University of Bayreuth
Chair of Scientific Computing
Chair of Applied Mathematics

# Physics Informed Neural Networks for Inverse Problems involving the heat equation

**Master's Thesis**

by

Mehdi Hassanbeigi

August 2026

Supervised by:
Prof. Dr. Lars Grüne
Prof. Dr. Vadym Aizinger

In association with:
Chair of Applied Mathematics

# Dedication:

*To my Mom and Dad,*

*I hope that this dissertation will reflect, even in a small way, some of the love and sacrifices that you have made for me, and that this work will bring you happiness.*

*I cannot thank you enough for what you have done for me. I am sorry for the times that I have worried you, when I have been absent from your life, and for all the things that you have sacrificed for me.*

*Words can't describe how much I love you, and I hope that this work will bring you a bit of the happiness and joy that you have brought me.*

## Acknowledgments

I'd like to thank Prof.Dr Lars Grüne, it has been an honor to work under the guidance of a world class supervisor. His expertise and patience have shaped not only this dissertation, but also my growth as a researcher. I am grateful for the opportunity to learn from him and for his support throughout this journey.

Also I'd like to thank Prof.Dr Vadym Aizinger, thank you for reviewing and correcting this Dissertation. I appreciate the time you dedicated to checking my work.


## Declaration of Originality

I hereby declare that this dissertation is my own work and that I am the only author.

No external tools or assistance were used in the preparation of this dissertation beyond those explicitly stated herein.

Artificial Intelligence (AI) tools were used sometimes only for formatting assistance, translation, and improving clarity. NO AI tools were used to generate the research, proofs, results, or conclusions presented in this dissertation.

# Abstract

The inverse heat source problem is studied both computationally and theoretically in this Dissertation. The stationary heat source $\varphi^\dagger$ on the real line is identified through a single noisy observation of the terminal temperature distribution $v(\cdot, 1)$, satisfying the differential equation $\partial_t v - \partial_{xx} v = \varphi$ with initial condition $v(\cdot, 0) = 0$. We are presenting our methodology (Hermite variational physics-informed neural networks), where we first train the network by using data obtained by intentionally perturbing our forward model, and then we provide mathematical support for what we found.

In Chapter 1 we discuss what is needed later in the analysis. We begin our discussion with introduction to PDEs and their classification. The main equation is parabolic in the domain of $\mathbb{R} \times (0, 1]$. Data and operator belong to $L^2(\mathbb{R})$, while the source belongs to Schwartz class $\mathcal{S}(\mathbb{R})$, and its dual enables definition of Sobolev spaces of negative or fractional order. The $H^s(\mathbb{R})$ spaces with the weight $(1 + \xi^2)^s$ characterize the scale for the entire problem, while the a-priori ball $\mathcal{W}_{\ell,E}$ in $H^\ell(\mathbb{R})$ allows estimating any stability since the inverse $A^{-1}$ is unbounded in $L^2(\mathbb{R})$. Bochner Space, $C([0, 1]; L^2(\mathbb{R}))$ must be defined to ensure that the operator $A$ exists. Scaled Hermite basis constitutes an orthonormal system which helps to make penalty diagonalizable by solving the eigenvalue problem thus eliminating the source term analytically. Duhamel's formula and contraction of the heat semigroup make sure that the small residue leads to small final error. Data is calculated by the Crank Nicolson approach, and there are two Gauss quadratures in the loss calculation.

In Chapter 2, we begin with a brief description of Hadamard's notions of well-posedness. Uniqueness is ensured by $m > 0$; existence is in $\mathrm{Ran}(A)$, which is dense, but not closed. Stability, however, is not achieved as $m(\xi)$ tends to zero. Such an instability depends only quadratically (polynomial order 2) rather than exponentially, and, as $A$ is not compact, we cannot apply the singular value grading. Hence, Sobolev-penalised Tikhonov regularization for $\mathcal{W}_{\ell,E}$ leads to a conditional stability estimate with exponent $\ell/(\ell + 2)$ but with an unknown constant $C(E, \ell)$.

In Chapter 3, we explain the elements related to the Computational Method. The approximation of the state is done using a neural network whose activation function is the tanh function and therefore its residual can easily be obtained through automatic differentiation. The Universal Approximation Theorem allows such an approximation without giving convergence rates, which is why we proceed a posteriori. The variational approach tests for the residual in a weak form through Hermite basis functions. The residual is an affine function of the source vector, which makes the loss function strictly convex in the source coefficients. This gives rise to an analytical minimizer of the source, making the determination of the source map possible.

Main results are stated in Chapter 4. Bounded, self-adjoint, Injective operator is not compact. Spectrum lies on segment $[0, 1]$ and it is absolutely continuous. Also there are no eigenvalues in this spectrum. Therefore, for proving sharpness, we have to analyze Weyl sequences. $A$ defines an isomorphism from $H^s(\mathbb{R})$ to $H^{s+2}(\mathbb{R})$. These estimates are optimal and are independent of $s$. The inverse operator has operator norm equal to one from $H^{s+2}(\mathbb{R})$ to $H^s(\mathbb{R})$ whereas the operator has the operator norm $c_\sharp$ from $H^s(\mathbb{R})$ to $H^{s+2}(\mathbb{R})$ where $c_\sharp = 1.2984\ldots$ The stability constant in the mentioned stability estimate is taken to be $(2E)^{2/(\ell+2)}$, and we show that this constant is sharp. That is because the considered function is the Laplace transform function, and hence, the square of the considered function, which is $m^2$, is the completely monotone function and is strictly convex with respect to the weight $(1+\xi^2)^\ell$. In this case, the moment inequality involving two atoms is reduced to the moment inequality involving one atom using the one point tangent line certificate. This results in the accurate modulus of continuity in the presence of any noise level (as opposed to asymptotically small), in the optimal estimation accuracy over all reconstruction schemes, and also an accurate square root rate in $J$ for several observation times. Thus, the source map produced by the network, which was eliminated, will be the same as Tikhonov's regularization for the minimax optimal filter in Hermite Sobolev spaces, and as the certificate is a posteriori, both hypotheses are discharged from the log.

In Chapter 5 we will compare our approach to two traditional Classical methods. The first is the mollification method with the use of Dirichlet Kernel, which is a low pass filter with only one parameter and lacking a Hilbert scale structure.Second is Adjoint state inversion under Sobolev Tikhonov regularization which is performed with our basis and truncation in grid form without neural state. We discuss advantages and disadvantages of each technique by running various experiments, including noise sweep, sparse sensing, limited domain observation, unknown diffusivity, non-smooth source, high noise, and some other experiments.

# Contents


**1 Analytical, Spectral, and Numerical Foundations** **5**

1.1 Introduction to Partial Differential Equations . . . . . 6
1.2 The Heat Equation with a Time Independent Source . . . . . 6
1.3 Function Spaces and Fourier-Hermite Analysis . . . . . 8
1.3.1 Lebesgue, Schwartz, and Sobolev Spaces . . . . . 8
1.3.2 The Fourier Transform . . . . . 9
1.3.3 Hermite Functions and the Hermite-Sobolev Structure . . . . . 10
1.3.4 Bochner Spaces and the Heat Semigroup . . . . . 12
1.4 Analytical, Classical, and Spectral Numerical Methods . . . . . 14
1.4.1 Fourier Multiplier Representation of the Forward Operator . . . . . 14
1.4.2 Finite Difference Methods . . . . . 15
1.4.3 Gauss-Hermite and Gauss-Legendre Quadrature . . . . . 16
1.5 From Classical Methods to Modern Methods . . . . . 17

**2 Inverse Problems and Regularization Theory** **19**

2.1 The Forward Problem and Its Inverse . . . . . 20
2.2 Classification and Analysis . . . . . 20
2.2.1 Classification of Inverse PDE Problems . . . . . 20
2.2.2 The Picard Criterion and Degree of Ill-Posedness . . . . . 21
2.3 Tikhonov Regularization . . . . . 22
2.4 Optimal Convergence Rates and Conditional Stability . . . . . 24
2.4.1 Source Conditions and Conditional Stability . . . . . 24
2.4.2 Optimal A-Priori Parameter Choice . . . . . 24

**3 Neural Networks and the Hermite Variational PINN** **26**

3.1 Tanh Feedforward Networks and Default Initialization . . . . . 27
3.2 Universal Approximation Theory for tanh Networks . . . . . 28
3.3 Physics-Informed Neural Networks and Variational PINNs . . . . . 28
3.3.1 Strong Form PINNs and Variational PINNs . . . . . 28
3.3.2 Variational PINNs . . . . . 28
3.3.3 Trial Source and Trial State . . . . . 29
3.3.4 The Continuous and Discrete Losses . . . . . 30
3.3.5 Bilevel Structure: Source Elimination . . . . . 30

**4 A Hermite Variational PINN for the Inverse Heat Source Problem on $\mathbb{R}$: Analysis and Numerical Experiments 32**
4.1 Introduction . . . 33
4.1.1 Literature Review . . . 33
4.1.2 Research Questions . . . 36
4.1.3 Problem Formulation . . . 38
4.1.4 Nonlinear Extension . . . 38
4.1.5 Anti Inverse Crime . . . 39
4.2 The Hermite V-PINN Algorithm: Linear and Nonlinear Implementations . . . 40
4.2.1 Architecture . . . 40
4.2.2 The Hybrid Discrete Loss . . . 41
4.2.3 The Hermite to Sobolev Norm Bound . . . 44
4.2.4 Algorithm 1: Linear Training . . . 46
4.2.5 Algorithm 2: Nonlinear Training . . . 46
4.3 The Forward Operator: Analysis and Spectral Properties . . . 48
4.3.1 Definition and Its Properties . . . 48
4.3.2 The Smoothing Property $A : H^s(\mathbb{R}) \to H^{s+2}(\mathbb{R})$ . . . 49
4.3.3 Spectral Structure . . . 51
4.3.4 From Extremizers to Weyl Sequences . . . 51
4.3.5 The Multiplier Inequality . . . 52
4.4 Sharp Hölder Stability and Minimax Optimality . . . 53
4.4.1 The Sharp Interpolation Inequality . . . 54
4.4.2 Sharp Hölder Stability . . . 55
4.4.3 Modulus of Continuity and Minimax Optimality . . . 56
4.4.4 Several Observation Times (an Exact $\sqrt{J}$ Law) . . . 64
4.4.5 The Hermite V-PINN as Hilbert Scale Regularization . . . 66
4.4.6 A-Posteriori Control of the Hypotheses . . . 69
4.4.7 On the Choice of Regularization Exponent . . . 70
4.5 Numerical Verification . . . 72
4.5.1 Experimental Setup . . . 72
4.5.2 Forward Solver Validation . . . 73
4.5.3 Noise Free Baselines . . . 74
4.5.4 Empirical Hölder Stability Under Noise . . . 77
4.5.5 Nonlinear Variant . . . 82
4.5.6 Verification of the Statements . . . 83

**5 Comparison with Two Classical Methods 92**
5.1 Mollification with a Dirichlet Kernel . . . 93
5.1.1 Problem Formulation . . . 93
5.1.2 Discrete Realisation . . . 94
5.2 Adjoint-State Inversion with a Sobolev Tikhonov Penalty . . . 96
5.2.1 Problem Formulation . . . 96
5.3 Experiments . . . 99
5.3.1 Noise Sweep . . . 99

5.3.2 Sparse Sensors . . . 100
5.3.3 Partial Domain Observation . . . 101
5.3.4 Unknown Diffusivity . . . 101
5.3.5 Multiple Time Measurements . . . 102
5.3.6 Non Smooth Source . . . 103
5.3.7 High Noise . . . 103
5.3.8 Experiments Unique to the Mollification Study . . . 104
5.3.9 Experiments Unique to the Adjoint Study . . . 106
5.4 Synthesis . . . 108

**References** **111**

**Appendix A Notations** **116**

**Appendix B Figures** **121**
B.1 Experiment 0: Standard Noise Sweep . . . 121
B.2 Experiment 1: Sparse Sensors . . . 122
B.3 Experiment 2: Partial-Domain Observation . . . 123
B.4 Experiment 3: Unknown Diffusivity . . . 125
B.5 Experiment 4: Multiple Time Measurements . . . 126
B.6 Experiment 5: Non-Smooth Source . . . 127
B.7 Experiment 6: High-Noise Regime . . . 127
B.8 Mollification-Specific Experiments . . . 128
B.9 Adjoint-Specific Experiments . . . 129

# Chapter 1

# Analytical, Spectral, and Numerical Foundations

*Mathematics knowns no races or geographic boundaries, for Mathematics the cultural world is one country*

David Hilbert
*International Congress of Mathematicians (ICM)* (1928)

*In the first chapter we introduce PDEs and their classifications. we focus on the parabolic PDE, whose forward problem is well posed while its inverse is not. This problem is set up in Remark 1.3, and then Proposition 1.4 allows us to derive the existence and uniqueness of solutions to (1.3), without which the operator $A\varphi := v(\cdot, 1)$ is meaningless. Section 1.3.1 discusses the spaces. We deal with noisy measurements. So, we are not able to achieve stability without assuming that the source is smooth. In order to do so, we consider the space $\mathcal{W}_{\ell,E}$ introduced in Definition 1.8, whose parameters $\ell$ and $E$ are the ones appearing in Theorem 4.11. Fourier transform is introduced in Definition 1.9 with the symmetric definition such that Plancherel formula (1.9) does not have a $2\pi$, otherwise it will appear in front of each of the numbers found in Chapter 4. Hermite functions (Section 1.3.3), however, form the basis used for the expansion of the source and are orthonormal and hence, the source can be determined using a linear system (Proposition 3.7) as opposed to an optimization approach. Two important points regarding the heat semigroup are that it is always not expensive (Proposition 1.17), as well as Duhamel's principle (Proposition 1.18), which implies that any errors made by the network will be captured by the semigroup, hence a small training error results in small final time errors. As mentioned in Proposition 1.19, for the Fourier transform the equation splits into an ODE for each frequency $\xi$ and its solution for $t = 1$ corresponds to the operation of multiplication on $m(\xi) = (1 - e^{-\xi^2})/\xi^2$. This multiplier takes a value 1 when $\xi = 0$, and decays at large frequencies like $\xi^{-2}$. Thus, the inverse operation corresponds to division by a number close to zero, which amplifies high-frequency noise without bound. in Section 1.4.2, the data is generated using the Crank Nicolson approach; and the quadrature functions of two Gauss points are given in Section 1.4.3.*

## 1.1 Introduction to Partial Differential Equations

**Definition 1.1.** Suppose that $\Omega$ is an open subset of $\mathbb{R}^n$ and let $k$ be a natural number. Consider the notion of multi-index notation $\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}$, where $\alpha \in \mathbb{N}_0^n$; then PDEs of order $k$ can be written as

$$F\big(x,\, u(x),\, Du(x),\, \ldots,\, D^k u(x)\big) = 0, \qquad x \in \Omega. \tag{1.1}$$

In this context, $u$ is an unknown function from the domain $\Omega$ to the real numbers, while $D^j u$ represents all partial derivatives of order $j$. [81, §1.1] [70, §1.1]

Throughout this dissertation, the equation of our interest will be $\partial_t v - \partial_{xx} v = \varphi$, and it is linear in the state and its derivatives.

**Definition 1.2.** For a second-order linear scalar PDE in two independent variables $x$ and $y$,

[59, §1.2] $$A(x,y)\, u_{xx} + 2B(x,y)\, u_{xy} + C(x,y)\, u_{yy} + (\text{lower-order terms}) = 0. \tag{1.2}$$

$\Delta$ is referred to as a discriminant, and it is represented as $\Delta := B^2 - AC$. The point by point classification of the equation is as follows:

- *elliptic* for $\Delta < 0$, where the canonical form is Laplace's equation;
- *parabolic* for $\Delta = 0$, where the canonical form is the heat equation; and
- *hyperbolic* for $\Delta > 0$, where the canonical form is the wave equation.

[103] In the case of $n$-dimensional equations, classification is performed by the signature of the principal-symbol matrix $(a_{ij}(x))_{i,j=1}^n$. This is the higher-dimensional analogue of the two-variable discriminant criterion; see [132, §1.6].

**Remark 1.3.** This investigation deals exclusively with the parabolic case. The dimension of the space is $n = 1$, the spatial domain is $\Omega = \mathbb{R}$, and the temporal domain is $I = (0, 1]$. The source function $\varphi : \mathbb{R} \to \mathbb{R}$ is independent of time, a condition that characterizes the inverse problem of Chapter 4 (recovery of $\varphi$ from a single terminal-time snapshot of $v$).

## 1.2 The Heat Equation with a Time Independent Source

All that follows is a consequence of the solution to one Cauchy problem. It is given below in its exact form as it appears in Chapter 4, equation (4.1),

$$\partial_t v(x,t) - \partial_{xx} v(x,t) = \varphi(x), \quad v(x,0) = 0, \quad \lim_{|x|\to\infty} v(x,t) = 0, \qquad (x,t) \in \mathbb{R} \times (0,1]. \tag{1.3}$$

In this particular problem there is a unique solution and it is well posed, and it is this that makes sure that the forward operator is well defined. Here are three key things about

the solution. Firstly, the source $\varphi$ does not have any dependence on time, and therefore it gives significance to recovering it from a single measurement. The initial datum vanishes, and Chapter 4 sets this in concrete in its trial state

$$v_\theta(x,t) := t\, e^{-\alpha_v x^2}\, \mathcal{N}_\theta(x,t),$$

where the factor $t$ is already present, rather than being forced through a penalty term. Furthermore, the whole space $\mathbb{R}$ itself is the domain, which means that there is no truncation-related boundary influence.

**Proposition 1.4.** *Given any $\varphi \in L^2(\mathbb{R})$, there exists a unique solution $v$ of (1.3), such that*

$$v \in C\big([0,1]; L^2(\mathbb{R})\big) \cap L^2\big(0,1; H^1(\mathbb{R})\big).$$

**Proof.** In relation to the Gelfand triple $H^1(\mathbb{R}) \subset L^2(\mathbb{R}) \subset H^{-1}(\mathbb{R})$, define the bilinear form

$$a(u,v) = \int_{\mathbb{R}} u'(x)\, v'(x)\, dx.$$

The bilinear form $a$ is bounded by

$$|a(u,v)| \le \|u\|_{H^1(\mathbb{R})} \cdot \|v\|_{H^1(\mathbb{R})},$$

and satisfies the Gårding inequality [97]

$$a(u,u) = \|u\|^2_{H^1(\mathbb{R})} - \|u\|^2_{L^2(\mathbb{R})} \ge c\|u\|^2_{H^1(\mathbb{R})} - d\|u\|^2_{L^2(\mathbb{R})},$$

with $c = d = 1$, irrespective of whether the domain is bounded. Since $a$ does not depend on $t$, it follows that conditions (H2*) of Corollary 23.26 in [97, Ch. 23] are satisfied trivially. regarding $\varphi$ as an input which does not depend on time and is an element of $L^2(0,1; H^{-1}(\mathbb{R}))$, Corollary 23.26 gives a unique

$$v \in W^1_2\big(0,1; H^1(\mathbb{R}), L^2(\mathbb{R})\big).$$

From Proposition 23.23(ii) in the same book, we obtain that this space is continuously embedded into $C([0,1]; L^2(\mathbb{R}))$, hence

$$v \in C([0,1]; L^2(\mathbb{R})) \cap L^2(0,1; H^1(\mathbb{R})).$$

□

## 1.3 Function Spaces and Fourier-Hermite Analysis

Here we will discuss the spaces in which all the results stated in Chapter 4 are stated. The spaces and their norms in which all the results hold need to be fixed before the operator, source condition, and spectral construction.

### *1.3.1 Lebesgue, Schwartz, and Sobolev Spaces*

Data and operator are in $L^2(\mathbb{R})$. The Schwartz space is suitable in terms of Fourier transform and also provides a space where the smooth source $\varphi^\dagger$ lives. Its dual space $\mathcal{S}'(\mathbb{R})$ enables us to consider Sobolev spaces with negative and fractional order. The $H^s(\mathbb{R})$ norm defined through Fourier and its weight $(1+\xi^2)^s$ plays an essential role because it defines the Sobolev scale; in particular, in the special case $s = \ell$, we get $H^\ell(\mathbb{R})$, which serves as the space the source ball $\mathcal{W}_{\ell,E}$ of Definition 1.8 is built on.

**Definition 1.5.** The Lebesgue space $L^2(\mathbb{R})$ comprises equivalence classes of measurable functions $f : \mathbb{R} \to \mathbb{R}$, the equivalence relation being that two functions are equal up to a set of measure zero. For the $L^2(\mathbb{R})$ norm, we have the definition

$$\|f\|_{L^2(\mathbb{R})} := \left(\int_{\mathbb{R}} |f(x)|^2 \, \mathrm{d}x\right)^{1/2} < \infty. \tag{1.4}$$

The inner product is defined as $\langle f, g\rangle := \int_{\mathbb{R}} f(x)\,\overline{g(x)}\,\mathrm{d}x$ (conjugate-linear in the second coordinate). $L^2(\mathbb{R})$ is a separable Hilbert space [79, In terms of Hilbert Space see §5.5,§6.1], Seperability also can be proven by

$$\text{Find a countable set } D \subset L^2(\mathbb{R}) \text{ such that } \overline{D} = L^2(\mathbb{R}).$$

Also see [69, Thm 4.13]

**Definition 1.6.** The *Schwartz space* [80, §2.2],[146] is

$$\mathcal{S}(\mathbb{R}) := \left\{ f \in C^\infty(\mathbb{R}) : \sup_{x\in\mathbb{R}} |x^a \partial^b f(x)| < \infty \ \text{ for all } a, b \in \mathbb{N}_0 \right\}, \tag{1.5}$$

The set of functions in the space is equipped with the semi-norms $p_{a,b}(f) := \sup_{x\in\mathbb{R}} |x^a \partial^b f(x)|$, and it becomes a Fréchet space, being dense in $L^2(\mathbb{R})$ [79, Proposition 8.17]. The dual space to the Schwartz space $\mathcal{S}(\mathbb{R})$ is the space of tempered distributions, $\mathcal{S}'(\mathbb{R})$ [127, §V.3, p. 133].

Our Sobolev spaces are based on the Fourier transform, according to the convention in Definition 1.9.

**Definition 1.7.** For any real number $s$, the fractional Sobolev space $H^s(\mathbb{R})$ is the collection of all tempered distributions $f \in \mathcal{S}'(\mathbb{R})$ [105] for which

$$\|f\|^2_{H^s(\mathbb{R})} := \int_{\mathbb{R}} (1+\xi^2)^s |\hat{f}(\xi)|^2 \, \mathrm{d}\xi < \infty. \tag{1.6}$$

When $s = k$, where $k$ is from the set of natural numbers including 0, the space is $W^{k,2}(\mathbb{R})$. $H^0(\mathbb{R})$ is equal to $L^2(\mathbb{R})$; see [70, §5.2]. The Fourier characterization of the space $W^{s,2}(\mathbb{R}^n)$ is stated in [2, §7.62]. For integer values of $k$, the space $W^{k,2}$ is a separable Hilbert space by [2, Thm. 3.6].

Afterward, We discuss the a-priori class $\mathcal{W}_{\ell,E}$, which is the class of assumed true sources. Further down, we will realize that although the operator $A$ is injective, the inverse operator $A^{-1}$ is unbounded. What this implies is that when $A^{-1}$ is regarded as a map over $L^2(\mathbb{R})$, the inverse problem is entirely ill-posed since arbitrarily large perturbations in the source could produce arbitrarily small perturbations in the data. There is no meaningful bound for the entire $L^2(\mathbb{R})$ space. The only thing we can do is limit the sources to a compact like set of functions; this is exactly $\mathcal{W}_{\ell,E}$, which lives inside $H^\ell(\mathbb{R})$.

$$\mathcal{W}_{\ell,E} \subset H^\ell(\mathbb{R}) \subset L^2(\mathbb{R}).$$

**Definition 1.8** (Source class)**.** For $\ell \geq 1$ and $E > 0$, the *source class* is

$$\mathcal{W}_{\ell,E} := \left\{ \varphi \in H^\ell(\mathbb{R}) : \|\varphi\|^\ell_H(\mathbb{R}) \leq E \right\}. \tag{1.7}$$

$\ell$ and $E$ are both constants in Chapters 2 to 4. $\ell$ is responsible for the rate $\delta^{\ell/(\ell+2)}$, while $E$ is found in the constant $(2E)^{2/(\ell+2)}$ of Chapter 4, Theorem 4.11.

### *1.3.2 The Fourier Transform*

In this section, We address the topic of the Fourier Transform which plays a significant role in Chapter 4. The Fourier Transform simplifies the difficult integral operator to a simple multiplication. Some questions include: Is the operator $A$ injective? To what extent is it ill-posed? What is the stability constant exactly? These are just some of the questions which should be addressed in Chapter 4 and are difficult to solve in physical space.

The selected convention is the symmetric normalization. The reason is that one of the main result of Chapter 4 is a precise Hölder stability estimate with an explicit constant. This constant is obtained via multiple passes from the physical side (function norms) to the Fourier side (integral estimates with the symbol of the operator), and back, with each pass utilizing Plancherel. It ensures that the Plancherel formula does not involve any additional constants, and therefore retains all the constants present in Chapter 4 without any further terms such as $2\pi$.

**Definition 1.9.** If a function $f$ belongs to the Schwartz class $\mathcal{S}(\mathbb{R})$, then its Fourier transform is given by [146]:

$$\hat{f}(\xi) := \mathcal{F}f := (2\pi)^{-1/2} \int_{\mathbb{R}} f(x)\, e^{-i\xi x}\, \mathrm{d}x, \qquad \xi \in \mathbb{R}, \tag{1.8}$$

with inverse $\mathcal{F}^{-1}[g](x) = (2\pi)^{-1/2} \int_{\mathbb{R}} g(\xi) e^{i\xi x}\, \mathrm{d}\xi$; both extend by continuity to $L^2(\mathbb{R})$ and by duality to $\mathcal{S}'(\mathbb{R})$ [126, §IX.1, Thms. IX.1–IX.2, pp. 3–6; §IX.2, Thm. IX.6, p. 10].

**Proposition 1.10.** *According to the convention* (1.8)*, we have*:

(i) (Topological isomorphism $\mathcal{S} \to \mathcal{S}$) *The Fourier transform $\mathcal{F}$ is a topological isomorphism of the Schwartz space $\mathcal{S}(\mathbb{R})$ onto itself, with inverse $\mathcal{F}^{-1}$.*

(ii) (Plancherel identity) *For all functions $f$ belonging to $L^2(\mathbb{R})$, the $L^2(\mathbb{R})$ norm of $f$ is equal to the $L^2(\mathbb{R})$ norm of its Fourier transform [70, §4.3.1] :*

$$\|f\|_{L^2(\mathbb{R})} = \|\hat{f}\|_{L^2(\mathbb{R})}. \tag{1.9}$$

(iii) (Differentiation rule) *If $f$ is a function belonging to the Schwartz class $\mathcal{S}(\mathbb{R})$ and $k$ is any nonnegative integer, then*

$$\widehat{\partial_x^k f}(\xi) = (i\xi)^k \hat{f}(\xi) \tag{1.10}$$

*holds true.*

**Proof.** Every item above is classical. The isomorphism is due to the Fourier inversion formula on Schwartz space, as proved in [126, Thm. IX.1, pp. 3–4]. The Plancherel formula is given by [126, Thm. IX.6, p. 10] in the same book. The differentiation formula is deduced using integration by parts in the Schwartz space $\mathcal{S}(\mathbb{R})$, which holds for the dual $\mathcal{S}'(\mathbb{R})$ too by duality. □

In such an expression, $H^s(\mathbb{R})$ is the pre-image of $L^2(\mathbb{R})$ with respect to the weight $(1+\xi^2)^{s/2}$, the connection between the theory of multipliers that Chapter 4 uses and $A$ in Theorem 4.4.

### *1.3.3 Hermite Functions and the Hermite-Sobolev Structure*

In Chapter 4, the source is expanded not in terms of a grid but in terms of scaled Hermite functions. Hermite functions are the eigenfunctions of the harmonic oscillator and constitute an orthonormal basis of $L^2(\mathbb{R})$ with Gaussian decay. Their values at the residual nodes provide the design matrix $\Phi$, whereas the oscillator eigenvalues $2j+1$, raised to the penalty order $\ell$, give rise to the mass matrix $\Lambda$. These two matrices determine the contribution of the source coefficient vector $c$ to all terms of the discrete loss function (3.4) and the regularised Gram matrix (4.17), which are introduced in Chapter 4.

**Definition 1.11.** Physicists' Hermite polynomials are expressed in terms of

$$H_j(x) := (-1)^j e^{x^2} \frac{\mathrm{d}^j}{\mathrm{d}x^j} e^{-x^2}, \qquad j \in \mathbb{N}_0.$$

And their associated Hermite functions in terms of

$$\psi_j(x) := \frac{1}{\sqrt{2^j j! \sqrt{\pi}}} H_j(x)\, e^{-x^2/2}, \qquad j \in \mathbb{N}_0. \tag{1.11}$$

The family $\{\psi_j\}_{j\geq 0}$ is an orthonormal basis of $L^2(\mathbb{R})$ [137, §1.1], [133, Ch. V, §5.5 (Orthonormality),§5.7 (Completeness)].

**Proposition 1.12.** *The operator $B := -\partial_{xx} + x^2$ is defined with respect to $L^2(\mathbb{R})$, its domain being*

$$\mathrm{Dom}(B) := \Big\{ f = \sum_{j=0}^{\infty} c_j \psi_j \in L^2(\mathbb{R}) : \sum_{j=0}^{\infty} (2j+1)^2 |c_j|^2 < \infty \Big\}.$$

*$B$ is self-adjoint, strictly positive, and its spectrum is $\sigma(B) = \{2j+1 : j \in \mathbb{N}_0\}$. Its eigenfunctions are the Hermite functions*:

$$B\psi_j = (2j+1)\psi_j, \qquad j \in \mathbb{N}_0. \tag{1.12}$$

**Proof.** The operators $a^{\pm} := \mp\partial_x + x$ factor $B$ as $a^+a^- + 1$, and they operate on $\psi_j$ according to $a^-\psi_j = \sqrt{2j}\,\psi_{j-1}$ and $a^+\psi_j = \sqrt{2(j+1)}\,\psi_{j+1}$, which gives (1.12) at once. Since $\{\psi_j\}$ forms an orthonormal basis for $L^2(\mathbb{R})$, $B$ is unitarily equivalent to multiplication by $2j+1$ on $\ell^2(\mathbb{N}_0)$. Hence $B$ is a self-adjoint operator that is strictly positive and whose spectrum is discrete. □

The rescaled one is the actual basis used in Chapter 4, and it is here we will make mention of it alongside the oscillator.

**Definition 1.13.** For a positive scale $\sigma$ and a truncation order $N_\varphi \in \mathbb{N}$, the scaled Hermite functions are

$$\psi_j^\sigma(x) := \sigma^{-1/2} \psi_j(x/\sigma), \qquad j \in \mathbb{N}_0. \tag{1.13}$$

$\{\psi_j^\sigma\}_{j\geq 0}$ is an orthonormal basis of $L^2(\mathbb{R})$ [51], and, scaling the result of Proposition 1.12, we obtain that the scaled harmonic oscillator $L_\sigma := -\sigma^2\partial_{xx} + \sigma^{-2}x^2$ possesses the eigenfunctions $\psi_j^\sigma$ with the eigenvalues $2j+1$:

$$L_\sigma \psi_j^\sigma = (2j+1)\psi_j^\sigma, \qquad j \in \mathbb{N}_0.$$

**Definition 1.14.** Choose arbitrary $\sigma > 0$, $N_\varphi \in \mathbb{N}$, regularity $\ell \geq 1$, and the Gauss-Hermite quadrature nodes $\{x_i^H\}$ and Gauss-Legendre quadrature nodes $\{t_k^L\}$. This generates a tensor product grid of residual nodes $\{(x_i^H, t_k^L)\}$ having size $M_r = N_{\mathrm{GH}} N_{\mathrm{GL}}$.

(i) The *Hermite-Sobolev mass matrix* is

$$\Lambda := \operatorname{diag}\big((2j+1)^{\ell}\big)_{j=0}^{N_\varphi-1} \in \mathbb{R}^{N_\varphi \times N_\varphi}. \tag{1.14}$$

For $c \in \mathbb{R}^{N_\varphi}$, the squared mass norm becomes $\|c\|_\Lambda^2 = c^\top \Lambda c$. The design matrix according to the Hermite basis can be written as

$$\Phi_{(i,k),j} := \psi_j^\sigma(x_i^H), \qquad \Phi \in \mathbb{R}^{M_r \times N_\varphi}, \tag{1.15}$$

where $j$ ranges from 0 to $N_\varphi - 1$.

(ii) There exists $C_{\sigma,\ell} > 0$ such that for all $c \in \mathbb{R}^{N_\varphi}$,

$$\|\varphi^c\|_H^\ell(\mathbb{R}) \le C_{\sigma,\ell}\|c\|_\Lambda, \qquad \varphi^c := \sum_{j=0}^{N_\varphi-1} c_j \psi_j^\sigma. \tag{1.16}$$

The explicit value $C_{2,2} \le \sqrt{105}/8 \approx 1.281$ is computed in Chapter 4, Lemma 4.2.

This is precisely Chapter 4, equation (4.12): $\Lambda$ carries the Tikhonov regularization term $\lambda\|c\|_\Lambda^2$ of the discrete loss (4.18), and it is also present in the regularised Gram matrix discussed in Chapter 4, equation (4.17).

Even a simple one-line bound for the tail estimation explains why $N_\varphi$ modes suffice: assuming that $L_\sigma^{r/2} f$ belongs to $L^2(\mathbb{R})$ and $f_{N_\varphi}$ denotes the projection of $f$ on the first $N_\varphi$ basis functions, i.e. $f_{N_\varphi} := \sum_{j=0}^{N_\varphi-1} \langle f, \psi_j^\sigma\rangle \psi_j^\sigma$, with Parseval applied to $L_\sigma \psi_j^\sigma = (2j+1)\psi_j^\sigma$ (Proposition 1.12; cf. [137, §1.1]) we obtain the estimate

$$\|f - f_{N_\varphi}\|_L^2(\mathbb{R}) \le (2N_\varphi + 1)^{-r/2} \|L_\sigma^{r/2} f\|_L^2(\mathbb{R}) \quad \text{(cf. [40, Lem. 2.1])},$$

so a smooth (Schwartz) source is approximated at a spectral rate.

### 1.3.4 Bochner Spaces and the Heat Semigroup

In this part of the dissertation, We discuss Bochner spaces, the heat semigroup, contractivity of the heat semigroup, and Duhamel's principle. $A\varphi = v(\cdot, 1)$ is our forward operator. It refers to the value of the state at $t = 1$. For this to make sense and to be defined correctly, it is necessary to have the state at any time $t$ being an element of $L^2(\mathbb{R})$. It is a well-defined $L^2(\mathbb{R})$ function at all times, and computing its value when $t = 1$ is meaningful. This is assured by the existence of the Bochner space $C([0,1]; L^2(\mathbb{R}))$. Otherwise, $v(\cdot, 1)$ is not meaningful, and the forward operator will not exist. Another critical element is the heat semigroup. The symbol $m(\xi) = (1 - e^{-\xi^2})/\xi^2$ that Chapter 4 revolves around comes directly from the integration of the semigroup multiplier $e^{-t\xi^2}$ with respect to $t \in [0,1]$. In the absence of the semigroup, there can be no Fourier multiplier; in the absence of the Fourier multiplier, there will be no Fourier expression for $A$, and the theory of injectivity, ill-posedness, and the stability constant fails.

Moreover, this chapter explains the contractivity of the heat semigroup and Duhamel's principle. This is because contractivity implies that the heat semigroup will never amplify. What goes into the semigroup comes out the same size or smaller. There are two results from this:

1. The operator norm of the semigroup never exceeds 1 in $L^2(\mathbb{R})$ at any point in time.

2. The Error Control Bridge via Duhamel (the most significant implication). The neural net from Chapter 4 does not provide an exact solution to the PDE but rather leaves us with a residual $R$. The question is: how wrong is the end result $v(\cdot\,,1)$ due to that residual? According to Duhamel, the state at time 1 is

$$v(\cdot\,,1) = \int_0^1 e^{(1-s)\Delta}\big(\varphi + R(\cdot\,,s)\big)\,\mathrm{d}s.$$

The residual $R$ sits inside the semigroup. The semigroup being a contraction means that it cannot amplify $R$. Hence,

$$\text{small PDE residual} \implies \text{small terminal error at } t = 1.$$

**Definition 1.15.** Let $(X, \|\cdot\|_X)$ be a Banach space and $I = (0,1)$ an interval. The Bochner-integrable functions, $L^p(I;X)$ for $1 \le p < \infty$, are defined to be the set of all strongly measurable $u : I \to X$ satisfying

$$\|u\|_{L^p(I;X)} := \Big(\int_I \|u(t)\|_X^p \,\mathrm{d}t\Big)^{1/p} < \infty.$$

Also, we need to point out that $C([0,1];X)$ is endowed with the supremum norm $\|u\|_{C([0,1];X)} := \sup_{t\in[0,1]}\|u(t)\|_X$ [70, §5.9.2, pp. 285–288].

**Definition 1.16.** We will use the following definition of the *heat semigroup* $\{e^{t\Delta}\}_{t\ge0}$ acting on $L^2(\mathbb{R})$, given for an $f \in L^2(\mathbb{R})$ at a $t \ge 0$ through its Fourier multiplier,

$$\widehat{e^{t\Delta}f}(\xi) := e^{-t\xi^2}\hat{f}(\xi), \qquad \xi \in \mathbb{R}. \tag{1.17}$$

For $t > 0$, this leads to the heat-kernel convolution of $f$,

$$(e^{t\Delta}f)(x) = \int_{\mathbb{R}} G(x-y,t)\,f(y)\,\mathrm{d}y, \qquad G(x,t) := \frac{e^{-x^2/(4t)}}{\sqrt{4\pi t}}. \tag{1.18}$$

$\{e^{t\Delta}\}_{t\ge0}$ constitutes a strongly continuous semigroup of contractions on $L^2(\mathbb{R})$ or, in other words, a $C_0$-semigroup; see [113, §II, 2.13] and [119, §7.2, Theorem 2.5 & Remark 2.9]. The heat kernel representation (1.18) for $t > 0$ follows from [113, §II, 2.13, (2.10)] and [70, §2.3.a, Thm. 1].

**Proposition 1.17.** *For all $f \in L^2(\mathbb{R})$ and all $t \ge 0$, the heat semigroup is a contraction operator on $L^2(\mathbb{R})$:*

$$\|e^{t\Delta}f\|_L^2(\mathbb{R}) \le \|f\|_L^2(\mathbb{R}). \tag{1.19}$$

**Proof.** Using the Fourier formulation (1.17) and the result from Plancherel (1.9), we have the

following inequality: $\|e^{t\Delta} f\|_L^2(\mathbb{R})^2 = \int_{\mathbb{R}} e^{-2t\xi^2} |\hat{f}(\xi)|^2 \, \mathrm{d}\xi \leq \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \, \mathrm{d}\xi = \|f\|_L^2(\mathbb{R})^2$, which is the contraction property of the heat semigroup (cf. [113, §II, 2.13]) □

**Proposition 1.18.** *Let us consider $\varphi$ belonging to $L^2(\mathbb{R})$ and $R$ belonging to $L^1(0,1;L^2(\mathbb{R}))$. The only solution in a mild sense of*

$$\partial_t v - \partial_{xx} v = \varphi + R(\,\cdot\,,t), \qquad v(\,\cdot\,,0) = 0,$$

*is given by*

$$v(\,\cdot\,,t) = \int_0^t e^{(t-s)\Delta}\big(\varphi + R(\,\cdot\,,s)\big)\,\mathrm{d}s, \qquad t \in [0,1]. \tag{1.20}$$

**Proof.** Using the transform in $x$, the problem becomes the scalar ODE $\hat{v}_t + \xi^2 \hat{v} = \hat{\varphi}(\xi) + \hat{R}(\xi,t)$, with $\hat{v}(\xi,0) = 0$. The solution by variation of constants inverts to the formula (1.20). This is the classical mild solution of the inhomogeneous Cauchy problem, see [119, §4.2, Def. 2.3, p. 106]. □

## 1.4 Analytical, Classical, and Spectral Numerical Methods

We will consider the Fourier multiplier as well as the numerical methods which are used (Gauss-Hermite and Gauss-Legendre quadrature) in this section. For testing our method, we solve the heat equation numerically. An appropriate solver needs to be stable and accurate. Error should be low enough to ensure reliability of data. Among the least expensive solvers, which are unconditionally stable and second order accurate, is the Crank Nicolson scheme. Only Crank Nicolson was used in this problem, although We also explain the FTCS method for comparison purposes. The FTCS method has conditional stability, which cannot be satisfied by the fine mesh necessary in Chapter 4.

The loss function also contains space-time integrals that have to be computed numerically by applying two quadratures: the Gauss-Hermite and Gauss-Legendre quadratures. The tensor product of both quadratures forms the residual node grid in the loss.

### *1.4.1 Fourier Multiplier Representation of the Forward Operator*

In this case, the Fourier transform of the heat equation is considered. This results in an easy ODE in the time variable $t$ for each fixed value of the frequency $\xi$. Solving the ODE and substituting the value $t = 1$, the multiplier $m(\xi)$ becomes equal to $(1 - e^{-\xi^2})/\xi^2$. This symbol is of importance in Section 4.3 and Section 4.4. Its properties reveal immediately that the operator $A$ is bounded, injective, and ill-posed.

**Proposition 1.19.** *For every $\varphi \in L^2(\mathbb{R})$, the mild solution $v$ to (4.1) has the following*

*representation,*

$$\hat{v}(\xi,t) = \hat{\varphi}(\xi)\cdot\frac{1-e^{-\xi^2 t}}{\xi^2} \qquad (\xi\neq 0), \quad \hat{v}(0,t) = t\,\hat{\varphi}(0). \tag{1.21}$$

*Therefore, the operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, given by $A\varphi := v(\,\cdot\,,1)$, has the following feature: its Fourier transform is given by the formula*

$$\widehat{(A\varphi)}(\xi) = m(\xi)\,\hat{\varphi}(\xi), \quad \textit{where} \quad m(\xi) := \frac{1-e^{-\xi^2}}{\xi^2}\ (\xi\neq 0),\ m(0) := 1. \tag{1.22}$$

**Proof.** Apply the Fourier transform of Definition 1.9 to (4.1), then use the differentiation rule (1.10). This gives

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{v}(\xi,t) + \xi^2\hat{v}(\xi,t) = \hat{\varphi}(\xi), \qquad \hat{v}(\xi,0) = 0.$$

Fixing $\xi$, this is a linear ODE of order one in $t$ with the constant coefficient $\xi^2$ and constant forcing. Multiplying by the integrating factor $e^{\xi^2 t}$ and setting $\xi\neq 0$ gives

$$\hat{v}(\xi,t) = \hat{\varphi}(\xi)\,e^{-\xi^2 t}\int_0^t e^{\xi^2 s}\,\mathrm{d}s = \hat{\varphi}(\xi)\,\frac{1-e^{-\xi^2 t}}{\xi^2}.$$

For $\xi = 0$, we have $\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{v}(0,t) = \hat{\varphi}(0)$, thus $\hat{v}(0,t) = t\,\hat{\varphi}(0)$. Choosing $t = 1$, and using $m(0) = \lim_{\xi\to 0}(1-e^{-\xi^2})/\xi^2 = 1$, we get (1.22). Moreover, because $|m(\xi)|\le 1$, by Plancherel (1.9) we have $\|A\varphi\|_L^2(\mathbb{R}) \le \|\varphi\|_L^2(\mathbb{R})$, and therefore $A$ is a bounded operator on $L^2(\mathbb{R})$. □

### *1.4.2 Finite Difference Methods*

Both schemes (FTCS, Crank Nicolson) follow [96, Ch. 9].

**FTCS (Forward Time, Centred Space):**

$$\frac{u_j^{n+1}-u_j^n}{\Delta t} = \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta x^2} + f_j^n, \qquad r := \frac{\Delta t}{\Delta x^2}. \tag{1.23}$$

With the ansatz $u_j^n = g^n\exp(i\xi j\Delta x)$, we get $g(\xi) = 1 - 4r\sin^2(\xi\Delta x/2)$, so that $|g(\xi)|\le 1$ for all $\xi$ if and only if

$$r \le \tfrac{1}{2}, \tag{1.24}$$

the classical condition of stability of the explicit Euler scheme [96, Ch. 9, §9.3, Example 9.2].

**Crank Nicolson ($\theta$-method with $\theta = 1/2$):**

$$\frac{u_j^{n+1}-u_j^n}{\Delta t} = \frac{1}{2}\left(\frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta x^2} + \frac{u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}}{\Delta x^2}\right) + f_j^{n+1/2}. \tag{1.25}$$

The trapezoidal discretisation of the centred second difference is second order in both $\Delta t$ and $\Delta x$. The amplification factor of this finite difference scheme is $g(\xi) = \frac{1-2r\sin^2(\xi\Delta x/2)}{1+2r\sin^2(\xi\Delta x/2)}$, so $|g(\xi)|\le 1$ for any $r > 0$ (strict for $\xi\neq 0$). This is the scheme Chapter 4 uses to generate data [96, Ch. 9, §9.1, §9.6].

**Proposition 1.20** (Lax–Richtmyer equivalence)**.** *Stability is necessary and sufficient for convergence when the finite difference method is consistent for a linear well-posed initial value problem.*

**Proof.** See [128, Ch. III, §5]. □

**Remark 1.21.** In Chapter 4, Section 4.1.5, for the reference data $g^{\delta}$ the equation (1.25) is used with $N_x = 10\,000$ and $N_t = 5\,000$ on the domain $[-30, 30]$. It is used for the problem where the diffusion coefficient is perturbed, $a(x) = 1 + \varepsilon e^{-x^2/4}$, to avoid the inverse crime. Unconditional stability implies that these step sizes are not constrained by the CFL condition (1.24), which is imposed on FTCS. Then, by using stability along with consistency, Proposition 1.20 ensures convergence when $\Delta t$ and $\Delta x$ tend to zero.

### *1.4.3 Gauss-Hermite and Gauss-Legendre Quadrature*

Integrals of loss functions in Chapter 4 are evaluated through Gauss quadrature: Gauss-Hermite is applied to space in the domain $\mathbb{R}$, while Gauss-Legendre is used for time in the domain $[0, 1]$, and their product constitutes a tensor product grid that is applied for evaluating the weak-form residual, recorded here with the rates that make the quadrature error negligible.

**Definition 1.22.** The Gauss-Hermite quadrature formula for integration over the real line employs the node points $\{x_i^H\}_{i=1}^{N_{\mathrm{GH}}}$ (zeros of $H_{N_{\mathrm{GH}}}$) and positive weights $\{w_i^H\}_{i=1}^{N_{\mathrm{GH}}}$ satisfying the relation

$$\int_{\mathbb{R}} f(x)\, e^{-x^2}\, \mathrm{d}x = \sum_{i=1}^{N_{\mathrm{GH}}} w_i^H f(x_i^H) \tag{1.26}$$

for all polynomials $f$ with degrees up to $2N_{\mathrm{GH}} - 1$. See [60, §3.4, p.174]; for analytic integrands on the real axis, the convergence is root-exponential in $N_{\mathrm{GH}}$ [141, §5, p. 10]. In Chapter 4, the Gaussian/Hermite nature of the trial functions makes it possible to write the quadrature terms with such weights.

**Definition 1.23.** For the $N_{\mathrm{GL}}$-point Gauss-Legendre quadrature on the interval $[0, 1]$, the nodes $t_k^L$ and positive weights $w_k^L$ (for $k = 1, \ldots, N_{\mathrm{GL}}$) are used to achieve the following result for any polynomial $g$ of degree up to $2N_{\mathrm{GL}} - 1$:

$$\int_0^1 g(t)\, \mathrm{d}t = \sum_{k=1}^{N_{\mathrm{GL}}} w_k^L g(t_k^L). \tag{1.27}$$

See [60, §2.7, pp. 73–76]; for integrands analytic in a Bernstein ellipse containing $[0, 1]$ the convergence is geometric in $N_{\mathrm{GL}}$ [140, Ch. 19, Thm. 19.3].

## 1.5 From Classical Methods to Modern Methods

Classical methods such as finite difference (FDM), finite element (FEM), and spectral methods remain the standard approaches to analyse computational PDEs due to their well-founded convergence theory [96, Ch. 9, §§9.1–9.5][37][45]. The Classical methods will not disappear completely in the future because of their proven convergence rates, error bounds, and stability theory (Lax Richtmyer equivalence, Céa's lemma [37], and etc). They are also optimal for their niche. For example, for a well posed PDE on a known, low dimensional geometry, mesh based methods are computationally cheaper, faster to set up, and maybe even more accurate than training a neural network.Neural network training involves randomness (initialisation [82, §2.2], stochastic optimisation [93]), reproducibility and debuggability are harder (Remark 4.27), which matters considerably for safety critical or regulated engineering contexts.

**But** what if we have a high dimensional PDE, where mesh based methods suffer the curse of dimensionality? A method of order $k$ attaining accuracy $\varepsilon$ in $d$ dimensions requires $O(h^{-d}) = O(\varepsilon^{-d/k})$ unknowns, so we have exponential growth with the dimension [1], and it is this that motivated the deep solvers for high-dimensional parabolic problems [67].

What if we need to solve a family of PDEs repeatedly with varying parameters or initial conditions? [116]

What if we are dealing with inverse PDEs or parameter estimation, where modern methods such as PINNs can incorporate data and physics simultaneously without remeshing [29, 121]?

What if we have sparse, noisy, or non uniform sensor data scattered across the domain, not just on the boundary?

What if we have complex or irregular geometries with unknown or evolving boundaries? [25] Learning unknown terms, parameters, or even entire governing equations from data is fundamentally something classical solvers cannot do (they require the PDE to already be known) [3].

So the role of modern methods such as PINNs also cannot be neglected [6], due to the growing volume of messy, real-world data, problems where the physics itself is incomplete, and the curse of dimensionality, which is a mathematical fact, not an engineering limitation.

So the realistic future is probably going to be **hybrid**. Neither can classical methods be considered finished, nor can the role of modern methods be ignored.

The motivation behind the method we have used in this dissertation (Hermite variational PINN) for the inverse source problem, however, is not the curse of dimensionality or having complex or irregular geometries [64]. The main reason is generality and robustness to messier or extended settings, because our method holds up under conditions that some classical methods such as Fourier methods [65, 111] were not built for. For example, our method easily deals with a nonlinear version of the current problem in the dissertation with minor changes (Section 4.1.4), while it is impossible for some classical Fourier methods such as mollification with a Dirichlet kernel [147]. Or our method, as shown in Chapter 5, works pretty well with irregular data, but some classical methods such as Fourier truncation [65] presuppose a uniform grid and cannot be applied directly (Table 5.9).

Also, modern methods are not limited to neural network training. We draw attention to operator learning methods [13] such as DeepONet [76], Fourier neural operator (FNO) [75], MeshGraphNets / graph neural simulators [77], and convolutional/transformer-based PDE sur-

rogates [73, 74]; or equation/model discovery that is regression-based, such as SINDy (sparse identification of nonlinear dynamics) [71] or PDE-FIND [78], an extension of SINDy specifically for discovering PDEs from spatiotemporal data.

# Chapter 2

# Inverse Problems and Regularization Theory

*A man is like a fraction whose numerator is what he is and whose denominator is what he thinks of himself. The larger the denominator, the smaller the fraction.*

Leo Tolstoy
Letter, (1910)

*In this chapter, we start from the Hadamard three conditions (Definition 2.1), and see that the solution has uniqueness, existence happens only within* $\mathrm{Ran}(A)$*, which is dense, however not closed, and also no continuous dependence takes place at all. The reason is that* $m(\xi) > 0$ *at all frequencies, nevertheless* $m(\xi)$ *approaches zero, and hence,* $1/m(\xi)$ *becomes unbounded (Remark 2.2). The model for noise, given in* (2.1)*, is not affected by us in any way. The classification is presented in Section 2.2. Remark 2.3 fixes the perturbed model all our data come from. The second step involves the degree of instability. As Proposition 2.4 relies on* $A$ *to be compact, which is not the case in our dissertation, Remark 2.6 uses truncations to derive the degree from Definition 2.5. The degree is equal to two. This extra step through the truncations is also the reason for not basing the proof on eigenvectors in Chapter 4, since no eigenvectors exist, but rather by using Weyl sequences. Stability is achieved by using the* $H^{\ell}(\mathbb{R})$ *penalty term in Section 2.3 that is seen in Definition 2.7. The filter form* (2.5) *inverts the low frequency content and damps the high frequency one, where* $\lambda$ *determines the crossover. As noted in Remark 2.10, the penalty term can be expressed via the Hermite basis as* $\lambda\|c\|_{\Lambda}^2$*, and, based on Proposition 2.11, we have convergence but no rate. Secondly, in Section 2.4, we have Theorem 2.12 concerning the Hölder bound for* $\mathcal{W}_{\ell,E}$ *with parameter* $\ell/(\ell+2)$*, and Proposition 2.13 regarding Tikhonov regularization, which converges at this rate under the a-priori choice* (2.8)*. Both results contain constant factors that are never evaluated here: in the former it is* $C(E,\ell)$*, in the latter one it is an absolute constant. Their evaluation is the subject of Chapter 4 (Remark 2.14).*

## 2.1 The Forward Problem and Its Inverse

**Definition 2.1.** Consider two Banach spaces $X$ and $Y$. If $A : X \to Y$ is a linear operator, then the problem $A\varphi = g$ is *well-posed in the sense of Hadamard* when the following conditions hold true:

(H1) *(Existence)* For each $g \in Y$, there exists a $\varphi \in X$ such that $A\varphi = g$.

(H2) *(Uniqueness)* The solution $\varphi$ is unique.

(H3) *(Stability)* The solution depends continuously on the data: $\varphi_n \to \varphi$ in $X$ provided $g_n \to g$ in $Y$.

A problem which fails to satisfy at least one of (H1)-(H3) is said to be *ill posed*. The three conditions (H1)-(H3) can be found in [68, Ch. 2][90, §2.1].

In this case, the answer would be (H3). Existence and uniqueness are there, while continuity is not; the latter is evident from the very symbol.

**Remark 2.2.** $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is a Fourier multiplier with the symbol $m(\xi) = (1 - e^{-\xi^2})/\xi^2$. There are two important properties that should be mentioned: $m(\xi) > 0$ for any $\xi$, and $m(\xi) \to 0$ for $|\xi| \to \infty$ (Proposition 1.19). The inversion in the formal sense is simply multiplication by the unbounded $1/m(\xi)$. This is precisely how a small high-frequency perturbation in the input data can swing the reconstructed source by an arbitrary amount. This is the violation of condition (H3) of Definition 2.1.

The magnitude of these perturbations will be the focus of the current chapter, and hence, we settle on what constitutes noise. The data in Chapters 2 through 4 have an error of deterministic nature whose size is known:

$$g^\delta = A\varphi^\dagger + \delta\eta, \qquad \eta \in L^2(\mathbb{R}),\ \|\eta\|_{L^2(\mathbb{R})} \le 1,\ \delta > 0. \tag{2.1}$$

This is the model (4.4) used in Chapter 4, taken here as a standing assumption. In particular, $\|g^\delta - A\varphi^\dagger\|_{L^2(\mathbb{R})} \le \delta$, the standard noise-level condition in the theory of regularization [68, Ch. 3, Def. 3.1].

## 2.2 Classification and Analysis

### 2.2.1 Classification of Inverse PDE Problems

Inverse problems associated with partial differential equations are classified by which coefficient, datum, source, or boundary is unknown.

*Source identification* (this dissertation): recovery of the time-independent source $\varphi^\dagger$ from the observed terminal-time data $g^\delta = A\varphi^\dagger + \delta\eta$ for the heat equation. In the one-dimensional whole-space case the multiplier satisfies $m(\xi) \asymp \xi^{-2}$ as $|\xi| \to \infty$, so it is mildly ill-posed of polynomial type.

*Recovery of the initial datum*: the backward heat equation, which is severely ill-posed since the high-frequency modes are exponentially damped [114].

*Coefficient identification*: determining an unknown diffusion coefficient $a(x)$ in the equation $\partial_t u - \partial_x(a(x)\partial_x u) = f$. It is typically a nonlinear problem [90, Ch.9].

*Boundary and geometry identification*: from boundary or far-field data, as in the case of the Dirichlet-to-Neumann approach [21], or from inverse obstacle scattering [46].

> **Remark 2.3.** In Chapter 4, the data is deliberately created by slightly perturbing the forward model, $a(x) = 1+\varepsilon e^{-x^2/4}$, and then solving the problem based on the unperturbed model $a \equiv 1$. This is done deliberately because it makes sure that the inverse method does not get graded against data generated by essentially the model it inverts with, thus avoiding the *inverse crime*; see Wirgin [145].

### 2.2.2 The Picard Criterion and Degree of Ill-Posedness

The knowledge that the inversion is unstable is not sufficient; we need to quantify the degree of instability, since that will determine the best achievable rate. The singular-system approach gives us the language of grades with only one minor exception for Chapter 4: our operator $A$ is not compact, and so the terminology applies to its truncations, which foreshadows the continuous-spectrum investigation Chapter 4 carries out directly.

It should be remembered that this is the singular-system theory for a *generic* compact operator $T$. This case is distinct from our operator $A$, which is non compact: its spectrum is the continuum $[0, 1]$ with no eigenvalues (Chapter 4, Theorem 4.6), so the framework below applies to its truncations $A_L$ (Chapter 4, Section 4.3.4). Consider $T : X \to Y$ such that $T$ is a compact operator with the following singular value decomposition:

$$T = \sum_{n=1}^{\infty} \sigma_n \langle \cdot\,,\, u_n \rangle_X \, v_n, \qquad \sigma_1 \geq \sigma_2 \geq \cdots > 0, \tag{2.2}$$

with $\{u_n\}, \{v_n\}$ orthonormal bases of $\overline{\mathrm{Ran}(T^*)}$ and $\overline{\mathrm{Ran}(T)}$ [68, §2.2, pp. 36–37].

> **Proposition 2.4.** *In case $T : X \to Y$ is a compact operator which is also injective and whose range is infinite dimensional, then its inverse $T^{-1} : \mathrm{Ran}(T) \subset Y \to X$ is unbounded. Therefore, condition* (H3) *of Definition 2.1 does not hold.*

**Proof.** Let us take $g_n := \sigma_n v_n$. It will then be true that the norm of $g_n$ in $Y$ is $\|g_n\|_Y = \sigma_n \to 0$, whereas the norm of $T^{-1}g_n$, which is equal to $u_n$, is $\|T^{-1}g_n\|_X = 1$. Thus $T^{-1}$ fails to be continuous at 0 and, being linear, is therefore unbounded [68, Prop. 2.7, p. 38]. □

The Picard criterion makes this quantitative: $T\varphi = g$ has a solution if and only if $g \in \overline{\mathrm{Ran}(T)}$ and the sum over all $n$ of $|\langle g,\, v_n\rangle_Y|^2/\sigma_n^2$ is finite. In that case, the minimum-norm solution can be computed as $\varphi^\dagger = \sum_n \sigma_n^{-1}\langle g,\, v_n\rangle_Y u_n$ [68, Thm. 2.8, pp. 38–39]. The division by $\sigma_n \to 0$ is the mechanism: an error in $\langle g^\delta,\, v_n\rangle$ is amplified by the factor $1/\sigma_n$, which is why the quicker $\sigma_n$ tends to zero, the greater the error will be.

**Definition 2.5.** The operator $T$ is referred to as having *degree of ill-posedness* $s > 0$ when its singular values satisfy $\sigma_n \asymp n^{-s}$ as $n \to \infty$. If the singular values behave like $e^{-cn^{\alpha}}$, the problem is classified as being *severely* ill-posed [48, §2, Def. 1].

**Remark 2.6.** The symbol of the operator, $m(\xi) = (1 - e^{-\xi^2})/\xi^2 \asymp \xi^{-2}$ (Proposition 1.19), smooths by two derivatives. This means that $A$ acts on $H^s(\mathbb{R})$ and takes values in $H^{s+2}(\mathbb{R})$. Being a non compact operator, we consider the singular values of $A_L$ instead. Since the eigenfunctions are $\sin(n\pi x/L)$, we have $\sigma_n = m(n\pi/L) \asymp n^{-2}$. The problem has polynomial instability of *degree* 2, not an exponential one. This is consistent with the classification on the Hilbert scale from Chapter 4, Corollary 4.9, with Lemma 4.7 proving the precise lower bound $m(\xi) \geq c_0/(1 + \xi^2)$ behind it.

