Linear methods for non-linear inverse problems

Geerten Koers (Delft University of Technology)

Botond Szabó (Bocconi University)
Aad van der Vaart (Delft University of Technology)

StatMathAppli 2022

2022-08-29

Outline

Inverse problems

- **Goal:** To recover an infinite-dimensional parameter f_0 from the observation

$$Y = \mathcal{A}f_0 + \epsilon,$$

with $\mathcal{A}$ a known, injective map and ϵ noise.
- If $\mathcal{A}^{-1}$ is not continuous, simply inverting Y may amplify noise.
- **Usual approach:** Model the data with a prior Π:

$$f \sim \Pi$$
$$Y \mid f \sim P_f$$

and then analyse the posterior $f \mid Y$.
- For linear $f \mapsto \mathcal{A}f$, there are good (asymptotic) results.

Inverse problems

- **Goal:** To recover an infinite-dimensional parameter f_0 from the observation

$$Y = \mathcal{A} f_0 + \epsilon,$$

with $\mathcal{A}$ a known, injective map and ϵ noise.

- If $\mathcal{A}^{-1}$ is not continuous, simply inverting Y may amplify noise.
- **Usual approach:** Model the data with a prior Π:

$$\begin{aligned} f &\sim \Pi \\ Y \mid f &\sim P_f \end{aligned}$$

and then analyse the posterior $f \mid Y$.

- For linear $f \mapsto \mathcal{A} f$, there are good (asymptotic) results.

Inverse problems

- **Goal:** To recover an infinite-dimensional parameter f_0 from the observation

$$Y = \mathcal{A}f_0 + \epsilon,$$

with $\mathcal{A}$ a known, injective map and ϵ noise.
- If $\mathcal{A}^{-1}$ is not continuous, simply inverting Y may amplify noise.
- **Usual approach:** Model the data with a prior Π:

$$f \sim \Pi$$
$$Y \mid f \sim P_f$$

and then analyse the posterior $f \mid Y$.
- For linear $f \mapsto \mathcal{A}f$, there are good (asymptotic) results.

Inverse problems

- **Goal:** To recover an infinite-dimensional parameter f_0 from the observation

$$Y = \mathcal{A}f_0 + \epsilon,$$

with $\mathcal{A}$ a known, injective map and ϵ noise.
- If $\mathcal{A}^{-1}$ is not continuous, simply inverting Y may amplify noise.
- **Usual approach:** Model the data with a prior Π:

$$f \sim \Pi$$
$$Y \mid f \sim P_f$$

and then analyse the posterior $f \mid Y$.
- For linear $f \mapsto \mathcal{A}f$, there are good (asymptotic) results.

Inverse problems

- **Setup:** We consider

$$\begin{aligned} \mathcal{L}u_f &= c(f, u_f) && \text{on } \mathcal{O}, \\ u_f &= g && \text{on } \Gamma \subseteq \partial\mathcal{O}, \end{aligned}$$

  with $\mathcal{L}$ a linear differential operator, and $c : \mathbb{R}^2 \to \mathbb{R}$ and $g : \partial\mathcal{O} \to \mathbb{R}$ are known functions.
- We assume we can invert:

$$\mathbf{f} = \mathbf{e}(\mathcal{L}\mathbf{u}_\mathbf{f}, \mathbf{u}_\mathbf{f})$$

  for a known map $e : \mathbb{R}^2 \to \mathbb{R}$.
- **Our approach:** Put a prior on $\mathcal{L}u_{f_0}$.

Inverse problems

- **Setup:** We consider

$$\begin{aligned} \mathcal{L} u_f &= c(f, u_f) && \text{on } \mathcal{O}, \\ u_f &= g && \text{on } \Gamma \subseteq \partial\mathcal{O}, \end{aligned}$$

  with $\mathcal{L}$ a linear differential operator, and $c : \mathbb{R}^2 \to \mathbb{R}$ and $g : \partial\mathcal{O} \to \mathbb{R}$ are known functions.
- We assume we can invert:

$$\mathbf{f} = \mathbf{e}(\mathcal{L}\mathbf{u_f}, \mathbf{u_f})$$

  for a known map $e : \mathbb{R}^2 \to \mathbb{R}$.
- **Our approach:** Put a prior on $\mathcal{L} u_{f_0}$.

