

Data-driven modeling for wave-propagation

Tristan van Leeuwen, Peter Jan van Leeuwen and Sergiy Zhuk

Abstract Many imaging modalities, such as ultrasound and radar, rely heavily on the ability to accurately model wave propagation. In most applications, the response of an object to an incident wave is recorded and the goal is to characterize the object in terms of its physical parameters (e.g., density or soundspeed). We can cast this as a joint parameter and state estimation problem. In particular, we consider the case where the inner problem of estimating the state is a weakly constrained data-assimilation problem. In this paper, we discuss a numerical method for solving this variational problem.

1 Introduction

Many imaging modalities, such as ultrasound, geophysical exploration, and radar, rely heavily on the ability to accurately model wave propagation. In most applications, the response of an object to an incident wave is recorded and the goal is to characterize the object in terms of its physical parameters (e.g., density or soundspeed). We can capture this setup in terms of a process and measurement model

$$\mathcal{L}(c)u = q, \tag{1}$$

$$d = \mathcal{P}u, \tag{2}$$

where u denotes the wavefield, $\mathcal{L}(c) = \partial_t^2 - c^2 \nabla^2$ represents the wave equation depending on the physical parameters, c , and \mathcal{P} is the sampling operator that models

Tristan van Leeuwen

Utrecht University, Utrecht, the Netherlands e-mail: t.vanleeuwen@uu.nl

Peter Jan van Leeuwen

Colorado State University, Colorado, U.S.A. e-mail: Peter.vanLeeuwen@colostate.edu

Sergiy Zhuk

IBM Research, Dublin, Ireland

the measurement process. In particular, consider the case where we are given a *finite* number of samples of the state. We are ultimately interested in estimating the parameters c from the measurements d . There are different ways to go about this;

PDE-constrained optimization: Eliminate the process model and set up a non-linear data-fitting problem to match the solution of (1) to the data [12].

Equation-error approach: Estimate the state directly from the measurements by solving (2) and subsequently solve for c from (1) [8, 2].

Joint parameter and state estimation: Find the parameter and state that satisfy both (1) and (2) approximately [13].

The former two can be thought of as limiting cases of the latter where the state is estimated either completely determined by the process model or determined solely from the data. The joint approach gives rise to a data-driven modelling problem, where one aims to estimate a state that satisfies both the data and the physics to some extent. We can formally express this as a variational problem

$$\min_u \|\mathcal{P}u - d\|^2 + \rho \|\mathcal{L}(c)u - q\|^2, \quad (3)$$

where ρ is a parameter that controls the trade-off between the two terms. How well we are able to approximate the *true* state by solving (3) depends on how many measurements are available, the observability of the system, how close c is to the true parameter and requires an appropriate choice of ρ . We will not address this issue here and focus solely on solving (3) for given c and ρ .

A straightforward approach to solving (3) would be to derive the Euler-Lagrange equations, and solve the resulting PDE numerically. Alternatively, one could discretize the wave-equation first and set up a large sparse system of equations for the state [3, 7, 9, 4]. In both cases, the dimensionality of the problem is governed by the numerical discretization. In this paper, we cast the problem in a reproducing kernel Hilbert space, allowing us to express the solution of (3) as a finite linear combination of kernel functions [11, 5]. This leads to a system of linear equations involving a kernel matrix. The dimension of this system is given by the number of measurements and is thus independent of the underlying numerical discretisation of the PDE. We discuss a preconditioned iterative method for solving this system. Finally, we present some numerical examples and conclude the paper.

2 Theory

We consider a scalar wave-equation in $[0, T] \times \mathbb{R}^d$ of the form $\mathcal{L}u = q$ with $\mathcal{L}(c) = \partial_t^2 - c^2 \nabla^2$ and initial conditions $u(0, x) = \partial_t u(0, x) = 0$. Without elaborating on the details, we assume that this problem is well-posed for all parameters c of interest and that the solution is given by $u = \mathcal{G}q$ with

$$\mathcal{G}q(t, x) = \int_0^t \int_{\mathbb{R}^d} g(t-s, x, y) q(s, y) dy ds.$$

The corresponding adjoint problem $\mathcal{L}^* v = r$ (with $u(T, x) = \partial_t u(T, x) = 0$) has solution $v = \mathcal{G}^* r$

$$\mathcal{G}^* r(t, x) = \int_t^T \int_{\mathbb{R}^d} g(s - t, x, y) r(s, y) dy ds.$$

For further details regarding the well-posedness of variable-coefficient wave-equations we refer to [1].

