Stochastic inverse problems with impulsive noise

Christian Clason[1] Laurent Demaret[2]

[1]Faculty of Mathematics, Universität Duisburg-Essen
[2]HelmholtzZentrum München

SIAM Conference on Uncertainty Quantification
Lausanne, April 5, 2016

Motivation: impulsive noise

Impulsive noise

- appears in digital image acquisition, processing (hardware defects, cosmic rays, …)
- characterization: noise is "sparse", localized
- e.g., random-valued impulsive noise

$$\eta(x_i) = \begin{cases} \xi_i & \text{with probability } \lambda \\ 0 & \text{with probability } 1 - \lambda \end{cases}$$

$\xi_i \in \mathcal{N}(0, \sigma^2)$ i.i.d. Gaussian, $\lambda > 0$, σ large

Motivation: impulsive noise

Impulsive noise

- appears in digital image acquisition, processing (hardware defects, cosmic rays, …)
- characterization: noise is "sparse", localized
- e.g., random-valued impulsive noise

$$\eta(x_i) = \begin{cases} \xi_i & \text{with probability } \lambda \\ 0 & \text{with probability } 1-\lambda \end{cases}$$

$\xi_i \in \mathcal{N}(0, \sigma^2)$ i.i.d. Gaussian, $\lambda > 0$, σ large

- meaningless in function space!

Motivation

Goal:

- rigorous definition of continuous impulsive noise model
- analysis of stochastic inverse problems with impulsive noise
- conforming discretization reproducing discrete noise

Approach:

- model impulsive noise as point process $\rightsquigarrow$ random measure
- relate noise level to noise parameters
- discretization by averaging $\rightsquigarrow$ linear combination of Diracs

Noise process: definition

Poisson point process:

- random countable set $\Pi \subset \Omega \subset \mathbb{R}^n$
- intensity measure μ (here: $\mu(A) = \lambda|A|$ for $\lambda > 0$)
- counting measure $N : A \mapsto \#(\Pi \cap A)$

satisfying

1. $A_i \subset \Omega$ disjoint, measurable $\Rightarrow N(A_i)$ independent
2. $A \subset \Omega$ measurable $\Rightarrow N(A)$ Poisson distributed with mean $\mu(A)$,

$$\mathbb{P}\left[N(A) = k\right] = e^{-\mu(A)} \frac{\mu(A)^k}{k!}$$

Noise process: definition

Marked Poisson point process:

$$\Pi^* = \left\{(x, \xi_x) : x \in \Pi,\ \xi_x \in \mathcal{N}(0, \sigma^2)\right\}$$

- $x \in \Pi$ denotes location of corrupted point
- ξ_x i.i.d denotes magnitude of corruption
- statistical model for physical cause (e.g., cosmic rays)
- Poisson point process on $\Omega \times \mathbb{R}$
- defines random measure ($\rightsquigarrow$ impulse noise)

$$\eta = \sum_{(x,\xi_x)\in\Pi^*} \xi_x \delta_x$$

- Ω bounded $\rightsquigarrow$ Π finite, $\eta \in \mathcal{M}(\Omega) = C(\overline{\Omega})^*$ almost surely

Noise process: moments

- Expectation: for $A \subset \Omega$,

$$\mathbb{E}[\eta(A)] = \sum_{k=1}^{\infty} \mathbb{P}[N(A) = k] \sum_{x \in \Pi \cap A} \int_{\mathbb{R}} \xi_x \, dv = 0$$

- Variance: for $A \subset \Omega$,

$$\begin{aligned} \mathrm{Var}[\eta(A)] &= \sum_{k=1}^{\infty} \mathbb{P}[N(A) = k] \sum_{x \in \Pi \cap A} \int_{\mathbb{R}} \xi_x^2 \, dv \\ &= \sum_{k=1}^{\infty} e^{-\lambda |A|} \frac{(\lambda |A|)^k}{k!} k \sigma^2 \\ &= \lambda \sigma^2 |A| \end{aligned}$$

