A model based data driven dictionary learning for seismic data representation

Can Evren Yarman[1], Rajiv Kumar[2], James Rickett[3]

March 7, 2017

[1] cyarman@slb.com. Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge, UK CB3 0EL

[2] rajmittal09@gmail.com. University of British Columbia, Department of Earth, Ocean and Atmospheric Sciences 2020 - 2207 Main Mall Vancouver, BC, CAN V5X1S1 6043557908

[3] jrickett@slb.com. Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge, UK CB3 0EL

Abstract

Planar waves events recorded in a seismic array can be represented as lines in the Fourier domain. However, in the real-world seismic events usually have curvature or amplitude variability that means their Fourier transforms are no longer strictly linear, but rather occupy conic regions of the Fourier domain that are narrow at low frequencies, but broaden at high frequencies where the effect of curvature becomes more pronounced. One can consider these regions as localised "signal cones". In this work, we consider a space-time variable signal cone to model the seismic data. The variability of the signal cone is obtained through scaling, slanting and translation of the kernel for cone-limited (C-limited) functions (functions whose Fourier transform lives within a cone) or C-Gaussian function (a multivariate function whose Fourier transform decays exponentially with respect to slowness and frequency) which constitutes our dictionary. We find a discrete number of scaling, slanting and translation parameters from a continuum by optimally matching the data. This is a nonlinear optimization problem which we address by a fixed point method which utilizes a variable projection method with ℓ_1 constraints on the linear parameters and bound constraints on the nonlinear parameters. We observe that slow decay and oscillatory behavior of the kernel for C-limited functions constitute bottlenecks for the optimization problem which we partially overcome by the C-Gaussian function. We demonstrate our method through an interpolation example. We present the interpolation result using the estimated parameters obtained from the proposed method and compare it with those obtained using sparsity promoting curvelet decomposition, matching pursuit Fourier interpolation and sparsity promoting plane wave decomposition methods.

Keywords— dictionary learning, kernel method, reproducing kernel, variational method, nonlinear optimization, variable projection

List of Figures

List of Tables

1 Introduction

Seismic data are modeled through the wave equation. Under the acoustic wave equation assumption, the bounds for the dispersion relationship obtained from the wave equation defines a combination of triangular or conic regions in the Fourier domain. This combined region is referred to as the signal cone. Adopting the term R-limited functions (functions whose Fourier transform is supported within a region R) coined in Slepian (1964), we refer to functions whose Fourier transforms are supported within the signal cone as C-limited functions. A C-limited function, for example, seismic data, can be expressed in terms of a convolution using a kernel, whose Fourier transform is equal to unity within the signal cone and zero everywhere else.

The concept of C-limited functions can be applied in a broader context as well. Plane-waves arriving at the surface can be represented as a linear function in the Fourier domain. However, curvature and amplitude variability causes leakage, meaning that, in practice, seismic events occupy conic regions of the Fourier domain that are broader at high frequencies than at low frequencies. Although these quasi-planar events may not be strictly C-limited, they can be well modeled with functions of this form.

Having explicit forms for the kernels eases the design of the algorithms in the measurement domain for forward and inverse problems in seismic data processing. Examples are filter design, data regularization and interpolation of single as well as multiple component measurements, slant stack and least square slant stack transformation, directional filtering, etc. For example, once the seismic data are approximated as a discrete sum of a kernel, the slant stack can be computed analytically Yarman and Flagg (2014).

Approximation of the seismic data as a discrete sum of kernels for C-limited function can be performed in many ways. There are many potential parameterizations for the kernel functions, and many potential optimisation schemes. The goal of this work is to capture the salient features of seismic data with high resolution, while maintaining a discrete sum of fixed size. This necessitates an adaptive algorithm, where the coefficients and parametrization of the kernels are both optimized to fit the data.

In this regard, our main contributions in this work are

- Formalizing the concept of C-limited functions and analytic representation of the corresponding kernels in time and space.
- Introducing C-Gaussian functions and their analytic representations or approximations. C-Gaussian functions are a generalization of Gaussians intended to approximate C-limited functions.

- The concept of effectively C-limited functions, which are defined in terms of a convolution with C-Gaussian functions.

- A physics motivated model based dictionary learning for seismic data processing. Unlike the approaches presented in Cai et al. (2014); Beckouche and Ma (2014); Liang et al. (2014), where discretized basis functions are obtained using K-SVD, a clustering algorithm guided by local coherence inferred from singular value decomposition, we have, by construction, analytic basis functions that are in the form of a paraxial ray based solution of the wave equation (see Section 5.8 in Červený (2001)).

The outline of our discussion is as follows. In Section 2, we state the problem formulation. Section 3, starts with the wave equation and introduce the signal cone, which is used to define C-limited functions. This is followed by definition of the C-Gaussian function, which consequently is used to define effectively C-limited functions. After the introduction of C-limited and effectively C-limited functions, we present the representation problem for the sampled seismic data and our proposed solution, which is the fixed point method obtained by modification of variable projection method. We present numerical results in Section 4. Finally we conclude our discussion in Section 5.

In Appendix A, we provide explicit expression for the primitive forms of K_C and K_G in Tables 1 and 3. Derivation of the kernel for C-limited functions in dimensions $(1+1)\,D$ and $(1+2)\,D$ are presented in Appendices B and D. Derivation of the C-Gaussian function for different dimensions are presented in Appendices C, E and F. Review of the variable projection method is given in Appendix G.

2 Problem formulation

2.1 Representation of seismic data

Let $K(t, \mathbf{x}, \mathbf{y})$ denote the kernel, $d(t, \mathbf{s}, \mathbf{r})$ be the measured data observed at a receiver $\mathbf{r} \in \mathbb{R}^{N_\mathbf{r}}$ due to a source initiated at $\mathbf{s} \in \mathbb{R}^{N_\mathbf{s}}$, $N_{\mathbf{s}/\mathbf{r}} = 0, 1, 2, 3$. We consider an approximation of the data as a sum of shifted versions of the kernel $K(t, \mathbf{x}, \mathbf{y})$ centered around $(t_m, \mathbf{x}_m, \mathbf{y}_m)$, slanted with source and receiver slownesses $\mathbf{p}_{\mathbf{s},m}$ and $\mathbf{p}_{\mathbf{r},m}$, and scaled in time and space by $\rho_{\omega/p_\mathbf{s},p_\mathbf{r},m}$:

$$\hat{d}_{(\mathcal{D},\mathcal{P})}(t, \mathbf{s}, \mathbf{r}) = \sum_{m=1}^{M} d_m K\left(\tau(t, P_m), \mathbf{x}(t, P_m), \mathbf{y}(t, P_m)\right) \tag{1}$$

where

$$\tau\left(t, P_m\right) = \rho_{\omega,m}\left(\left[t - t_m\right] - \left(\mathbf{p}_{\mathbf{s},m} \cdot \left[\mathbf{s} - \mathbf{x}_m\right] - \mathbf{p}_{\mathbf{r},m} \cdot \left[\mathbf{r} - \mathbf{y}_m\right]\right)\right), \tag{2}$$

$$\mathbf{x}\left(\mathbf{s}, P_m\right) = \rho_{k_{\mathbf{s}},m}\left[\mathbf{s} - \mathbf{x}_m\right] = \rho_{\omega,m}\rho_{p_{\mathbf{s}},m}\left[\mathbf{s} - \mathbf{x}_m\right], \tag{3}$$

$$\mathbf{y}\left(\mathbf{r}, P_m\right) = \rho_{k_{\mathbf{r}},m}\left[\mathbf{r} - \mathbf{y}_m\right] = \rho_{\omega,m}\rho_{p_{\mathbf{r}},m}\left[\mathbf{r} - \mathbf{y}_m\right]. \tag{4}$$

Here $\tau\left(t, P_m\right)$ is the scaled, slanted and shifted time around t_m, $\mathbf{x}\left(\mathbf{s}, P_m\right)$ and $\mathbf{y}\left(\mathbf{r}, P_m\right)$ are the shifted and scaled spatial coordinates centered around $\mathbf{x}_m$ and $\mathbf{y}_m$, respectively. The linear and nonlinear parameter sets $\mathcal{D}$ and $\mathcal{P}$ are defined by

$$\mathcal{D} = \left\{d_m\right\}_{m=1,\ldots,M} \tag{5}$$

and

$$\mathcal{P} = \left\{P_m = \left(t_m, \mathbf{x}_m, \mathbf{y}_m, \mathbf{p}_{\mathbf{s},m}, \mathbf{p}_{\mathbf{r},m}, \rho_{\omega,m}, \rho_{p_{\mathbf{s}},m}, \rho_{p_{\mathbf{r}},m}\right) \in \mathbb{R}^{4+2N_{\mathbf{s}}+2N_{\mathbf{r}}}\right\}_{m=1,\ldots,M}. \tag{6}$$

It should be noted that $\rho_{p_{\mathbf{s}/\mathbf{r}},m} \in \mathbb{R}$ may be replaced with matrices $R_{p_{\mathbf{s}/\mathbf{r}},m} \in GL_{N_{\mathbf{x}/\mathbf{y}}}\left(\mathbb{R}\right)$, however, for the sake of brevity of the discussion we will consider the scalar case.

As a result, we consider the following minimization problem

$$\underset{(\mathcal{D},\mathcal{P})}{\operatorname{argmin}}\, J\left(\mathcal{D}, \mathcal{P}\right) \tag{7}$$

with constraints on linear and nonlinear parameter sets $\mathcal{D}$ and $\mathcal{P}$, where

$$J\left(\mathcal{D}, \mathcal{P}\right) = \frac{1}{2}\int \left[d\left(t, \mathbf{s}, \mathbf{r}\right) - \hat{d}_{(\mathcal{D},\mathcal{P})}\left(t, \mathbf{s}, \mathbf{r}\right)\right]^2 d\mu\left(t, \mathbf{s}, \mathbf{r}\right) \tag{8}$$

is some least square error between the data and its approximation $\hat{d}_{(\mathcal{D},\mathcal{P})}\left(t, \mathbf{s}, \mathbf{r}\right)$ with respect to some measure μ.

In the absence of constraints, linear parameter set $\mathcal{D}$ depends on the nonlinear parameter set $\mathcal{P}$ and the least square minimization problem (7) can be solved using variable projection algorithm Golub and Pereyra (1973, 2003). In the presence of certain constraints, such as sparsity constraints on linear parameters and bound constraints on nonlinear parameters, one can utilize fixed point algorithms Bauschke et al. (2011); Bolte et al. (2014); Parikh and Boyd (2013). We used an alternating linearized minimization algorithm obtained through modification of the variable projection method to address

the minimization problem (7) which will be discussed in Section 3.4.2.

2.2 Literature

The minimization problem (7) can be handled in two ways. The first way is to find $\mathcal{P}$ as a subset of a predefined countably finite set that satisfies the constraints on $\mathcal{P}$. This is achieved by prediscretization of the nonlinear parameter space and then solve the constrained minimization problem (7) for the linear parameters $\mathcal{D}$. This is analogous to performing a ***global search*** from a predefined library with countable finite number of elements. The advantage of this approach is that it reduces the minimization problem to a convex optimization problem with constraints. For this task there are mainly three classes of approaches: (i) greedy methods such as matching pursuit and its variants Mallat and Zhang (1993); Pati et al. (1993); Tropp et al. (2007); Donoho et al. (2012); (ii) iterative hard thresholding Blumensath and Davies (2008); Blumensath (2012); Fornasier and Rauhut (2008); and (iii) basis pursuit, also referred to as iterative soft thresholding Chen et al. (1998); Candès and Wakin (2008); Yin et al. (2008); Candès et al. (2011); Andersson et al. (2012) or its variants Andersson et al. (2012). In order to avoid local minima of the cost function with respect to $\mathcal{P}$, one may need to globally sample the nonlinear parameter, sometimes densely, depending on the oscillatory behavior of J with respect to $\mathcal{P}$. As the number of samples for nonlinear parameters increases so does the computational complexity of finding the optimal values for $\mathcal{D}$. This especially constitutes a challenge with the introduction of curvature, window size and other parameters into formation of the library Hoecht et al. (2009).

On the other hand, a ***descent*** type of a search method can be used to build up the library on the fly from an uncountably infinite basis set. The computational cost of these methods are proportional to the number of elements you would utilize in representation of your data which makes them appealing for large-scale machine learning applications consuming big data sets Hinton and Salakhutdinov (2006); Cevher et al. (2014); Fan et al. (2014); Richtárik and Takáč (2016). These methods can be classified under (i) continuous basis pursuit Ekanadham et al. (2011); Knudson et al. (2014), where the continuous infinite dictionary is approximated through a discrete subset and its variations which reduces the problem to a basis pursuit problem; and (ii) variational methods Candès and Fernandez-Granda (2013, 2014); Duval and Peyré (2015); Chauffert et al. (2015), where the nonlinear optimization problem is tackled to estimate the basis parameters. Continuous basis pursuit can be conceptualized as an approximation to the class of variational methods. Our approach falls into the variational methods where we considered multidimensional signals and a multivariate basis.

2.3 Choice of kernel

Kernels for the C-limited functions are analogous to sine cardinal function (sinc), which is the kernel for the band-limited functions. Similar to the sinc, they are oscillatory and decay slowly, which limits their localization properties in space and time. One way to increase the decay rate is finding kernels that are analogous to Gaussian function. We extend the kernels for C-limited functions by replacing the characteristic functions in frequency and slowness domains that define the signal cone with Gaussian functions. We refer to these kernels as C-Gaussian functions. Oscillations to C-Gaussian functions can be introduced by taking temporal derivatives. The advantage of C-Gaussian functions compared to the kernels for C-limited functions are:

1. Their Fourier transforms are supported over the entire frequency-wavenumber domain;
2. They decay fast both in the frequency-wavenumber domain as well as in the time-space domain;
3. They can be tailored to have the majority of their energy lie within the signal cone;
4. Their derivatives also decay fast both in the Fourier domain as well as in the time-space domain;
5. The amount of oscillations they contain can be controlled by the order of temporal derivatives;
6. They can be efficiently approximated as a finite sum of Gaussian functions;
7. Their temporal and spatial derivatives can be computed or approximated analytically;
8. Unlike kernels for C-limited functions, they don't have removable singularities;
9. Some of the bound constraints, which are required when kernels for the C-limited are used, may be removed to reduce the steps in optimization algorithms.

A possible disadvantage of the C-Gaussian functions is that they are not strictly C-limited. However, this is not a problem when fitting data based on a quasi-plane-wave/localised-signal-cone model, as in this case the C-limits do not represent strict physical bounds.

3 Theoretical background

3.1 Wave equation and signal cone

Considering an acoustic medium, the measured seismic data $d(t, \mathbf{s}, \mathbf{r})$, $(t, \mathbf{s}, \mathbf{r}) \in \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}^3$, is modeled as the wave field $u(t, \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^3$, observed at $\mathbf{r} \in \mathbb{R}^3$ due to a source at $\mathbf{s} \in \mathbb{R}^3$ fired at time $t = 0$ that

satisfies the wave equation:

$$\left[\partial_t^2 - c^2(\mathbf{x})\nabla_{\mathbf{x}}^2\right] u(t,\mathbf{x}) = \delta(\mathbf{x}-\mathbf{s})\,\delta(t) \tag{9}$$

Away from source, the wave-field satisfies the homogeneous wave equation

$$\left[\partial_t^2 - c^2(\mathbf{x})\nabla_{\mathbf{x}}^2\right] u(t,\mathbf{x}) = 0 \tag{10}$$

For a homogeneous medium $c(\mathbf{x}) = c$, in the Fourier domain (10) leads to the dispersion relationship:

$$\frac{\omega^2}{c^2} = |\mathbf{k}_{\mathbf{x}}|^2 \tag{11}$$

where ω and $\mathbf{k}_{\mathbf{x}}$ are duals of time and space, respectively, referred to as the frequency and wavenumber

Receivers in a marine streamer can measure $k_{\mathbf{x},h}$, the horizontal component of $\mathbf{k}_{\mathbf{x}}$ in the direction parallel to the cable. For propagating waves, where $k_{\mathbf{x},v}$, the vertical component of $\mathbf{k}_{\mathbf{x}}$ is real, this leads to the inequality:

$$k_{\mathbf{x},h} \leq \sqrt{\frac{\omega^2}{c^2}} \tag{12}$$

Similarly an areal array of receivers can measure the two horizontal components, $k_{\mathbf{x},h_1}$ and $k_{\mathbf{x},h_2}$, and the inequality becomes:

$$\sqrt{k_{\mathbf{x},h_1} + k_{\mathbf{x},h_2}} \leq \sqrt{\frac{\omega^2}{c^2}} \tag{13}$$

These define triangular and conic regions of the $\omega - \mathbf{k}$ domain that are known as the signal cone, which we denote by C.