The measurements are obtained by sampling the state at given locations $\{(t_i, x_i)\}_{i=1}^M$:

$$\mathcal{P}u = \{u(t_i, x_i)\}_{i=1}^M.$$

We introduce a Hilbert space \mathcal{U} with inner product:

$$\langle u, v \rangle_{\mathcal{U}} = \langle \mathcal{L}u, \mathcal{L}v \rangle_{L^2(\mathbb{R}^{d+1})}.$$

We can think of this as the space of solutions of the wave equation with square integrable source term. The space \mathcal{U} is a *Reproducing Kernel Hilbert Space* (RKHS) [6]. A special property of an RKHS is that point-evaluation is a bounded linear functional with Riesz representation $k_{t,x} \in \mathcal{U}$ so that $\langle k_{t,x}, u \rangle_{\mathcal{U}} = u(t, x)$. The reproducing kernel of \mathcal{U} is given by $k(t, x, t', x') = \langle k_{t,x}, k_{t',x'} \rangle_{\mathcal{U}}$. It is the Greens function of $\mathcal{L}^* \mathcal{L}$ and is thus given by

$$k(t, x, t', x') = \int_0^T \int_{\mathbb{R}^d} g(t - s, x, y) g(t' - s, y, x') dy ds.$$

We can think of k as a spline that is tailored to represent solutions of the wave equation.

The aim is to solve a variational problem of the form

$$\min_{u \in \mathcal{U}} \sum_{i=1}^M (u(t_i, x_i) - d_i)^2 + \rho \|\mathcal{L}u - q\|_{L^2(\mathbb{R}^{d+1})}^2. \quad (4)$$

By splitting the solution as $u = \mathcal{G}q + w$ and introducing $r = d - \mathcal{G}q$, we can re-write this as

$$\min_{w \in \mathcal{U}} \sum_{i=1}^M (w(t_i, x_i) - r_i)^2 + \rho \|w\|_{\mathcal{U}}^2. \quad (5)$$

Utilizing the Representer Theorem [10], we know that the solution to this variational problem has the following form

$$w(t, x) = \sum_{i=1}^M w_i k(t_i, x_i, t, x).$$

We can use this finite-dimensional representation of the solution to express (4) as a finite-dimensional least-squares problem

$$\min_{\mathbf{w} \in \mathbb{R}^M} \|K\mathbf{w} - \mathbf{r}\|_2^2 + \rho \mathbf{w}^T K \mathbf{w}, \quad (6)$$

where K is the kernel matrix with elements $k_{ij} = k(t_i, x_i, t_j, x_j)$. The kernel matrix is guaranteed to be positive definite, ensuring a unique solution given by

$$\hat{\mathbf{w}} = (K + \rho I)^{-1} \mathbf{r}.$$

Example: constant coefficients

With $c(x) = 1$ we can express all quantities in the spatial Fourier domain. The Greens function is then given by

$$\hat{g}(t, \xi) = \frac{\sin((t-s)|\xi|)}{|\xi|},$$

giving

$$\hat{k}(t, t', \xi) = |\xi|^{-2} \int_0^{\min(t, t')} \sin((t-s)|\xi|) \sin((t'-s)|\xi|) ds,$$

which yields

$$\hat{k}(t, t', \xi) = \begin{cases} t|\xi|^{-2} \cos((t-t')|\xi|) - |\xi|^{-3} \cos(t'|\xi|) \sin(t|\xi|) & t \leq t' \\ t'|\xi|^{-2} \cos((t-t')|\xi|) - |\xi|^{-3} \cos(t|\xi|) \sin(t'|\xi|) & t > t'. \end{cases} \quad (7)$$

To get some insight into the properties of the continuous kernel operator defined by

$$\mathcal{K}\hat{u}(t) = \int_0^\infty \hat{k}(t, t', \xi) \hat{u}(t') dt',$$

we take $\hat{u}(t) = \sin(\omega t)$ and find that this is an eigenfunction with eigenvalue $\lambda = (\omega^2 - |\xi|^2)^{-2}$. The continuous operator can thus have an arbitrary large norm due to modes with $\omega \approx \pm |\xi|$. The corresponding kernel matrix, K , can thus be extremely ill-conditioned.