Noise process: noise level

$$\varepsilon(\eta) := \|\eta\|_{\mathcal{M}(\Omega)} = \sup_{\|\varphi\|_{C(\overline{\Omega})} \leqslant 1} \sum_{(x,\xi_x) \in \Pi^*} \xi_x \langle \delta_x, \varphi \rangle = \sum_{(x,\xi_x) \in \Pi^*} |\xi_x|$$

Campbell's theorem, $|\xi_x|$ i.i.d. and half-normal $\rightsquigarrow$

$$\mathbb{E}[\varepsilon(\eta)] = \int_\Omega \int_{\mathbb{R}} |\xi_x| \, d\mu dv = \lambda|\Omega| \int_{\mathbb{R}} |\xi| \, dv = \lambda\sigma|\Omega|\sqrt{\frac{2}{\pi}}$$

$$\mathrm{Var}[\varepsilon(\eta)] = \int_\Omega \int_{\mathbb{R}} |\xi_x|^2 \, d\mu dv = \lambda|\Omega| \int_{\mathbb{R}} |\xi|^2 \, dv = \lambda\sigma^2|\Omega| \left(1 - \frac{2}{\pi}\right)$$

Noise process: convergence

Consider $\{\eta_n\}_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega) \quad \text{for} \quad \lambda_n, \sigma_n > 0$

1. If $\lambda_n \sigma_n \to 0$:

$$\mathbb{E}[\varepsilon(\eta_n)] = \mathcal{O}(\lambda_n \sigma_n) \to 0$$

2. If also $\lambda_n \sigma_n^2 = \mathcal{O}(n^{-r})$ for $r > 1$ (e.g., subsequence):

$$\varepsilon(\eta_n) \to 0 \quad \text{almost surely}$$

Proof:

- Chebyshev concentration inequality + Borel–Cantelli
- not constructive $\rightsquigarrow$ no uniform a priori bounds, no rates

Inverse problem

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

$$\min_{u \in X} \|F(u) - y^{\varepsilon}(\omega)\|_{\mathcal{M}(\Omega)} + \alpha \mathcal{R}(u),$$

- X Banach space, $\mathcal{R}$ convex, l.s.c., weakly sequentially precompact sublevel sets
- e.g., $\mathcal{R}(u) = \frac{1}{2}\|u\|_X^2$
- $F : X \to \mathcal{M}(\Omega)$ bounded, completely continuous (compact embedding $F : X \to Y \hookrightarrow \mathcal{M}(\Omega)$)
- $y^{\varepsilon} = F(u^{\dagger}) + \eta$ random noisy data, $y^{\varepsilon}(\omega)$ realization

Inverse problem

$$\min_{u \in X} \|F(u) - y^\varepsilon(\omega)\|_{\mathcal{M}(\Omega)} + \alpha \mathcal{R}(u),$$

Standard arguments: for every $\alpha > 0$ and realization $y^\varepsilon(\omega) \in \mathcal{M}(\Omega)$:

- existence of minimizer $u_\alpha^\varepsilon(\omega)$
- $y_n \to y^\varepsilon(\omega)$ implies $u_\alpha^n \to u_\alpha^\varepsilon(\omega)$
- if $\mathcal{R}$ strictly convex, $u_\alpha^\varepsilon(\omega)$ unique

$\rightsquigarrow$ defines random field u_α^ε

Consider

- sequence $\{\eta_n\}$ for λ_n, σ_n with

$$\lambda_n \sigma_n \to 0$$

- noisy data $y_n := F(u^\dagger) + \eta_n$, minimizer $u_n := u_{\alpha_n}^{\varepsilon_n}$

If $$\alpha_n \to 0 \quad \text{and} \quad \frac{\lambda_n \sigma_n}{\alpha_n} \to 0$$

then subsequence $$\mathbb{E}[u_n] \rightharpoonup u^\dagger$$

- proof: standard deterministic arguments + convergence of ε_n [Bissantz/Hohage/Munk '04]
- full sequence if $u^\dagger$ unique, strong convergence if $\mathcal{R}$ Kadec–Klee

Inverse problem: convergence

Consider

- sequence $\{\eta_n\}$ for λ_n, σ_n with

$$\{\lambda_n\}, \{\sigma_n\} \text{ bounded}, \qquad \lambda_n \sigma_n = \mathcal{O}(n^r) \text{ for } r > 1$$