Inverse problems

- **Setup:** We consider

$$\begin{aligned} \mathcal{L}u_f &= c(f, u_f) && \text{on } \mathcal{O}, \\ u_f &= g && \text{on } \Gamma \subseteq \partial\mathcal{O}, \end{aligned}$$

  with $\mathcal{L}$ a linear differential operator, and $c : \mathbb{R}^2 \to \mathbb{R}$ and $g : \partial\mathcal{O} \to \mathbb{R}$ are known functions.
- We assume we can invert:

$$\mathbf{f} = \mathbf{e}(\mathcal{L}\mathbf{u_f}, \mathbf{u_f})$$

  for a known map $e : \mathbb{R}^2 \to \mathbb{R}$.
- **Our approach:** Put a prior on $\mathcal{L}u_{f_0}$.

Example: Schrödinger Equation

- For $f_0 > 0$, consider u_{f_0} the solution to the PDE

$$\frac{1}{2}\Delta u_{f_0} = f u_{f_0}$$

  on $\mathcal{O} \subset \mathbb{R}^d$, with $u_{f_0} = g$ on $\partial\mathcal{O}$.
- We observe $Y = u_{f_0} + \frac{1}{\sqrt{n}}\epsilon$, with ϵ Gaussian noise indexed by $L^2(\mathcal{O})$.
- How can we recover f_0 in an optimal way, and quantify the uncertainty, as $n \to \infty$?

Example: Schrödinger Equation

- For $f_0 > 0$, consider u_{f_0} the solution to the PDE

$$\frac{1}{2}\Delta u_{f_0} = f u_{f_0}$$

  on $\mathcal{O} \subset \mathbb{R}^d$, with $u_{f_0} = g$ on $\partial\mathcal{O}$.
- We observe $Y = u_{f_0} + \frac{1}{\sqrt{n}}\epsilon$, with ϵ Gaussian noise indexed by $L^2(\mathcal{O})$.
- How can we recover f_0 in an optimal way, and quantify the uncertainty, as $n \to \infty$?

Example: Schrödinger Equation

- For $f_0 > 0$, consider u_{f_0} the solution to the PDE

$$\frac{1}{2}\Delta u_{f_0} = f u_{f_0}$$

  on $\mathcal{O} \subset \mathbb{R}^d$, with $u_{f_0} = g$ on $\partial\mathcal{O}$.
- We observe $Y = u_{f_0} + \frac{1}{\sqrt{n}}\epsilon$, with ϵ Gaussian noise indexed by $L^2(\mathcal{O})$.
- How can we recover f_0 in an optimal way, and quantify the uncertainty, as $n \to \infty$?

Intuition

The method consists of three steps:

- Put a prior on Δu_{f_0} and compute the posterior using the observations.
- Using the posterior for Δu_{f_0} and g, recover u_{f_0}.
- Recover f_0 by using the structural equation coming from the PDE.

Intuition

The method consists of three steps:

- Put a prior on Δu_{f_0} and compute the posterior using the observations.
- Using the posterior for Δu_{f_0} and g, recover u_{f_0}.
- Recover f_0 by using the structural equation coming from the PDE.

Intuition

The method consists of three steps:

- Put a prior on Δu_{f_0} and compute the posterior using the observations.
- Using the posterior for Δu_{f_0} and g, recover u_{f_0}.
- Recover f_0 by using the structural equation coming from the PDE.