The slowness is defined as $\mathbf{p}_{\mathbf{x}} = \omega^{-1}\mathbf{k}_{\mathbf{x}}$ and is bounded by $c_{\max}^{-1} = p_{\min} \leq |\mathbf{p}| \leq p_{\max} = c_{\min}^{-1}$, where $c_{\min} \leq c(\mathbf{x}) \leq c_{\max}$. As a consequence, the Fourier transform of the data is expected to reside within the signal cone, which we denote by C, is defined by

$$C = \left\{(\omega, \mathbf{k}_{\mathbf{x}} = \omega\mathbf{p}_x) \in \mathbb{R}\times\mathbb{R}^3 | \omega \in [-\omega_{\max}, \omega_{\max}], |\mathbf{p}_{\mathbf{x}}| \leq p_{\max}\right\}. \tag{14}$$

We refer to the functions whose Fourier transform is supported within the signal cone C as C-limited functions. Considering the reciprocity of the wave equation with respect to source and receiver loca-

tions, the cone can be generalized as

$$C = \{(\omega, \mathbf{k_x} = \omega \mathbf{p_x}, \mathbf{k_y} = \omega \mathbf{p_y}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m | \ |\mathbf{p_x}| \leq p_{\mathbf{x},\max}, \ |\mathbf{p_y}| \leq p_{\mathbf{y},\max}, \ \omega \in [-\omega_{\max}, \omega_{\max}]\} \tag{15}$$

for $n, m \in \{0, 1, 2, 3\}$. Here $\mathbf{x}$ and $\mathbf{y}$ can be associated with source and receiver locations, respectively, n and m may be associated to the dimensions of the source and receiver acquisition surfaces. Without loss of generality we assume that $p_{x,\max} = p_{y,\max} = p_{\max}$.

We denote dimensions by $(1+n+m)$D where the first dimension, denoted by 1, is temporal and further dimensions, denoted by n and m, are spatial. The need for two different spatial dimensions emerges with the source and receiver dependency of the seismic data . These two spatial dimensions may possess different symmetry properties. If either of the dimensions n or m are zero, we denote the dimension by $(1+m)$D or $(1+n)$D, respectively.

For the rest of our discussion we will introduce the concepts for the general case of $(1+n+m)$D. We present analytic expressions of kernels for C-limited and C-Gaussian functions in 1D and $(1+1)$D and their analytic approximation for $(1+2)$D. Our examples are shown for a temporal and a single spatial dimension in 1D, i.e. $(1+1)$D. The computations can easily be extended to $(1+n+m)$D.

3.2 C-limited functions

A C-limited function can be defined by functions whose Fourier transform are supported within the cone C:

$$f_C(t, \mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{1+n+m}} \int_C \tilde{f_C}(\omega, \mathbf{k_x}, \mathbf{k_y}) \, \mathrm{e}^{\mathrm{i}(\omega t - [\mathbf{k_x} \cdot \mathbf{x} + \mathbf{k_y} \cdot \mathbf{y}])} d\omega \, d\mathbf{k_x} \, d\mathbf{k_y}, \tag{16}$$

where $(t, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$, or using the convolution integral

$$f_C(t, \mathbf{x}, \mathbf{y}) = \int_{\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m} f(\tau, \mathbf{x}', \mathbf{y}') \, K_C(t - \tau, \mathbf{x} - \mathbf{x}', \mathbf{y} - \mathbf{y}') \, d\tau \, d\mathbf{x}' \, d\mathbf{y}' \tag{17}$$

where

$$K_C(t, \mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{1+n+m}} \int_C \mathrm{e}^{\mathrm{i}(\omega t - [\mathbf{k_x} \cdot \mathbf{x} + \mathbf{k_y} \cdot \mathbf{y}])} d\omega \, d\mathbf{k_x} \, d\mathbf{k_y} \tag{18}$$

Given samples of $f(t, \mathbf{x}, \mathbf{y})$ for some $\{t_k, \mathbf{x}_k, \mathbf{y}_k\}_{k=1}^K \subset \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$, a C-limited function $g_C(t, \mathbf{x}, \mathbf{y})$

that agrees with these sample points can be constructed by first discretizing the integral (17)

$$g_C(t,\mathbf{x},\mathbf{y}) = \sum_{l=1}^{K} g_l K_C(t-t_l, \mathbf{x}-\mathbf{x}_l, \mathbf{y}-\mathbf{y}_l) \tag{19}$$

for some $g_l \in \mathbb{R}$ and then solving the following linear system for g_l:

$$g_C(t_k,\mathbf{x}_k,\mathbf{y}_k) = f(t_k,\mathbf{x}_k,\mathbf{y}_k) = \sum_{l=1}^{K} g_l K_C(t_k-t_l, \mathbf{x}_k-\mathbf{x}_l, \mathbf{y}_k-\mathbf{y}_l). \tag{20}$$

The least square solution of (20) becomes a generalization of Theorem IV in Yen (1956) for C-limited functions.

In Table 1, we present primitive forms of $K_C(t,\mathbf{x},\mathbf{y})$ ($\omega_{\max} = p_{\max} = 1$) for dimensions 1D, $(1+1)$D and $(1+2)$D. An extended table for $(1+n+m)$D, $n,m \in \{0,1,2,3\}$ is derived and presented in Yarman (2015). Deriving a desired $K_C(t,\mathbf{x},\mathbf{y})$ from these primitive forms is straightforward via scaling, stretching, slanting or other linear transformations.

3.3 C-Gaussian functions and effectively C-limited functions

If we consider $K_C(t,\mathbf{x},\mathbf{y})$ as a generalization of $\mathrm{sinc}(t)$, rewriting $K_C(t,\mathbf{x},\mathbf{y})$ as

$$\begin{aligned} K_C(t,\mathbf{x},\mathbf{y}) &= \frac{1}{(2\pi)^{1+n+m}} \int_C \mathrm{e}^{\mathrm{i}(\omega t - [\mathbf{k_x}\cdot\mathbf{x}+\mathbf{k_y}\cdot\mathbf{y}])} d\omega\, d\mathbf{k_x}\, d\mathbf{k_y} \\ &= \frac{1}{(2\pi)^{1+n+m}} \int_{-\omega_{\max}}^{\omega_{\max}} \int_{|\mathbf{p_x}|\le p_{x,\max}} \int_{|\mathbf{p_y}|\le p_{y,\max}} \mathrm{e}^{\mathrm{i}\omega(t-[\mathbf{p_x}\cdot\mathbf{x}+\mathbf{p_y}\cdot\mathbf{y}])} |\omega|^{n+m} d\omega\, d\mathbf{p_x}\, d\mathbf{p_y} \end{aligned} \tag{21}$$

we can define a generalization of the Gaussian:

$$K_G(t,\mathbf{x},\mathbf{y}) = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\frac{|\mathbf{p}_x|^2}{2\sigma_{p_x}^2}} \mathrm{e}^{-\frac{|\mathbf{p_y}|^2}{2\sigma_{p_y}^2}} \mathrm{e}^{\mathrm{i}\omega(t-[\mathbf{p_x}\cdot\mathbf{x}+\mathbf{p_y}\cdot\mathbf{y}])} |\omega|^{n+m} d\omega\, d\mathbf{p_x}\, d\mathbf{p_y}, \tag{22}$$

which we refer to as a ***C-Gaussian function***. Here $\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}}$ and $\mathrm{e}^{-\frac{|\mathbf{p_x}|^2}{2\sigma_{p_x}^2}} \mathrm{e}^{-\frac{|\mathbf{p_y}|^2}{2\sigma_{p_y}^2}}$ controls the frequency and slowness decays, respectively. Consequently, we relax C-limited function definition and define ***effectively C-limited*** functions through the convolution integral

$$f_G(t,\mathbf{x},\mathbf{y}) = \int_{\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^m} f(\tau,\mathbf{x}',\mathbf{y}')\, K_G(t-\tau,\mathbf{x}-\mathbf{x}',\mathbf{y}-\mathbf{y}')\, d\tau\, d\mathbf{x}'\, d\mathbf{y}'. \tag{23}$$

Given samples of $f(t,\mathbf{x},\mathbf{y})$ for some $\{t_k,\mathbf{x}_k,\mathbf{y}_k\}_{k=1}^K \subset \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^m$, an effective C-limited function $g_G(t,\mathbf{x},\mathbf{y})$ that agrees with these sample points can be constructed by first discretizing the integral

(23)

$$g_G(t,\mathbf{x},\mathbf{y}) = \sum_{l=1}^{K} g_l K_G(t-t_l, \mathbf{x}-\mathbf{x}_l, \mathbf{y}-\mathbf{y}_l) \tag{24}$$

for some $g_l \in \mathbb{R}$ and then solving the following linear system for :

$$g_G(t_k,\mathbf{x}_k,\mathbf{y}_k) = f(t_k,\mathbf{x}_k,\mathbf{y}_k) = \sum_{l=1}^{K} g_l K_G(t_k-t_l, \mathbf{x}_k-\mathbf{x}_l, \mathbf{y}_k-\mathbf{y}_l). \tag{25}$$

In Table 3, we present primitive forms $K(t,\mathbf{x},\mathbf{y})$ of $K_G(t,\mathbf{x},\mathbf{y})$ for dimensions 1D, $(1+1)$D and $(1+2)$D. Extensions to $(1+m+n)$D, for $n,m \in \{0,1,2,3\}$ is provided in Appendix F.

3.4 Representing the sampled seismic data

Considering that seismic data is C-limited and is sampled discretely, construction of a C-limited function that agrees with the data requires solving (20) for g_l . In this regard, an explicit representation of $K_C(t,\mathbf{x},\mathbf{y})$ will make it easy to develop a solution for computation of g_l in the native acquisition geometry. However, like bandlimitedness, C-limitedness is a global property. Expressing C-limited functions using (19) does not reflect the temporal and spatial variability of the signal cone.

Furthermore, for poorly sampled seismic data, the signal-cone defined in Section 3.1 above does not provide a sufficient constraint with which we can unambiguously reconstruct the data. Further constraints are needed, and these can be provided by making additional assumptions about the nature of seismic data.

Many reconstruction algorithms assert that the data consists of a small number of locally planar events. This leads to algorithms that try to fit the observed data with a sparse set of plane-wave basis functions. We try to relax this assumption, and assert that seismic data can be modeled efficiently with a sparse set of either kernels of C-limited functions or C-Gaussians. These kernels relax the concept of a plane-wave.

3.4.1 Capturing temporal and spatial variability of signal cone

In order to capture temporal and spatial variability of the signal cone, we consider the representation of the following form

$$d(t,\mathbf{s},\mathbf{r}) = \sum_m d_m K_C \begin{pmatrix} \rho_{\omega,m}\left([t-t_m] - (\mathbf{p}_{\mathbf{s},m}\cdot[\mathbf{s}-\mathbf{x}_m] + \mathbf{p}_{\mathbf{r},m}\cdot[\mathbf{r}-\mathbf{y}_m])\right), \\ \rho_{k_\mathbf{s},m}[\mathbf{s}-\mathbf{x}_m], \\ \rho_{k_\mathbf{r},m}[\mathbf{r}-\mathbf{y}_m] \end{pmatrix}, \tag{26}$$

which captures the variability of frequency and slowness content with respect to temporal and spatial translations. The parameter $\rho_{\omega,m}\omega_{\max}$ controls the maximum frequency centered around zero and $p_{\mathbf{s},m} = p_{\max}\rho^{-1}_{\omega,m}\rho_{k_\mathbf{s},m}$ and $p_{\mathbf{r},m} = p_{\max}\rho^{-1}_{\omega,m}\rho_{k_\mathbf{r},m}$ controls slownesses centered around $\mathbf{p}_{\mathbf{s},m}$ and $\mathbf{p}_{\mathbf{r},m}$, respectively. For $\mathbf{p}_{\mathbf{s},m} = \mathbf{p}_{\mathbf{r},m} = 0$ and $\rho_{\omega,m} = \rho_{k_\mathbf{s},m} = \rho_{k_\mathbf{r},m} = 1$, (26) simplifies to (19). It should be noted with (26) one can also capture more than one singal cone for at a given temporal and spatial translation (see Figure 1).

(a) Cross sections of absolute value of the Fourier transform of the filter $K_F(t,\mathbf{x})$

m	t_m	x_m	d_m	$\sigma_{\omega,m}$	$\sigma_{p,m}$	$\mathbf{p}_l = [p_{x_1,m}, p_{x_2,m}]$
1	0	0	1	.9	.4	$[0.25, 0]$
2	0	0	-1	.9	.2	$[0.25, 0]$
3	0	0	-1	.5	.4	$[0.25, 0]$
4	0	0	1	.5	.2	$[0.25, 0]$
5	0	0	1	1	.25	$[0, 0.75]$
6	0	0	-1	.2	.25	$[0, 0.75]$

(b) Parameters used to generate $K_F(t,\mathbf{x})$.

Figure 1: Absolute value of the Fourier transform of a filter $K_F(t,\mathbf{x})$ obtained by weighted summation of $(1+2)$D kernel for $\omega_0 = p_{\max} = 1$: $K_F(t,\mathbf{x}) = \sum_i w_i \sigma^3_{\omega,i}\sigma^2_{p,i} K_C(\sigma_{\omega,i}[t - \mathbf{p}_i\cdot\mathbf{x}], \sigma_{\omega,i}\sigma_{p,i}\mathbf{x})$.

Alternatively, one can relax the C-limited constraint in (26) with C-Gaussian constraint and use

the following representation

$$d(t,\mathbf{s},\mathbf{r}) = \sum_m d_m K_G \begin{pmatrix} \rho_{\omega,m}\left([t-t_m] - \left(\mathbf{p}_{\mathbf{s},m}\cdot[\mathbf{s}-\mathbf{x}_m] + \mathbf{p}_{\mathbf{r},m}\cdot[\mathbf{r}-\mathbf{y}_m]\right)\right), \\ \rho_{k_\mathbf{s},m}[\mathbf{s}-\mathbf{x}_m], \\ \rho_{k_\mathbf{r},m}[\mathbf{r}-\mathbf{y}_m] \end{pmatrix}, \tag{27}$$

which captures leaky band-pass filtering as well as leaky fan filtering where the leaks are governed by the decay rate of the Gaussian functions in frequency and slownesses. The parameter $\rho_{\omega,m}$ controls the frequency decay rate around zero relative to $\sigma_{\omega,m}$ and $\rho_{p_\mathbf{s},m} = \rho_{\omega,m}^{-1}\rho_{k_\mathbf{s},m}$ and $\rho_{p_\mathbf{r},m} = \rho_{\omega,m}^{-1}\rho_{k_\mathbf{r},m}$ control slowness decay rates centered around $\mathbf{p}_{\mathbf{s},m}$ and $\mathbf{p}_{\mathbf{r},m}$, respectively, relative to $\sigma_p = \sigma_\omega^{-1}\sigma_k$. In this regard, $\rho_{\omega,m}$, $\rho_{k_\mathbf{s},m}$ and $\rho_{k_\mathbf{r},m}$ used in (27) are analogous to the ones used in (26). Consequently, maximum frequency and slowness, $\omega_{\max}$ and $p_{\max}$, are analogous to frequency and slowness decay rates, σ_ω and σ_p, respectively.