- noisy data $y_n := F(u^\dagger) + \eta_n$, minimizer $u_n := u_{\alpha_n}^{\varepsilon_n}$

If
$$\alpha_n \to 0 \qquad \text{and} \qquad \frac{\lambda_n \sigma_n^2}{\alpha_n} \to 0$$

then subsequence $\quad u_n \rightharpoonup u^\dagger$ almost surely

- proof: standard deterministic arguments + convergence of ε_n [Bissantz/Hohage/Munk ’04]
- full sequence if $u^\dagger$ unique, strong convergence if $\mathcal{R}$ Kadec–Klee

Inverse problem: convergence rates

Under usual assumptions (source condition, nonlinearity):

1. A priori choice: $\alpha \sim (\lambda\sigma)^{\tau}$ for $\tau \in (0,1)$

$$\mathbb{E}\left[\|u_\alpha^\varepsilon - u^\dagger\|_X\right] \leqslant c(\lambda\sigma)^{\frac{1-\tau}{2}}$$

2. A posteriori choice: $\|F(u_\alpha^\varepsilon) - y^\varepsilon\|_{\mathcal{M}(\Omega)} \sim \tau\lambda\sigma$

$$\mathbb{E}\left[\|u_\alpha^\varepsilon - u^\dagger\|_X\right] \leqslant c(\lambda\sigma)^{\frac{1}{2}}$$

- no almost sure rates, since no such rates for ε_n
- for σ bounded: rates independent of σ

  $\rightsquigarrow$ λ essentially characterizes noise level; robustness

Discretization

UNIVERSITÄT DUISBURG ESSEN Open-Minded

Approach: start with discretization of $C(\overline{\Omega})$ [Casas/C./Kunisch '12]

- $\{x_j\}_{j=1}^{N_h} \subset \Omega$ nodes (sampling points, pixel midpoints, vertices)
- $\{e_j\}_{j=1}^{N_h}$ nodal basis of continuous functions (FEM basis, point spread functions)
- $h := \max\limits_{1 \leqslant j \leqslant N} h_j, \quad h_j := |\operatorname{supp} e_j|$

$$C_h := \left\{ v_h \in C(\overline{\Omega}) : v_h = \sum_{j=1}^{N_h} v_j e_j, \text{ where } \{v_j\}_{j=1}^{N_h} \subset \mathbb{R} \right\}$$

Discretization

$$M_h := \left\{ \mu_h \in \mathcal{M}(\Omega) : \mu_h = \sum_{j=1}^{N_h} \mu_j \delta_{x_j}, \text{ where } \{\mu_j\}_{j=1}^{N_h} \subset \mathbb{R} \right\}$$

with norm

$$\|\mu_h\|_{\mathcal{M}(\Omega)} = \sup_{\|v\|_{C(\overline{\Omega})}=1} \sum_{j=1}^{N_h} \mu_j \langle \delta_{x_j}, v \rangle = \sum_{j=1}^{N_h} |\mu_j| =: |\vec{\mu}_h|_1$$

$\rightsquigarrow$ M_h topological dual of C_h with respect to duality pairing

$$\langle \mu_h, v_h \rangle = \sum_{j=1}^{N_h} \mu_j v_j = \vec{\mu}_h^T \vec{v}_h$$

Discretization: interpolation operators

$$\Pi_h : C(\overline{\Omega}) \to C_h, \qquad \Pi_h v = \sum_{j=1}^{N_h} \langle v, \delta_{x_j} \rangle e_j$$

$$\Lambda_h : \mathcal{M}(\Omega) \to M_h, \qquad \Lambda_h \mu = \sum_{j=1}^{N_h} \langle \mu, e_j \rangle \delta_{x_j}$$

$\rightsquigarrow$ For all $\mu \in \mathcal{M}(\Omega)$, $\quad v \in C(\overline{\Omega})$, $\quad v_h \in C_h$:

1. $\langle \mu, v_h \rangle = \langle \Lambda_h \mu, v_h \rangle$ and $\langle \mu, \Pi_h v \rangle = \langle \Lambda_h \mu, v \rangle$
2. $\|\Lambda_h \mu\|_{\mathcal{M}(\Omega)} \leqslant \|\mu\|_{\mathcal{M}(\Omega)}$
3. $\Lambda_h u \rightharpoonup^{\star} u$ in $\mathcal{M}(\Omega)$ and $\|\Lambda_h u\|_{\mathcal{M}(\Omega)} \to \|u\|_{\mathcal{M}(\Omega)}$

Discretization: noise

Define discretized noise η_h via

$$
\begin{aligned}
\eta_h(\omega) := \Lambda_h[\eta(\omega)] &= \sum_{j=1}^{N_h} \langle \eta(\omega), e_j \rangle \, \delta_{x_j} \\
&= \sum_{j=1}^{N_h} \left(\sum_{x \in \Pi \cap \operatorname{supp} e_j} e_j(x) \xi_x(\omega) \right) \delta_{x_j} \\
&=: \sum_{j=1}^{N_h} \eta_j(\omega) \delta_{x_j}
\end{aligned}
$$

- nodes x_j deterministic $\rightsquigarrow$ identify η_h with $(\eta_1, \ldots, \eta_j) \in \mathbb{R}^{N_h}$
- averaging $\rightsquigarrow$ model of physical image acquisition by sensors

Discrete noise: moments

Case differentiation:

1. $\eta_j = 0$: iff $\operatorname{supp} e_j \cap \Pi = \emptyset$ (a.s.) $\rightsquigarrow$

$$\mathbb{P}(\mu_j = 0) = \mathbb{P}\left(N(\operatorname{supp}(e_j)) = 0\right) = e^{-\lambda h_j}$$

2. $\eta_j \neq 0$: then

$$\eta_j(\omega) = \sum_{x \in \Pi \cap \operatorname{supp}(e_j)} e_j(x)\xi_x(\omega)$$

a.s. finite linear combination of Gaussian $\rightsquigarrow$ Gaussian, $\mathbb{E}[\eta_h] = 0$,

$$\operatorname{Var}[\mu_j] = \lambda \int_\Omega e_j(x)^2\, dx \int_{\mathbb{R}} \xi^2\, dv =: \lambda s_j \sigma^2$$

with $s_j \leqslant h_j \leqslant h$ (Campbell's theorem)

Discrete noise: comparison

Discrete noise model in uniform case $s_j \equiv s \approx h$:

$$\eta_h(x_j) = \eta_j = \begin{cases} 0 & \text{with probability } 1 - \lambda_h \\ \xi_j \in \mathcal{N}(0, \sigma_h^2) & \text{with probability } \lambda_h \end{cases}$$

$$\lambda_h := 1 - e^{-\lambda h}, \qquad \sigma_h \approx \lambda \sigma^2 h$$

- effective noise parameters λ_h, σ_h discretization dependent
- σ_h depends on σ and λ
- note: taking $h \to 0$ here meaningless since $\eta_h \rightharpoonup^* \eta$

Discrete noise: level

$$\varepsilon_h := \|\eta_h\|_{\mathcal{M}(\Omega)} = \sum_{j=1}^{N_h} |\eta_j|$$

- $|\eta_j|$ half-normal random variable (not independent!)
- Λ_h interpolation $\quad\rightsquigarrow\quad$ $\varepsilon_h \leqslant \varepsilon$ almost surely,

$$\mathbb{E}[\varepsilon_h] \leqslant \mathbb{E}[\varepsilon]$$

- $\rightsquigarrow$ convergence $\varepsilon_h \to 0$ as $\lambda, \sigma \to 0$

Discrete inverse problem

$$\min_{u \in X} \|F_h(u) - y_h^\varepsilon\|_{\mathcal{M}(\Omega)} + \alpha \mathcal{R}(u)$$

- $F_h := (\Lambda_h \circ F) : X \to M_h$
- $y_h^\varepsilon := \Lambda_h y^\varepsilon = F_h(u^\dagger) + \eta_h \in M_h$
- semi-discretization (discretization of X independent)
- conforming discretization $\rightsquigarrow$ well-posed, solution $u_h := u_\alpha^{\varepsilon_h}$
- ε_h uniformly bounded $\rightsquigarrow$ convergence, rates (uniform in h)