## 2.3 Tikhonov Regularization

We locate the point of instability and the extent of it and deal with the problem through a simple compromise: a small error in data fit is acceptable in order to make the solution more stable. This is significant in the context of Chapter 4 since after discretizing using the Hermite basis, the regularization term is simply the coefficient $\lambda$ multiplied by the squared norm of $c$ in $\Lambda$ within the objective of the network, and the filter it produces is the object Chapter 4 identifies the trained source map with.

**Definition 2.7.** For $\lambda > 0$ and noisy input data $g^{\delta}$ which satisfies the noise model (2.1), the *Tikhonov functional with $H^{\ell}$-penalty* is given by

$$J_{\lambda}(\varphi) := \|A\varphi - g^{\delta}\|^2_{L^2(\mathbb{R})} + \lambda \|\varphi\|^2_{H^{\ell}(\mathbb{R})}, \qquad \varphi \in H^{\ell}(\mathbb{R}). \tag{2.3}$$

$L^2(\mathbb{R})$ penalization is what was used directly by Tikhonov [138, 139] with the regular Hilbert space framework [68, Thm. 5.1, pp. 117–118]. The Sobolev, or Hilbert-scale, penalization term in the Tikhonov functional is the one given by Engl, Hanke, and Neubauer [68, §8.5, p. 215] and also Natterer [115]. The optimization of this functional is the standard quadratic problem.

**Proposition 2.8.** *For all $\lambda > 0$ and all $g^{\delta} \in L^2(\mathbb{R})$, the functional $J_{\lambda}$ admits a unique minimizer $\varphi^{\delta}_{\lambda} \in H^{\ell}(\mathbb{R})$ satisfying the* normal equation

$$(A^*A + \lambda \mathcal{L}^{\ell})\varphi^{\delta}_{\lambda} = A^* g^{\delta}, \tag{2.4}$$

*where $\mathcal{L}^{\ell} := (I - \partial_{xx})^{\ell}$ is the $H^{\ell}$-regularizer.*

**Proof.** The functional $J_{\lambda}$ is strictly convex and coercive on $H^{\ell}(\mathbb{R})$ and hence admits a unique minimizer. Setting the first variation of $J_{\lambda}$ to zero will yield the normal equation (2.4). See [68, Ch. 5, §5.1] for details of this Hilbert space formulation. The Hilbert-scale penalty is detailed in [68, Ch. 8, §8.5]. □

Because all of our operators are Fourier multipliers, we can use the normal equation on a frequency by frequency basis.

**Proposition 2.9.** *For the Tikhonov solution, we have the following equation in the Fourier domain:*

$$\hat{\varphi}_\lambda^\delta(\xi) = \frac{\overline{m(\xi)}}{|m(\xi)|^2 + \lambda(1+\xi^2)^\ell}\, \hat{g}^\delta(\xi) =: q_\lambda(\xi)\, \hat{g}^\delta(\xi), \tag{2.5}$$

*In this case, $q_\lambda(\xi)$ is a regularization filter whose range is $(0, m(\xi)^{-1})$. For any given $\xi$, $q_\lambda(\xi)$ tends to $m(\xi)^{-1}$ as $\lambda \to 0^+$.*

**Proof.** Fourier transform the normal equation (2.4). Here is what you need to use: the Fourier transform of $A^* f$ is $\overline{m}\, \hat{f}$, and the Fourier transform of $\mathcal{L}^\ell f$ is $(1+\xi^2)^\ell \hat{f}$. The scalar equation becomes

$$\big(|m(\xi)|^2 + \lambda(1+\xi^2)^\ell\big)\hat{\varphi}_\lambda^\delta(\xi) = \overline{m(\xi)}\, \hat{g}^\delta(\xi).$$

Solve the resulting equation pointwise to obtain (2.5), which is the Fourier-multiplier form of the classical Tikhonov filter [68, §5.1], but with the $H^\ell$ penalty [68, §8.5]. □

There are two reasons for using the $q_\lambda$ filter. First, at large values of the multiplier $m(\xi)$, i.e. at low frequencies, $q_\lambda$ is equal to about $1/m(\xi)$, and the inversion proceeds easily. Second, at small $m(\xi)$, i.e. at high frequencies, where the noise is amplified, the expression $\lambda(1+\xi^2)^\ell$ is dominant in the denominator and damps the reconstruction. In this case, the single parameter $\lambda$ is responsible for the crossover point the same trade Chapter 4 makes, in the Hermite basis.

**Remark 2.10.** In Chapter 4, the discussion about discretization of the penalty term using the Hermite basis for a diagonal expression is considered. In particular,

$$\lambda \|\varphi^c\|^2_{H^\ell(\mathbb{R})} \approx \lambda \|c\|^2_\Lambda = \lambda \sum_{j=0}^{N_\varphi - 1} (2j+1)^\ell |c_j|^2.$$

This stems immediately from the norm equivalence (1.16) and the expression for the mass matrix $\Lambda$ as stated in (1.14). Hence, the penalizing factor $\lambda\|c\|^2_\Lambda$ in (4.18) is just the Hermite-discrete counterpart of the $H^\ell$-Tikhonov penalty (2.3).

**Proposition 2.11.** *We define $\varphi^\dagger$ to be the exact solution of the problem, lying in $H^\ell(\mathbb{R})$, and further suppose that $g^\delta$ satisfies the noise model (2.1). Under the conditions*

$$\lambda(\delta) \to 0, \qquad \frac{\delta^2}{\lambda(\delta)} \to 0 \qquad \text{as } \delta \to 0^+,$$

*we get*

$$\|\varphi_\lambda^\delta - \varphi^\dagger\|_{L^2(\mathbb{R})} \xrightarrow{\delta \to 0^+} 0. \tag{2.6}$$

**Proof.** The two assumptions offset each of the two error components. By letting $\lambda$ approach 0, we can neglect the approximation error $\|\varphi_\lambda^0 - \varphi^\dagger\|$, while by letting $\delta^2/\lambda$ approach 0 we negate the propagated data error $\|\varphi_\lambda^\delta - \varphi_\lambda^0\|$. Embedding $H^\ell(\mathbb{R})$ into $L^2(\mathbb{R})$, we get our result

for $L^2(\mathbb{R})$. It is the classical a-priori convergence theorem [68, Thm. 5.2] in the framework of Hilbert scales given in [68, Ch. 8, §8.5, Remark 8.24]. □

## 2.4 Optimal Convergence Rates and Conditional Stability

But just convergence does not say anything about the rate. In our case, there is an issue with the concept of "rate," which actually indicates how fast the error goes to zero with a smaller $\delta$. To define the "rate" of convergence, it is essential to know the function $\varphi^\dagger$, which is defined via $\mathcal{W}_{\ell,E}$ in Definition 1.8.

### *2.4.1 Source Conditions and Conditional Stability*

It represents the core achievement obtained throughout the course of the dissertation. The derived bound is conditionally stable; moreover, in the event the source originates from $\mathcal{W}_{\ell,E}$, the bound for the reconstruction error in terms of the data error will be a power of the latter with the best-possible exponent. It forms the template used in Chapter 4, where $C$ in it is replaced by $(2E)^{2/(\ell+2)}$.

**Theorem 2.12.** *There exists a constant $C = C(E,\ell) > 0$ such that for all $\varphi_1, \varphi_2 \in \mathcal{W}_{\ell,E}$,*

$$\|\varphi_1 - \varphi_2\|_{L^2(\mathbb{R})} \le C \, \|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)}. \tag{2.7}$$

**Proof sketch.** Consider $\varphi$ to be the difference between $\varphi_1$ and $\varphi_2$ such that $\varphi = \varphi_1 - \varphi_2$. Then $\varphi$ belongs to the space $\mathcal{W}_{\ell,2E}$. Consider $\varepsilon$ to be the $L^2(\mathbb{R})$ norm of $A\varphi$. Therefore, we have $\varepsilon = \|A\varphi\|_{L^2(\mathbb{R})}$. Split the Plancherel integral at some frequency $K$. In case of low frequencies, use $m(\xi) \ge c_0/(1+\xi^2)$, while for high frequencies, use the source bound. Once the optimal value of $K$ is selected, we have $\|\varphi\|_{L^2(\mathbb{R})} \le C(E,\ell)\, \varepsilon^{\ell/(\ell+2)}$. This holds because the method follows the conventional interpolation approach for conditional stability [136, Thm. 2.1]. However, the precise value of the constant $C$ has not been mentioned, but it is calculated exactly in Chapter 4. □

### *2.4.2 Optimal A-Priori Parameter Choice*

Combining the filter described in Proposition 2.9 and the bound provided by Theorem 2.12 leads to the well-known fact that under appropriate selection of $\lambda$ to correspond to the level of noise, one gets an optimal rate on $\mathcal{W}_{\ell,E}$. This is recorded because Chapter 4 uses this tuning and identifies the trained network with the regularizer responsible for it.

**Proposition 2.13.** *Suppose $\varphi^\dagger \in \mathcal{W}_{\ell,E}$. With the a-priori parameter choice*

$$\lambda^\star(\delta) := \frac{1}{2}\Big(\frac{\delta}{E}\Big)^2, \tag{2.8}$$

*the Tikhonov regularization estimator satisfies the following bound:*

$$\|\varphi_{\lambda^\star}^{\delta} - \varphi^{\dagger}\|_{L^2(\mathbb{R})} \leq C_{\mathrm{abs}}\, E^{2/(\ell+2)}\, \delta^{\ell/(\ell+2)}, \tag{2.9}$$

*where $C_{\mathrm{abs}} > 0$ is an absolute constant. The power $\ell/(\ell+2)$ is the optimal one on $\mathcal{W}_{\ell,E}$ in the modulus-of-continuity sense.*

**Proof.** The selection in (2.8) ensures that the scales for the data fit and the penalty are identical. Equation (2.9) is based on inserting the obtained residual bound in Theorem 2.12. This is the conventional order optimal a-priori approach in case of a source constraint as proved by [135, Thm. 1; Thm. 2]. The optimality of the exponent is the modulus of continuity result of [136, §2.3]. □

**Remark 2.14.** Constants appearing in Theorem 2.12 and Proposition 2.13 are not stated here. They can be found in Chapter 4. The additional input is the multiplier estimate $m(\xi)(1+\xi^2) \geq 1$ given by Lemma 4.7 (c). By using the sharp interpolation estimate stated in Theorem 4.10, we have $C = (2E)^{2/(\ell+2)}$ in Theorem 4.11, and at Corollary 4.12 the central estimate is the following:

$$\|\widehat{\varphi} - \varphi^{\dagger}\|_{L^2(\mathbb{R})} \leq (2E)^{2/(\ell+2)}(\delta+\varepsilon)^{\ell/(\ell+2)}.$$

The corresponding result from the a-priori and training point of view is presented in Corollary 4.13. The optimality approach underlying all of the above is due to Tautenhahn, Hämarik, Hofmann, and Shao [136, §2.3].

# Chapter 3

# Neural Networks and the Hermite Variational PINN

*Science is a differential equation. Religion is a boundary condition.*

---

Alan Turing
Letter to Robin Gandy (1954)

*In this chapter we provide an overview of the background behind the Computational approach used in the dissertation, including the motivation behind each part of the architecture. The state function is as a neural network with* tanh *activation functions (Definition 3.1), so the residual of the PDE is obtained through automatic differentiation, and the parameters are initialized as in Definition 3.2, giving the warm start for training in Chapter 4. The Universal Approximation Theorem (Theorem 3.3) ensures that such an approximation is possible but provides no convergence rate, hence we need the a posteriori approach. In Section 3.3 we start by imposing the equation as a penalty at certain nodes while also recovering the source via minimization. The weak form of the residual against the Hermite basis functions leads to three advantages, namely, the Gram matrix of the test functions is equal to the identity, the stiffness matrix is diagonal according to Proposition 1.12 and the penalty term decomposes into a weighted sum of squares. With Definition 3.4, the factor t guarantees that the problem is initialized with a zero state as per Proposition 1.4, directly and not through a penalty. Moreover, since by Definition 3.5 the residual is an affine function of the source vector, the loss defined in Definition 3.6 is strictly convex in the source coefficients. That loss requires not only fitting the data but also consistency with the differential equation while ensuring the boundedness of the source coefficients, and it is given in both a discrete and a continuous form, whose difference is bounded in Chapter 4. Therefore, the source has an analytic minimizer (Proposition 3.7), which is exact for each set of network parameters; these parameters are computed by means of gradient descent on the reduced loss (3.13); this bilevel structure (Section 3.3.5) is what makes the source mapping computable, and since the source originates from an explicit formula and not an optimiser, Theorem C recognises the source map as Tikhonov regularisation in the Hermite Sobolev scale, which is why the source is a spectral expansion and not a network.*

## 3.1 Tanh Feedforward Networks and Default Initialization

The state is a neural network, and almost everything that we require from it, differentiability of the residual and a reasonable choice of the initial value is determined by two things: the choice of activation function and initialization; both are fixed here in the form Chapter 4 uses.

**Definition 3.1.** Let $D \geq 1$ be the number of hidden layers used for the network, and $n_0, n_1, \ldots, n_{D+1} \in \mathbb{N}$ denote the layer widths. The definition of a *tanh feedforward network* $\mathcal{N}_\theta : \mathbb{R}^{n_0} \to \mathbb{R}^{n_{D+1}}$ is recursive, and is stated as follows:

$$x^{(0)} := x, \qquad x^{(k)} := \tanh\big(W^{(k)}x^{(k-1)}+b^{(k)}\big),\ k = 1, \ldots, D, \qquad \mathcal{N}_\theta(x) := W^{(D+1)}x^{(D)}+b^{(D+1)}. \tag{3.1}$$

In this case $W^{(k)} \in \mathbb{R}^{n_k \times n_{k-1}}$, $b^{(k)} \in \mathbb{R}^{n_k}$, and tanh acts entrywise. The vector of parameters is $\theta := (W^{(1)}, b^{(1)}, \ldots, W^{(D+1)}, b^{(D+1)}) \in \mathbb{R}^P$, with $P = \sum_{k=1}^{D+1} n_{k-1}n_k + n_k$ components.

The choice of using tanh is deliberate [121, §2]. Due to the infinite smoothness of the tanh function over the real line, and hence the infinite smoothness of $\mathcal{N}_\theta$ in $x$ for every parameter $\theta$, the smoothness property is transferred to $v_\theta(x,t) = t\, e^{-\alpha_v x^2}\mathcal{N}_\theta(x,t)$. Consequently, the PDE residual $R_{\mathrm{PDE}}(\theta, c)$ becomes differentiable by backpropagation [121, §3.1].

**Definition 3.2.** The trial-state network is initialized in accordance with the default linear layer initialization in the deep learning framework, namely the Kaiming-uniform initialization [82] (the uniform variant, the PyTorch convention [31]), with the framework's default negative-slope value $a = \sqrt{5}$. The weight values and bias values are independently drawn from the fan-in-scaled uniform law

$$W_{ij}^{(k)},\ b_i^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{U}\left[-\frac{1}{\sqrt{n_{k-1}}},\ \frac{1}{\sqrt{n_{k-1}}}\right], \qquad k = 1, \ldots, D+1, \tag{3.2}$$

where $n_{k-1}$ is the fan-in of layer $k$. The obtained $\theta_0$ acts as the warm start for Chapter 4. During its measurement warm-up phase (Algorithm 1, Phase 1), it drives the state to the data before the source enters, so the reconstruction is insensitive to the initial scale.

The concrete dimensions are specified in Chapter 4. The mapping $\mathcal{N}_\theta$ receives $(x,t) \in \mathbb{R}^2$ and produces a real value, so $n_0 = 2$ and $n_{D+1} = 1$, with $D = 7$ hidden layers of width $W$ (hence $n_1 = \cdots = n_7 = W$ and $n_8 = 1$). In Chapter 4, $W$ is chosen to be 192.

## 3.2 Universal Approximation Theory for tanh Networks

**Theorem 3.3.** *Let $f$ be a continuous function on $[0,1]^d$, and let $\varepsilon > 0$. There exist $N \in \mathbb{N}$ and parameters $\theta$ such that the one-hidden-layer network $\mathcal{N}_\theta(x) = \sum_{k=1}^{N} c_k \tanh(\langle a_k, x\rangle + b_k)$ approximates $f$ on $[0,1]^d$, which means that $\sup_{x\in[0,1]^d} |f(x) - \mathcal{N}_\theta(x)| < \varepsilon$.*

**Proof.** [61, Thm. 2]; the only property of tanh used is that it is a non-constant, bounded, continuous (*sigmoidal*) function. □

## 3.3 Physics-Informed Neural Networks and Variational PINNs

### 3.3.1 Strong Form PINNs and Variational PINNs

Following design of the neural network architecture for the particular problem being considered, such that it can capture the truth in the state, the other question that comes up is how to enforce the PDE on it. That has already been done in the case of physics-informed modeling, which imposes the PDE in the cost function, but there is an inverse variant too wherein the unknown source term can come along in the same minimisation process, and this is the coupling Chapter 4 makes use of, with the source in Hermite coefficients rather than a second network.

The *physics-informed neural network* (PINN) uses the PDE as a soft loss constraint. Suppose we have a problem of the form $\partial_t u - \partial_{xx} u = f$ on $\mathbb{R} \times (0,1]$ with noisy terminal data $g^\delta$ [3, 88]:

$$\mathcal{L}_{\text{PINN}}(\theta) := \frac{1}{M_d} \sum_{i=1}^{M_d} |v_\theta(x_i, 1) - g_i^\delta|^2 + \frac{\mu}{M_r} \sum_{(i,k)} |\partial_t v_\theta - \partial_{xx} v_\theta - f|^2_{(x_i,t_k)}. \tag{3.3}$$

If $f = \varphi$, and $\varphi$ is the function which is *unknown*, We try to minimize the loss function (3.3) simultaneously w.r.t. $(\theta, \varphi)$, where $\varphi$ is represented using another neural network or in this case by the Hermite expansion $\varphi^c = \sum_j c_j \psi_j^\sigma$.

### 3.3.2 Variational PINNs

This is the change to a weak-form residual being applied in relation to the pointwise residual in (3.3). The usefulness of this technique comes in due to the test functions: the scaled Hermite functions of Section 1.3.3, which diagonalize the algebraic structure, making the source elimination of Section 3.3.5 almost cost-free.

A *variational PINN* (V-PINN) replaces the pointwise residual in (3.3) by a weak-form residual against test functions $\{\eta_k\}_{k\geq 1}$:

$$\mathcal{L}_{\text{VPINN}}(\theta) := \sum_k \Big| \int_{\mathbb{R}\times(0,1)} R_{\text{PDE}}(\theta, c)(x,t)\, \eta_k(x,t)\, \mathrm{d}x\, \mathrm{d}t \Big|^2 + (\text{data terms}). \tag{3.4}$$

[92]

In Chapter 4, the test functions are the scaled Hermite functions $\psi_j^\sigma$ (Definition 1.13, §1.3.3). The choice of these functions is due to three features that make the source elimination of §3.3.5 both transparent and cheap: orthonormality, so the test-function Gram matrix is the identity; the eigenfunction relation $L_\sigma \psi_j^\sigma = (2j+1)\psi_j^\sigma$ (Proposition 1.12), so the stiffness matrix is diagonal; and $\langle \varphi^c, \psi_k^\sigma \rangle = c_k$, so the regularizer $\lambda \|c\|_\Lambda^2$ is a weighted Euclidean norm (Definition 1.14).

### 3.3.3 Trial Source and Trial State

The two trial objects optimized by Chapter 4 are kept as two distinct entities: a spectral source and a neural state. Such an approach makes it possible to solve for the spectral source in closed form and to train the neural state by gradient descent.

**Definition 3.4.** Set a positive scaling factor $\sigma$, truncation parameter $N_\varphi \in \mathbb{N}$, decay factor $\alpha_v > 0$, and network parameter $\theta \in \mathbb{R}^P$.

(i) The *trial source* is

$$\varphi^c(x) := \sum_{j=0}^{N_\varphi - 1} c_j \psi_j^\sigma(x), \qquad c := (c_0, \dots, c_{N_\varphi - 1})^\top \in \mathbb{R}^N_\varphi. \tag{3.5}$$

(ii) The *trial state* is

$$v_\theta(x,t) := t \cdot e^{-\alpha_v x^2} \cdot \mathcal{N}_\theta(x,t). \tag{3.6}$$

The trial state (3.6) employs two constraints. Firstly, the presence of the term $t$ guarantees that $v_\theta(x,0) = 0$ for each parameter $\theta$, thus embedding the constraint of zero initial conditions rather than learning it. Secondly, the presence of the envelope $e^{-\alpha_v x^2}$ ensures that $v_\theta(\,\cdot\,, t)$ belongs to $L^2(\mathbb{R})$ provided $\mathcal{N}_\theta$ is bounded, in the same way enforcing far-field decay (Chapter 4 takes $\alpha_v = 0.08$). What is left to $\mathcal{N}_\theta : \mathbb{R}^2 \to \mathbb{R}$ is the expressive work.

**Definition 3.5.** The *PDE residual* of the pair $(\theta, c)$ is given by

$$R_{\mathrm{PDE}}(\theta, c)(x,t) := \partial_t v_\theta(x,t) - \partial_{xx} v_\theta(x,t) - \varphi^c(x). \tag{3.7}$$

Since the *residual generator* $\mathcal{G}_\theta := \partial_t v_\theta - \partial_{xx} v_\theta$ is completely defined by $\theta$, we have $R_{\mathrm{PDE}}(\theta, c) = \mathcal{G}_\theta - \varphi^c$. When the residual nodes $\{(x_i^H, t_k^L)\}$ are taken into account, this reads

$$R_{\mathrm{PDE}}(\theta, c)\big|_{\mathrm{nodes}} = \mathbf{g}_\theta - \Phi c, \tag{3.8}$$

where $\mathbf{g}_\theta := (\mathcal{G}_\theta(x_i^H, t_k^L)) \in \mathbb{R}^{M_r}$ and $\Phi$ is the Hermite design matrix (1.15).

The crucial concept underlying the entire structure is the separation $R_{\mathrm{PDE}} = \mathcal{G}_\theta - \Phi c$, where the residual is *affine* in $c$, and the neural network contributes only the constant vector $\mathbf{g}_\theta$ to it. The fact that it is linear with respect to $c$ makes the loss quadratic with respect to $c$, hence solvable in closed form.

### 3.3.4 The Continuous and Discrete Losses

These three messages are communicated concurrently through the loss function: fit the data, adhere to the equation, and preserve regularity. Through expressing the loss function first as its continuous form and then as its discrete form, we maintain the analytical formulation, in the former, distinct from the one used by the optimizer, in the latter, which is essential in bounding their difference in Chapter 4.

**Definition 3.6.** The definition of the *continuous loss* is:

$$\mathcal{L}(\theta, c) := \|v_\theta(\,\cdot\,, 1) - g^\delta\|^2_{L^2(\mathbb{R})} + \mu\|R_{\mathrm{PDE}}(\theta, c)\|^2_{L^2(\mathbb{R}\times(0,1))} + \lambda\|c\|^2_\Lambda. \tag{3.9}$$

The *discrete loss* (see also Chapter 4, equation (4.18)) is the one where the $L^2$ integrals are approximated using the Gauss–Hermite and Gauss–Legendre quadrature (Definitions 1.22 and 1.23). It is calculated as follows:

$$L(\theta, c) := \frac{1}{M_d}\sum_{i=1}^{M_d}|v_\theta(x_i, 1) - g_i^\delta|^2 + \frac{\mu}{M_r}\sum_{i,k} w_i^H w_k^L |R_{\mathrm{PDE}}(\theta, c)(x_i^H, t_k^L)|^2 + \lambda\|c\|^2_\Lambda. \tag{3.10}$$

In matrix notation, we have $\mathbf{v}_\theta := (v_\theta(x_i, 1))_{i=1}^{M_d} \in \mathbb{R}^{M_d}$ (the trial state at the $M_d$ data sensors), $\mathbf{g}^\delta := (g_i^\delta) \in \mathbb{R}^{M_d}$, the matrix $\mathbf{W}_d := \frac{1}{M_d} I_{M_d}$, and $\mathbf{W}_r := \mathrm{diag}(w_i^H w_k^L / M_r)$. The discrete loss takes the compact form

$$L(\theta, c) = (\mathbf{v}_\theta - \mathbf{g}^\delta)^\top \mathbf{W}_d (\mathbf{v}_\theta - \mathbf{g}^\delta) + \mu(\mathbf{g}_\theta - \Phi c)^\top \mathbf{W}_r (\mathbf{g}_\theta - \Phi c) + \lambda c^\top \Lambda c. \tag{3.11}$$

### 3.3.5 Bilevel Structure: Source Elimination

This is where the affine-in-$c$ structure comes in handy. Assuming $\theta$ to be fixed, the mapping of $c$ to $L(\theta, c)$ is a strictly convex quadratic, whose minimizer can be explicitly expressed rather than found by searching. This is the important component of the approach, as it converts the optimization into one that involves only the neural network and, as Chapter 4 shows, makes the source map a Tikhonov filter in disguise.

**Proposition 3.7.** *For every parameter vector $\theta \in \mathbb{R}^P$, $\mu > 0$, and $\lambda > 0$, the mapping of $c$ to $L(\theta, c)$ is strongly convex and has a unique minimizer*

$$c^\star(\theta) := \Gamma(\mu, \lambda)^{-1} \cdot \mu\, \Phi^\top \mathbf{W}_r \mathbf{g}_\theta, \qquad \Gamma(\mu, \lambda) := \mu\, \Phi^\top \mathbf{W}_r \Phi + \lambda\Lambda. \tag{3.12}$$

*The matrix $\Gamma(\mu, \lambda)$ (cf. Chapter 4, equation (4.17)) is symmetric positive definite for all $\mu, \lambda > 0$.*

**Proof.** From (3.11), it is clear that the dependence on $c$ is quadratic, which can be written as $c^\top \Gamma(\mu, \lambda) c - 2c^\top[\mu\, \Phi^\top \mathbf{W}_r \mathbf{g}_\theta] + \mathrm{const}(\theta)$. As all diagonal entries $(2j+1)^\ell$ of $\Lambda$ are strictly positive, we can conclude that $\Gamma(\mu, \lambda) \succeq \lambda\lambda_{\min}(\Lambda) I \succ 0$. This finding implies that the form is strongly convex, and that its gradient vanishes only at (3.12). □

By cancelling $c$ through its optimum counterpart $c^\star(\theta)$, we can write our *reduced loss* $J(\theta)$

as

$$J(\theta) := L(\theta, c^\star(\theta)). \tag{3.13}$$

In fact, what we have here is a bilevel approach, where the source is solved exactly for each parameter $\theta$, and at the same time $\theta$ is optimized by gradient descent on $J$. The outer problem takes place in dimension $P$ instead of $P + N_\varphi$, which is especially relevant to Chapter 4, since $\theta \mapsto c^\star(\theta)$ is the exact, differentiable object the next chapter analyses.

# Chapter 4

# A Hermite Variational PINN for the Inverse Heat Source Problem on $\mathbb{R}$: Analysis and Numerical Experiments

*Mathematics is the poetry of logical ideas.*

inspired by Albert Einstein
Letter to New York times (1935)

*In this chapter we start in Section 4.1 with the literature review, the source class, the definition of noise and the nonlinear model, and we generate all training and testing data for the algorithm from a perturbed model in order not to evaluate the algorithm on data generated from the same model (to avoid the inverse crime). The architecture and the loss function are presented in Section 4.2. The operator presented in Section 4.3 is proved to be bounded, self-adjoint, injective, and non compact. It is a smoothing operator of order two for which the sharp constants have been established. As the spectrum of this operator is continuous, there are no eigenvalues; therefore, we analyze sequences approximating the worst function (Weyl sequences). The operator also fulfills the pointwise multiplier inequality which carries the rest of the chapter. Since m is a Laplace transform, $m^2$ is completely monotone and strictly convex w.r.t. the Sobolev weight (Proposition 4.17). In particular, there is always a unique worst frequency. This leads us to conclude an interpolation inequality with constant one, the sharp prefactor of Theorem 4.11, the exact error at each specific noise level and not only asymptotically (Theorem A), that no reconstruction operator does better, and that even a linear strategy performs as well (Theorem B), as well as the exact square-root law of J in the case of multisnapshot measurements (Theorem 4.22). Theorem C proves that the learned source map is identical to classical Tikhonov regularization on the Hermite Sobolev scale. The assumptions in this theorem are discharged directly from our training log, and we will provide an explanation as to why we use a certain decay rate for the penalty weight. We confirm the setting, the forward solver, the baselines, and the theoretical results in Section 4.5.*

## 4.1 Introduction

### *4.1.1 Literature Review*

Conditional stability for source densities is well studied in the work of [53], where the observation is a boundary trace and the modulus of continuity obtained is only logarithmic, of the form $\|f\|_{L^2(\Omega)} = O\big((\log 1/\eta)^{-\beta}\big)$ as the data noise level $\eta \downarrow 0$. Identifiability from terminal data is proved by Isakov [89],[90, Ch.9]. In both cases, only order estimates are provided, with no explicit constants.

Dou et al. consider the same forward operator. Specifically, in Theorems 3.4 and 4.1, Dou et al. [65, Thm. 3.4, Thm. 4.1] consider the multiplier $m(\xi) = (1 - e^{-\xi^2})/\xi^2$ with the same Sobolev ball assumption. Instead of finding a sharp constant for the operator, they obtain an order optimal bound for Fourier regularisation. Most relevant to this problem is the result of Umbricht [19, Thm. 2], dealing with the determination of the source from terminal observation for the equation on $\mathbb{R}^n$, $u_t = \alpha^2 \Delta u - \beta \cdot \nabla u - \nu u + f$. In our case $\alpha = t_0 = 1$, $\beta = 0$, $n = 1$, $\nu \to 0$, and the exponent obtained by Umbricht is also $p/(p+2)$ on an $H^p$ ball, with an explicit constant. The companion bounds are given in [142, Thm. 4]. Neither of the two statements is sharp, while the first order filter considered in [19] is $(1 + \mu^2 \|\xi\|^2)^{-1}$, which has qualification equal to one. This implies that under the stated parameter rule the guaranteed exponent is $\min\{p, 2\}/(p+2)$, peaking at $\frac{1}{2}$ and degrading for $p > 2$. All of these are linear and translation invariant. The variational technique considered herein also applies to the nonlinear problem mentioned in Section 4.1.4 as well as the perturbation due to a variable coefficient in Section 4.1.5. In none of these cases is there a fixed multiplier that can be applied. The problem of identifying the initial source, treated by an inverse moment formulation, is discussed in Liu and Zuazua [20], but in the Kantorovich-Wasserstein metric as opposed to the $L^2$ Sobolev framework used here.

*Modulus of continuity and minimax interval.* The conditional stability estimate is the modulus of an inverse that is restricted, and the comparison $\Omega(\delta) \le e_{\mathrm{mm}}(\delta) \le 2\,\Omega(\delta)$ is proved in Hofmann, Mathé and Schieck [87]. Moduli of inverses of operators having a non-closed range are discussed in Hegland and Hofmann [22]. The assertion is used as stated in Remark 4.21.

*The optimality of linear smoothing.* The minimax error with respect to all reconstruction procedures, either linear or nonlinear, equals a centred extremal value which is attained by linear smoothing between two ellipsoids. This theorem, proved in [106, Thm. 3.1] and reviewed in [108], has a statistical counterpart by Donoho [63], with the parameter selection problem solved in [38]. In combination, these theorems provide the upper bound described in Theorem B.

*Scales of Hilbert spaces, interpolation, and convergence rates.* Variable scales of Hilbert spaces are introduced in the compact operator case by Hegland [122, §3, Thm. 3.1–3.4, Cor. 3.5] and in the case of an unbounded operator by Hegland [124, §2, Thm. 2.1, Thm. 2.2]. The index function "covering" calculus was developed afterwards, see [104]. Sharp interpolation results for compact operators appear in Tautenhahn [136, §2, Thm. 2.1], while the exponent $\ell/(\ell+2)$ on a Sobolev ball is given in [136, §12, Thm. 12.2] of the same work, although without the constant being stated. The power type inequality of Tautenhahn [134] enforces smoothness via $(A^*A)^{p/2}$, resulting in $p/(p+1)$. The rate on the order optimal Hilbert scale is provided by Natterer [115], while the rate $\delta^{s/(s+2)}$ in this context follows from Hermite truncation, as shown

by Meng and Zhao [107, Thm. 6].

*The non compact multiplication operators*, as well as other operators of such type, is considered in the work by Hofmann, Hofmann and Kindermann, Mathé, Nair and Hofmann, and Kindermann and Hofmann [23, 83, 123].

Following up from Bot and Hofmann's [24] investigations regarding conditional stability in connection with monotonicity and self-adjointness, the operator under consideration is not new to this research topic. The square of the operator $m^2$ is given by Werner and Hofmann in [144, Ex. 12], where the degree of ill-posedness is $2/d$, the same as Corollary 4.9 for $d = 1$. A double-sided equivalence between Sobolev regularity and a source condition on the multiplier is obtained by Mathé et al., the latter being formulated via Proposition 4.5. In these cases, the constants are not identified explicitly.

*measurement repetitions.* That $J$ measurements performed within the same budget are equivalent to one measurement at $\delta/\sqrt{J}$ is standard in information based complexity, see [54, §2.10.1]. In addition, the exact optimal recovery of heat equation solutions – even in the case where two time instances are measured inaccurately – is classical, see Magaril-Il'yaev and Osipenko [101] and Osipenko and Wedenskaya [26].

*Error analysis of physics informed neural networks.* The generalisation error bound is investigated in Mishra and Molinaro [109], while a posteriori bounds are provided in Hillebrecht and Unger [27] and Berrone et al. [28]. The convergence analysis for an elliptic inverse source problem can be found in Zhang and Liu [29]. The Tikhonov-PINN of Ji et al. [30] for an inverse potential problem is the closest to Theorem C, contrasted in (N4).

Bernstein–Widder [131, Ch. 1, Thm. 1.4]; moment problems and their extreme points [5, 10, 91, 129]; Lebesgue decomposition [127, Vol. I, §VII.2]; the harmonic oscillator scale [34]; Danskin's theorem [36, §4.3.1, Thm. 4.13].

The findings by Tautenhahn [134] prove that the exact determination of the modulus is possible for all levels of noise and not only to first order, provided the requirement of convexity is satisfied: the function $\varrho(\lambda) = \lambda\varphi^{-1}(\lambda)$ must be convex. It implies a strictly weaker form of concavity that is not equivalent to it, which acts as the hypothesis for [87, Thm. 2], to be verified on a case by case basis; where it has been discharged in general, it has been discharged for compact operators [87, Prop. 1]. The authors Hofmann and Mathé [85] generalize the bare concavity of the profile to non compact operators, but do not treat the closed-form hypothesis. Further, Boţ, Hofmann and Mathé [86, Conj. 3.11] record what is still open for compact operators too, i.e., if the right continuity of the modulus $\omega_{M(G)}$ – and not just the right continuity of the distribution function – needs compactness of $G$. It has only been proven for commuting $G$ and $H$. The operator considered here is non compact (Theorem 4.4 (vi)), so bare concavity follows from [85], but the closed-form condition does not, and that is exactly what Section 4.4.3 supplies.

Accurate constants and the spectral structure are obtained (refer to Theorems 4.4 and 4.6 and Proposition 4.5). The level of ill-posedness is shown in Werner and Hofmann [144, Ex.12]. It is important to mention that no theory based on growth estimates gives a constant. Proposition 4.5 indicates that the lower isomorphism constant is equal to 1, whereas the upper one is equal to $c_\sharp \approx 1.2984$, both of which are sharp and scale-index-independent. Theorem 4.6 substitutes the term "purely continuous" with the Lebesgue decomposition, while the spectrum

of $A$ is found to be

$$\sigma(A) = \sigma_{\mathrm{ac}}(A) = [0, 1],$$

without any point spectrum or singular continuous spectrum; i.e., $\sigma_{\mathrm{pp}}(A) = \sigma_{\mathrm{sc}}(A) = \emptyset$, and $A$ is bounded with operator norm 1. The multiplier has no singularity since $m(\xi)$ converges to 1 when $\xi$ goes to zero. Regarding the methodology, sharpness can be obtained in case that the point spectrum is empty using only Weyl sequences (Section 4.3.4).

Indeed, the precise value of the modulus of continuity at every level of noise (Theorem A) is obtained by discharging the extra assumption for a Non compact operator. Namely, as $m(\xi) = \int_0^1 e^{-s\xi^2}\,\mathrm{d}s$, its square $m^2$ is completely monotone in $\xi^2$ and, hence, strictly convex under the weight $(1+\xi^2)^{\ell}$ (Proposition 4.17). The tangent dual certificate touches at one frequency and drops the upper bound on the moment of two atoms to one atom (Lemma 4.18). Two atoms can be constructed solely out of the extreme points [9, 10]. The certificate makes the decrease necessary. This technique is classical and is used in [101], however, there the convexity requirement comes automatically while in Proposition 4.17 it is a necessity. The constant $(2E)^{2/(\ell+2)}$ appearing in Theorem 4.11 is sharp, not only correct (as observed in Remark 4.19) and that is the question the order-optimal results in [65, 107] leave open.

Exact minimax error and exact $\sqrt{J}$ law (Theorem B and Theorem 4.22). Both the endpoints of the interval of Hofmann, Mathe, Schieck [87] have been obtained, which closes the gap of a factor of two. The corollary of the exact minimax error is an application of the results from the ellipsoid theory developed by Melkman and Micchelli [106], Foucart and Liao [38] to the current class of operators, and not an extension of that. The exact value may be obtained only through making use of the property of complete monotonicity proved above. The scaling in case of multiple snapshots is the standard scaling as mentioned in [54, §2.10.1]; but the exact relation $\omega_{\oplus}(\delta) = \delta / m_{\oplus}(\xi^{\star}_{\oplus})$ and the complete collapse with noise on each channel are new results. The exact recovery result of Magaril-Il'yaev and Osipenko [101, Thm. 1] cannot be used in the present case, since in that case all the observation functions and the target belong to the same heat semigroup; hence the convexity in the time parameter is automatic. In the present case, the stacked symbols do not belong to one semigroup, so that convexity is not automatic, while the target is the identity function.

None of those bounds represents a new rate; rather, the novelty consists in the fact that both order and constant are set. The order-only bound does not contain any information regarding an implicit multiplicative constant and thus cannot prove closeness to the optimum. In contrast, Corollary 4.12 contains an exact worst-case constant. Indeed, for the case $\ell = p = 2$ with chapter parameters, the bound by [19, Thm. 2] is $3 \cdot E^{1/2} \cdot \delta^{1/2}$, whereas the minimax quantity from Theorem B equals $E^{1/2} \cdot \delta^{1/2}$. Three times too far off from optimality, an error impossible to reveal with the help of any order-only bound. Thus, the benchmarking of Chapter 5 must be conducted against the optimum and not against any rate. As for the nonlinear case of Section 4.1.4, we could not locate any comparable result due to the absence of any other tool for conversion from a linear parabolic modulus to a semilinear one besides the exponentially dependent on the Lipschitz constant transfer of [125], which requires the nonlinearity to be globally Lipschitz, which the convective and cubic terms in our problem are not.

### *4.1.2 Research Questions*

In this dissertation, we are working with a **Non Compact** heat source operator. Writing $v$ for the solution of

$$\partial_t v - \partial_{xx} v = \varphi(x) \quad \text{on } \mathbb{R} \times (0,1), \qquad v(\,\cdot\,,0) = 0,$$

the operator is the terminal data map

$$A : L^2(\mathbb{R}) \to L^2(\mathbb{R}), \qquad A\varphi := v(\,\cdot\,,1), \qquad \widehat{A\varphi}(\xi) = m(\xi)\,\widehat{\varphi}(\xi), \qquad m(\xi) = \frac{1 - e^{-\xi^2}}{\xi^2},$$

whose properties are discussed in Section 4.3.

Below are the following research questions, which have been considered in the current dissertation; all the research questions are answered in detail in Chapter 4.

1. **What is the spectrum of the operator $A$ without a singular system and its degree of ill posedness?** This research question has been answered by Theorem 4.6, Proposition 4.5, and Corollary 4.9. In particular, the spectrum is given by $\sigma(A) = \sigma_{\mathrm{ac}}(A) = [0,1]$ and the point and singular continuous spectra are: $\sigma_{\mathrm{pp}}(A) = \sigma_{\mathrm{sc}}(A) = \varnothing$. The exact two sided Hilbert scale constants for the operator are $\|A^{-1}\| = 1$ and $\|A\| = c_\sharp$, and its degree of ill posedness is exactly two.

2. **What is the modulus of continuity of the inverse on $\mathcal{W}_{\ell,E}$ for any finite level of noise $\delta$, and not just asymptotically?** Theorem A, which uses the complete monotonicity certificate from Proposition 4.17.

3. **Is the conditional Hölder bound on $\mathcal{W}_{\ell,E}$ tight in terms of the constant, or only in the rate?** Theorem 4.11 and Theorem A show that the constant $(2E)^{2/(\ell+2)}$ is precisely the worst-case constant, and there can be no improvement.

4. **What is the minimum worst case error of any reconstruction map, and is it achieved?** The answer to this question is in Theorem B, where we can compute it explicitly for the current Non compact source case, and it is achieved through linear Hilbert scale smoothing.

5. **In what way is the quality of reconstruction impacted if one knows the forward operator only to within an $\varepsilon$ perturbation?** Answer: As shown in Corollary 4.12, $\delta$ and $\varepsilon$ occur together in the sharp bound.

6. **What does one gain from having $J$ observation times versus a single observation time, and can this be negated in any way?** Answer: As stated in Theorem 4.22, there is an exact $\sqrt{J}$ improvement law, which comes along with an exact collapse law for the worst possible coherent noise.

7. **Is the source map defined by the training of Hermite variational PINNs exactly the same as classical spectral regularization(Tikhonov Regularization in this case), or does this only hold true in an approximate sense?** An answer comes from Theorem C, in fact, there is an exact algebraic identity with a Tikhonov regularization in the Hermite Sobolev scale, even discretely and without requiring compactness or SVD.

8. **Are there ways to confirm the assumptions under which that estimator is order optimal for the actual trained network, instead of making assumptions?** An answer comes from Section 4.4.6: in fact, both assumptions go down to a-posteriori quantities which can be obtained from the training log, and then we can derive a certified error bound using Corollary 4.26.

9. **Is the proposed methodology still valid in the nonlinear forward problem setting where Fourier multiplier methods, such as mollification with the Dirichlet kernel or other classical methods, cannot be applied or are very difficult to apply and implement?** This question is considered in Section 4.1.4 and Algorithm 2, where it is demonstrated that the only differences between the linear and nonlinear approaches in our Method are in a residual term and a data generator, while the two formulations give almost the same accuracy in the source recovery.

10. **Do the exact statements hold true if the computational algorithm uses only definitions and never the closed forms?** The answer can be found in Section 4.5.6, and the resulting evaluation of the methodology's place in comparison with classical approaches is given in Chapter 5.

Variable Hilbert scales as studied in [87, 104]; the exactness mechanism proposed by [134]; the optimal order Fourier analysis method [65, 142]; the non compact degree of ill posedness approach [23, 83, 144]; the optimal recovery approach [63, 106]; the complete monotonicity approach [131], the Hermite scale and corresponding rate of convergence [34, 35, 107]; and the PINN generalisation theory [109]. There is no claim of novelty made regarding these or the existence of the Hermite Sobolev scale or the rate $\delta^{s/(s+2)}$.

### 4.1.3 Problem Formulation

In this work, $\varphi : \mathbb{R} \to \mathbb{R}$ is the unknown source that we want to recover which is time independent. given $\varphi$ the resulting state $v$ satisfies the Cauchy problem:

$$\partial_t v(x,t) - \partial_{xx} v(x,t) = \varphi(x), \qquad v(x,0) = 0, \qquad \lim_{|x|\to\infty} v(x,t) = 0, \tag{4.1}$$

on $\mathbb{R} \times (0,1]$. By Proposition 1.4 we showed that for every $\varphi \in L^2(\mathbb{R})$, this equation has a unique mild solution, $v \in C([0,1]; L^2(\mathbb{R})) \cap L^2(0,1; H^1(\mathbb{R}))$.

The terminal time map is then defined as:

$$A : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}), \qquad A\varphi := v(\,\cdot\,, 1), \tag{4.2}$$

is a well defined bounded linear operator. Multiplier representation, spectral resolution, and ill posedness profile will be discussed in Section 4.3.

The inverse problem is to recover $\varphi^\dagger$ from a single noisy observation of the terminal state, It is the first of the two inverse source problems studied by [147], that we investigate in this dissertation.the source is restricted to the a-priori source class of Definition 1.8:

$$\varphi^\dagger \in \mathcal{W}_{\ell,E} = \big\{\varphi \in H^\ell(\mathbb{R}) : \|\varphi\|_{H^\ell(\mathbb{R})} \le E\big\}, \qquad \ell \ge 1,\ E > 0 \text{ fixed}, \tag{4.3}$$

by [66, Assumption (A1)] we define:

$$g^\delta = A\varphi^\dagger + \delta\,\eta, \qquad \eta \in L^2(\mathbb{R}),\ \|\eta\|_{L^2(\mathbb{R})} \le 1,\ \delta > 0. \tag{4.4}$$

$\delta$ measures sensor noise, and the source ball is defined with a norm of a Hilbert scale [68, §8.4], following the source sets $\mathcal{M}_\rho^s$ of [66, §1].

> **Remark 4.1.** in Section 4.5 experiments use [147]:
>
> $$\varphi^\dagger(x) := \left(\tfrac{x^3}{4} - \tfrac{3x}{2}\right) e^{-x^2/4}, \qquad \|\varphi^\dagger\|_{L^2(\mathbb{R})} \approx 1.533. \tag{4.5}$$
>
> At the natural scale $\sigma = \sqrt{2}$ it lies in the two dimensional span $\operatorname{span}\{\psi_1^{\sqrt{2}}, \psi_3^{\sqrt{2}}\} \subset \mathcal{S}(\mathbb{R})$ of the scaled Hermite functions of Definition 1.13, so $\varphi^\dagger \in \mathcal{W}_{\ell,E}$ for every $\ell \ge 0$, a Schwartz class source with, in effect, arbitrarily high Hilbert scale smoothness.

### 4.1.4 Nonlinear Extension

The Hermite V-PINN **easily** extends to the hybrid equation combining the nonlinear convection term of the viscous Burgers equation [41] with a cubic reaction term inspired by the Allen Cahn equation [39] inverse source problem:

$$\partial_t v - \partial_{xx} v - v\,\partial_x v + v^3 = \varphi(x), \qquad v(x,0) = 0, \tag{4.6}$$

on $\mathbb{R} \times (0,1]$, only three changes are enough:

In the strong form residual we have a new term: $-v\,\partial_x v + v^3$.

And a new term in the weak form residual: $+\frac{1}{2}\int_0^1\int_{\mathbb{R}} v^2\partial_x\psi\, q\, \mathrm{d}x\, \mathrm{d}t.$ is derived from rewriting $-v\partial_x v$ as $-\partial_x(v^2/2)$, then integrating by parts to move the derivative to $\psi$, and because the residual starts with $-v\partial_x v$, that move makes it positive.

The data generation changes too. Here we replace Crank Nicolson (Section 1.4.2) by the IMEX method of Ascher, Ruuth, and Wetton [11, §3.1, $\gamma = \frac{1}{2}$]. Diffusion is still handled as before (by Crank Nicolson tridiagonal solve), but the advection term $v\partial_x v$ and the cubic term $v^3$ are calculated from the previous step instead, on the same grid $N_x = 10{,}000$, $N_t = 5{,}000$. The bilevel structure, the Tikhonov and hybrid loss, and the Hermite source stay the same.

### *4.1.5 Anti Inverse Crime*

An inverse crime occurs when an inverse method is tested using data generated from the exact same forward model that was used for reconstruction. This leads to an idealized scenario where errors are able to cancel each other. Consequently, the method might seem more accurate than it really is. To avoid this, testing should be performed with realistic data or a different forward model.[145],[32]

So we introduce a mismatch between the model that generates the data and the model that inverts it. this is because any error appearing in Section 4.5 can be trusted to be a property of the method itself, not an artifact of a favorable experiment.

So All sensor data in Section 4.5 are generated from the perturbed variable coefficient model

$$\partial_t v_\varepsilon - \partial_x\big(a(x)\,\partial_x v_\varepsilon\big) = \varphi(x), \qquad a(x) := 1 + \varepsilon\, e^{-x^2/4}, \quad \varepsilon = 10^{-3}, \tag{4.7}$$

on the truncated domain $[-30, 30]$, while the V-PINN inverts the constant coefficient equation (4.1). Equation (4.7) is integrated by the Crank Nicolson scheme of Section 1.4.2 with $N_x = 10{,}000$ and $N_t = 5{,}000$.

Three more safties also reinforce the mismatch (finite differences for the data, Gauss-Hermite × Gauss-Legendre for the inversion, Section 1.4.3), distinct domains ($[-30, 30]$ vs $\mathbb{R}$); and, we hold out 60 of the 600 uniformly spaced measurements on $[-10, 10]$

we can measure the gap with $w := v_\varepsilon - v$ which satisfies $w(\cdot, 0) = 0$ and $\partial_t w - \partial_x(a(x)\partial_x w) = \partial_x((a(x) - 1)\partial_x v)$, By standard parabolic energy estimate [70, Ch. 7, §7.1.2] and uniform ellipticity with $\|a - 1\|_{L^\infty(\mathbb{R})} \le \varepsilon$ we get $\|w(\cdot, 1)\|_{L^2(\mathbb{R})} \le C\varepsilon\|\varphi\|_{L^2(\mathbb{R})}$, so

$$\|A_\varepsilon - A\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \le C\varepsilon = \mathcal{O}(\varepsilon). \tag{4.8}$$

For $\varepsilon = 10^{-3}$ this basically keeps the operator mismatch below the percent level, which is confirmed numerically in Section 4.5.

## 4.2 The Hermite V-PINN Algorithm: Linear and Nonlinear Implementations

In this section we provide a description of our approach in four parts: architecture, discrete loss, norm bound and algorithms. The architecture of Section 4.2.1 involves parameterizing into two pieces, a tanh network (4.10) for the state variable, and a Hermite expansion (4.11) for the source, which is linear in $c$. Section 4.2.2 includes the aggregation of five terms to form (4.18); since each residual is affine in $c$, the loss is quadratic in $c$, making it strongly convex for a fixed $\theta$, as per Proposition 3.7. The optimum point for $c$, given by (3.12) and built on the Gram matrix (4.17), is the Hilbert scale regulariser as given by Theorem C. Section 4.2.3 links the penalty norm $\|c\|_\Lambda$ and the norm $\|\varphi^c\|_{H^\ell(\mathbb{R})}$ used in Section 4.4,

$$\|\varphi^c\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}\,\|c\|_\Lambda, \tag{4.9}$$

with $C_{2,2} \le \sqrt{105}/8$ for $(\sigma,\ell) = (2,2)$, as Theorem 4.11 needs. From this, Section 4.2.4 and Section 4.2.5 give Algorithm 1 and Algorithm 2, which vary only in their residual terms and data solvers.

### *4.2.1 Architecture*

The architecture has two parts: there is a neural network for the state and a fixed Hermite expansion that represents the source, which is linear in its coefficients, making the loss function quadratic in $c$. The trial state is the tanh network from Definition 3.1, with depth $D$ and width $W$, and parameters that follow the initialisation from Definition 3.2:

$$v_\theta(x,t) := t\cdot e^{-\alpha_v x^2}\cdot \mathcal{N}_\theta(x,t), \qquad \alpha_v = 0.08,\ W = 192,\ D = 7. \tag{4.10}$$

Two important things are that, first of all, the factor $t$ enforces $v_\theta(\,\cdot\,,0) = 0$, as a hard constraint and not a loss term [18], and secondly that the envelope $e^{-\alpha_v x^2}$ ensures $v_\theta(\,\cdot\,,t) \in H^k(\mathbb{R})$ for all $k \ge 0$ on bounded $\theta$-sets. So the trial state is smooth, with Gaussian decay, and its residual derivatives are obtained through automatic differentiation. The trial source is the Hermite spectral expansion (Meng and Zhao [107]):

$$\varphi^c(x) := \sum_{j=0}^{N_\varphi - 1} c_j\,\psi_j^\sigma(x), \qquad \sigma = 2,\ N_\varphi = 64, \tag{4.11}$$

in the orthonormal basis defined in Definition 1.13, where the mass and design matrices are:

$$\Lambda := \operatorname{diag}\big((2j+1)^\ell\big)_{j=0}^{N_\varphi-1}, \qquad \Phi_{(i,k),j} := \psi_j^\sigma(x_i^H), \tag{4.12}$$

so $\Lambda \in \mathbb{R}^{N_\varphi\times N_\varphi}$ and $\Phi \in \mathbb{R}^{M_r\times N_\varphi}$, the objects of Definition 1.14 at order $\ell = \ell_{\text{reg}} = 2$, with $\|c\|_\Lambda^2 = c^\top \Lambda c$ and $\|\varphi^c\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}\|c\|_\Lambda$ (Lemma 4.2). We then integrate the weak form residual with respect to the tensor test space,

$$\mathbb{V}_K := \operatorname{span}\{\psi_0^\sigma,\dots,\psi_{K-1}^\sigma\}, \qquad \mathbb{Q}_{M_L} := \operatorname{span}\{q_0,\dots,q_{M_L-1}\}, \tag{4.13}$$

with $K = 32$ Hermite modes in space and $M_L = 10$ fixed weighting functions in time, the constant function $q_0(t) \equiv 1$ and nine Gaussian functions localised around $t = 1$,

$$q_0(t) \equiv 1, \qquad q_m(t) = \exp\!\big(-((t - c_m)/\tau)^2\big), \quad m = 1, \dots, 9, \tag{4.14}$$

of width $\tau = 0.05$ centred at $\{c_m\} = \{1.0, 0.97, 0.93, 0.88, 0.82, 0.74, 0.64, 0.52, 0.38\}$; these are the test functions; they do not include the quadrature rule that appears below, where $\Pi$ is the orthogonal projection onto $\mathbb{V}_K \otimes \mathbb{Q}_{M_L}$. Gauss Hermite quadrature is applied with $N_{\mathrm{GH}} = 64$ nodes $(z_i, w_i)$ (Definition 1.22), rescaled with $\sigma_q = 2.5$ and de-weighted; thus the spatial integral is a Lebesgue rule on $\mathbb{R}$,

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x \;\approx\; \sum_{i=1}^{N_{\mathrm{GH}}} w_i^H\, f(x_i^H), \qquad x_i^H = \sigma_q z_i, \quad w_i^H = \sigma_q\, w_i\, e^{z_i^2}, \tag{4.15}$$

along with $N_{\mathrm{GL}} = 64$ Gauss Legendre nodes on $[0, 1]$ in time (Definition 1.23;). The errors are computed on the sensors and not on the quadrature nodes: $N_{\mathrm{sensor}} = 600$ sensors are uniform over $[-10, 10]$ (Section 4.1.5), where $M_d = 540$ are used for training and the other 60 are kept aside. For the strong form residual, we use $N_{\mathrm{coll}} = 8192$ points, redrawn at each iteration on $[-L_{\mathrm{box}}, L_{\mathrm{box}}] \times (0, 1)$, with $L_{\mathrm{box}} = 8$.

### *4.2.2 The Hybrid Discrete Loss*

The PDE residual associated with the trial pair $(\theta, c)$ is:

$$R_{\mathrm{PDE}}(\theta, c)(x, t) := \partial_t v_\theta(x, t) - \partial_{xx} v_\theta(x, t) - \varphi^c(x), \tag{4.16}$$

and the regularised Gram matrix associated with the inner $c$-quadratic is:

$$\Gamma(\mu_{\mathrm{coll}}, \mu_{\mathrm{wf}}, \lambda) := \mu_{\mathrm{coll}}\, \mathbf{H}_{\mathrm{coll}} + \mu_{\mathrm{wf}}\, \mathbf{H}_{\mathrm{wf}} + \lambda\, \Lambda \;\in\; \mathbb{R}^{N_\varphi \times N_\varphi}. \tag{4.17}$$

This is the matrix that is inverted to solve for $c$. There are 2 blocks, $\mathbf{H}_{\mathrm{coll}}$ and $\mathbf{H}_{\mathrm{wf}}$, that are both symmetric and positive semidefinite. $\mathbf{H}_{\mathrm{coll}}$ comes from checking $R_{\mathrm{PDE}}$ at scattered points, which is the strong form check; $\mathbf{H}_{\mathrm{wf}}$ comes from checking $R_{\mathrm{PDE}}$ against the test functions $\psi_k^\sigma, q_m$, which is the weak form check. Both are built using $\Phi$, which turns plugging $c$ into $R_{\mathrm{PDE}}$ into one matrix multiply.

in Proposition 3.7 we proved the closed form solve that is valid in the case of one residual block but here we have two blocks (the strong form and weak form, each with their own weight), but splitting one block into two does not bring any changes, the closed form answer (3.12) with all the same guarantees remains true, and since $\Lambda \succ 0$, $\Gamma$ is strictly positive definite for every $\lambda > 0$.

The total loss is :

$$L(\theta, c) := w_{\mathrm{m}}\, L_{\mathrm{d}}(\theta) + \mu_{\mathrm{coll}}\, L_{\mathrm{c}}(\theta, c) + \mu_{\mathrm{wf}}\, L_{\mathrm{w}}(\theta, c) + w_{\mathrm{b}}\, L_{\mathrm{b}}(\theta) + \lambda\, \|c\|_\Lambda^2, \tag{4.18}$$

with the terms:

$$L_{\mathrm{d}}(\theta) := \frac{1}{M_d}\sum_{i=1}^{M_d}|v_\theta(x_i,1)-g_i^\delta|^2, \qquad \{x_i\}_{i=1}^{M_d} \text{ the 540 training sensors,} \tag{4.19}$$

$$L_{\mathrm{c}}(\theta,c) := \frac{1}{N_{\mathrm{coll}}}\sum_{p=1}^{N_{\mathrm{coll}}}|R_{\mathrm{PDE}}(\theta,c)(x_p,t_p)|^2, \quad (x_p,t_p)\sim\mathcal{U}\big([-L_{\mathrm{box}},L_{\mathrm{box}}]\times(0,1)\big), \tag{4.20}$$

$$L_{\mathrm{w}}(\theta,c) := \frac{1}{KM_L}\sum_{k=0}^{K-1}\sum_{m=0}^{M_L-1}|b_{k,m}(\theta,c)|^2, \tag{4.21}$$

$$b_{k,m}(\theta,c) := \int_0^1\int_{\mathbb{R}} R_{\mathrm{PDE}}(\theta,c)(x,t)\,\psi_k^\sigma(x)\,q_m(t)\,\mathrm{d}x\,\mathrm{d}t, \tag{4.22}$$

$$L_{\mathrm{b}}(\theta) := \frac{1}{N_b}\sum_{n=1}^{N_b}\Big[v_\theta(-L_{\mathrm{box}},t_n)^2+v_\theta(L_{\mathrm{box}},t_n)^2\Big], \quad t_n\in[0,1] \text{ uniform.} \tag{4.23}$$

The data and strong form terms are unweighted Monte Carlo averages (4.19) over the 540 training sensors, The average in (4.20) is also computed over 8192 points. Unlike the sensor locations, however, these points are not fixed, they are randomly chosen again at every training step. the integrals (4.22) are instead computed on the fixed Gauss Hermite × Gauss Legendre grid of (4.15), not by random sampling; $L_{\mathrm{w}}$ is normalized by the total number of test function pairs so that its natural size matches the other four terms, which basically means that the weight $\mu_{\mathrm{wf}} = 1$ keeps this term balanced. The boundary term (4.23) just checks the guess is near zero at the far edges but network's Gaussian shape already forces that so this term is kept only as a backup at $x = \pm L_{\mathrm{box}}$.

The coefficients $b_{k,m}$ from (4.22) are not computed by just plugging $R_{\mathrm{PDE}}$ straight into the integral, with $\partial_{xx}v_\theta$ left as it was defined. Instead, that term is rewritten first, using integration by parts. For the linear PDE (4.1), this gives:

$$b_{k,m}^{\mathrm{lin}}(\theta,c) = \int_0^1\int_{\mathbb{R}}\Big[\partial_t v_\theta\,\psi_k^\sigma+\partial_x v_\theta\,\partial_x\psi_k^\sigma\Big]q_m\,\mathrm{d}x\,\mathrm{d}t-\int_0^1\int_{\mathbb{R}}\varphi^c\,\psi_k^\sigma q_m\,\mathrm{d}x\,\mathrm{d}t. \tag{4.24}$$

Here, the second $x$-derivative that used to sit on $v_\theta$ has been moved onto $\psi_k^\sigma$ instead (By Integration by parts).normally there will also be an extra boundary term, but that term is zero here, because both $v_\theta$ and $\psi_k^\sigma$ shrink to zero fast as $x\to\pm\infty$.