3.4.2 Fitting seismic data with modified variable projection

Given the data $d(t,\mathbf{s},\mathbf{r})$, using either (26) or (27), its approximation $\hat{d}_{(\mathcal{D},\mathcal{P})}(t,\mathbf{s},\mathbf{r})$ can be captured by the following model

$$\hat{d}_{(\mathcal{D},\mathcal{P})}(t,\mathbf{s},\mathbf{r}) = \sum_{m=1}^{M} d_m K\left(\tau(t,P_m), \mathbf{x}(t,P_m), \mathbf{y}(t,P_m)\right) \tag{28}$$

with

$$\tau(t,P_m) = \rho_{\omega,m}\left([t-t_m] - \left(\mathbf{p}_{\mathbf{s},m}\cdot[\mathbf{s}-\mathbf{x}_m] + \mathbf{p}_{\mathbf{r},m}\cdot[\mathbf{r}-\mathbf{y}_m]\right)\right), \tag{29}$$

$$\mathbf{x}(\mathbf{s},P_m) = \rho_{k_\mathbf{s},m}[\mathbf{s}-\mathbf{x}_m] = \rho_{\omega,m}\rho_{p_\mathbf{s},m}[\mathbf{s}-\mathbf{x}_m] \tag{30}$$

$$\mathbf{y}(\mathbf{r},P_m) = \rho_{k_\mathbf{r},m}[\mathbf{r}-\mathbf{y}_m] = \rho_{\omega,m}\rho_{p_\mathbf{r},m}[\mathbf{r}-\mathbf{y}_m] \tag{31}$$

The linear and nonlinear parameter sets $\mathcal{D}$ and $\mathcal{P}$ are defined by

$$\mathcal{D} = \{d_m\}_{m=1,\ldots,M} \tag{32}$$

and

$$\mathcal{P} = \left\{P_m = (t_m, \mathbf{x}_m, \mathbf{y}_m, \mathbf{p}_{\mathbf{s},m}, \mathbf{p}_{\mathbf{r},m}, \rho_{\omega,m}, \rho_{p_\mathbf{s},m}, \rho_{p_\mathbf{r},m}) \in \mathbb{R}^{4+2N_\mathbf{s}+2N_\mathbf{r}}\right\}_{m=1,\ldots,M}. \tag{33}$$

In (26) and (27), it is expected that

- the temporal and spatial translation parameters t_m, $\mathbf{x}_m$, $\mathbf{y}_m$ lie within a region of interest determined by the acquisition geometry;
- the slowness vectors $\mathbf{p}_{\mathbf{s},m}$ and $\mathbf{p}_{\mathbf{r},m}$ lie within the corresponding signal cone;
- Given $\left|\delta\mathbf{p}_{\mathbf{s}/\mathbf{r}}\right| = p_{\max,\mathbf{s}/\mathbf{r}}$, perturbations around the slowness vectors, $\mathbf{p}_{\mathbf{s},m} + \rho_{p_\mathbf{s},m}\delta\mathbf{p}_\mathbf{s}$ and $\mathbf{p}_{\mathbf{r},m} + \rho_{p_\mathbf{r},m}\delta\mathbf{p}_\mathbf{r}$, lie within the signal cone. This condition can be removed for effectively C-limited functions.

In addition to the bound constraints on the nonlinear parameters, one can introduce constraints on the linear parameters. For the aim of finding a short representation, we employ ℓ_1 constraint on the linear parameters. As a result, for discretely sampled data we consider the following two minimization problems.

Problem 1. Find a C-limited approximation of the sampled data by

$$\text{minimize } J\left(\mathcal{D},\mathcal{P}\right) = \frac{1}{2}\sum_{n_t,n_\mathbf{s},n_\mathbf{r}=1}^{N_t,N_\mathbf{s},N_\mathbf{r}}\left[d\left(t_{n_t},\mathbf{s}_{n_\mathbf{s}},\mathbf{r}_{n_\mathbf{r}}\right) - \hat{d}_{(\mathcal{D},\mathcal{P})}\left(t_{n_t},\mathbf{s}_{n_\mathbf{s}},\mathbf{r}_{n_\mathbf{r}}\right)\right]^2 \tag{34}$$

$$\text{subject to } \left\|\mathcal{D}\right\|_1 \leq \lambda_\mathcal{D}$$

$$0 \leq \rho_{p_\mathbf{s},m}, \rho_{p_\mathbf{r},m} \leq 1 \tag{35}$$

$$\begin{aligned}&\mathbf{p}_{\mathbf{s},m} + \rho_{p_\mathbf{s},m}\delta\mathbf{p}_\mathbf{s} \in C_\mathbf{s} \text{ for } \left|\delta\mathbf{p}_\mathbf{s}\right| = p_{\max,\mathbf{s}}\\ &\mathbf{p}_{\mathbf{r},m} + \rho_{p_\mathbf{r},m}\delta\mathbf{p}_\mathbf{r} \in C_\mathbf{r} \text{ for } \left|\delta\mathbf{p}_\mathbf{r}\right| = p_{\max,\mathbf{r}}\end{aligned} \tag{36}$$

$$0 \leq \rho_\omega < 1 \tag{37}$$

$$\begin{aligned}&t_m \in [-T_{\min}, T_{\max}]\\ &(\mathbf{x}_m, \mathbf{y}_m) \in \Omega_\mathbf{s} \times \Omega_\mathbf{r}\end{aligned} \tag{38}$$

for some desired signal cones $C_\mathbf{s}$ and $C_\mathbf{r}$, slownesses $p_{\max,\mathbf{s}}$ and $p_{\max,\mathbf{r}}$, recording time $[-T_{\min}, T_{\max}]$ and acquisition domain $\Omega_\mathbf{s} \times \Omega_\mathbf{r} \subset \mathbb{R}^{N_\mathbf{s}} \times \mathbb{R}^{N_\mathbf{r}}$.

Here n_t, $n_\mathbf{s}$ and $n_\mathbf{r}$ denote the time, source and receiver sample number, N_t, $N_\mathbf{s}$ and $N_\mathbf{r}$ total number of time, source and receiver samples, respectively.

Algorithm 1 Modified variable projection

1. Initialize $\mathcal{P}^0$
2. For $l > 0$,
   (a) Compute $\mathcal{D}^{l+1} = \text{argmin}_{\mathcal{D}} \left\{ \frac{1}{2} \left\| d - K\left(\mathcal{P}^l\right)\mathcal{D} \right\|_2^2 \right\}$ subject to $\|\mathcal{D}\|_1 \leq \lambda_{\mathcal{D}}$.
   (b) Compute $\mathcal{P}^{l+1} = \mathcal{P}^l - t_l P^{\perp}_{K(\mathcal{P}^l)} \left[\partial_{\mathcal{P}} K\left(\mathcal{P}\right)\right]_{\mathcal{P}=\mathcal{P}^l} \mathcal{D}^{l+1}$ subject to constraints:
      i. (35), (36), (37) and (38) for C-limited approximation,
      ii. (40), (41), (42) and (38) for effectively C-limited approximation,

   until stopping criteria is met.

Problem 2. Find an effectively C-limited approximation of the sampled data by

$$\text{minimize } J\left(\mathcal{D},\mathcal{P}\right) = \frac{1}{2} \sum_{n_t,n_{\mathbf{s}},n_{\mathbf{r}}} \left[d\left(t_{n_t},\mathbf{s}_{n_{\mathbf{s}}},\mathbf{r}_{n_{\mathbf{r}}}\right) - \hat{d}_{(\mathcal{D},\mathcal{P})}\left(t_{n_t},\mathbf{s}_{n_{\mathbf{s}}},\mathbf{r}_{n_{\mathbf{r}}}\right) \right]^2 \tag{39}$$

$$\text{subject to } \|\mathcal{D}\|_1 \leq \lambda_{\mathcal{D}}$$

$$\begin{aligned} &\mathbf{p}_{\mathbf{s},m} \in C_{\mathbf{s}} \\ &\mathbf{p}_{\mathbf{r},m} \in C_{\mathbf{r}} \end{aligned} \tag{40}$$

$$0 \leq \rho_{p_{\mathbf{s}},m}, \rho_{p_{\mathbf{r}},m} < \infty \tag{41}$$

$$0 \leq \rho_{\omega} < \infty \tag{42}$$

$$\begin{aligned} &t_m \in [-T_{\min}, T_{\max}] \\ &(\mathbf{x}_m, \mathbf{y}_m) \in \Omega_{\mathbf{s}} \times \Omega_{\mathbf{r}} \end{aligned}$$

Note that the bound constraints (35), (36) and (37) required for C-limited approximation are relaxed for the effectively C-limited approximation.

The nonlinear least square problems (34) and (39) can be tackled using fixed point methods. In this regard, we employ a modification of variable projection method summarized in Algorithm 1. In computation of iterations of linearized parameters (Step 2a of Algorithm 1), we used SPGL1 van den Berg and Friedlander (2008), a modification of limited memory-BFGS (see Chapter 9 of Nocedal and Wright (1999)) which incorporates ℓ_1 constraints Schmidt et al. (2009).

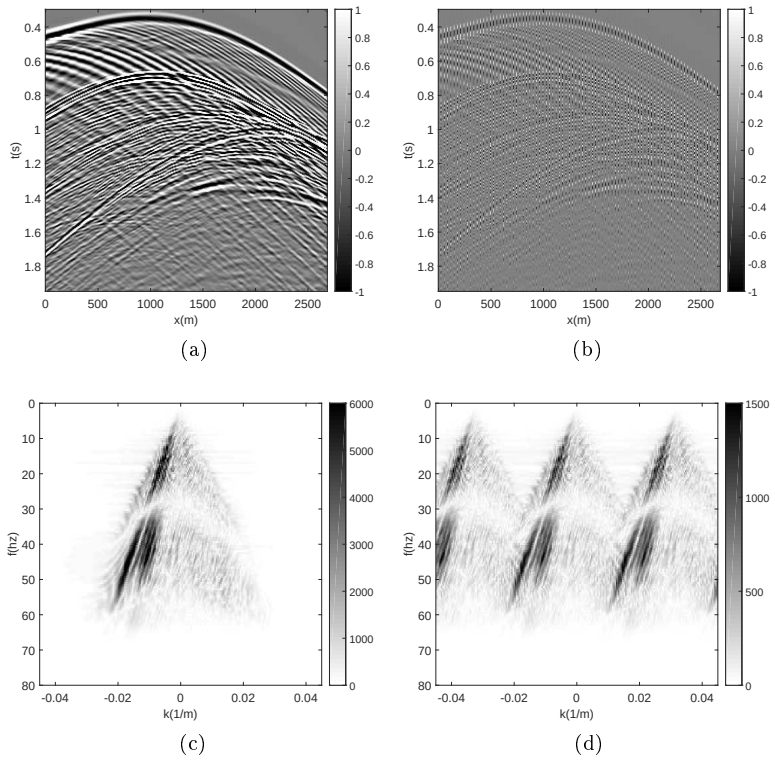


Figure 2: Single X-line extracted from a synthetic 3D seismic data acquisition generated on SEAM model. (a,c) Ground truth and (b,d) 6-fold spatially subsampled data in time-space and frequency-wavenumber domain respectively.

4 Numerical example in $(1+1)$D

4.1 Data

In order to evaluate the proposed method, we used part of a single shot gather from the SEAM data set Fehler and Keliher (2011) , where $N_t = 275$ and $N_x = 430$ are the number of time and space samples with a sampling interval of 6ms and 6.25m, respectively. Figures 2 (a) and (c) show the original sampled seismic data and corresponding absolute value of the Fourier transform in frequency-wavenumber (f-k) domain, respectively. For our numerical test, we spatially subsampled this data by a factor of six. The subsampled data has a spatial sampling rate of 37.5m. Figures 2 (b) and (d) show the subsampled data and corresponding absolute value of the Fourier transform in frequency-wavenumber (f-k) domain, respectively.

4.2 Kernels

By looking at the temporal Fourier transform of the subsampled data, which is unaliased, one can infer about the minimum and maximum frequency content of the data. Similarly, looking at the low frequency content of the subsampled data, one can infer about the minimum and maximum slownesses contained in the data.. This defines a trapezoidal region which agrees with the support of the Fourier transform of the original sampled data. This region can be characterized by the kernel

$$K(t,x) = K_C(t,x) - \rho_\omega^2 K_C(\rho_\omega t, \rho_\omega x), \tag{43}$$

where $\rho_\omega = \omega_{\max}^{-1}\omega_{\min}$ is the ratio between the minimum and maximum frequency content of the data. After substitution of this kernel into (28), Problem 1 leads to a C-limited approximation of the data. Similarly, Problem 2 leads to effectively C-limited approximation when (43) is effectively approximated by

$$K(t,x) = \partial_t^2 K_G(t,x). \tag{44}$$

for some suitable choice of σ_ω and σ_p. See Figure 3 for an example of approximating (43) using (44). Besides effectively capturing the support of (43), (44) is a multivariate generalization of Ricker wavelet, whose practical success was supported with the recent theoretical study Gholamy and Kreinovich (2014).

Remark. In consideration of the basis pursuit methods shearlets and curvelets, (43) provides an explicit way to compute the ideal response of a tile in the 2D shearlet or discrete curvelet decomposition Guo and Labate (2007); Kutyniok and Labate (2012); Candes et al. (2006); Ying et al. (2005). In implementation of shearlet and curvelets, these tiles are usually smoothed to improve the decay properties which we analytically capture with (44). However, instead of tiling in the temporal direction, one can use higher order derivatives of K_G,$\{\partial_t^n K_G(t,x)\}_{n=1}^{\infty}$ which constitutes a basis, for an alternative construction for multiresolution decomposition. $(1+1+1)$D analogues of K_C and K_G are analogous to 3D shearlets or discrete curvelets. On the other hand, in $(1+2)$D, neither K_C nor K_G corresponds to the tiles of 3D shearlets or discrete curvelets due to their circular symmetries in space. Basis elements in the existing implementations of shearlets and curvelets are either defined interms of discretization of the (smoothed) Fourier tiles in the Fourier domain or the discrete inverse Fourier transform of these discretized tiles in the space-time domain. Unlike shearlets and discrete curvelets basis, K_C and K_G have analytic representation in the space time domain which enables them to be

easily utilized in variational methods. Compared to the Gaussian wave packets used in Andersson et al. (2012), whose Fourier transform is concentrated within a translated, rotated and dilated elliptical region in frequency-wavenumber domain, Fourier transform of K_G is concentrated within a slanted conic region, which corresponds to a elliptical region in frequency-slowness domain. Oscillatory nature of the aforementioned Gaussian wave packets are controlled by a harmonic planewave. On the other hand higher order oscillations can be captured by taking temporal derivatives of K_G. The resulting functions will be a generalization of Herminate functions. Like Hermite functions, these functions can be used to form an orthonormal basis with respect to the temporal variables. In this work, we only considered second order derivative of K_G due to its relationship to Ricker wavelet and leave exploration of practical benefits of higher order derivatives to a future study.

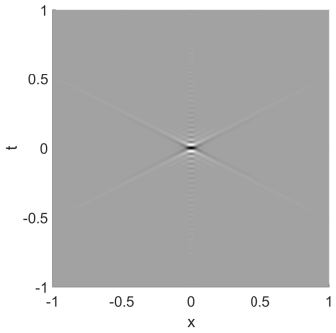
4.3 Method

4.3.1 Parameter estimation and interpolation

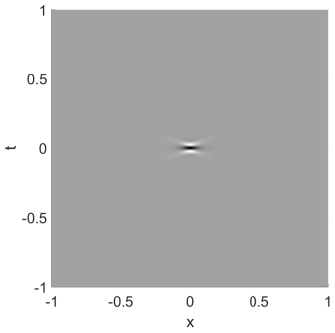
We subdivided the subsampled data into 12 overlapping windows using a partition of unity where each window size is 780ms $\times$ 812.5m with 180ms temporal and 187m spatial overlap. We refer to each of these windows as a region of interest. Patching the regions of interest is obtained by solving a linear problem after non-linear parameters are estimated. We performed the proposed method on each of these regions of interest to estimate linear and nonlinear parameters. Once the parameters are obtained, we evaluate (28) on the same grid as of the original data's to perform interpolation.