Discrete solution: convergence

Consider

- noise parameters λ, σ fixed
- discretization parameter $h \to 0$

Then: $\left\{u_h^\varepsilon\right\}_{h>0}$ contains subsequences with

1. $\mathbb{E}[u_\alpha^{\varepsilon_h}] \rightharpoonup \mathbb{E}[u_\alpha^\varepsilon]$
2. $u_\alpha^{\varepsilon_h} \rightharpoonup u_\alpha^\varepsilon$ almost surely

- whole sequence if u_α unique, strong convergence if $\mathcal{R}$ Kadec–Klee
- proof: boundedness of Λ_h, standard arguments

Numerical example: comparison

Illustrate behavior of discretized vs. discrete noise

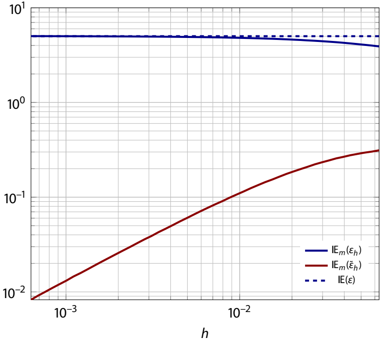
- $\Omega = [0, 2\pi]$
- e_j linear B-spline basis (hat) functions
- $\lambda \in \{1, 100\}, \quad \sigma \in \{0.1, 1\}$ fixed
- $N_h \in [10^2, 10^4]$

Compare empirical mean (average over $m = 1000$ realizations) for

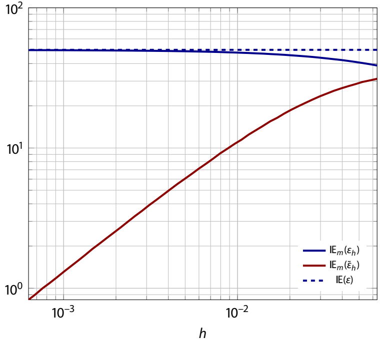
- $\mathbb{E}_m[\varepsilon_h] = \frac{1}{m}\sum_{i=1}^m \|\Lambda_h \eta(\omega_i)\|_{\mathcal{M}(\Omega)}$
- $\mathbb{E}_m[\tilde{\varepsilon}_h] = \frac{1}{m}\sum_{i=1}^m \|\tilde{\eta}_h(\omega_i)\|_{\mathcal{M}(\Omega)}$

  for $\tilde{\eta}_h$ discrete impulsive noise with rate λ_h, variance σ_h
- $\mathbb{E}[\varepsilon] = \lambda\sigma|\Omega|\sqrt{\frac{2}{\pi}}$

Numerical example: noise

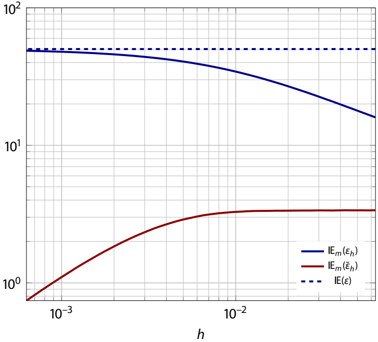


(a) $\lambda = 10, \sigma = 0.1$

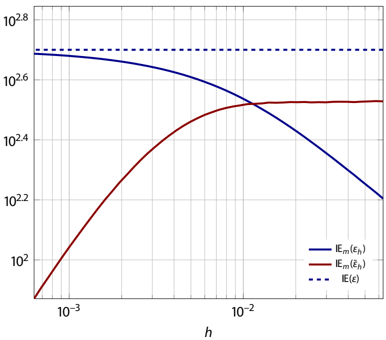


(b) $\lambda = 10, \sigma = 1$

Numerical example: noise



(c) $\lambda = 100, \sigma = 0.1$



(d) $\lambda = 100, \sigma = 1$

Conclusions

Continous impulsive noise:

- Poisson point process is appropriate model
- conforming discretization reproduces standard discrete noise
- convergence of stochastic inverse problem

Outlook:

- adaptive discretization & regularization
- heuristic parameter choice
- fitting with probability metrics
- Bayesian inverse problems