3 Algorithm

We discretize all quantities on a regular grid and introduce the notation $\mathbf{u}_k = (u(k \cdot \Delta t, x_1), u(k \cdot \Delta t, x_2), \dots, u(k \cdot \Delta t, x_{n_x}))$. A second order finite-difference discretization of \mathcal{L} on $[0, T] \times [-D, D]$ leads to forward and adjoint systems of the form $L\mathbf{u} = \mathbf{q}$, $L'\mathbf{v} = \mathbf{r}$ with $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_{n_t})$, $\mathbf{q} = (\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{n_t-1})$ and

$$L = (\Delta t)^{-2} \begin{pmatrix} 2I & & & & \\ S & I & & & \\ I & S & I & & \\ & & \ddots & \ddots & \\ & & & I & S & I \end{pmatrix}, \quad L' = (\Delta t)^{-2} \begin{pmatrix} I & S^T & I & & \\ & \ddots & & & \\ S^T & I & S^T & I & \\ & & I & S^T & \\ & & & 2I \end{pmatrix}, \quad (8)$$

where $S = -2I - (\Delta t)^2 A$ and $A \in \mathbb{R}^{n_x \times n_x}$ is a second order central finite-difference discretization of $c^2 \nabla^2$. Note that $L' \neq L^T$. The sampling operator is discretized using piecewise linear interpolation, yielding a matrix $P \in \mathbb{R}^{M \times N}$. Using adjoint interpolation is an appropriate way to represent the point source [14]. The combination of a second-order finite-difference approximation of \mathcal{L} and linear interpolation ensures an overall second order approximation of the elements of K .

Storing the full matrix $K = P(L'L)^{-1}P^T$ may not be very attractive, but we can compute matrix-vector products with K by solving one forward and one adjoint problem;

$$K\mathbf{w} = P\mathbf{u},$$

with $L\mathbf{u} = \mathbf{v}$ and $L'\mathbf{v} = P^T\mathbf{w}$. Since the matrix is symmetric we can apply CG to solve the system $(K + \rho I)\mathbf{w} = \mathbf{r}$. As this system becomes increasingly ill-conditioned as ρ decreases, preconditioning is of paramount importance. Due to the specific form of K , we propose a preconditioner of the form $K^{-1} \approx M = QL'LQ^T$, where Q is chosen so that $Q^TP\mathbf{u} \approx \mathbf{u}$ for solutions of $L\mathbf{u} = \mathbf{q}$. When P samples the solution on a grid we can take Q to be a high-order interpolation operator (e.g. cubic splines).

4 Numerical results

4.1 Harmonic oscillator

For a single spatial Fourier mode, the kernel is given by (7). Figure 1 (a) shows an example of $\hat{k}(t, t')$ for $\xi = 20$, $t' = \frac{1}{3}$ and $t = \frac{1}{2}$. We take samples on a regular grid with $m = 20$ samples in $(0, 1)$. The spectrum of the corresponding kernel matrix is shown in figure 1 (b). Also shown is the Fourier approximation of the spectrum. The kink in the spectrum is due to the singularity in the spectrum of the continuous operator at $\omega = \|\xi\|$. Figure 1 (c) shows the absolute error as a function of Δt when using the numerical approximation described above. The effect of the preconditioner is shown in figure 2. We see that when using 1D spline interpolation, most of the eigenvalues of MK are clustered around one.