Due to non-convex nature of the objective function, after numerous tests we conclude that initial parameter estimation shall be performed in small windows. Based on this experience, for initialization, within each region of interest, we estimated the nonlinear parameters $P_m = \{t_m, x_m, p_m, \rho_{\omega,m}, \rho_{p,m}\}$ on smaller windows of 60ms $\times$ 375m centered around each sample location $(t_{n_t}, x_{n_x})_{n_t,n_x=1}^{N_t,\lfloor N_x/6 \rfloor}$ on the subsampled grid using Algorithm 1. For each small window we initialized with one $\rho_{\omega/p,m}$ and 10 p_m uniformly distributed within the slowness range of the signal cone. Once the parameters are initilized, we perform non-linear optimization in each of the regions of interest.

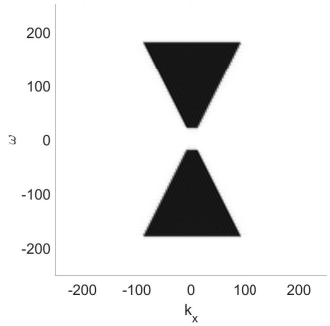
Once the nonlinear parameters $\mathcal{P} = \{P_m\}$ are estimated, we perform interpolation within each of the 12 windows by first solving the convex constrained optimization problem $\operatorname{argmin}_{\mathcal{D}} \left\{ \frac{1}{2} \left\| d - K(\mathcal{P})\mathcal{D} \right\|_2^2 \right\}$ for $\|\mathcal{D}\|_1 \leq \lambda_{\mathcal{D}}$ followed by evaluation $\hat{d} = K(\mathcal{P})\mathcal{D}$, which is the interpolation part. Our method is summarized in Algorithm 2. The computational cost of Algorithm 2 is combination of many local non-convex optimization problems to estimate the non-linear parameters (Step 2 of Algorithm 2) and a global convex optimization problem to estimate the linear parameters (Step 3 of Algorithm 2).



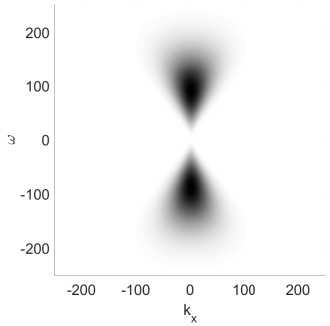
(a) K obtained from (43)



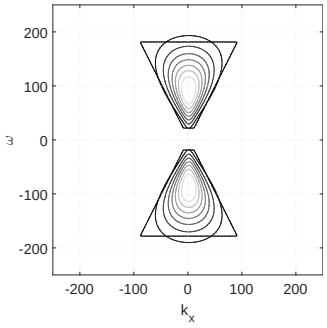
(b) K obtained from (44)



(c) Fourier transform of K in Figure 3(a).



(d) Fourier transform of K o in Figure 3(b).



(e) Contour plots of Figures 3 (c) and 3 (d)

Figure 3: Plots of $K(t,x) = K_C(t,x) - K_C(t/8, x/8)/64$, for $\omega_{\max} = 170$ and $p_{\max} = .5$, and $K(t,x) = \partial_t^2 K_G(t,x)$, for $\sigma_\omega = 40$ and $\sigma_p = .25$, in time-space in (1+1)D (a,c) and absolute value of their Fourier transforms in frequency-wavenumber (b,d) domain, respectively (see Appendix A for definitions of K_C and K_G). Overlaid contour plots of absolute value of the Fourier transform of K_C and K_G in the frequency-wavenumber domain (e) shows that K_C can be effectively approximated by K_G.

Algorithm 2 Representation of seismic data by (28)

Given the sampled seismic data $d\left(t_{n_t}, \mathbf{s}, x_{n_x}\right)_{n_t,n_x=1}^{N_t,N_x}$ for a fixed source and regions of interest :

Initialization

1. Choose a window size smaller than region of interest.
2. For each $\left(t_{n_t}, x_{n_x}\right)_{n_t,n_x=1}^{N_t,N_x}$, estimate $\left\{t_{n_t}, x_{n_x}, p_{0,n_t,n_x,n_p}, \rho_{\omega,n_t,n_x}, \rho_{p,n_t,n_x}\right\}_{n_t,n_x,n_p=1}^{N_t,N_x,N_p}$ within the small window centered at $\left(t_{n_t}, x_{n_x}\right)$ using Algorithm 1.

Parameter estimation

1. For each $\left(t_{n_t}, x_{n_x}\right)_{n_t,n_x=1}^{N_t,N_x}$ within the regions of interest initialize $\left\{p_{0,n_t,n_x,n_p}, \rho_{\omega,n_t,n_x}, \rho_{p,n_t,n_x}\right\}_{n_t,n_x,n_p=1}^{N_t,N_x,N_p}$. This means we initialize decomposition with $N_t N_x N_p$ terms.
2. Compute $P_m = \left\{t_m, x_m, p_m, \rho_{\omega,m}, \rho_{p,m}\right\}_{m=1}^{M}$, where $M \leq N_t N_x N_p$ within each region of interest using Algorithm 1. Note that $\left(t_m, x_m\right)$ does not need to be in $\left\{\left(t_{n_t}, x_{n_x}\right)_{n_t,n_x=1}^{N_t,N_x}\right\}$
3. Patch the regions of interest by solving the linear least squares problem $\mathcal{D} = \operatorname{argmin}_{\mathcal{D}} \left\{\frac{1}{2}\left\|d - K\left(\mathcal{P}\right)\mathcal{D}\right\|_2^2\right\}$ subject to $\|\mathcal{D}\|_1 \leq \lambda_{\mathcal{D}}$ for regions of interests.

Interpolation

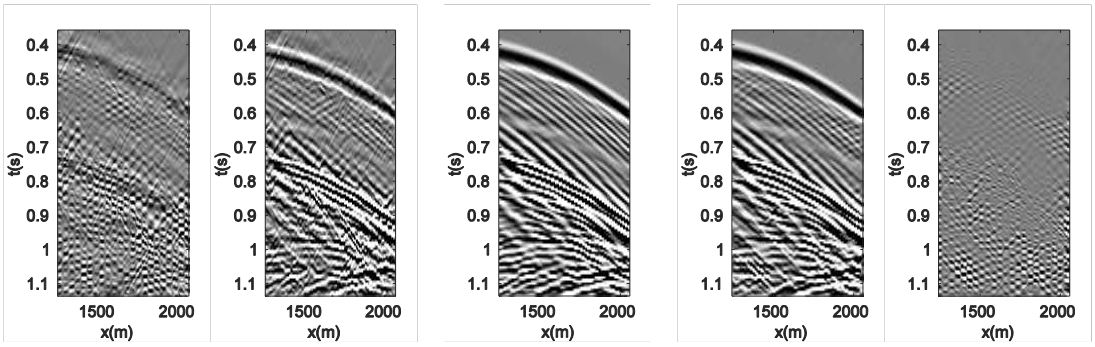
1. Evaluate $\hat{d}_{(\mathcal{D},\mathcal{P})}\left(t, \mathbf{s}, \mathbf{r}\right)$ at desired locations using

4.3.2 Choosing the basis

For one of the 12 windows, the interpolation results using (43) and (44) are presented in Figure 4. The interpolation obtained using (43) presents more artifacts when compared to the one obtained using (44). Especially in the region where the signal is zero. This is due to the slow decaying property of (43) combined with its oscillatory behavior. In order to capture the zero part of the data, temporal and spatial translates of (43) come with alternating signs, which indicates presence of cancellations in the representation as well as inefficiency of (43) to approximate the seismic signal. On the other hand, because (44) decays temporally and spatially faster than (43), it has better localization properties. For the translates of (44) falling into zero part of the signal, the corresponding linear coefficients are either zero or close to zero. Therefore, when compared to (43), (44) captures the zero part of the signal better and efficiently. This is why we decided to use (44) for interpolating the whole data.

4.4 Numerical results

We compare the proposed interpolation algorithm with sparsity promoting curvelet decomposition (CD) Herrmann et al. (2008), sparsity promoting plane-wave decomposition (PWD) algorithms Rickett



(a) Interpolated using (43) (right) and corresponding error (left)

(b) Original

(c) Interpolated using (44) (left) and corresponding error (right)

Figure 4: Comparison of the interpolation using (43) and (44) and corresponding difference errors.

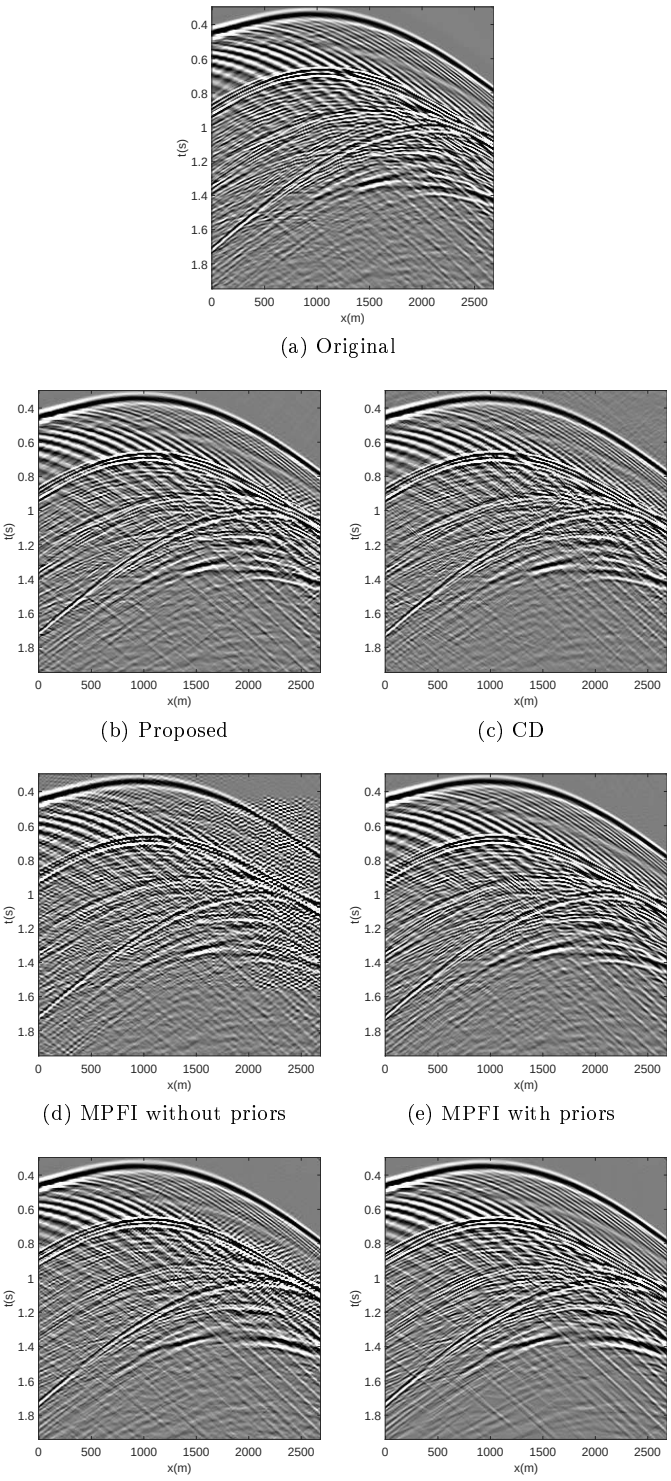
(2014) and matching pursuit Fourier interpolator (MPFI) Mallat and Zhang (1993). MPFI is performed on moving windows of the same size as the proposed method, 12 overlapping windows where each window size is 780ms $\times$ 812.5m with 180ms temporal and 187mspatial overlap. PWD was parametrized using 65 plane wave basis elements, uniformly distributed within the signal cone, for each time sample.

In case of MPFI and PWD we considered two versions of these methods, namely with and without priors. The priors are chosen to favor slownesses present in the low to mid frequency content of the data. Figures 5 and 6 show the interpolated results using all algorithms in the time-space and frequency-wavenumber domain. We present the difference between the original and the interpolations obtained by proposed method, CD, MPFI and PWD with priors in Figure 7. Based on visual inspection of results, proposed method performs better when compared to the MPFI and PWD without priors and comparable to CD, MPFI. While PWD with priors performs better than PWD without priors and captures kinematics as good as the proposed method, CD and MPFI, it is not able to capture the amplitudes as accurately as other methods. We associate this with the large spatial extent of the planewave basis used for creating a smaller dictionary for PWD, hence fewer degree of freedom, compared to the other methods.

All space-time and wavenumber-frequency plots use the same color scale as in Figure 2a and 2c, respectively.

5 Conclusions and discussion

In this work, we presented a new method to decompose seismic data into a model based dictionary where the dictionary elements are motivated by the physics of the wave equation and are analytically



(a) Original

(b) Proposed

(c) CD

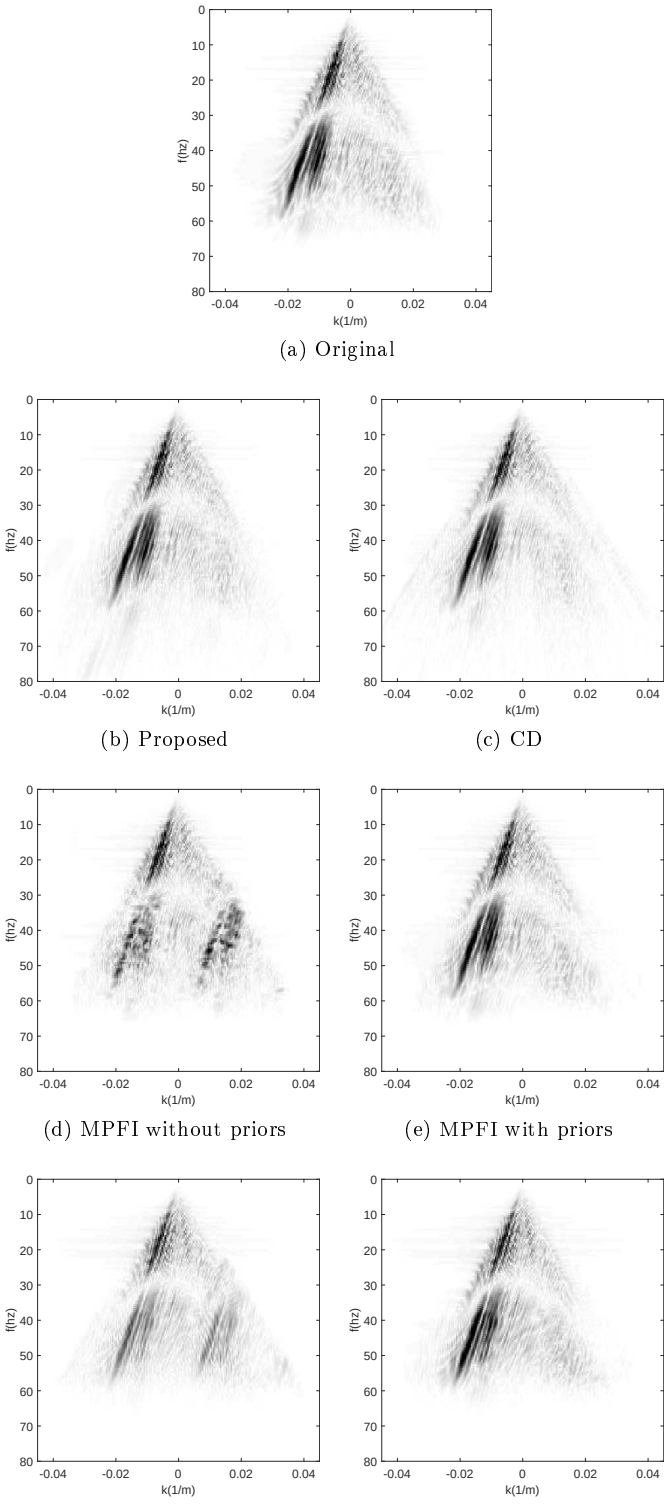
(d) MPFI without priors

(e) MPFI with priors

(f) PWD without priors

(g) PWD with priors

Figure 5: Interpolation results obtained using proposed method sparsity promoted curvelet decomposition, matching pursuit Fourier interpolator (MPFI) with and without priors, and sparsity promoting plane-wave decomposition (PWD) with and without priors.