Linearisation

- Let K be the integral operator

$$\begin{cases} \Delta K(h) &= h \qquad \text{on } \mathcal{O}; \\ K(h) &= 0 \qquad \text{on } \partial\mathcal{O}. \end{cases}$$

- Take $\tilde{g}$ such that $\Delta\tilde{g} = 0$ and $\tilde{g}|_{\partial\mathcal{O}} = g$ and write

$$Y = K(\Delta u_{f_0}) + \tilde{g} + \frac{1}{\sqrt{n}}\epsilon.$$

- The problem becomes linear for $\hat{Y} := Y - \tilde{g}$.

Linearisation

- Let K be the integral operator

$$\begin{cases} \Delta K(h) & = h & \text{on } \mathcal{O}; \\ K(h) & = 0 & \text{on } \partial\mathcal{O}. \end{cases}$$

- Take $\tilde{g}$ such that $\Delta \tilde{g} = 0$ and $\tilde{g}|_{\partial\mathcal{O}} = g$ and write

$$Y = K(\Delta u_{f_0}) + \tilde{g} + \frac{1}{\sqrt{n}}\epsilon.$$

- The problem becomes linear for $\hat{Y} := Y - \tilde{g}$.

Linearisation

- Let K be the integral operator

$$\begin{cases} \Delta K(h) &= h \qquad \text{on } \mathcal{O}; \\ K(h) &= 0 \qquad \text{on } \partial\mathcal{O}. \end{cases}$$

- Take $\tilde{g}$ such that $\Delta\tilde{g} = 0$ and $\tilde{g}|_{\partial\mathcal{O}} = g$ and write

$$Y = K(\Delta u_{f_0}) + \tilde{g} + \frac{1}{\sqrt{n}}\epsilon.$$

- The problem becomes linear for $\hat{Y} := Y - \tilde{g}$.

Inversion method

- For any $v \in L^2(\mathcal{O})$, using the boundary conditions, we can construct a function $\hat{u}_v = K(v) + \tilde{g}$ satisfying

$$\begin{cases} \Delta \hat{u}_v &= v \qquad \text{on } \mathcal{O}; \\ \hat{u}_v &= g \qquad \text{on } \partial\mathcal{O}. \end{cases}$$

- Using the inversion $f_0 = \frac{\Delta u_{f_0}}{2u_{f_0}}$, we construct $\hat{f} := \frac{v}{2\hat{u}_v}$.

Inversion method

- For any $v \in L^2(\mathcal{O})$, using the boundary conditions, we can construct a function $\hat{u}_v = K(v) + \tilde{g}$ satisfying

$$\begin{cases} \Delta \hat{u}_v & = v \qquad \text{on } \mathcal{O}; \\ \hat{u}_v & = g \qquad \text{on } \partial\mathcal{O}. \end{cases}$$

- Using the inversion $f_0 = \frac{\Delta u_{f_0}}{2 u_{f_0}}$, we construct $\hat{f} := \frac{v}{2\hat{u}_v}$.

Equivalent model

- The eigenbasis h_i of $K^\top K$ is convenient to model the observations.
- We observe

$$\hat{Y}_i = \kappa_i \; \mu_{0,i} + \frac{1}{\sqrt{n}} Z_i, i = 1, 2, \ldots$$

  with

  * κ_i^2 the eigenvalues of $K^\top K$;
  * $\mu_{0,i} := \langle \mathcal{L} u_{f_0}, h_i \rangle$ the coefficients of $\mathcal{L} u_{f_0}$ with respect to the basis;
  * $Z_i \sim N(0, 1)$ i.i.d.

Equivalent model

- The eigenbasis h_i of $K^\top K$ is convenient to model the observations.
- We observe

$$\hat{Y}_i = \kappa_i \ \mu_{0,i} + \frac{1}{\sqrt{n}} Z_i, i = 1, 2, \ldots$$

with

* κ_i^2 the eigenvalues of $K^\top K$;
* $\mu_{0,i} := \langle \mathcal{L} u_{f_0}, h_i \rangle$ the coefficients of $\mathcal{L} u_{f_0}$ with respect to the basis;
* $Z_i \sim N(0, 1)$ i.i.d.