For the nonlinear PDE (4.6), the residual has an extra term, $-v\,\partial_x v$. Since $-v\,\partial_x v = -\partial_x(v^2/2)$, it also has a derivative, so the same move can be applied to it too, giving:

$$b_{k,m}^{\mathrm{nl}}(\theta,c) = b_{k,m}^{\mathrm{lin}}(\theta,c)+\tfrac{1}{2}\int_0^1\int_{\mathbb{R}}v_\theta^2\,\partial_x\psi_k^\sigma\,q_m\,\mathrm{d}x\,\mathrm{d}t+\int_0^1\int_{\mathbb{R}}v_\theta^3\,\psi_k^\sigma\,q_m\,\mathrm{d}x\,\mathrm{d}t. \tag{4.25}$$

The weight values we used in the implementation are:

$$w_{\mathrm{m}} = 500, \quad \mu_{\mathrm{coll}} = 5, \quad \mu_{\mathrm{wf}} = 1, \quad w_{\mathrm{b}} = 5, \quad \lambda\approx 10^{-9}. \tag{4.26}$$

As observed, there is a huge gap between $w_{\mathrm{m}}$ and the residual weights, and the gap is not really arbitrary, it is justified by Corollary 4.12, which tells us the reconstruction error is bounded by

$$\|\widehat{\varphi}-\varphi^\dagger\|_{L^2(\mathbb{R})} \le (2E)^{2/(\ell+2)}\,(\delta+\varepsilon)^{\ell/(\ell+2)}, \tag{4.27}$$

with $\varepsilon := \|A\widehat{\varphi} - g^{\delta}\|_{L^2(\mathbb{R})}$ is precisely what $L_{\mathrm{d}}$ is trying to minimize. Since this corollary of the sharp stability estimate shows that $\varepsilon$ controls how close the reconstruction gets to the truth, it makes sense to weight $L_{\mathrm{d}}$ heavily.

Everything so far ($w_{\mathrm{m}} = 500$, $\lambda \approx 10^{-9}$) was for the noise free case, $\delta = 0$. The moment noise is added, two of the five weights get changed to different formulas, $\lambda$ and $w_{\mathrm{m}}$ before training begins.

$$\lambda(\delta) = \max\bigl(10^{-9},\ 3{\cdot}10^{-4}\,\delta^{2/3}\bigr), \qquad w_{\mathrm{m}}(\delta) = \frac{500}{1 + (\delta/0.02)^2}. \tag{4.28}$$

$\lambda(\delta)$ as the noise $\delta$ grows, $\lambda$ grows too because noisier data needs more regularisation, but it never really falls below $10^{-9}$, the numerical safety floor from before. also $w_{\mathrm{m}}(\delta)$ shrinks as $\delta$ grows

One of the main difficulties in PINN training is to ensure that loss terms with different scales are balanced properly. if we do not, one term can quietly dominate the others [143]. what we did is as $\delta$ increases, the data become less trustworthy, so we reduce the weight of the data term, to avoid fitting noise by the network instead of the desired solution. for example, at $\delta = 0.02$, $w_{\mathrm{m}}(\delta) = 500/(1+1) = 250$ which is half the baseline. we only $\lambda$ and $w_{\mathrm{m}}$. the two PDE residual weights $\mu_{\mathrm{coll}}, \mu_{\mathrm{wf}}$ are kept fixed.

### 4.2.3 The Hermite to Sobolev Norm Bound

The loss function penalizes the coefficients according to:

$$\|c\|_\Lambda^2 = \sum_j (2j+1)^\ell c_j^2.$$

However, in theory, things are different. The Sobolev norm of the function $\varphi^c(x)$ is what is used,

$$\|\varphi^c\|_{H^\ell(\mathbb{R})},$$

and this is the norm that defines the source class $\mathcal{W}_{\ell,E}$, this can be seen in the stability bound of Theorem 4.11. This is the reason why we need to connect the two norms, which is done in Lemma 4.2.

**Lemma 4.2.** *Let $\sigma > 0$ and $\ell \in \{0,1,2\}$. then For all $c \in \mathbb{R}^{N_\varphi}$ the truncated Hermite source $\varphi^c = \sum_{j=0}^{N_\varphi - 1} c_j \psi_j^\sigma$ satisfies:*

$$\|\varphi^c\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}\, \|c\|_\Lambda, \tag{4.29}$$

*with:*

$$C_{\sigma,0}^2 = 1, \qquad C_{\sigma,1}^2 \le 1 + \sigma^{-2}, \qquad C_{\sigma,2}^2 \le 1 + 2\sigma^{-2} + \tfrac{9}{4}\sigma^{-4}. \tag{4.30}$$

*In case $(\sigma,\ell) = (2,2)$,*

$$C_{2,2}^2 \le \tfrac{105}{64}, \qquad C_{2,2} \le \tfrac{\sqrt{105}}{8} \approx 1.281, \tag{4.31}$$

*and the multiplicative inflation $C_{2,2}^{2/(\ell+2)} = C_{2,2}^{1/2} \le (105/64)^{1/4} \approx 1.132$ enters the assembled bound of Theorem 4.11 via the substitution $E \leftarrow C_{\sigma,\ell}\, E_\Lambda$.*

**Proof.** The case $\ell = 0$ is only orthonormality of $\{\psi_j^\sigma\}_{j\ge 0}$ (Definition 1.13). With an orthonormal basis, we see that $\|\varphi^c\|_{L^2(\mathbb{R})}^2 = \sum_j c_j^2$, without any approximation involved, and since $(2j+1)^0 = 1 \le (2j+1)^\ell$ for all $\ell \ge 0$, the sum will be smaller than or equal to $\|c\|_\Lambda^2$. Consequently, we get $\|\varphi^c\|_{L^2(\mathbb{R})}^2 \le \|c\|_\Lambda^2$.

For the case of $\ell = 1$, one observation from calculus that is required is that differentiating a Hermite function never spreads it across the whole basis, it only reaches its two immediate neighbours:

$$\frac{\mathrm{d}}{\mathrm{d}x}\psi_j^\sigma = \frac{1}{\sigma}\Big(\sqrt{\tfrac{j}{2}}\,\psi_{j-1}^\sigma - \sqrt{\tfrac{j+1}{2}}\,\psi_{j+1}^\sigma\Big), \qquad j \ge 0, \tag{4.32}$$

with $\psi_{-1}^\sigma \equiv 0$. This is a classical ladder identity, based on the standard Hermite derivative relations [62, §18.9(iii), eqs. (18.9.25)-(18.9.26)] [137, §1.1], so it is cited here instead of proving it from the beginning.

We apply (4.32) to all terms of $\varphi^c$ and then collect according to which Hermite basis function each term contributes to. We obtain $\partial_x \varphi^c$, now written as a Hermite series:

$$\partial_x \varphi^c = \sum_{k\ge 0} \frac{1}{\sigma}\Big(\sqrt{\tfrac{k+1}{2}}\, c_{k+1} - \sqrt{\tfrac{k}{2}}\, c_{k-1}\Big)\psi_k^\sigma,$$

with $c_j := 0$ for all indices not belonging to the set $\{0, \dots, N_\varphi - 1\}$, hence the summation can be done over all $k \geq 0$. By squaring each coefficient and using the bound $(a-b)^2 \leq 2(a^2+b^2)$, a summation and the combination of the two terms of each $c_j^2$ yields:

$$\|\partial_x \varphi^c\|^2_{L^2(\mathbb{R})} \leq \sigma^{-2} \sum_{k\geq 0} \Big( (k+1)c_{k+1}^2 + k\, c_{k-1}^2 \Big) = \sigma^{-2} \sum_{j\geq 0} (2j+1) c_j^2 \leq \sigma^{-2} \|c\|^2_\Lambda,$$

Finally, observe that $(2j+1) \leq (2j+1)^\ell$. Since the $\|f\|^2_{H^1(\mathbb{R})}$ is simply the $L^2(\mathbb{R})$ norm of the function itself and the $L^2(\mathbb{R})$ norm of its derivative, $\|f\|^2_{H^1(\mathbb{R})} = \|f\|^2_{L^2(\mathbb{R})} + \|\partial_x f\|^2_{L^2(\mathbb{R})}$, we add this bound to the $\ell = 0$ case to get $C^2_{\sigma,1} \leq 1 + \sigma^{-2}$.

And that is what we do for the situation $\ell = 2$ as well, just one level deeper. If we differentiate the ladder equation twice, then another basis function two steps away also contributes:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} \psi_j^\sigma = \sigma^{-2} \Big( \tfrac{1}{2}\sqrt{(j-1)j}\, \psi_{j-2}^\sigma - \tfrac{2j+1}{2}\, \psi_j^\sigma + \tfrac{1}{2}\sqrt{(j+1)(j+2)}\, \psi_{j+2}^\sigma \Big),$$

with the cross terms cancelling in the same ladder algebra as before. Now the coefficient of $\psi_k^\sigma$ in $\partial_{xx}\varphi^c$ is made up of three terms, $c_{k-2}, c_k, c_{k+2}$, rather than two, and we use Cauchy Schwarz to bound the following:

$$\|\partial_{xx}\varphi^c\|^2_{L^2(\mathbb{R})} \leq \frac{3}{4\sigma^4} \sum_{j\geq 0} \Big( (j-1)j + (2j+1)^2 + (j+1)(j+2) \Big) c_j^2.$$

It's all algebra from here. The three terms in the summation add up and are equal to $(j-1)j + (2j+1)^2 + (j+1)(j+2) = 3(2j^2+2j+1)$, and when comparing $2j^2+2j+1$ against $(2j+1)^2 = 4j^2+4j+1$ one finds that the difference between them is $2j(j+1) \geq 0$ for all $j \geq 0$, so that $2j^2+2j+1 \leq (2j+1)^2$ always, with equality only at $j=0$. Therefore $\|\partial_{xx}\varphi^c\|^2_{L^2(\mathbb{R})} \leq (9/(4\sigma^4))\|c\|^2_\Lambda$. The $H^2(\mathbb{R})$ norm puts weight on the first derivative with a factor of 2 and the second derivative with a factor of 1, $\|f\|^2_{H^2(\mathbb{R})} = \|f\|^2_{L^2(\mathbb{R})} + 2\|\partial_x f\|^2_{L^2(\mathbb{R})} + \|\partial_{xx} f\|^2_{L^2(\mathbb{R})}$, thus by putting all three inequalities together, we have: $C^2_{\sigma,2} \leq 1 + 2\sigma^{-2} + \frac{9}{4}\sigma^{-4}$.

Finally, we substitute the true values, $(\sigma, \ell) = (2, 2)$: $\sigma^{-2} = \frac{1}{4}$, so $2\sigma^{-2} = \frac{1}{2}$, and $\sigma^{-4} = \frac{1}{16}$, so $\frac{9}{4}\sigma^{-4} = \frac{9}{64}$. Putting $1 + \frac{1}{2} + \frac{9}{64}$ over a common denominator 64, this equals $C^2_{2,2} \leq \frac{105}{64}$, the constant used later in the a-posteriori bound. □

### 4.2.4 Algorithm 1: Linear Training

**Algorithm 1** Hermite V-PINN - linear inverse heat source problem.

**Require:** noisy sensor data $\{(x_i, g_i^\delta)\}_{i=1}^{N_{\text{sensor}}}$ on $[-10, 10]$, with $N_{\text{sensor}} = 600$; architectural and numerical hyperparameters as in (4.10)–(4.26).

Split sensors: 540 training, 60 held out validation.

Initialise $\theta_0$ by the framework default Kaiming-uniform scheme (Definition 3.2); set $c_0 = \mathbf{0} \in \mathbb{R}^{N_\varphi}$.

*All Adam phases use one cosine schedule [100] over the full* 34,000 *steps,* $10^{-3} \to 5\times10^{-5}$*, with global gradient norm clipping at* 1 *[118, §3.2, Algorithm 1].*

**Phase 1 - Measurement warm up** (5,000 **Adam steps,** $c$ **frozen).**

**for** $n = 1, \dots, 5000$ **do**

  Update $\theta$ by Adam [93] on $w_{\text{m}} L_{\text{d}}(\theta) + w_{\text{b}} L_{\text{b}}(\theta)$.

**end for**

**Phase 2 - Source focus** (4,000 **Adam steps,** $\theta$ **frozen).**

**for** $n = 1, \dots, 4000$ **do**

  Update $c$ by Adam on $\mu_{\text{coll}} L_{\text{c}}(\theta, c) + \lambda \|c\|_\Lambda^2$.

**end for**

**Phase 3 - Hybrid joint training** (25,000 **Adam steps).**

**for** $n = 1, \dots, 25000$ **do**

  Ramp factor $r_n := \min(1, n/1500)$. ▷ linear warm-in of *both* PDE terms over 1,500 steps

  Assemble $L := w_{\text{m}} L_{\text{d}} + w_{\text{b}} L_{\text{b}} + r_n \mu_{\text{coll}} L_{\text{c}} + \lambda \|c\|_\Lambda^2$.

  **if** $n \bmod 5 = 0$ **or** $n = 1$ **then**

    Recompute $L_{\text{w}}(\theta, c)$ on the Gauss tensor grid and add $r_n \mu_{\text{wf}} L_{\text{w}}$ to $L$. ▷ the weak-form term enters the gradient only on these steps

  **end if**

  Update $(\theta, c)$ by Adam on $L$.

**end for**

**Phase 4 - L-BFGS endgame (full-batch,** $\leq$ 2,000 **steps).**

Refine $(\theta, c)$ by L-BFGS [99] on the full loss (4.18) (all weights at $r_n \equiv 1$), strong Wolfe [16] line search, history size 100, gradient/step tolerances $10^{-14}$.

**return** $\widehat{\varphi} := \varphi^c = \sum_{j=0}^{N_\varphi - 1} c_j \psi_j^\sigma$.

### 4.2.5 Algorithm 2: Nonlinear Training

**Algorithm 2** Hermite V-PINN - nonlinear inverse source for $\partial_t v - \partial_{xx} v - v\,\partial_x v + v^3 = \varphi$.

**Modifications relative to Algorithm 1:**

Replace the residual $R_{\text{PDE}}$ in (4.16) by

$$R_{\text{PDE}}^{\text{nl}}(\theta, c) := \partial_t v_\theta - \partial_{xx} v_\theta - v_\theta \partial_x v_\theta + v_\theta^3 - \varphi^c.$$

Replace the weak form coefficient $b_{k,m}^{\text{lin}}$ in (4.24) by $b_{k,m}^{\text{nl}}$ in (4.25).

Replace the data generator: integrate (4.6) rather than (4.7) on $[-30, 30]$ by an implicit-explicit (IMEX) scheme [11]: Crank Nicolson for the diffusion, explicit (lagged) treatment of the advection and cubic reaction $v\,\partial_x v - v^3$, with $N_x = 10{,}000$ and $N_t = 5{,}000$.

**everything else and all other steps (Phases 1-4, hyperparameters,** 540+60 **sensor split, optimiser sequence) do not change.**

Training the network based on Algorithm 1 consists of four phases, Phase 1 is fitting the state to the observations keeping the source zero. Phase 2 is freezing the state and fitting the source to the state through the PDE. Phase 3 is the simultaneous fitting of both, gradually including the PDE terms. At the final phase, 4, L-BFGS refinement is done. The output is the Hermite source $\widehat{\varphi}$. Algorithm 2 works the same way as the linear algorithm, but with three differences: two additional terms in the residual, the integration by parts of the convective term in the weak formulation, and IMEX discretisation of the data rather than the Crank Nicolson one. All the rest is the same, which is the reason why the accuracy comparisons make sense.

*Adam.* Iterative algorithm with gradient descent approach where step size is independently changed for each parameter. It is quite insensitive to the initial choice of point, but cannot achieve high accuracy.

*L-BFGS.* This algorithm computes an approximate inverse Hessian based on the last 100 update pairs. It has a bad initial convergence rate, but it is extremely accurate as we approach the optimum.

*Strong Wolfe condition.* This is the acceptance condition that L-BFGS uses, and it requires a sufficient reduction in the value of the function as well as the slope magnitude being reduced sufficiently. It requires the function to be the same function at each test point.

*Cosine Annealing.* In this method, the learning rate decreases from $10^{-3}$ to $5 \times 10^{-5}$ using the cosine schedule throughout all 34,000 Adam steps.

*Clipping the gradient to* 1. If the norm of the gradient exceeds 1, it will be reduced to 1 to ensure that one bad batch doesn't ruin all of our parameters.

*Kaiming-uniform initialisation.* Default randomly generated values, divided by the square root of the number of inputs for each layer, to avoid any vanishing/exploding gradients.

*Frozen.* It is frozen and not updated during the process. The lock on $c$ occurs during Phase 1 while the lock on $\theta$ occurs in Phase 2.

*Scaling Ramp* $r_n = \min(1, n/1{,}500)$. A scaling factor that increases linearly from 0 to 1 over 1,500 training steps, letting the PDE terms come into effect progressively.

*Full batch.* The computation of the loss function takes into account all the data points, without sampling randomly. This is required by the L-BFGS method.

*Sensor hold-out.* Out of the 600 sensors, there are sixty that never contribute to the loss function; their misfit represents an out-of-sample check.

*IMEX.* Implicit-explicit scheme: the implicit component (the Crank Nicolson scheme) is employed to handle the stiff diffusion term, while the nonlinear component is calculated based on the data of the previous time step, as this is less stiff than the stiff diffusion term. Hence, there is no need for a nonlinear solve at each time step.

*Strong formulation / Weak formulation Residuals.* For the strong formulation, the residual is tested at 8192 random points, while in the weak formulation the PDE residual is tested in the integral sense against test functions at fixed quadrature points.

## 4.3 The Forward Operator: Analysis and Spectral Properties

In this section, we shall consider $A$, smoothing of $A$, spectrum of $A$, and one particular inequality for $m$.

From Theorem 4.4, the operator $A$ is bounded with $\|A\| = 1$ approached but not attained, self adjoint, positive, injective, of dense but not closed range, and non compact: the middle two lead to Hadamard ill posedness, and non compactness implies that there are no eigenvectors. From Proposition 4.5, we have that the operator $A\colon H^s(\mathbb{R}) \to H^{s+2}(\mathbb{R})$ is an isomorphism with sharp constants 1 and $c_\sharp \approx 1.2984$. Therefore, ill posedness results from the disparity between the $L^2(\mathbb{R})$ norm of the data and the $H^{s+2}(\mathbb{R})$ norm of the range, measured in Section 4.4.3.

From Theorem 4.6, we have that the spectrum is absolutely continuous, thus the extrema from the minimax characterization will not be reached, but approached. This is different from the compact bounded domain, where eigenfunctions exist and the extrema are attained, so on $\mathbb{R}$ Weyl sequences take their place. Throughout, using Lemma 4.7(c), we have that $m(\xi)(1+\xi^2) \geq 1$ with equality if and only if $\xi = 0$. This will give us the sharp constant of Theorem 4.10, the maximiser of Theorem A, and degree two in Corollary 4.9.

### *4.3.1 Definition and Its Properties*

**Definition 4.3.** The forward operator of the inverse heat source problem is:

$$A : L^2(\mathbb{R}) \to L^2(\mathbb{R}), \qquad A\varphi := v(\,\cdot\,, 1),$$

which is a bounded linear map, where $v$ is the unique mild solution of (4.1) (Proposition 1.4). In Fourier space, $A$ acts as multiplication by a function of frequency:

$$\widehat{A\varphi}(\xi) = m(\xi)\,\hat{\varphi}(\xi), \qquad m(\xi) := \frac{1 - e^{-\xi^2}}{\xi^2}\ (\xi \neq 0), \qquad m(0) := 1, \tag{4.33}$$

In physical space, we can write it in the form of a Duhamel integral:

$$A\varphi = \int_0^1 e^{(1-s)\Delta}\,\varphi\,\mathrm{d}s, \tag{4.34}$$

with $\{e^{t\Delta}\}_{t\geq 0}$ the heat semigroup of Definition 1.16.

They describe the same operator. The multiplier form (4.33) is the Fourier symbol from Proposition 1.19; the Duhamel form (4.34) is the variation of constants formula of Proposition 1.18 at $t = 1$.

**Theorem 4.4.** *The forward operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ has the following properties:*

(i) Boundedness. $\|A\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} = 1$ *(cannot be attained by any $L^2(\mathbb{R})$ function, can only be approached in the limit where Fourier mass tends towards $\xi = 0$)*

(ii) *Self-adjointness.* $A = A^*$.

(iii) *Positivity.* $A \succeq 0$ *in the operator order;* $A = A^{1/2}\,A^{1/2}$ *with* $\widehat{A^{1/2}\varphi}(\xi) = m(\xi)^{1/2}\,\hat{\varphi}(\xi)$.

(iv) *Injectivity.* $A$ *is injective,* $\ker(A) = \{0\}$.

(v) *Range.* $\mathrm{Ran}(A)$ *is dense (but not closed) in* $L^2(\mathbb{R})$*, and hence inverse* $A^{-1} : \mathrm{Ran}(A) \to L^2(\mathbb{R})$ *is well defined but unbounded.*

(vi) Non compactness. $A$ *is not a compact operator, which implies it has no singular value decomposition with discrete singular values that accumulate only at zero.*

**Proof.** By [80, §2.2.4], we know that The Fourier transform $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is unitary [52, Exercise 15.41], so writing $A = \mathcal{F}^{-1}M_m\mathcal{F}$ shows $A$ as unitarily equivalent to $M_m$ (the operator that multiplies by $m(\xi)$) we know Fourier transforming preserves size, $A$ and $M_m$ share what's about sizes, norms, spectra, self-adjointness. hence they are really the same operator, and instead of proving six things about $A$, it is enough to prove six things about $M_m$.

**(i)** By [58, Ch. II, §1] we know that:

$$\|M_g\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} = \operatorname*{ess\,sup}_{\xi\in\mathbb{R}} |g(\xi)|$$

The general statement is used for the operator $A = M_m$, where $M_m$ is the operator of multiplication by the symbol $m(\xi)$. Due to Lemma 4.7 (a) below, $0 < m \leq 1$ and $\operatorname{ess\,sup} m = 1$, which is attained only at the point $\xi = 0$ of zero Lebesgue measure. Hence

$$\|A\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} = \operatorname*{ess\,sup}_{\xi\in\mathbb{R}} m(\xi) = 1,$$

and this is approached but never attained.

**(ii)-(iii)** by (Lemma 4.7) $m$ is real and non negative, also multiplication by a real non negative function on $L^2(\mathbb{R})$ is self adjoint and positive. [47, §32, §33] The square root $M_{m^{1/2}}$ is well defined and satisfies $M_{m^{1/2}}M_{m^{1/2}} = M_m$.[47, §120,§134].

**(iv)** If $A\varphi = 0$ then $m\hat{\varphi} = 0$ in $L^2(\mathbb{R})$; since $m(\xi) > 0$ pointwise, $\hat{\varphi} = 0$ a.e., hence $\varphi = 0$.

**(v)** The closure of the range of a self-adjoint operator is the same as the orthogonal complement of its kernel [58, Ch. II, §2, Thm 2.19]. By (ii) the operator $A$ is self-adjoint, and by (iv) its kernel is $\ker(A) = \{0\}$; there is nothing left to be perpendicular to except the whole space. So plugging that in, $\overline{\mathrm{Ran}(A)} = L^2(\mathbb{R})$. That is what dense means: the closure of the range is the entire space.

**(vi)** If the operator is compact and self adjoint in a separable Hilbert space, then the spectrum of the operator is discrete with the set of eigenvalues accumulating only at zero [58, Ch. II, §5]. Below we show that the set $\sigma(A) = [0, 1]$ is a continuum (Theorem 4.6); hence $A$ is not compact. □

### *4.3.2 The Smoothing Property $A : H^s(\mathbb{R}) \to H^{s+2}(\mathbb{R})$*

The Fourier multiplier (symbol) decays $\xi^{-2}$, thats why$A$ gains exactly two derivatives on the Sobolev scale. It can not gain more than two derivatives, because thats need a further factor $\xi^{-1}$ the symbol does not have. The following proposition states this fact in a two sided

estimate, with the constants being explicitly calculated and $s$-independent, and improves, along the way, the estimate $m(\xi)(1+\xi^2) \le 2$ of Lemma 4.7 (d) to its exact value $c_\sharp$.

**Proposition 4.5.** *For $s \in \mathbb{R}$. $A$ is a bounded isomorphism $A : H^s(\mathbb{R}) \to H^{s+2}(\mathbb{R})$ with:*

$$\|\varphi\|_{H^s(\mathbb{R})} \;\le\; \|A\varphi\|_{H^{s+2}(\mathbb{R})} \;\le\; c_\sharp\, \|\varphi\|_{H^s(\mathbb{R})} \qquad \text{for all } \varphi \in H^s(\mathbb{R}), \tag{4.35}$$

*and both constants are sharp and none of them depend on $s$:*

$$\|A^{-1}\|_{H^{s+2}(\mathbb{R})\to H^s(\mathbb{R})} = 1, \qquad \|A\|_{H^s(\mathbb{R})\to H^{s+2}(\mathbb{R})} = c_\sharp := \sup_{\xi\in\mathbb{R}} m(\xi)\,(1+\xi^2) = \frac{(1+t_\sharp)^2}{1+t_\sharp+t_\sharp^2} \approx 1.2984, \tag{4.36}$$

*where $t_\sharp \approx 1.7933$ is the unique positive root of the equation:*

$$e^{t_\sharp} = 1 + t_\sharp + t_\sharp^2. \tag{4.37}$$

**Proof.** By Definition 4.3 and Plancherel,

$$\|A\varphi\|^2_{H^{s+2}(\mathbb{R})} = \int_{\mathbb{R}} \big[(1+\xi^2)\, m(\xi)\big]^2 (1+\xi^2)^s\, |\hat\varphi(\xi)|^2 \, \mathrm{d}\xi,$$

Therefore the two operator norms coincide with the maximum value of $(1+\xi^2)\, m(\xi)$ and the maximum value of $[(1+\xi^2)\, m(\xi)]^{-1}$ respectively. The lower bound in (4.35) may be rewritten as in Lemma 4.7 (c) as $m(\xi)(1+\xi^2) \ge 1$, where equality holds for $\xi = 0$ only. Taking the inverse of $A$ leads to

$$\|A^{-1}\| = \sup_{\xi} \big[m(\xi)(1+\xi^2)\big]^{-1} = 1.$$

To prove the surjectivity of $A$, take any $w$ from $H^{s+2}(\mathbb{R})$. Using the reverse multiplication approach, $\hat\varphi := \hat w/m$. It follows that $A\varphi = w$. Also, from (c), we have $\|\varphi\|_{H^s(\mathbb{R})} \le \|w\|_{H^{s+2}(\mathbb{R})}$, which implies that $\varphi$ belongs to $H^s(\mathbb{R})$. Since $w$ was arbitrary, $A$ is surjective.

For the upper bound constant, we rename $t = \xi^2$ and continue, because in this case $(1+\xi^2)\, m(\xi)$ can be rewritten as $g(t) := (1+t)(1-e^{-t})/t$, where $g(0) := 1$. Therefore, finding $c_\sharp$ as the maximal value of $(1+\xi^2)\, m(\xi)$ equals finding the maximal value of $g(t)$, which equals $c_\sharp = \sup_{t\ge 0} g(t)$; and $g(t) \to 1$ when $t \downarrow 0$ and when $t \to \infty$. In this way, the maximum cannot be attained on either side of the interval, so the maximum is attained inside the interval. Taking the derivative we have

$$g'(t) = \frac{e^{-t}(1+t+t^2) - 1}{t^2},$$

and the critical points are the solutions of the equation $N(t) := e^{-t}(1+t+t^2) - 1 = 0$, that is $e^t = 1 + t + t^2$. As $N(0) = 0$ and $N'(t) = e^{-t}\, t\,(1-t)$, then $N$ is increasing on $(0,1)$ and decreasing on $(1,\infty)$ with the limit $-1$, so $N$ has a unique positive root $t_\sharp > 1$ such that $N(t) > 0$ for $t < t_\sharp$ and $N(t) < 0$ for $t > t_\sharp$; hence $g$ attains its maximum exactly at $t_\sharp$. From the numerical point of view, $t_\sharp = 1.79328\ldots$, and putting $e^{-t_\sharp} = (1+t_\sharp+t_\sharp^2)^{-1}$ into $g$, we find

$$c_\sharp = \frac{(1+t_\sharp)^2}{1+t_\sharp+t_\sharp^2} = 1.29842\ldots$$

□

### 4.3.3 Spectral Structure

**Theorem 4.6.** *The forward operator satisfies the following:*

(i) $\sigma(A) = [0, 1]$, *the essential range of* $m$.

(ii) $\sigma_{\mathrm{pp}}(A) = \varnothing$; *$A$ has no eigenvectors.*

(iii) $\sigma_{\mathrm{sc}}(A) = \varnothing$.

(iv) $\sigma_{\mathrm{ac}}(A) = [0, 1]$ *(here the spectral measure behaves just like Lebesgue measure on* $[0, 1]$*)*

(v) *For every bounded Borel function* $f : [0, 1] \to \mathbb{C}$, $f(A) = \mathcal{F}^{-1} M_{f\circ m} \mathcal{F}$.

**Proof.** **(i)** By [58, Ch. IX, §2], we know that The spectrum of $M_m$ on $L^2(\mathbb{R}, \mathrm{d}\xi)$ equals the essential range of $m$ By Lemma 4.7 below, $m$ is continuous and even, strictly decreasing on $[0, \infty)$, with $m(0) = 1$ and $m(\xi) \to 0$ as $|\xi| \to \infty$. So Its image is $(0, 1]$, whose closure, the essential range, is $[0, 1]$.

**(ii)** $\lambda$ is counted as an eigenvalue of $M_m$ only when the set $\{m = \lambda\}$ has positive Lebesgue measure. But $m$ is real analytic (Lemma 4.7) and non constant, so every level set is discrete, meaning every level set has zero measure. So $\sigma_{\mathrm{pp}}(A) = \varnothing$, no eigenvalues exist.

**(iii)-(iv)** For $g \in L^2(\mathbb{R})$ the scalar spectral measure of $M_m$ is:

$$\mu_g^m(\Delta) = \langle E_m(\Delta) g, g \rangle = \int_{m^{-1}(\Delta)} |g(\xi)|^2 \, \mathrm{d}\xi = m_*\big(|g|^2 \, \mathrm{d}\xi\big)(\Delta),$$

since the spectral projections of a multiplication operator are $E_m(\Delta) = M_{\mathbf{1}_{m^{-1}(\Delta)}}$ [58, Ch. IX, §2]. Because $m$ is even and strictly monotone on each of $(-\infty, 0]$ and $[0, \infty)$ (Lemma 4.7), the pushforward under $m$ of any absolutely continuous measure stays absolutely continuous on $[0, 1]$, by the one dimensional change of variables formula [130, Thm. 7.26]. Hence $\sigma_{\mathrm{sc}}(A) = \varnothing$ and $\sigma_{\mathrm{ac}}(A) = [0, 1]$.

**(v)** By [58, Ch. IX, §2] we know that the bounded Borel functional calculus for multiplication operators gives $f(M_m) = M_{f\circ m}$, and unitary conjugation means that the same identity applies to $A$. □

### 4.3.4 From Extremizers to Weyl Sequences

On a bounded interval, the truncated operator $A_L$ is diagonal in the sine basis with eigenvalues $\widetilde{m}_n = m(n\pi/L) \downarrow 0$, so compact, and its largest stretch factor is attained at the leading eigenfunction (Chapter 2, Remark 2.6, compact self adjoint theory [58, Ch. II, §5,Lemma 5.9], with Courant Fischer giving the variational characterisation [33, Cor. III.1.2]). On $\mathbb{R}$, Theorem 4.6 (ii) does not leave an eigenvector to point at, so two substitutions are made.

**(a)** Weyl sequences are the analogue of eigenvectors, for $\lambda \in \sigma(A)$, and for no other $\lambda$, there is a normalised sequence $\varphi_n$ along which $\|(A - \lambda I)\varphi_n\|_{L^2(\mathbb{R})} \to 0$ [127, Vol. I, Ch. VII.3, Thm. VII.12]; it never converges in $L^2(\mathbb{R})$.

**(b)** Concentration instead of spectral projection. Since the spectrum is absolutely continuous, there is no mass at any single point, so if we project onto a single exact value we get nothing. in fact , only the projection to a whole interval, $E_A([\lambda - \varepsilon, \lambda])$, tells us something, and

this is realized concretely in Section 4.4 by squeezing a the Fourier content of a function tightly around one frequency $\xi_\star$ such that $m(\xi_\star) = \lambda$.

### 4.3.5 The Multiplier Inequality

The analytic fact that does most of the work in Section 4.4, is the pointwise inequality $m(\xi)(1+\xi^2) \geq 1$. It is what fixes the sharp prefactor $(2E)^{2/(\ell+2)}$ of Theorem 4.11 and determines the modulus of continuity in Theorem A.Here we state it on its own, and along with the other bounds on $m$ that it belongs with.

**Lemma 4.7.** *The function $m$, defined in (4.33), has four properties, it is positive, even, real analytic on $\mathbb{R}$, and strictly decreasing on $(0,\infty)$. For every $\xi \in \mathbb{R}$,*

(a) $m(\xi) \leq \min\{1, \xi^{-2}\}$;

(b) $m(\xi) \geq (1 - e^{-1})/(1 + \xi^2)$;

(c) $m(\xi)(1 + \xi^2) \geq 1$, *with equality if and only if* $\xi = 0$;

(d) $m(\xi)(1 + \xi^2) \leq 2$.

*Moreover $m(\xi) \sim \xi^{-2}$ as $|\xi| \to \infty$.*

**Proof.** We start with the Taylor series of $1 - e^{-t}$, then evaluate it at $t = \xi^2$, and obtain $1 - e^{-\xi^2} = \sum_{k\geq 1}(-1)^{k-1}\xi^{2k}/k!$, hence:

$$m(\xi) = \sum_{k\geq 0} \frac{(-1)^k}{(k+1)!}\,\xi^{2k} = 1 - \tfrac{1}{2}\xi^2 + \tfrac{1}{6}\xi^4 - \cdots,$$

which is convergent for all $\xi \in \mathbb{R}$, and positivity, evenness, and real analyticity follow.

We now define the auxiliary function as follows:

$$G(t) := (1 + t)\, e^{-t} - 1, \qquad t \geq 0. \tag{4.38}$$

Then $G(0) = 0$ and $G'(t) = -t\, e^{-t} \leq 0$, strictly for $t > 0$, so $G$ is strictly decreasing on $[0,\infty)$ and $G(t) < 0$ for $t > 0$.

We can rearrange this to obtain:

$$m(\xi)(1 + \xi^2) - 1 = \left.\frac{(1 - e^{-t})(1 + t) - t}{t}\right|_{t=\xi^2} = -\frac{G(\xi^2)}{\xi^2} \geq 0, \tag{4.39}$$

equality holding only at $\xi = 0$, which is (c).

The strict monotonicity of $m$ on $(0,\infty)$ follows from $m'(\xi) = 2\xi^{-3}G(\xi^2) < 0$ for $\xi > 0$.

*Parts* (a), (b), (d). With $t = \xi^2$: (a) is a consequence of $1 - e^{-t} \leq \min\{1, t\}$; (b) is obtained by splitting at $|\xi| = 1$, using $m \geq m(1) = 1 - e^{-1}$ on the inner range and $1 - e^{-t} \geq 1 - e^{-1}$, $\xi^2 \leq 1 + \xi^2$ on the outer; (d) follows from $(1 - e^{-t})(1 + t)/t \leq \min\{1 + t,\, 1 + t^{-1}\} \leq 2$, splitting at $t = 1$. Finally $m(\xi)\xi^2 = 1 - e^{-\xi^2} \to 1$ yields the asymptotics.

□

**Remark 4.8.** Let $\alpha = \ell/(\ell+2)$. Now let's see why this choice is important: $1-\alpha = 2/(\ell+2)$, so $\ell(1-\alpha) = 2\ell/(\ell+2)$, and $2\alpha$ comes out to that exact same value too.

We refer to this as the spectral balance function:

$$h(\xi) := m(\xi)^{2\alpha}\,(1+\xi^2)^{\ell(1-\alpha)} = [m(\xi)(1+\xi^2)]^{2\ell/(\ell+2)} \tag{4.40}$$

Since $2\alpha = \ell(1-\alpha) = 2\ell/(\ell+2)$, the two exponents are equal, and we can directly apply $x^p y^p = (xy)^p$, with $p = 2\ell/(\ell+2)$.

And $h$ behaves as we already know $m(\xi)(1+\xi^2)$ does. By Lemma 4.7 (c), $h \geq 1$ everywhere, and equals 1 only at $\xi = 0$. And since $m \sim \xi^{-2}$, $h$ also comes back down to 1 out at infinity too. Together, they show that $\inf_\xi h(\xi) = 1$, the one fact on which the sharpness theory of Section 4.4 depends. The same exponent $\alpha = \ell/(\ell+p)$ which agrees with the degree-$p$ ill posedness is the same tool underlying the Nirenberg interpolation inequalities. [117, Lecture II, pp. 124-126].

**Corollary 4.9.** *The inverse problem (4.1)-(4.2) is ill posed in the sense of Hadamard [68, Ch. 2],[72],[95] . Since we have already seen that the inverse operator $A^{-1} : \mathrm{Ran}(A) \to L^2(\mathbb{R})$ is unbounded, we can also claim that $\mathrm{Ran}(A)$ is not closed [68, Ch. 2, §2.1][87]. The Hilbert scale degree of ill posedness of this problem is 2 because of $m(\xi) \asymp \xi^{-2}$ when $|\xi| \to \infty$. This corresponds to the two sided estimate on Hilbert scale which is used for the interpretation of the order of smoothing as the degree of ill posedness [68, §8.4 - Cor. 8.22,§8.5][94, §9][84], realised here with the exact constants of Proposition 4.5.*

## 4.4 Sharp Hölder Stability and Minimax Optimality

All results presented in this section concern one inequality, i.e., the source error bound $\leq C\delta^{\ell/(\ell+2)}$. Here $C$ is the unknown coefficient while the power is already known, and we do not know whether either can be improved.

The constant emerges at each single frequency. For each individual frequency, the smoother the source is, the less of it can exist at that frequency, and $m$ tells us how much of it the measurement data see; together this gives Theorem 4.10, with the constant being exactly 1. Summing over all frequencies leads to the Hölder estimate on $\mathcal{W}_{\ell,E}$ (Theorem 4.11), rephrased for the computed reconstruction in Corollary 4.12 and, under $\varepsilon \leq \delta$, in Corollary 4.13.

Section 4.4.3 determines whether the constant may be reduced further, which means finding the biggest possible error that any algorithm might produce. This is generally known only in the limit as $\delta \to 0$, which fixes the exponent but not the constant; here it is known at every $\delta$.

The convexity of $m^2$ against the Sobolev weight (Proposition 4.17), due to its complete monotonicity, guarantees that a unique worst frequency exists. Lemma 4.18 then transforms the conventional two-frequency estimate into an equality, which takes the closed form of Theorem A, approached by Definition 4.15 but attained by no $L^2(\mathbb{R})$ function.

These results follow: a linear method already attains it (Theorem B), so nonlinearity provides no benefit; Theorem 4.22 prices additional observation instants; and Theorem C tells us

that the trained source map is Tikhonov in a Hilbert scale in (a), with its rate in (b) under (B1) and (B2). This makes it possible to apply everything to the executed algorithm itself.

### *4.4.1 The Sharp Interpolation Inequality*

**Theorem 4.10.** *For every $\ell \geq 1$ and every $\varphi \in H^\ell(\mathbb{R}) \setminus \{0\}$,*

$$\|\varphi\|_{L^2(\mathbb{R})} \leq \|A\varphi\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)} \, \|\varphi\|_{H^\ell(\mathbb{R})}^{2/(\ell+2)}, \tag{4.41}$$

*and the constant 1 on the RHS cannot be replaced by any smaller value.*

**Proof.** Let $\alpha := \ell/(\ell+2) \in (0,1)$. The exponents satisfy the identity:

$$2\alpha = \ell(1-\alpha) = \frac{2\ell}{\ell+2}, \tag{4.42}$$

which is what was discussed in Remark 4.8. The important step now is this: for a.e. $\xi$, the following pointwise factorisation holds:

$$|\hat\varphi(\xi)|^2 = \big[m(\xi)^2|\hat\varphi(\xi)|^2\big]^\alpha \big[(1+\xi^2)^\ell|\hat\varphi(\xi)|^2\big]^{1-\alpha} \cdot h(\xi)^{-1} \;\leq\; \big[m(\xi)^2|\hat\varphi(\xi)|^2\big]^\alpha \big[(1+\xi^2)^\ell|\hat\varphi(\xi)|^2\big]^{1-\alpha}, \tag{4.43}$$

where $h(\xi) = [m(\xi)(1+\xi^2)]^{2\ell/(\ell+2)} \geq 1$ is the spectral balance function of Remark 4.8, which is bounded below by one via Lemma 4.7 (c). Since $h(\xi) \geq 1$, its reciprocal can be at most 1. That is the reason dropping the factor $h(\xi)^{-1}$ turns the identity above into the inequality. Now lets integrate over $\mathbb{R}$ and use Hölder's inequality, with conjugate exponents $1/\alpha$ and $1/(1-\alpha)$, which holds since $\alpha + (1-\alpha) = 1$,

$$\begin{aligned}
\|\varphi\|_{L^2(\mathbb{R})}^2 &= \int_{\mathbb{R}} |\hat\varphi(\xi)|^2 \, \mathrm{d}\xi \\
&\leq \Big( \int_{\mathbb{R}} m(\xi)^2 |\hat\varphi(\xi)|^2 \, \mathrm{d}\xi \Big)^\alpha \Big( \int_{\mathbb{R}} (1+\xi^2)^\ell |\hat\varphi(\xi)|^2 \, \mathrm{d}\xi \Big)^{1-\alpha} \\
&= \|A\varphi\|_{L^2(\mathbb{R})}^{2\alpha} \, \|\varphi\|_{H^\ell(\mathbb{R})}^{2(1-\alpha)}.
\end{aligned}$$

The first integral is $\|A\varphi\|_{L^2(\mathbb{R})}^2$, by Plancherel and the Fourier symbol of $A$ in (4.33). The second is $\|\varphi\|_{H^\ell(\mathbb{R})}^2$, by the Fourier domain Sobolev norm (1.6) of Definition 1.7.

Taking the square roots and substituting $\alpha = \ell/(\ell+2)$,

$$\begin{aligned}
\|\varphi\|_{L^2(\mathbb{R})}^2 \leq \|A\varphi\|_{L^2(\mathbb{R})}^{2\alpha} \, \|\varphi\|_{H^\ell(\mathbb{R})}^{2(1-\alpha)} \implies \|\varphi\|_{L^2(\mathbb{R})} \leq \|A\varphi\|_{L^2(\mathbb{R})}^{\alpha} \, \|\varphi\|_{H^\ell(\mathbb{R})}^{1-\alpha} \\
\overset{\alpha=\ell/(\ell+2)}{\implies} \|\varphi\|_{L^2(\mathbb{R})} \leq \|A\varphi\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)} \, \|\varphi\|_{H^\ell(\mathbb{R})}^{2/(\ell+2)},
\end{aligned}$$

which is (4.41). One thing this proof does not yet show that the constant 1 here is indeed the best possible one, not just a valid one.

□

### *4.4.2 Sharp Hölder Stability*

In this section, we prove one of this chapter's most important results, Theorem 4.11. We take the difference of two admissible sources, feed it into Theorem 4.10, and bound its Sobolev norm using the ball diameter $2E$. What matters here is not the inequality itself, its exponent is already standard, but the best possible prefactor in front of it, a claim Theorem A proves it later in the analysis.

**Theorem 4.11.** *For every $\ell \geq 1$ and $E > 0$, and for all $\varphi_1, \varphi_2 \in \mathcal{W}_{\ell,E}$,*

$$\|\varphi_1 - \varphi_2\|_{L^2(\mathbb{R})} \leq (2E)^{2/(\ell+2)} \, \|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)}. \tag{4.44}$$

*Both the Hölder exponent $\ell/(\ell+2)$ and the prefactor $(2E)^{2/(\ell+2)}$ are sharp on $\mathcal{W}_{\ell,E}$.*

**Proof.** Let $\psi := \varphi_1 - \varphi_2$. The Sobolev triangle inequality and the source class definition (1.7) yields:

$$\|\psi\|_{H^\ell(\mathbb{R})} \leq \|\varphi_1\|_{H^\ell(\mathbb{R})} + \|\varphi_2\|_{H^\ell(\mathbb{R})} \leq 2E.$$

Linearity of $A$ and Theorem 4.10 applied to $\psi$ then gives:

$$\begin{aligned} \|\psi\|_{L^2(\mathbb{R})} &\leq \|A\psi\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)} \, \|\psi\|_{H^\ell(\mathbb{R})}^{2/(\ell+2)} \\ &\leq (2E)^{2/(\ell+2)} \, \|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)}, \end{aligned}$$

which is (4.44).

by [104, §3,Thm. 1, Cor. 1] The sharpness of the exponent $\ell/(\ell+2)$ is in agreement with the variable Hilbert scale minimax rate and with the heat source Fourier regularisation analysis of [65].The sharpness of the prefactor $(2E)^{2/(\ell+2)}$, is a different matter, which we discuss in details in Theorem A, the exact modulus of continuity is indeed this constant, and it is realised asymptotically by the high frequency Weyl sequence $w_n^{\mathrm{hi}}$ of Definition 4.15, built at the optimal scale $K_\star$. □

**Corollary 4.12.** *Suppose $\widehat{\varphi} \in H^\ell(\mathbb{R})$ satisfies $\|\widehat{\varphi}\|_{H^\ell(\mathbb{R})} \leq E$ and $\|A\widehat{\varphi} - g^\delta\|_{L^2(\mathbb{R})} \leq \varepsilon$, where $g^\delta$ is the noisy datum of (4.4). Then*

$$\|\widehat{\varphi} - \varphi^\dagger\|_{L^2(\mathbb{R})} \leq (2E)^{2/(\ell+2)} \, (\delta + \varepsilon)^{\ell/(\ell+2)}. \tag{4.45}$$

**Proof.** By the triangle inequality together with the noise model (4.4) we obtain:

$$\|A\widehat{\varphi} - A\varphi^\dagger\|_{L^2(\mathbb{R})} \leq \|A\widehat{\varphi} - g^\delta\|_{L^2(\mathbb{R})} + \|g^\delta - A\varphi^\dagger\|_{L^2(\mathbb{R})} \leq \varepsilon + \delta.$$

$\widehat{\varphi}$ and $\varphi^\dagger$ belong to $\mathcal{W}_{\ell,E}$, $\varphi^\dagger$ by the assumption on the true source and by (1.7), $\widehat{\varphi}$ by the assumption of this corollary itself, Theorem 4.11 then yields (4.45). □

**Corollary 4.13.** *Under the assumptions of Corollary 4.12, for the case where the training*

*residual matches the noise level, $\varepsilon \le \delta$, it holds that*

$$\|\widehat{\varphi} - \varphi^\dagger\|_{L^2(\mathbb{R})} \le (2E)^{2/(\ell+2)}\,(2\delta)^{\ell/(\ell+2)} = 2^{\ell/(\ell+2)}\,(2E)^{2/(\ell+2)}\,\delta^{\ell/(\ell+2)}. \tag{4.46}$$

*The exponent and the dependence on $E$ are unchanged; only the $\delta$-multiplier has the factor $2^{\ell/(\ell+2)} < 2$.*

**Proof.** Substitute $\varepsilon \le \delta$ into (4.45): then $\delta+\varepsilon \le 2\delta$, and the Hölder exponent $\ell/(\ell+2) \in (0,1)$ is monotone. □

The balance $\varepsilon \asymp \delta$ is the training-side analogue of the a-priori balance $\lambda^\star(\delta) \asymp (\delta/E)^2$ of Proposition 2.13.

### 4.4.3 Modulus of Continuity and Minimax Optimality

Theorem 4.11 has already given us an upper bound for the modulus of continuity of $A^{-1}$ on $\mathcal{W}_{\ell,E}$, by $(2E)^{2/(\ell+2)}\delta^{\ell/(\ell+2)}$. However, what is missing is the lower bound, which is what we will address in this section. On $\mathbb{R}$ we require two things, a high frequency Weyl sequence tuned to the optimal spectral scale, and a proof that no source in $\mathcal{W}_{\ell,E}$ can do better than it. When both are available, everything on $\mathcal{W}_{\ell,E}$ is completely determined, and not just in terms of proving the constant to be sharp. The modulus is known at every single level of noise, not just asymptotically. And due to this, the minimax error is derived from this quite easily, requiring just a two point argument and the optimality of linear smoothing. These same tools can be used for another purpose, to determine how much extra observation times are worth.

**Definition 4.14.** For $\delta > 0$, the modulus of continuity of $A^{-1}$ on the source class $\mathcal{W}_{\ell,E}$ is:

$$\omega(\delta, \mathcal{W}_{\ell,E}) := \sup\Big\{\|\varphi_1 - \varphi_2\|_{L^2(\mathbb{R})} : \varphi_1, \varphi_2 \in \mathcal{W}_{\ell,E},\ \|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})} \le \delta\Big\}. \tag{4.47}$$

**Definition 4.15.** For $\varepsilon_n \downarrow 0$ and a sequence $K_n \to \infty$ to be calibrated, define $w_n^{\mathrm{hi}} \in L^2(\mathbb{R})$ by:

$$\widehat{w_n^{\mathrm{hi}}}(\xi) := (4\varepsilon_n)^{-1/2}\left[\chi_{[K_n-\varepsilon_n,K_n+\varepsilon_n]}(\xi) + \chi_{[-K_n-\varepsilon_n,-K_n+\varepsilon_n]}(\xi)\right], \tag{4.48}$$

i.e. $\widehat{w_n^{\mathrm{hi}}}$ is the symmetrised, $L^2$ normalised indicator of two narrow $\varepsilon_n$-windows at $\pm K_n$.

The following lemma describes what the sharpness proof relies on, how the $L^2(\mathbb{R})$, $H^\ell(\mathbb{R})$, and $\|A\cdot\|$ norms of $w_n^{\mathrm{hi}}$ behave as the bands become narrower and further away from the origin.

**Lemma 4.16.** *for $\varepsilon_n \le K_n/2$, the sequence $w_n^{\mathrm{hi}}$ satisfies*

(1) $\|w_n^{\mathrm{hi}}\|_{L^2(\mathbb{R})} = 1$;

(2) $\|w_n^{\mathrm{hi}}\|^2_{H^\ell(\mathbb{R})} = (1+K_n^2)^\ell\big(1+\mathcal{O}((\varepsilon_n/K_n)^2)\big)$;

(3) $\|Aw_n^{\mathrm{hi}}\|^2_{L^2(\mathbb{R})} = m(K_n)^2\big(1+\mathcal{O}((\varepsilon_n/K_n)^2)\big) = K_n^{-4}\big(1+\mathcal{O}(K_n^{-2})\big)\big(1+\mathcal{O}((\varepsilon_n/K_n)^2)\big)$.

**Proof.** **(1)** $\|w_n^{\mathrm{hi}}\|^2_{L^2(\mathbb{R})} = \|\widehat{w_n^{\mathrm{hi}}}\|^2_{L^2(\mathbb{R})} = (4\varepsilon_n)^{-1} \cdot (2 \cdot 2\varepsilon_n) = 1$.

**(2)** For $\xi$ in the support of $\widehat{w_n^{\mathrm{hi}}}$, $|\xi^2 - K_n^2| \le 2K_n\varepsilon_n + \varepsilon_n^2 \le 3K_n\varepsilon_n$ when $\varepsilon_n \le K_n/2$. Expanding, $(1+\xi^2)^\ell = (1+K_n^2)^\ell\big(1+\mathcal{O}(\varepsilon_n/K_n)\big)^\ell$, the first order correction averages to zero by symmetry about $\pm K_n$, leaving the second-order term.

**(3)** We apply the same calculation, with $m$ taking the place of $(1+\xi^2)^{\ell/2}$. This gives $m(K_n)^2$, up to the same $\big(1+\mathcal{O}((\varepsilon_n/K_n)^2)\big)$ correction as before. Then apply the asymptotic $m(K_n) \sim K_n^{-2}$ from Lemma 4.7 as $K_n \to \infty$, which turns $m(K_n)^2$ into $K_n^{-4}\big(1+\mathcal{O}(K_n^{-2})\big)$. □

The lower bound needs more work. On the Fourier side, the modulus becomes a linear optimisation problem over nonnegative measures, subject to two linear constraints, and the classical extreme point theory for such optimisation problems [10] allows only two atoms. Reducing this to the point of a single atom marks the point where the fine structure of the symbol kicks in, $m^2$ being completely monotone in $t = \xi^2$ makes it strictly convex in terms of the Sobolev weight $(1+t)^\ell$, this convexity condition allows for the formulation of a tangent line argument which ensures that no admissible source can outperform a single atom placed at the balance frequency. Thus, we obtain this convexity statement in a format that will be reused later on without any changes for the multi snapshot operators in Section 4.4.4.

**Proposition 4.17.** *Assume that $\mu$ is a positive Borel measure which is also finite and, with compact support in $[0,\infty)$ and such that $\mu((0,\infty)) > 0$, and set:*

$$Q(t) := \int_{[0,\infty)} e^{-rt}\,\mu(\mathrm{d}r), \qquad t \ge 0. \tag{4.49}$$

(i) *$Q$ is strictly completely monotone in the following sense: $(-1)^k Q^{(k)}(t) = \int r^k e^{-rt}\,\mu(\mathrm{d}r) > 0$ for all $k \ge 1$ and $t \ge 0$, and $Q > 0$; in particular $Q' < 0 < Q''$ everywhere.*

(ii) *For every $\ell \ge 1$, $Q$ is strictly convex and strictly decreasing as a function of the weight $P(t) := (1+t)^\ell$: writing $q$ for the function on $[1,\infty)$ with $Q(t) = q(P(t))$,*

$$q''(P(t)) = \frac{(1+t)\,Q''(t) - (\ell-1)\,Q'(t)}{\ell^2\,(1+t)^{2\ell-1}} > 0, \qquad t \ge 0. \tag{4.50}$$

*therefore, for all $t^\star \ge 0$ there exist $a^\star, b^\star > 0$ such that $a^\star = -q'(P^\star)/\varkappa$ and $b^\star = 1/\varkappa$ with $P^\star := P(t^\star)$ and $\varkappa := Q(t^\star) - q'(P^\star)P^\star > 0$ Such that the dual bound*

$$a^\star P(t) + b^\star Q(t) \;\ge\; 1 \qquad \text{for all } t \ge 0 \tag{4.51}$$

*holds, with equality $\iff t = t^\star$..*

(iii) *The squared symbol of the forward operator has the form (4.49): in the variable $t = \xi^2$,*

$$m = \int_0^1 e^{-st}\,\mathrm{d}s, \qquad m^2 = \int_0^2 e^{-rt}\,w_\triangle(r)\,\mathrm{d}r, \qquad w_\triangle(r) := \min\{r,\,2-r\}, \tag{4.52}$$

*the triangular density $w_\triangle$ on $[0,2]$ being the self convolution of the uniform density*

*on* $[0,1]$.

**Proof.** **(i)** As $\mu$ is a finite measure with compact support, differentiation under the integral sign is justified, giving us $Q^{(k)}(t) = \int (-r)^k e^{-rt}\,\mu(\mathrm{d}r)$, and therefore $(-1)^k Q^{(k)}(t) = \int r^k e^{-rt}\,\mu(\mathrm{d}r)$,

The above expression is positive for all $k \geq 1$ as $r^k e^{-rt} > 0$ for all $r \in (0,\infty)$ and $\mu((0,\infty)) > 0$; in addition, $Q > 0$ since $\mu \neq 0$. Hence, this proves complete monotonicity via Bernstein Widder theorem. For more information, refer to [131, Ch. 1,Thm. 1.4].

**(ii)** With $P'(t) = \ell(1+t)^{\ell-1} > 0$ and $P''(t) = \ell(\ell-1)(1+t)^{\ell-2}$, the chain rule gives $q'(P) = Q'/P' < 0$ and

$$q''(P) = \frac{Q''P' - Q'P''}{(P')^3} = \frac{(1+t)\,Q'' - (\ell-1)\,Q'}{\ell^2(1+t)^{2\ell-1}} > 0,$$

since $Q'' > 0$, $-Q' > 0$, and $\ell \geq 1$, the numerator is positive for all $P \geq 1$, Therefore, since we have strict convexity, the graph of $q$ lies on or above its own tangents, except at the point of tangency.

$$q(P) \;\geq\; q(P^\star) + q'(P^\star)\,(P - P^\star),$$

holds if and only if $P = P^\star$. rearranging this can be re-written as: $-q'(P^\star)\,P + q(P) \geq \varkappa := Q(t^\star) - q'(P^\star)P^\star$, where $\varkappa > 0$ since $Q(t^\star) > 0$ and $-q'(P^\star)P^\star \geq 0$. After Dividing by $\varkappa$ we obtain (4.51), wherein the equality case changes from $P = P^\star$ to $t = t^\star$ since $P$ is strictly increasing.

**(iii)** We compute the integral directly: $\int_0^1 e^{-st}\mathrm{d}s = (1-e^{-t})/t = m$ for $t > 0$, and at $t = 0$, it is 1, as expected from (4.33),in both aspects. squaring the integral and interchanging the order of integration by Tonelli's theorem [79, Thm. 2.37]., we obtain:

$$m^2 = \int_0^1\!\!\int_0^1 e^{-(s+u)t}\,\mathrm{d}s\,\mathrm{d}u = \int_0^2 e^{-rt}\,w_\triangle(r)\,\mathrm{d}r,$$

where $w_\triangle$ is the density of the sum $s+u$ when $s,u$ are independent uniform variables on $[0,1]$, having the triangular density $\min\{r, 2-r\}$, We use the well-known fact that products of completely monotone functions are completely monotone [131, Cor. 1.6] □

**Lemma 4.18.** *For any $\delta > 0$, we have the following:*

$$\omega(\delta, \mathcal{W}_{\ell,E})^2 = \sup_{\xi \geq 0} \min\Big\{ \frac{(2E)^2}{(1+\xi^2)^\ell},\, \frac{\delta^2}{m(\xi)^2} \Big\}. \tag{4.53}$$

*For $0 < \delta < 2E$ the supremum is attained at the unique solution $\xi^\star = \xi^\star(\delta)$ of the balance equation:*

$$\frac{m(\xi^\star)}{(1+\xi^{\star 2})^{\ell/2}} = \frac{\delta}{2E}, \tag{4.54}$$

*and this modulus is known, not just approximated, for all levels of noise.,*

$$\omega(\delta, \mathcal{W}_{\ell,E}) = \frac{2E}{(1+\xi^{\star 2})^{\ell/2}} = \frac{\delta}{m(\xi^\star)} \qquad \textit{exactly, for all } 0 < \delta < 2E; \tag{4.55}$$

*for $\delta \geq 2E$ the data constraint is inactive and $\omega(\delta, \mathcal{W}_{\ell,E}) = 2E$.*

**Proof.** Because $\mathcal{W}_{\ell,E}$ is convex and symmetric, we have that its difference set is $\mathcal{W}_{\ell,E} - \mathcal{W}_{\ell,E} = 2\mathcal{W}_{\ell,E} = \{\varphi : \|\varphi\|_{H^\ell(\mathbb{R})} \leq 2E\}$, and consequently writing out $\varphi = \varphi_1 - \varphi_2$ in (4.47) we see that this results in the single constraint form:

$$\omega(\delta, \mathcal{W}_{\ell,E}) = \sup\big\{\|\varphi\|_{L^2(\mathbb{R})} : \|\varphi\|_{H^\ell(\mathbb{R})} \leq 2E,\ \|A\varphi\|_{L^2(\mathbb{R})} \leq \delta\big\}. \tag{4.56}$$

Moving on to the Fourier side, let us introduce $\nu(\mathrm{d}\xi) := |\widehat{\varphi}(\xi)|^2 \mathrm{d}\xi$, which is a finite nonnegative measure on $[0, \infty)$, as two half-lines can be replaced with a single line due to evenness. Plancherel's formula implies that $\|\varphi\|_{L^2(\mathbb{R})}$ equals the total mass of $\nu$, In terms of the Sobolev norm, it corresponds to the integration of $(1 + \xi^2)^\ell$ against $\nu$, and separately, the multiplier representation (4.33) transforms $\|A\varphi\|_{L^2(\mathbb{R})}$ into the integral of $m(\xi)^2$ against $\nu$.