(a) Original

(b) Proposed

(c) CD

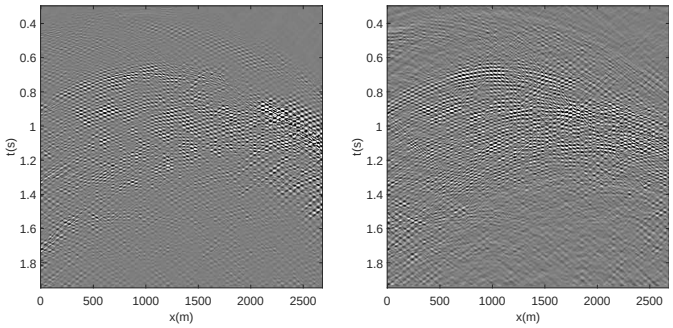
(d) MPFI without priors

(e) MPFI with priors

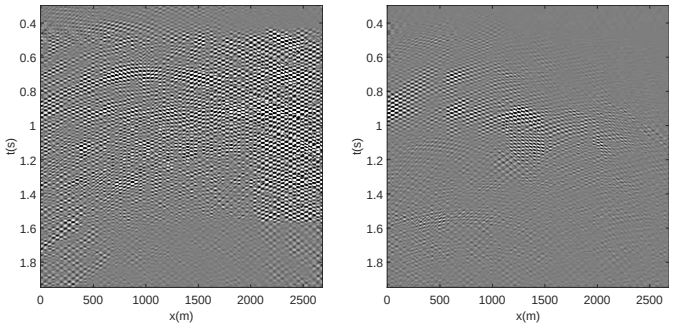
(f) PWD without priors

(g) PWD with priors

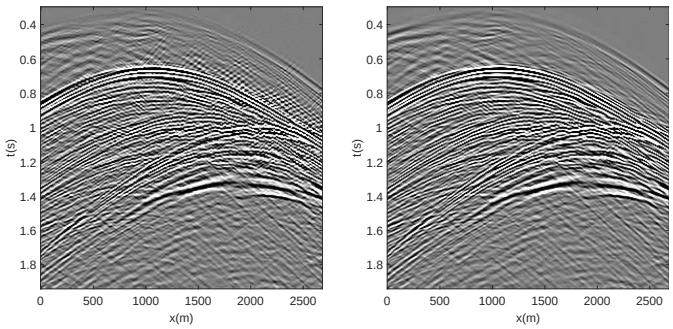
Figure 6: Absolute value of the Fourier transform of interpolation results obtained using proposed method, sparsity promoted curvelet decomposition, matching pursuit Fourier interpolator (MPFI) with and without priors, and sparsity promoting plane-wave decomposition (PWD) with and without priors.



(a) Proposed method (b) CD



(c) MPFI without priors (d) MPFI with priors



(e) PWD without priors (f) PWD with priors

Figure 7: Difference between the original and the (a) proposed method, (b) CD, (c,e) MPFI and PWD without priors, (d,f) MPFI and PWD with priors methods in time-space domain.

defined. Starting with the wave equation and associated signal cone, we formalized the concepts of C-limited and effectively C-limited functions. Identifying seismic data as a C-limited function, we provide a variational method to decompose it as a sum of kernels for C-limited functions, K_C, or C-Guassian function, K_G. Each term in the sum is made of a linear parameter multiplied by a modification of K_C or K_G via nonlinear parameters. We estimate the linear and nonlinear parameters, hence learn the dictionary, from the provided sampled seismic data using a fixed point method, namely a modification of variable projection method. We provide analytic expressions for both K_C and K_G, or their approximations. These approximations enable development of fast evaluation of these (approximate) kernels and their derivatives which are essential for computation of the Jacobian required for efficient non-convex optimization.

We demonstrated the potential of the proposed method through numerical examples. Based on numerical examples, we conclude that, by relaxing the C-limit requirements, K_G provides a better basis for representation of the seismic data as well as a less constrained optimization problem when compared to K_C. K_G decays faster than K_C which is essential for capturing the causal nature, i.e. no signal before first arrival, of the seismic data during interpolation.

Our numerical examples also demonstrated that we were able to approximate seismic data from its subsampled version. The results obtained were comparable to the ones that are obtained using sparsity promoting curvelet decomposition (SPCD) or sparsity promoting plane wave decomposition (SPPWD) with priors or matching pursuit Fourier interpolators (MPFI) with priors. All SPCD, SPPWD and MPFI are global search type methods. The first two methods are basis pursuit type methods and the third one is a greedy method. Although our results demonstrate that a variational method, like the one proposed here, can produce results as good as the existing methods, further studies are needed to explore the limitations and potentials of variational methods. This will also require detailed comparative studies against the existing methods. However, one potential advantage of a variational type method over the global search methods is they can be extended to libraries with more number of nonlinear parameters with reasonable computational cost. Introduction of curvature and other parameters into the basis functions are currently under investigation and will be reported in the near future.

Acknowledgments

We thank Zhimei Yan and Michel Schonewille for the MPFI results, Michel Schonewille for valuable feedback during the review of the manuscript, Thomas Pock for pointing out the literature on proximal

algorithms, Carola Schönlieb, Kemal Özdemir and Ali Özbek for illuminating discussions. We also thank Kemal Özdemir for inspiring the terminology "effectively C-limited", which is a generalization of effective bandlimit concept associated with Gabor's "effective frequency width" definition (see equation (1.22) in Gabor (1946)).

References

Milton Abramowitz and Irene A Stegun, 1965. *Handbook of Mathematical Function with Formulas, Graphs, and Mathematical Tables*. Dover Publications.

Fredrik Andersson, Marcus Carlsson, and Luis Tenorio, 2012. On the representation of functions with gaussian wave packets. *Journal of Fourier Analysis and Applications*, 18(1):146–181.

Heinz H Bauschke, Regina S Burachik, Patrick L Combettes, Veit Elser, D Russell Luke, and Henry Wolkowicz, 2011. Fixed-point algorithms for inverse problems in science and engineering. *Springer Optimization and Its Applications*.

Simon Beckouche and Jianwei Ma, 2014. Simultaneous dictionary learning and denoising for seismic data. *Geophysics*, 79(3):A27–A31.

Thomas Blumensath, 2012. Accelerated iterative hard thresholding. *Signal Processing*, 92(3):752–756.

Thomas Blumensath and Mike E Davies, 2008. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle, 2014. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494.

Jian-Feng Cai, Hui Ji, Zuowei Shen, and Gui-Bo Ye, 2014. Data-driven tight frame construction and image denoising. *Applied and Computational Harmonic Analysis*, 37(1):89–105.

Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying, 2006. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899.

Emmanuel J Candès and Carlos Fernandez-Granda, 2013. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254.

Emmanuel J Candès and Carlos Fernandez-Granda, 2014. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956.

Emmanuel J Candès and Michael B Wakin, 2008. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30.

Emmanuel J Candès, Yonina C Eldar, Deanna Needell, and Paige Randall, 2011. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1): 59–73.

Volkan Cevher, Stephen Becker, and Mark Schmidt, 2014. Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics. *IEEE Signal Processing Magazine*, 31 (5):32–43.

Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre Weiss, 2015. A projection algorithm on measures sets. *arXiv preprint arXiv:1509.00229*.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders, 1998. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61.

David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck, 2012. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 58(2):1094–1121.

Vincent Duval and Gabriel Peyré, 2015. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, pages 1–41.

Chaitanya Ekanadham, Daniel Tranchina, and Eero P Simoncelli, 2011. Recovery of sparse translation-invariant signals with continuous basis pursuit. *IEEE Transactions on Signal Processing*, 59(10): 4735–4744.

Jianqing Fan, Fang Han, and Han Liu, 2014. Challenges of big data analysis. *National science review*, 1(2):293–314.

Michael Fehler and P Joseph Keliher, 2011. *SEAM Phase 1: Challenges of Subsalt Imaging in Tertiary Basins, with Emphasis on Deepwater Gulf of Mexico*. Society of Exploration Geophysicists Tulsa.

Massimo Fornasier and Holger Rauhut, 2008. Iterative thresholding algorithms. *Applied and Computational Harmonic Analysis*, 25(2):187–208.

Dennis Gabor, 1946. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441.

David M Gay and Linda Kaufman, 1991. Tradeoffs in algorithms for separable nonlinear least squares. In *Proc. of the 13th World Congress on Computational and Applied Mathematics, IMACS*, volume 91, pages 157–158.

Afshin Gholamy and Vladik Kreinovich, 2014. Why ricker wavelets are successful in processing seismic data: Towards a theoretical explanation. In *Computational Intelligence for Engineering Solutions (CIES), 2014 IEEE Symposium on*, pages 11–16. IEEE.

Gene Golub and Victor Pereyra, 2003. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, 19(2):R1.

Gene H Golub and Victor Pereyra, 1973. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis*, 10(2):413–432.

I. S. Gradshteyn and I. M. Ryzhik, 2007. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition. ISBN 978-0-12-373637-6; 0-12-373637-4. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

Kanghui Guo and Demetrio Labate, 2007. Optimally sparse multidimensional representation using shearlets. *SIAM journal on mathematical analysis*, 39(1):298–318.

Felix J. Herrmann, Deli Wang, Gilles Hennenfent, and Peyman P. Moghaddam, 03 2008. Curvelet-based seismic data processing: a multiscale and nonlinear approach. *Geophysics*, 73(1):A1–A5. doi: 10.1190/1.2799517. URL `https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2008/herrmann08GEOcbs/herrmann0`

Geoffrey E Hinton and Ruslan R Salakhutdinov, 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

German Hoecht, Patrice Ricarte, Steffen Bergler, and Evgeny Landa, 2009. Operator-oriented crs interpolation. *Geophysical Prospecting*, 57(6):957–979.

Linda Kaufman, 1975. A variable projection method for solving separable nonlinear least squares problems. *BIT Numerical Mathematics*, 15(1):49–57.

Karin C Knudson, Jacob Yates, Alexander Huk, and Jonathan W Pillow, 2014. Inferring sparse representations of continuous signals with continuous orthogonal matching pursuit. In *Advances in neural information processing systems*, pages 1215–1223.

Gitta Kutyniok and Demetrio Labate, 2012. *Shearlets: Multiscale analysis for multivariate data*. Springer Science & Business Media.

Jingwei Liang, Jianwei Ma, and Xiaoqun Zhang, 2014. Seismic data restoration via data-driven tight frame. *Geophysics*, 79(3):V65–V74.

Stéphane G Mallat and Zhifeng Zhang, 1993. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415.

Katharine M Mullen, Mikas Vengris, and Ivo HM van Stokkum, 2007. Algorithms for separable nonlinear least squares with application to modelling time-resolved spectra. *Journal of Global Optimization*, 38(2):201–213.

Jorge Nocedal and Stephen J Wright, 1999. *Numerical optimization*. Springer New York, 2nd edition edition.

Neal Parikh and Stephen Boyd, 2013. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231.

Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad, 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE.

Peter Richtárik and Martin Takáč, 2016. Parallel coordinate descent methods for big data optimization. *Mathematical Programming*, 156(1-2):433–484.

James Rickett, 2014. Successes and challenges in 3D interpolation and deghosting of single-component marinestreamer data. *2014 SEG Annual Meeting*.

M. Schmidt, E. van den Berg, M. P. Friedlander, and K. Murphy, April 2009. Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. In D. van Dyk and M. Welling, editors, *Proceedings of The Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS) 2009*, volume 5, pages 456–463, Clearwater Beach, Florida.

David Slepian, 1964. Prolate spheroidal wave functions, fourier analysis and uncertainty - IV: extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal*, 43(6):3009–3057.

Joel Tropp, Anna C Gilbert, et al., 2007. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666.

E. van den Berg and M. P. Friedlander, 2008. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912. doi: 10.1137/080714488. URL `http://link.aip.org/link/?SCE/31/890`.

V. Červený, 2001. *Seismic Ray Theory*. Cambridge University Press.

Can Evren Yarman. Table of representation kernels for (signal) cone-limited functions. Internal communication, Schlumberger Cambridge Research, 2015.

Can Evren Yarman and Garret Flagg. A new way to calculate the sine integral function. Internal communication, Schlumberger Cambridge Research, 2014.

Can Evren Yarman and Garret Flagg, 2015. Generalization of padé approximation from rational functions to arbitrary analytic functions. *Mathematics of Computation*, 84(294):1835–1860.

Jui L Yen, 1956. On nonuniform sampling of bandwidth-limited signals. *Circuit Theory, IRE Transactions on*, 3(4):251–257.

Wotao Yin, Stanley Osher, Donald Goldfarb, and Jerome Darbon, 2008. Bregman iterative algorithms for l_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1 (1):143–168.

Lexing Ying, Laurent Demanet, and Emmanuel Candes, 2005. 3d discrete curvelet transform. In *Optics & Photonics 2005*, pages 591413–591413. International Society for Optics and Photonics.

Appendix A: Primitive kernels

Dimensions	Kernel
$1+0$	$K_C(t) = (2\pi)^{-1} 2\omega_{\max} \operatorname{sinc}(\omega_{\max} t)$
$1+1$	$K_C(t,x) = (2\pi)^{-2} \frac{2\omega_{\max}}{x} \left(\operatorname{cosinc}(\omega_{\max}[t+p_{\max}x]) - \operatorname{cosinc}(\omega_{\max}[t-p_{\max}x])\right)$
$1+2$	$K_C(t,\mathbf{x}) \approx (2\pi)^{-3} \frac{4\omega_{\max}}{r^2} \sum_{m=1}^{M} \frac{\alpha_m}{\gamma_m} \left(\operatorname{sinc}(\omega_{\max} t) - \frac{1}{2}\begin{bmatrix} \operatorname{sinc}(\omega_{\max}[t-\gamma_m p_{\max} r]) \\ +\operatorname{sinc}(\omega_{\max}[t+\gamma_m p_{\max} r]) \end{bmatrix}\right)$
$1+3$	$K_C(t,\mathbf{x}) = -(2\pi)^{-4} \frac{4\pi\omega_{\max}}{r} \partial_r \left\{\frac{1}{r}\left[\operatorname{cosinc}(\omega_{\max}[t+p_{\max}r]) - \operatorname{cosinc}(\omega_{\max}[t-p_{\max}r])\right]\right\}$
$1+1+1$	$K_C(t,x,y) = -(2\pi)^{-3} 2\frac{\omega_{\max}}{xy} \begin{pmatrix} \operatorname{sinc}(\omega_{\max}[t+p_{x,\max}x+p_{y,\max}y]) \\ -\operatorname{sinc}(\omega_{\max}[t+p_{x,\max}x-p_{y,\max}y]) \\ -\operatorname{sinc}(\omega_{\max}[t-p_{x,\max}x+p_{y,\max}y]) \\ +\operatorname{sinc}(\omega_{\max}[t-p_{x,\max}x-p_{y,\max}y]) \end{pmatrix}$
$1+1+2$	$K_C(t,\mathbf{x},y) \approx (2\pi)^{-4} \frac{2\omega_{\max}}{y\, r_x^2} \sum_{m_x=1}^{M} \frac{\alpha_{m_x}}{\gamma_{m_x}} \begin{pmatrix} \operatorname{cosinc}(\omega_{\max}[t+p_{y,\max}y]) \\ -\operatorname{cosinc}(\omega_{\max}[t-p_{y,\max}y]) \\ -\operatorname{cosinc}(\omega_{\max}[t+\gamma_{m_x} r_x p_{x,\max}+p_{y,\max}y]) \\ -\operatorname{cosinc}(\omega_{\max}[t-\gamma_{m_x} r_x p_{x,\max}+p_{y,\max}y]) \\ +\operatorname{cosinc}(\omega_{\max}[t+\gamma_{m_x} r_x p_{x,\max}-p_{y,\max}y]) \\ +\operatorname{cosinc}(\omega_{\max}[t-\gamma_{m_x} r_x p_{x,\max}-p_{y,\max}y]) \end{pmatrix}$
$1+1+3$	$K_C(t,\mathbf{x},y) = (2\pi)^{-5} \frac{16\pi\omega_{\max}}{r_x^3 y} \begin{bmatrix} f_1(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \omega_{\max}p_{y,\max}y) \\ -\omega_{\max}p_{x,\max}r_x\, f_2(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \omega_{\max}p_{y,\max}y) \end{bmatrix}$
$1+2+2$	$K_C(t,\mathbf{x},\mathbf{y}) \approx (2\pi)^{-5} \frac{4\omega_{\max}}{r_x^2 r_y^2} \sum_{m_x=1}^{M}\sum_{m_y=1}^{M} \times \frac{\alpha_{m_x}}{\gamma_{m_x}}\frac{\alpha_{m_y}}{\gamma_{m_y}} \begin{pmatrix} \operatorname{sinc}(\omega_{\max}t) \\ -\frac{1}{2}\begin{bmatrix} \operatorname{sinc}(\omega_{\max}t - \gamma_{m_x} r_x \omega_{\max} p_{x,\max}) \\ +\operatorname{sinc}(\omega_{\max}t + \gamma_{m_x} r_x \omega_{\max} p_{x,\max}) \end{bmatrix} \\ -\frac{1}{2}\begin{bmatrix} \operatorname{sinc}(\omega_{\max}t - \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \\ +\operatorname{sinc}(\omega_{\max}t + \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \end{bmatrix} \\ +\frac{1}{4}\begin{pmatrix} \operatorname{sinc}(\omega_{\max}t - \gamma_{m_x} r_x \omega_{\max} p_{x,\max} - \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \\ +\operatorname{sinc}(\omega_{\max}t + \gamma_{m_x} r_x \omega_{\max} p_{x,\max} - \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \end{pmatrix} \\ +\frac{1}{4}\begin{pmatrix} \operatorname{sinc}(\omega_{\max}t - \gamma_{m_x} r_x \omega_{\max} p_{x,\max} + \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \\ +\operatorname{sinc}(\omega_{\max}t + \gamma_{m_x} r_x \omega_{\max} p_{x,\max} + \gamma_{m_y} r_y \omega_{\max} p_{y,\max}) \end{pmatrix} \end{pmatrix}$
$1+2+3$	$K_C(t,\mathbf{x},\mathbf{y}) \approx (2\pi)^{-6} \frac{8\pi\omega_{\max}}{r_x^3 r_y^2} \sum_{m=1}^{M} \frac{\alpha_m}{\gamma_m} \times \begin{pmatrix} g_0(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x) \\ -g_1(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \gamma_m\omega_{\max}p_{y,\max}r_y) \\ -\omega_{\max}p_{x,\max}r_x g_2(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x) \\ +\omega_{\max}p_{x,\max}r_x g_3(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \gamma_m\omega_{\max}p_{y,\max}r_y) \end{pmatrix}$
$1+3+3$	$K_C(t,\mathbf{x},\mathbf{y}) = (2\pi)^{-7} \frac{2^2(2\pi)^2\omega_{\max}}{r_x^3 r_y^3} \begin{pmatrix} f_1(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \omega_{\max}p_{y,\max}r_y) \\ -p_{x,\max}r_x\omega_{\max}f_2(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \omega_{\max}p_{y,\max}r_y) \\ -p_{y,\max}r_y\omega_{\max}f_2(\omega_{\max}t, \omega_{\max}p_{y,\max}r_y, \omega_{\max}p_{x,\max}r_x) \\ +p_{x,\max}p_{y,\max}r_x r_y \omega_{\max}^2 f_3(\omega_{\max}t, \omega_{\max}p_{x,\max}r_x, \omega_{\max}p_{y,\max}r_y) \end{pmatrix}$