Prior

- **Smoothness:** The smoothness of $\mathcal{L}u_{f_0}$ is determined by the decay of $\mu_{0,i}$ for $i \to \infty$.
- We say that $\mu \in S^\beta$ if $\sum_{i=1}^{\infty} \mu_i^2 i^{2\beta} < \infty$.
- **Prior:** Take a Gaussian prior

$$\mu_{i,0} \sim N(0, i^{-1-2\alpha}).$$

- We wish to match α with the unknown β.

Prior

- **Smoothness:** The smoothness of $\mathcal{L}u_{f_0}$ is determined by the decay of $\mu_{0,i}$ for $i \to \infty$.
- We say that $\mu \in S^\beta$ if $\sum_{i=1}^{\infty} \mu_i^2 i^{2\beta} < \infty$.
- **Prior:** Take a Gaussian prior

$$\mu_{i,0} \sim N(0, i^{-1-2\alpha}).$$

- We wish to match α with the unknown β.

Prior

- **Smoothness:** The smoothness of $\mathcal{L}u_{f_0}$ is determined by the decay of $\mu_{0,i}$ for $i \to \infty$.
- We say that $\mu \in S^\beta$ if $\sum_{i=1}^{\infty} \mu_i^2 i^{2\beta} < \infty$.
- **Prior:** Take a Gaussian prior

$$\mu_{i,0} \sim N(0, i^{-1-2\alpha}).$$

- We wish to match α with the unknown β.

Prior

- **Smoothness:** The smoothness of $\mathcal{L}u_{f_0}$ is determined by the decay of $\mu_{0,i}$ for $i \to \infty$.
- We say that $\mu \in S^\beta$ if $\sum_{i=1}^{\infty} \mu_i^2 i^{2\beta} < \infty$.
- **Prior:** Take a Gaussian prior

$$\mu_{i,0} \sim N(0, i^{-1-2\alpha}).$$

- We wish to match α with the unknown β.

Adaptation

- **Empirical Bayes:** The log-likelihood of $\hat{Y}$ given α is

$$\ell_n(\alpha) = \frac{1}{2}\sum_{i=1}^{\infty}\left(\log\left(1 + \frac{n}{i^{1+2\alpha}\kappa_i^{-2}}\right) - \frac{n^2}{i^{1+2\alpha}\kappa_i^{-2} + n}\hat{Y}_i^2\right).$$

- The maximizer

$$\hat{\alpha}_n = \operatorname{argmax}_{\alpha\in[0,\log n]}\ell_n(\alpha)$$

is used as a plug-in estimator, i.e. use the prior $\Pi_\alpha|_{\alpha=\hat{\alpha}_n}$.[1][2]

- **Hierarchical Bayes:** We put a prior τ on α such that $\tau(\alpha) \asymp \alpha^{-c1}e^{-c_2\alpha}$.

---

[1] B. Knapik, A. van der Vaart, and H. van Zanten (2011). "Bayesian Inverse Problems with Gaussian Priors". In: *The Annals of Statistics* 39.5, pp. 2626–2657. DOI: 10.1214/11-AOS920

[2] B. Knapik, B. Szabo, A. van der Vaart, and H. van Zanten (2015). "Bayes procedures for adaptive inference in inverse problems for the white noise model". In: *Probability Theory and Related Fields* 164.3, pp. 771–813. DOI: 10.1007/s00440-015-0619-7

Adaptation

- **Empirical Bayes:** The log-likelihood of $\hat{Y}$ given α is

$$\ell_n(\alpha) = \frac{1}{2}\sum_{i=1}^{\infty}\left(\log\left(1 + \frac{n}{i^{1+2\alpha}\kappa_i^{-2}}\right) - \frac{n^2}{i^{1+2\alpha}\kappa_i^{-2} + n}\hat{Y}_i^2\right).$$

- The maximizer

$$\hat{\alpha}_n = \text{argmax}_{\alpha\in[0,\log n]}\ell_n(\alpha)$$

is used as a plug-in estimator, i.e. use the prior $\Pi_\alpha|_{\alpha=\hat{\alpha}_n}$.[1][2]

- **Hierarchical Bayes:** We put a prior τ on α such that $\tau(\alpha) \asymp \alpha^{-c1}e^{-c_2\alpha}$.