So (4.56), In other words, by representing each of the norms as such, we arrive at a problem of classical analysis, referred to as the moment problem [98],[129], maximize the mass of a nonnegative measure, subject to two constraints on its integrals against fixed weight functions,

$$V(\delta) := \sup_{\nu \geq 0} \int \mathrm{d}\nu \quad \text{subject to} \quad \int (1 + \xi^2)^\ell \, \mathrm{d}\nu \leq (2E)^2, \ \int m(\xi)^2 \, \mathrm{d}\nu \leq \delta^2, \tag{4.57}$$

where the supremum on the right is taken over all finite nonnegative Borel measures, of which the admissible spectral densities are a subclass; we show that the two suprema coincide and equal the right-hand side of (4.53).

*Lower bound:* Let $\xi$ be fixed such that $\xi \geq 0$ and $p < p(\xi) := \min\{(2E)^2(1+\xi^2)^{-\ell}, \delta^2 m(\xi)^{-2}\}$, it is the maximum mass of a single atom placed at $\xi$ which will satisfy both constraints. let $w_n^{\mathrm{hi}}$ from (4.48) with fixed centre $K = \xi$ and width $\varepsilon \downarrow 0$ (one window centered at the origin when $\xi = 0$), scale to mass $p$, the constraint integrals will converge to $p(1 + \xi^2)^\ell$ and $pm(\xi)^2$ by continuity, thus it is admissible for sufficiently small $\varepsilon$, it follows that $\omega(\delta, \mathcal{W}_{\ell,E})^2 \geq p$, and taking $p \uparrow p(\xi)$ and the supremum over $\xi$ yields $\omega(\delta, \mathcal{W}_{\ell,E})^2 \geq \sup_{\xi \geq 0} p(\xi)$.

We consider the functions $A(\xi)$ and $B(\xi)$ individually. Set $A(\xi) := (2E)^2(1 + \xi^2)^{-\ell}$ that is a strictly decreasing function from $(2E)^2$ to 0 while $B(\xi) := \delta^2 m(\xi)^{-2}$ increases strictly from $\delta^2$ (Lemma 4.7). For $0 < \delta < 2E$ the map $\min\{A, B\}$ increases while $B \leq A$ and then decreases, therefore, its maximum sits at the unique intersection, which is (4.54): the ratio $m(\xi)(1 + \xi^2)^{-\ell/2}$ is continuous, strictly decreasing, and maps $[0, \infty)$ onto $(0, 1]$, so it intersects the value $\delta/(2E) \in (0, 1)$ at exactly one point. The common value at the intersection is $p^\star := p(\xi^\star) = (2E)^2(1 + \xi^{\star 2})^{-\ell} = \delta^2 m(\xi^\star)^{-2}$.

*Upper bound:* By [9, Thm. 3] for a two constraint moment problem of the current kind the true optimum does not need to have more than two atoms. This bound alone does not tell us whether a single atom already suffices, that's why we need the specific convexity property of $m(\xi)^2$ with respect to the Sobolev weight $(1 + \xi^2)^\ell$, established in Proposition 4.17.

Let $t = \xi^2$, $P(t) = (1 + t)^\ell$, and $Q = m^2$, since $Q$ is shown to be the Laplace transform of the triangular density $w_\triangle$, by (4.52), which means that Proposition 4.17 can be applied.

applying part(ii) to $t^\star = \xi^{\star 2}$ yields $a^\star, b^\star > 0$ such that $a^\star P + b^\star Q \geq 1$ on $[0, \infty)$ and equality

holds only at $t^\star$. For every admissible $\nu$ in (4.57),

$$\nu\big([0,\infty)\big) = \int 1\,\mathrm{d}\nu \le a^\star \int P\,\mathrm{d}\nu + b^\star \int Q\,\mathrm{d}\nu \le a^\star (2E)^2 + b^\star \delta^2.$$

The balance equation indicates that $(2E)^2 = p^\star P(t^\star)$ and $\delta^2 = p^\star Q(t^\star)$. Substituting these, the right-hand side comes out to be: $p^\star\big(a^\star P(t^\star) + b^\star Q(t^\star)\big)$, which is equal to $p^\star$, because the expression in brackets is 1 by the equality condition of the tangent line inequality. So one obtains $V(\delta) \le p^\star$.

When this is combined with the previously proven lower bound, we get $p^\star \le \omega(\delta, \mathcal{W}_{\ell,E})^2 \le V(\delta) \le p^\star$, so everything is forced equal: $\omega(\delta, \mathcal{W}_{\ell,E})^2 = V(\delta) = p^\star$, which is (4.53) and (4.55).

*The case $\delta \ge 2E$.* Here $\|A\varphi\|_{L^2(\mathbb{R})} \le \|\varphi\|_{H^\ell(\mathbb{R})} \le 2E \le \delta$ ($\|A\| = 1$, Theorem 4.4): the data constraint is inactive, localizing mass at low frequency gives $\omega(\delta, \mathcal{W}_{\ell,E}) = 2E$, and this is confirmed by the closed-form expression, its supremum attained at $\xi = 0$. □

**Theorem A (Exact Modulus of Continuity).**

*Let $0 < \delta < 2E$ and let $\xi^\star = \xi^\star(\delta)$ be the unique root of the balance equation (4.54).*

(i) (Closed form at every noise level.) *The modulus of continuity (4.47) comes out as:*

$$\omega(\delta, \mathcal{W}_{\ell,E}) = \frac{2E}{(1+\xi^{\star 2})^{\ell/2}} = \frac{\delta}{m(\xi^\star)}, \qquad 0 < \delta < 2E, \tag{4.58}$$

*and $\omega(\delta, \mathcal{W}_{\ell,E}) = 2E$ once $\delta \ge 2E$. The content of (4.58) is the quantifier: it is an equality for* every *$\delta \in (0, 2E)$, not the asymptotic statement*

$$\omega(\delta, \mathcal{W}_{\ell,E}) = (2E)^{2/(\ell+2)}\, \delta^{\ell/(\ell+2)}\big(1 + o(1)\big), \qquad \delta \downarrow 0,$$

*which part* (ii) *derives from it.*

(ii) (Asymptotics and the sharp constant.) *As $\delta \downarrow 0$,*

$$\omega(\delta, \mathcal{W}_{\ell,E}) = (2E)^{2/(\ell+2)}\, \delta^{\ell/(\ell+2)}\Big(1 - \tfrac{\ell}{\ell+2}\,\big(\delta/2E\big)^{2/(\ell+2)} + \mathcal{O}\big(\delta^{4/(\ell+2)}\big)\Big). \tag{4.59}$$

*so $(2E)^{2/(\ell+2)}$ is the leading constant, and it is also the smallest $C$ that makes $\omega(\delta, \mathcal{W}_{\ell,E}) \le C\,\delta^{\ell/(\ell+2)}(1 + o(1))$ hold, anything smaller than it fails. The balance frequency itself satisfies:*

$$K_\star = \big(2E/\delta\big)^{1/(\ell+2)}\big(1 + \mathcal{O}(\delta^{2/(\ell+2)})\big), \tag{4.60}$$

*so $K_\star \to \infty$, which puts the extremal frequency out in the ill posed high frequency range. that is worth noticing, because it says the sharp constant of Theorem 4.11 is not something that only happens near $\xi \to 0$, where $A \approx \mathrm{Id}$ anyway.*

(iii) (Attainment.) *For each fixed $\delta \in (0, 2E)$ the value (4.58) is reached in the limit by the Weyl windows of Definition 4.15, taken with centre fixed at $\xi^\star(\delta)$ and width shrinking, but no $L^2(\mathbb{R})$ function ever attains it.*

**Proof.** *Part (i)* matches Lemma 4.18, along with equation (4.55); the case of an inactive constraint $\delta \geq 2E$ is also discussed here. Whats really important is the tangent line inequality stated in Proposition 4.17, where the classic extreme point reduction of two atoms is reduced to one balance frequency, this is how the identity is created at all levels of noise.

*Part (ii)* Lets recall on the balance equation (4.54), The left hand side decreases from 1 to 0, thus $\delta \downarrow 0$ forces $\xi^\star \to \infty$, where $m(\xi) = \xi^{-2}(1 + \mathcal{O}(e^{-\xi^2}))$ (Lemma 4.7) and $(1 + \xi^2)^{\ell/2} = \xi^\ell(1 + \frac{\ell}{2}\xi^{-2} + \mathcal{O}(\xi^{-4}))$. substitution into the balance equation, (4.54) we obtain $\xi^{\star-(\ell+2)}(1 - \frac{\ell}{2}\xi^{\star-2} + \mathcal{O}(\xi^{\star-4})) = \delta/(2E)$, Thus, through the inversion, we get equation (4.60). Upon substitution of Equation (4.60) back into $\omega = 2E(1 + \xi^{\star2})^{-\ell/2}$ and expand again to second order in $\delta^{2/(\ell+2)}$, which yields (4.59). What happens here is that the two distinct $\mathcal{O}(\xi^{\star-2})$ corrections become one with the coefficient being $\frac{\ell}{\ell+2}$, so $(2E)^{2/(\ell+2)}$ is actually the leading constant, and this is approached from below. Minimality is evident, since the admissible constants have to be asymptotic upper bounds of the left hand side as $\delta \downarrow 0$.

*Part (iii)* Fix some $\delta \in (0, 2E)$ and choose some $w_n^{\mathrm{hi}}$ with a high frequency centered at $\xi^\star(\delta)$ with a small width $\varepsilon_n \downarrow 0$. then define $\beta_n$ such that $\beta_n := 2E/\|w_n^{\mathrm{hi}}\|_{H^\ell(\mathbb{R})}$, and the antipodal pair $\pm\frac{1}{2}\beta_n w_n^{\mathrm{hi}}$ are contained in $\mathcal{W}_{\ell,E}$, which is Lemma 4.16 (2). Then we consider both the data and the source separations. By Lemma 4.16 (3), keeping the centre fixed, and using the balance equation (4.54), the data separation converges to $2E\, m(\xi^\star)(1 + \xi^{\star2})^{-\ell/2}$, which is $\delta$, whereas the source separation $\beta_n$ converges to $2E(1 + \xi^{\star2})^{-\ell/2}$, which is the same as (4.58). So the supremum is approached. Nevertheless, none of the members of the sequence reaches this state. To have equality in the certificate (4.51) the whole spectral mass would have to sit on the two points $\pm\xi^\star$, which is not possible for any $L^2(\mathbb{R})$ function, the reason being that $A$ has no point spectrum at all (Theorem 4.6 (ii)). □

**Remark 4.19.** Theorem 4.11 was proven earlier using the pointwise interpolation inequality, however it can also be derived from the closed form above, in just three easy steps. Dividing (4.58) by $\delta^\ell$ and using the balance equation to eliminate $\delta$,

$$\frac{\omega(\delta, \mathcal{W}_{\ell,E})^{\ell+2}}{(2E)^2\,\delta^\ell} = \big[m(\xi^\star)(1 + \xi^{\star2})\big]^{-\ell} \leq 1, \tag{4.61}$$

the inequality being Lemma 4.7 (c). Rearranging gives $\omega(\delta, \mathcal{W}_{\ell,E}) \leq (2E)^{2/(\ell+2)}\delta^{\ell/(\ell+2)}$, which is (4.44), and it holds at every $\delta$ rather than only asymptotically. thus the sharp constant of Theorem 4.11 and the equality condition of Lemma 4.7 (c) are the same statement seen from two sides. Section 4.5.6 checks (4.61) numerically at 240 combinations of $(\ell, E, \delta)$.

**Remark 4.20.** The structure of (4.58), which represents an exact modulus of the Sobolev ball under action of a Fourier multiplier operator presented as a solution of a scalar equation of equilibrium, is the same as the one derived by Regińska and Tautenhahn [43, Thm. 4.2] for the Helmholtz Cauchy problem, based on [134]. Two things separate the present statement. To start with, the convexity assumption is not proved for $m$ anywhere in that

literature; however, Proposition 4.17 provides it through complete monotonicity. Second, [43, Eq. (4.4)] gives just the leading-order term $(1 + o(1))$, and [43, Rem. 3.3] presents the Hölder consequence as an upper bound, whereas (4.59) carries the second-order term and part (ii) identifies $(2E)^{2/(\ell+2)}$ as the least admissible constant. The attainment statement of part (iii) has no counterpart there.

**Remark 4.21.** The modulus, which was previously obtained, fixes the actual minimax error. In the case of the centrally symmetric and convex source set, set:

$$e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) := \inf_{T:L^2(\mathbb{R})\to L^2(\mathbb{R})} \sup_{\varphi\in\mathcal{W}_{\ell,E}} \sup_{\|\eta\|_{L^2(\mathbb{R})}\le\delta} \|T(A\varphi+\eta)-\varphi\|_{L^2(\mathbb{R})},$$

This infimum is over all measurable, possibly nonlinear reconstruction operators $T$. There are two closely related but essentially different definitions of this modulus, and they are important to keep apart. Alongside the difference set modulus $\omega(\delta, \mathcal{W}_{\ell,E})$ from (4.47), built from pairs of sources, there is a second, centred version, built from a single source inside the class itself,

$$\Omega(\delta) := \sup\big\{\|\varphi\|_{L^2(\mathbb{R})} : \varphi \in \mathcal{W}_{\ell,E},\ \|A\varphi\|_{L^2(\mathbb{R})} \le \delta\big\}, \tag{4.62}$$

and this second one, $\Omega(\delta)$, is the object the abstract calculus of Hofmann, Mathé, and Schieck is built on [87, §1, eq. (6)]. The two are not independent, though. The difference set reduction (4.56) together with the fact that scaling a source by 2 scales both constraints by 2 as well, gives an exact relationship between them

$$\omega(2\delta, \mathcal{W}_{\ell,E}) = 2\,\Omega(\delta), \qquad \text{i.e.} \qquad \Omega(\delta) = \tfrac{1}{2}\,\omega(2\delta, \mathcal{W}_{\ell,E}), \tag{4.63}$$

so as observed there is no approximation but an exact identity. and $\Omega(\delta)$ gives both a lower and upper bound on the true minimax error,

$$\Omega(\delta) \ \le\ e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) \ \le\ 2\,\Omega(\delta) = \omega(2\delta, \mathcal{W}_{\ell,E}), \tag{4.64}$$

the lower bound is obtained from an antipodal two point argument, as demonstrated in the proof of Theorem B, the upper bound is obtained from any Chebyshev central reconstruction, because the difference between two sources compatible with the same datum does not exceed $\omega(2\delta, \mathcal{W}_{\ell,E})$. That is the two sided comparison provided in [87, Prop. 3]. Evaluating (4.58) at $2\delta$, one gets that both bounds take on real values at each noise level and differ only by the universal factor 2; thus, $\ell/(\ell + 2)$ is the irreducible minimax rate of convergence on $\mathcal{W}_{\ell,E}$, with the constant in Theorem 4.11 proven to be *exactly* equal to the modulus, rather than being just an upper bound on it. The remaining factor 2 is eliminated in Theorem B

**Theorem B (Exact Minimax Error).**

*For every $0 < \delta < E$,*

$$e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) = \Omega(\delta) = \frac{E}{\big(1 + (\xi^\circ)^2\big)^{\ell/2}} = \tfrac{1}{2}\,\omega(2\delta, \mathcal{W}_{\ell,E}), \tag{4.65}$$

*where $\xi^\circ = \xi^\circ(\delta)$ is the unique root of the centred balance equation*

$$\frac{m(\xi^\circ)}{\big(1 + (\xi^\circ)^2\big)^{\ell/2}} = \frac{\delta}{E}; \tag{4.66}$$

*for $\delta \geq E$, $e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) = E$. Asymptotically,*

$$e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) = E^{2/(\ell+2)}\,\delta^{\ell/(\ell+2)}\,(1 + o(1)) = 2^{-2/(\ell+2)}\,\omega(\delta, \mathcal{W}_{\ell,E})\,(1 + o(1)), \qquad \delta \downarrow 0. \tag{4.67}$$

*The infimum over all reconstruction operators is achieved via a linear scheme: Tikhonov regularization in the Hilbert scale, the family of spectral filters $q_\lambda$ of (2.5) with an appropriately chosen parameter. Therefore, the optimal worst case error over all nonlinear schemes equals that of the best linear one.*

*This particular worst case is the one defining $\Omega$ in (4.62), taking into account only the constraint set $\{\varphi \in \mathcal{W}_{\ell,E} : \|A\varphi\|_{L^2(\mathbb{R})} \leq \delta\}$. It is not the sum of a bias evaluated on $\mathcal{W}_{\ell,E}$ and a noise part taken separately on the ball of radius $\delta$, which would be the most natural measure of performance of a linear filter (the quantity minimised in Table 4.7).*

**Proof.** The scalar reduction established in Lemma 4.18 can be applied directly, now replacing $2E$ with $E$, because there is no difference set step and the reduction is done immediately on $\mathcal{W}_{\ell,E}$. The result is that $\Omega(\delta) = E(1 + (\xi^\circ)^2)^{-\ell/2}$, along with the balancing equation (4.66) for $0 < \delta < E$, and $\Omega = E$ beyond. The scale relation (4.63) gives the final equality in (4.65), and the asymptotic formula (4.59) with inputs $(E, \delta)$ gives (4.67).

*Lower bound* Let us take a reconstruction operator $T : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ and consider a function $\varphi$ from the class $\mathcal{W}_{\ell,E}$ such that $\|A\varphi\|_{L^2(\mathbb{R})} \leq \delta$. The key point is that the same observation $y = 0$ can be generated by two different sources. For example, if we have the source $+\varphi$ with noise $\eta = -A\varphi$, the observed data will be equal to zero. Conversely, the source $-\varphi$ with noise $\eta = +A\varphi$ will also generate zero data. As the norms of both noises are bounded by $\delta$, both cases are possible, and $\pm\varphi \in \mathcal{W}_{\ell,E}$ because of central symmetry. Hence the error of $T$ can be bounded below by $\max\{\|T(0) - \varphi\|, \|T(0) + \varphi\|\} \geq \frac{1}{2}\big(\|T(0) - \varphi\| + \|T(0) + \varphi\|\big) \geq \|\varphi\|_{L^2(\mathbb{R})}$. The supremum over all such $\varphi$ gives $e_{\mathrm{mm}}(\delta, \mathcal{W}_{\ell,E}) \geq \Omega(\delta)$.

*Upper bound* Here we show that $\Omega(\delta)$ is not only a lower bound but a value that can be achieved. First of all, we reparameterize the Sobolev ball as the image of the unit $L^2(\mathbb{R})$ ball via the bounded injection $E\,(I - \partial_{xx})^{-\ell/2}$. Having completed the reparameterization, we have a canonical estimation problem, since it is now necessary to recover $\varphi$ from $A\varphi + \eta$ with $\|\eta\| \leq \delta$. This problem concerns the optimal estimation of a bounded operator with respect to two ellipsoids. Melkman and Micchelli [106, Thm. 3.1] prove that smoothing the data with Tikhonov regularization followed by the solution map is the optimal recovery method in terms of worst case error, which coincides with the centred extremal value $\Omega(\delta)$ [87, Prop. 3, Thm. 3,

Cor. 2]; see [38] for the framework and [63] for the Gaussian white noise counterpart. Here the optimal linear smoother is the Hilbert scale Tikhonov filter $q_\lambda$ of (2.5), matched to $\xi^\circ(\delta)$, and it is exactly the family Theorem C identifies inside the trained estimator. On combining the two bounds one obtains (4.65). In the case $\delta \geq E$ the estimator $T \equiv 0$ produces an error no greater than $E$, matching $\Omega(\delta) = E$. □

### 4.4.4 Several Observation Times (an Exact $\sqrt{J}$ Law)

The theoretical basis of the empirical result that has already been observed is discussed first. The comparative experiment introduced in Chapter 5 provides the method with snapshots of the system state at multiple intermediate points in time (Section 5.3.5), and the accuracy does improve. This work presents the worst case theory of this improvement in closed form.

First, it is important to emphasize what has already been proven. Exact optimal recovery of the solution of the heat equation is classical (it is known within the Osipenko school, [101]), and identifiability of sources on terminal or time distributed data has also been solved [53, 55, 89].

In fact, the idea of the scaling effect itself, where the effective noise decreases to $\delta/\sqrt{J}$ when there are $J$ measurements within a single noise budget, is an established notion in the theory of information based complexity [54, §2.10.1], and therefore this part is not innovative here. What is innovative here is only one exact relationship, which is $\omega_\oplus(\delta) = \delta/m_\oplus(\xi^\star_\oplus)$. The stacked multiplier built from $m_\tau(\xi) = (1 - e^{-\tau\xi^2})/\xi^2$ does not constitute the power of some certain fixed function. The latter fact alone makes it impossible for this particular case to be treated within the framework of the exact modulus theorem of [101, Thm. 1], as that theorem is actually stated; so obtaining an exact answer here seems, to the best of our knowledge, to be novel.

Notation aside: in all cases within this section, $J$ is nothing but the total number of observation times and applies only locally; it has nothing whatsoever to do with the reduced loss function $J(\theta)$ discussed in Section 3.3.5.

Suppose that observation times are ordered such that $0 < t_1 < \cdots < t_J = 1$, and for any $\tau \in (0, 1]$, define $A_\tau\varphi := v(\cdot\,, \tau)$, a Fourier multiplier whose symbol is

$$m_\tau(\xi) := \frac{1 - e^{-\tau\xi^2}}{\xi^2} \ (\xi \neq 0), \qquad m_\tau(0) := \tau, \qquad m_1 = m, \tag{4.68}$$

Proceeding in a similar way to Definition 4.3, but this time stopping at time $\tau$, instead of 1, leads to the definition of $A_\tau$. Stacking, which concatenates all the channels together, gives us the following operator and symbol:

$$A_\oplus\varphi := \big(A_{t_1}\varphi, \ldots, A_{t_J}\varphi\big) \in \big(L^2(\mathbb{R})\big)^J, \qquad m_\oplus(\xi) := \Big(\sum_{j=1}^{J} m_{t_j}(\xi)^2\Big)^{1/2}, \tag{4.69}$$

with the product norm $\|(g_j)_j\|_\oplus^2 := \sum_j \|g_j\|_{L^2(\mathbb{R})}^2$. Plancherel applied separately on each channel gives us $\|A_\oplus\varphi\|_\oplus^2 = \int_{\mathbb{R}} m_\oplus^2 |\hat{\varphi}|^2 \mathrm{d}\xi$, and there is no essential difference except that $m$ is replaced by $m_\oplus$. The modulus, computed with a common noise energy budget, is defined by:

$$\omega_\oplus(\delta, \mathcal{W}_{\ell,E}) := \sup\Big\{ \|\varphi_1 - \varphi_2\|_{L^2(\mathbb{R})} : \varphi_1, \varphi_2 \in \mathcal{W}_{\ell,E},\ \|A_\oplus(\varphi_1 - \varphi_2)\|_\oplus \leq \delta \Big\}. \tag{4.70}$$

**Theorem 4.22.** *Let $0 < t_1 < \cdots < t_J = 1$ be fixed and let $T_\oplus := \left(\sum_{j=1}^J t_j^2\right)^{1/2} \geq 1$.*

(i) (Exact modulus.) *For any $0 < \delta < 2E\,T_\oplus$,*

$$\omega_\oplus(\delta, \mathcal{W}_{\ell,E}) = \frac{2E}{\left(1+\xi_\oplus^{\star 2}\right)^{\ell/2}} = \frac{\delta}{m_\oplus(\xi_\oplus^\star)}, \tag{4.71}$$

*where $\xi_\oplus^\star = \xi_\oplus^\star(\delta)$ is the unique solution to the balance equation for the stacked configuration,*

$$\frac{m_\oplus(\xi_\oplus^\star)}{\left(1+\xi_\oplus^{\star 2}\right)^{\ell/2}} = \frac{\delta}{2E}; \tag{4.72}$$

*and for $\delta \geq 2E\,T_\oplus$ the constraint for the stacked configuration is inactive, so $\omega_\oplus(\delta, \mathcal{W}_{\ell,E}) = 2E$.*

(ii) (The $\sqrt{J}$ law.) *As $\delta \downarrow 0$,*

$$\omega_\oplus(\delta, \mathcal{W}_{\ell,E}) = (2E)^{2/(\ell+2)} \left(\frac{\delta}{\sqrt{J}}\right)^{\ell/(\ell+2)} \big(1 + o(1)\big) : \tag{4.73}$$

*for the same amount of energy spent on noise, $J$ snapshots correspond to a single snapshot of noise strength $\delta/\sqrt{J}$.*

(iii) (Collapse per channel.) *On the other hand, if every channel has its own budget, so that the condition in (4.70) is substituted by $\|A_{t_j}(\varphi_1 - \varphi_2)\|_{L^2(\mathbb{R})} \leq \delta$ for every $j$, then the feasible region is identical to the region corresponding to just the final snapshot. Hence the modulus is $\omega(\delta, \mathcal{W}_{\ell,E})$, as asserted in Theorem A, exactly, for all $\delta > 0$: when there is worst case coherent noise, multiple snapshots provide no benefit.*

**Proof.** **(i)** The reduction of the difference set and moment problems used to prove Lemma 4.18 is not required to be developed here. All it used was Plancherel's formula and the fact that the operator acts as a multiplier; both are still true, and hence the reduction holds again, but with $m$ replaced by $m_\oplus$ everywhere. The only thing that has to be checked is whether $Q_\oplus := m_\oplus^2$, as a function of $t = \xi^2$, can be written in the Laplace form given by (4.49).

It can. Since $m_\tau = \int_0^\tau e^{-st} \mathrm{d}s$ in the variable $t$, the same self convolution computation as in Proposition 4.17 (iii) gives

$$m_\tau^2 = \int_0^{2\tau} e^{-rt}\, \min\{r,\, 2\tau - r\}\, \mathrm{d}r, \qquad \text{and hence} \qquad Q_\oplus = \int_0^2 e^{-rt}\, W_\oplus(r)\, \mathrm{d}r,$$

where $W_\oplus(r) := \sum_{j=1}^J \min\{r,\, 2t_j - r\}\, \chi_{[0,2t_j]}(r) \geq 0$ is a nonzero compactly supported density.

Thus, using Proposition 4.17, we see that $Q_\oplus$ is strictly convex w.r.t. $P = (1+t)^\ell$, along with the tangent line bound at $t^\star$. The crossing argument remains exactly the same. The ratio $Q_\oplus(t)/P(t)$ is strictly decreasing and continuous since the logarithmic derivative of the function satisfies $Q_\oplus'/Q_\oplus - \ell/(1+t) < 0$. At $t = 0$ the value of the function is $Q_\oplus(0) = \sum_j m_{t_j}(0)^2 = T_\oplus^2$, and it decreases to 0. Thus, there is one unique solution to the balance equation (4.72), and it exists precisely when $\delta/(2E) < T_\oplus$. Using the tangent line inequality, we can obtain the value of the moment problem at the balance frequency, which is (4.71).

The constraint does not impose any further effect for $\delta \geq 2E\,T_\oplus$, since $\|A_\tau\| = \sup_\xi m_\tau = \tau$ (the integrands of $m_\tau = \int_0^\tau e^{-s\xi^2}\mathrm{d}s$ get smaller when $|\xi|$ gets larger). Therefore $\|A_\oplus\varphi\|_\oplus \leq T_\oplus\, 2E \leq \delta$, and the modulus achieves the value $2E$, which is the same conclusion as in Lemma 4.18.

**(ii)** At high frequencies, all channels become equivalent, and thus $m_{t_j}(\xi)^2 = \xi^{-4}(1 + \mathcal{O}(e^{-t_1\xi^2}))$ as $|\xi| \to \infty$. Hence, the stacked symbol brings nothing else but the $\sqrt{J}$ factor, that is $m_\oplus = \sqrt{J}\,\xi^{-2}(1 + \mathcal{O}(e^{-t_1\xi^2}))$. Substituting in the balance formula (4.72) and using the inversion procedure from Theorem A (ii): $\xi_\oplus^\star = (2E\sqrt{J}/\delta)^{1/(\ell+2)}(1 + o(1))$, and hence $\omega_\oplus(\delta, \mathcal{W}_{\ell,E}) = 2E\,\xi_\oplus^{\star\,-\ell}(1 + o(1)) = (2E)^{2/(\ell+2)}(\delta/\sqrt{J})^{\ell/(\ell+2)}(1 + o(1))$.

**(iii)** Since the integral in $m_\tau(t) = \int_0^\tau e^{-st}\mathrm{d}s$ is taken over a larger interval, $m_\tau$ is strictly increasing in $\tau$ for all fixed $t \geq 0$. This means that $m_{t_j} \leq m_{t_J} = m$ pointwise, and therefore $\|A_{t_j}\psi\|_{L^2(\mathbb{R})} \leq \|A\psi\|_{L^2(\mathbb{R})}$ for all $\psi \in L^2(\mathbb{R})$ and all $j$. Thus the channel constraint at the end time already satisfies all the other channel constraints, and the feasible set per channel is $\{\psi : \|\psi\|_{H^\ell(\mathbb{R})} \leq 2E,\ \|A\psi\|_{L^2(\mathbb{R})} \leq \delta\}$. This is the same as the feasible set of (4.56), and Theorem A (i) then finishes the proof. □

### *4.4.5 The Hermite V-PINN as Hilbert Scale Regularization*

Up to this point, we have derived our results regarding $A$ and the source class $\mathcal{W}_{\ell,E}$. In what follows we discuss the method.

The bilevel elimination technique of Proposition 3.7 *is* the Tikhonov regularisation method in the Hermite-Sobolev scale, with the spectral filter given by $\gamma_k/(\gamma_k + \rho)$. Given this, the recovered source function possesses the convergence rate provided in Theorem 4.11 on the basis of the two conditions provided in Theorem C (b); furthermore, this rate is optimal as demonstrated in Theorem A and Theorem B.

In this section, the source identified using the eliminated coefficient map (3.12) is expressed as $\widehat{\varphi} := \varphi^{c^\star(\theta)} = \sum_{j=0}^{N_\varphi - 1} c_j^\star(\theta)\,\psi_j^\sigma$.

Let $\rho$ be the ratio of penalty to residual weight, that is $\rho := \lambda/\mu$. We let $\mathbf{G}$ be the $\mathbf{W}_r$-weighted Hermite Gram matrix, which is the matrix $\mathbf{G} := \Phi^\top \mathbf{W}_r \Phi \in \mathbb{R}^{N_\varphi \times N_\varphi}$. This is a symmetric positive definite matrix due to the full column rank property of the Hermite design $\Phi$ at the residual points.

We now state the definition of the inner product for the vector space endowed with the $\Lambda$ matrix, $\langle a, b\rangle_\Lambda := a^\top \Lambda b$, as well as the norm $\|a\|_\Lambda^2 = \langle a, a\rangle_\Lambda$. In a similar manner, the norm of the residual generator $\mathbf{g}_\theta$ in the $\mathbf{W}_r$-weighted space is $\|\mathbf{g}_\theta\|_{\mathbf{W}_r}^2 := \mathbf{g}_\theta^\top \mathbf{W}_r \mathbf{g}_\theta$.

**Theorem C (Identification and Order Optimality of the Hermite V-PINN as Hilbert Scale Tikhonov Regularization).**

*Let $\mu$ and $\lambda$ to represent positive real constants and set $\rho = \lambda/\mu$.*

**(a) Identification of filter.** *There is a set of generalized eigenpairs $(\gamma_k, u_k)$, for $k = 0, \dots, N_\varphi - 1$, related to the matrix pair $(\mathbf{G}, \Lambda)$ satisfying $\mathbf{G}u_k = \gamma_k \Lambda u_k$, where the eigenvalues are sorted in the order $\gamma_0 \geq \cdots \geq \gamma_{N_\varphi - 1} \geq 0$ and the eigenvectors are orthonormal under the $\Lambda$ inner product, that is $\langle u_j, u_k \rangle_\Lambda = \delta_{jk}$. The map of eliminated coefficients (3.12) in this basis acts diagonally, meaning that*

$$\langle c^\star(\theta), u_k \rangle_\Lambda = f_k(\rho)\, \langle \widetilde{c}(\theta), u_k \rangle_\Lambda, \qquad f_k(\rho) = \frac{\gamma_k}{\gamma_k + \rho}, \tag{4.74}$$

*where $\widetilde{c}(\theta) := \mathbf{G}^{-1}\Phi^\top \mathbf{W}_r \mathbf{g}_\theta$ is the unregularised $\mathbf{W}_r$ weighted least squares projection of the residual generator (3.8) onto the Hermite source space.*

*The $f_k(\rho)$ form the Tikhonov filter factors, such that $0 \leq f_k(\rho) \leq 1$ for any $k$ and $f_k(\rho)$ increase monotonically to 1 as $\rho \downarrow 0$. The regularisation of the weighted generalised inverse of the design map takes the form of the solution operator $R_\rho : \mathbf{g}_\theta \mapsto c^\star(\theta) = (\mathbf{G} + \rho\Lambda)^{-1}\Phi^\top \mathbf{W}_r \mathbf{g}_\theta$, in the sense of [68, Ch. 3, Def. 3.1], with the stability bound*

$$\|c^\star(\theta)\|_\Lambda \leq \tfrac{1}{2}\, \rho^{-1/2}\, \|\mathbf{g}_\theta\|_{\mathbf{W}_r} = \tfrac{1}{2}\sqrt{\mu/\lambda}\, \|\mathbf{g}_\theta\|_{\mathbf{W}_r}. \tag{4.75}$$

*In fact, since $\lambda\|c\|_\Lambda^2$ is the squared discrete $H^\ell(\mathbb{R})$ penalty for $\varphi^c$ (Lemma 4.2), the assignment $\mathbf{g}_\theta \mapsto \widehat{\varphi}$ represents a Tikhonov regularization method in the Hilbert scale defined by the Hermite number operator, with penalty of order $\ell$ and qualification order$\ell$ [115][68, Ch. 8, §8.5].*

**(b) Order-optimality.** *Let us suppose further that*

(B1) *the estimated parameters define the forward map of the recovered source in the sense of the data error, that is $\|A\widehat{\varphi} - g^\delta\|_{L^2(\mathbb{R})} \leq \varepsilon$ with $\varepsilon \leq \delta$, where $g^\delta$ is the observed datum of (4.4); and*

(B2) *the source has effective Hilbert-scale smoothness of degree $s \geq \ell$, meaning that both $\varphi^\dagger$ and its approximation $\widehat{\varphi}$ lie in the ball $\mathcal{W}_{s,E} = \{\varphi \in H^s(\mathbb{R}) : \|\varphi\|_{H^s(\mathbb{R})} \leq E\}$.*

*Then*

$$\|\widehat{\varphi} - \varphi^\dagger\|_{L^2(\mathbb{R})} \leq (2E)^{2/(s+2)}\, (\delta + \varepsilon)^{s/(s+2)} \leq 2^{s/(s+2)}\, (2E)^{2/(s+2)}\, \delta^{s/(s+2)}, \tag{4.76}$$

*and from the exact value of modulus and minimax as defined by Theorem A and Theorem B at smoothness index $s$, there is no linear or nonlinear reconstruction operator, that can attain a smaller worst case $L^2(\mathbb{R})$ error rate on $\mathcal{W}_{s,E}$ at the smoothness parameter $s$. Thus the Hermite V-PINN is order optimal, with rate $\delta^{s/(s+2)}$, the minimax rate of the class.*

**Proof.** **(a)** Because both $\Lambda$ and $\mathbf{G}$ are positive definite matrices, the definite matrix pair $(\mathbf{G}, \Lambda)$ possesses an $\Lambda$-orthonormal generalised eigenpair system. Let $q_k$ be the orthonormal eigenvectors of $\Lambda^{-1/2}\mathbf{G}\Lambda^{-1/2}$, and let $u_k := \Lambda^{-1/2} q_k$. Then $\mathbf{G}u_k = \gamma_k \Lambda u_k$ and $\langle u_j, u_k \rangle_\Lambda = \delta_{jk}$. This represents the simultaneous diagonalisation of a definite matrix pair [56, §12.3],[42].

Let $r := \Phi^\top \mathbf{W}_r \mathbf{g}_\theta$ and expand $c^\star(\theta) = \sum_j a_j u_j$. By the equation of (3.12), $(\mu\mathbf{G} + \lambda\Lambda)c^\star = \mu r$, along with $(\mu\mathbf{G} + \lambda\Lambda)u_k = (\mu\gamma_k + \lambda)\Lambda u_k$, left multiplication by $u_k^\top$ gives

$$a_k\,(\mu\gamma_k + \lambda) = \mu\, u_k^\top r, \qquad \text{i.e.} \qquad \langle c^\star(\theta), u_k\rangle_\Lambda = a_k = \frac{u_k^\top r}{\gamma_k + \rho}.$$

If we take $\widetilde{c} = \mathbf{G}^{-1} r$, then the same test produces $\langle \widetilde{c}, u_k\rangle_\Lambda = u_k^\top r/\gamma_k$, from which we obtain $a_k = \frac{\gamma_k}{\gamma_k+\rho}\langle \widetilde{c}, u_k\rangle_\Lambda$, which is (4.74); $0 \le f_k \le 1$ and $f_k \uparrow 1$ are immediate.

In this case we write $y := \mathbf{W}_r^{1/2}\mathbf{g}_\theta$ and denote $B := \mathbf{W}_r^{1/2}\Phi$, the weighted design matrix, such that $\mathbf{G} = B^\top B$ and $r = B^\top y$. We have $\widetilde{B} := B\Lambda^{-1/2}$. Hence $\Lambda^{1/2}c^\star = (\widetilde{B}^\top\widetilde{B} + \rho I)^{-1}\widetilde{B}^\top y$, which represents ordinary Tikhonov regularisation with squared singular values $\gamma_k$, and the scalar filter estimates $\sup_\gamma \gamma(\gamma+\rho)^{-1} \le 1$ and $\sup_\gamma \sqrt{\gamma}(\gamma+\rho)^{-1} = \frac{1}{2}\rho^{-1/2}$ [8, §6][4, §3.3.2] identify $R_\rho$ as a regularisation [8, §6, p. 58],[68, Ch. 3, Def. 3.1] and yield $\|c^\star(\theta)\|_\Lambda \le \frac{1}{2}\rho^{-1/2}\|y\|_2$, which is (4.75).

Finally, due to Lemma 4.2, the penalty $\lambda\|c\|_\Lambda^2$ controls $\lambda\, C_{\sigma,\ell}^{-2}\|\varphi^c\|_{H^\ell(\mathbb{R})}^2$. Thus the scheme is Tikhonov regularisation with respect to the Hermite generated Hilbert scale, with penalty order $\ell$ [115][68, Ch. 8, §8.5].

**(b)** From the noise model (4.4) it follows that $\|g^\delta - A\varphi^\dagger\|_{L^2(\mathbb{R})} \le \delta$. Through the triangle inequality and the discrepancy condition (B1), one gets

$$\|A\widehat{\varphi} - A\varphi^\dagger\|_{L^2(\mathbb{R})} \le \|A\widehat{\varphi} - g^\delta\|_{L^2(\mathbb{R})} + \|g^\delta - A\varphi^\dagger\|_{L^2(\mathbb{R})} \le \varepsilon + \delta.$$

Since we assume (B2), both $\widehat{\varphi}$ and $\varphi^\dagger$ belong to $\mathcal{W}_{s,E}$. Hence, invoking Theorem 4.11 at smoothness parameter $s$, or alternatively Corollary 4.12 with $\ell \leftarrow s$, whose proof uses only Theorem 4.11 and the triangle inequality, we obtain the first inequality of (4.76).

The second claim is a consequence of $\varepsilon \le \delta$ and the monotonicity of the mapping $t \mapsto t^{s/(s+2)}$. The associated lower bound matches exactly the minimax quantity of Theorem B with parameter $s$: each part of Section 4.4.3, namely the certificate of Proposition 4.17, the reduction of the problem to a one-dimensional one, and the Weyl windows of Definition 4.15 calibrated at $K_n^\star \asymp (E/\delta)^{1/(s+2)}$, depends on $\ell$ only through the index of the class and thus works as stated with $\ell \leftarrow s$, giving $e_{\mathrm{mm}}(\delta, \mathcal{W}_{s,E}) = E^{2/(s+2)}\delta^{s/(s+2)}(1+o(1))$; no reconstruction scheme can achieve a better worst case accuracy. This implies that the upper and lower bounds have the same exponent $s/(s+2)$, which is order optimality. □

**Remark 4.23.** The scheme in (a) makes use of one residual block, but the loss function defined by (4.18) consists of two residual blocks, which are put together in (4.17) The rationale behind it is similar. We first define $\mathbf{G}_{\mathrm{hyb}} := \mu_{\mathrm{coll}}\mathbf{H}_{\mathrm{coll}} + \mu_{\mathrm{wf}}\mathbf{H}_{\mathrm{wf}} \succeq 0$,and the matrix will be positive definite if all the columns have full rank. We construct $B_{\mathrm{hyb}}$ using both the weighted residual matrices, so we have $\mathbf{G}_{\mathrm{hyb}} = B_{\mathrm{hyb}}^\top B_{\mathrm{hyb}}$ and $r_{\mathrm{hyb}} = B_{\mathrm{hyb}}^\top y_{\mathrm{hyb}}$. Consequently, we have $c^\star = (\mathbf{G}_{\mathrm{hyb}} + \lambda\Lambda)^{-1} r_{\mathrm{hyb}}$, since the matrix pair $(\mathbf{G}_{\mathrm{hyb}}, \Lambda)$ can be simultaneously diagonalized according to the method stated in part (a). This results in the fact that the filters have the structure $f_k(\lambda) = \gamma_k/(\gamma_k + \lambda)$, and the stability bound (4.75) takes the form $\|c^\star\|_\Lambda \le \frac{1}{2}\lambda^{-1/2}\|y_{\mathrm{hyb}}\|_2$. The square of the Euclidean norm of $y_{\mathrm{hyb}}$ is

$\|y_{\mathrm{hyb}}\|_2^2 = \mu_{\mathrm{coll}}\|y_{\mathrm{coll}}\|^2 + \mu_{\mathrm{wf}}\|y_{\mathrm{wf}}\|^2$. This derivation depends only on the definiteness of the pair and the filter structure, both of which still apply.

**Remark 4.24.** An unconditional identity for the discrete map (3.12) is provided in part (a). Without regard to how the source map of the V-PINN is seen from outside, it is always a classical Hilbert scale Tikhonov filter, whose factors (4.74) damp exactly those high Hermite modes which are most strongly amplified by the ill posedness.

Part (b) is contingent on both of its assumptions. The assumption (B1) serves as the consistency criterion that can later be verified based on the misfit obtained in Section 4.5.3. Hypothesis (B2) defines the source condition, in which the improved qualification due to the order-$\ell$ penalty allows the attainment of exponents above the $L^2(\mathbb{R})$ Tikhonov saturation of 2/3 [115],[7, §4,§5]. The exponent attained is then determined by the effective smoothness of the source, expressed by the formula $\hat{\alpha} = s_{\mathrm{eff}}/(s_{\mathrm{eff}} + 2)$, corresponding exactly to what is measured in Section 4.5.4.

### 4.4.6 A-Posteriori Control of the Hypotheses

It is worth mentioning that the proof of Part (b) of Theorem C depends on assumptions (B1) and (B2). The main problem is how each of these assumptions can be verified to hold true for a trained neural network. Both conditions are derived from quantities available from the training log itself. Linking the two parts is accomplished through the continuum loss defined by (3.9) and the Duhamel contraction bound proved in Section 1.3.4, referred to as the "error control bridge" there, as well as the norm bound from Lemma 4.2. The following reasoning will not make any assumptions about whether the optimiser succeeded; instead, it will take whatever loss the optimiser did reach and turn it into hard statements about the recovered source.

**Proposition 4.25.** *Let $(\theta, c)$ be any admissible trial pair, with $v_\theta$ the trial state of Definition 3.4 and $\varphi^c$ the Hermite trial source. Assume that the continuum loss of Definition 3.6 is bounded by $\mathcal{L}(\theta, c) \le \vartheta^2$ for some $\vartheta > 0$, where the residual weight is $\mu > 0$ and the penalty weight is $\lambda > 0$. Then*

$$\|A\varphi^c - g^\delta\|_{L^2(\mathbb{R})} \le \left(1 + \mu^{-1/2}\right)\vartheta, \tag{4.77}$$

$$\|c\|_\Lambda \le \vartheta\,\lambda^{-1/2}, \qquad \|\varphi^c\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}\,\vartheta\,\lambda^{-1/2}. \tag{4.78}$$

**Proof.** The trial state meets the condition $v_\theta(\,\cdot\,, 0) = 0$ everywhere, and this is the exact reason for the presence of the factor $t$ in the trial state as defined in (3.6). It is also smooth with Gaussian decay, and thus becomes the only mild solution of $\partial_t v - \partial_{xx} v = \varphi^c + R$, where $v(\,\cdot\,, 0) = 0$ and $R := R_{\mathrm{PDE}}(\theta, c) \in L^2(\mathbb{R} \times (0,1)) \subset L^1(0,1; L^2(\mathbb{R}))$.

Using Duhamel's principle (Proposition 1.18) at the terminal time and (4.34), we get

$$v_\theta(\,\cdot\,, 1) - A\varphi^c = \int_0^1 e^{(1-s)\Delta}\, R(\,\cdot\,, s)\, \mathrm{d}s.$$

As the heat semigroup is contractive (Proposition 1.17), and by using the Cauchy Schwarz

inequality on the unit time interval, we get

$$\|v_\theta(\cdot,1) - A\varphi^c\|_{L^2(\mathbb{R})} \le \int_0^1 \|R(\cdot,s)\|_{L^2(\mathbb{R})}\,\mathrm{d}s \le \|R\|_{L^2(\mathbb{R}\times(0,1))} \le \vartheta\,\mu^{-1/2},$$

where this last step is based on the residual block of (3.9). From the data block of (3.9) we get $\|v_\theta(\cdot,1) - g^\delta\|_{L^2(\mathbb{R})} \le \vartheta$, and from this, using the triangle inequality, we get (4.77). From the penalty block we obtain $\lambda\|c\|_\Lambda^2 \le \vartheta^2$, and Lemma 4.2 converts this $\Lambda$ bound into the $H^\ell(\mathbb{R})$ bound of (4.78). □

**Corollary 4.26.** *Given that $\varphi^\dagger \in \mathcal{W}_{\ell,E}$, suppose that our learned parameters $(\widehat{\theta},\widehat{c})$ meet the constraint $\mathcal{L}(\widehat{\theta},\widehat{c}) \le \vartheta^2$, with the following additional restrictions:*

$$\left(1+\mu^{-1/2}\right)\vartheta \;\le\; \delta \qquad \text{and} \qquad \lambda = (\vartheta/E)^2. \tag{4.79}$$

*Then assumption (B1) holds with $\varepsilon := (1+\mu^{-1/2})\vartheta \le \delta$, and the solution obtained obeys $\|\varphi^{\widehat{c}}\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}E$, implying that both $\varphi^{\widehat{c}}$ and $\varphi^\dagger$ lie in the expanded ball $\mathcal{W}_{\ell,C_{\sigma,\ell}E}$ (using $C_{\sigma,\ell} \ge 1$, (4.30)). Hence Theorem 4.11 at index $\ell$ gives the fully checkable bound*

$$\|\varphi^{\widehat{c}} - \varphi^\dagger\|_{L^2(\mathbb{R})} \;\le\; \left(2\,C_{\sigma,\ell}\,E\right)^{2/(\ell+2)}(2\delta)^{\ell/(\ell+2)}. \tag{4.80}$$

*In this case the rate $\delta^{\ell/(\ell+2)}$ is the minimax rate of $\mathcal{W}_{\ell,E}$ (Theorem A), and the constant exceeds the sharp bound of Theorem 4.11 only by $C_{\sigma,\ell}^{2/(\ell+2)}$, at the deployed $(\sigma,\ell) = (2,2)$ at most $(105/64)^{1/4} \approx 1.13$ ((4.31)). Every hypothesis in (4.79) is a number from the training log.*

**Proof.** It can be seen from the relations (4.77) and (4.79) that $\|A\varphi^{\widehat{c}} - g^\delta\|_{L^2(\mathbb{R})} \le \varepsilon \le \delta$, and therefore the assumption (B1) is fulfilled. In addition to that, with $\lambda = (\vartheta/E)^2$, (4.78) leads to $\|\varphi^{\widehat{c}}\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}E$, and $\varphi^\dagger \in \mathcal{W}_{\ell,E} \subseteq \mathcal{W}_{\ell,C_{\sigma,\ell}E}$ since $C_{\sigma,\ell} \ge 1$; hence both sources belong to the same expanded ball. From this it follows, using (4.4) and the triangle inequality, that $\|A\varphi^{\widehat{c}} - A\varphi^\dagger\|_{L^2(\mathbb{R})} \le 2\delta$, and Theorem 4.11 on the expanded ball gives (4.80). □

### 4.4.7 On the Choice of Regularization Exponent

There can be some confusion about the two separate exponents for the shrinking of $\lambda$, which is discussed in this section. The two exponents reappear in Chapter 5, so it is necessary to carefully define the meaning of each.

The first is the matched a-priori optimum, $\lambda^\star \asymp \delta^2$ (Corollary 4.13, Proposition 2.13). In the case where the $H^\ell(\mathbb{R})$ regularisation is adapted to the $H^\ell(\mathbb{R})$ ball $\mathcal{W}_{\ell,E}$ on which it is being used, an optimal bias-variance trade off is achieved. Following this approach leads to the minimax rate that has already been proved in Theorem A to be theoretically the best in this setting.

Second, the classical rate of convergence is defined as $\lambda \sim \delta^{2/(2\nu+1)}$, where the special case $\nu = 1$ yields $\delta^{2/3}$. This is another older, alternative result [57, §4,§5] on the $L^2(\mathbb{R})$ penalised Tikhonov regularisation method, formulated under an alternative definition of smoothness,

$\varphi^\dagger = (A^*A)^\nu z$. It is known that this rate of convergence is limited: no matter how big the smoothness parameter $\nu$ becomes, once it reaches 1 the rate cannot get any better.

These two statements are not in contradiction, because they simply answer two different questions.

The question, then, is why the deployed schedule (4.28) uses $\delta^{2/3}$, the second of the two, despite the fact that the penalty in this case is Sobolev matched, corresponding to the $H^2(\mathbb{R})$ mass matrix $\Lambda$. The usage of $\delta^{2/3}$ is not the result of a misunderstanding but a conscious numerical choice: a conservative floor chosen to stay safely above the machine precision level of $10^{-9}$. At the noise levels considered, $\delta^2$ based regularisation is too soft to be numerically safe, although it is theoretically correct. The speed of decay of the error depends on the effective smoothness of $\varphi^\dagger$ and not on this mismatch (Remark 4.29).

Another important aspect to note is that the method of adjoints, as explained in Section 5.2, utilises the same numerical decision. In addition, since it undergoes Sobolev penalisation, it experiences the same point of saturation with sufficiently smooth sources (Tables 5.1 and 5.7).

## 4.5 Numerical Verification

### *4.5.1 Experimental Setup*

The objective function is the target function (4.5), $\varphi^\dagger(x) = (x^3/4 - 3x/2)e^{-x^2/4}$, whose $L^2(\mathbb{R})$ norm is $\approx 1.533$; the configuration is as specified in Table 4.1, the same for both variants except for the residual and the forward integrator. The errors are computed based on the relative discrete $L^2(\mathbb{R})$ norms defined on an equispaced grid $\{x_n\}_{n=1}^{N_{\text{eval}}}$, $N_{\text{eval}} = 800$, on $[-10, 10]$:

$$\varphi_{\text{rel}} := \frac{\|\widehat{\varphi} - \varphi^\dagger\|_\Delta}{\|\varphi^\dagger\|_\Delta}, \qquad v_{\text{rel}} := \frac{\|v_\theta(\cdot, 1) - v_{\text{ref}}\|_\Delta}{\|v_{\text{ref}}\|_\Delta}, \qquad \|f\|_\Delta^2 := \Delta x \sum_{n=1}^{N_{\text{eval}}} f(x_n)^2, \tag{4.81}$$

with $\Delta x = 20/(N_{\text{eval}} - 1)$ and $v_{\text{ref}}$ the perturbed finite difference field at $t = 1$ interpolated onto the same evaluation grid. Using Proposition 3.7, $c$ is eliminated in closed form at fixed $\theta$; the runs below train it jointly instead. This happens thanks to the strict convexity of the inner problem, and (4.93) serves as an indicator for that. The positions of the sensors are depicted in Figure 4.1.

**Table 4.1:** Deployment configuration shared by Algorithms 1 and 2: the model and discretization (top block); data, loss weighting, and schedule (bottom block).

| Component | Symbol | Value |
|---|---|---|
| State network $v_\theta$ | $D, W, \alpha_v$ | depth 7, width 192 (tanh); envelope rate $\alpha_v = 0.08$ |
| Hermite source $\varphi^c$ | $\sigma, N_\varphi, \ell$ | scale $\sigma = 2$, $N_\varphi = 64$ modes, penalty order $\ell = 2$ |
| Spatial test space $\mathbb{V}_K$ | $K$ | 32 scaled Hermite functions $\psi_k^\sigma$ |
| Temporal test space $\mathbb{Q}_{M_L}$ | $M_L, \tau$ | one constant + nine Gaussian bumps, width $\tau = 0.05$ (Equation (4.14)) |
| Weak-form quadrature | $N_{\text{GH}} \times N_{\text{GL}}, \sigma_q$ | $64 \times 64$ Gauss-Hermite/Legendre; node scale $\sigma_q = 2.5$ (Equation (4.15)) |
| Strong-form collocation | $N_{\text{coll}}, L_{\text{box}}$ | 8192 points drawn i.i.d. uniformly on $[-8, 8] \times (0, 1)$, redrawn at every loss evaluation (Remark 4.27) |
| Sensors (total/train/held-out) | $N_{\text{sensor}}, M_d$ | 600 / 540 / 60, uniform on $[-10, 10]$ |
| Loss weights ($\delta = 0$) | $w_{\text{m}}, \mu_{\text{coll}}, \mu_{\text{wf}}, w_{\text{b}}, \lambda$ | 500, 5, 1, 5, $10^{-9}$ (Equation (4.26)) |
| Adaptive weights ($\delta > 0$) | $\lambda(\delta)$, $w_{\text{m}}(\delta)$ | $\max(10^{-9}, 3{\times}10^{-4}\delta^{2/3})$; $500/(1 + (\delta/0.02)^2)$ |
| Anti-inverse-crime model | $\varepsilon$ | diffusivity $a = 1 + 10^{-3}e^{-x^2/4}$ on $[-30, 30]$ (Section 4.1.5) |
| Forward integrator | $N_x, N_t$ | $10^4 \times 5{\times}10^3$; Crank Nicolson (linear) / IMEX (nonlinear) |
| Optimisation | — | 5000+4000+25000 Adam, then $\leq$2000 L-BFGS; cosine $10^{-3}{\to}5{\times}10^{-5}$; grad-clip 1; seed 42 |

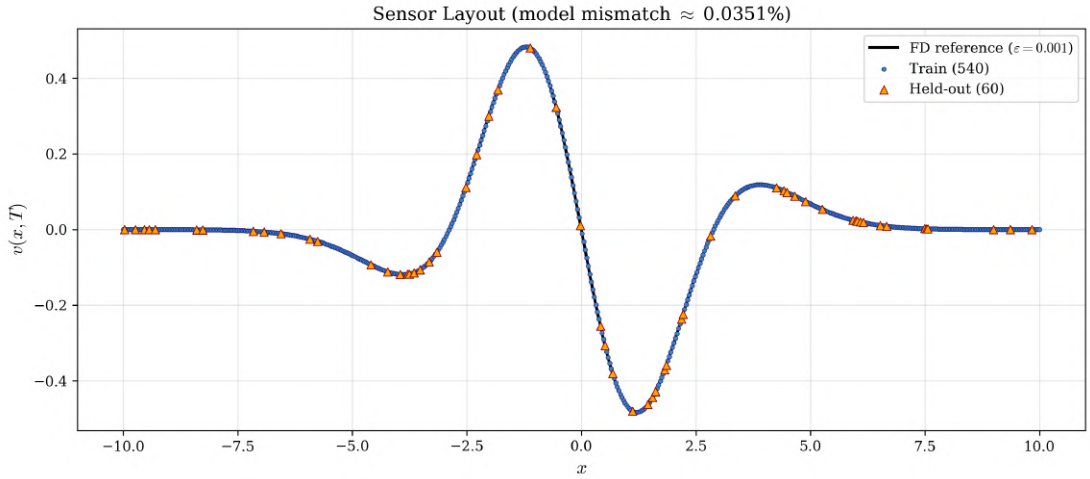


**Figure 4.1:** Sensor placement: 540 sensors for training (blue circles), and 60 sensors for validation (orange triangles) are considered. The aforementioned sensors are placed uniformly in $[-10, 10]$, with respect to the perturbed reference field $v(\cdot, 1)$ of the Crank Nicolson scheme, where $\varepsilon = 10^{-3}$. The set aside sensors never enter the loss; their misfit is reported as an out of sample check. The mismatch value of about 0.0351% is the operator mismatch floor of (4.8).

### *4.5.2 Forward Solver Validation*

However, in the case where the coefficients are constant, an explicit solution is possible: since $\varphi^\dagger = \partial_{xx}\big(x\, e^{-x^2/4}\big)$ and $x\, s^{-3/2} e^{-x^2/(4s)}$ solves the homogeneous equation,

$$v^\dagger(x, 1) = x\, 2^{-3/2} e^{-x^2/8} - x\, e^{-x^2/4} = A\varphi^\dagger(x). \tag{4.82}$$

In the case of Crank Nicolson (Section 1.4.2), the corresponding error in terms of $L^2(\mathbb{R})$ norm estimate takes a value of $2.57 \times 10^{-6}$. With $\varepsilon = 10^{-3}$, the discrepancy between the computed field and the actual solution reaches a value of $3.52 \times 10^{-4}$. In this specific instance, the error floor related to the inconsistency of operators, $\|(A_\varepsilon - A)\varphi^\dagger\|_{L^2(\mathbb{R})} / \|A\varphi^\dagger\|_{L^2(\mathbb{R})}$, measured between the two finite difference fields, is equal to $3.51 \times 10^{-4}$, or $0.35\,\varepsilon$, which is expected to be $\mathcal{O}(\varepsilon)$ as anticipated in Section 4.1.5, below the percent level (Figure 4.2).

The nonlinear equation cannot be solved analytically, and therefore the comparison between the perturbed and unperturbed cases is made using IMEX: $3.11 \times 10^{-4}$. The cubic reaction limits the effect of the Burgers equation on steepening; hence, the state remains bounded, $\max_x |v(x, 1)| \approx 0.461$, and no blow-up occurs.

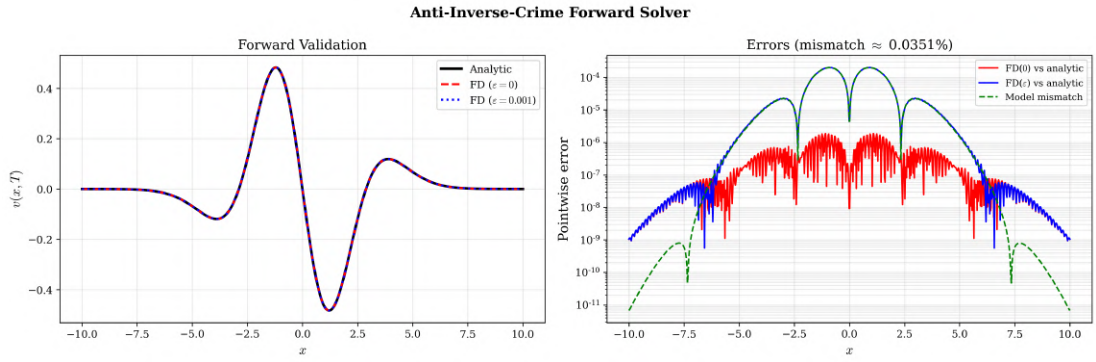


**Figure 4.2:** Forward solver of anti-inverse crime for the linear case. *Left:* The results obtained with Crank Nicolson scheme for $\varepsilon = 0$ and $\varepsilon = 10^{-3}$ compared to the closed form solution of (4.82). Visually they look the same. *Right:* Pointwise error between finite differences scheme with $\varepsilon = 0$ (red) and analytical solution is about $10^{-6}$, the solver's floor. Error between finite differences scheme with $\varepsilon > 0$ (blue) and analytical solution follows the model-mismatch curve (green), of roughly 0.0351%.