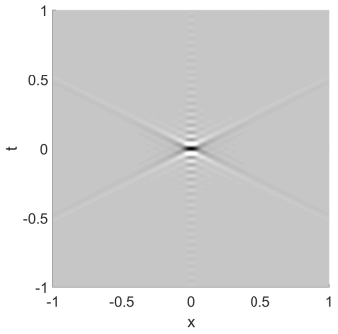
Table 1: Computed and approximated kernels for C-limited functions. We used the notation $r = r_x = |\mathbf{x}|$ and $r_y = |\mathbf{y}|$. See Table 2 for the definition of the functions f_i, $i = 1, 2, 3$ and g_i, $i = 0, 1, 2, 3$. See Appendices B and D for derivation of $(1+1)\,D$ and approximation of $(1+2)\,D$ kernels.

Dimensions	Functions used
$1+2+3$	$g_0(a,b) = \frac{1}{2}\left(\operatorname{cosinc}(a+b) - \operatorname{cosinc}(a-b)\right)$ $g_1(a,b,c) = \frac{1}{4}\begin{pmatrix} \operatorname{cosinc}(a+b+c) + \operatorname{cosinc}(a+b-c) \\ -\operatorname{cosinc}(a-b+c) - \operatorname{cosinc}(a-b-c) \end{pmatrix}$ $g_2(a,b) = \frac{1}{2}\left(\operatorname{cosinc}'(a+b) + \operatorname{cosinc}'(a-b)\right)$ $g_3(a,b,c) = \frac{1}{4}\begin{pmatrix} \operatorname{cosinc}'(a+b+c) + \operatorname{cosinc}'(a+b-c) \\ +\operatorname{cosinc}'(a-b+c) + \operatorname{cosinc}'(a-b-c) \end{pmatrix}$
$1+1+3$ $1+3+3$	$f_1(a,b,c) = \frac{1}{2}\begin{pmatrix} \operatorname{sinc}(a+b-c) + \operatorname{sinc}(a-b+c) \\ -\operatorname{sinc}(a-b-c) - \operatorname{sinc}(a+b+c) \end{pmatrix}$ $f_2(a,b,c) = -\frac{1}{2}\begin{pmatrix} \operatorname{sinc}'(a+b+c) - \operatorname{sinc}'(a+b-c) \\ +\operatorname{sinc}'(a-b+c) - \operatorname{sinc}'(a-b-c) \end{pmatrix}$ $f_3(a,b,c) = -\frac{1}{2}\begin{pmatrix} \operatorname{sinc}^{(2)}(a+b+c) + \operatorname{sinc}^{(2)}(a+b-c) \\ +\operatorname{sinc}^{(2)}(a-b+c) + \operatorname{sinc}^{(2)}(a-b-c) \end{pmatrix}$

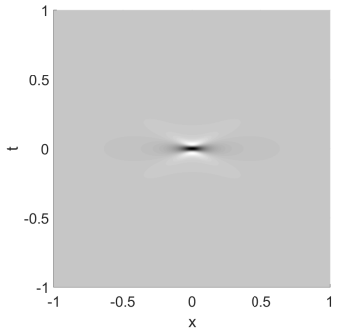
Table 2: Functions used in Table 1.

Dimensions	Kernel
$1+0$	$K_G(t) = (2\pi)^{-1} \exp\left(-\left[\sigma_\omega t\right]^2/2\right)$
$1+1$	$K_G(t,x) = -(2\pi)^{-2} 4\sqrt{\frac{\pi}{\left(1+[\sigma_k x]^2\right)}}\sigma_k \partial_t \left\{F\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+[\sigma_k x]^2\right)}}\right)\right\}$
$1+2$	$K_G(t,\mathbf{x}) = -(2\pi)^{-3} 8\left(\frac{\pi}{2}\right)^{3/2} \frac{\sigma_\omega \sigma_p^2}{\sqrt{1+\sigma_k^2[x^2+y^2]}} \partial_t^2 \left\{\mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\sigma_k^2[x^2+y^2]\right)}}\right\}$
$1+m+n$	$K_G(t,\mathbf{x},\mathbf{y}) = (2\pi)^{-(1+m+n)} \times \begin{cases} \left(\begin{array}{l} \frac{\left(\sqrt{2\pi}\right)^{m+n+1}}{\sqrt{2}^{m+n}} \sigma_{p_x}^m \sigma_{p_y}^n (-1)^{\mod(m+n,4)/2} \left[\frac{\sigma_\omega}{\sqrt{1+\lvert\sigma_{k_x}\mathbf{x}\rvert^2+\lvert\sigma_{k_y}\mathbf{y}\rvert^2}}\right]^{m+n+1} \\ \times H_{m+n}\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\lvert\sigma_{k_x}\mathbf{x}\rvert^2+\lvert\sigma_{k_y}\mathbf{y}\rvert^2\right)}}\right) \mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\lvert\sigma_{k_x}\mathbf{x}\rvert^2+\lvert\sigma_{k_y}\mathbf{y}\rvert^2\right)}} \end{array}\right), & \mod(m+n,2)=0 \\ \left(\begin{array}{l} \frac{2\sqrt{2}\left(\sqrt{2\pi}\right)^{m+n}}{\sqrt{2}^{m+n}} \sigma_{p_x}^m \sigma_{p_y}^n (-1)^{(\mod(m+n,4)+1)/2} \left[\frac{\sigma_\omega}{\sqrt{1+\lvert\sigma_{k_x}\mathbf{x}\rvert^2+\lvert\sigma_{k_y}\mathbf{y}\rvert^2}}\right]^{m+n+1} \\ \times \partial_t^{m+n}[F]\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\lvert\sigma_{k_x}\mathbf{x}\rvert^2+\lvert\sigma_{k_y}\mathbf{y}\rvert^2\right)}}\right) \end{array}\right), & \mod(m+n,2)=1 \end{cases}$

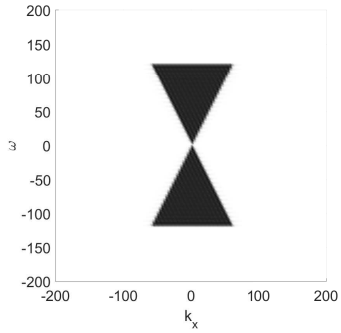
Table 3: Kernels for C-Gaussian functions. Here $H_n(t)$ is the Hermite polynomial defined by (56) and $F(t)$ is the is the Dawson's integral defined by (49). See Appendices C, E and F for the derivation of the kernels.



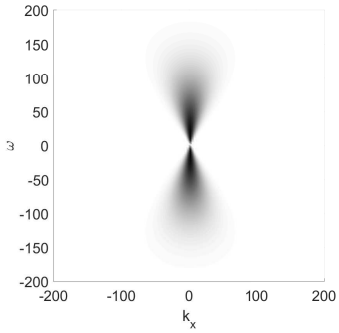
(a) K_C for $\omega_{\max} = 120$ and $p_{\max} = .5$



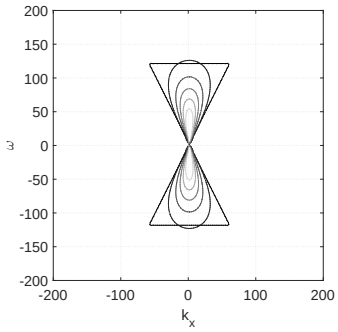
(b) K_G for $\sigma_\omega = 40$ and $\sigma_p = .25$



(c)

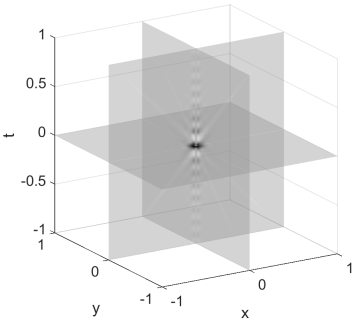


(d)

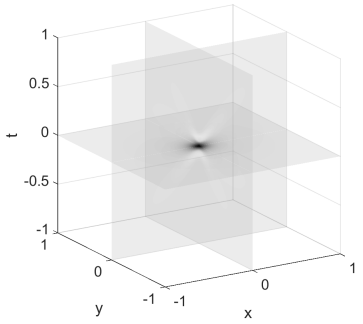


(e)

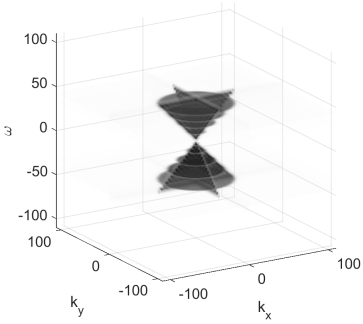
Figure 8: Plots of (1+1)D K_C, for $\omega_{\max} = 124$ and $p_{\max} = .5$, and K_G, for $\sigma_\omega = 40$ and $\sigma_p = .25$, in time-space (a,c) and absolute value of their Fourier transforms in frequency-wavenumber (b,d) domain, respectively (see Appendix A for definitions of K_C and K_G). Overlaid contour plots of absolute value of the Fourier transform of K_C and K_G in the frequency-wavenumber domain (e) shows that K_C can be effectively approximated by K_G.



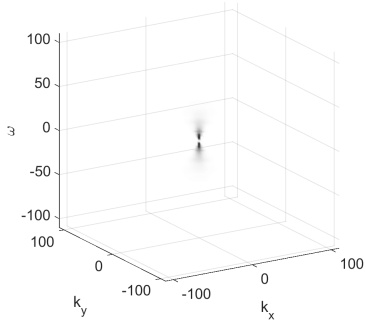
(a) K_C for $\omega_{\max} = 120$ and $p_{\max} = .5$



(b) K_G for $\sigma_\omega = 40$ and $\sigma_p = .25$



(c)



(d)

Figure 9: Plots of (1+2)D K_C, for $\omega_{\max} = 50$ and $p_{\max} = .5$, and K_G, for $\sigma_\omega = 25$ and $\sigma_p = .25$, in time-space (a,c) and absolute value of their Fourier transforms in frequency-wavenumber (b,d) domain, respectively (see Appendix A for definitions of K_C and K_G).

Appendix B: C-limited Kernel for $(1+1)$D

Consider kernel $K(t,x)$, $t \in \mathbb{R}$ and $x \in \mathbb{R}$, for $(1+1)$D whose Fourier transform, $\tilde{K}(\omega,k)$, is equal to one within $C = \{(\omega,k) \in \mathbb{R}\times\mathbb{R} | \omega \in [-\omega_0,\omega_0], |k| \leq |\omega| p_{\max}\}$. Then

$$\begin{aligned} K(t,x) &= (2\pi)^{-2} \int_C \mathrm{e}^{\mathrm{i}(\omega t - kx)} d\omega\, dk \\ &= (2\pi)^{-2} \int_{-p_{\max}}^{p_{\max}} \int_{-\omega_0}^{\omega_0} \mathrm{e}^{\mathrm{i}\omega(t-px)} |\omega|\ d\omega\, dp \\ &= (2\pi)^{-2} \frac{2\omega_0}{x} \left[\mathrm{cosinc}\left(\omega_0 [t + p_{\max} x]\right) - \mathrm{cosinc}\left(\omega_0 [t - p_{\max} x]\right)\right] \end{aligned} \tag{45}$$

where $\mathrm{cosinc}(x) = \int_0^1 \sin(\omega x)\, d\omega = \frac{1-\cos(x)}{x}$.