[1] B. Knapik, A. van der Vaart, and H. van Zanten (2011). "Bayesian Inverse Problems with Gaussian Priors". In: *The Annals of Statistics* 39.5, pp. 2626–2657. DOI: 10.1214/11-AOS920

[2] B. Knapik, B. Szabo, A. van der Vaart, and H. van Zanten (2015). "Bayes procedures for adaptive inference in inverse problems for the white noise model". In: *Probability Theory and Related Fields* 164.3, pp. 771–813. DOI: 10.1007/s00440-015-0619-7

Adaptation

- **Empirical Bayes:** The log-likelihood of $\hat{Y}$ given α is

$$\ell_n(\alpha) = \frac{1}{2}\sum_{i=1}^{\infty}\left(\log\left(1+\frac{n}{i^{1+2\alpha}\kappa_i^{-2}}\right) - \frac{n^2}{i^{1+2\alpha}\kappa_i^{-2}+n}\hat{Y}_i^2\right).$$

- The maximizer

$$\hat{\alpha}_n = \mathrm{argmax}_{\alpha\in[0,\log n]}\ell_n(\alpha)$$

  is used as a plug-in estimator, i.e. use the prior $\Pi_\alpha|_{\alpha=\hat{\alpha}_n}$.[1][2]
- **Hierarchical Bayes:** We put a prior τ on α such that $\tau(\alpha) \asymp \alpha^{-c1}e^{-c_2\alpha}$.

[1] B. Knapik, A. van der Vaart, and H. van Zanten (2011). "Bayesian Inverse Problems with Gaussian Priors". In: *The Annals of Statistics* 39.5, pp. 2626–2657. DOI: 10.1214/11-AOS920

[2] B. Knapik, B. Szabo, A. van der Vaart, and H. van Zanten (2015). "Bayes procedures for adaptive inference in inverse problems for the white noise model". In: *Probability Theory and Related Fields* 164.3, pp. 771–813. DOI: 10.1007/s00440-015-0619-7

Prior

- We now consider $\mathcal{O} = [0,1]^d$.
- The eigenfunctions and eigenvalues of $K^\top K$ are

$$h_{i_1,\ldots,i_d} = 2^{d/2}\prod_{j=1}^{d}\sin(i_j\pi x_j),$$
$$\kappa^2_{i_1,\ldots,i_d} = \frac{1}{(\sum_{j=1}^d i_j^2)^2\pi^4}.$$

Prior

- We now consider $\mathcal{O} = [0,1]^d$.
- The eigenfunctions and eigenvalues of $K^\top K$ are

$$h_{i_1,\dots,i_d} = 2^{d/2}\prod_{j=1}^{d}\sin(i_j\pi x_j),$$
$$\kappa^2_{i_1,\dots,i_d} = \frac{1}{(\sum_{j=1}^{d} i_j^2)^2\pi^4}.$$

Convergence rate

Theorem

For $\beta > \frac{d}{2}$, let $f_0 \in S^\beta$ be compactly supported. For any sequence $M_n \to \infty$, we have

$$\Pi\left(v : \left\|\frac{v}{2\hat{u}} - f_0\right\|_{L^2} \geq M_n\epsilon_n \mid Y_1, \dots, Y_n\right) \xrightarrow{P} 0,$$

where ϵ_n is given by

$$\epsilon_n = L_n n^{-\beta/(d+2\beta+4)}, \tag{1}$$

with L_n a log-factor.

Credible sets

We can use the credible sets of the posterior for Δu_{f_0} to construct credible sets for f.

