Stochastic asymptotical regularization for inverse problems

Ye Zhang
ye.zhang@smbu.edu.cn

Shenzhen MSU-BIT University
&
Beijing Institute of Technology

25 March 2023

深圳北理莫斯科大學
УНИВЕРСИТЕТ МГУ–ППИ В ШЭНЬЧЖЭНЕ
SHENZHEN MSU–BIT UNIVERSITY

Outline

Settings

$$Ax = y, \tag{1}$$

- A: a compact linear operator acting between infinite dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$.
- Ill-posedness of type II in Nashed: $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$.
- Deterministic noise model: $\|y^\delta - y\| \leq \delta$.
- Model (1) $+\ y^\delta \Rightarrow x^\delta$ (Regularization: $x^\delta \to x^\dagger$ as $\delta \to 0$.)
- Variational regularization vs Iterative regularization.

- Tikhonov, Bakushinskiy, Yagola, Vainikko, Tautenhahn, etc.

Asymptotic regularization

- Landweber iteration ($\min_x \|Ax - y\|^2$):

$$x_{k+1}^{\delta} = x_k^{\delta} + \Delta t A^*(y^{\delta} - Ax_k^{\delta}), \quad x_0^{\delta} = x_0, \quad \Delta t \in (0, 2/\|A\|^2), \tag{2}$$

- Asymptotic regularization (Showalter's method)

$$\dot{x}^{\delta}(t) = A^*(y^{\delta} - Ax^{\delta}(t)) \tag{3}$$

- Hölder-type source conditions: $x^{\dagger} \in \mathcal{R}\left((A^*A)^p\right) \Rightarrow$
- Order optimal convergence rate & Morozov's discrepancy principle:

$$\|x^{\delta}(T_*) - x^{\dagger}\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}}) \quad \text{and} \quad \boxed{T_* = \mathcal{O}(\delta^{-\frac{2}{2p+1}})} \quad \text{as} \quad \delta \to 0. \tag{4}$$

- Generalized asymptotical regularization:

$$\mathcal{D}\, x^{\delta}(t) = A^*(y^{\delta} - Ax^{\delta}(t)) \tag{5}$$

Outline

Heavy ball model

Motivation: fast dynamic for $\boxed{\min_x \|Ax - y\|^2}$.

Heavy ball model

$$\ddot{x}^\delta(t) + \eta \dot{x}^\delta(t) + A^* A x(t) = A^* y^\delta. \tag{6}$$

The optimal convergence rate + damped symplectic integrators [a].

[a] Zhang Y, Hofmann B, On the second-order asymptotical regularization of linear ill-posed inverse problems. Applicable Analysis, 2020, 99, 1000-1025.

- Hölder-type source conditions: $x^\dagger \in \mathcal{R}\left((A^*A)^p\right) \Rightarrow$
- Order optimal convergence rate & Total energy (Morozov's) discrepancy principle:

$$\|x^\delta(T_*) - x^\dagger\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}}) \quad \text{and} \quad T_* = \mathcal{O}(\delta^{-\frac{2}{2p+1}}) \quad \text{as} \quad \delta \to 0.$$

Literature on asymptotical regularization

- Bot & O. Scherzer [Foundations of Computational Mathematics, 2021; Optimization]
- S. Lu [The CSIAM Transactions on Applied Mathematics, 20; SIAM/ASA J. Uncertainty Quantification, 2021]
- W. Wang & Q. Jin [Inverse Problems, 2022; etc.]
- ZY [Inverse Problems, 2018; IMA Journal of Applied Mathematics, 2019; Journal of Computational and Applied Mathematics, 2020; Inverse Problems, 2020; SIAM Journal on Imaging Sciences, 2021]

Accelerated regularization

$$\|x^\delta(T_*) - x^\dagger\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}}) \quad \text{and} \quad T_* = \mathcal{O}(\delta^{-\frac{2}{2p+1}}) \quad \text{as} \quad \delta \to 0.$$

Accelerated regularization

An iterative/dynamical method is called an *accelerated order-optimal regularization algorithm* if it exhibits the order-optimal convergence rate of the approximate solution, but requires far fewer iterations than needed for an ordinary Landweber iteration/asymptotical regularization.

Acceleration factor

A method has an acceleration factor σ if it is an order-optimal reg. algorithm under Hölder-type source conditions, and the iteration number/stopping time has the asymptotic $\mathcal{O}(\delta^{-\frac{2}{(\sigma+1)(2p+1)}})$.

Acceleration asymptotical regularization

Fractional asymptotical regularization (⇝ Kaczmarz)

$$\left({}^{C}D_{0+}^{\theta}x^{\delta}\right)(t) + A^{*}Ax^{\delta}(t) = A^{*}y^{\delta}, \quad D^{k}x^{\delta}(0) = b_{k}, \ k = 0, ..., n-1. \tag{7}$$

Optimal accuracy $\|x^{\delta}(T_{*}) - x^{\dagger}\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}})$ + Time cost $t^{*} = \mathcal{O}(\delta^{-\frac{2}{\theta(2p+1)}})$ for $p \leq 1$ [a].

