XP operating system.

6 Conclusions

In this paper, we show how the issue of overlapping peaks can be considered as an MAP-posed problem and solved by minimizing an object functional that includes a likelihood term and two *prior* terms. The major novelty of the proposed algorithm is that it can estimate the kernel slit width and latent spectrum simultaneously. The method does not require a known PSF in advance but models as a parametric function by combining *a prior* knowledge about the instrumental response characteristic. Moreover, the Huber function is introduced to adaptively describe the steep and plain regions characteristic of spectral data. Experimental results suggest that the method can recover spectral structural details as well as suppress noise effectively. In particular, owing to no need for a known PSF, the new method has considerable value in practice. It is worth noting that although the Gaussian kernel only is discussed here, any other kernels with parametric forms, such as Lorentzian or Voigt functions can be applied to the HMSBD method.

We recognize an unknown molecular based on this work by extracting spectrum characteristics such as band numbers, positions, widths, area and so on. We will examine this finding in future research.

Acknowledgements

This research was supported by the Project of the National Natural Science Foundation of China under grant no. 60736010 and 60902060. The authors would like to thank the reviewers for their valuable and helpful suggestions.

References

1 A. Economou, P. R. Fielden and A. J. Packham, *Analyst*, 1996, **121**, 1015–1018.
2 A. Economou, P. R. Fielden and A. J. Packham, *Analyst*, 1996, **121**, 97–104.
3 J. K. Kauppinen, D. J. Moffatt, H. H. Mantsch and D. G. Cameron, *Anal. Chem.*, 1981, **53**, 1454–1457.
4 V. A. Lorenz-Fonfria and E. Padros, *Analyst*, 2004, **129**, 1243–1250.
5 X. Q. Zhang, J. B. Zheng and H. Gao, *Analyst*, 2000, **125**, 915–919.
6 J. Ottaway, J. H. Kalivas and E. Andries, *Appl. Spectrosc.*, 2010, **64**, 1388–1395.
7 J. Yuan, Z. Hu and J. Sun, *Appl. Opt.*, 2005, **44**, 7595–7601.
8 T. M. D. Ebbels, J. C. Lindon and J. K. Nicholson, *Appl. Spectrosc.*, 2001, **55**, 1214–1224.
9 R. Narayan and R. Nityananda, *Annu. Rev. Astron. Astrophys.*, 1986, **24**, 127–170.
10 V. A. Lórenz-Fonfría and E. Padrós, *Appl. Spectrosc.*, 2005, **59**, 474–486.
11 Y. Senga, K. Minami, S. Kawata and S. Minami, *Appl. Opt.*, 1984, **23**, 1601–1608.
12 P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley and Sons, New York, 2nd edn, 2007.
13 R. K. Singh and A. L. Verma, *J. Raman Spectrosc.*, 1997, **28**, 301–304.
14 V. A. Lórenz-Fonfría and E. Padrós, *Appl. Spectrosc.*, 2009, **63**, 791–799.
15 V. A. Lórenz-Fonfría and E. Padrós, *Appl. Spectrosc.*, 2008, **62**, 689–700.
16 P. A. Jansson, *Deconvolution of Images and Spectra*, Dover Science, Mineola, New York, 2nd edn, 2007.
17 J. Yuan and Z. Hu, *Appl. Spectrosc.*, 2006, **60**, 692–697.
18 J. KatraSnik, F. Pernu and B. T. Likar, *Appl. Spectrosc.*, 2010, **64**, 1265–1273.
19 A. Mohammad-Djafari, J. F. Giovannelli, G. Demoment and J. Idier, *Int. J. Mass Spectrom.*, 2002, **215**, 175–193.
20 Q. Wang, D. D. Allred and J. González-Hernández, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 6119.
21 R. Pan and S. J. Reeves, *IEEE Trans. Image Process.*, 2006, **15**, 3728–3735.
22 Y. Yu-Li and M. Kaveh, *IEEE Trans. Image Process.*, 1996, **5**, 416–428.
23 T. F. Chan and W. Chiu-Kwong, *IEEE Trans. Image Process.*, 1998, **7**, 370–375.
24 R. R. Schultz and R. L. Stevenson, *IEEE Trans. Image Process.*, 1996, **5**, 996–1011.
25 A. A. Urbas and S. J. Choquette, *Appl. Spectrosc.*, 2011, **65**, 665–677.
26 P. L. Davies and A. Kovac, *Ann. Stat.*, 2001, **20**, 65.
27 H. G. Schulze, M. M. L. Yu, C. J. Addison, M. W. Blades and R. F. B. Turner, *Appl. Spectrosc.*, 2006, **60**, 820–825.
28 P. R. Griffiths and L. Shao, *Appl. Spectrosc.*, 2009, **63**, 916–919.
29 S. B. Engelson, *Raman Spectral of Xanthan*, http://www.models.kvl.dk/specarb.
30 S. B. Engelson, *Raman Spectral of D(+)-Glucopyranose*, http://www.models.kvl.dk/specarb.