Theorem

If $\alpha < \beta$ is fixed, then there exists a set C_γ such that

$$\lim_{n\to\infty} \sup_{f_0 : \|\Delta u_{f_0}\|_\beta \leq 1} \mathbb{P}_{f_0}(f_0 \in C_\gamma) = 1.$$

If $\alpha = \beta$, then for each $\Delta u_{f_0} \in S^\beta$,

$$\lim_{n\to\infty} \mathbb{P}_{f_0}(f_0 \in C_\gamma) = 1.$$

Simulations

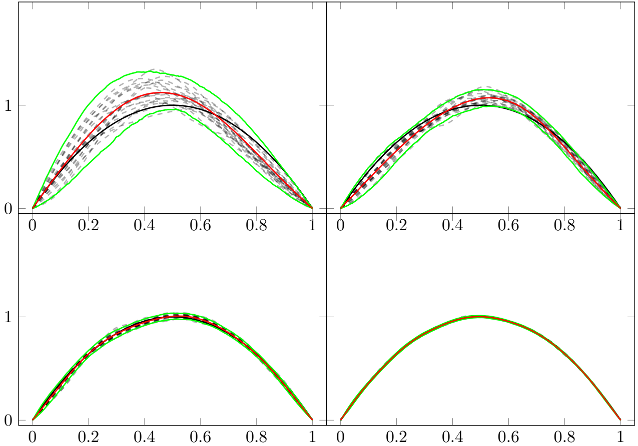


Figure: $n = 10^4, 10^6, 10^8, 10^{10}$ with $f_0(x) = 1 - 4(x - 1/2)^2$.

Simulations

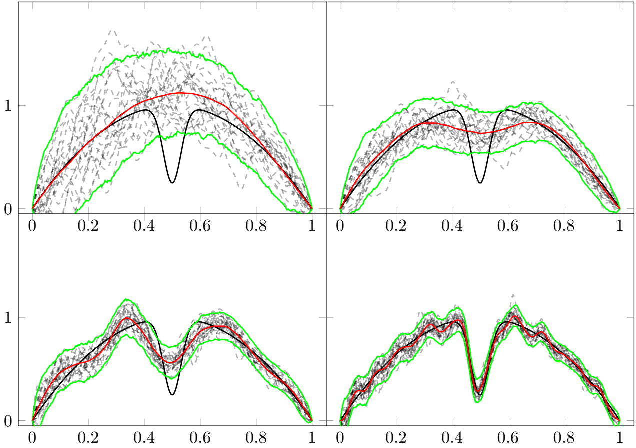
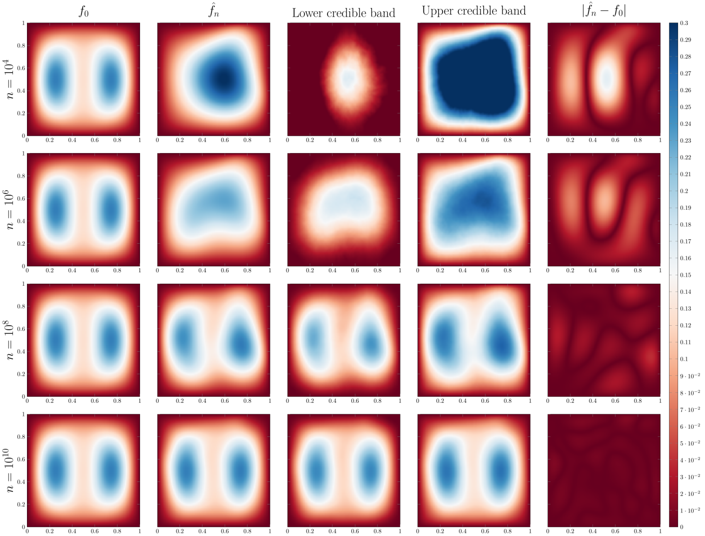


Figure: $n = 10^4, 10^6, 10^8, 10^{10}$ with
$f_0(x) = 1 - 4(x - 1/2)^2 - \frac{3}{4}\exp(-500(x - 1/2)^2)$.