### *4.5.3 Noise Free Baselines*

The recovery performance for $\delta = 0$ is presented in Table 4.2, and for source and state in Figure 4.3.

**Table 4.2:** Noiseless baselines ($\delta = 0$) under the anti-inverse-crime setup of Section 4.1.5: linear algorithm (Algorithm 1) and nonlinear algorithm (Algorithm 2). Training time is 38.3 minutes and 39.0 minutes, correspondingly. In Phase 4, the line search accepts no step for either problem: the objective is left at $L = 8.47 \times 10^{-6}$ and $L = 6.40 \times 10^{-6}$ respectively, and $\varphi_{\text{rel}}$ is unchanged from the end of Phase 3. This represents a flaw with the schedule, not with convergence (see Remark 4.27). Both columns are from the same uncorrected schedule and thus may be directly compared to each other. There is no nonlinear equivalent for the corrected linear value $\varphi_{\text{rel}} = 1.47 \times 10^{-3}$ from (4.84), and the two may not be compared.

| Quantity | Linear | Nonlinear |
|---|---|---|
| Source error $\|\widehat{\varphi} - \varphi^\dagger\|_\Delta$ | $6.670 \times 10^{-3}$ | $5.160 \times 10^{-3}$ |
| Relative source error $\varphi_{\text{rel}}$ | $4.350 \times 10^{-3}$ | $3.370 \times 10^{-3}$ |
| Relative state error $v_{\text{rel}}$ | $2.090 \times 10^{-4}$ | $1.530 \times 10^{-4}$ |
| Operator-mismatch floor | $3.510 \times 10^{-4}$ | $3.110 \times 10^{-4}$ |
| Held-out misfit (MSE) | $1.800 \times 10^{-9}$ | $7.950 \times 10^{-10}$ |

If we take $\delta = 0$, then the relevant term in (4.45) is $\varepsilon$. Since $\ell/(\ell+2) = \frac{1}{2}$, Corollary 4.12 becomes

$$\|\widehat{\varphi} - \varphi^\dagger\|_{L^2(\mathbb{R})} \lesssim (2E)^{1/2}\varepsilon^{1/2}.$$

The radius will be calculated by $\widehat{\varphi^\dagger}(\xi) = 2\sqrt{2}\, i\, \xi^3 e^{-\xi^2}$:

$$E = \|\varphi^\dagger\|_{H^\ell(\mathbb{R})} = 4.452844, \qquad (2E)^{1/2} = 2.984240, \qquad \|A\varphi^\dagger\|_{L^2(\mathbb{R})} = 0.814852. \tag{4.83}$$

As mentioned before, the value of $3.51 \times 10^{-4}$ in Section 4.5.2 is a relative value; however, in the formula for (4.45), the absolute value is used. Hence $\varepsilon = 3.51 \times 10^{-4} \times 0.814852 = 2.86 \times 10^{-4}$ and $\varepsilon^{1/2} = 1.69 \times 10^{-2}$. The observed $6.67 \times 10^{-3}$ gives an empirical constant 0.394, 13.2% of

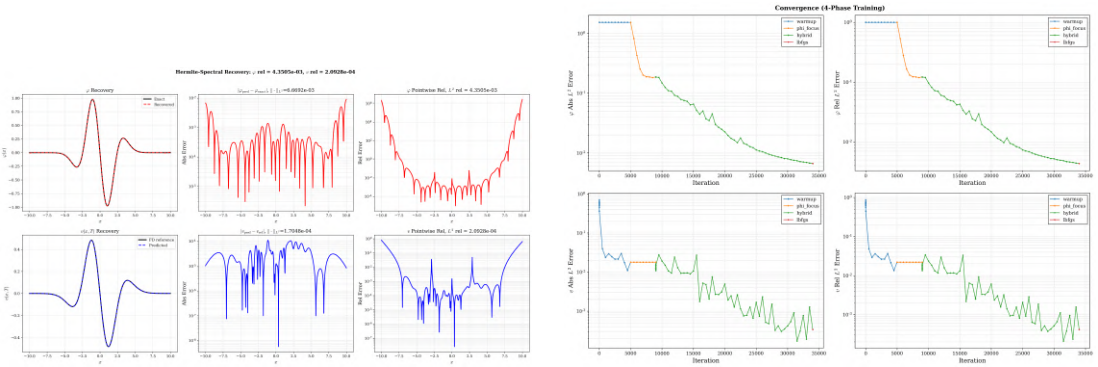

**Figure 4.3:** *Left:* Linear noiseless reconstruction. It represents the source reconstruction against $\varphi^\dagger$ (upper graph; absolute error $6.67\times10^{-3}$) and the terminal state versus the finite difference result (lower graph; relative error $2.09\times10^{-4}$). The error increases only in the far field regime $|x|\gtrsim 7$, where both are exponentially small. *Right:* In the four-phase schedule shown in Algorithm 1, the $L^2(\mathbb{R})$ error for both source (in top) and the state (in bottom) is shown in each of the phases. The $L^2(\mathbb{R})$ error for the source remains unchanged in the state only warm-up phase because the values of all coefficients remain fixed at zero in that phase, falls during `phi_focus` and `hybrid`, and is left unchanged by `lbfgs` which accepts no step (Remark 4.27).

the sharp prefactor 2.984. The nonlinear run, whose own absolute misfit is $\varepsilon = 2.45\times10^{-4}$, yields 0.330, or 11.1%, while the corrected schedule in Remark 4.27 produces 0.133, or 4.5%. All three have landed at different points in the interval, but the overall result remains unchanged. It was not surprising that they landed in the interval since $E$ represents a worst case for $\mathcal{W}_{\ell,E}$ compared to $\varphi^\dagger$, which is smooth (see Section 4.5.6). Held-out misfits of order $10^{-9}$ place the data fit far below the floor value, implying that the residual source error is a result of model misspecification and not optimization. The relative state error $v_{\text{rel}}$ dips below the floor value as a result of the fact that the state is estimated based on the perturbed observations, the floor bounding recovery of the *source* through the ill-posed inverse.

The nonlinearity does not cost anything because the two columns are nearly the same with the nonlinear being marginally better. It is a good comparison between apples to apples because the same scheduling scheme and cut-off Phase 4 apply. The Burgers-Allen-Cahn source is found to have linear accuracy whereas this is not possible using Fourier multiplier inversion at all [65]. One property of the spectrum is that $\sigma$ is equal to 2, whereas the natural value is $\sqrt{2}$, at which $\varphi^\dagger$ consists of only two Hermite terms (Remark 4.1). This means the expansion with $N_\varphi = 64$ only approximates the source: the coefficients decay over the first thirty modes and then flatten into a numerical floor rather than continuing to zero (Figure 4.4), consistent with the numerical rank 61/64 of $\mathbf{G}$ reported at (4.94). The pair satisfies the equation and not only the terminal data (Figure 4.5).

> **Remark 4.27.** For the baselines, the application of Algorithm 1 is done precisely as described, with the strong-form term re-sampling its collocations each time it is called. It does not impact on Adam, but the full batch quasi-Newton optimization algorithm will face an issue. The objective function varies from one evaluation to another in the line search phase, which means that the strong Wolfe conditions cannot validate the step taken. Phase 4 therefore terminates without an accepted step, leaving the values of Table 4.2. In

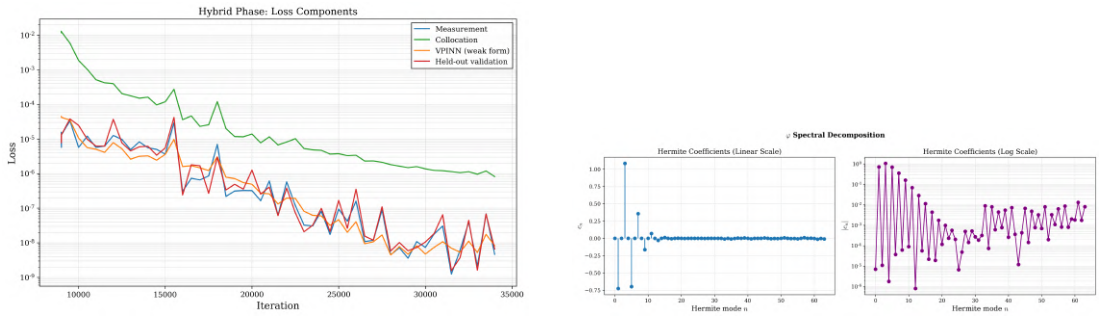


**Figure 4.4:** *Left:* Phase fitting loss terms: data fitting, strong form collocation, weak form residual and out-of-sample fitting, shown as a function of the number of iterations. The out-of-sample fitting loss term is equal to the training fitting loss term at $10^{-9}$; this implies a very good data fitting. *Right:* Hermite coefficients $c_n$ reconstruction for $\sigma = 2$ with respect to both linear and logarithmic axis. Here, we have mass allocation only for odd $n$, because of the oddness of the source itself, starting from $n = 1$ and going up to $n = 13$, while all even $n$ carry no significant mass (4-6 orders of magnitude lower). As soon as $n$ gets equal to or greater than 32, the number of spatial test functions, the alternation ends up getting absent, and for both parities, these numbers lie in the range from $10^{-3}$ to $10^{-2}$. In simple words, these modes cannot be recovered using the weak form and are held only by $\lambda = 10^{-9}$; and hence, truncation is governed by conditioning rather than by $\sigma = 2 \neq \sqrt{2}$ alone.

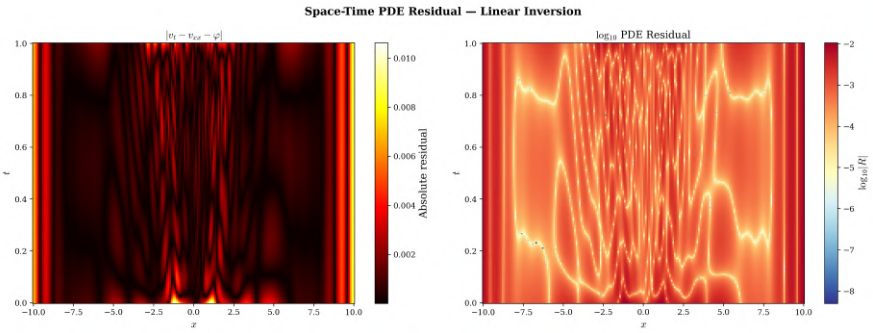


**Figure 4.5:** Space-time PDE residual. $|\partial_t v - \partial_{xx} v - \widehat{\varphi}|$ for the converged linear reconstruction on the left and its log (base 10) on the right. The residual in the interior is around 0.001, but close to $t = 0$ and the far-field truncation, the residual jumps to 0.01, those being the regions least constrained by the terminal data.

the entire period of phase four, freezing one sample according to the L-BFGS approach [99] lets it run to 2081 evaluations, with the objective declining monotonically and $\varphi_{\text{rel}}$ reaching its best value at iterate 1800:

$$\varphi_{\text{rel}}: \ 3.16 \times 10^{-3} \to 1.47 \times 10^{-3}, \qquad L: \ 3.00 \times 10^{-5} \to 2.21 \times 10^{-6}, \tag{4.84}$$

a factor 2.16, which results in an adjusted baseline that is 1.41 times the floor (4.86), not 4.18; both are lower bounds since the floor is an upper bound. This figure is the best Phase 4 iterate on $\varphi_{\text{rel}}$. Considering phase 4 of $\varphi_{\text{rel}}$, the final figure is $1.52 \times 10^{-3}$, which is off by 3.5%. At the budget, Adam's tail was dropping at $2.0 \times 10^{-2}$ per checkpoint, hence it has not converged either. Tables 4.2 and 4.3 precede the corrections and remain unchanged since they employ exponents that are dependent on ratios of the noise level rather than being affected by the improvement of the floor in a uniform manner, their levels referring to the uncorrected schedule. The two assertions in Section 4.5.6 make use

of data collected from a trained experiment, (4.93) and Proposition 4.25, and rely on the corrected schedule.

**Remark 4.28.** Iteration number 200 is when the refinement cycle stops for Table 4.3, and the ratio $\varphi_{\text{rel}}$ is still rising at that moment. In case of the starting configuration of noises, the ratio $\varphi_{\text{rel}}$ raises by 5.6% between iteration numbers 150 and 200 where $\delta$ is 0.008, and by 13.1% where $\delta$ is 0.040. For this same test trial at readings taken at 50, 100, 150, and 200, the values for the slopes prior to the removal of the floor are 0.354, 0.440, 0.487 and 0.530, and after its removal 1.209, 0.818, 0.794 and 0.822. Both conditions are parts of this particular stopping criterion, and the bigger $\delta$ is, the more there is of the drift, which means that with increased length of the refinement phase, one would increase the slope even more. This is an example of the noisy version of the scheduling problem discussed in Remark 4.27, where the cold starting solution stops short of convergence, as the sweep does here.

### 4.5.4 Empirical Hölder Stability Under Noise

For the noise $\varepsilon$ that appears in $(2E)^{2/(\ell+2)}(\delta+\varepsilon)^{\ell/(\ell+2)}$, this quantity does not go to zero as $\delta$ approaches zero, and thus the irreducible floor $F$ persists. This error measure is only concerned with errors greater than this threshold. There are four separate experiments performed that differentiate between the two: the sweep performed for $\sigma = 2$ (Table 4.3), and three at $\sigma = \sqrt{2}$, the scale at which the Hermite trial space is exact. Here, the floor is measured by means of Algorithm 3. Then, two matching scans are done: in one of them, the smooth source is left intact, hence $\varphi^\dagger \in \operatorname{span}\{\psi_1, \psi_3\}$ exactly, and in the other scan, the smooth source is replaced by one at the boundary of the class (Algorithm 4), which the same basis also represents without truncation; the only difference being the smoothness of the target.

With $\delta$ being fixed, the error term has a bias term $F$ and the noise is zero-mean. If the two are not correlated, then the sample average of the pattern will to leading order be

$$\varphi_{\text{rel}}(\delta) \;\approx\; \sqrt{F^2 + \left(\hat{C}\,\delta^{\,\hat{\alpha}}\right)^2}, \tag{4.85}$$

It follows that bias and noise add to the problem in quadrature. It would be better to get an estimate of $(\hat{C}, \hat{\alpha})$ from $F$ (which may be either free or measured) than to subtract a cold baseline and regress against $\log \delta$, which forces a slope whenever the baseline is under-resolved. The error present in Table 4.3 without correction is proportional to $\delta^{0.51}$, which is approximately equal to one-half, whereas the error with cold correction is raised to around 0.79 at the same stopping point (cf. Remarks 4.28 and 4.29).

This setup gives $F_{\sigma=2} = 4.35 \times 10^{-3}$ (Table 4.2). To make sure that this is indeed the limit of model mismatch and not an outcome of the optimization process, we repeat the same experiment without adding any noise on the natural scale, better conditioned and with a larger budget.

The probe converges to

$$F \;=\; 1.04 \times 10^{-3} \qquad (0.10\%), \tag{4.86}$$

**Table 4.3:** Linear reconstruction with additive noise at the deployed value of $\sigma = 2$. The mean and standard deviation are taken over five realizations of the noise, the sensor split and the collocation streams; the choice of weights was made using (4.28). The starting point for each is the noiseless reconstruction presented in Table 4.2, after which a short joint fine-tuning step follows using 200 Adam iterations at a one-tenth base learning rate; Phases 1-2 and the L-BFGS fine-tuning are not repeated. The noise vector is always rescaled in order to have a unit discrete $L^2(\mathbb{R})$ norm; hence, $\delta$ actually denotes the value of the noise norm, namely $\|\eta\|_{L^2(\mathbb{R})} = \delta$, and does not represent a part of $\|A\varphi^\dagger\|_{L^2(\mathbb{R})} = 0.814852$. Relative errors are expressed in units of $10^{-3}$. The source errors exhibit a pattern similar to $\varphi_{\mathrm{rel}} \sim \delta^{0.51}$ for this range; natural scale observations that incorporate the floor effect are presented in Table 4.4.

| $\delta$ | $\lambda(\delta)$ | $w_{\mathrm{m}}(\delta)$ | $\varphi_{\mathrm{rel}}$ $(\times 10^{-3})$ | $v_{\mathrm{rel}}$ $(\times 10^{-3})$ |
|---|---|---|---|---|
| 0.008 | $1.20 \times 10^{-5}$ | 431.0 | $8.31 \pm 0.87$ | $1.68 \pm 0.33$ |
| 0.012 | $1.57 \times 10^{-5}$ | 367.6 | $10.52 \pm 1.24$ | $2.46 \pm 0.49$ |
| 0.017 | $1.98 \times 10^{-5}$ | 290.3 | $12.74 \pm 1.58$ | $3.39 \pm 0.70$ |
| 0.022 | $2.35 \times 10^{-5}$ | 226.2 | $14.56 \pm 1.85$ | $4.27 \pm 0.91$ |
| 0.028 | $2.76 \times 10^{-5}$ | 168.9 | $16.25 \pm 2.04$ | $5.27 \pm 1.15$ |
| 0.033 | $3.08 \times 10^{-5}$ | 134.3 | $17.38 \pm 2.06$ | $6.05 \pm 1.33$ |
| 0.040 | $3.51 \times 10^{-5}$ | 100.0 | $18.73 \pm 2.02$ | $7.10 \pm 1.58$ |

**Algorithm 3** Floor probe: measurement of the noise-free floor $F$ directly.

**Changes made in comparison with Algorithm 1:**

1: The selection of the source scaling factor $\sigma = \sqrt{2}$ is obvious, because in this case, the function $\varphi^\dagger$ belongs precisely to the subspace generated by functions $\{\psi_1, \psi_3\}$, and the $N_\varphi = 64$ expansion is free from truncation bias.

2: The value of the Tikhonov regularization coefficient is raised from $\lambda = 10^{-9}$ to $\lambda = 2 \times 10^{-6}$ so that the threshold for the conditioning of the Gram matrix problem can be exceeded.

3: The optimization algorithm applies the four-stage process to cold start: Stage 3 goes up to 60,000 Adam iterations, and Stage 4 to 4,000 L-BFGS iterations.

4: This operation is carried out without applying any noise ($\delta = 0$), resulting in $F$ value of $\varphi_{\mathrm{rel}}$.

**The rest is to remain the same.**

One-fourth of the size of the deployed floor with leakage equal to $10^{-4}$ and spectrum localized on $\{\psi_1, \psi_3\}$: the deployed floor does not depend on the mismatch bound but on the regularization constant and the schedule length. The following are the two main reasons that justify why the result produced by (4.86) cannot be considered to be converged, but rather it is an upper bound, a feature shared by all other results based on it. As described in Remark 4.27, the resampling problem also emerges in Phase 4: the line search stops after just a few iterations, leading to an improvement in $\varphi_{\mathrm{rel}}$ of only 0.08%, and hence 4,000 iterations of the polishing procedure have not been fully utilized. Moreover, no plateau can be observed for Adam: in the previous 11 checkpoints, its value drops by 8.4% for every increase of 5,000 steps, as seen on the logarithmic regression trendline where the slope is $-1.75 \times 10^{-5}$ per step ($R^2 = 0.87$). The drop rate between 46,500 and 60,000 steps increases to 15.7% per 5,000 steps, while the error outside of sample is $1.3 \times 10^{-10}$. The optimization, not the information, bounds $F$ (Remark 4.30).

The smooth participant in the pair checks out eleven $\delta$ values ranging from $5 \times 10^{-4}$ to $4 \times 10^{-2}$ for $\sigma = \sqrt{2}$. The number of patterns checked out for each value is five. The run reaches about ten times lower than Table 4.3. Using (4.85) with $F$ being held constant to the

measured floor value (4.86) and per-level standard deviation weighting,

$$\hat{\alpha}_{\text{smooth}} \;=\; 0.677 \pm 0.039 \qquad (\chi^2/\text{dof} = 0.93). \tag{4.87}$$

In respect to the floor observed on the basis of which this analysis is made, Tikhonov's saturation of $L^2(\mathbb{R})$ is equal to $\frac{2}{3}$, and therefore it is located 0.3 standard errors below that value, while it is 4.5 standard errors better than $\frac{1}{2}$. This means that the smooth source is recovered faster than the worst case, hence the regularizer will have a qualification higher than that of the source class. The robustness comes from the strict inequality; the value is not robust (see Remark 4.30).

For $\delta = 5 \times 10^{-4}$, the sweep is at $2.51 \times 10^{-3}$; however, without noise, the short-schedule simulation value comes to $7.20 \times 10^{-3}$. Thus, it can be said that noise appears to improve the reconstruction threefold. Four out of eleven data points are below the baseline in the absence of noise and, on subtracting the baseline, we have $\hat{\alpha} = 1.32$. This cannot be a stability exponent at all: for any effective smoothness $s \geq 0$ one has $\hat{\alpha} = s/(s+2) < 1$, so no value above one is attainable on any Sobolev class. The worst-case exponent $\frac{1}{2}$ of Theorem 4.11 is the rate *uniform over* $\mathcal{W}_{\ell,E}$; individual smooth sources may exceed it, and (4.87) shows that this one does. For small $\delta$, it is more an optimization problem than a noise problem; Table 4.3 begins above the point, and the fits that take the floor into account are unaffected since $F$ is estimated, not imposed.

**Worst case source:** the exponent $\frac{1}{2}$ comes up for high-frequency sequences, because the smoothing operator communicates the least amount of information. To achieve such an exponent, one needs to have a very slowly falling spectrum.

**Algorithm 4** Worst case stress test - rough source of effective smoothness $s \approx 2$.

**Changes made in comparison with Algorithm 1:**

1: Modify the test source to $\varphi_{\text{rough}} = \sum_{j=0}^{63} c_j \, \psi_j^{\sqrt{2}}$, where $c_j \propto (2j+1)^{-3/2}$, normalised so that $\|\varphi_{\text{rough}}\|_{L^2(\mathbb{R})} = \|\varphi^\dagger\|_{L^2(\mathbb{R})} = 1.533$. As the decay is of power law nature, $\varphi_{\text{rough}} \in H^s(\mathbb{R})$ for all $s < 2$ but not for $s = 2$. The effective smoothness of the source considered is $s \approx 2$, which is the borderline case the worst-case estimate targets.

2: This scale is applied to the source using $\sigma = \sqrt{2}$ in both the data generation process and the recovery basis, to make sure that the spectral basis is an exact replica of $\varphi_{\text{rough}}$.

3: Perform the $\delta$ test for all eleven grid points of (4.87) using five realizations of noise for each $\delta$. The adaptive parameter is defined as $\lambda(\delta) = \max(2 \times 10^{-6}, 3 \times 10^{-4}\delta^{2/3})$. For each pairing, the tuning process should be started from the noise-free reconstruction.

**All other aspects of the algorithm, including architecture, sensor segmentation, and the five-term loss function, are unchanged as explained above. Each noisy point is a 1,000-step hybrid fine-tune at one tenth the base learning rate; Phases 1-2 and the L-BFGS polish are run only for the $\delta = 0$ baseline.**

The difficult case is that of the rough source. Since there is no obvious probe in this case, its floor comes from the sweep,

$$F_{\text{rough}} \;=\; (9.593 \pm 0.144) \times 10^{-3} \qquad (0.96\%), \tag{4.88}$$

The fitted floor is $(9.593 \pm 0.144) \times 10^{-3}$, as opposed to the noiseless run of the same sweep which is $8.957 \times 10^{-3}$; the fitted floor is the warm starting plateau, approximately 7% above the cold starting one. In both cases, again using the same scale and costing approach, it amounts

to about ten times the smooth floor. The floor obliterates the window, and hence the naked slope of 0.12 (for eleven levels) and 0.26 (for the six upper levels) gets overwhelmed by the floor. The free-floor quadrature gives

$$\hat{\alpha}_{\text{rough}} \;=\; 0.456 \pm 0.066 \qquad (\chi^2/\text{dof} = 0.13), \tag{4.89}$$

This number is 0.67 standard errors below 0.5, which means that it cannot be distinguished from 0.5 but is less than the smooth number from (4.87). For the case $\hat{\alpha} = s_{\text{eff}}/(s_{\text{eff}} + 2)$, we find $s_{\text{eff}} = 1.67$, consistent with the targeted $s \approx 2$.

**Table 4.4:** The smoothness condition implies a recovery condition, more precisely for $\sigma = \sqrt{2}$, with eleven noise levels and five patterns per noise level. The *floor* corresponds to the noise-free relative error, while the *raw* rate is contrasted with $\varphi_{\text{rel}}$. The *floor-aware* rate fits with (4.85). There are two consistent observations: first, there is a tenfold difference between the floors; second, $\hat{\alpha}_{\text{rough}} < \hat{\alpha}_{\text{smooth}}$.

| source | effective $s$ | floor $F$ | raw $\hat{\alpha}$ | floor-aware $\hat{\alpha}$ |
|---|---|---|---|---|
| smooth<br>$\text{span}\{\psi_1, \psi_3\}$ | $\to \infty$ | $0.10\%^{\dagger}$ | 0.60 | $0.677 \pm 0.039$ |
| rough<br>$c_j \propto (2j{+}1)^{-3/2}$ | $\approx 2$ | 0.96% | 0.12–0.26 | $0.456 \pm 0.066$ |

Refer to exponent: the value of $\ell/(\ell+2)$ is $\frac{1}{2}$ in the worst case scenario, whereas, at the point of saturation for $L^2(\mathbb{R})$ norms, the value is $\frac{2}{3}$.

$^{\dagger}$ shows that the value is fixed to the measured upper bound as stated in (4.86); the rough floor is fitted. See Remark 4.30.

In Theorem 4.11, there is a bound for the whole set of sources in terms of $\delta^{1/2}$, and not for each one individually. The smooth source is included in the bound ($\hat{\alpha} \approx 0.68$), while the rough source, built using the discretization of sequences used in the construction of the bound, is close to it ($\hat{\alpha} \approx 0.46$). The former achieves performance above the worst case, while the latter does not; thus lending proof that the claim is indeed sharp. The upper bound for $\hat{\alpha}$ depends on the $H^2(\mathbb{R})$ property of the first source and the smoothness of the second source; the floor gap factor of ten is because of the smoothing discussed in Section 4.3.2.

**Remark 4.29.** The two different values that have been obtained from the same experiment need to remain separate. The value of 0.79 obtained from the excess over floor approach (Section 5.3.1) is the result of the difference between cold and noisy experiments. Consequently, the artificial slope becomes steeper. The floor-corrected fit of (4.85) that uses the floor information provided by the probe should be used, and we obtain $\hat{\alpha}_{\text{smooth}} = 0.677 \pm 0.039$. The measurement done over the floor provides $0.7939 \pm 0.0481$ with $R^2 = 0.9820$. Both techniques converge on the classification of a gradient which is above the second-degree worst case value of $\frac{1}{2}$, and which agrees with the higher qualification of the penalty of order $\ell$. It is framed in such a way as to make Chapter 5 into a *relative* comparative analysis of the techniques rather than an absolute one. The scanning from the rough source, given in (4.89), provides us with an important benchmark in which the exponent depends on the level of smoothness, and our estimate for the exponent is $\hat{\alpha} = 0.456 \pm 0.066$ (alternatively, $s \approx 1.7$). The qualitative information above is the content of Theorem 4.11.

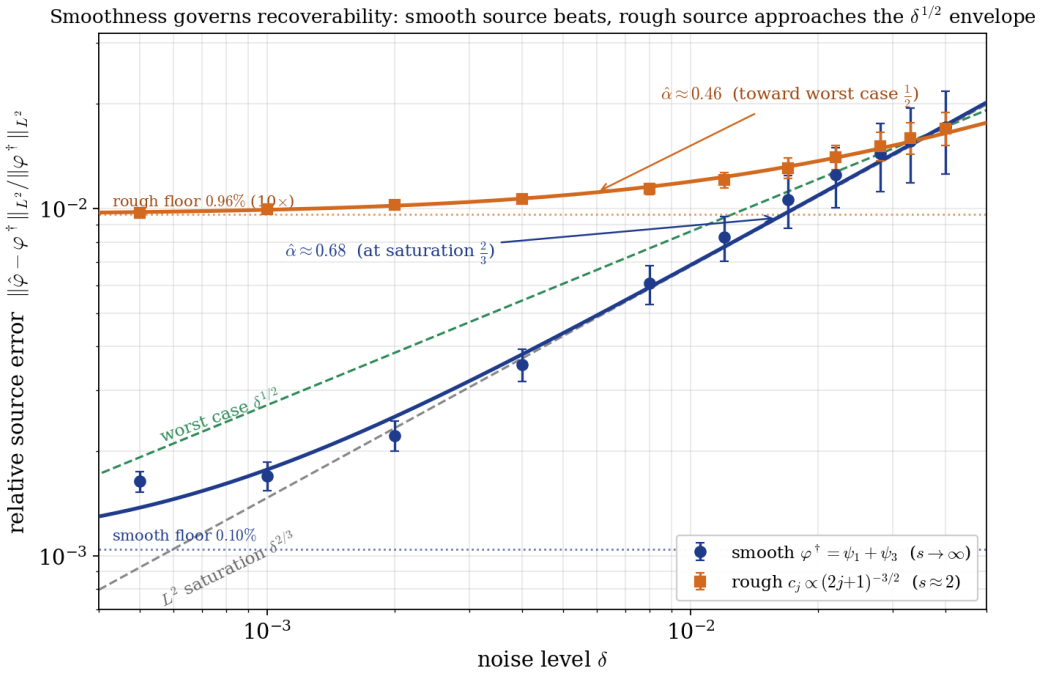


**Figure 4.6:** Comparison between smooth and rough reconstructions under noisy conditions with $\sigma = \sqrt{2}$. Error illustrated here is plotted on a log-log plot basis according to (4.85). In the case of the smooth source (i.e., floor of 0.10% that has been applied and sustained at its maximum possible value according to (4.86)), the degradation would be equal to $\hat{\alpha} = 0.677 \pm 0.039$, which is greater than $\frac{1}{2}$. While the rugged source is at a floor of 0.96%, that is, ten times higher and fitted, its value of $\hat{\alpha} = 0.456 \pm 0.066$ is undeniably closer to $\delta^{1/2}$. The dashed lines represent the slopes $\frac{1}{2}$ and $\frac{2}{3}$. The robust estimate is in the gap between floors of one order of magnitude and in the same order of the rate estimates but without being dependent on the values of the exponents (Remark 4.29); companion to Table 4.4.

**Remark 4.30.** The floor and the exponent of (4.85) become almost degenerate if the floor is predominant compared to the window. In case of a smooth source, with the floor fixed to its maximum value defined by (4.86), moving it from $0.60 \times 10^{-3}$ to $1.64 \times 10^{-3}$ moves the exponent monotonically:

| $F$ $(\times 10^{-3})$ | 0.60 | 0.80 | 1.044 | 1.30 | 1.64 |
|---|---|---|---|---|---|
| $\hat{\alpha}_{\text{smooth}}$ | 0.588 | 0.621 | 0.677 | 0.749 | 0.850 |
| s.e. | 0.033 | 0.035 | 0.039 | 0.046 | 0.060 |

While allowing $F$ to vary leads to $\hat{\alpha} = 0.799 \pm 0.069$ for $F = 1.47 \times 10^{-3}$. In the case of the rough source, we have very large degeneracy, and in terms of the fitted correlation, we obtain $+0.84$. By fixing $F$ at $8.00 \times 10^{-3}$, at the noise-free run $8.96 \times 10^{-3}$, at the fit value $9.59 \times 10^{-3}$, or even at the minimum error value $9.73 \times 10^{-3}$, we get $\hat{\alpha} = 0.164$, 0.276, 0.456, 0.515. Although an estimate based on the additive model as opposed to the quadrature model comes out to be $0.803 \pm 0.122$, three things can be concluded from this. The first thing is that all approaches will make sure that the estimated value of $\hat{\alpha}_{\text{smooth}}$ is greater than $\frac{1}{2}$ by 2.7 to 6 standard errors, so *that* inequality is robust. In fact, by the same process,

its location falls in the interval between 0.59 and 0.85, spanning the $L^2(\mathbb{R})$ saturation $\frac{2}{3}$. The value "at saturation" is obtained at the measured floor and would be invalid if a corrected probe produced a value of $F$ appreciably below $10^{-3}$, which Section 4.5.4 argues it would. In addition to this, the ten-to-one ratio between the floor numbers and the ranking $\hat{\alpha}_{\text{rough}} < \hat{\alpha}_{\text{smooth}}$ is unaffected by all of these choices; it is the combination of these factors, and not the third digit of either of the exponents, that underpins the sharpness argument in Theorem 4.11.

The selection of the fitting norm does not matter much: the difference between absolute and relative errors is a constant $\|\varphi^\dagger\|_{L^2(\mathbb{R})}$, and the same holds for the baselines. A multiplicative constant will affect the intercept but not the slope of a log-log fit. We find $\hat{\alpha} = 0.7939 \pm 0.0481$ with $R^2 = 0.9820$ when fitting Table 4.3 in the absolute norm; the prefactor varies from $1.99 \times 10^{-1}$ to $3.06 \times 10^{-1}$. Only the latter is comparable to $(2E)^{2/(\ell+2)}$, an actual absolute constant. The state error varies more drastically with the sweep since $v_{\text{rel}}$ increases by $\approx 4.2$ while $\varphi_{\text{rel}}$ increases by $\approx 2.3$. This is the signature of the ill-posed inverse, whose *source* error Theorem 4.11 controls.

### 4.5.5 Nonlinear Variant

The Burgers Allen Cahn equation of (4.6) is captured with near-linear accuracy, yielding $\varphi_{\text{rel}} = 3.37 \times 10^{-3}$ against $4.35 \times 10^{-3}$. Without an analytical solution for validation, accuracy is determined from the solver, the field extracted, and the PDE residual. Figure 4.7 shows the complete space-time evolution that was replicated, rather than just the end slice which was used for training, and it has an overall pointwise accuracy of $10^{-4}$.

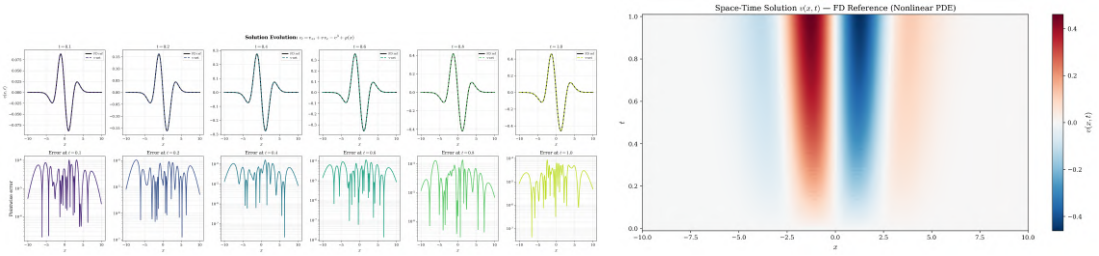


**Figure 4.7:** *(Left)* The nonlinear forward precision is shown using the graph of the recovered states (dashed line) versus the IMEX benchmark solution (solid line) for each $t \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$. The second figure below depicts the pointwise error, which stays at $10^{-4}$ for all $t$. This indicates that the neural net learned the complete dynamics of the Burgers-Allen-Cahn equation (4.6) from only one terminal datum. *(Right)* Reference space-time field. Cubic damping prevents divergence of the solution, and two lobes, positive at $x \approx -1.27$ and negative at $x \approx +1.27$ build toward $t = 1$, the instant when the snapshot was taken.

The term $-v^3$ reduces the steepness of the Burgers equation, which means that the solution is bounded ($\max_x |v(x,1)| \approx 0.46$). The field is odd, so it has the positive hump around $x = -1.27$ and the negative hump around $x = +1.27$. Both humps become more pronounced as the observation moment gets closer. Space-time residuals are about $10^{-3}$, localized around $t = 0$ and the boundaries of the domain, just like in the linear problem (Figure 4.5). Convection and cubic terms reach 10% to 20% of the diffusion and source terms (Figure 4.8), and therefore the inversion is clearly a nonlinear problem.

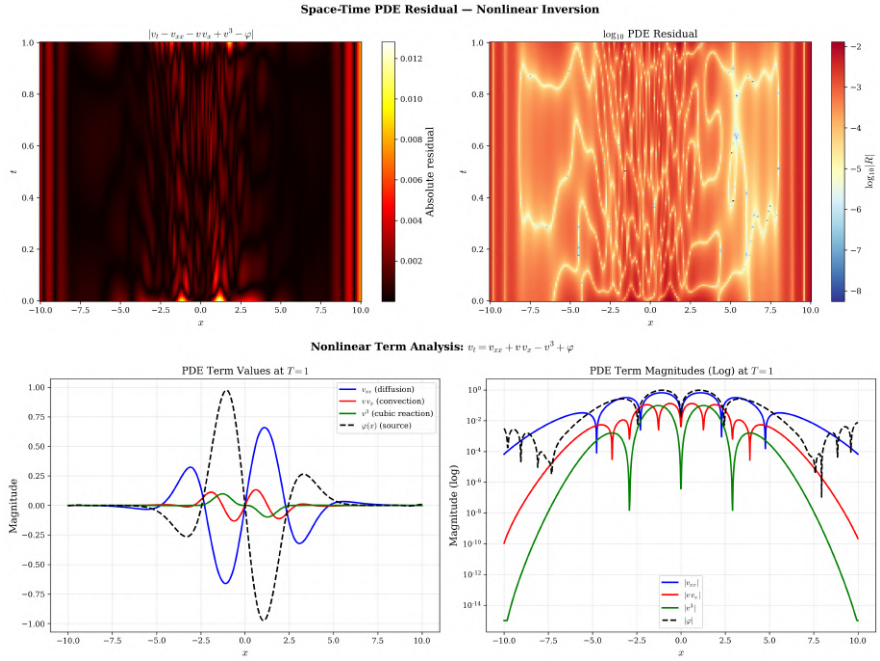


**Figure 4.8:** *(Top)* PDE residual in space-time $|\partial_t v - \partial_{xx} v - v\,\partial_x v + v^3 - \widehat{\varphi}|$ for the converged nonlinear reconstruction (absolute and $\log_{10}$ values). The residual has an order of $10^{-3}$ and it is localized around $t = 0$ and the boundaries of the domain, similarly to the linear PDE problem. *(Bottom)* Relative sizes of the terms in the PDE at time $t = 1$: diffusion ($v_{xx}$), advection ($v\,\partial_x v$), nonlinear reaction ($v^3$), and source ($\varphi$). The diffusion and source terms are by far the largest, whereas the other two reach 10 to 20 percent of the dominant terms non-trivial and representative of a truly nonlinear regime.

### *4.5.6 Verification of the Statements*

Since a rate can be confirmed by the slope of the value but not a constant, the analysis for each statement in Section 4.4 is done using only the *definitions*. There are three computer programs that help do this: one operates spectrally and does not train the network at all; one constructs Weyl windows according to Definition 4.15 using their Fourier-based definition; and another computes $\mathbf{G}$ and $r$ using the termination of a run. The linear optimization is done using the dual revised simplex method in HiGHS [15]. Scalar roots are computed using Brent's method, while integrals are done using adaptive quadrature. The symmetric-definite eigenproblem and weighted nonlinear least squares are also solved using LAPACK and MINPACK drivers from SciPy [14]. All the above techniques are summarized in Table 4.5; below we explain what is, and what is not, shown by each of them.

#### *The exact modulus*

The modulus is the moment problem (4.57) thanks to Lemma 4.18. Over a spectral grid, the moment problem can be posed as a linear program with one non-negative variable for each node, with the objective of maximizing the mass under two weighted integral constraints. The grid consists of 40 001 equally spaced grid points in the interval $[0, 6]$ and 20 001 geometrically spaced grid points in $[10^{-6}, 60]$ at a separation of $1.5 \times 10^{-4}$ about $\xi^\star$; neither the balance equation (4.54) nor the reduction of Section 4.4.3 enters Algorithm 5.

This can be attributed to two reasons. The value of LP lies below the closed form value

**Table 4.5:** Independent verification of the exact statements. Each computation receives the definitions alone; the closed forms are withheld and used only for comparison.

| Statement | What the computation is handed | Solver, resolution | Outcome |
|---|---|---|---|
| Theorem A (i) | (4.57) as a linear program, one nonnegative variable per spectral node; $\xi^\star$ and (4.58) withheld | HiGHS simplex, 60 002 nodes, spacing $1.5 \times 10^{-4}$ | seven digits or better at six noise levels, argmax on $\xi^\star$ (Table 4.6) |
| (4.61) | (4.58) and Lemma 4.7 (c) at 240 combinations of $(\ell, \text{radius}, \delta)$ | Brent, xtol $= 10^{-15}$ | Envelope respected at 240/240, and at 200/200 levels of a second sweep |
| Theorem 4.11, prefactor | Definition 4.15 in Fourier form; $C_n$ formed from three quadratures | Adaptive quadrature, $K$ to $10^7$ | $C_n \uparrow (2E)^{1/2}$ from below; deficit $\frac{4}{3}(\varepsilon_n/K_n)^2$ to six digits (4.90) |
| Theorem A (iii) | The same windows at $K = \xi^\star(\delta)$, $\delta = 0.05, 0.15, 0.40$, $\varepsilon_n : 0.5 \to 8 \times 10^{-4}$ | Adaptive quadrature | Data and source separations reach $\delta$ and $\omega(\delta, \mathcal{W}_{\ell,E})$ to seven digits, at second order in $\varepsilon_n$ |
| Proposition 4.5 | $m(\xi)(1+\xi^2)$ on a grid; the root of $e^t = 1+t+t^2$ withheld | $4\times10^6$ nodes, refined to $4 \times 10^5$ locally | 1.2984256075 at $\xi =$ 1.33913482 against $c_\sharp$ at $\xi =$ 1.33913484; infimum 1.0000000000 |
| Theorem 4.6 | Convolution kernel of $A$ truncated to $[-L, L]$, $L = 10, 20, 40, 80$ | LAPACK, $h = 0.05$, $n = 400 \ldots 3200$ | $L \times$ gap $\to 0.833$; pushforward law (4.91) to 2.50% at $L = 80$ |
| Theorem 4.22 (iii) | Both moment problems as separate programs, $t_j \in \{0.35, 0.7, 1\}$ | HiGHS simplex, 2 against 4 constraints | $3.42 \times 10^{-8}$ at $\delta = 0.05$, $3.76 \times 10^{-16}$ at $\delta = 0.80$ |
| Theorem B | Four filter families, each minimised over its own parameter against $\Omega(\delta)$ | Bounded Brent, xatol $= 10^{-12}$ | No entry below $\Omega$; the best exceeds it by $\leq 0.69\%$ (Table 4.7) |
| Theorem C (a) | $\mathbf{G}$, $r$ assembled from a completed run by the training routines themselves; (3.12) solved in closed form | LAPACK definite pencil, $N_\varphi = 64$ | $5.49 \times 10^{-3}$ (4.93); filter factors match to r.m.s. $8.0 \times 10^{-4}$ on the resolvable modes |
| Proposition 4.25 | Final loss at $N_{\mathrm{GH}} = N_{\mathrm{GL}} = 64, 128, 256$ on a fourfold collocation sample; trained $\|c\|_\Lambda$ | Gauss-Hermite $\times$ Gauss-Legendre | Loss stable to 0.01%, so $\vartheta = 1.551449 \times 10^{-3}$; 12.80 against the admissible 49.06 |

**Algorithm 5** Recovery of $\omega(\delta, \mathcal{W}_{\ell,E})$ from (4.57) by linear programming.

**Require:** noise level $\delta$, radius $2E$, penalty order $\ell$, spectral grid $\{\xi_i\}_{i=1}^N$ with $N = 60\,002$.
1: $P_i \leftarrow (1+\xi_i^2)^\ell$, $\quad Q_i \leftarrow m(\xi_i)^2$ ▷ $m$ by its series for $|\xi| < 10^{-4}$
2: Solve $p^\star = \max\{\sum_i \nu_i \ : \ \sum_i P_i\nu_i \leq (2E)^2,\ \sum_i Q_i\nu_i \leq \delta^2,\ \nu \geq 0\}$
3: $\widehat{\omega} \leftarrow \sqrt{p^\star}$, $\quad \hat{\xi} \leftarrow \xi_{\arg\max_i \nu_i}$
4: Record the number of loaded nodes $\#\{i : \nu_i > 10^{-9} \max_k \nu_k\}$
5: **return** $\widehat{\omega}$, $\hat{\xi}$, atom count ▷ compare against (4.58) only now

by a relative margin of $1.8 \times 10^{-10}$ to $4.4 \times 10^{-8}$: as a result of considering only nodes on the grid, the feasible region is smaller, and hence we get a lower bound, while the tangent line certificate in Proposition 4.17 is the upper bound. The values found by these two methods are approaching each other up to the eighth digit with a difference of the order of squared distance between grid lines. Secondly, the optimal loading contains precisely two nodes at each level. For two constraints, there is an optimal solution that is basic with at most two nonzero elements [16, §13]. Consequently, the precise figure is immaterial, but what is important is that each vertex lies within $6 \times 10^{-4}$ of $\xi^\star$, which amounts to a single atom divided by the discretization process, not two. If for some reason the certificate failed to succeed, then the algorithm would

**Table 4.6:** Recovery of $\omega(\delta, \mathcal{W}_{\ell,E})$ by linear programming, $E = 4.452844$, $\ell = 2$. "Atoms" counts grid nodes carrying mass above $10^{-9}$ of the maximum.

| $\delta$ | LP value | (4.58) | rel. diff. | LP argmax | $\xi^\star(\delta)$ | Atoms |
|---|---|---|---|---|---|---|
| 0.05 | 0.642765299 | 0.642765301 | $4.23 \times 10^{-9}$ | 3.58545 | 3.58543 | 2 |
| 0.15 | 1.083593354 | 1.083593401 | $4.37 \times 10^{-8}$ | 2.68650 | 2.68676 | 2 |
| 0.40 | 1.709360357 | 1.709360405 | $2.78 \times 10^{-8}$ | 2.05170 | 2.05182 | 2 |
| 1.00 | 2.628991069 | 2.628991118 | $1.86 \times 10^{-8}$ | 1.54530 | 1.54515 | 2 |
| 2.00 | 3.713692572 | 3.713692609 | $9.83 \times 10^{-9}$ | 1.18245 | 1.18240 | 2 |
| 4.00 | 5.421820300 | 5.421820301 | $1.79 \times 10^{-10}$ | 0.80160 | 0.80160 | 2 |

have yielded a value strictly greater by distributing mass at two distant frequencies. Hence, the single atom decomposition is accomplished as a result of the computation itself. The Hölder envelope (4.61), which was validated independently from (4.58) and Lemma 4.7 (c), still applies for all 240 possible combinations of $\ell \in \{1, 2, 4\}$, $R \in \{2, 2E\}$ and forty noise levels, and at all 200 levels of a second sweep spanning $10^{-4}$ to $1.99E$.

### *The sharp prefactor*

The validation of the constant is carried out by the object for which the realization of the constant is the purpose; the object is constructed according to Definition 4.15 in the form of a boxcar with height squared $(4\varepsilon_n)^{-1}$ on $[K_n - \varepsilon_n, K_n + \varepsilon_n] \cup [-K_n - \varepsilon_n, -K_n + \varepsilon_n]$, so that $\|w_n^{\mathrm{hi}}\|_{L^2(\mathbb{R})} = 1$ holds exactly by construction. The other two are adaptive quadratures based on the actual support, and without any assistance at all from the asymptotic properties of Lemma 4.16. With scaling as $\beta_n = 2E/\|w_n^{\mathrm{hi}}\|_{H^\ell(\mathbb{R})}$, we see that $\pm\frac{1}{2}\beta_n w_n^{\mathrm{hi}}$ is precisely on the boundary of $2\mathcal{W}_{\ell,E}$, and the empirical prefactor is

$$C_n = \frac{\|\varphi_1 - \varphi_2\|_{L^2(\mathbb{R})}}{\|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)}} = \frac{\beta_n}{\big(\beta_n \|A w_n^{\mathrm{hi}}\|_{L^2(\mathbb{R})}\big)^{1/2}}, \qquad \varepsilon_n = \min\big\{K_n^{1/2}, K_n/2.0001\big\}.$$

The sequence $\{\varepsilon_n\}$ so constructed has the property $\varepsilon_n \le K_n/2$ and $\lim_{n\to\infty} \varepsilon_n/K_n = 0$. Since any pair is admissible, every $C_n$ is a lower bound for the best possible constant in (4.44):

| $K_n$ | $2 \times 10^2$ | $10^3$ | $10^5$ | $10^7$ |
|---|---|---|---|---|
| $C_n$ | 2.9643807 | 2.9802624 | 2.9842001 | 2.9842395 |
| $1 - C_n/(2E)^{1/2}$ | $6.655 \times 10^{-3}$ | $1.333 \times 10^{-3}$ | $1.333 \times 10^{-5}$ | $1.333 \times 10^{-7}$ |
| $K_n\big(1 - C_n/(2E)^{1/2}\big)$ | 1.3310 | 1.3329 | 1.33333 | 1.33333 |

(4.90)

On the other hand, with respect to $(2E)^{1/2} = 2.984240$, the increase is strictly monotone and always below the desired value in the case of all $K_n$. This allows one to understand that the "approached but never attained" clause of Theorem A (ii) is made visible, so no smaller constant serves. The final line provides additional information: band averaging of $(1 + \xi^2)^\ell$ along with consideration of the fact that $m^2 \sim \xi^{-4}$ results in an underestimation of the amount of $\frac{4}{3}(\varepsilon_n/K_n)^2 + O(K_n^{-2})$. For $\varepsilon_n = K_n^{1/2}$, this amounts to $\frac{4}{3}K_n^{-1}$, and the measured product settles on $4/3$ to six digits. Fixing the aspect ratio constitutes a negative control, exhibiting its

typical plateau feature in that the defect freezes into the values of $3.328\times10^{-3}$, $1.324\times10^{-2}$, and $5.185\times10^{-2}$ for the ratios $\varepsilon_n/K_n = \frac{1}{20}, \frac{1}{10}, \frac{1}{5}$, flat in $K_n$ to five digits from $10^3$ to $10^7$. These are to be compared to the expected plateaus at $3.333\times10^{-3}$, $1.333\times10^{-2}$, and $5.333\times10^{-2}$; our hypothesis $\varepsilon_n/K_n \to 0$ is therefore doing work.

When there is a specific $\delta$, then it is precisely the construction that tests the attainment. In other words, if $K = \xi^\star(\delta)$, then for $\varepsilon_n = 0.5$ and decreasing down to $8\times10^{-4}$ at $\delta = 0.05, 0.15, 0.40$, then we have the data separation as follows: $5.000000\times10^{-2}$, $1.500000\times10^{-1}$, $4.000000\times10^{-1}$, and the source separation $6.427653\times10^{-1}$, 1.083593, 1.709360: seven digits compared to $\delta$, and compared to $\omega(\delta, \mathcal{W}_{\ell,E})$. Deficit decreases get close to 25 when we divide by five, which is of order $\varepsilon_n^2$ (Figure 4.9, left panel). The witness is the object in the theorem statement, thus Theorem A (iii) is checkable.

### *Operator norms and the spectrum*

Optimisation of a single function leads to the constants in Proposition 4.5. For $4\times10^6$ grid points in the range $[10^{-9}, 12]$, along with finer meshing with $4\times10^5$ grid points locally, the value of the quantity $\sup m(\xi)(1+\xi^2)$ is 1.2984256075 at $\xi = 1.33913482$, that is $t = 1.79328207$. Using the approach from Brent for the equation $e^t = 1+t+t^2$ one gets that $c_\sharp = 1.2984256075$ at $t = 1.79328213$, and hence they agree in all digits of double precision accuracy. On the other hand, in this grid we have the infimum 1.0000000000, which is Lemma 4.7 (c), and thus $\|A^{-1}\| = 1$ on the scale.

Theorem 4.6 does not directly apply since the discrete system is always compact and hence has a point spectrum, which Theorem 4.6 rules out. All we have left after the truncation is the spectral measure. Since the spectral frequency spacing in the domain $[-L, L]$ is $\pi/(2L)$,

$$\tfrac{1}{L}\,\#\{\lambda_k > \lambda\} \;\longrightarrow\; \tfrac{2}{\pi}\,m^{-1}(\lambda), \tag{4.91}$$

Szegő asymptotic theory for truncated convolution operators [12], applied to the operator $A$ that has a kernel given by $k(x) = \int_0^1 (4\pi s)^{-1/2}\exp(-x^2/(4s))\mathrm{d}s$. Upon discretizing the operator via its symmetric Toeplitz matrix using step size $h = 0.05$ and diagonalizing it for different values of $L$, namely $L = 10, 20, 40, 80$, the largest gap in $(0.05, 0.95)$ narrows from 0.07880 to 0.04071 to 0.02068 to 0.01042, while $L\times$ gap rises to 0.833. Gaps diminish like $1/L$ and there remains no isolated eigenvalue; whereas in case of a compact operator with fixed discrete spectrum, it would have exhibited flat behavior. Counting rule works for six levels, where the percentage value comes to 2.50% at $L = 80$. This happens in the case of $\lambda = 0.9$, which comes to 0.2875 as opposed to 0.294883. The function for counting proceeds in steps of $1/L = 0.0125$, which means that error equals six-tenths of an eigenvalue for such an experiment. The maximum eigenvalue increases from 0.98937804 to 1.00001861 in the process of approaching $\|A\| = 1$. Additional value of $1.9\times10^{-5}$ at $L = 80$ is a $O(h^2)$ Toeplitz error and not an attained supremum.

Part (iii) of Theorem 4.22 does not express an inequality but equality between two programs. In case of any error, there will be an obvious error in the problem formulation: there are only two constraints in the single snapshot problem, but in the multi-snapshot problem, there is the same $H^\ell(\mathbb{R})$ constraint together with one budget constraint $\delta$ per channel at times $t_j \in \{0.35, 0.7, 1\}$. Two programs match up to a tolerance of $3.42\times10^{-8}$ if $\delta = 0.05$, of $6.06\times10^{-9}$ if $\delta = 0.20$,

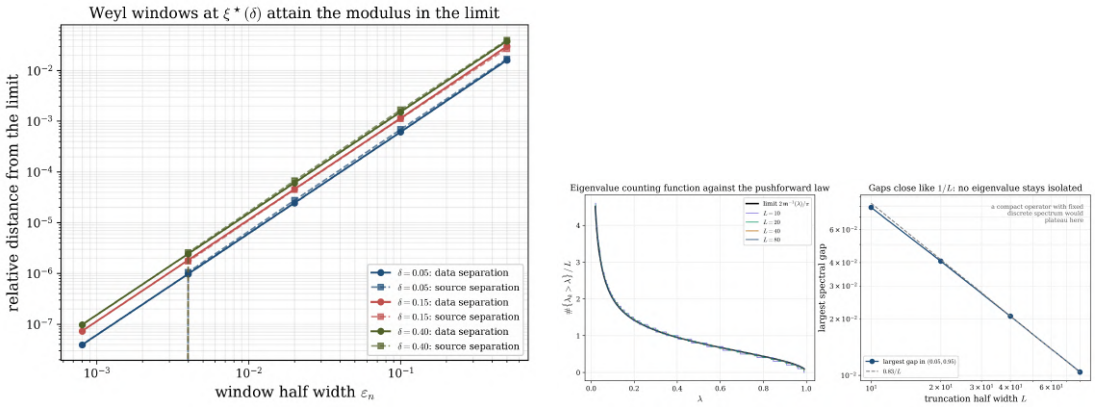


**Figure 4.9:** *Left:* Weyl windows at $\xi^\star(\delta)$; relative distance of the data and source separations from $\delta$ and $\omega(\delta, \mathcal{W}_{\ell,E})$ against the window half width. The common slope is 2, the second-order correction of Lemma 4.16. *Right:* the truncated spectrum against the pushforward law (4.91), and the largest spectral gap closing like $0.83/L$.

and of $3.76 \times 10^{-16}$ if $\delta = 0.80$, that is, to solver precision.

### *The minimax floor*

An infimum among all the maps cannot be attained through sampling, therefore we have to aim for Theorem B. There are four filter families, each one minimizing itself with respect to their parameter in the worst case scenario for linear mappings that is: $E \sup_\xi |1 - q(\xi)|(1 + \xi^2)^{-\ell/2} + \delta \sup_\xi q(\xi)/m(\xi)$, relative to $\Omega(\delta)$ in the centered moment problem with $E$ radius. All of them are no less than $\Omega$. The optimum one is even greater by 0.12% to 0.69%, thus keeping the floor in place. Two categories of reading can be distinguished between: the former corresponds to the minimal distance of the $L^2(\mathbb{R})$ Tikhonov regularization, whereas the latter corresponds to being 13.8% to 15.2% larger than $\Omega$, however it does not violate the attainment clause, since it consists of the summation of bias plus noise on the Cartesian product of both balls. The latter gives the upper bound of the worst-case scenario over the single-constraint set that defines $\Omega$, and such attainment was proved analytically in [106]. The only things that the table sets out here are falsification tests and measurements of near-attainment.

**Table 4.7:** Worst-case error of four filter families against $\Omega(\delta)$. Any entry below would falsify Theorem B; none is. Spectral cut-off [8] and Landweber [8], [17].

| $\delta$ | $\Omega(\delta)$ | Hilbert-scale | $L^2(\mathbb{R})$ Tikhonov | cut-off | Landweber | best/$\Omega$ |
|---|---|---|---|---|---|---|
| 0.05 | 0.4475391 | 0.5091102 | 0.4480722 | 0.8937472 | 0.4721188 | 1.0012 |
| 0.15 | 0.7481084 | 0.8521914 | 0.7498611 | 1.4920108 | 0.7937982 | 1.0023 |
| 0.40 | 1.1815738 | 1.3528904 | 1.1831388 | 2.3601716 | 1.2597551 | 1.0013 |
| 1.00 | 1.8568463 | 2.1399440 | 1.8696397 | 3.7114108 | 1.9887142 | 1.0069 |

### *Where the measurement sits*

On the other hand, when it comes to $\omega$, the data belong to the interval $[4.8\%, 5.2\%]$, and thus constitute an underestimation of the result, since $\omega$ corresponds to the most pessimistic of

all the cases considered on the $H^\ell(\mathbb{R})$ ball of radius $E$, and since $\varphi^\dagger$ is $C^\infty$ and is in the $H^s(\mathbb{R})$ ball of radius $E_s = \|\varphi^\dagger\|_{H^s(\mathbb{R})}$ for all $s$, each of which gives a valid bound. The sharpest is

$$\omega_\star(\delta) := \min_{s \geq \ell} \omega_s(\delta), \qquad \text{where} \qquad \omega_s(\delta) = \frac{2E_s}{\left(1 + \xi_s^{\star 2}\right)^{s/2}}, \tag{4.92}$$

Let $\xi_s^\star$ be the root of (4.54) having the index $s$ with the radius of $E_s$. Such a minimization problem involves some kind of trade-off, since the higher value of $s$ means the larger value of the power $s/(s+2)$, but the higher values of $s$ will lead to a radius growing from $E_2 = 4.45$ to $E_{16} = 1.29 \times 10^6$. As against $\omega_\star$, the measurements have the range from 11.6% to 20.1% in Table 4.3, while for the highly noisy cases, the measurements have the range from 4.5% to 11.2%, a factor of four closer. The chosen index decreases from $s^\star = 14$ in case of the minimum value to $s^\star = 8$ at $\delta = 0.15$. The sharpest upper bound obtained from the scan implies that the effective smoothness belongs to the range of values $[8, 14]$, thus validating the choice made in Section 5.3.5. Another component of the gap does not allow experimental approach since $\omega_s$ involves the supremum over a ball where an extremizer can be found within a small vicinity around $\xi_s^\star$, whereas $\varphi^\dagger$ is just one element of this ball. This is a structural property, which is precisely what we get in Theorem C (b), provided that the effective smoothness is high enough. Here, we do not try to show that this statement is sharp; sharpness is carried by the window terms and the boundary source.