[a] Zhang Y, Hofmann B, On fractional asymptotical regularization of linear ill-posed problems in Hilbert spaces. Fractional Calculus and Applied Analysis, 2019, 22, 699-721.

Second order asymptotical regularization (⇝ Nesterov)

$$\ddot{x}^{\delta}(t) + \frac{1+2s}{t}\dot{x}^{\delta}(t) + A^{*}Ax(t) = A^{*}y^{\delta}, \quad x(0) = x_{0}, \ \dot{x}(0) = 0, \quad s > -1/2. \tag{8}$$

Optimal accuracy $\|x^{\delta}(T_{*}) - x^{\dagger}\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}})$ + Time cost $t^{*} = \mathcal{O}(\delta^{-\frac{1}{2p+1}})$ [a].

[a] Gong R, Hofmann B, Zhang Y. A new class of accelerated regularization methods, with application to bioluminescence tomography. Inverse Problems, 2020, 36, 055013.

Super acceleration asymptotical regularization

Super Acceleration Regularization of order n (SAR^n). $n > -1$

$$t\ddot{x}^\delta(t) + \left(t^{-n} - n\right)\dot{x}^\delta(t) + t^{n+1}A^*A\dot{x}^\delta(t) + A^*Ax^\delta(t) = A^*y^\delta, \quad x(0) = x_0,\ \dot{x}(0) = 0. \tag{9}$$

Optimal accuracy $\|x^\delta(T_*) - x^\dagger\| = \mathcal{O}(\delta^{\frac{2p}{2p+1}})$ + Time cost $t^* = \mathcal{O}(\delta^{-\frac{2}{(2p+1)(n+1)}})$ [a].

[a] Zhang Y. On the acceleration of optimal regularization algorithms for linear ill-posed inverse problems. Calcolo, 2022, 60(1).

Outline

Outline

Stochastic Asymptotical Regularization (SAR)

SAR (linear: $Ax = y$)

$$\boxed{dx^{\delta} = A^{*}(y^{\delta} - Ax^{\delta})dt + f(t)dB_t, \quad x^{\delta}(0) = x_0,}\,{}^{a} \tag{10}$$

where $x_0 \in \mathcal{X}$ is non-random, B_t is an $\mathcal{X}$-valued cylindrical Wiener process $B_t = \sum_{i=1}^{\infty} u_j \beta_j(t)$.
$\{u_j\}$: the orthorgomal basis of $\mathcal{X}$. $\{\beta_j\}$: independent $\mathbb{R}$-valued Brownian motions.

[a] Ye Zhang, Chuchu Chen, Stochastic asymptotical regularization for linear inverse problems, Inverse Problems, 39(1), 2022, 015007.

SAR (nonlinear: $F(x) = y$)

$$\boxed{\mathrm{d}x^{\delta}(t) = F'\left(x^{\delta}(t)\right)^{*}\left[y^{\delta} - F\left(x^{\delta}(t)\right)\right]\mathrm{d}t + f(t)\,\mathrm{d}B_t, \quad x^{\delta}(0) = x_0.}\,{}^{a} \tag{11}$$

[a] Haie Long, Ye Zhang, Stochastic asymptotical regularization for nonlinear ill-posed problems, 2022.

- Goal: $\mathbb{E}(\left\|x^{\delta}(t^{*}(\delta)) - x^{\dagger}\right\|^{2}) \to 0$ as $\delta \to 0$, and convergence rates?

Outline

Uncertainty Quantification: I

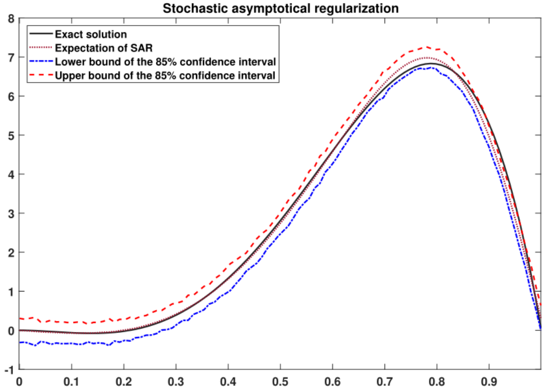


Figure: The expectation of SAR and the 85% confidence interval for problem (with $\delta = 1\%$)

$$Ax(s) := \int_0^1 K(s,t)x(t)dt = y(s), \quad K(s,t) = s(1-t)\chi_{s\le t} + t(1-s)\chi_{s>t}.$$

Uncertainty Quantification: II

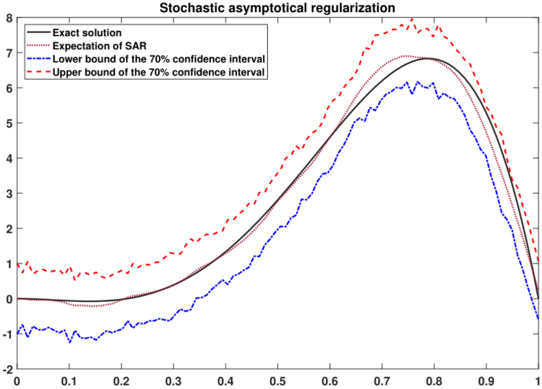


Figure: The expectation of SAR and the 70% confidence interval for problem (with $\delta = 5\%$)

$$Ax(s) := \int_0^1 K(s,t)x(t)dt = y(s), \quad K(s,t) = s(1-t)\chi_{s\le t} + t(1-s)\chi_{s>t}.$$