Simulations



Examples

- **Heat equation with absorption:**

$$\begin{cases} \frac{d}{dt}u - \frac{1}{2}\Delta u & = fu \qquad (x,t) \in [0,1]^d \times (0,1]; \\ u(t,x) & = g \qquad x \in \partial[0,1]^d; \\ u(0,\cdot) & = u_0 \qquad x \in [0,1]^d. \end{cases} \tag{2}$$

- **Darcy's problem:**

$$\begin{cases} \frac{d}{dx}\left(f \frac{d}{dx}u_f\right) & = g \qquad x \in (0,1]; \\ u_f & = 0 \qquad x = 0. \end{cases} \tag{3}$$

- **First order ODE:**

$$\begin{cases} \frac{d}{dx}u_f & = fu_f \qquad x \in (0,1]; \\ u_f & = c \qquad x = 0. \end{cases} \tag{4}$$

Examples

- **Heat equation with absorption:**

$$\begin{cases} \frac{d}{dt}u - \frac{1}{2}\Delta u & = fu & (x,t) \in [0,1]^d \times (0,1]; \\ u(t,x) & = g & x \in \partial[0,1]^d; \\ u(0,\cdot) & = u_0 & x \in [0,1]^d. \end{cases} \tag{2}$$

- **Darcy's problem:**

$$\begin{cases} \frac{d}{dx}\left(f \frac{d}{dx} u_f\right) & = g & x \in (0,1]; \\ u_f & = 0 & x = 0. \end{cases} \tag{3}$$

- **First order ODE:**

$$\begin{cases} \frac{d}{dx}u_f & = fu_f & x \in (0,1]; \\ u_f & = c & x = 0. \end{cases} \tag{4}$$

Examples

- **Heat equation with absorption:**

$$\begin{cases} \frac{d}{dt}u - \frac{1}{2}\Delta u & = fu \qquad (x,t) \in [0,1]^d \times (0,1]; \\ u(t,x) & = g \qquad x \in \partial[0,1]^d; \\ u(0,\cdot) & = u_0 \qquad x \in [0,1]^d. \end{cases} \tag{2}$$

- **Darcy's problem:**

$$\begin{cases} \frac{d}{dx}\left(f \frac{d}{dx} u_f\right) & = g \qquad x \in (0,1]; \\ u_f & = 0 \qquad x = 0. \end{cases} \tag{3}$$

- **First order ODE:**

$$\begin{cases} \frac{d}{dx} u_f & = fu_f \qquad x \in (0,1]; \\ u_f & = c \qquad x = 0. \end{cases} \tag{4}$$

Discussion

- Can it be applied to discrete observational models as well:

$$Y_i = (\mathcal{A}f_0)(x_i) + \epsilon_i, i = 1, \ldots, n?$$

- Can this be extended PDE's without (explicit) inversion formulas?

Discussion

- Can it be applied to discrete observational models as well:

$$Y_i = (\mathcal{A}f_0)(x_i) + \epsilon_i, i = 1, \ldots, n?$$

- Can this be extended PDE's without (explicit) inversion formulas?

Discussion

- Can it be applied to discrete observational models as well:

$$Y_i = (\mathcal{A}f_0)(x_i) + \epsilon_i, i = 1, \ldots, n?$$

- Can this be extended PDE's without (explicit) inversion formulas?

References I

Knapik, B., B. Szabo, A. van der Vaart, and H. van Zanten (2015). “Bayes procedures for adaptive inference in inverse problems for the white noise model”. In: *Probability Theory and Related Fields* 164.3, pp. 771–813. DOI: 10.1007/s00440-015-0619-7.

Knapik, B., A. van der Vaart, and H. van Zanten (2011). “Bayesian Inverse Problems with Gaussian Priors”. In: *The Annals of Statistics* 39.5, pp. 2626–2657. DOI: 10.1214/11-AOS920.

Nickl, R. (2018). “Bernstein-von Mises Theorems for statistical inverse problems I: Schrödinger Equation”. In: *Journal of the European Mathematical Society* 22.8, pp. 2697–2750. DOI: https://doi.org/10.4171/JEMS/975.