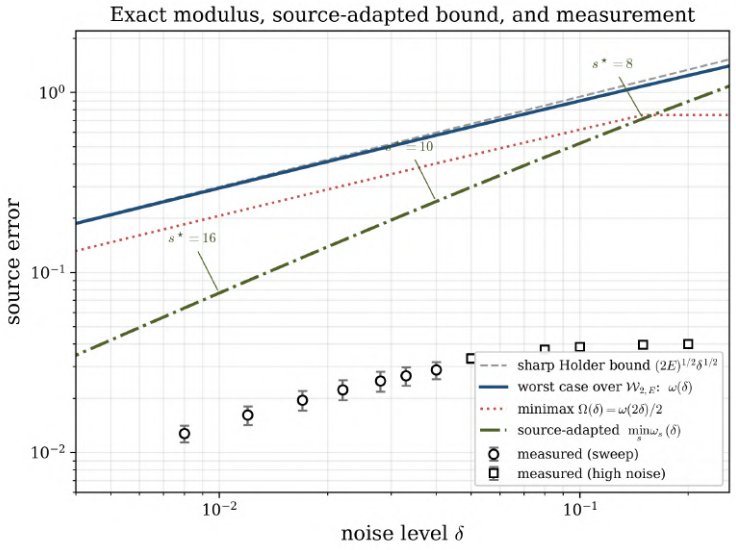


**Figure 4.10:** Exact modulus, source-adapted bound, and measurement. Solid: $\omega(\delta, \mathcal{W}_{\ell,E})$ from (4.58), the worst case over $\mathcal{W}_{\ell,E}$. Dashed: the sharp Hölder bound, approached from below. Dotted: $\Omega(\delta) = \frac{1}{2}\omega(2\delta, \mathcal{W}_{\ell,E})$ of Theorem B. Dash-dotted: the source-adapted bound (4.92), with the selected index $s^\star$ marked. Circles and squares: Table 4.3 and the high-noise extension of Section 5.3.7, error bars one standard deviation over five patterns. Both axes absolute; each noise pattern is normalised to unit discrete $L^2(\mathbb{R})$ norm, so $\|\eta\|_{L^2(\mathbb{R})} = \delta$.

### *Verification on the trained network*

Only one of these theorems needs training: Theorem C (a). It is also the only theorem which could possibly go wrong because of problems related to the training process. There are two

reasons why this theorem differs from its coded counterpart: optimization of the vector $c$ occurs at the same time as network training, and training produces only an approximate minimiser. In the current experiment, both $\mathbf{G}$ and $r$ are obtained using the whole training process; the matrix is thus the true $c$-Hessian of equation (4.18). The strong formulation implies blocking of the matrix $\Phi^\top\Phi/N_{\mathrm{coll}}$ with a fixed collocation sample, while in the case of weak form, one needs to block $B\otimes S$, where $S_{kj}=\sum_i w_i\psi_k(x_i)\psi_j^\sigma(x_i)$, and $B_m=\int_0^1 q_m$, because the losses are formed in this manner after the time integration is separated. The solution of equation (3.12), taken against the trained vector, gives:

$$\|c-c^\star\|_2/\|c\|_2 = 5.49\times 10^{-3}, \tag{4.93}$$

Thus, the concept of strict convexity introduced in Section 4.5.1 is transformed into an explicitly quantitative notion. Indeed, the pencil $(\mathbf{G},\Lambda)$ can be diagonalized, and thus the rationale is clear: the pencil resides in the range $3.09\times 10^{-1}$ to $10^{-20}$, so that

$$\mathrm{cond}(\mathbf{G}) = 6.4\times 10^{16}, \qquad \mathrm{cond}(\mathbf{G}+\lambda\Lambda) = 6.7\times 10^{4}\,: \tag{4.94}$$

Double precision results show that the numerical rank of $\mathbf{G}$ is 61/64, while the penalty term brings the matrix back to positive definiteness. The gain term of the order 12 gives a formula for the $\lambda\Lambda$ term of equation (4.17). Only then, if the projection $|u_k^\top r|$ is greater than the threshold value $\vartheta$, would such comparisons between the filter factors (4.74) become sensible. In regard to the six modes under consideration, the calculated value has the agreement within the root mean square error of $8.0\times 10^{-4}$ and has the maximum error of $1.9\times 10^{-3}$. Regarding the concept of "below threshold," the comparison of two zeros is meaningless because it only speaks to the limitation of the experiment and not the limitation of the theorem itself. Part (a) is simply an algebraic identity involving the exact minimiser, which requires no compactness.

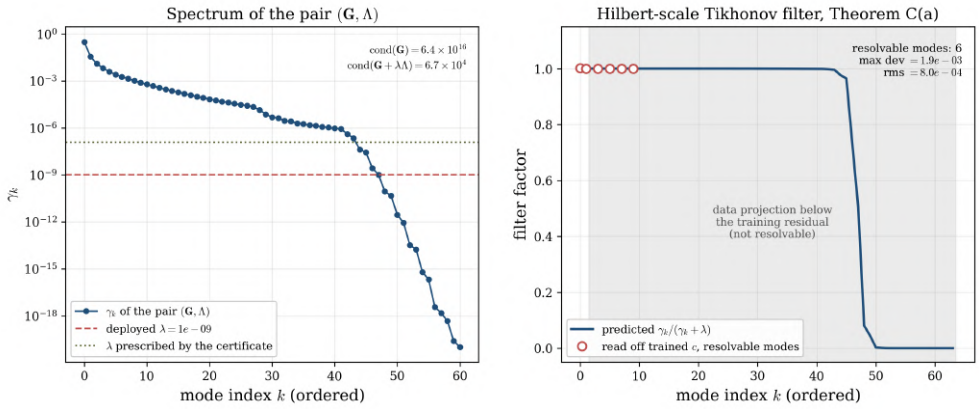


**Figure 4.11:** *Left:* spectrum of the pair $(\mathbf{G},\Lambda)$, with the deployed $\lambda$ and the value prescribed by Corollary 4.26 marked. *Right:* the filter factors (4.74), predicted against read off the trained coefficients; shaded are the modes whose data projection falls below the training residual $\vartheta$, where no comparison is resolvable.

The hypotheses of Proposition 4.25 are discharged from the log rather than assumed. For the convergence loss for $N_{\mathrm{GH}} = N_{\mathrm{GL}} = 64, 128, 256$, where four times collocation samples have been used, we have the following values 2.406993, 2.406752, and 2.406820, respectively, each multiplied by $10^{-6}$. The spread in this case is 0.01%, which shows that the process of

convergence of the quadrature has occurred. The value $\vartheta = 1.551449 \times 10^{-3}$ is read off without any initial guess for $\vartheta$. In case (ii), we get $\|c\|_\Lambda \le \vartheta\lambda^{-1/2} = 49.06$, whereas the value found from learning is 12.80. Therefore, the margin equals 3.83, not depending on the scaling of the objective. Concerning (i), as all losses are nonnegative, we may bound each loss from above by their sum; that lets us formulate the losses of sensors and collocation in terms of the norms, introduced in Definition 3.6, for $[-10, 10]$ and $[-8, 8] \times (0, 1)$ correspondingly.

$$\|v_\theta(\cdot,1)-g^\delta\|_{L^2(\mathbb{R})} \le 3.10\times10^{-4}, \quad \|R_{\mathrm{PDE}}\|_{L^2(\mathbb{R}\times(0,1))} \le 2.78\times10^{-3}, \quad \text{hence} \quad \|A\varphi^c-g^\delta\|_{L^2(\mathbb{R})} \le 3.09\times10^{-3} \tag{4.95}$$

With $\delta$ being 0.008, 0.015, and 0.04 and having a safety factor of 2.6, we can clear them according to the bound from the Duhamel inequality. Therefore, the results obtained fall within the certified rate stated in Corollary 4.26. When selecting the value of $C_{\sigma,\ell} = \sqrt{105}/8$, then this selection will guarantee 0.427, 0.585, and 0.955 against the observed range from $1.27 \times 10^{-2}$ to $2.87 \times 10^{-2}$, which is a factor of about thirty-three. The cost of this assurance consists in its uniformity across $\mathcal{W}_{\ell,E}$. Lambda may be estimated using the expression $\lambda = (\vartheta/E)^2 = 1.21 \times 10^{-7}$, which according to the corollary is two magnitudes greater than the current one. Both values are contained in the resolvable range indicated in the left graph of Figure 4.11, therefore the currently used parameter removes fewer modes than needed to satisfy the condition for the certificate. Hence, the guarantee turns out to be conservative with respect to the certificate condition.

### *The source at the boundary of the class*

As far as the estimation of this exponent of Theorem 4.11 is concerned, the important factor turns out to be the rough source sweep as given in Algorithm 4. Especially when we have $c_j$ proportional to $(2j+1)^{-3/2}$,

$$\|c\|_\Lambda^2 \;\propto\; \sum_j (2j+1)^{s-3} \quad \text{converges for every } s < 2, \quad \text{and diverges at } s = 2, \tag{4.96}$$

As mentioned above, this divergence is harmonic (Figure 4.12). The norms of the 64 modes employed have the following values for $s = 1$, 1.5, 1.9, and 2 respectively: 1.658, 1.891, 2.396, and 2.615.

The source sits at the exact edge of the class and, the balanced set of 64 modes, which belongs to the trial space where $\sigma = \sqrt{2}$, does not possess a representation floor. On the other hand, when trying to solve equation (4.85) through satisfying the unconstrained floor for all eleven levels (with five patterns each) weighted with the per-level standard deviation values,

$$\hat\alpha = 0.456 \pm 0.066, \quad F = (9.593 \pm 0.144) \times 10^{-3}, \quad C = (5.80 \pm 1.6) \times 10^{-2}, \quad \chi^2/\mathrm{dof} = 0.13. \tag{4.97}$$

There is a distance of 0.67 standard errors from the critical ratio $\ell/(\ell+2) = 1/2$, and it is situated below this value: the statistic is indistinguishable from 1/2. However, for the well-behaved source, we have $0.677 \pm 0.039$, that is 4.5 standard errors above 1/2. The first

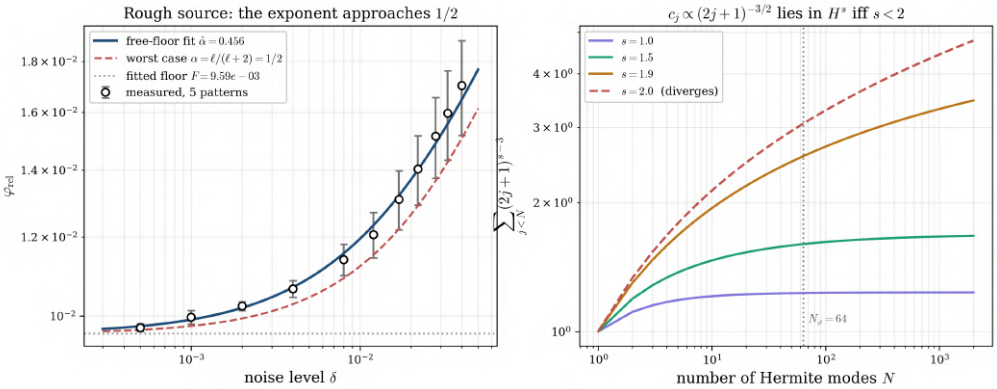


**Figure 4.12:** *Left:* the rough-source sweep, eleven noise levels with five patterns each, against the inverse-variance weighted free-floor fit $\sqrt{F^2 + (C\delta^{\hat{\alpha}})^2}$ of (4.85) and the worst-case slope $\ell/(\ell+2) = \frac{1}{2}$; error bars are the per-level standard deviations, which span a factor of twenty across the window and therefore carry the weighting. *Right:* partial sums of (4.96); every $s < 2$ curve converges while the $s = 2$ curve diverges harmonically, so the source sits on the boundary of the class of Theorem 4.11.

approach, which uses the assumption of an extremely smooth function, outperforms the worst case dramatically. However, for a case in which the function is at the borderline, then such a claim will be invalid. There is nothing that makes such two interpretations mutually exclusive, they are all related to Theorem C (b) as read in Remark 4.29. It should be noted that the exponent needs to be properly quoted as the relationship between the exponent and floor value becomes close to degenerate when the correlation between $F$ and $\hat{\alpha}$ is $+0.84$. The $+0.84$ correlation is discussed in Remark 4.30; the corresponding raw slopes, 0.124 over eleven levels and 0.255 over the top six, are those of Table 4.4. The main point here, however, comes from the floor effect evident in 0.96% compared to 0.10%, and the ordering of the two, rather than the third figure in either.

### *What has been established*

Four constants denied their closed form values can be expressed in numerical terms: the norm of $A^{-1}$ is equal to one, and the constant $c_\sharp$ is recovered to machine precision. The modulus of continuity, i.e., $\omega(\delta, \mathcal{W}_{\ell,E})$, has seven digits found using the upper bound and the lower bound approaches. And the Hölder coefficient $(2E)^{2/(\ell+2)}$, approached under the $\frac{4}{3}(\varepsilon_n/K_n)^2$ law. Collapse in multisnapshot is an identity to the accuracy of the solver, whereas the trained one satisfies the conditions of Theorem C (a) concerning Hilbert scale identification, where the error rate is about 0.55%, and both the hypotheses in Proposition 4.25 are discharged from the log. They are naturally delicate, for that reason they are said as follows: the theorem of spectral structure is correct only until the truncation precision since every finite section is compact. Thus, the final gap indicates, but it does not prove.

The result for Theorem B passes the falsifiability check and comes close to being attained, but one cannot obtain the infimum over all maps through sampling alone. For Theorem C (b), the prerequisites are met, although the rate still has to be stated in accordance with Theorem B. All other results are derivable from the symbol itself and do not require numerics.

# Chapter 5

# Comparison with Two Classical Methods

*The test of a method is not whether it succeeds where everything is easy, but where the data are few, the window is short, and the model is wrong.*

FOLKLORE OF INVERSE PROBLEMS

*In this chapter, our algorithm called Hermite Variational PINN is compared to two other classical approaches to tackling the problem of inverse PDEs. One approach involves the use of mollification with a Dirichlet kernel (Section 5.1) where the frequencies that are greater than the cutoff frequency are completely filtered out. Since this approach has just one adjustable parameter, it becomes easy to strike a balance between accuracy and noise filtering, and we then set out how that parameter is chosen and the reconstruction computed on the sensor grid. The second technique is the adjoint-state method with a Sobolev Tikhonov regularization term (Section 5.2), which again uses the same Hermite basis and truncation scheme as ours, though at a different length scale. It needs only one forward solve and one backward solve per iteration to compute the gradient. Both approaches will use the same data set, with all of the settings discussed in Chapter 4 left unchanged. Seven comparisons between the three approaches follow (see Section 5.3): a noise sweep, a minimal number of sensors, partial domain observation, unknown diffusivity, multiple observation times, a non-smooth source, and extremely noisy data; four other tests examine just the mollifier, while two other tests check just the adjoint, in settings where the assumptions of the other comparator do not apply. In Section 5.4, there are two main statements made. Firstly, our algorithm easily extends to nonlinear problems, which is impossible for some classical methods such as mollification with a Dirichlet kernel; also our approach works pretty well wherever the data are not clean, dense and uniform and at high noise levels. Secondly, the adjoint performs optimally when the data is very well cleaned up and plentiful, and it also takes significantly less time. The comparison experiments are shown and discussed in this chapter in detail.*

## 5.1 Mollification with a Dirichlet Kernel

### *5.1.1 Problem Formulation*

The first comparator operates in the Fourier domain and addresses the ill posedness in this domain. There is one parameter that needs to be set, i.e., the frequency threshold, making the bias variance tradeoff straightforward. As for Section 4.3, let us observe the forward operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $\varphi \mapsto v(\,\cdot\,, 1)$, where $v$ satisfies $\partial_t v - \partial_{xx} v = \varphi$, $v(\,\cdot\,, 0) = 0$. This map is a Fourier multiplier,

$$\widehat{A\varphi}(\xi) = m(\xi)\,\widehat{\varphi}(\xi), \qquad m(\xi) = \frac{1 - e^{-\xi^2}}{\xi^2}, \tag{5.1}$$

In the context of the Fourier transform given in Definition 1.9, the function $m$ is infinitely differentiable and strictly positive. Furthermore, $m(\xi)$ approaches 1 as $\xi$ approaches 0 and also $m(\xi)$ is similar to $\xi^{-2}$ when $|\xi|$ goes to infinity. The given problem is therefore ill posed of degree two (Corollary 4.9).

As the part of the formal inversion, we need to multiply our data by $1/m(\xi)$, which in turn is equal to $\xi^2/(1-e^{-\xi^2})$. In this case, $1/m(\xi)$ resembles $\xi^2$ for big $\xi$s, providing unbounded amplification of high frequency noise. The mollification technique addresses this difficulty through the smoothing process via the use of the Dirichlet kernel. It is equivalent to the spectral truncation of the data to a finite band of frequencies [112]. Our algorithm with the Dirichlet kernel for our instance is precisely that of [147], which will serve as our classical comparator subsequently. The Dirichlet kernel and its corresponding transform for radius $\mu > 0$ are expressed as follows:

$$D_{1/\mu}(x) = \frac{\sin(x/\mu)}{\pi x}, \tag{5.2}$$

$$\beta_\mu(\xi) = \widehat{D_{1/\mu}}(\xi) = \chi_{\{|\xi|<1/\mu\}}(\xi), \tag{5.3}$$

Concisely put, $\beta_\mu$ is the hard low pass filter which only allows the passage of frequencies less than the cut off value $1/\mu$. Consequently, the mollified inversion formula can be represented by

$$\widehat{\varphi}^\delta_\mu(\xi) = \frac{\xi^2}{1 - e^{-\xi^2}}\,\beta_\mu(\xi)\,\widehat{g^\delta}(\xi), \tag{5.4}$$

amplifying by the exact inverse symbol only those frequencies which are retained, while throwing away the rest of the frequencies above $1/\mu$. Because $1/m$ is bounded on the frequencies kept, (5.4) is bounded for any fixed $\mu > 0$. It follows then that there is just one parameter in the tuning process, balancing the loss of actual high frequency data, the bias, against amplification of the noise, the variance.

$\mu$ is a very important parameter in this case. In order to choose this parameter, there are two techniques available. One of the techniques that can be used here is the a priori technique, where one balances both types of error through the noise level $\delta$ and an assumed regularity bound $E$,

$$\mu^\star_{\text{apri}} = \left(\frac{\delta}{E}\right)^{1/(\ell+2)}, \qquad \ell = 2,\ E = 5, \tag{5.5}$$

The exponent $1/(\ell+2)$ is the balance behind the rate $\delta^{\ell/(\ell+2)}$ of Theorem 4.11 [49, Thm. 2.1]. In the a posteriori strategy, Morozov's discrepancy principle [110] can be used. In this case, $\mu$

can take the maximum value possible, that is maximal smoothing, with the condition that the discarded energy still matches the noise level in accordance with a factor $\tau > 1$,

$$\mu^{\star}_{\text{Mor}} = \sup\Big\{\mu > 0 : \ \|(\beta_\mu - 1)\widehat{g^\delta}\|_{L^2(\mathbb{R})} \le \tau\,\delta\,\big(N_{\text{sens}}\,\Delta x_s\big)^{1/2}\Big\}, \qquad \tau = 1.5, \tag{5.6}$$

The Parseval weight depends upon the number of sensors, $N_{\text{sens}}$, and their spacing, $\Delta x_s$. The residual norm increases as $\mu$ increases since it would just get larger after eliminating more modes. Thus, (5.6) has a unique solution; it can be found through bisection in the range from $10^{-10}$ to 10.

### 5.1.2 Discrete Realisation

The relations that satisfy (5.4) are determined according to Morozov's criterion (5.6). Sampling of the terminal field is performed for uniformly distributed values of $x$,

$$x_i = -L_{\text{obs}} + (i-1)\,\Delta x_s, \qquad i = 1, \dots, N_{\text{sens}},$$

with $N_{\text{sens}} = 600$ and $L_{\text{obs}} = 10$. Let $\mathcal{F}_N$ denote the discrete Fourier transform on the angular grid defined by $\xi_k = 2\pi k/(N_{\text{sens}}\Delta x_s)$. Here, the coefficients will be scaled by $\Delta x_s$ such that $\widehat{g^\delta} = \Delta x_s\,\mathcal{F}_N g^\delta$. Therefore, we combine $\mathcal{F}_N$ with the trapezoidal weight $\Delta x_s$ to approximate the continuous Fourier transform of Definition 1.9. In this case, the discretized inverse symbol and cutoff vectors take the form

$$a_k = \begin{cases} \dfrac{\xi_k^2}{1 - e^{-\xi_k^2}}, & |\xi_k| \ge 10^{-10}, \\ 1, & |\xi_k| < 10^{-10}, \end{cases} \qquad (\beta_\mu)_k = \chi_{\{|\xi_k| < 1/\mu\}}, \tag{5.7}$$

In this case, the removable singularity is expressed as follows,

$$\lim_{\xi \to 0} \frac{1}{m(\xi)} = m(0)^{-1} = 1.$$

With respect to the mollified reconstruction on the sensor grid, the real vector is

$$\varphi_\mu^\delta = \frac{1}{\Delta x_s}\,\Re\,\mathcal{F}_N^{-1}\big[a \odot \beta_\mu \odot \widehat{g^\delta}\big] \in \mathbb{R}^{N_{\text{sens}}}, \tag{5.8}$$

This is element wise multiplication represented by $\odot$. The output of this operation is then resampled using linear interpolation at $N_{\text{eval}} = 800$ points in (4.81) and extended by zero outside $[-L_{\text{obs}}, L_{\text{obs}}]$.

In the case of the Morozov principle (5.6), our discarded energy will be evaluated as the sum over all $k$ of $(1 - (\beta_\mu)_k)|\widehat{g^\delta}_k|^2$, and the comparison is made with the square of the right hand side of (5.6).

For the state error that is commensurate with the metric in Section 3.3.3, we need to solve $\partial_t v - \partial_{xx} v = \varphi_\mu^\delta$ one time for $\varepsilon = 0$ and $\kappa = 1$ using the Crank Nicolson scheme (1.25). The problem is to solve in the interval $[-25, 25]$, with $N_x = 2000$ and $N_t = 1000$. This results in obtaining $v_\mu^\delta(\,\cdot\,, 1)$, that is to compare with the perturbed reference field on the evaluation grid. The result is to obtain the reconstruction that would be the oracle choice between the two, that

is

$$\varphi_\star^\delta = \underset{\mu\in\{\mu_{\text{apri}}^\star,\,\mu_{\text{Mor}}^\star\}}{\arg\min} \|\varphi_\mu^\delta - \varphi^\dagger\|_{L^2(\mathbb{R})}, \tag{5.9}$$

However, this requires that one know the true value $\varphi^\dagger$, which is not something we have access to, so it gives only a mildly optimistic error for the mollification. We use the same definition consistently and point out when it materially affects a comparison. Algorithm 6 demonstrates how the reconstruction of the function can be done from (5.7) up to (5.9), whereas Algorithm 7 performs the direct comparison.

**Algorithm 6** Mollification with a Dirichlet kernel (single reconstruction).

**Require:** sensors $\{(x_i, g_i^\delta)\}_{i=1}^{N_{\text{sens}}}$ on $[-L_{\text{obs}}, L_{\text{obs}}]$; noise level $\delta$; rule $\in \{\text{APRIORI}, \text{MOROZOV}\}$; bound $E = 5$, order $\ell = 2$, safety factor $\tau = 1.5$.

$\Delta x_s \leftarrow x_2 - x_1$; $\xi_k \leftarrow 2\pi\,\text{fftfreq}(N_{\text{sens}}, \Delta x_s)$; $\widehat{g^\delta} \leftarrow \Delta x_s\,\mathcal{F}_N g^\delta$.
**if** rule = APRIORI **then**
  $\mu \leftarrow (\delta/E)^{1/(\ell+2)}$ if $\delta > 0$, else $\mu \leftarrow 10^{-10}$. ▷ (5.5)
**else if** rule = MOROZOV **then**
  bisect $\mu \in [10^{-10}, 10]$ for the largest $\mu$ with $\sum_k (1-(\beta_\mu)_k)|\widehat{g^\delta}_k|^2 \le \tau^2\delta^2 N_{\text{sens}}\Delta x_s$. ▷ (5.6)
**end if**
form $a$ and $\beta_\mu$ by (5.7); $\varphi_\mu^\delta \leftarrow \Delta x_s^{-1}\,\Re\,\mathcal{F}_N^{-1}[a \odot \beta_\mu \odot \widehat{g^\delta}]$. ▷ (5.8)
interpolate onto $x_e$, extend by zero outside $[-L_{\text{obs}}, L_{\text{obs}}]$.
**return** $\varphi_\mu^\delta$ and $\mu$.

**Algorithm 7** Hermite V-PINN vs mollification.

**Require:** true source $\varphi^\dagger$ (or $\varphi^\triangle$); noise levels $\{\delta_j\}$; $P = 5$ noise patterns; seed 42.

Generate the perturbed terminal field by Crank Nicolson with $\varepsilon = 10^{-3}$ on $[-30, 30]$, $N_x = 10^4$, $N_t = 5000$ (Section 4.1.5); sample $g^{\text{clean}}$ at the $N_{\text{sens}} = 600$ sensors; record the model-mismatch floor against the $\varepsilon = 0$ field; draw $P$ unit-norm patterns $\{\eta^{(p)}\}$ (Equation (4.4)).
**Baselines ($\delta = 0$).**
Train the full four-phase V-PINN (Algorithm 1) on $g^{\text{clean}}$; cache the converged state as a warm start.
Run Algorithm 6 under both rules on $g^{\text{clean}}$; keep the oracle selection (5.9).
**Noise sweep.**
**for** each $\delta_j$ **do**
  Set $\lambda(\delta_j)$, $w_{\text{m}}(\delta_j)$ by the adaptive schedule (4.28).
  **for** $p = 1, \ldots, P$ **do**
    $g^\delta \leftarrow g^{\text{clean}} + \delta_j\,\eta^{(p)}$.
    **V-PINN:** warm-start from the cached baseline and fine-tune 200 Adam steps at $0.1\times$ the base rate.
    **Mollification:** Algorithm 6 under both rules; take the oracle selection; transport once ($\varepsilon = 0$) for the state error.
    record $\varphi_{\text{rel}}, v_{\text{rel}}$ for both.
  **end for**
  aggregate mean±std over the $P$ patterns.
**end for**
fit each method's empirical Hölder slope by regressing $\log(\varphi_{\text{rel}}(\delta) - c_0)$ on $\log\delta$, with $c_0$ the method's $\delta = 0$ floor (Equation (4.85)).
**return** all per-level errors, predictions, and slopes.

In terms of the V-PINN side of Algorithm 7, the same exact method from Chapter 4 is used,

that is the VNet with depth of 7 and width of 192, the same Hermite source with $N_\varphi = 64$ modes at scale $\sigma_{\text{vpinn}} = 2$, the five term loss (4.18), and the adaptive weights (4.28). Only the inversion under comparison changes. The noise free V-PINN baseline reproduces Table 4.2 to the digit ($\varphi_{\text{rel}} = 4.35 \times 10^{-3}$), confirming that the comparison harness and the verification harness of Section 4.5 exercise the identical method.

## 5.2 Adjoint-State Inversion with a Sobolev Tikhonov Penalty

### 5.2.1 Problem Formulation

The second comparison is against adjoint-state inversion with Sobolev Tikhonov regularization [102, 120]. This implies that the source term is parametrized. A least squares objective function is formulated which is then optimized through a quasi-Newton technique based on the gradient derived by the adjoint state. The current technique works on the same basis and the same truncation but with different length scales, therefore, the comparison focuses on the inversion algorithm itself and not on the representation (5.10). The source is expanded in the scaled Hermite functions $\psi_j^\sigma$ of Definition 1.13,

$$\varphi^c(x) = \sum_{j=0}^{N_\Psi - 1} c_j \, \psi_j^{\sigma_{\text{adj}}}(x) = \big(\Psi^{\sigma_{\text{adj}}} c\big)(x), \qquad N_\Psi = 64, \quad \sigma_{\text{adj}} = 2.5, \tag{5.10}$$

Using the vector $c \in \mathbb{R}^{N_\Psi}$ along with the basis $\Psi^{\sigma_{\text{adj}}}$ sampled on the grid, this is the same set and truncation as the one used for the deployed source in (4.11), but with different length scales ($\sigma_{\text{adj}} = 2.5$ as opposed to $\sigma_{\text{vpinn}} = 2$) and carrying no neural state.

Mesh points inside the $\Omega_{\text{adj}} = [-L_{\text{adj}}, L_{\text{adj}}]$, where $L_{\text{adj}} = 10$, are taken $N_x = 500$ with spacing $\Delta x$, and the number of time points is $N_t = 1000$, with $\Delta t = 1/(N_t - 1)$. Define $v_h^c$ as the solution to the heat equation with constant coefficients,

$$\partial_t v = \kappa \, \partial_{xx} v + \varphi^c,$$

where the boundary conditions are homogeneous and Dirichlet, computed using the Crank Nicolson scheme given in (1.25). Where the three-point Laplacian stencil is represented by the matrix $T = \text{tridiag}(1, -2, 1)$ and $r = \kappa \, \Delta t / (2 \, \Delta x^2)$, the calculation for one time step is as follows:

$$\underbrace{(I - r\,T)}_{A_{\text{impl}}} v_h^{c,\,n+1} = \underbrace{(I + r\,T)}_{A_{\text{expl}}} v_h^{c,\,n} + \Delta t \, \varphi^c, \qquad v_h^{c,\,n+1}\big|_{\partial\Omega_{\text{adj}}} = 0, \tag{5.11}$$

We denote $A_{\text{impl}} = I - rT$ and $A_{\text{expl}} = I + rT$. We define the initial value as $v_h^{c,0} = 0$. The discrete forward map is associated with the terminal time slice, namely

$$F_h(c) := v_h^c(\,\cdot\,, 1) = v_h^{c,\,N_t - 1},$$

and the map here is a linear map in $c$ that stores its entire history for computing the gradient.

Points of the terminal field are gathered at the raw sensor abscissae $x_i$ ($i = 1, \ldots, N_{\text{sens}}$)

through the linear interpolation operator $S$. Each row represents the weights associated with the two adjacent nodes bracketing the sampling point. The recovered coefficients minimize

$$J_\lambda(c) = \tfrac{1}{2}\,\Delta x\,\|S\,F_h(c) - g^\delta\|_2^2 + \tfrac{\lambda}{2}\,c^\top D\,c, \qquad D = \operatorname{diag}\big((1+j)^{2s}\big)_{j=0}^{N_\Psi-1}, \quad s = 1, \tag{5.12}$$

Let us define $\Delta x$ as the Parseval weight in the residual and $D$ as the diagonal Sobolev penalty on the high indexed coefficients. This means that this regularization is done on the Hilbert scale and its qualification goes beyond the $L^2(\mathbb{R})$-Tikhonov saturation [50]. Because the functional is quadratic, it will be strictly convex for all $\lambda > 0$, and hence L-BFGS-B [44] can minimize the functional starting with $c = 0$.

Gradient computation for the data term is accomplished with only a single backward pass, rather than using finite differences. We have the following expression,

$$r_h = S\,F_h(c) - g^\delta,$$

where $r_h$ is the residual at the terminal time, and the final time condition for $\eta_h$ is defined as $S^\top r_h$. As the diffusion operator is a self-adjoint operator, the recurrence relation will have the same structure as (5.11), just in reverse direction and without any source term.

$$A_{\text{impl}}\,\eta_h^{n-1} = A_{\text{expl}}\,\eta_h^n, \qquad \eta_h^{N_t-1} = S^\top r_h, \qquad \eta_h^n\big|_{\partial\Omega_{\text{adj}}} = 0. \tag{5.13}$$

Finding the gradient of the coefficient is done by applying the adjoint approach. In this case, the adjoint variable is computed by integrating over time by means of the trapezoidal rule and is further projected onto the basis, along with the exact differentiation of the penalty term,

$$\nabla_c J_\lambda = \Delta t\,\Delta x\left(\tfrac{1}{2}\eta_h^0 + \sum_{n=1}^{N_t-2}\eta_h^n + \tfrac{1}{2}\eta_h^{N_t-1}\right)^\top \Psi^{\sigma_{\text{adj}}} + \lambda\,D\,c, \tag{5.14}$$

The gradient is exact to the discretization, in that calculating the objective function and the gradient needs only a single solution of the forward problem (5.11) and a single solution of the backward problem (5.13). That is the reason why the adjoint technique is superior to finite differencing when it comes to $N_\Psi$ unknowns.

Similarly to our algorithm, we determine the value of the regularization parameter using the magnitude of the noise and adjust it independently, due to the different scaling of $J_\lambda$ as compared to that of the V-PINN loss,

$$\lambda_{\text{adj}}(\delta) = 10^{-5}\Big(\frac{\delta}{0.016}\Big)^{0.667}, \qquad \lambda_{\text{adj}}(0) = 10^{-6}, \tag{5.15}$$

The exponent 0.667, which is approximately two-thirds, agrees with the exponent used in the equation for adaptive weights (see (4.28)). The schedule is anchored at $\delta \approx 0.016$, from where $\lambda_{\text{adj}} \approx 10^{-5}$ at this point.

As a supplement, the non-smooth source applied in Experiment 5, i.e., the triangular function, is provided in

$$\varphi^{\triangle}(x) = \max\big(5 - |x|,\,0\big), \qquad x \in \mathbb{R}, \tag{5.16}$$

It is supported on $[-5, 5]$, having its corner at the origin. It does not belong to the Schwartz

class of $\varphi^\dagger$ (Remark 4.1), and this serves as an example where neither of the bases resolves the source exactly.

Evaluation of (5.12) for the sake of its minimization is achieved through repeated computations of the tuple $\big(J_\lambda(c), \nabla_c J_\lambda(c)\big)$. This computation process comprises exactly:

forward recursion based on (5.11), calculation of the residual vector $r_h = S\,F_h(c) - g^\delta$, computation of the cost functional value in the form $J_\lambda(c) = \frac{1}{2}\Delta x\|r_h\|_2^2 + \frac{\lambda}{2}c^\top Dc$, recursive back step based on (5.13) with initial point $S^\top r_h$, and application of the gradient contraction (5.14).

Since $F_h$ is linear in $c$ and $J_\lambda$ is strictly convex, the L-BFGS-B algorithm with the starting point $c^{(0)} = 0$ converges to the unique minimiser $c^\star$. The maximum number of iterations allowed is 500 and the gradient tolerance $10^{-10}$. The reconstructed source is obtained through the re-expansion $\varphi^{c^\star} = \Psi^{\sigma_{\text{adj}}} c^\star$, evaluated on the $N_{\text{eval}} = 800$-point grid of (4.81). As the Hermite basis is mesh-independent, it means that we do deal with an exact re-expansion of the same coefficients, not with interpolation of them. The terminal field $F_h(c^\star)$ is transported once and interpolated onto the same grid for the state error. Two specific experimental modifications are purely set-theoretic on the sensor index. Partial domain: the restriction applied on $S$ is that sensors $x_i$ are in $\Omega_{\text{obs}}$ prior to assembly. Unknown diffusivity: in (5.11), we set $\kappa = 1$.

Algorithm 8 states the adjoint solver in isolation, while Algorithm 9 gives the full head-to-head driver. Similarly to Section 5.1, the part corresponding to the V-PINN uses the exact same method of Chapter 4, and its noise-free baseline reproduces Table 4.2 to the digit.

**Algorithm 8** Adjoint-state inversion with a Sobolev Tikhonov penalty.

**Require:** sensors $\{(x_i, g_i^\delta)\}_{i=1}^{N_{\text{sens}}}$; penalty $\lambda$; diffusivity $\kappa$; modes $N_\Psi = 64$, scale $\sigma_{\text{adj}} = 2.5$; grid $N_x = 500$, $N_t = 1000$ on $[-10, 10]$; Sobolev order $s = 1$; optional window $\Omega_{\text{obs}}$.

$\Delta x \leftarrow 2L_{\text{adj}}/(N_x - 1)$, $\Delta t \leftarrow 1/(N_t - 1)$, $r \leftarrow \kappa\,\Delta t/(2\Delta x^2)$.
assemble $A_{\text{impl}} = I - rT$, $A_{\text{expl}} = I + rT$; $\Psi \leftarrow [\psi_j^{\sigma_{\text{adj}}}(x_{\text{grid}})]$; $D \leftarrow \text{diag}((1+j)^{2s})$.
**if** $\Omega_{\text{obs}}$ given **then** drop sensors outside $\Omega_{\text{obs}}$.
**end if**
build linear-interpolation operator $S$ at the (retained) sensors; $c \leftarrow 0$.
**function** $J$-AND-GRAD($c$)
    $\varphi \leftarrow \Psi c$; march (5.11) for $n = 0, \ldots, N_t - 2$ to get $v_h^c$; $v_T \leftarrow v_h^{c,\,N_t-1}$.
    $r_h \leftarrow S v_T - g^\delta$; $J \leftarrow \frac{1}{2}\Delta x\|r_h\|_2^2 + \frac{\lambda}{2}c^\top Dc$. ▷ (5.12)
    seed $\eta_h^{N_t-1} \leftarrow S^\top r_h$; march (5.13) backward for $\eta_h$.
    $g_{\text{data}} \leftarrow \Delta t\,\Delta x\,(\frac{1}{2}\eta_h^0 + \sum_{n=1}^{N_t-2}\eta_h^n + \frac{1}{2}\eta_h^{N_t-1})^\top \Psi$.
    **return** $J$ and $g_{\text{data}} + \lambda Dc$. ▷ (5.14)
**end function**
$c^\star \leftarrow$ L-BFGS-B($J$-AND-GRAD, $c$; maxiter 500, gtol $10^{-10}$).
**return** $\varphi^{c^\star} = \Psi^{\sigma_{\text{adj}}} c^\star$ re-expanded on the eval grid; transport once for the state.

**Algorithm 9** Hermite V-PINN vs adjoint Sobolev Tikhonov.

**Require:** true source $\varphi^{\dagger}$ (or $\varphi^{\triangle}$); noise levels $\{\delta_j\}$; $P = 5$ patterns; seed 42.

Generate the perturbed terminal field by Crank Nicolson with $\varepsilon = 10^{-3}$ on $[-30, 30]$, $N_x = 10^4$, $N_t = 5000$ (Section 4.1.5); sample $g^{\text{clean}}$ at the $N_{\text{sens}} = 600$ sensors; draw $P$ unit-norm patterns $\{\eta^{(p)}\}$.

**Baselines ($\delta = 0$).**

Train the full four-phase V-PINN (Algorithm 1) on $g^{\text{clean}}$; cache as a warm start.

Run Algorithm 8 with $\lambda = 10^{-6}$, $\kappa = 1$ on $g^{\text{clean}}$.

**Noise sweep.**

**for** each $\delta_j$ **do**

  set $\lambda(\delta_j), w_{\text{m}}(\delta_j)$ by (4.28); $\lambda_{\text{adj}}(\delta_j)$ by (5.15).

  **for** $p = 1, \dots, P$ **do**

    $g^{\delta} \leftarrow g^{\text{clean}} + \delta_j \, \eta^{(p)}$.

    **V-PINN:** warm-start and fine-tune 200 Adam steps at $0.1\times$ the base rate.

    **Adjoint:** Algorithm 8 with $\lambda_{\text{adj}}(\delta_j)$; transport for the state.

    record $\varphi_{\text{rel}}, v_{\text{rel}}$ for both.

  **end for**

  aggregate mean±std over the $P$ patterns.

**end for**

fit each method's empirical Hölder slope by regressing $\log(\varphi_{\text{rel}}(\delta) - c_0)$ on $\log \delta$ (Equation (4.85)).

**return** all per-level errors, predictions, slopes, and solve times.

## 5.3 Experiments

### 5.3.1 Noise Sweep

**Setup.** the noise level over the seven values $\delta \in \{0.008, 0.012, 0.017, 0.022, 0.028, 0.033, 0.040\}$. $N_{\text{sens}} = 600$ uniform sensors on $[-10, 10]$, the adaptive V-PINN schedule (4.28), $P = 5$ noise patterns. Purpose: measure the conditional stability of all three inversions and fit the empirical Hölder slope (4.85) of each.

**Results.** Table 5.1 contains results for both relative error norms across the seven level sweep. Using the common measure of excess over floor (to be read as a relative ranking, see also Remark 4.29) the obtained source error slopes are $\hat{\alpha} \approx 0.79$ ($R^2 \approx 0.98$) for the V-PINN, $\hat{\alpha} \approx 0.61$ ($R^2 = 0.78$) for mollification, and $\hat{\alpha} = 0.764$ ($R^2 = 0.9996$) for the adjoint.

**Verdict.** In terms of the source case, the adjoint leads through $\delta = 0.017$ and has an improvement factor of $2.9\times$ with regard to the accuracy at $\delta = 0$, while the V-PINN leads from $\delta = 0.022$ onward and achieves a gain of 16.5% at $\delta = 0.040$. The point of crossover marks the end of the floor advantage of the adjoint, not a bettering of the exponent. Finally, mollification comes last at six out of seven levels and beats the adjoint method only when $\delta = 0.040$, and is the only method which is not monotone in $\delta$, which explains its $R^2 = 0.78$. For the state variable, the V-PINN outperforms the other two techniques in all seven cases, by 33.5% to 43.8% for the adjoint method and by 22.4% to 45.3% for mollification. The results of the recovery process are given in Figure B.1 and Figure B.2, and the three fitted Hölder laws in Figure 4.6, Figure B.3 and Figure B.4.

**Table 5.1:** Experiment 0 (all methods): relative source and state errors across the seven-level noise sweep, mean over five patterns. Baselines ($\delta = 0$): V-PINN $\varphi_{\text{rel}} = 4.35 \times 10^{-3}$, $v_{\text{rel}} = 2.09 \times 10^{-4}$; mollification $\varphi_{\text{rel}} = 5.61 \times 10^{-3}$, $v_{\text{rel}} = 7.20 \times 10^{-4}$; adjoint $\varphi_{\text{rel}} = 1.48 \times 10^{-3}$, $v_{\text{rel}} = 4.14 \times 10^{-4}$. The adjoint penalty $\lambda_{\text{adj}}(\delta)$ rises with noise from $6.30 \times 10^{-6}$ at $\delta = 0.008$ through $8.25 \times 10^{-6}$, $1.04 \times 10^{-5}$, $1.24 \times 10^{-5}$, $1.45 \times 10^{-5}$, $1.62 \times 10^{-5}$ to $1.84 \times 10^{-5}$ at $\delta = 0.040$.

| | $\varphi_{\text{rel}}$ | | | $v_{\text{rel}}$ | | |
|---|---|---|---|---|---|---|
| $\delta$ | V-PINN | Mollif. | Adjoint | V-PINN | Mollif. | Adjoint |
| 0.008 | $8.31 \times 10^{-3}$ | $1.12 \times 10^{-2}$ | $7.55 \times 10^{-3}$ | $1.68 \times 10^{-3}$ | $3.07 \times 10^{-3}$ | $2.99 \times 10^{-3}$ |
| 0.012 | $1.05 \times 10^{-2}$ | $1.59 \times 10^{-2}$ | $9.92 \times 10^{-3}$ | $2.46 \times 10^{-3}$ | $3.53 \times 10^{-3}$ | $4.07 \times 10^{-3}$ |
| 0.017 | $1.27 \times 10^{-2}$ | $1.96 \times 10^{-2}$ | $1.26 \times 10^{-2}$ | $3.39 \times 10^{-3}$ | $4.83 \times 10^{-3}$ | $5.35 \times 10^{-3}$ |
| 0.022 | $1.46 \times 10^{-2}$ | $2.23 \times 10^{-2}$ | $1.49 \times 10^{-2}$ | $4.27 \times 10^{-3}$ | $6.31 \times 10^{-3}$ | $6.56 \times 10^{-3}$ |
| 0.028 | $1.62 \times 10^{-2}$ | $1.90 \times 10^{-2}$ | $1.75 \times 10^{-2}$ | $5.27 \times 10^{-3}$ | $6.79 \times 10^{-3}$ | $7.96 \times 10^{-3}$ |
| 0.033 | $1.74 \times 10^{-2}$ | $2.24 \times 10^{-2}$ | $1.96 \times 10^{-2}$ | $6.05 \times 10^{-3}$ | $8.00 \times 10^{-3}$ | $9.10 \times 10^{-3}$ |
| 0.040 | $1.87 \times 10^{-2}$ | $2.23 \times 10^{-2}$ | $2.24 \times 10^{-2}$ | $7.10 \times 10^{-3}$ | $9.42 \times 10^{-3}$ | $1.07 \times 10^{-2}$ |

### *5.3.2 Sparse Sensors*

**Setup.** the number of measurements $N_{\text{sens}} \in \{15, 30, 50, 100, 200, 300, 600\}$. measurement error magnitude $\delta = 0.015$, three noise patterns. Purpose: investigate the stability of our inversion method in conditions of poor data sampling, where a Fourier-based inversion and a grid-interpolating adjoint are most exposed.

**Results.** The absolute error of the source term and the solve times of the V-PINN and the adjoint depending on the number of sensors are presented in Table 5.2.

**Table 5.2:** Experiment 1 (all schemes): absolute source error $\varphi_{\text{abs}}$ with respect to the number of sensors for $\delta = 0.015$ (mean $\pm$ standard deviation for the V-PINN and the adjoint scheme over three patterns; mean for mollification), and mean solve time of the V-PINN and the adjoint.

| | | $\varphi_{\text{abs}}$ | | time (s) | |
|---|---|---|---|---|---|
| $N_{\text{sens}}$ | V-PINN | Mollif. | Adjoint | V-PINN | Adj. |
| 15 | $3.19 \times 10^{-2} \pm 1.10 \times 10^{-3}$ | $5.46 \times 10^{-1}$ | $3.09 \times 10^{-1} \pm 2.76 \times 10^{-3}$ | 16.0 | 11.1 |
| 30 | $3.18 \times 10^{-2} \pm 1.65 \times 10^{-3}$ | $1.49 \times 10^{-1}$ | $1.53 \times 10^{-1} \pm 1.25 \times 10^{-3}$ | 14.5 | 10.0 |
| 50 | $3.63 \times 10^{-2} \pm 3.11 \times 10^{-3}$ | $9.01 \times 10^{-2}$ | $1.15 \times 10^{-1} \pm 5.01 \times 10^{-3}$ | 18.0 | 11.8 |
| 100 | $2.67 \times 10^{-2} \pm 4.18 \times 10^{-3}$ | $5.86 \times 10^{-2}$ | $6.95 \times 10^{-2} \pm 6.04 \times 10^{-3}$ | 16.2 | 12.4 |
| 200 | $2.07 \times 10^{-2} \pm 2.64 \times 10^{-3}$ | $4.08 \times 10^{-2}$ | $3.73 \times 10^{-2} \pm 1.76 \times 10^{-3}$ | 16.4 | 14.7 |
| 300 | $1.87 \times 10^{-2} \pm 1.46 \times 10^{-3}$ | $2.90 \times 10^{-2}$ | $3.01 \times 10^{-2} \pm 3.62 \times 10^{-3}$ | 16.4 | 15.7 |
| 600 | $1.97 \times 10^{-2} \pm 1.63 \times 10^{-3}$ | $2.75 \times 10^{-2}$ | $1.88 \times 10^{-2} \pm 2.26 \times 10^{-3}$ | 16.3 | 17.3 |

**Verdict.** In the case of few data, the V-PINN beats both the other methods. For $N_{\text{sens}} = 15$, it delivers an output of $3.19 \times 10^{-2}$, whereas the mollifier delivers an ineffective result of $5.46 \times 10^{-1}$ (the FFT of the mollifier requires a dense, evenly-spaced grid that it no longer has) and the adjoint outputs $3.09 \times 10^{-1}$ (putting fifteen samples on a 500-node grid injects large error). The discrepancies diminish as more data is provided the mollifier only as $N_{\text{sens}} \to 600$, where the adjoint edges slightly ahead and both iterative techniques exhibit comparable solving times.

Figures B.5 and B.6 present the error frontiers for the source errors in relation to the number of sensors, while Figures B.7 and B.8 present the reconstructions at the sparsest setting $N_{\text{sens}} = 15$.

### 5.3.3 Partial Domain Observation

**Setup.** the observation half width $L_{\text{obs}} \in \{2, 4, 6, 8, 10\}$. The sensor locations range between $-L_{\text{obs}}$ and $+L_{\text{obs}}$, but the error ranges over the whole domain. Fixed: $\delta = 0.015$. The mollifier utilizes information contained within the window, while the adjoint does not use the data from the outside sensors when creating its normal operator. Purpose: to examine how well the source extrapolates in the tails.

**Results.** Table 5.3 reports both error norms against $L_{\text{obs}}$.

**Table 5.3:** Experiment 2 (all techniques): source absolute and relative errors vs observation range $L_{\text{obs}}$ at $\delta = 0.015$.

| | V-PINN | | Mollification | | Adjoint | |
|---|---|---|---|---|---|---|
| $L_{\text{obs}}$ | $\varphi_{\text{abs}}$ | $\varphi_{\text{rel}}$ | $\varphi_{\text{abs}}$ | $\varphi_{\text{rel}}$ | $\varphi_{\text{abs}}$ | $\varphi_{\text{rel}}$ |
| 2.0 | $2.04 \times 10^{-2}$ | $1.55 \times 10^{-2}$ | $1.43 \times 10^{0}$ | $9.27 \times 10^{-1}$ | $4.60 \times 10^{-1}$ | $2.82 \times 10^{-1}$ |
| 4.0 | $1.77 \times 10^{-2}$ | $1.14 \times 10^{-2}$ | $6.61 \times 10^{-1}$ | $4.30 \times 10^{-1}$ | $1.84 \times 10^{-2}$ | $1.00 \times 10^{-2}$ |
| 6.0 | $1.91 \times 10^{-2}$ | $1.24 \times 10^{-2}$ | $1.18 \times 10^{-1}$ | $7.90 \times 10^{-2}$ | $1.63 \times 10^{-2}$ | $8.78 \times 10^{-3}$ |
| 8.0 | $1.93 \times 10^{-2}$ | $1.16 \times 10^{-2}$ | $2.13 \times 10^{-2}$ | $1.56 \times 10^{-2}$ | $1.76 \times 10^{-2}$ | $9.95 \times 10^{-3}$ |
| 10.0 | $1.97 \times 10^{-2}$ | $1.20 \times 10^{-2}$ | $2.75 \times 10^{-2}$ | $1.71 \times 10^{-2}$ | $1.88 \times 10^{-2}$ | $1.20 \times 10^{-2}$ |

**Verdict.** Both the comparators work in opposite manner when using windows. The mollifier diverges to 1.43 at $L_{\text{obs}} = 2$; hence, a global spectral inversion cannot recover a source whose support it never observes. This becomes twice as close to the V-PINN only when the window spans the whole support. This clearly indicates that the adjoint is superior with its windowing: for observation half-widths $L_{\text{obs}}$ greater than or equal to 4, it is better than the V-PINN (for $L_{\text{obs}} = 6$, its performance is $8.78 \times 10^{-3}$ compared to $1.24 \times 10^{-2}$). For $L_{\text{obs}} = 2$ it breaks down (its performance is $2.82 \times 10^{-1}$), where the V-PINN's collocation prior still extrapolates. In fact, the V-PINN is mostly insensitive to the window, holding $\varphi_{\text{abs}} \approx 2 \times 10^{-2}$ throughout.

Figures B.9 and B.10 plot the error against the observation window, and Figures B.11 and B.12 the reconstructions at the most severe truncation $L_{\text{obs}} = 2$.

### 5.3.4 Unknown Diffusivity

**Setup.** Varied: the level of noise in the data $\delta \in \{0, 0.007, 0.015, 0.033, 0.040\}$. Fixed: the true value of the diffusion coefficient $\kappa_{\text{true}} = 1.05$ is unknown to all three models, but the V-PINN infers it together with the source, while the other two use $\kappa = 1$. Purpose: explore the price for modeling errors, and examine the protective effect of joint inference on the source when the diffusivity is wrong.

**Results.** Table 5.4 reports the error in the source term and the recovered $\hat{\kappa}$ across the sweep.

**Verdict.** This is somehow a disadvantage of our method, although the reason is the lack of identifiability of the experiment, not our optimizer. It is not possible to obtain the pair $(\kappa, \varphi)$ from one terminal snapshot alone: our network settles at $\widehat{\kappa} \approx 0.46$ and $\varphi_{\text{abs}} \approx 0.50$, while

**Table 5.4:** Experiment 3 (all methods): unknown diffusivity, $\kappa_{\text{true}} = 1.05$. In the V-PINN, the diffusivity is trained together with the source coefficients, while the comparators assume $\kappa = 1$.

| $\delta$ | V-PINN $\varphi_{\text{abs}}$ | V-PINN $\widehat{\kappa}$ | Mollif. $\varphi_{\text{abs}}$ | Adjoint $\varphi_{\text{abs}}$ |
|---|---|---|---|---|
| 0.000 | $4.96 \times 10^{-1}$ | 0.4630 | $4.71 \times 10^{-2}$ | $4.66 \times 10^{-2}$ |
| 0.007 | $5.02 \times 10^{-1}$ | 0.4591 | $4.98 \times 10^{-2}$ | $5.15 \times 10^{-2}$ |
| 0.015 | $5.05 \times 10^{-1}$ | 0.4581 | $5.53 \times 10^{-2}$ | $5.64 \times 10^{-2}$ |
| 0.033 | $5.09 \times 10^{-1}$ | 0.4581 | $6.31 \times 10^{-2}$ | $6.63 \times 10^{-2}$ |
| 0.040 | $5.11 \times 10^{-1}$ | 0.4582 | $6.31 \times 10^{-2}$ | $7.01 \times 10^{-2}$ |

each of the fixed-$\kappa$ comparators suffers only a minor bias of roughly 0.05-0.07 due to the wrong choice of constant. Neither of the traditional approaches makes any effort to retrieve both parameters, and hence neither has to pay for it; the flexibility of the V-PINN works against it absent a second time slice (Experiment 4).

Figures B.13 and B.14 show the $\kappa$ and source recoveries against each comparator.

### *5.3.5 Multiple Time Measurements*

**Setup.** Varied: the set of observation times, ranging from the terminal snapshot alone to three slices $\{0.3, 0.6, 1\}$. Fixed: $\delta = 0.015$, three patterns. Purpose: determine the value of additional temporal data for each method.

**Results.** Table 5.5 reports the absolute source error by time schedule.

**Table 5.5:** The relationship between absolute source error and the set of observation times for Experiment 4 (adjoint mean±std), at $\delta = 0.015$.

| Observation times | V-PINN $\varphi_{\text{abs}}$ | Mollif. $\varphi_{\text{abs}}$ | Adjoint $\varphi_{\text{abs}}$ |
|---|---|---|---|
| $t = 1$ | $1.97 \times 10^{-2}$ | $2.75 \times 10^{-2}$ | $1.88 \times 10^{-2} \pm 2.26 \times 10^{-3}$ |
| $t = 0.5,\ 1$ | $1.46 \times 10^{-2}$ | $2.75 \times 10^{-2}$ | $1.88 \times 10^{-2} \pm 2.26 \times 10^{-3}$ |
| $t = 0.3,\ 0.6,\ 1$ | $1.27 \times 10^{-2}$ | $2.75 \times 10^{-2}$ | $1.88 \times 10^{-2} \pm 2.26 \times 10^{-3}$ |

**Verdict.** These extra data points result in a better accuracy of the V-PINN by 35%: from 1.97 to $1.27 \times 10^{-2}$. These data cannot be utilized by either of the two comparators: the mollifier, since it is fixed on the terminal symbol and still retains $2.75 \times 10^{-2}$, and the adjoint, since it is based on $F_h(c) = v_h^c(\,\cdot\,, 1)$ and thus stays at $1.88 \times 10^{-2}$. Time-distributed data is one of the advantages that the variational methodology provides us structurally, and it now has an exact worst-case theory: Theorem 4.22 makes predictions of a factor reaching $(1/\sqrt{3})^{\hat{\alpha}}$, where $\hat{\alpha}$ is the realised exponent. The realised value is 0.68, as can be found from Section 4.5.4, and $(1/\sqrt{3})^{0.68} = 0.688$, compared to the measurement $1.27/1.97 \approx 0.645$. This difference falls within 6%, and the exponent was measured independently rather than fitted here. Taking the reverse of this ratio, one would get $\hat{\alpha} = 0.799$ and therefore $s \approx 8$, but then quoting this prediction at this value of $s$ would make the agreement look exact while assuming the answer

in advance, although $s \approx 8$ is not arbitrary because (4.92) selects $s^\star \in [8, 14]$ for this source over the noise range swept.

Figures B.15 and B.16 plot the error by time schedule against each comparator.

### 5.3.6 Non Smooth Source

**Setup.** Varied: the noise level $\delta$ from the set $\{0, 0.007, 0.015, 0.033, 0.040\}$; the source is substituted by the triangular source $\varphi^{\triangle}$ of (5.16). Fixed: $N_{\text{sens}} = 600$, three patterns. Purpose: to investigate the behaviour of the three methods in the case of a non-Schwartz-type source, which is what both spectral bases are built for.

**Results.** Table 5.6 presents the absolute source error for varying noise levels.

**Table 5.6:** Experiment 5 (all techniques): non-smooth triangular source. Baselines ($\delta = 0$): V-PINN $6.46 \times 10^{-2}$, mollification $6.99 \times 10^{-3}$, adjoint $8.85 \times 10^{-2}$.

| $\delta$ | V-PINN $\varphi_{\text{abs}}$ | Mollif. $\varphi_{\text{abs}}$ | Adjoint $\varphi_{\text{abs}}$ |
|---|---|---|---|
| 0.000 | $6.46 \times 10^{-2}$ | $6.99 \times 10^{-3}$ | $8.85 \times 10^{-2}$ |
| 0.007 | $7.58 \times 10^{-2}$ | $7.27 \times 10^{-2}$ | $1.12 \times 10^{-1}$ |
| 0.015 | $7.92 \times 10^{-2}$ | $1.04 \times 10^{-1}$ | $1.24 \times 10^{-1}$ |
| 0.033 | $8.27 \times 10^{-2}$ | $1.14 \times 10^{-1}$ | $1.38 \times 10^{-1}$ |
| 0.040 | $8.33 \times 10^{-2}$ | $1.26 \times 10^{-1}$ | $1.42 \times 10^{-1}$ |

**Verdict.** However, when it comes to the corner, everything changes since at $\delta = 0$ the mollifier is superior to our smooth functions due to its abrupt cutoff, $6.99 \times 10^{-3}$ compared to our $6.46 \times 10^{-2}$, its one clean win. Noise then takes care of that, with the curves meeting close to $\delta = 0.007$. From here onwards the V-PINN overtakes the mollifier, and it beats the adjoint everywhere, where the latter's Sobolev penalty gobbles up its corner advantage.

Figure B.17 depicts the noise-free baselines, whereas Figure B.18 depicts the recovery of the triangle for various levels of noise, in which the mollifier's Gibbs oscillations are visible.

### 5.3.7 High Noise

**Setup.** Varied: the noise level over the high range, with $\delta$ taking the values $\{0.05, 0.08, 0.10, 0.15, 0.20\}$. Fixed: $N_{\text{sens}} = 600$, five patterns. Purpose: the saturation regime, and the high-noise Hölder slope.

**Results.** The absolute source error is presented in Table 5.7. The fitted high-noise values for the slopes are:

- V-PINN: $\hat{\alpha} = 0.160 \pm 0.034$, $R^2 = 0.878$;
- Mollification: $\hat{\alpha} = 0.749 \pm 0.239$, $R^2 = 0.765$;
- Adjoint: $\hat{\alpha} = 0.721 \pm 0.000$, $R^2 = 1.0000$.

**Table 5.7:** Experiment 6 (all methods): absolute source error in the high-noise regime, mean over five patterns (V-PINN and adjoint mean±std).

| $\delta$ | V-PINN $\varphi_{\text{abs}}$ | Mollif. $\varphi_{\text{abs}}$ | Adjoint $\varphi_{\text{abs}}$ |
|---|---|---|---|
| 0.05 | $3.33 \times 10^{-2} \pm 1.92 \times 10^{-3}$ | $4.72 \times 10^{-2}$ | $4.27 \times 10^{-2} \pm 6.44 \times 10^{-3}$ |
| 0.08 | $3.73 \times 10^{-2} \pm 1.79 \times 10^{-3}$ | $4.14 \times 10^{-2}$ | $5.91 \times 10^{-2} \pm 9.36 \times 10^{-3}$ |
| 0.10 | $3.85 \times 10^{-2} \pm 1.69 \times 10^{-3}$ | $4.99 \times 10^{-2}$ | $6.90 \times 10^{-2} \pm 1.11 \times 10^{-2}$ |
| 0.15 | $3.98 \times 10^{-2} \pm 1.30 \times 10^{-3}$ | $8.77 \times 10^{-2}$ | $9.16 \times 10^{-2} \pm 1.52 \times 10^{-2}$ |
| 0.20 | $4.01 \times 10^{-2} \pm 1.02 \times 10^{-3}$ | $1.00 \times 10^{-1}$ | $1.12 \times 10^{-1} \pm 1.89 \times 10^{-2}$ |

**Verdict.** V-PINN's error stabilizes in a range of 3.3 to $4.0 \times 10^{-2}$ due to the adaptive weight of the V-PINN which restricts the data term. Both the two other methods continue to increase, the mollifier goes up to $1.00 \times 10^{-1}$ and the adjoint up to $1.12 \times 10^{-1}$. The slope of the V-PINN error of 0.16 represents a relatively flat region. On the other hand, the slope of the mollifier is high and quite rough (0.75) indicating fast degradation, whereas the adjoint has almost an ideal Hilbert scale law ($R^2 = 1.0000$) seems good **But** unbounded. The key is not the slope but rather the level, which, for the V-PINN, remains considerably lower than either of the two comparators across the whole range.