Appendix C: C-Gaussian function for $(1+1)$D

We assume we have Gaussian in frequency and in $p = k_x \omega^{-1}$:

$$\begin{aligned} K_G(t,x) &= (2\pi)^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\frac{p^2}{2\sigma_p^2}} \mathrm{e}^{\mathrm{i}\omega(t-px)} \underbrace{|\omega|\, d\omega dp}_{d\omega dk_x} \\ &= (2\pi)^{-2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\frac{p^2}{2\sigma_p^2}} \mathrm{e}^{\mathrm{i}\omega(t-px)} |\omega|\, d\omega dp \\ &= (2\pi)^{-2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\mathrm{i}\omega t} \left\{ \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\omega p x} \mathrm{e}^{-\frac{p^2}{2\sigma_p^2}} dp \right\} |\omega|\, d\omega \\ &= (2\pi)^{-2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\mathrm{i}\omega t} \left\{ 2\int_{0}^{\infty} \mathrm{e}^{-\frac{p^2}{2\sigma_p^2}} \cos(\omega x p)\, dp \right\} |\omega|\, d\omega \end{aligned} \tag{46}$$

which using the identity

$$\int_0^\infty \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \cos(\omega t)\, d\omega = \sqrt{\frac{\pi}{2}} \sigma_\omega \mathrm{e}^{-\frac{\sigma_\omega^2 t^2}{2}} \tag{47}$$

becomes

$$\begin{aligned} K_G(t,x) &= (2\pi)^{-2} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\mathrm{i}\omega t} \left\{ 2\int_{0}^{\infty} \mathrm{e}^{-\frac{p^2}{2\sigma_p^2}} \cos(\omega x p)\, dp \right\} |\omega|\, d\omega \\ &= (2\pi)^{-2}\, 2 \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \mathrm{e}^{-\mathrm{i}\omega t} \left\{ \sqrt{\frac{\pi}{2}} \sigma_p \mathrm{e}^{-\frac{\omega^2 x^2 \sigma_p^2}{2}} \right\} |\omega|\, d\omega \\ &= (2\pi)^{-2}\, 2\sqrt{\frac{\pi}{2}} \sigma_p \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2}\left(\sigma_\omega^{-2} + x^2\sigma_p^2\right)} \mathrm{e}^{-\mathrm{i}\omega t} |\omega|\, d\omega \end{aligned} \tag{48}$$

Using the definition of Dawson's integral $F(x)$ Abramowitz and Stegun (1965),

$$\int_0^\infty \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}} \sin(\omega t)\, d\omega = \sqrt{2}\sigma_\omega F\left(\frac{\sigma_\omega t}{\sqrt{2}}\right), \tag{49}$$

we can write

$$\begin{aligned} K_G(t,x) &= (2\pi)^{-2}\, 2\sqrt{\frac{\pi}{2}} \sigma_p \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2}\left(\sigma_\omega^{-2} + x^2\sigma_p^2\right)} \mathrm{e}^{-\mathrm{i}\omega t} |\omega|\, d\omega \\ &= -(2\pi)^{-2}\, 2\sqrt{\frac{\pi}{2}} \sigma_p \partial_t \left\{ 2\int_0^\infty \mathrm{e}^{-\frac{\omega^2}{2}\left(\sigma_\omega^{-2} + x^2\sigma_p^2\right)} \sin(\omega t)\, d\omega \right\} \\ &= -(2\pi)^{-2}\, 4\sqrt{\frac{\pi}{(1 + x^2\sigma_k^2)}} \sigma_k \partial_t \left\{ F\left(\frac{\sigma_\omega t}{\sqrt{2\,(1 + x^2\sigma_k^2)}}\right) \right\} \end{aligned} \tag{50}$$

Then $K_{G,n}(t,x) = \partial_t^n K_G(t,x)$ becomes

$$\begin{aligned} K_{G,n}(t,x) &= -(2\pi)^{-2} 4\sqrt{\frac{\pi}{(1+x^2\sigma_k^2)}}\sigma_k \partial_t^{n+1}\left\{F\left(\frac{\sigma_\omega t}{\sqrt{2(1+x^2\sigma_k^2)}}\right)\right\} \\ &= -(2\pi)^{-2}\frac{2^{\frac{3-n}{2}}\sqrt{\pi}\sigma_k\sigma_\omega^{n+1}}{(1+x^2\sigma_k^2)^{\frac{n+2}{2}}} F^{(n+1)}\left(\frac{\sigma_\omega t}{\sqrt{2(1+x^2\sigma_k^2)}}\right) \end{aligned} \tag{51}$$

where $\sigma_k = \sigma_p\sigma_\omega$. This expression can be efficiently approximated through approximation of Dawson function as a sum of product of Hankel functions and exponentials, which we discuss next.

Approximating Dawson's integral

Considering the Taylor series of the Dawson's integral $F(z)$ at zero, we can construct an approximation of the following form using a generalization of Padé approximation Yarman and Flagg (2015):

$$\begin{aligned} F(z) &= \sum_{n=0}^{\infty}\frac{(-1)^n 2^n}{(2n+1)!!}z^{2n+1} \\ &\approx z\sum_{m=1}^{M}\alpha_m \mathrm{e}^{-\gamma_m z^2} \end{aligned} \tag{52}$$

which is based on the following moment problem

$$\frac{(-1)^n 2^n}{(2n+1)!!} \approx \sum_{m=1}^{M}\alpha_m\frac{\gamma_m^n}{n!}. \tag{53}$$

A solution to the moment problem is given in Table 4.

Using the identity

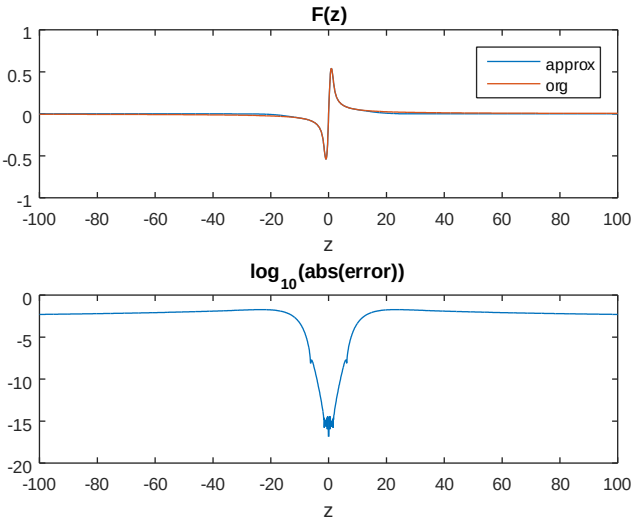
$$d_t^n\left[tf(t)\right] = nf^{(n-1)}(t) + tf^{(n)}(t) \tag{54}$$

we approximate derivatives of the Dawson integral by

$$\begin{aligned} d_z^n\hat{F}(z) &= \sum_{m=1}^{M}\alpha_m\left[nd_z^{n-1}\left\{\mathrm{e}^{-\gamma_m z^2}\right\} + zd_z^n\left\{\mathrm{e}^{-\gamma_m z^2}\right\}\right] \\ &= \sum_{m=1}^{M}\alpha_m\left[(-1)^{n-1}n\gamma_m^{(n-1)/2}H_{n-1}\left(\gamma_m^{1/2}z\right) + (-1)^n z\gamma_m^{n/2}H_n\left(\gamma_m^{1/2}z\right)\right]\mathrm{e}^{-\gamma_m z^2} \end{aligned} \tag{55}$$

Table 4: Table of (α_m, γ_m) used in approximation $F(z) \approx z \sum_{m=1}^{M=13} \alpha_m \mathrm{e}^{-\gamma_m z^2}$.

m	α_m	γ_m
1	0.115572712123670	-0.996654032557164
2	0.114181249809118	-0.970127281589401
3	0.111398554247821	-0.918350873603871
4	0.107222887331758	-0.843840584108076
5	0.101646558555868	-0.750277809703885
6	0.094652872760151	-0.642406973572952
7	0.086223071634045	-0.525900374357792
8	0.076359092260680	-0.407164113626245
9	0.065112215902274	-0.293046590019713
10	0.010785057749242	-0.008387187289610
11	0.025075427693923	-0.043780721833197
12	0.039149878504339	-0.105616569378803
13	0.052620421427110	-0.190420598936569



Top: Dawson integral (blue) and its approximation (red). Bottom: Logarithm of the absolute error.

Here $H_n(t)$ is the Hermite polynomial which is defined by

$$H_n(x) = (-1)^n \mathrm{e}^{x^2} d_x^n \left\{ \mathrm{e}^{-x^2} \right\} \tag{56}$$

and satisfies the three term recurrence relation

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x) \tag{57}$$

with $H_0(x) = 1$ and $H_1(x) = 2x$.

The least square error is computed to be

$$\begin{aligned} &\int_{-\infty}^{\infty} \left| d_x^n F(x) - d_x^n \hat{F}(x) \right|^2 dx \\ &= \pi 2^{n-5/2} \Gamma\left(n + \frac{1}{2}\right) + \sum_{m=1}^{M} \sum_{n=1}^{M} \frac{4^n \alpha_m \alpha_n (\gamma_m \gamma_n)^n}{(\gamma_m + \gamma_n)^{n+3/2}} \Gamma\left(n + \frac{3}{2}\right) - \sqrt{\pi} 4^n \sum_{m=1}^{M} \frac{\alpha_m \gamma_m^{n-1/2}}{(1+\gamma_m)^{n+1}} \Gamma(n+1) \end{aligned} \tag{58}$$

We can prove (58) using the identities

$$\int_0^{\infty} x^a \mathrm{e}^{-\gamma x^2} dx = \frac{1}{2} \gamma^{-(a+1)/2} \Gamma\left(\frac{a+1}{2}\right) \tag{59}$$

$$\int_{-\infty}^{\infty} \sin^{(a)}(kx) \sin^{(a)}(lx)\, dx = \pi \delta(k-l) \tag{60}$$

Then, the least square error is given by

$$\int_{-\infty}^{\infty} \left| d_x^a F(x) - d_x^a \hat{F}(x) \right|^2 dx = J_1 + J_2 - 2J_3 \tag{61}$$

where

$$
\begin{aligned}
J_1 &= \frac{1}{4}\int_0^\infty\int_0^\infty k^a l^a \mathrm{e}^{-\frac{k^2+l^2}{4}}\left[\int_{-\infty}^{\infty}\sin^{(a)}(kx)\sin^{(a)}(lx)\,dx\right]dk\,dl \\
&= \frac{\pi}{4}\int_0^\infty k^{2a}\mathrm{e}^{-\frac{2k^2}{4}}dk \\
&= \frac{\pi}{4}\frac{1}{2}\left(\frac{1}{2}\right)^{-(a+1/2)}\Gamma\left(a+\frac{1}{2}\right) \\
&= \pi 2^{a-5/2}\Gamma(a+1/2) \qquad (62)
\end{aligned}
$$

$$
\begin{aligned}
J_2 &= \int_{-\infty}^{\infty}\left[\sum_{m=1}^{M}\frac{\alpha_m}{\sqrt{\pi\gamma_m}}\int_0^\infty \frac{k^{a+1}\gamma_m^{a/2}}{2}\mathrm{e}^{-\frac{k^2}{4}}\sin^{(a)}(k\sqrt{\gamma_m}x)\,dk\right] \\
&\qquad\times\left[\sum_{n=1}^{M}\frac{\alpha_n}{\sqrt{\pi\gamma_n}}\int_0^\infty \frac{l^{a+1}\gamma_n^{a/2}}{2}\mathrm{e}^{-\frac{l^2}{4}}\sin^{(a)}(l\sqrt{\gamma_n}x)\,dl\right]dx \\
&= \frac{\pi}{4}\sum_{m=1}^{M}\sum_{n=1}^{M}\frac{\alpha_m\alpha_n}{\pi(\gamma_m\gamma_n)^{3/2}}\int_0^\infty k^{2(a+1)}\mathrm{e}^{-\frac{\left(\gamma_m^{-1}+\gamma_n^{-1}\right)k^2}{4}}dk \\
&= \sum_{m=1}^{M}\sum_{n=1}^{M}\frac{4^a\alpha_m\alpha_n(\gamma_m\gamma_n)^a}{(\gamma_m+\gamma_n)^{a+3/2}}\Gamma(a+3/2) \qquad (63)
\end{aligned}
$$

$$
\begin{aligned}
J_3 &= \int_{-\infty}^{\infty}\left[\frac{1}{2}\int_0^\infty k^a\mathrm{e}^{-\frac{k^2}{4}}\sin^{(a)}(kx)\,dk\right]\left[\sum_{m=1}^{M}\frac{\alpha_m}{\sqrt{\pi}\gamma_m^{3/2}}\int_0^\infty\frac{l^{a+1}}{2}\mathrm{e}^{-\frac{\gamma_m^{-1}l^2}{4}}\sin^{(a)}(lx)\,dl\right]dx \\
&= \frac{1}{4}\sum_{m=1}^{M}\frac{\alpha_m}{\sqrt{\pi}\gamma_m^{3/2}}\int_0^\infty\int_0^\infty k^a l^{a+1}\mathrm{e}^{-\frac{k^2+\gamma_m^{-1}l^2}{4}}\left[\int_{-\infty}^{\infty}\sin^{(a)}(kx)\sin^{(a)}(lx)\,dx\right]dk\,dl \\
&= \frac{\sqrt{\pi}}{4}\sum_{m=1}^{M}\frac{\alpha_m}{\gamma_m^{3/2}}\int_0^\infty k^{2a+1}\mathrm{e}^{-\frac{\left(1+\gamma_m^{-1}\right)k^2}{4}}dk \\
&= \sqrt{\pi}\sum_{m=1}^{M}\frac{2^{2a-1}\alpha_m\gamma_m^{a-1/2}}{(1+\gamma_m)^{(a+1)}}\Gamma(a+1) \qquad (64)
\end{aligned}
$$

which leads to (58).

Given (α_m,γ_m) as in Table 4, for n equal to 0, 1, 2, 3, 4 and 5, the least square errors normalized with respect to $\int_{-\infty}^{\infty}\left|d_x^n F(x)\right|^2 dx = \pi 2^{n-5/2}\Gamma\left(n+\frac{1}{2}\right)$ are 0.0258, 7.3923×10^{-5}, 1.1645×10^{-6}, 3.2947×10^{-8}, 1.4733×10^{-9}, and 1.0006×10^{-10}, respectively. Thus, Dawson function and its derivatives can be accurately and efficiently approximated.

Appendix D: C-limited Kernel for $(1+2)$D

Consider kernel $K(t,\mathbf{x})$, $t\in\mathbb{R}$ and $\mathbf{x}\in\mathbb{R}^2$, for $(1+2)$D whose Fourier transform, $\tilde{K}(\omega,\mathbf{k})$, is equal to one within $C=\left\{(\omega,\mathbf{k})\in\mathbb{R}\times\mathbb{R}^2|\omega\in[-\omega_0,\omega_0],|\mathbf{k}|\leq\omega p_{\max}\right\}$:

$$\begin{aligned}
K(t,\mathbf{x}) &= (2\pi)^{-3}\int_C \mathrm{e}^{\mathrm{i}(\omega t-\mathbf{k}\cdot\mathbf{x})}d\omega\,d\mathbf{k},\quad C=\left\{(\omega,\mathbf{k})\in\mathbb{R}\times\mathbb{R}^2|\omega\in[-\omega_0,\omega_0],|\mathbf{k}|\leq\omega p_{\max}\right\}\\
&= (2\pi)^{-3}\int_0^{2\pi}\int_0^{p_{\max}}\int_{-\omega_0}^{\omega_0}\mathrm{e}^{\mathrm{i}\omega(t-pr\cos\theta)}\omega^2 p\,d\omega\,dp\,d\theta,\quad |\mathbf{x}|=r\\
&= (2\pi)^{-3}\,4\int_0^1\left[\int_0^1 J_0\left(\omega r\omega_0 p_{\max}p\right)p\,dp\right]\cos(\omega\omega_0 t)\,\omega^2\omega_0^3\,p_{\max}^2\,d\omega\\
&= (2\pi)^{-3}\,\frac{4\omega_0^2\,p_{\max}}{r}\int_0^1 J_1(\omega r\omega_0 p_{\max})\cos(\omega\omega_0 t)\,\omega\,d\omega
\end{aligned}\tag{65}$$

where $J_n(t)$ is the n^{th} order Bessel function of the first kind. Following Yarman and Flagg (2015), $J_1(x)$ can be approximated by $J_1(x)\approx\sum_{m=1}^{M}\alpha_m\mathrm{cosinc}(\gamma_m x)$ where a solution for $(\alpha_m,\gamma_m)_{m=1}^{M=11}$ is given in Table 5.