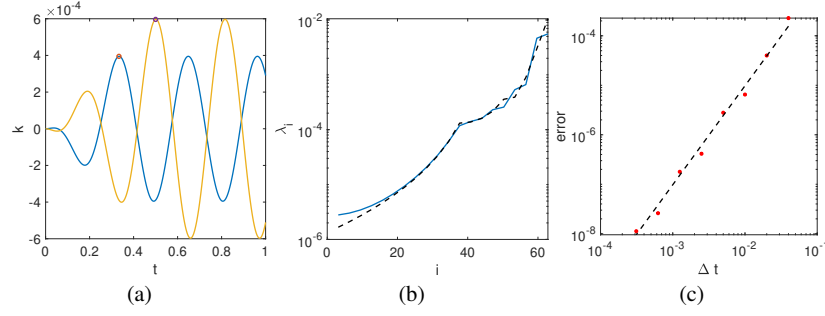


Fig. 1 (a) kernel $\hat{k}(t, t')$ for $\xi = 20$, $t' = \frac{1}{3}$ and $t' = \frac{1}{2}$. (b) Spectrum of K (solid) and its Fourier-approximation (dashed). (c) absolute error as a function of Δt

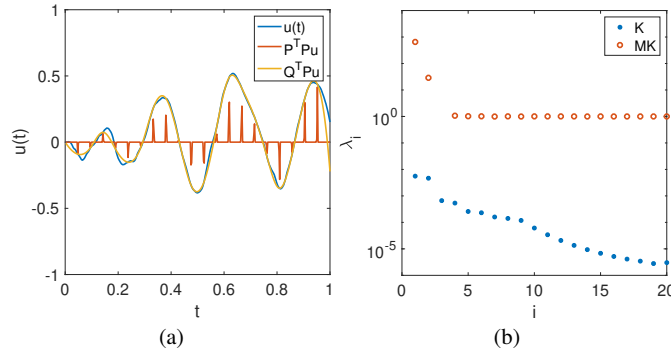


Fig. 2 (a) Illustration of the action of the interpolation operator Q . The solution to $L\mathbf{u} = \mathbf{q}$ with random source term is depicted, as well as its interpolated result $P^T P \mathbf{u}$ and $Q^T P \mathbf{u}$. (b) Eigenvalues of K and MK . The preconditioner nicely clusters most of the eigenvalues.

4.2 1D wave-equation

We generate data by solving the non-constant coefficient wave equation and sampling the solution on a regular grid. We then solve the variational problem for a constant reference soundspeed. The grid and velocity profiles are depicted in figure 3. The true state, reference state and reconstructed state for $\rho = 10^{-4} \|K\|_2$ are depicted in figure 4. As preconditioner for small ρ we use $(K + \rho I)^{-1} \approx K^{-1} \approx M$ while for large ρ we use $(K + \rho I)^{-1} \approx \rho^{-1} I - \rho^{-2} K$ with $M = QL' LQ^T$ and Q is 2D spline interpolation. The convergence history of CG, with and without preconditioner, for various values of ρ is shown in figure 5.

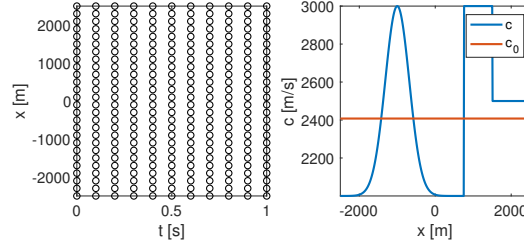


Fig. 3 Left: Grid used to define the sampling operator. Right: velocity profiles used to generate data and estimate state.

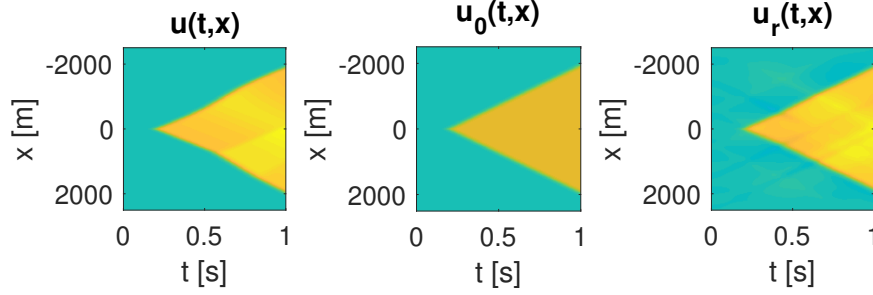


Fig. 4 True, reference and estimated states for $\rho = 10^{-4}$.