Figure B.19 shows the high-noise recoveries, Figure B.20 the combined high-noise Hölder fit, and Figure B.21 the adjoint and mollifier Hölder-stability fits.

### *5.3.8 Experiments Unique to the Mollification Study*

The four cases that lack the adjoint counterpart are those where the assumptions of the Fourier inversion approach do not hold true: a model mismatch sweep, a sensor geometry case, a problem of coloured noise, and the data-efficiency frontier in the noise sensor plane. The V-PINN uses the same approach deployed everywhere else.

#### *Model Mismatch*

**Setup.** Varied: the disturbance $\varepsilon$ for the anti-inverse crime coefficient function $a(x) = 1+\varepsilon \exp(-x^2/4)$ as given by (4.7), where $\varepsilon \in \{0.001, 0.005, 0.010, 0.030, 0.050, 0.100\}$. Both procedures invert under $\varepsilon = 0$, but otherwise we consider the case of misspecification, and in that case we also compute the induced relative $L^2(\mathbb{R})$ mismatch of the terminal field. Fixed: $\delta = 0.015$, three patterns. Purpose: robustness to a misspecified operator.

**Results.** Table 5.8 gathers the row data.

**Verdict.** In the case of small $\varepsilon$, both methods absorb the discrepancy within the residual budget. In the case of larger $\varepsilon$, the V-PINN error increases at a slower pace compared to the mollifier error. Increasing $\varepsilon$ a hundred times results in an error that is doubled, whereas the mollifier triples it. The cut-off frequency of the mollifier is tuned according to the noise $\delta$ and does not identify any systematic mismatch. Despite the mismatch growing linearly with respect to $\varepsilon$ (around 35 times $\varepsilon$), the V-PINN can incorporate the perturbation into its weak-form residual and still degrade sub-linearly.

Figure B.22 plots the source error against the perturbation $\varepsilon$.

**Table 5.8:** Model-mismatch sweep at $\delta = 0.015$ (V-PINN vs mollification). "Mismatch" is the relative $L^2(\mathbb{R})$ norm of the difference between the data generated at $\varepsilon$ and the data that would have been generated at $\varepsilon = 0$.

| $\varepsilon$ | mismatch (%) | V-PINN $\varphi_{\text{rel}}$ | Mollif. $\varphi_{\text{rel}}$ |
|---|---|---|---|
| 0.001 | 0.035 | $1.20 \times 10^{-2}$ | $1.71 \times 10^{-2}$ |
| 0.005 | 0.175 | $1.31 \times 10^{-2}$ | $1.74 \times 10^{-2}$ |
| 0.010 | 0.349 | $1.39 \times 10^{-2}$ | $1.81 \times 10^{-2}$ |
| 0.030 | 1.040 | $1.86 \times 10^{-2}$ | $2.35 \times 10^{-2}$ |
| 0.050 | 1.719 | $2.15 \times 10^{-2}$ | $3.12 \times 10^{-2}$ |
| 0.100 | 3.371 | $2.47 \times 10^{-2}$ | $5.28 \times 10^{-2}$ |

### *Sensor Geometry*

**Setup.** Varied: the sensor layout for a fixed number of sensors, evenly spaced, one centre cluster, and one centre gap. Fixed: $\delta = 0.015$, three noise patterns. Purpose: to investigate how sensitive the methods are to the sampling geometry, isolated from sparsity.

**Results.** Table 5.9 reports both error norms by layout.

**Table 5.9:** Absolute and relative source errors versus sensor layout, at $\delta = 0.015$ (V-PINN vs mollification).

| | V-PINN | | Mollification | |
|---|---|---|---|---|
| Layout | $\varphi_{\text{abs}}$ | $\varphi_{\text{rel}}$ | $\varphi_{\text{abs}}$ | $\varphi_{\text{rel}}$ |
| Uniform | $1.97 \times 10^{-2}$ | $1.20 \times 10^{-2}$ | $2.75 \times 10^{-2}$ | $1.71 \times 10^{-2}$ |
| Centre-heavy | $1.67 \times 10^{-2}$ | $1.08 \times 10^{-2}$ | $8.06 \times 10^{-1}$ | $5.26 \times 10^{-1}$ |
| Centre gap | $2.92 \times 10^{-2}$ | $1.96 \times 10^{-2}$ | $2.72 \times 10^{0}$ | $1.77 \times 10^{0}$ |

**Verdict.** Layout variations have minimal influence on the V-PINN due to the fact that the quadrature formula employed in it does not consider the location of the sensor, and it has no mesh for computations. In case there is a gap in the middle, the relative error of the mollifier increases to 1.77 due to the violation of the FFT's implicit uniform-grid assumption.

Figure B.23 shows the recoveries by sensor layout.

### *Coloured Noise*

**Setup.** Varied: the correlation length of the additive noise, $\ell_c$, chosen from $\{0, 0.1, 0.5, 1.0, 2.0\}$, where $\ell_c = 0$ corresponds to white noise having no correlation and $\ell_c = 2.0$ means highly correlated noise. Fixed: $\delta = 0.015$, three patterns. Purpose: to check robustness against deviations from the white noise assumption of (4.4).

**Results.** Table 5.10 reports the source error against $\ell_c$.

**Verdict.** Both approaches become less efficient at small values of the correlation length and become better as $\ell_c$ tends to 2, yet the mollifier is more strongly influenced at the correlation

**Table 5.10:** Absolute source error versus the noise correlation length $\ell_c$, at $\delta = 0.015$ (V-PINN vs mollification).

| $\ell_c$ | V-PINN $\varphi_{abs}$ | Mollif. $\varphi_{abs}$ |
|---|---|---|
| 0.0 | $1.97 \times 10^{-2}$ | $2.75 \times 10^{-2}$ |
| 0.1 | $2.96 \times 10^{-2}$ | $8.48 \times 10^{-2}$ |
| 0.5 | $2.87 \times 10^{-2}$ | $5.52 \times 10^{-2}$ |
| 1.0 | $2.38 \times 10^{-2}$ | $2.49 \times 10^{-2}$ |
| 2.0 | $2.11 \times 10^{-2}$ | $2.19 \times 10^{-2}$ |

length being equal to 0.1 ($8.48 \times 10^{-2}$ compared to $2.96 \times 10^{-2}$) because correlated noise introduces in-band energy which the abrupt cutoff cannot separate from signal.

Figure B.24 plots the source error against the noise correlation length.

#### *Data-Efficiency Frontier*

**Setup.** Varied jointly: $\delta$ and $N_{sens}$, where values for $\delta$ are 0.007, 0.015, 0.033 and 0.040 and those for $N_{sens}$ are 15, 30, 50, 100, 200, 300 and 600. This generates a full set of combinations on a $4 \times 7$ grid. Purpose: to map the accuracy frontier in the noise-data plane and highlight the regions of greatest difference between the algorithms.

**Results.** Table 5.11 gives the relative source error in every cell, paired V-PINN / mollification.

**Table 5.11:** Relative source error $\varphi_{rel}$ on the noise sensor grid (V-PINN vs mollification). Each cell pairs V-PINN / mollification.

| $N_{sens}$ | $\delta = 0.007$<br>V / Mollif. | $\delta = 0.015$<br>V / Mollif. | $\delta = 0.033$<br>V / Mollif. | $\delta = 0.040$<br>V / Mollif. |
|---|---|---|---|---|
| 15 | $1.66\times10^{-2}$ / $3.56\times10^{-1}$ | $2.11\times10^{-2}$ / $3.56\times10^{-1}$ | $2.86\times10^{-2}$ / $3.57\times10^{-1}$ | $2.99\times10^{-2}$ / $3.58\times10^{-1}$ |
| 30 | $1.22\times10^{-2}$ / $8.56\times10^{-2}$ | $2.12\times10^{-2}$ / $8.77\times10^{-2}$ | $2.98\times10^{-2}$ / $1.12\times10^{-1}$ | $3.10\times10^{-2}$ / $1.24\times10^{-1}$ |
| 50 | $1.49\times10^{-2}$ / $3.65\times10^{-2}$ | $2.33\times10^{-2}$ / $6.29\times10^{-2}$ | $3.02\times10^{-2}$ / $7.38\times10^{-2}$ | $3.12\times10^{-2}$ / $7.25\times10^{-2}$ |
| 100 | $9.16\times10^{-3}$ / $1.56\times10^{-2}$ | $1.44\times10^{-2}$ / $5.03\times10^{-2}$ | $2.13\times10^{-2}$ / $7.59\times10^{-2}$ | $2.31\times10^{-2}$ / $7.66\times10^{-2}$ |
| 200 | $8.14\times10^{-3}$ / $1.26\times10^{-2}$ | $1.23\times10^{-2}$ / $2.06\times10^{-2}$ | $1.76\times10^{-2}$ / $2.62\times10^{-2}$ | $1.90\times10^{-2}$ / $2.98\times10^{-2}$ |
| 300 | $7.52\times10^{-3}$ / $1.34\times10^{-2}$ | $1.17\times10^{-2}$ / $1.97\times10^{-2}$ | $1.81\times10^{-2}$ / $3.96\times10^{-2}$ | $1.98\times10^{-2}$ / $3.53\times10^{-2}$ |
| 600 | $7.69\times10^{-3}$ / $1.18\times10^{-2}$ | $1.20\times10^{-2}$ / $1.71\times10^{-2}$ | $1.80\times10^{-2}$ / $3.09\times10^{-2}$ | $1.95\times10^{-2}$ / $2.88\times10^{-2}$ |

**Verdict.** The V-PINN has a lower relative error than the mollifier in all twenty-eight cells. This benefit is most pronounced at the point where there is very little data, that is, $(\delta, N_{sens}) = (0.007, 15)$. At this point, the value of the V-PINN is $1.66 \times 10^{-2}$ while for the mollifier it is $3.56 \times 10^{-1}$, a factor above twenty.

Figure B.27 shows the data-efficiency frontier.

### *5.3.9 Experiments Unique to the Adjoint Study*

Two contrasts illustrate how the mesh-free approach is different from the grid-based adjoint: randomly scattered sensors and different levels of grid resolution for the adjoint problem solution. In both cases, there is no mollification analogue.

### *Randomly Scattered Sensors*

**Setup.** Varied: the number of randomly scattered sensors, $N_{\text{sens}} \in \{15, 30, 50, 100, 200\}$. Fixed: $\delta = 0.015$, three patterns. Purpose: to show the mesh-free contrast in the cleanest possible way, where the grid-based solver has to interpolate the scattered values onto its grid.

**Results.** Table 5.12 reports the source error against the scattered sensor count.

**Table 5.12:** Absolute source error (mean±std over three patterns) versus the number of randomly scattered sensors, at $\delta = 0.015$ (V-PINN vs adjoint).

| $N_{\text{sens}}$ | V-PINN $\varphi_{\text{abs}}$ | Adjoint $\varphi_{\text{abs}}$ |
|---|---|---|
| 15 | $4.74 \times 10^{-2} \pm 9.33 \times 10^{-3}$ | $8.92 \times 10^{-1} \pm 4.14 \times 10^{-1}$ |
| 30 | $3.96 \times 10^{-2} \pm 7.12 \times 10^{-3}$ | $3.55 \times 10^{-1} \pm 1.46 \times 10^{-1}$ |
| 50 | $4.15 \times 10^{-2} \pm 9.89 \times 10^{-3}$ | $1.40 \times 10^{-1} \pm 3.06 \times 10^{-2}$ |
| 100 | $3.56 \times 10^{-2} \pm 8.61 \times 10^{-3}$ | $6.97 \times 10^{-2} \pm 2.23 \times 10^{-2}$ |
| 200 | $3.07 \times 10^{-2} \pm 7.12 \times 10^{-3}$ | $4.67 \times 10^{-2} \pm 2.14 \times 10^{-3}$ |

**Verdict.** The errors in the meshless technique degrade gracefully ($4.74 \times 10^{-2}$ at $N_{\text{sens}} = 15$). The grid-based approach requires interpolation, which is why it performs poorly (0.892, with a very large variability in values). Placement of sensors has little impact on the V-PINN.

Figure B.25 plots the error against the number of randomly placed sensors.

### *Adjoint Grid-Resolution Sensitivity*

**Setup.** Varied: the spatial discretization of the adjoint solver, where $N_x$ is one of $\{50, 100, 200, 500, 1000, 2000\}$ with $N_t = \max(200, 2N_x)$, and the data set is kept constant. Fixed: $\delta = 0.015$ and three patterns; the V-PINN is mesh-free and so does not move. Purpose: the response of the grid-based inversion to variations in the grid size, compared with a technique that has no grid.

**Results.** Relative error and time to solve for the adjoint at all resolutions are given in Table 5.13

**Table 5.13:** Relative source error (mean ± standard deviation over three patterns) and solve time as a function of the adjoint discretization size $N_x$, computed at $\delta = 0.015$. Note that the V-PINN has no discretization and thus a constant relative source error across the sweep, $\varphi_{\text{rel}} = 1.20 \times 10^{-2}$.

| $N_x$ | Adjoint $\varphi_{\text{rel}}$ | time (s) |
|---|---|---|
| 50 | $1.65 \times 10^{-2} \pm 2.78 \times 10^{-3}$ | 1.9 |
| 100 | $1.31 \times 10^{-2} \pm 2.56 \times 10^{-3}$ | 1.8 |
| 200 | $1.14 \times 10^{-2} \pm 2.09 \times 10^{-3}$ | 4.2 |
| 500 | $1.22 \times 10^{-2} \pm 1.47 \times 10^{-3}$ | 17.3 |
| 1000 | $1.81 \times 10^{-2} \pm 1.27 \times 10^{-3}$ | 58.3 |
| 2000 | $3.05 \times 10^{-2} \pm 1.12 \times 10^{-3}$ | 212.0 |

**Verdict.** One must note that the adjoint problem error does not have monotonic dependence on $N_x$. The minimum is attained at $N_x = 200$, when the error equals $1.14 \times 10^{-2}$. In the case of $N_x = 2000$, the error value equals $3.05 \times 10^{-2}$, while the solve time rises drastically to 212 seconds. With a given $\lambda$, when grid refinement is taking place, there come into play increasingly many modes, for which the influence of the Sobolev penalty term becomes insignificant. Therefore, the ill-posedness is once again intensified, with the discretization being the regularizer and an optimal grid.

In Figure B.26, we display the values of the relative errors and solve times for the adjoint problem versus $N_x$, with the recoveries across resolutions.

## 5.4 Synthesis

Figures 5.1 and 5.2 illustrate the individual cases where each comparator is plotted against the V-PINN, before we interpret their stability slopes relative to the sharp rate of Theorem 4.11.

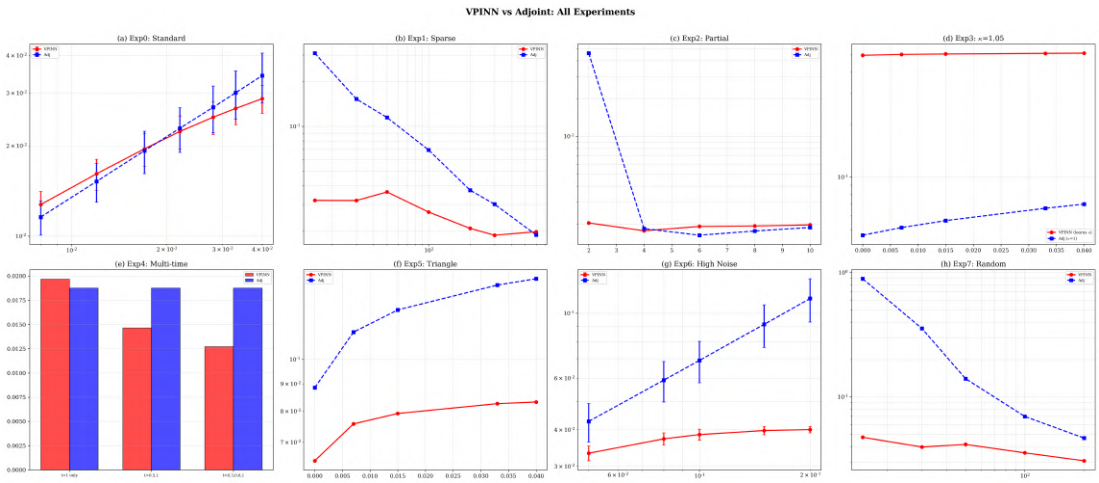


**Figure 5.1:** V-PINN against adjoint Sobolev Tikhonov for each of the eight direct comparisons between the two methods (source error versus the swept parameter; panel (e) depicts the multi-time bar graph). The V-PINN always performs equal to or better than the alternative, apart from the non-identifiable Experiment 3 (panel d).

With respect to the source ball $\mathcal{W}_{\ell,E}$, it has been proved in Theorem 4.11 that the $L^2(\mathbb{R})$ norm of the difference between $\varphi_1$ and $\varphi_2$ does not exceed some constant times $\|A\varphi_1 - A\varphi_2\|_{L^2(\mathbb{R})}^{\ell/(\ell+2)}$. Besides, according to Theorem A, this exponent cannot be improved; as a result, we obtain a $\delta^{1/2}$ rate for $\ell = 2$. Because of the saturation of $L^2(\mathbb{R})$-Tikhonov regularization at the $\delta^{2/3}$ rate, slopes above $\frac{2}{3}$ indicate use of the Hilbert-scale smoothness which the sharp theory rewards. Using the conventional definition of $\hat{\alpha}$ based on excess over floor, we see that Experiment 0 estimates $\hat{\alpha}$ as 0.79 ($R^2 = 0.98$) for the V-PINN, 0.76 ($R^2 = 0.9996$) for the adjoint, and 0.61 ($R^2 = 0.78$) for mollification. It is critical to understand these numbers as a *relative* ranking rather than as absolute exponents. The removal of the cold baseline, as highlighted in Remark 4.29, inflates the slope, and the floor-aware approximation places the V-PINN's smooth-source rate at the $L^2(\mathbb{R})$ saturation rate of $\frac{2}{3}$ (4.87) instead of 0.79. There are two important realizations here. On the one hand, both of the Hilbert-scale techniques, the V-PINN and the Sobolev-penalized adjoint formulation, track each other and recover more

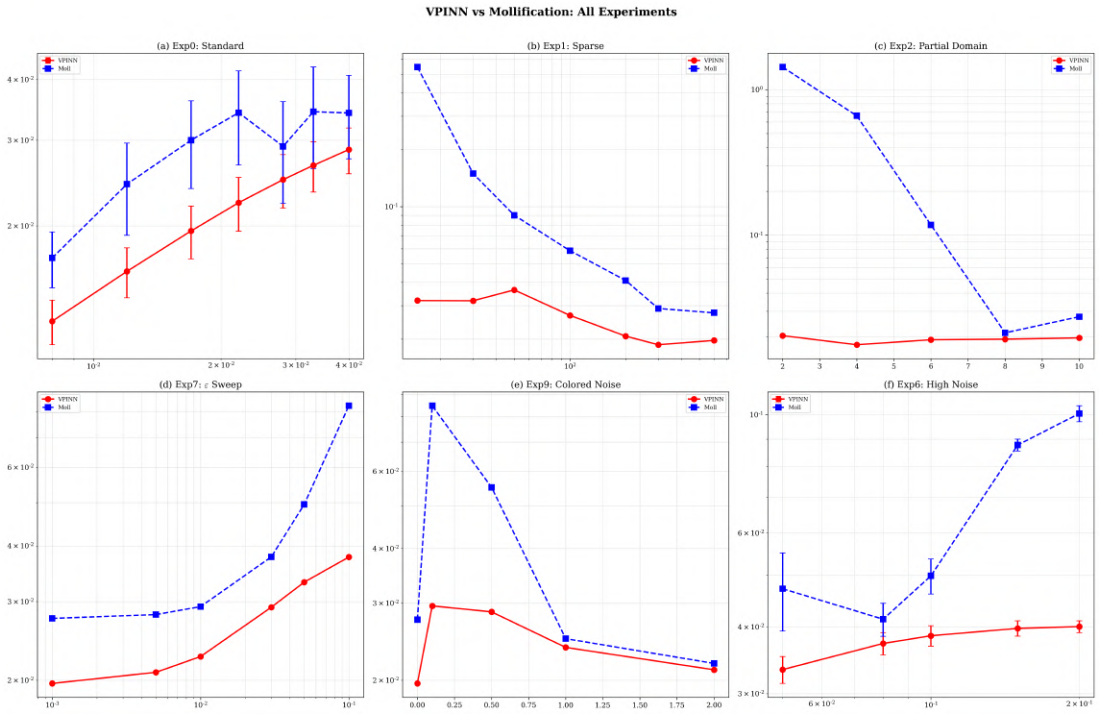


**Figure 5.2:** Comparison of V-PINN and Dirichlet-kernel mollification techniques using six different experiments that include standard and high noise, sparse and partial-domain sampling, the model-mismatch sweep, and coloured noise. In all subfigures, the V-PINN technique is found at or below the mollifier.

rapidly than the hard-truncation approach, which has no Hilbert-scale structure, just as the higher qualification of the Sobolev penalty indicates (Section 4.4.7); both clear the degree-two worst case of $\frac{1}{2}$. On the other hand, the $R^2$ value of the adjoint method is close to 1, which a well-defined quadratic penalty should have, whereas the V-PINN's scatter is due to optimization stochasticity and finite quadrature. For high noise settings its slope flattens out around 0.16 (Table 5.7). However, this does not represent a failure of the law itself, but an adaptive-weight plateau induced by the schedule $w_{\mathrm{m}}(\delta) = 500/(1 + (\delta/0.02)^2)$ of (4.28), where fidelity is traded for stability and the Hölder asymptotics no longer apply.

Figures 5.3 and 5.4 give the fitted Hölder laws comparing the V-PINN with each of its comparators in Experiment 0 (standard noise level) and the high-noise Experiment 6.

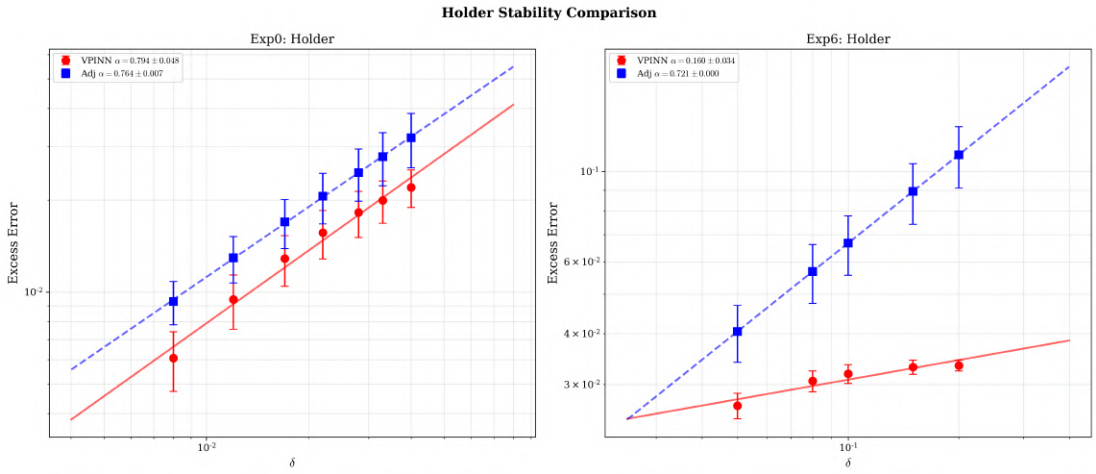


**Figure 5.3:** Empirical Hölder laws, V-PINN versus adjoint. In Experiment 0 the estimated exponent $\hat{\alpha}$ is equal to 0.794 for the V-PINN and 0.764 for the adjoint, on a common excess-over-floor basis; cf. Remark 4.29. In Experiment 6, which corresponds to the high noise case, the exponents are 0.160 and 0.721.

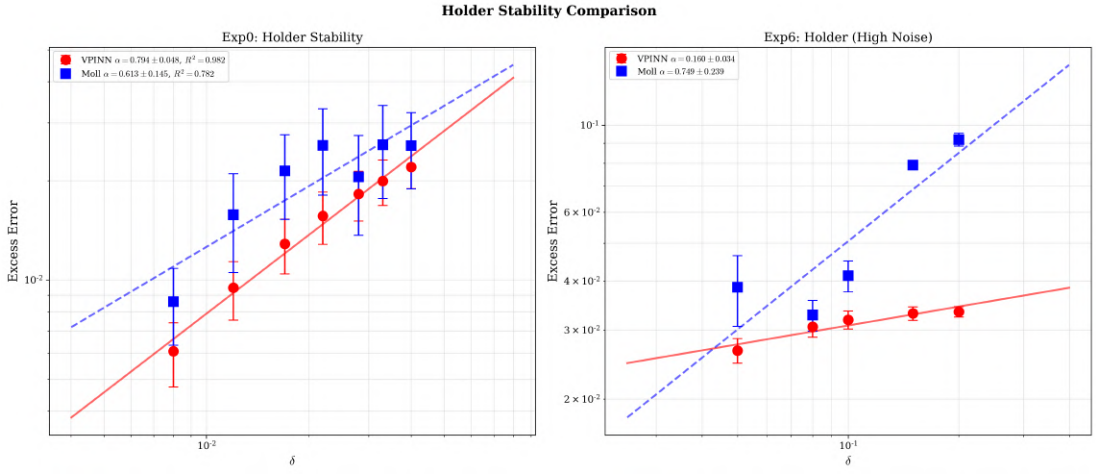


**Figure 5.4:** Empirical Hölder laws, for comparison between V-PINN and mollification. For Experiment 0, one gets $\hat{\alpha} = 0.794$ and $R^2 = 0.982$ compared to 0.613 and $R^2 = 0.782$ respectively. For Experiment 6 with high noise, one gets 0.160 compared to 0.749.

# References


[1] Bungartz, H.-J. and Griebel, M. (2004) 'Sparse grids', *Acta Numerica*, 13, pp. 147–269.

[2] Adams, R.A. and Fournier, J.J.F. (2003) *Sobolev Spaces. Pure and Applied Mathematics*, 140.

[3] Raissi, M. (2018) 'Deep hidden physics models: Deep learning of nonlinear partial differential equations', *Journal of Machine Learning Research*, 19(25), pp. 1–24.

[4] Korolev, Y. and Latz, J. (2021) *Lecture Notes, Michaelmas Term 2020, University of Cambridge*.

[5] Schmüdgen, K. (2017) *The Moment Problem*. Vol. 9. Cham: Springer.

[6] Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) 'Physics-informed machine learning', *Nature Reviews Physics*, 3(6), pp. 422–440.

[7] Gerth, D. (2021) 'A new interpretation of (Tikhonov) regularization', *Inverse Problems*, 37(6), 064002.

[8] Clason, C. (2020) 'Regularization of inverse problems', *arXiv preprint*, arXiv:2001.00617.

[9] de Klerk, E. and Laurent, M. (2019) 'A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis', in *World Women in Mathematics 2018: Proceedings of the First World Meeting for Women in Mathematics (WM)*$^2$, pp. 17–56. Cham: Springer International Publishing.

[10] Pinelis, I. (2016) 'On the extreme points of moments sets', *Mathematical Methods of Operations Research*, 83(3), pp. 325–349.

[11] Ascher, U.M., Ruuth, S.J. and Wetton, B.T.R. (1995) 'Implicit-explicit methods for time-dependent partial differential equations', *SIAM Journal on Numerical Analysis*, 32(3), pp. 797–823.

[12] Grenander, U. and Szegő, G. (1958) *Toeplitz Forms and Their Applications*. Berkeley: University of California Press.

[13] Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A. and Anandkumar, A. (2023) 'Neural operator: Learning maps between function spaces with applications to PDEs', *Journal of Machine Learning Research*, 24(89), pp. 1–97.

[14] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. et al. (2020) 'SciPy 1.0: Fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272. doi:10.1038/s41592-019-0686-2.

[15] Huangfu, Q. and Hall, J.A.J. (2018) 'Parallelizing the dual revised simplex method', *Mathematical Programming Computation*, 10(1), pp. 119–142. DOI: 10.1007/s12532-017-0130-5

[16] Nocedal, J. and Wright, S.J. (2006) *Numerical Optimization*. 2nd edn. Springer Series in Operations Research and Financial Engineering, New York: Springer.

[17] Landweber, L. (1951) 'An iteration formula for Fredholm integral equations of the first kind', *American Journal of Mathematics*, 73(3), pp. 615–624.

[18] Lagaris, I.E., Likas, A. and Fotiadis, D.I. (1998) 'Artificial neural networks for solving ordinary and partial differential equations', *IEEE Transactions on Neural Networks*, 9(5), pp. 987–1000.

[19] Umbricht, G.F. (2021) 'Identification of the source for full parabolic equations', *arXiv preprint* arXiv:2111.09819.

[20] Liu, K. and Zuazua, E. (2026) 'Moments, time-inversion and source identification for the heat equation', *Inverse Problems*, 42(1), p. 015009.

[21] Lassas, M. and Uhlmann, G. (2001) 'On determining a Riemannian manifold from the Dirichlet-to-Neumann map', *Annales scientifiques de l'École normale supérieure*, 34(5), pp. 771–787.

[22] Hegland, M. and Hofmann, B. (2010) 'Errors of regularisation under range inclusions using variable Hilbert scales', *arXiv preprint* arXiv:1005.3883.

[23] Mathé, P., Nair, M.T. and Hofmann, B. (2022) 'Regularization of linear ill-posed problems involving multiplication operators', *Applicable Analysis*, 101(2), pp. 714–732.

[24] Boţ, R.I. and Hofmann, B. (2016) 'Conditional stability versus ill-posedness for operator equations with monotone operators in Hilbert space', *Inverse Problems*, 32(12), p. 125003.

[25] Li, J., Wu, W. and Feng, X. (2023) 'Improved physics-informed neural networks combined with small sample learning to solve two-dimensional Stefan problem', *Entropy*, 25(4), p. 675.

[26] Osipenko, K.Y. and Wedenskaya, E.V. (2007) 'Optimal recovery of solutions of the generalized heat equation in the unit ball from inaccurate data', *Journal of Complexity*, 23(4–6), pp. 653–661.

[27] Hillebrecht, B. and Unger, B. (2023) 'Rigorous a posteriori error bounds for PDE-defined PINNs', *IEEE Transactions on Neural Networks and Learning Systems*, 36(1), pp. 1583–1593.

[28] Berrone, S., Canuto, C. and Pintore, M. (2022) 'Solving PDEs by variational physics-informed

neural networks: An a posteriori error analysis', *Annali dell'Università di Ferrara*, 68(2), pp. 575–595.

[29] Zhang, H. and Liu, J. (2023) 'Solving an inverse source problem by deep neural network method with convergence and error analysis', *Inverse Problems*, 39(7), p. 075013.

[30] Ji, X., Jiang, Z., Song, P. and Yuan, C. (2025) 'Potential identification via Tikhonov-PINNs', *Inverse Problems*, 41(11), p. 115008.

[31] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. et al. (2019) 'PyTorch: an imperative style, high-performance deep learning library', in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035.

[32] Colton, D. (2003) 'Inverse Acoustic and Electromagnetic', *Inside Out: Inverse Problems and Applications*, 47, p. 67.

[33] Bhatia, R. (2013) *Matrix Analysis*. Springer Science & Business Media.

[34] Bongioanni, B. and Torrea, J.L. (2006) 'Sobolev spaces associated to the harmonic oscillator', *Proceedings of the Indian Academy of Sciences–Mathematical Sciences*, 116(3), pp. 337–360.

[35] Comte, F. and Genon-Catalot, V. (2018) 'Laguerre and Hermite bases for inverse problems', *Journal of the Korean Statistical Society*, 47(3), pp. 273–296. doi:10.1016/j.jkss.2018.03.001.

[36] Bonnans, J.F. and Shapiro, A. (2013) *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media.

[37] Brenner, S.C. and Scott, L.R. (2008) *The Mathematical Theory of Finite Element Methods*. New York, NY: Springer New York.

[38] Foucart, S. and Liao, C. (2022) 'Optimal recovery from inaccurate data in Hilbert spaces: regularize, but what of the parameter?', *Constructive Approximation*.

[39] Allen, S.M. and Cahn, J.W. (1979) 'A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening', *Acta Metallurgica*, 27(6), pp. 1085–1095.

[40] Guo, B.Y., Shen, J. and Xu, C.L. (2003) 'Spectral and pseudospectral approximations using Hermite functions: Application to the Dirac equation', *Advances in Computational Mathematics*, 19(1), pp. 35–55.

[41] Burgers, J.M. (1948) 'A mathematical model illustrating the theory of turbulence', *Advances in Applied Mechanics*, 1, pp. 171–199.

[42] Golub, G.H. and Van Loan, C.F. (2013) *Matrix Computations*. Vol. 3. JHU Press.

[43] Regińska, T. and Tautenhahn, U. (2009) 'Conditional stability estimates and regularization with applications to Cauchy problems for the Helmholtz equation', *Numerical Functional Analysis and Optimization*, 30(9–10), pp. 1065–1097. doi:10.1080/01630560903393170.

[44] Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995) 'A limited memory algorithm for bound constrained optimization', *SIAM Journal on Scientific Computing*, 16(5), pp. 1190–1208.

[45] Canuto, C., Hussaini, M.Y., Quarteroni, A. and Zang, T.A. (2006) *Spectral Methods*. Vol. 285. Berlin: Springer.

[46] Colton, D. (2003) 'Inverse Acoustic and Electromagnetic', *Inside Out: Inverse Problems and Applications*, 47, p. 67.

[47] Semmes, S. (2010) *An Introduction to Some Aspects of Functional Analysis, Volume 2: Bounded Linear Operators*.

[48] Randrianarisoa, T. and Szabo, B. (2023) 'Variational Gaussian processes for linear inverse problems', *Advances in Neural Information Processing Systems*, 36, pp. 28960–28972.

[49] Yang, F. and Li, D.G. (2010) 'The Fourier regularization method for identifying the unknown source on Poisson equation', *International Journal of Mathematical Analysis*, 4, pp. 1357–1361.

[50] Egger, H. (2008) 'Y-scale regularization', *SIAM Journal on Numerical Analysis*, 46(1), pp. 419–436.

[51] Belomestny, D., Comte, F. and Genon-Catalot, V. (2019) 'Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$', *Annals of the Institute of Statistical Mathematics*, 71(1), pp. 29–62.

[52] Muscat, J. (2024) *Functional Analysis: An Introduction to Metric Spaces, Hilbert Spaces, and Banach Algebras*. Springer.

[53] Yamamoto, M. (1993) 'Conditional stability in determination of force terms of heat equations in a rectangle', *Mathematical and Computer Modelling*, 18(1), pp. 79–88.

[54] Plaskota, L. (1996) *Noisy information and computational complexity*. Cambridge University Press.

[55] Romanov, V. and Hasanov, A. (2022) 'Uniqueness and stability analysis of final data inverse source problems for evolution equations', *Journal of Inverse and Ill-posed Problems*, 30(3), pp. 425–446.

[56] Arbenz, P. and Kressner, D. (2012) *Lecture notes on solving large scale eigenvalue problems*. D-MATH, ETH Zurich.

[57] Neubauer, A. (2016) *Inverse and Ill-Posed Problems*. Lecture notes. Available at: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Lecture+Notes+Inverse+and+Ill-Posed+Problems+Andreas+Neubauer&btnG=.

[58] Conway, J.B. (1990) *A Course in Functional Analysis*. 2nd edn. Vol. 96 of *Graduate Texts in Mathematics*. New York: Springer.

[59] Wuryatmo, A.S. (1991) *Approximation Solutions of the Wave Problems*. PhD thesis. University of Tasmania.

[60] Davis, P.J. and Rabinowitz, P. (1984) *Methods of Numerical Integration*. 2nd edn. Orlando, FL: Academic Press.

[61] Hornik, K. (1991) 'Approximation capabilities of multilayer feedforward networks', *Neural Networks*, 4(2), pp. 251–257.

[62] *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.2.7 of 2026-06-15. F.W.J. Olver, A.B. Olde Daalhuis, D.W. Lozier, B.I. Schneider, R.F. Boisvert, C.W. Clark, B.R. Miller, B.V. Saunders, H.S. Cohl and M.A. McClain, eds.

[63] Donoho, D.L. (1994) 'Statistical estimation and optimal recovery', *The Annals of Statistics*, 22(1), pp. 238–270. DOI: 10.1214/aos/1176325367

[64] Lu, L., Meng, X., Mao, Z. and Karniadakis, G.E. (2021) 'DeepXDE: A deep learning library for solving differential equations', *SIAM Review*, 63(1), pp. 208–228.

[65] Dou, F.-F., Fu, C.-L. and Yang, F.-L. (2009) 'Optimal error bound and Fourier regularization for identifying an unknown source in the

heat equation', *Journal of Computational and Applied Mathematics*, 230(2), pp. 728–737.

[66] Egger, H. and Hofmann, B. (2018) 'Tikhonov regularization in Hilbert scales under conditional stability assumptions', *Inverse Problems*, 34(11), p. 115015. arXiv:1807.05807.

[67] Han, E. and Jentzen, A. (2017) 'Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations', *Communications in Mathematics and Statistics*, 5(4), pp. 349–380.

[68] Engl, H.W., Hanke, M. and Neubauer, A. (1996) *Regularization of Inverse Problems*. Vol. 375 of *Mathematics and Its Applications*. Dordrecht: Kluwer Academic Publishers.

[69] Brézis, H. (2011) *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Vol. 2. New York: Springer.

[70] Evans, L.C. (2010) *Partial Differential Equations*. 2nd edn. Vol. 19 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society.

[71] Brunton, S.L., Proctor, J.L. and Kutz, J.N. (2016) 'Discovering governing equations from data by sparse identification of nonlinear dynamical systems', *Proceedings of the National Academy of Sciences*, 113(15), pp. 3932–3937. DOI: 10.1073/pnas.1517384113

[72] Hadamard, J. (1964) *La théorie des équations aux dérivées partielles*. Éditions scientifiques.

[73] Cao, S. (2021) 'Choose a transformer: Fourier or Galerkin', *Advances in Neural Information Processing Systems*, 34, pp. 24924–24940. arXiv:2105.14995.

[74] Guo, X., Li, W. and Iorio, F. (2016) 'Convolutional neural networks for steady flow approximation', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 481–490. DOI: 10.1145/2939672.2939738

[75] Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A. and Anandkumar, A. (2021) 'Fourier neural operator for parametric partial differential equations', in *9th International Conference on Learning Representations (ICLR)*, Virtual Event, Austria. arXiv:2010.08895.

[76] Lu, L., Jin, P., Pang, G., Zhang, Z. and Karniadakis, G.E. (2021) 'Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators', *Nature Machine Intelligence*, 3(3), pp. 218–229. DOI: 10.1038/s42256-021-00302-5

[77] Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A. and Battaglia, P.W. (2021) 'Learning mesh-based simulation with graph networks', in *9th International Conference on Learning Representations (ICLR)*, Virtual Event, Austria. arXiv:2010.03409.

[78] Rudy, S.H., Brunton, S.L., Proctor, J.L. and Kutz, J.N. (2017) 'Data-driven discovery of partial differential equations', *Science Advances*, 3(4), p. e1602614. DOI: 10.1126/sciadv.1602614

[79] Folland, G.B. (1999) *Real Analysis: Modern Techniques and Their Applications*. 2nd edn. New York: John Wiley & Sons.

[80] Grafakos, L. (2008) *Classical Fourier Analysis*. 2nd edn. Vol. 249 of *Graduate Texts in Mathematics*. New York: Springer.

[81] Wyatt, Z. (2022) *Analysis of Partial Differential Equations*. University of Cambridge, Michaelmas 2021 lecture notes. Available at: https://aditya-sengupta.github.io/notes/pde.pdf

[82] He, K., Zhang, X., Ren, S. and Sun, J. (2015) 'Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification', in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034.

[83] Hofmann, B. and Kindermann, S. (2010) 'On the degree of ill-posedness for linear problems with non-compact operators', *Methods and Applications of Analysis*, 17(4), pp. 445–462. DOI: 10.4310/MAA.2010.v17.n4.a8

[84] Hofmann, B., Hofmann, C., Mathé, P. and Plato, R. (2022) 'Nonlinear Tikhonov regularization in Hilbert scales with oversmoothing penalty: inspecting balancing principles', in Jadamba, B., Khan, A.A., Migórski, S. and Sama, M. eds, *Deterministic and Stochastic Optimal Control and Inverse Problems*. Boca Raton: CRC Press, pp. 79–111.

[85] Hofmann, B. and Mathé, P. (2012) 'A note on the modulus of continuity for ill-posed problems in Hilbert space', *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 18(1), pp. 34–41.

[86] Boţ, R.I., Hofmann, B. and Mathé, P. (2013) 'Regularizability of ill-posed problems and the modulus of continuity', *Zeitschrift für Analysis und ihre Anwendungen*, 32(3), pp. 299–312.

[87] Hofmann, B., Mathé, P. and Schieck, M. (2008) 'Modulus of continuity for conditionally stable ill-posed problems in Hilbert space', *Journal of Inverse and Ill-Posed Problems*, 16(6), pp. 567–585.

[88] Anague, B., Hosseini, B., Karambal, I. and Ngnotchouye, J.M. (2025) 'Physics-Informed Neural Networks for Joint Source and Parameter Estimation in Advection-Diffusion Equations', *arXiv preprint*, arXiv:2512.07755.

[89] Isakov, V. (1991) 'Inverse parabolic problems with the final overdetermination', *Communications on Pure and Applied Mathematics*, 44(2), pp. 185–209.

[90] Isakov, V. (2017) *Inverse Problems for Partial Differential Equations*. 3rd edn. Cham: Springer.

[91] Karlin, S. and Studden, W.J. (1966) *Tchebycheff Systems: With Applications in Analysis and Statistics*. New York: Interscience Publishers.

[92] Kharazmi, E., Zhang, Z. and Karniadakis, G.E. (2021) 'hp-VPINNs: Variational physics-informed neural networks with domain decomposition', *Computer Methods in Applied Mechanics and Engineering*, 374, p. 113547.

[93] Kingma, D.P. and Ba, J. (2015) 'Adam: A method for stochastic optimization', in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

[94] Krein, M.G. and Petunin, Y.I. (1966) 'Scales of Banach spaces', *Russian Mathematical Surveys*, 21(2), pp. 85–159.

[95] Dileep, A. (2024) *Inverse Source Problems for the Damped Euler-Bernoulli Beam and Kirchhoff Plate Equations*. PhD thesis. Indian Institute of Space Science and Technology.

[96] LeVeque, R.J. (2007) *Finite Difference Methods for Ordinary and Partial Differential Equations*. Philadelphia: SIAM.

[97] Zeidler, E. (1990) *Nonlinear Functional Analysis and Its Applications II/A: Linear Monotone Operators*. New York: Springer-Verlag. Available at: https://link.springer.com/book/10.1007/978-1-4612-0985-0

[98] Shohat, J.A. and Tamarkin, J.D. (1943) *The Problem of Moments*. American Mathematical Society.

[99] Liu, D.C. and Nocedal, J. (1989) 'On the limited memory BFGS method for large scale optimization', *Mathematical Programming*, 45(1–3), pp. 503–528.

[100] Loshchilov, I. and Hutter, F. (2017) 'SGDR: Stochastic gradient descent with warm restarts', in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

[101] Magaril-Il'yaev, G.G. and Osipenko, K.Y. (2009) 'Optimal recovery of the solution of the heat equation from inaccurate data', *Sbornik: Mathematics*, 200(5), pp. 665–682. DOI: 10.1070/SM2009v200n05ABEH004014

[102] Protas, B., Bewley, T.R. and Hagen, G. (2004) 'A computational framework for the regularization of adjoint analysis in multiscale PDE systems', *Journal of Computational Physics*, 195(1), pp. 49–89.

[103] Evans, G., Blackledge, J. and Yardley, P. (2012) *Analytic Methods for Partial Differential Equations*. Springer Science & Business Media.

[104] Mathé, P. and Pereverzev, S.V. (2003) 'Geometry of linear ill-posed problems in variable Hilbert scales', *Inverse Problems*, 19(3), pp. 789–803.

[105] Sickel, W. and Ullrich, T. (2011) 'Spline interpolation on sparse grids', *Applicable Analysis*, 90(3–4), pp. 337–383.

[106] Melkman, A.A. and Micchelli, C.A. (1979) 'Optimal estimation of linear operators in Hilbert spaces from inaccurate data', *SIAM Journal on Numerical Analysis*, 16(1), pp. 87–105. DOI: 10.1137/0716007

[107] Meng, Z. and Zhao, Z. (2017) 'Hermite spectral method for solving inverse heat source problems in multiple dimensions', *Inverse Problems in Science and Engineering*, 25(9), pp. 1243–1258.

[108] Micchelli, C.A. and Rivlin, T.J. (1977) 'A survey of optimal recovery', in Micchelli, C.A. and Rivlin, T.J. eds, *Optimal Estimation in Approximation Theory*. New York: Plenum Press, pp. 1–54.

[109] Mishra, S. and Molinaro, R. (2022) 'Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs', *IMA Journal of Numerical Analysis*, 42(2), pp. 981–1022.

[110] Morozov, V.A. (1966) 'On the solution of functional equations by the method of regularization', *Soviet Mathematics Doklady*, 7, pp. 414–417.

[111] Murio, D.A. (1993) *The Mollification Method and the Numerical Solution of Ill-Posed Problems*. New York: John Wiley & Sons.

[112] Hào, D.N. (1996) 'A mollification method for a noncharacteristic Cauchy problem for a parabolic equation', *Journal of Mathematical Analysis and Applications*, 199(3), pp. 873–909.

[113] Engel, K.-J. and Nagel, R. (2000) *One-Parameter Semigroups for Linear Evolution Equations*. Vol. 194 of *Graduate Texts in Mathematics*. New York: Springer.

[114] Nair, M.T. (2004) 'Ill-Posedness of Backward Heat Conduction Problem', in National Seminar on 'Mathematical Modelling' at PSGR Krishnammal College for Women, Coimbatore, August 2004, Vol. 6.

[115] Natterer, F. (1984) 'Error bounds for Tikhonov regularization in Hilbert scales', *Applicable Analysis*, 18(1–2), pp. 29–37.

[116] Quarteroni, A., Manzoni, A. and Negri, F. (2015) *Reduced basis methods for partial differential equations: an introduction*. Springer.

[117] Nirenberg, L. (1959) 'On elliptic partial differential equations', *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze* (3), 13(2), pp. 115–162.

[118] Pascanu, R., Mikolov, T. and Bengio, Y. (2013) 'On the difficulty of training recurrent neural networks', in *Proceedings of the 30th International Conference on Machine Learning*, Vol. 28(3) of *PMLR*, pp. 1310–1318.

[119] Pazy, A. (1983) *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Vol. 44 of *Applied Mathematical Sciences*. New York: Springer.

[120] Plessix, R.-E. (2006) 'A review of the adjoint-state method for computing the gradient of a functional with geophysical applications', *Geophysical Journal International*, 167(2), pp. 495–503.

[121] Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) 'Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations', *Journal of Computational Physics*, 378, pp. 686–707.

[122] Hegland, M. (1992) 'An optimal order regularization method which does not use additional smoothness assumptions', *SIAM Journal on Numerical Analysis*, 29(5), pp. 1446–1461.

[123] Hofmann, B. (2006) 'Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators', *Mathematical Methods in the Applied Sciences*, 29(3), pp. 351–371.

[124] Hegland, M. (1995) 'Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization', *Applicable Analysis*, 59(1–4), pp. 207–223.

[125] Tabarintseva, E.V. (2015) 'On an estimate for the modulus of continuity of a nonlinear inverse problem', *Ural Mathematical Journal*, 1(1), pp. 87–92.

[126] Reed, M. and Simon, B. (1975) *Methods of Modern Mathematical Physics, Vol. II: Fourier Analysis, Self-Adjointness*. New York: Academic Press.

[127] Reed, M. and Simon, B. (1980) *Methods of Modern Mathematical Physics, Vol. I: Functional Analysis*. Revised and enlarged edn. New York: Academic Press.

[128] Richtmyer, R.D. and Morton, K.W. (1967) *Difference Methods for Initial-Value Problems*. 2nd edn. New York: Interscience Publishers.

[129] Rogosinski, W.W. (1958) 'Moments of non-negative mass', *Proceedings of the Royal Society of London, Series A*, 245, pp. 1–27.

[130] Rudin, W. (1987) *Real and Complex Analysis*. 3rd edn. New York: McGraw-Hill.

[131] Schilling, R.L., Song, R. and Vondráček, Z. (2012) *Bernstein Functions: Theory and Applications*. 2nd edn. Vol. 37 of *De Gruyter Studies*

*in Mathematics*. Berlin: De Gruyter. DOI: 10.1515/9783110269338

[132] Strauss, W.A. (2007) *Partial Differential Equations: An Introduction*. John Wiley & Sons.

[133] Szegő, G. (1975) *Orthogonal Polynomials*. 4th edn. Providence, RI: American Mathematical Society.

[134] Tautenhahn, U. (1998) 'Optimality for ill-posed problems under general source conditions', *Numerical Functional Analysis and Optimization*, 19(3–4), pp. 377–398.

[135] Tautenhahn, U. (2004) 'On order optimal regularization under general source conditions', *Proceedings of the Estonian Academy of Sciences, Physics, Mathematics*, 53(2), pp. 116–123.

[136] Tautenhahn, U., Hämarík, U., Hofmann, B. and Shao, Y. (2013) 'Conditional stability estimates for ill-posed PDE problems by using interpolation', *Numerical Functional Analysis and Optimization*, 34(12), pp. 1370–1417.

[137] Thangavelu, S. (1993) *Lectures on Hermite and Laguerre Expansions*. Vol. 42. Princeton University Press.

[138] Tikhonov, A.N. (1963) 'Solution of incorrectly formulated problems and the regularization method', *Soviet Mathematics Doklady*, 4, pp. 1035–1038.

[139] Tikhonov, A.N. and Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems*. Washington, DC: V.H. Winston & Sons.

[140] Trefethen, L.N. (2019) *Approximation Theory and Approximation Practice*. Extended edn. Philadelphia: SIAM.

[141] Trefethen, L.N. (2022) 'Exactness of quadrature formulas', *SIAM Review*, 64(1), pp. 132–150.

[142] Umbricht, G.F. and Rubio, D. (2024) 'Regularization Techniques for Estimating the Source in a Complete Parabolic Equation in $\mathbb{R}^n$', *arXiv preprint* arXiv:2404.13094.

[143] Wang, S., Yu, X. and Perdikaris, P. (2022) 'When and why PINNs fail to train: A neural tangent kernel perspective', *Journal of Computational Physics*, 449, p. 110768.

[144] Werner, F. and Hofmann, B. (2025) 'A unified concept of the degree of ill-posedness for compact and non-compact linear operator equations in Hilbert spaces under the auspices of the spectral theorem', *Numerical Functional Analysis and Optimization*, 46(4–5), pp. 322–347. arXiv:2408.01148. DOI: 10.1080/01630563.2025.2451922

[145] Wirgin, A. (2004) 'The inverse crime', *arXiv preprint* arXiv:math-ph/0401050.

[146] Hubmer, Simon, and Ronny Ramlau. "Andreas Neubauer." (2026).

[147] Yang, L., Zhu, L., He, S., Feng, X. and Zhao, B. (2024) 'Solving two kinds of inverse source problems for the heat equations by a mollification regularization method with Dirichlet kernel', *Mathematical Methods in the Applied Sciences*, 47(18), pp. 14024–14036.

# Appendix A

# Notations

Symbols are grouped by chapter, in order of first appearance; each chapter has its own accent colour.