Table 5: Table of (α_m,γ_m) used in approximation $J_1(x)\approx\sum_{m=1}^{M}\alpha_m\mathrm{cosinc}(\gamma_m x)$.

m	α_m	γ_m
1	-0.014657236787574	0.149849719646360
2	-0.030791593085749	0.295701374119717
3	-0.050322558926396	0.433883111513795
4	-0.076237543720937	0.561214541977292
5	-0.113915898598797	0.675085828291059
6	-0.174500782788577	0.773516867651912
7	9.820344725976016	1.001343018955562
8	-6.654085836116296	0.992693770137054
9	-1.256418697907451	0.965062653066941
10	-0.526619239685431	0.919148643455865
11	-0.284975362406715	0.855149098929969

Consequently, we have

$$\begin{aligned}
K(t,\mathbf{x})\approx\hat{K}(t,\mathbf{x}) &= (2\pi)^{-3}\,\frac{4\omega_0^2\,p_{\max}}{r}\int_0^1\left[\sum_{m=1}^{M}\alpha_m\frac{1-\cos(\gamma_m r\omega_0 p_{\max}\omega)}{\gamma_m r\omega_0 p_{\max}\omega}\right]\cos(\omega_0 t\omega)\,\omega\,d\omega\\
&= (2\pi)^{-3}\,\frac{4\omega_0}{r^2}\sum_{m=1}^{M}\frac{\alpha_m}{\gamma_m}\left(\frac{\sin(\omega_0 t)}{\omega_0 t}-\frac{1}{2}\left[\mathrm{sinc}\left(\omega_0\left[\gamma_m p_{\max}r-t\right]\right)+\mathrm{sinc}\left(\omega_0\left[\gamma_m p_{\max}r+t\right]\right)\right]\right)
\end{aligned}\tag{66}$$

Because

$$\hat{K}_p(t,\mathbf{x}) = \frac{1}{r^2}\left\{\operatorname{sinc}(\omega_0 t) - \frac{1}{2}\left[\operatorname{sinc}(\omega_0[pr-t]) + \operatorname{sinc}(\omega_0[pr+t])\right]\right\} \tag{67}$$

is C-limited within $\hat{C}_p = \left\{(\omega,\mathbf{k}) \in \mathbb{R}\times\mathbb{R}^2 | \omega \in [-\omega_0,\omega_0], |\mathbf{k}| \le \omega p\right\}$:

$$\begin{aligned}\int \hat{K}_p(t,\mathbf{x})\, \mathrm{e}^{\mathrm{i}(\omega t - \mathbf{k}\cdot\mathbf{x})} dt\, d\mathbf{x} &= \pi |\omega p|\, \chi_{[-1,1]}\left(\omega\omega_0^{-1}\right)\left\{\int_0^\infty \frac{1-\cos(|\omega p| r)}{|\omega p| r} J_0(|\mathbf{k}| r)\, dr\right\} \\ &= \pi |\omega p|\, \chi_{[-1,1]}\left(\omega\omega_0^{-1}\right) \operatorname{arcsinh}\left(\frac{\sqrt{(\omega p)^2 - |\mathbf{k}|^2}}{|\mathbf{k}|}\right), \quad 0 < |\mathbf{k}| \le |\omega p|, \end{aligned} \tag{68}$$

$\hat{K}(t,\mathbf{x})$ is C-limited within $\hat{C} = \left\{(\omega,\mathbf{k}) \in \mathbb{R}\times\mathbb{R}^2 | \omega \in [-\omega_0,\omega_0], |\mathbf{k}| \le \omega p_{\max} \max_m\{\gamma_m\}\right\}$. If $\max_m\{\gamma_m\} \approx 1$, then the cone-limit C of $K(t,\mathbf{x})$ is approximated by the cone-limit $\hat{C}$ of $\hat{K}(t,\mathbf{x})$. The least square error is given by

$$\begin{aligned}&\int \left|K(t,\mathbf{x}) - \hat{K}(t,\mathbf{x})\right|^2 dt\, d\mathbf{x} \\ &= (2\pi)^{-6} 4^2 \frac{2\pi\omega_0^3 p_{\max}^2}{3}\left(\frac{1}{2} - \sum_{m=1}^M \alpha_m\gamma_m + \sum_{m=1}^M\sum_{m'=1}^M \alpha_m\left[\int_{\mathbb{R}^+} \frac{\sin^2(\gamma_m r/2)}{\gamma_m r/2}\frac{\sin^2(\gamma_{m'} r/2)}{\gamma_{m'} r/2}\frac{dr}{r}\right]\alpha_{m'}\right).\end{aligned} \tag{69}$$

The integral inside square brackets can be explicitly computed using integration by parts and the identities 3.827-3.828 on pages 462-463 of Gradshteyn and Ryzhik (2007). Similar analysis can be extended to $(1+n+m)$D for n or m are even Yarman (2015).

Appendix E: C-Gaussian function for $(1+2)$D

We define the smooth $(1+2)\,D$ kernel by

$$K_{G,n}(t,x,y)=\partial_t^n K_G(t,x,y) \tag{70}$$

where

$$\begin{aligned} K_G(t,x,y) &= (2\pi)^{-3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}}\,\mathrm{e}^{-\frac{p_x^2+p_y^2}{2\sigma_p^2}}\,\mathrm{e}^{-\mathrm{i}\omega(t-[p_x x+p_y y])}\omega^2 d\omega dp_x dp_y \\ &= -(2\pi)^{-3}\,8\left(\frac{\pi}{2}\right)^{3/2}\frac{\sigma_\omega\sigma_p^2}{\sqrt{1+\sigma_k^2\left[x^2+y^2\right]}}\partial_t^2\left\{\mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\sigma_k^2[x^2+y^2]\right)}}\right\} \end{aligned} \tag{71}$$

Consequently we compute $K_{G,n}(t,x,y)$ by

$$\begin{aligned} &K_{G,n}(t,x,y)=\partial_t^n K_G(t,x,y) \\ &=(2\pi)^{-3}(-1)^{n+1}\pi^{3/2}2^{4+\frac{n-3}{2}}\sigma_p^2\left[\frac{\sigma_\omega}{\sqrt{1+\sigma_k^2\left[x^2+y^2\right]}}\right]^{n+3}H_{n+2}\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\sigma_k^2\left[x^2+y^2\right]\right)}}\right)\mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\sigma_k^2[x^2+y^2]\right)}} \end{aligned} \tag{72}$$

where $H_n(t)$ is the Hermite polynomial defined by (56).

Appendix F: C-Gaussian function for $(1+m+n)$D

$$
\begin{aligned}
&K_G(t,x)\\
&=(2\pi)^{-(1+m+n)}\int_{\mathbb{R}^{1+m+n}} \mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}}\mathrm{e}^{-\frac{|\mathbf{p_x}|^2}{2\sigma_{p_x}^2}}\mathrm{e}^{-\frac{|\mathbf{p_y}|^2}{2\sigma_{p_y}^2}}\mathrm{e}^{-\mathrm{i}\omega(t-\mathbf{p_x}\cdot\mathbf{x}+\mathbf{p_y}\cdot\mathbf{y})}|\omega|^{m+n}\,d\omega dp_x d\mathbf{p_y}\\
&=(2\pi)^{-(1+m+n)}\int_{\mathbb{R}}\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}}\left[\int_{\mathbb{R}^m}\mathrm{e}^{-\frac{|\mathbf{p_x}|^2}{2\sigma_{p_x}^2}}\mathrm{e}^{\mathrm{i}\omega\mathbf{p_x}\cdot\mathbf{x}}d\mathbf{p_x}\right]\left[\int_{\mathbb{R}^n}\mathrm{e}^{-\frac{|\mathbf{p_y}|^2}{2\sigma_{p_y}^2}}\mathrm{e}^{\mathrm{i}\omega\mathbf{p_y}\cdot\mathbf{y}}d\mathbf{p_y}\right]\mathrm{e}^{-\mathrm{i}\omega t}|\omega|^{m+n}\,d\omega\\
&=(2\pi)^{-(1+m+n)}\int_{\mathbb{R}}\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}}\left[\left(2\sqrt{\frac{\pi}{2}}\sigma_{p_x}\right)^m\mathrm{e}^{-\frac{\omega^2|\sigma_{p_x}\mathbf{x}|^2}{2}}\right]\left[\left(2\sqrt{\frac{\pi}{2}}\sigma_{p_y}\right)^n\mathrm{e}^{-\frac{\omega^2\left|\sigma_{p_y}\mathbf{y}\right|^2}{2}}\right]\mathrm{e}^{-\mathrm{i}\omega t}|\omega|^{m+n}\,d\omega\\
&=(2\pi)^{-(1+m+n)}\left(\sqrt{2\pi}\right)^{m+n}\sigma_{p_x}^m\sigma_{p_y}^n\int_{\mathbb{R}}\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}\left(1+|\sigma_{k_x}\mathbf{x}|^2+\left|\sigma_{p_y}\mathbf{y}\right|^2\right)}\mathrm{e}^{-\mathrm{i}\omega t}|\omega|^{m+n}\,d\omega\\
&=(2\pi)^{-(1+m+n)}\left(\sqrt{2\pi}\right)^{m+n}\sigma_{p_x}^m\sigma_{p_y}^n\\
&\quad\times\begin{cases}(-1)^{\mod(m+n,4)/2}\,\partial_t^{m+n}\left\{2\int_0^\infty\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}\left(1+|\sigma_{k_x}\mathbf{x}|^2+\left|\sigma_{p_y}\mathbf{y}\right|^2\right)}\cos(\omega t)\,d\omega\right\}, & \mod(m+n,2)=0\\ (-1)^{(\mod(m+n,4)+1)/2}\,\partial_t^{m+n}\left\{2\int_0^\infty\mathrm{e}^{-\frac{\omega^2}{2\sigma_\omega^2}\left(1+|\sigma_{k_x}\mathbf{x}|^2+\left|\sigma_{p_y}\mathbf{y}\right|^2\right)}\sin(\omega t)\,d\omega\right\}, & \mod(m+n,2)=1\end{cases}\\
&=(2\pi)^{-(1+m+n)}\left(\sqrt{2\pi}\right)^{m+n}\sigma_{p_x}^m\sigma_{p_y}^n\frac{\sigma_\omega}{\sqrt{1+|\sigma_{k_x}\mathbf{x}|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2}}\\
&\quad\times\begin{cases}(-1)^{\mod(m+n,4)/2}\sqrt{2\pi}\partial_t^{m+n}\left\{\mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}\right\}, & \mod(m+n,2)=0\\ (-1)^{(\mod(m+n,4)+1)/2}\,2\sqrt{2}\partial_t^{m+n}\left\{F\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+|\sigma_{k_x}\mathbf{x}|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}\right)\right\}, & \mod(m+n,2)=1\end{cases}
\end{aligned}
\tag{73}
$$

$$
\begin{aligned}
&K_G\left(t,\mathbf{x},\mathbf{y}\right)\\
&=\left(2\pi\right)^{-(1+m+n)}\left(\sqrt{2\pi}\right)^{m+n}\sigma_{p_x}^{m}\sigma_{p_y}^{n}\left[\frac{\sigma_\omega}{\sqrt{1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2}}\right]^{m+n+1}\\
&\quad\times\begin{cases}\frac{\sqrt{2\pi}}{\sqrt{2}^{m+n}}\left(-1\right)^{\mod(m+n,4)/2}\partial_t^{m+n}\left[\exp\left(-t^2\right)\right]\big|_{t=\frac{\sigma_\omega t}{\sqrt{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}}, & \mod\ (m+n,2)=0\\ \frac{2\sqrt{2}}{\sqrt{2}^{m+n}}\left(-1\right)^{(\mod(m+n,4)+1)/2}\partial_t^{m+n}\left[F\right]\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}\right), & \mod\ (m+n,2)=1\end{cases}\\
&=\left(2\pi\right)^{-(1+m+n)}\sqrt{\pi}^{m+n}\sigma_{p_x}^{m}\sigma_{p_y}^{n}\left[\frac{\sigma_\omega}{\sqrt{1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2}}\right]^{m+n+1}\\
&\quad\times\begin{cases}\sqrt{2\pi}\left(-1\right)^{\mod(m+n,4)/2}H_{m+n}\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}\right)\mathrm{e}^{-\frac{(\sigma_\omega t)^2}{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}, & \mod\ (m+n,2)=0\\ 2\sqrt{2}\left(-1\right)^{(\mod(m+n,4)+1)/2}\partial_t^{m+n}\left[F\right]\left(\frac{\sigma_\omega t}{\sqrt{2\left(1+\left|\sigma_{k_x}\mathbf{x}\right|^2+\left|\sigma_{k_y}\mathbf{y}\right|^2\right)}}\right), & \mod\ (m+n,2)=1\end{cases}
\end{aligned}
\tag{74}
$$

where $H_n\left(t\right)$ is the Hermite polynomial defined by (56).

Appendix G: Variable projection method

In Golub and Pereyra (1973, 2003), by reformulating minimization of $J(\mathcal{D},\mathcal{P})$ (see (8)) in two steps,

$$\min_{\mathcal{P}} J_{VP}(\mathcal{P}) = \min_{\mathcal{P}} \frac{1}{2} \|d - K(\mathcal{P})\mathcal{D}(\mathcal{P})\|_2^2 \tag{75}$$

where

$$\begin{aligned} \mathcal{D}(\mathcal{P}) &= \min_{\mathcal{D}} \frac{1}{2} \|d - K(\mathcal{P})\mathcal{D}\|_2^2 \\ &= K(\mathcal{P})^{-} d, \end{aligned} \tag{76}$$

it was shown that minimization of $J(\mathcal{D},\mathcal{P})$ with respect to $(\mathcal{D},\mathcal{P})$ is equivalent to minimization of the variable projection functional $J_{VP}(\mathcal{P})$ with respect to $\mathcal{P}$:

$$\min_{(\mathcal{D},\mathcal{P})} J(\mathcal{D},\mathcal{P}) = \min_{\mathcal{P}} J_{VP}(\mathcal{P}) \tag{77}$$

where

$$J_{VP}(\mathcal{P}) = \frac{1}{2} \left\| \left[I - K(\mathcal{P}) K(\mathcal{P})^{-} \right] d \right\|_2^2 = \frac{1}{2} \left\| P_{K(\mathcal{P})}^{\perp} d \right\|_2^2 \tag{78}$$

with $K(\mathcal{P})^{-}$ is the (pseudo) inverse of $K(\mathcal{P})$, $P_{K(\mathcal{P})}^{\perp} = \left[I - K(\mathcal{P}) K(\mathcal{P})^{-} \right]$ is the projection operator onto kernel of $K(\mathcal{P})$. The corresponding Gauss-Newton algorithm becomes (see (5.2) in Golub and Pereyra (1973))

$$\begin{aligned} \mathcal{P}^{l+1} &= \mathcal{P}^{l} - t_l \left[\partial_{\mathcal{P}} P_{K(\mathcal{P})}^{\perp} \right]_{\mathcal{P}=\mathcal{P}^l} d \\ &= \mathcal{P}^{l} - t_l \left[P_{K(\mathcal{P}^l)}^{\perp} \left[\partial_{\mathcal{P}} K(\mathcal{P}) \right]_{\mathcal{P}=\mathcal{P}^l} K\left(\mathcal{P}^l\right)^{-} + \left(P_{K(\mathcal{P}^l)}^{\perp} \left[\partial_{\mathcal{P}} K(\mathcal{P}) \right]_{\mathcal{P}=\mathcal{P}^l} K\left(\mathcal{P}^l\right)^{-} \right)^{T} \right] d \end{aligned} \tag{79}$$

where $0 < t_l \leq 1$ is the step length chosen to prevent divergence. For computational efficiency, (79) is approximated by Kaufman (1975); Gay and Kaufman (1991)

$$\begin{aligned} \mathcal{P}^{l+1} &\approx \mathcal{P}^{l} - t_l \left[P_{K(\mathcal{P}^l)}^{\perp} \left[\partial_{\mathcal{P}} K(\mathcal{P}) \right]_{\mathcal{P}=\mathcal{P}^l} K\left(\mathcal{P}^l\right)^{-} \right] d \\ &= \mathcal{P}^{l} - t_l P_{K(\mathcal{P}^l)}^{\perp} \left[\partial_{\mathcal{P}} K(\mathcal{P}) \right]_{\mathcal{P}=\mathcal{P}^l} \mathcal{D}\left(\mathcal{P}^l\right) \end{aligned} \tag{80}$$

Algorithm 3 Variable Projection

1. Initialize $\mathcal{P}^0$ and t_0
2. For $l > 0$,
   (a) compute $\mathcal{D}^{l+1} = \min_{\mathcal{D}} \frac{1}{2} \left\| d - K\left(\mathcal{P}^l\right)\mathcal{D} \right\|_2^2$
   (b) Compute $\mathcal{P}^{l+1} = \mathcal{P}^l - t_l P^{\perp}_{K(\mathcal{P}^l)} \left[\partial_{\mathcal{P}} K\left(\mathcal{P}\right)\right]_{\mathcal{P}=\mathcal{P}^l} \mathcal{D}^{l+1}$

   until converged.

The main difference between this approximation of variable projection and alternating least squares is projection onto the kernel of the $K\left(\mathcal{P}\right)$ using $P^{\perp}_{K(\mathcal{P}^l)}$ Mullen et al. (2007). We summarize the variable projection method in Algorithm 3 .