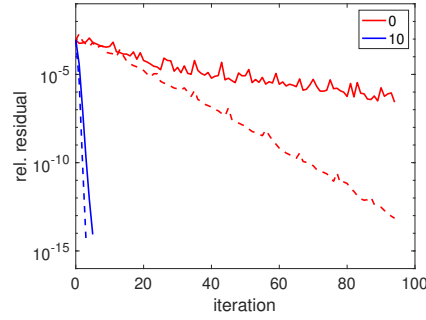


Fig. 5 Convergence of CG when solving the system for various values of ρ with (solid) and without (dashed) preconditioner.

5 Conclusion and discussion

We presented a numerical method for solving a variational data-assimilation problem involving the wave equation. By casting the problem in a Reproducing Kernel Hilbert Space, we derived a finite-dimensional system of equations involving a kernel matrix. Computing the action of this kernel on a given vector involves numerically solving a forward and adjoint wave equation and we described a non-self-adjoint second-order finite difference scheme for the wave equation to approximate the kernel. Using a simple Fourier analysis we show that the kernel matrix can be arbitrary

ill-conditioned. A simple preconditioner was proposed that appears to perform reasonably well in practice. The numerical examples presented in this paper involved a 1D wave equation and a relatively dense measurement grid. While the methodology described here can be easily extended to higher dimensions, the simple preconditioner will probably not perform as well on coarser measurement grids. Further analysis of the kernel for non-constant coefficients may shed some light on this issue.

References

1. Kirk D Blazek, Christiaan Stolk, and William W Symes. A mathematical framework for inverse wave problems in heterogeneous media. *Inverse Problems*, 29(6):065001, nov 2013.
2. S A L De Ridder and A Curtis. Seismic gradiometry using ambient seismic noise in an anisotropic Earth. *Geophysical Journal International Geophys. J. Int*, 209:1168–1179, 2017.
3. Ron Estrin, Dominique Orban, and Michael A. Saunders. LSLQ: An Iterative Method for Linear Least-Squares with an Error Minimization Property. *SIAM Journal on Matrix Analysis and Applications*, 40(1):254–275, jan 2019.
4. Melina A. Freitag and Daniel L.H. Green. A low-rank approach to the solution of weak constraint variational data assimilation problems. *Journal of Computational Physics*, 357:263–281, 2018.
5. Javier González, Ivan Vujačić, and Ernst Wit. Reproducing kernel Hilbert space based estimation of systems of ordinary differential equations. *Pattern Recognition Letters*, 45:26–32, aug 2014.
6. Vern I Paulsen and Raghupathi Mrinal. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. 2016.
7. Bas Peters and Felix J. Herrmann. A numerical solver for least-squares sub-problems in 3D wavefield reconstruction inversion and related problem formulations, pages 1536–1540. 2019.
8. C. Poppeliers, P. Punosevac, and T. Bell. Three-Dimensional Seismic-Wave Gradiometry for Scalar Waves. *Bulletin of the Seismological Society of America*, 103(4):2151–2160, aug 2013.
9. Gabrio Rizzuti, Mathias Louboutin, Rongrong Wang, Emmanouil Daskalakis, and Felix Herrmann. A dual formulation for time-domain wavefield reconstruction inversion, pages 1480–1485. 2019.
10. Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A Generalized Representer Theorem. *COLT/EuroCOLT*, 2111(2000 - 81):416–426, 2001.
11. Florian Steinke and Bernhard Schölkopf. Kernels, regularization and differential equations. *Pattern Recognition*, 41(11):3271–3286, 2008.
12. A Tarantola and A Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics*, 20(2):129–232, 1982.
13. T van Leeuwen and Felix J Herrmann. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007, jan 2016.
14. Johan Waldén. On the approximation of singular source terms in differential equations. *Numerical Methods for Partial Differential Equations*, 15(4):503–520, 1999.