## General Conventions

$\mathbb{R}$, $\mathbb{N}$, $\mathbb{C}$, $\mathbb{Z}$ – Real numbers, natural numbers (including 0), complex numbers, integers

$\mathbb{R} \times (0, 1]$ – Space–time domain; the terminal time is normalised to $t = 1$

$|\cdot|$ – Absolute value of a scalar; Euclidean modulus of a vector

$\|\cdot\|$ – Norm in $L^2(\mathbb{R})$ unless a subscript names another space

$\langle \cdot, \cdot \rangle$ – Inner product, conjugate-linear in the second argument

$\mathrm{d}x$, $\mathrm{d}\xi$, $\mathrm{d}t$ – Lebesgue measure element (upright Roman d)

$a \lesssim b$ – $a \le C\, b$ for an absolute constant $C > 0$ fixed at first use

$a \asymp b$ – $a \lesssim b$ and $b \lesssim a$ hold simultaneously

$\mathcal{O}(\cdot)$ – Landau big-$\mathcal{O}$ symbol

$x^\dagger$ – The exact (ground-truth) value of a quantity, e.g. $\varphi^\dagger$

$\delta_{jk}$ – Kronecker delta

$\mathbb{E}[\cdot]$ – Expectation with respect to the noise distribution

$\lambda_{\min}(M)$, $\lambda_{\max}(M)$ – Smallest and largest eigenvalue of a symmetric matrix $M$

## Chapter 1 · Analytical, Spectral, and Numerical Foundations

### *Function spaces and norms*

$L^2(\mathbb{R})$ – Lebesgue space of square-integrable functions on $\mathbb{R}$; $\|f\|_{L^2(\mathbb{R})} = \left(\int_{\mathbb{R}} |f(x)|^2 \,\mathrm{d}x\right)^{1/2}$ (Definition 1.5)

$H^s(\mathbb{R})$, $s \in \mathbb{R}$ – Fractional Sobolev space; $\|f\|^2_{H^s(\mathbb{R})} = \int_{\mathbb{R}} (1 + \xi^2)^s |\hat{f}(\xi)|^2 \,\mathrm{d}\xi$ (Definition 1.7)

$H^k(\mathbb{R})$, $H^\ell(\mathbb{R})$ – Integer-order Sobolev spaces: the $H^s(\mathbb{R})$ scale at integer $s$ ($H^\ell(\mathbb{R})$ at the fixed source-class order $\ell \ge 1$) (Definition 1.7)

$\mathcal{S}(\mathbb{R})$ – Schwartz space of rapidly decreasing smooth functions (Definition 1.6)

$p_{a,b}(f)$ – Schwartz seminorm $\sup_{x \in \mathbb{R}} |x^a \partial^b f(x)|$ (Definition 1.6)

$\partial^\alpha$ – Multi-index derivative $\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$ (Definition 1.2)

$\mathcal{S}'(\mathbb{R})$ – Tempered distributions; topological dual of $\mathcal{S}(\mathbb{R})$ (Definition 1.6)

$L^p(I; X)$ – Bochner space of $X$-valued maps; $\|u\|_{L^p} = \left(\int_I \|u(t)\|_X^p \,\mathrm{d}t\right)^{1/p}$ (Definition 1.15)

$C(I; X)$ – Continuous $X$-valued maps on $I = [0, 1]$, norm $\sup_t \|u(t)\|_X$ (Definition 1.15)

### *Fourier analysis*

$\hat{f} = \mathcal{F}[f]$ – Fourier transform, symmetric convention $\hat{f}(\xi) = (2\pi)^{-1/2} \int_{\mathbb{R}} f(x) e^{-i\xi x} \,\mathrm{d}x$ (Definition 1.9)

$\mathcal{F}^{-1}$ – Inverse Fourier transform; both extended to $\mathcal{S}'(\mathbb{R})$ by duality (Definition 1.9)

Plancherel identity – $\|f\|_{L^2(\mathbb{R})} = \|\hat{f}\|_{L^2(\mathbb{R})}$, with no factor of $2\pi$ (Proposition 1.10)

$\widehat{\partial_x^k f}(\xi)$ – Equals $(i\xi)^k \hat{f}(\xi)$; differentiation rule in Fourier space (Proposition 1.10)

### *Heat equation and the forward map*

$\Omega \subset \mathbb{R}^n$ – Open spatial domain on which a PDE is posed; throughout this work $\Omega = \mathbb{R}$ and the spatial dimension is $n = 1$ (Definition 1.1, Remark 1.3)

$v$ – State variable; solution of the heat equation on $\mathbb{R} \times (0, 1]$ (Definition 1.1)

$\varphi$ – Time-independent source term driving the heat equation (Definition 1.1)

$\Delta$ – Laplacian $\partial_{xx}$; generator of the heat semigroup (Definition 1.16). In the second-order classification of Definition 1.2, $\Delta = B^2 - AC$ instead denotes the discriminant—a distinct, purely local use

$G(x, t)$ – Gaussian heat kernel $(4\pi t)^{-1/2} e^{-x^2/4t}$ (Definition 1.16)

$\{e^{t\Delta}\}_{t \ge 0}$ – Heat semigroup on $L^2(\mathbb{R})$; Fourier multiplier $e^{-t\xi^2}$; contractive (Definition 1.16, Proposition 1.17)

$A$ – Forward (terminal-state) operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, $A\varphi := v(\cdot, 1)$

(Proposition 1.19, Definition 4.3)

$m(\xi)$ – Forward-operator Fourier symbol $(1-e^{-\xi^2})/\xi^2$ for $\xi \neq 0$, $m(0) := 1$ (Proposition 1.19)

$A_\varepsilon$ – Perturbed forward operator, generated with diffusivity $a(x) = 1 + \varepsilon e^{-x^2/4}$ and inverted with $a \equiv 1$; the anti-inverse-crime mismatch (Section 4.1.5)

***Hermite analysis***

---

$H_j$ – Hermite polynomial of degree $j$ (Definition 1.11)

$\psi_j$ – Hermite function $(2^j j!\sqrt{\pi})^{-1/2} H_j(x) e^{-x^2/2}$; orthonormal basis of $L^2(\mathbb{R})$ (Definition 1.11)

$B$ – Harmonic-oscillator operator $-\partial_{xx} + x^2$; $B\psi_j = (2j+1)\psi_j$ (Proposition 1.12)

$\psi_j^\sigma$ – Scaled Hermite function $\sigma^{-1/2}\psi_j(x/\sigma)$, $\sigma > 0$; orthonormal basis of $L^2(\mathbb{R})$ (Definition 1.13)

$L_\sigma$ – Scaled oscillator $-\sigma^2\partial_{xx} + \sigma^{-2}x^2$; $L_\sigma\psi_j^\sigma = (2j+1)\psi_j^\sigma$ (Definition 1.13)

$\sigma$ – Hermite length-scale parameter (Definition 1.13)

$N_\varphi$ – Hermite truncation level (number of retained modes) (Definition 1.14)

$\Lambda$ – Hermite–Sobolev mass matrix $\operatorname{diag}\big((2j+1)^\ell\big)_{j=0}^{N_\varphi-1}$ (Definition 1.14)

$\Phi$ – Hermite design matrix $\Phi_{(i,k),j} = \psi_j^\sigma(x_i^H)$ (Equation (1.15))

$c$ – Hermite coefficient vector $c = (c_0, \dots, c_{N_\varphi-1})^\top \in \mathbb{R}^{N_\varphi}$, with $\|c\|_\Lambda^2 = c^\top \Lambda c$ (Definition 1.14)

$\varphi^c$ – Truncated Hermite source expansion $\varphi^c = \sum_{j=0}^{N_\varphi-1} c_j \psi_j^\sigma$ (Definition 1.14)

$C_{\sigma,\ell}$ – Hermite-to-Sobolev norm-bound constant; $\|\varphi^c\|_{H^\ell(\mathbb{R})} \le C_{\sigma,\ell}\|c\|_\Lambda$ (Lemma 4.2)

$\operatorname{cond}(\cdot)$ – Spectral condition number of a matrix; used to compare $\mathbf{G}$ with the regularised $\mathbf{G} + \lambda\Lambda$ (Equation (4.94))

$V(\delta)$ – Value of the relaxed moment problem over all finite nonnegative Borel measures; equals $\omega(\delta, \mathcal{W}_{\ell,E})^2$ (Equation (4.57), Section 4.5.6)

***Numerical discretisation***

---

$\Delta x$, $\Delta t$ – Spatial and temporal grid spacings (Section 1.4.2)

$N_x$, $N_t$ – Numbers of spatial nodes and time steps (Remark 1.21)

$N_{\mathrm{GH}}$, $N_{\mathrm{GL}}$, $M_r$ – Numbers of Gauss–Hermite and Gauss–Legendre quadrature nodes, and the resulting count $M_r = N_{\mathrm{GH}}\,N_{\mathrm{GL}}$ of space–time residual nodes (Definition 1.14, Section 1.4.3)

$r$ – Mesh ratio $\Delta t/\Delta x^2$ governing the explicit-scheme (CFL) stability bound $r \le \frac{1}{2}$ (Equation (1.23), Equation (1.24))

FTCS – Forward-time centred-space explicit finite-difference scheme (Equation (1.23))

Crank Nicolson – Unconditionally stable, second-order implicit finite-difference scheme ($\theta$-method, $\theta = \frac{1}{2}$) (Equation (1.25))

Lax–Richtmyer theorem – For a consistent scheme, stability $\Leftrightarrow$ convergence (Proposition 1.20)

Gauss–Hermite, Gauss–Legendre quadrature – Quadrature rules on $\mathbb{R}$ and on a bounded interval, used to discretise the variational loss (Definition 1.22, Definition 1.23)

$x_i^H$, $w_i^H$ – Gauss–Hermite quadrature nodes (roots of $H_{N_{\mathrm{GH}}}$) and their positive weights (Definition 1.22)

$t_k^L$, $w_k^L$ – Gauss–Legendre quadrature nodes on $[0,1]$ and their positive weights (Definition 1.23)

$M_d$ – Number of terminal-time measurement sensors entering the data-misfit term ($M_d = 540$ in Chapter 4) (Equation (4.19))

spectral, finite-element methods – Global-basis and piecewise-polynomial discretisations reviewed as numerical background

## Chapter 2 · Inverse Problems and Regularization Theory

well-posed / ill-posed – Hadamard's trichotomy: existence, uniqueness, and continuous dependence on the data (Definition 2.1)

$g$, $g^\delta$ – Exact data $g = A\varphi^\dagger$ and noisy data $g^\delta = g + \delta\eta$ (Equation (2.1))

$\delta > 0$ – Noise level; deterministic bound on the data perturbation (Equation (2.1))

$\eta$ – Noise function, normalised by $\|\eta\|_{L^2(\mathbb{R})} \le 1$ (Equation (2.1))

$\varphi^\dagger$ – True (unknown) source to be reconstructed; equivalently the minimum-norm solution of $A\varphi = g$ (Equation (2.1))

$\mathcal{W}_{\ell,E}$ – A-priori source class $\{\varphi \in H^\ell(\mathbb{R}) : \|\varphi\|_{H^\ell(\mathbb{R})} \le E\}$ (Definition 1.8)

$E$ – Source-ball radius (a-priori energy bound), fixed throughout (Definition 1.8)

$\ell$ – Sobolev regularity index of the source class, fixed throughout (Definition 1.8)

$A^*$ – Hilbert-space adjoint of the forward operator; Fourier symbol $\overline{m(\xi)}$ (Equation (2.4))

$(\sigma_n; u_n, v_n)$ – Singular system of a compact operator: singular values and vectors (Equation (2.2))

Picard criterion – Solvability condition $\sum_n \sigma_n^{-2}|\langle g, v_n\rangle|^2 < \infty$ (Equation (2.2))

$\alpha > 0$ – Regularization parameter

$R_\alpha$ – Regularization operator; bounded approximate inverse of $A$

$r_\alpha(\cdot)$ – Filter function generating $R_\alpha$ through the spectral calculus

$\lambda > 0$ – Tikhonov penalty weight (Definition 2.7)

$J_\lambda$, Tikhonov functional – $\|A\varphi - g^\delta\|_{L^2(\mathbb{R})}^2 + \lambda\|\varphi\|_{H^\ell(\mathbb{R})}^2$, minimised by the regularised solution (Definition 2.7, Equation (2.3))

$\varphi_\lambda^\delta$ – Tikhonov-regularised solution; the unique minimiser of $J_\lambda$ (Proposition 2.8, Equation (2.4))

$\mathcal{L}^\ell$ – $H^\ell(\mathbb{R})$-regularizer $(I - \partial_{xx})^\ell$ appearing in the normal equation (Equation (2.4))

$q_\lambda(\xi)$, Tikhonov filter (Fourier) – Regularization filter $\hat\varphi_\lambda^\delta = q_\lambda\,\hat g^\delta$, $q_\lambda = \frac{\overline{m}}{|m|^2 + \lambda(1+\xi^2)^\ell}$ (Equation (2.5))

degree of ill-posedness – Order of the polynomial decay of the singular values, equivalently of the symbol (Definition 2.5)

$A_L$ – Bounded-domain truncation of the forward operator, to which the compact singular-system

theory is applied (Remark 2.6)

$\omega(\delta, \mathcal{W}_{\ell,E})$ – Modulus of continuity of the inverse problem on the source class

conditional stability – Continuous dependence restored on the a-priori class $\mathcal{W}_{\ell,E}$ (Theorem 2.12)

Hölder stability – Conditional-stability estimate of Hölder type (Theorem 2.12)

$\lambda^\star(\delta)$ – Order-optimal a-priori Tikhonov parameter $\frac{1}{2}(\delta/E)^2 \asymp \delta^2$ (Equation (2.8), Proposition 2.13)

Hilbert scale – Family of spaces interpolating Sobolev regularity; the setting for the optimal rate $\delta^{\ell/(\ell+2)}$

## Chapter 3 · Neural Networks and the Hermite Variational PINN

$\mathcal{N}_\theta$ – Feed-forward neural network with hyperbolic-tangent activations (Definition 3.1)

tanh – Hyperbolic-tangent activation function (Definition 3.1)

$D$, $W$ – Network depth (number of layers) and width (neurons per layer) (Definition 3.1)

$W^{(k)}$, $b^{(k)}$ – Weight matrix and bias vector of layer $k$ (Equation (3.1))

$\theta \in \mathbb{R}^P$ – Concatenated network parameter vector; $P$ is the parameter count (Definition 3.1)

$\theta_0$ – Default (Kaiming-uniform, $a = \sqrt{5}$) initialisation point of the network parameters (Definition 3.2)

PINN, V-PINN – Physics-informed neural network and its variational (weak-form) counterpart

$\varphi^c$ – Hermite trial source $\sum_{j=0}^{N_\varphi - 1} c_j \psi_j^\sigma$ (Equation (3.5))

$c \in \mathbb{R}^{N_\varphi}$ – Vector of Hermite source coefficients (Equation (3.5))

$v_\theta(x,t)$ – Trial state $t\, e^{-\alpha_v x^2} \mathcal{N}_\theta(x,t)$; the factor $t$ enforces $v(\,\cdot\,,0) = 0$ (Definition 3.4)

$\alpha_v$ – Decay rate of the Gaussian envelope of the trial state (Equation (3.6))

$R_{\mathrm{PDE}}(\theta, c)$ – PDE residual $\partial_t v_\theta - \partial_{xx} v_\theta - \varphi^c$ (Definition 3.5)

$\mathcal{G}_\theta$ – Residual generator $\partial_t v_\theta - \partial_{xx} v_\theta$; $R_{\mathrm{PDE}} = \mathcal{G}_\theta - \varphi^c$ (Equation (3.7))

test functions – Finite-dimensional functions against which the weak PDE residual is tested

$\mu$ – Relative weight of the PDE-residual term in the V-PINN loss (Definition 3.6)

$\mathcal{L}$ – Continuous (integral) V-PINN loss functional (Definition 3.6)

$L$ – Discrete (quadrature) V-PINN loss (Equation (3.10))

$J$ – Reduced loss after eliminating $c$, $J(\theta) = L(\theta, c^\star(\theta))$ (Equation (3.13))

$c^\star(\theta)$ – Inner minimiser; optimal coefficient vector for fixed $\theta$ (Proposition 3.7)

$\nabla J(\theta)$ – Reduced-loss gradient of the eliminated outer problem (Section 3.3.5)

$\Gamma$ – Regularised Gram matrix of the inner least-squares problem (Equation (3.12))

$\mathbf{g}_\theta$ – Residual generator sampled at the $M_r$ residual nodes, $\big(\mathcal{G}_\theta(x_i^H, t_k^L)\big) \in \mathbb{R}^{M_r}$ (Equation (3.8))

$\mathbf{v}_\theta$, $\mathbf{g}^\delta$ – Trial state and noisy data sampled at the $M_d$ data sensors (Equation (3.11))

$\mathbf{W}_d$, $\mathbf{W}_r$ – Diagonal data-averaging weight $\frac{1}{M_d} I_{M_d}$ and residual quadrature-weight matrix $\mathrm{diag}(w_i^H w_k^L / M_r)$ (Equation (3.11))

## Chapter 4 · The Hermite V-PINN for the Inverse Heat Source Problem

### *Forward operator and data generation*

$A_\varepsilon$ – Perturbed forward operator generating inverse-crime-free data (Equation (4.7), Equation (4.8))

$a(x)$ – Perturbed diffusion coefficient $1 + \varepsilon\, e^{-x^2/4}$ (Equation (4.7))

$\varepsilon$ – Anti-inverse-crime perturbation amplitude (Section 4.1.5)

$g^\delta$ – Noisy terminal datum, model $g^\delta = A\varphi^\dagger + \delta\eta$ (Equation (4.4))

$\varphi^\dagger$ – Smooth antisymmetric test source $(x^3/4 - 3x/2)\, e^{-x^2/4}$ (Equation (4.5))

### *Discrete loss, weights, and quadrature*

$L_{\mathrm{d}}$, $L_{\mathrm{c}}$, $L_{\mathrm{w}}$, $L_{\mathrm{b}}$ – Data-misfit, strong-form collocation, weak-form residual, and far-field boundary loss terms (Equation (4.19)–Equation (4.23))

$L$ – Hybrid discrete V-PINN loss combining the four terms with a Tikhonov penalty (Equation (4.18))

$w_{\mathrm{m}}, \mu_{\mathrm{coll}}, \mu_{\mathrm{wf}}, w_{\mathrm{b}}$ – Data, strong-form, weak-form, and boundary loss weights (Equation (4.18), Equation (4.26))

$\lambda$ – Hermite-coefficient (Tikhonov) penalty weight (Equation (4.18))

$\lambda(\delta)$, $w_{\mathrm{m}}(\delta)$ – Noise-adaptive weight schedules (Equation (4.28))

$\rho$ – Penalty-to-residual weight ratio $\rho = \lambda/\mu$ (Theorem C)

$r_n$ – Linear warm-in ramp $\min(1, n/n_{\mathrm{ramp}})$ on the strong- and weak-form terms during joint training (Algorithm 1)

$\Gamma(\mu_{\mathrm{coll}}, \mu_{\mathrm{wf}}, \lambda)$ – Three-weight regularised Gram matrix $\mu_{\mathrm{coll}}\mathbf{H}_{\mathrm{coll}} + \mu_{\mathrm{wf}}\mathbf{H}_{\mathrm{wf}} + \lambda\Lambda$; the $c$-Hessian of $L$ (Equation (4.17))

$\mathbf{H}_{\mathrm{coll}}$, $\mathbf{H}_{\mathrm{wf}}$ – Collocation and weak-form Gram blocks (Equation (4.17))

$\mathbf{G}$ – $\mathbf{W}_r$-weighted Hermite Gram matrix at the residual nodes, $\mathbf{G} = \Phi^\top \mathbf{W}_r \Phi$; the combined residual block of $\Gamma$ (Theorem C)

$f_k(\rho)$, $\gamma_k$ – Tikhonov filter factors $\gamma_k/(\gamma_k + \rho)$ and the generalised eigenvalues of the matrix pair $(\mathbf{G}, \Lambda)$ (Theorem C)

$\mathbf{G}_{\mathrm{hyb}}$, $y_{\mathrm{hyb}}$ – Hybrid two-block Gram matrix $\mu_{\mathrm{coll}}\mathbf{H}_{\mathrm{coll}} + \mu_{\mathrm{wf}}\mathbf{H}_{\mathrm{wf}}$ and its stacked weighted data vector (Remark 4.23)

$\vartheta$ – A-posteriori loss threshold, $\mathcal{L} \le \vartheta^2$, read off the training log (Proposition 4.25)

$N_{\mathrm{GH}}$, $N_{\mathrm{GL}}$ – Gauss–Hermite and Gauss-Legendre quadrature orders (Definition 1.22, Definition 1.23)

$\sigma_q$ – Length-scale of the Gauss-Hermite spatial quadrature nodes, $\sigma_q = 2.5$ (Equation (4.15))

$K$ – Number of spatial Hermite test functions spanning $\mathbb{V}_K$ (Equation (4.13))

$\Pi$ – Orthogonal projector onto the spatial test space

$\mathbb{V}_K$ (Equation (4.13))

$M_d$ – Number of training-sensor data nodes entering the misfit ($M_d = 540$), distinct from the quadrature nodes (Equation (4.19))

$N_{\text{coll}}$, $L_{\text{box}}$ – Number (8192) and half-width (8) of the Monte-Carlo strong-form collocation points on $[-L_{\text{box}}, L_{\text{box}}] \times (0,1)$ (Equation (4.20))

$\mathbb{V}_K$, $\mathbb{Q}_{M_L}$ – Spatial Hermite ($\dim K$) and temporal ($\dim M_L$) test spaces (Equation (4.13))

$b_{k,m}$ – Weak-form (Galerkin) residual coefficient: $R_{\text{PDE}}$ integrated against $\psi_k^\sigma \otimes q_m$ (Equation (4.21))

$q_m$, $\tau$, $\{c_m\}$ – Temporal test profiles (one constant and Gaussian bumps of width $\tau = 0.05$ centred at $\{c_m\}$) spanning $\mathbb{Q}_{M_L}$ (Equation (4.14))

$\Pi$ – Orthogonal projection onto the finite-dimensional test space

***Spectral structure and sharp stability***

$\sigma(A)$ – Spectrum of the forward operator; $\sigma(A) = [0,1]$ (Theorem 4.6)

$\sigma_{\text{pp}}, \sigma_{\text{sc}}, \sigma_{\text{ac}}$ – Point, singular-continuous, and absolutely-continuous spectrum of $A$; here $\sigma_{\text{pp}} = \sigma_{\text{sc}} = \varnothing$, $\sigma_{\text{ac}} = [0,1]$ (Theorem 4.6)

$M_m$ – Multiplication operator by the symbol $m$ on $L^2(\mathbb{R}, d\xi)$; $A = \mathcal{F}^{-1} M_m \mathcal{F}$ (Theorem 4.4)

$w_n^{\text{hi}}$ – High-frequency Weyl sequence realising the sharp constant (Definition 4.15)

$C^\star$ – Sharp interpolation / stability constant, $C^\star = 1$ (Theorem 4.10, Theorem 4.11)

$h(\xi)$ – Spectral-balance function in the sharp interpolation inequality (Theorem 4.10)

degree of ill-posedness – Equal to 2 for the present problem (Corollary 4.9)

$\alpha$ – Hölder / minimax exponent $\ell/(\ell+2)$ (Theorem 4.11, Theorem A)

$K_n^\star$ – Optimal Hermite-truncation level $\asymp (E/\delta)^{1/(\ell+2)}$ (Equation (4.60))

$\omega$ – Modulus of continuity on the source class $\mathcal{W}_{\ell,E}$; computed in closed form at every noise level (Definition 4.14, Theorem A)

$\xi^\star$ – Balance frequency solving $m(\xi^\star)/(1+\xi^{\star 2})^{\ell/2} = \delta/(2E)$ (Equation (4.54))

$K_\star$ – Leading-order form of $\xi^\star(\delta)$ as $\delta \downarrow 0$, $K_\star = (2E/\delta)^{1/(\ell+2)}(1 + \mathcal{O}(\delta^{2/(\ell+2)}))$ (Equation (4.60))

$e_{\text{mm}}$ – Minimax-optimal reconstruction error over $\mathcal{W}_{\ell,E}$; evaluated exactly (Theorem B)

$t$, $P(t)$, $Q(t)$ – Spectral coordinates of the moment problem: variable $t = \xi^2$, Sobolev weight $P = (1+t)^\ell$, squared symbol $Q = m^2$ (Proposition 4.17)

$w_\triangle$ – Triangular Laplace density of $m^2$, $w_\triangle(r) = \min\{r, 2-r\}$ (Equation (4.52))

$a^\star$, $b^\star$, $\varkappa$ – Tangent-line dual-certificate coefficients, $a^\star P + b^\star Q \geq 1$ with equality only at the balance point (Proposition 4.17)

$\Omega(\delta)$ – Centred modulus over $\mathcal{W}_{\ell,E}$ itself; $\Omega(\delta) = \frac{1}{2}\,\omega(2\delta, \mathcal{W}_{\ell,E})$ (Equation (4.62))

$\xi^\circ$ – Centred balance frequency of the exact minimax error, $m(\xi^\circ)/(1+(\xi^\circ)^2)^{\ell/2} = \delta/E$ (Theorem B)

$t_\sharp$, $c_\sharp$ – Maximiser and value of $\sup_\xi m(\xi)(1+\xi^2)$: $e^{t_\sharp} = 1 + t_\sharp + t_\sharp^2$, $c_\sharp = \|A\|_{H^s(\mathbb{R}) \to H^{s+2}(\mathbb{R})} \approx 1.2984$ (Proposition 4.5)

$A_\tau$, $m_\tau$ – Time-$\tau$ observation operator and its symbol, $m_\tau(0) = \tau$ (Equation (4.68))

$A_\oplus$, $m_\oplus$, $\|\cdot\|_\oplus$ – Stacked multi-snapshot operator, its symbol $(\sum_j m_{t_j}^2)^{1/2}$, and the product norm (Equation (4.69))

$J$, $T_\oplus$ – Number of observation times (local use in Section 4.4.4, distinct from the reduced loss $J(\theta)$) and $T_\oplus = (\sum_j t_j^2)^{1/2}$ (Theorem 4.22)

$\omega_\oplus$, $\xi_\oplus^\star$ – Multi-snapshot modulus under an aggregate noise budget and its balance frequency; obeys the exact $\sqrt{J}$ law (Theorem 4.22)

***Numerical verification***

$D, W, \alpha_v$ – Deployed network depth 7, width 192, envelope rate 0.08

$\ell_{\text{eff}}$ – Effective source regularity $2\hat{\alpha}/(1-\hat{\alpha})$ governing the observed convergence rate; the realised index $s$ of Theorem C (Remark 4.24)

$\hat{\alpha}$, $\hat{C}$, $R^2$ – Empirical Hölder slope, prefactor, and coefficient of determination (Equation (4.85))

$c_0$ – Noise-free relative error floor $c_0 = \varphi_{\text{rel}}(0) = 4.35 \times 10^{-3}$ subtracted in the Hölder fit; empirically set by the operator (model) mismatch rather than optimisation or quadrature (Equation (4.85))

$c^\dagger$ – Best-approximation coefficient vector of $\varphi^\dagger$ in the Hermite trial space

$\varepsilon_{\text{app}}$ – Network approximation error for the trial state, the term the quantitative approximation theorem of Section 3.2 bounds and drives to zero as the width grows (Chapter 3)

$\varphi_{\text{rel}}$, $v_{\text{rel}}$ – Relative discrete $L^2(\mathbb{R})$ errors of the recovered source and state, evaluated on the $N_{\text{eval}}$-point grid (Equation (4.81))

$\|\cdot\|_\Delta$ – Discrete $L^2(\mathbb{R})$ norm $\|f\|_\Delta^2 = \Delta x \sum_n f(x_n)^2$ on the evaluation grid (Equation (4.81))

$N_{\text{sensor}}$ – Total measurement sensors $N_{\text{sensor}} = 600$ (540 training + 60 held-out), uniform on $[-10, 10]$ (Table 4.1)

$N_{\text{eval}}$ – Uniform evaluation grid on $[-10, 10]$ for the error metrics ($N_{\text{eval}} = 800$, Equation (4.81))

seed – Fixed pseudo-random seed (42) and deterministic kernels; single-run baselines, multi-seed mean±std for the noise sweep

## Chapter 5 · Comparison with Two Classical Methods

***Mollification comparator***

$D_{1/\mu}(x)$ – Sharp Fourier-cutoff Dirichlet kernel $\sin(x/\mu)/(\pi x)$ (Equation (5.2))

$\beta_\mu(\xi)$ – Sharp low-pass projector $\chi_{|\xi|<1/\mu}$, the Fourier transform of $D_{1/\mu}$ (Equation (5.3))

$\mu$ – Mollification width; the cutoff frequency is $1/\mu$

$\mu^\star_{\text{apri}}$ – A-priori choice of $\mu$ from the explicit rate formula (Equation (5.5))

$\mu^\star_{\text{Mor}}$ – A-posteriori choice of $\mu$ by Morozov's discrepancy principle (Equation (5.6))

$\tau$ – Discrepancy-principle safety factor ($\tau = 1.5$)

$\mathcal{F}_N$ – Discrete Fourier transform on the sensor grid; with the trapezoidal weight it approximates the continuous transform of Definition 1.9

(Equation (5.8))

***Adjoint-state comparator***

$\Psi^{\sigma_{\mathrm{adj}}}$ – Hermite design matrix of the adjoint ansatz $\varphi^c = \Psi^{\sigma_{\mathrm{adj}}} c$, in the trial-source basis of Definition 1.13 at scale $\sigma_{\mathrm{adj}}$ (Equation (5.10))

$N_\Psi$ – Number of Hermite modes in the adjoint ansatz ($N_\Psi = 64$) (Equation (5.10))

$\sigma_{\mathrm{adj}}$ – Hermite length-scale of the adjoint ansatz ($\sigma_{\mathrm{adj}} = 2.5$) (Equation (5.10))

$\Omega_{\mathrm{adj}}$, $L_{\mathrm{adj}}$ – Bounded forward domain $[-L_{\mathrm{adj}}, L_{\mathrm{adj}}]$ of the adjoint solver ($L_{\mathrm{adj}} = 10$) (Equation (5.11))

$F_h(c)$ – Discrete Crank Nicolson forward map $c \mapsto v_h^c(\cdot, 1)$ (Equation (5.11))

$S$ – Linear-interpolation observation operator sampling the forward state at the raw sensor abscissae (Equation (5.12))

$D$ – Diagonal Sobolev penalty matrix $\mathrm{diag}\big((1+n)^{2s}\big)$, $s = 1$ (Equation (5.12))

$J_\lambda(c)$ – Regularised least-squares functional $\frac{1}{2}\|SF_h(c) - g^\delta\|^2 + \frac{\lambda}{2} c^\top D c$, minimised by L-BFGS-B (Equation (5.12))

$\eta_h$ – Discrete adjoint state; solves the backward heat equation supplying the gradient (Equation (5.13), Equation (5.14))

$\lambda_{\mathrm{adj}}$ – Sobolev Tikhonov (Hilbert-scale) penalty weight of the adjoint method (Equation (5.15))

***Experimental parameters and error metrics***

$N_{\mathrm{sens}}$ – Number of sensors sampling the terminal datum ($N_{\mathrm{sens}} = 600$, the total sensor count of Table 4.1) (Section 5.1.2)

$L_{\mathrm{obs}}$ – Half-width of the spatial observation window ($L_{\mathrm{obs}} = 10$) (Section 5.1.2)

$\kappa$ – Diffusivity coefficient (unknown-diffusivity experiment)

$\ell_{\mathrm{c}}$ – Correlation length of the coloured-noise model

$\sigma_{\mathrm{vpinn}}$ – Hermite length-scale of the deployed V-PINN ($\sigma_{\mathrm{vpinn}} = 2$, the $\sigma$ of Table 4.1)

$\varphi^{\triangle}$ – Non-smooth triangular test source $\max(5 - |x|, 0)$ on $[-5, 5]$ (Equation (5.16))

$\varphi_{\mathrm{abs}}$, $\varphi_{\mathrm{rel}}$ – Absolute and relative discrete $L^2(\mathbb{R})$ errors of the recovered source

$v_{\mathrm{abs}}$, $v_{\mathrm{rel}}$ – Absolute and relative discrete $L^2(\mathbb{R})$ errors of the recovered state

$\hat{\alpha}$, $R^2$ – Empirical Hölder slope and fit quality of an error-noise sweep

# Appendix B

# Figures

## B.1 Experiment 0: Standard Noise Sweep

Standard noise sweep of Section 5.3.1; the V-PINN's own Hölder fit is Figure 4.6.

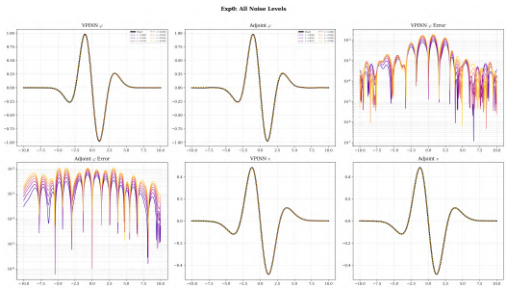

**Figure B.1:** Experiment 0: V-PINN versus adjoint across the noise sweep $\delta \in [0.008, 0.040]$ (source and state, with pointwise errors). Companion to Table 5.1.

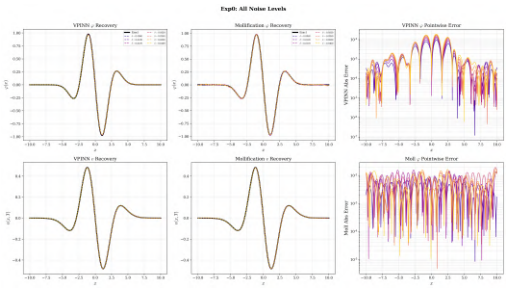

**Figure B.2:** Experiment 0: V-PINN versus mollification across the same sweep; the mollifier error is uniformly larger at every $\delta$.

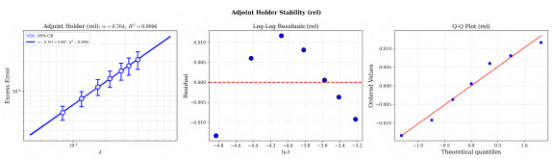

**Figure B.3:** Empirical Hölder fit of the adjoint (relative error): near-perfect law ($R^2 = 0.9996$), slope above the 2/3 saturation.

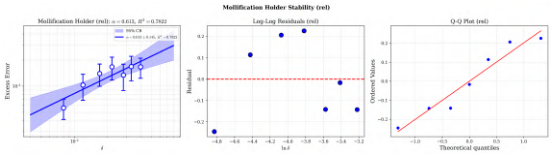


**Figure B.4:** Empirical Hölder fit of the mollifier: slope $\hat{\alpha} = 0.613$ below $2/3$, ragged fit ($R^2 = 0.782$).

## B.2 Experiment 1: Sparse Sensors

The sparse-sensor study of Section 5.3.2.

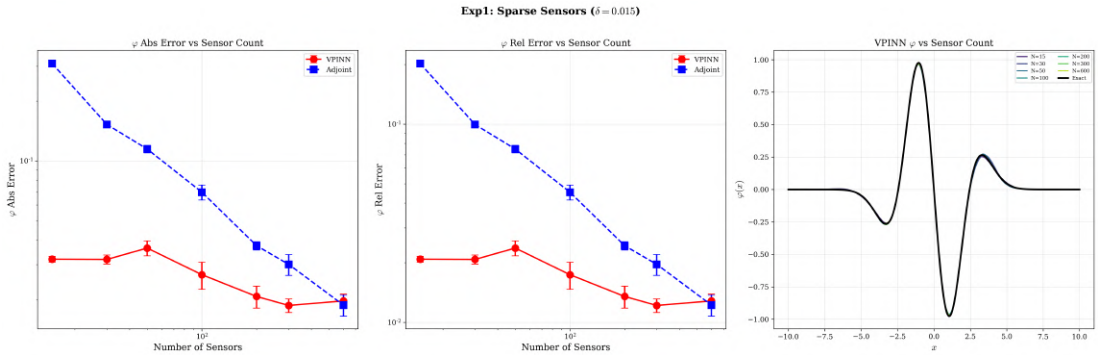


**Figure B.5:** Experiment 1: V-PINN versus adjoint source error against sensor count $N_{\text{sens}}$ at $\delta = 0.015$ (log-log, mean±std), with V-PINN recoveries. The V-PINN is nearly flat while the adjoint deteriorates as the data thin. Companion to Table 5.2.

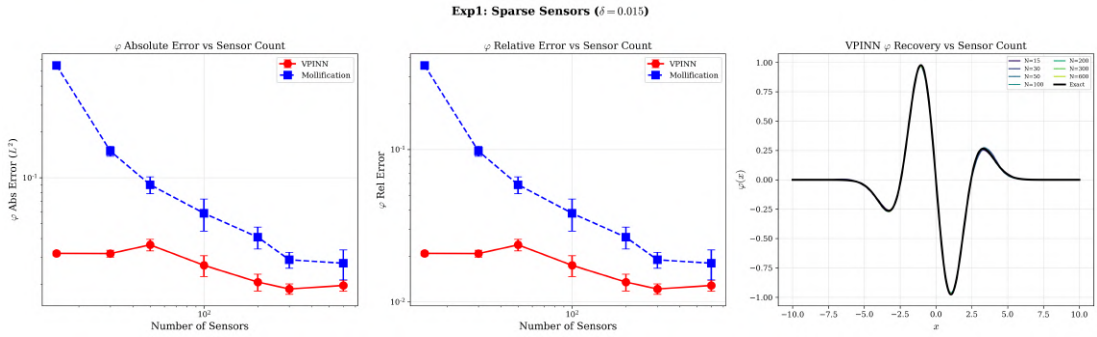


**Figure B.6:** Experiment 1: V-PINN versus mollification, the same data-efficiency frontier. The FFT-based mollifier implicitly demands a dense uniform grid and is an order of magnitude worse at small $N_{\text{sens}}$.

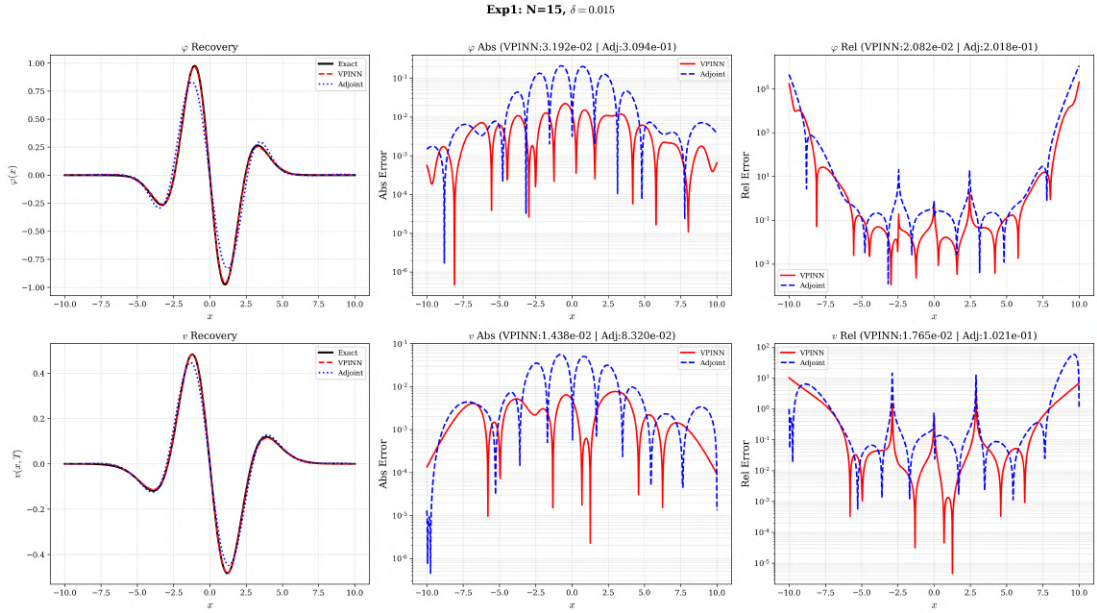


**Figure B.7:** Experiment 1 at the sparsest setting $N_{\text{sens}} = 15$: the V-PINN holds $\varphi_{\text{rel}} = 2.08 \times 10^{-2}$, while interpolating fifteen readings onto the adjoint's 500-node grid degrades it to $2.02 \times 10^{-1}$.

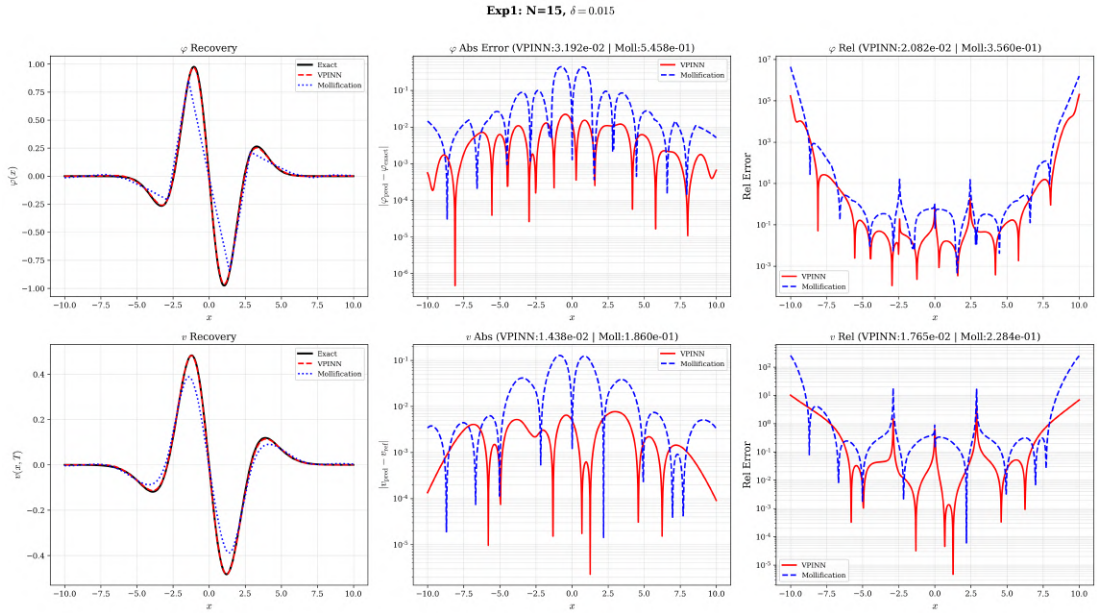


**Figure B.8:** Experiment 1 at $N_{\text{sens}} = 15$: the mollifier returns a badly distorted source ($\varphi_{\text{rel}} = 3.56 \times 10^{-1}$) while the V-PINN is essentially unaffected.

## B.3 Experiment 2: Partial-Domain Observation

The partial-domain study of Section 5.3.3.

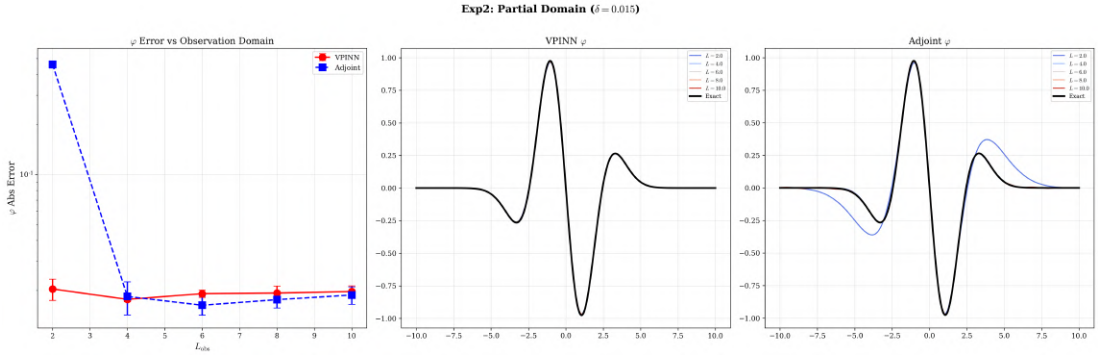


**Figure B.9:** Experiment 2: V-PINN versus adjoint source error against the observation half-width $L_{\text{obs}}$ at $\delta = 0.015$, with recoveries. The V-PINN is nearly insensitive to the window; the adjoint matches it for $L_{\text{obs}} \geq 4$ but breaks at $L_{\text{obs}} = 2$. Companion to Table 5.3.

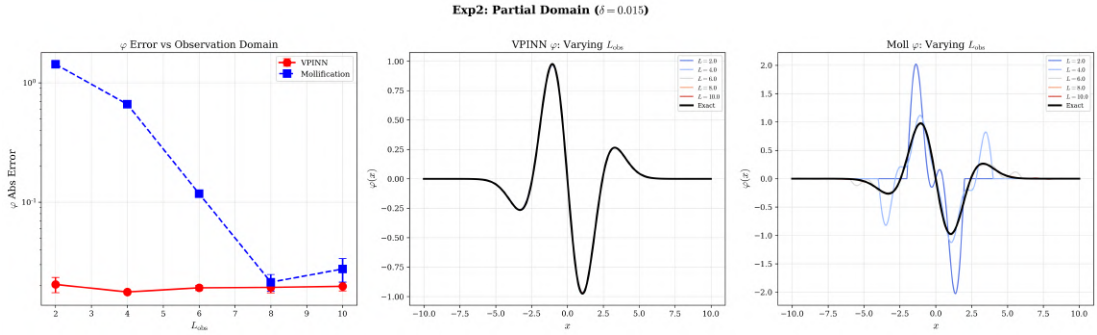


**Figure B.10:** Experiment 2: V-PINN versus mollification under the same windowing; the mollifier's global spectral inversion cannot recover a source it never observes, its error growing by over an order of magnitude as $L_{\text{obs}}$ shrinks.

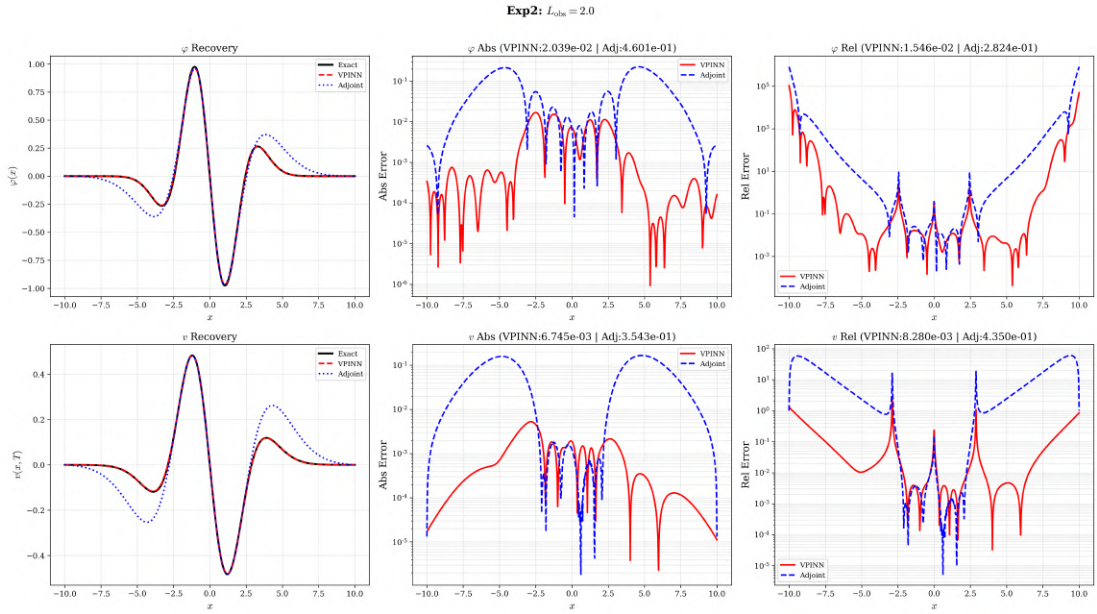


**Figure B.11:** Experiment 2 at the severe truncation $L_{\text{obs}} = 2$: the adjoint overshoots the unobserved tails ($\varphi_{\text{rel}} = 2.82 \times 10^{-1}$), whereas the V-PINN extrapolates and holds $1.55 \times 10^{-2}$.

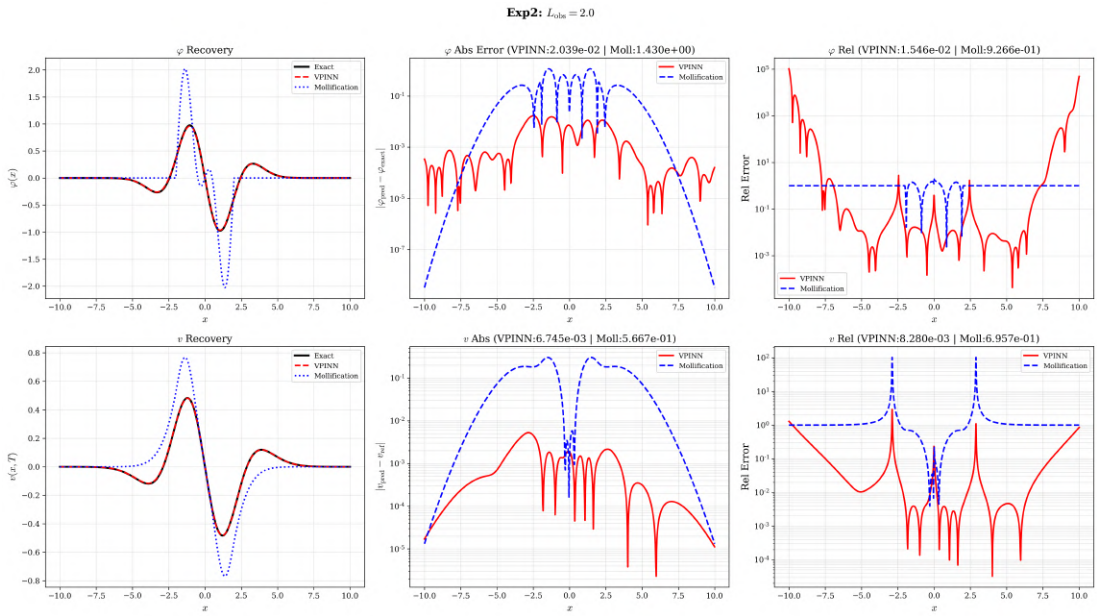


**Figure B.12:** Experiment 2 at $L_{\text{obs}} = 2$: global inversion of a windowed datum produces large spurious oscillations (mollifier $\varphi_{\text{rel}} = 9.27 \times 10^{-1}$); the V-PINN remains accurate.

## B.4 Experiment 3: Unknown Diffusivity

The unknown-diffusivity study of Section 5.3.4.

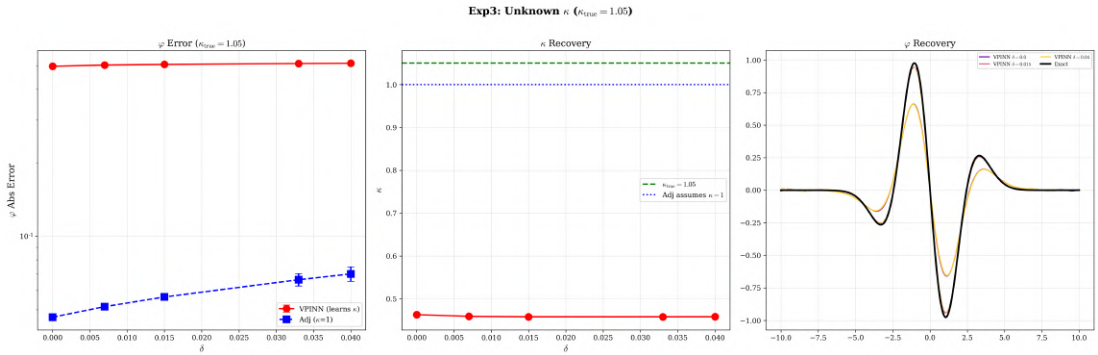


**Figure B.13:** Experiment 3 ($\kappa_{\text{true}} = 1.05$): V-PINN versus adjoint. *Centre:* the V-PINN's learned $\kappa$ collapses to $\approx 0.46$ a single snapshot does not identify $(\kappa, \varphi)$ while the adjoint fixes $\kappa = 1$. *Right:* V-PINN recoveries. Companion to Table 5.4.

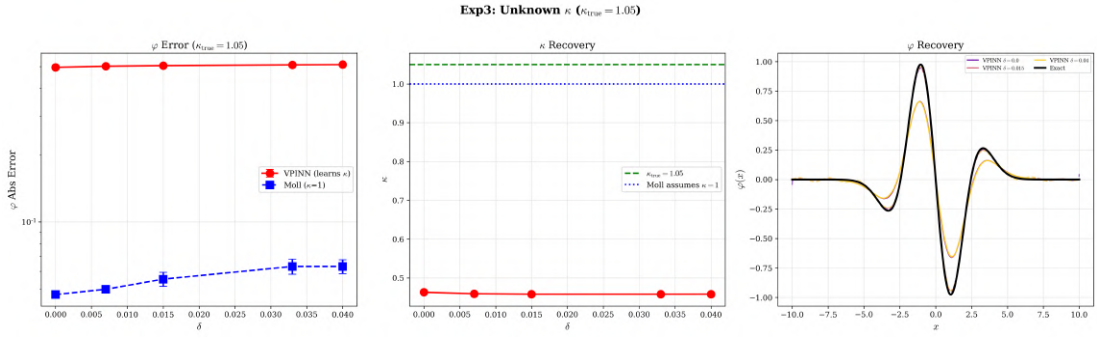


**Figure B.14:** Experiment 3: V-PINN versus mollification; as against the adjoint, the joint $(\kappa, \varphi)$ recovery is non-identifiable from one snapshot, whereas the fixed-$\kappa$ mollifier carries only the mild bias of the wrong constant.

## B.5 Experiment 4: Multiple Time Measurements

The multiple-time-measurement study of Section 5.3.5.

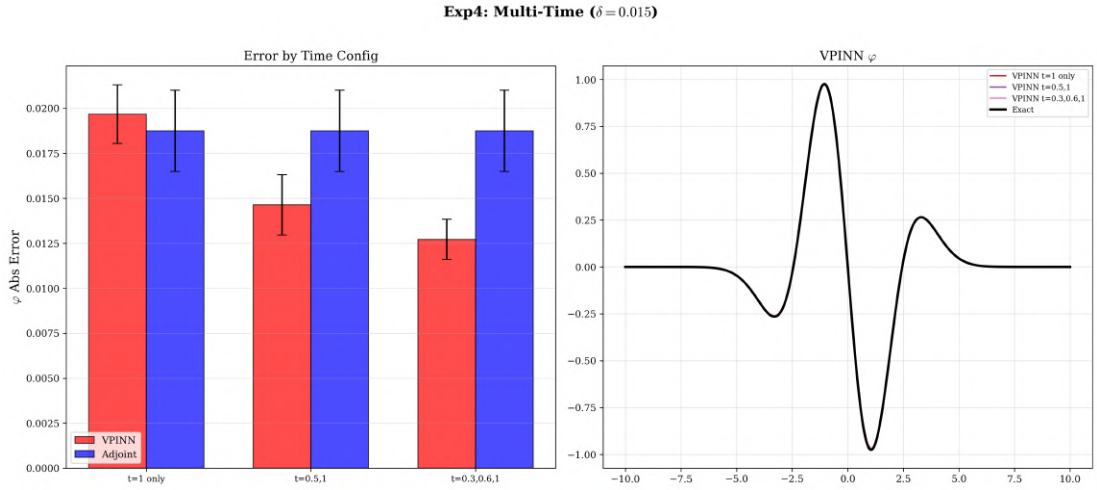


**Figure B.15:** Experiment 4 ($\delta = 0.015$): V-PINN versus adjoint. The V-PINN error falls as snapshots are added ($t$=1 → $t$=0.5, 1 → $t$=0.3, 0.6, 1), while the adjoint, tied to the terminal forward map, is flat. Companion to Table 5.5.

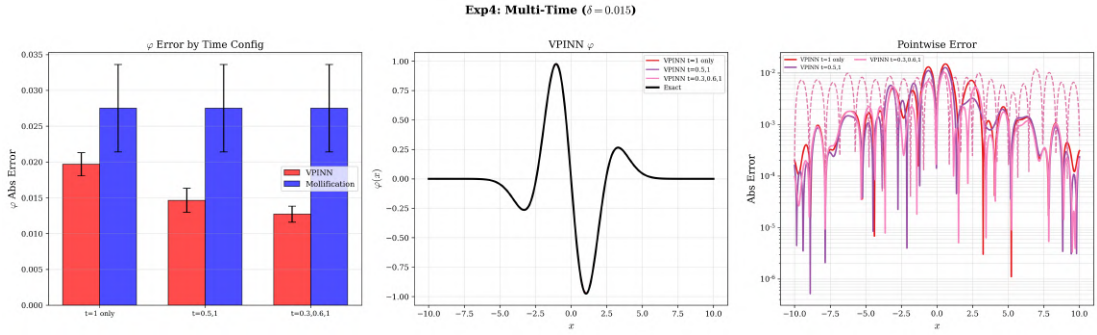


**Figure B.16:** Experiment 4: V-PINN versus mollification; the mollifier is anchored to the single terminal symbol and pinned at $2.75 \times 10^{-2}$, while the V-PINN converts the extra slices into accuracy.

## B.6 Experiment 5: Non-Smooth Source

The non-smooth (triangular) source of Section 5.3.6.

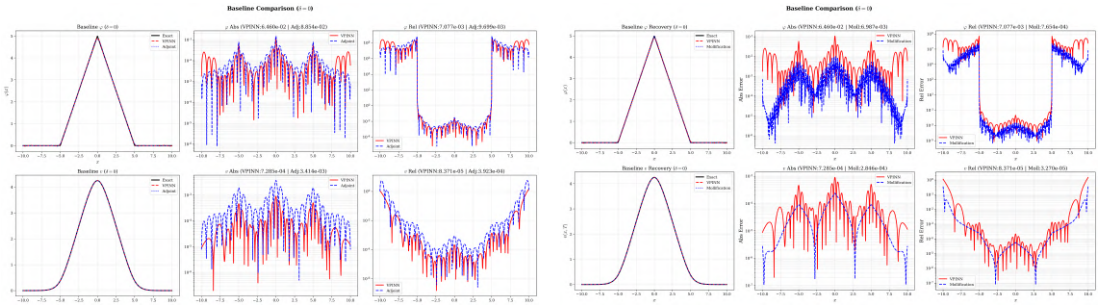

**Figure B.17:** Experiment 5 baseline ($\delta = 0$): V-PINN versus adjoint *(left)* and versus mollification *(right)*. The V-PINN beats the adjoint (6.46 versus $8.85 \times 10^{-2}$); the mollifier's cutoff resolves the corner best ($6.99 \times 10^{-3}$). Companion to Table 5.6.

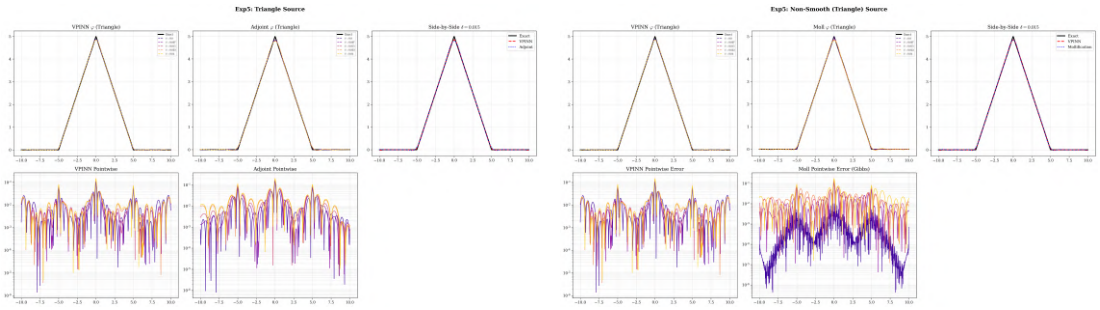

**Figure B.18:** Experiment 5, triangular source across noise: V-PINN versus adjoint *(left)* and versus mollification *(right)*. The V-PINN leads at every level; the mollifier shows Gibbs oscillations at the corner.

## B.7 Experiment 6: High-Noise Regime

The high-noise regime of Section 5.3.7, including the high-noise Hölder fits.

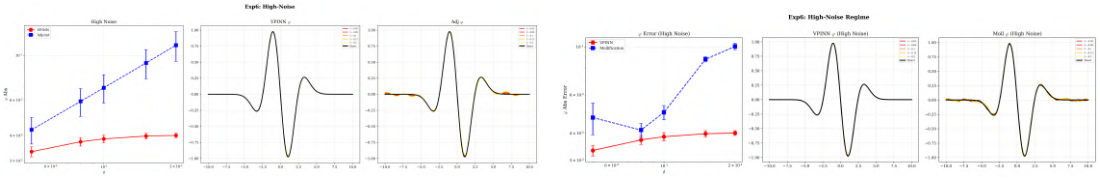

**Figure B.19:** Experiment 6, high noise $\delta \in [0.05, 0.20]$: V-PINN versus adjoint *(left)* and versus mollification *(right)*. The V-PINN saturates near $4 \times 10^{-2}$ while both comparators climb. Companion to Table 5.7.

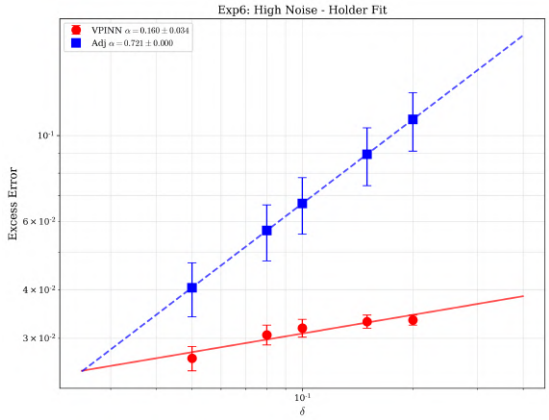


**Figure B.20:** Experiment 6 high-noise Hölder fit: V-PINN $\hat{\alpha} = 0.160$ versus adjoint 0.721.

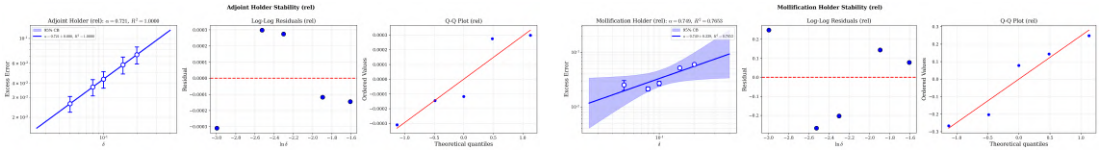

**Figure B.21:** Experiment 6 high-noise Hölder stability (relative error): adjoint *(left)*, $\hat{\alpha} = 0.721$, $R^2 = 1.0000$; mollifier *(right)*, $\hat{\alpha} = 0.749$, $R^2 = 0.765$.

## B.8 Mollification-Specific Experiments

Three contrasts unique to the mollification study (Section 5.3.8); the joint data-efficiency frontier is Figure B.27, grouped with the adjoint-specific figures.

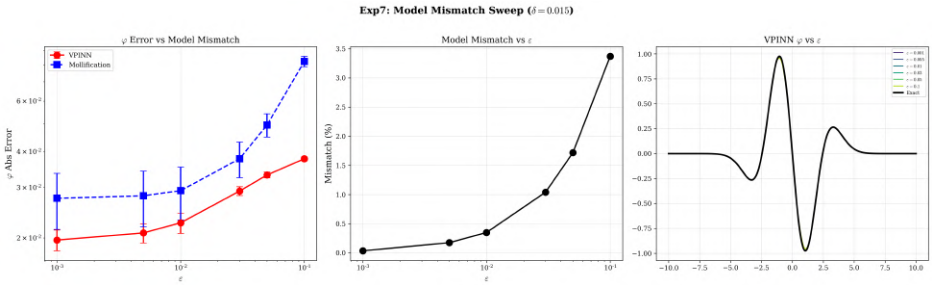


**Figure B.22:** Model-mismatch sweep ($\delta = 0.015$): V-PINN versus mollification source error against the perturbation $\varepsilon$ (left) and the induced data mismatch (centre). The V-PINN degrades sub-linearly. Companion to Table 5.8.

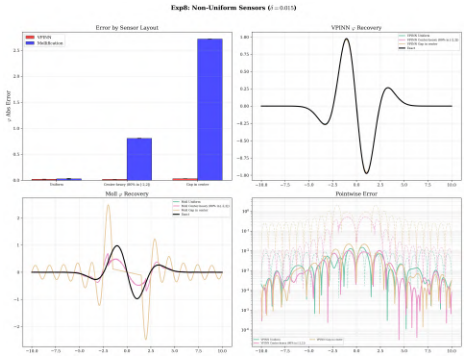

**Figure B.23:** Sensor geometry ($\delta = 0.015$): error by layout (top left), with V-PINN (top right) and mollifier (bottom) recoveries. A centre gap drives the mollifier to a relative error of 1.77. Companion to Table 5.9.

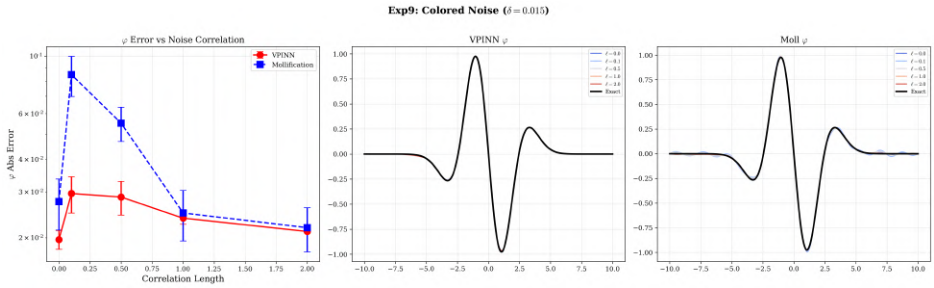


**Figure B.24:** Coloured noise ($\delta = 0.015$): source error against the correlation length $\ell_c$ (left), with recoveries. The mollifier is hit hardest at short correlation lengths. Companion to Table 5.10.

## B.9 Adjoint-Specific Experiments

Two contrasts unique to the adjoint study (Section 5.3.9) randomly scattered sensors and the adjoint's own grid-resolution sensitivity followed by the V-PINN-mollification data-efficiency frontier of Section 5.3.8.

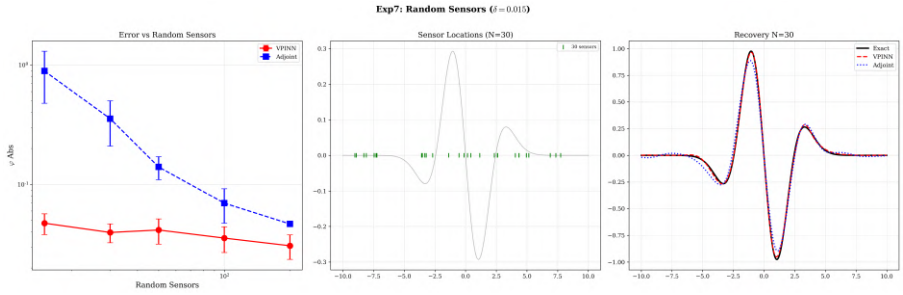


**Figure B.25:** Randomly scattered sensors ($\delta = 0.015$): V-PINN versus adjoint error against scattered-sensor count, a sample $N_{\text{sens}} = 30$ layout, and the recovery. The mesh-free V-PINN degrades gracefully. Companion to Table 5.12.

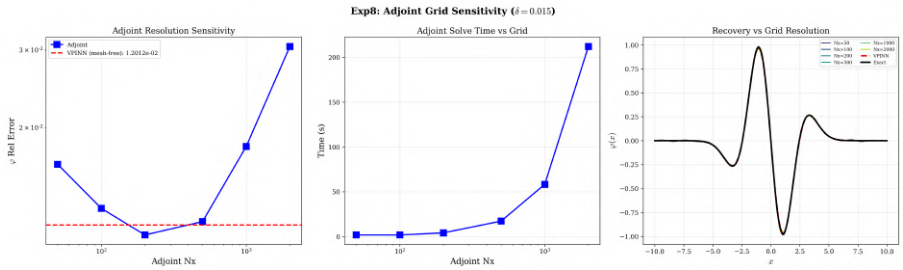


**Figure B.26:** Adjoint grid-resolution sensitivity ($\delta = 0.015$): relative error versus $N_x$ (optimum near 200), solve time, and recoveries, against the mesh-free V-PINN. Companion to Table 5.13.

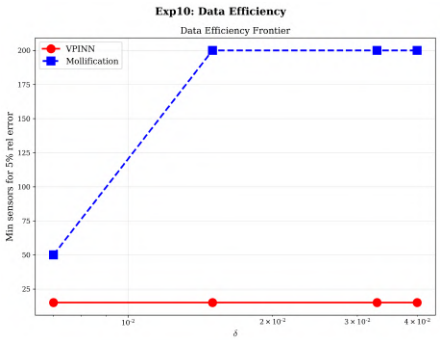


**Figure B.27:** Data-efficiency frontier (companion to Table 5.11): minimum sensors for 5% relative error versus $\delta$ (V-PINN vs mollification)-$\approx$ 15 for the V-PINN, 50 rising to 200 for the mollifier.