Nonlinear problem: the model problem

$$
\begin{aligned}
-\Delta u + cu &= w \quad \text{in} \quad \Omega, \\
\frac{\partial u}{\partial n} &= 0 \quad \text{on} \quad \partial\Omega,
\end{aligned} \tag{12}
$$

Problem (12) can be described as

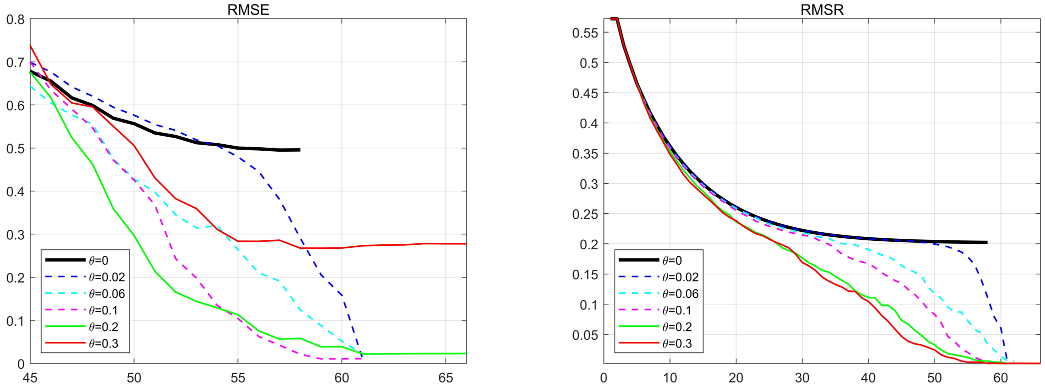
$$F(c) = u(c) \tag{13}$$

$$D(F) := \left\{ c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \le \zeta_0 \text{ for some } \hat{c} \ge 0,\ a.e. \right\}$$

$$F'(c)\, q = -A(c)^{-1} (qF(c)), \quad F'(c)^* \omega = -u(c)\, A(c)^{-1} \omega,$$

where $A(c) : H^2 \cap D(F) \to L^2$ is defined by $A(c)\, u = -\Delta u + cu$.

Nonlinear problem: escaping from local minima



$$\text{RMSE} := \mathbb{E}[\|x^\dagger - x_k^\delta\|^2]^{\frac{1}{2}}, \quad \text{RMSR} := \mathbb{E}[\|F(x_k^\delta) - y^\delta\|^2]^{\frac{1}{2}}.$$

Merits of SAR

In comparison with the Bayesian methods, our method has the following merits:

- The basic model is quite general. Though it is proposed in infinite dimensional Hilbert spaces, it can be easily extended to some more general abstract spaces e.g. Banach spaces, metric spaces. The Bayesian method is usually constructed in a finite Euclidian space. An infinite dimensional generalization is difficulty since there is no Lebesgue measure on infinite dimensional spaces.
- The operator equation serves as a deterministic model. The noise structure is almost arbitrary, and we only require the noise level assumption for the noisy data. The Bayesian method usually requires a strong assumption on the noise structure, e.g. the Gaussian noise.
- The Bayesian method requires a prior probability distribution of the exact solution (unavailable in practice). Only in very special cases, e.g. a Gaussian prior, one can obtain the closed form of the posteriori distribution. The requirements of SAR are much more slight: it does not need any *a priori* distribution for the unknown $x^\dagger$. For rigorously quantifying the uncertainty of the SAR solution, some source conditions of $x^\dagger$ are required.

Outline

Well-posedness of SAR

Proposition

For any $f \in L^\infty(\mathbb{R}_+)$, the stochastic differential equation (50) has a unique mild solution $x^\delta(t) \in \mathcal{X}$, given by

$$x^\delta(t) = e^{-A^*At}x_0 + \int_0^t e^{-A^*A(t-s)} A^* y^\delta ds + \int_0^t e^{-A^*A(t-s)} f(s) dB(s). \tag{14}$$

The random variable $x^\delta(t)$ is Gaussian on $\mathcal{X}$ with mean

$$\mathbb{E}x^\delta(t) = e^{-A^*At}x_0 + \int_0^t e^{-A^*A(t-s)} A^* y^\delta ds \tag{15}$$

and variance operator given by

$$\mathrm{Var}(x^\delta(t)) = \int_0^t e^{-A^*A(t-s)} Q e^{-A^*A(t-s)} [f(s)]^2 ds. \tag{16}$$

Quantity $\ell(x^\delta(t^*))$

- ℓ: Dirac delta distribution $\delta_{\vec{\xi}}$ at the point $\vec{\xi}$.
- It is a unbounded linear functional on $L^2(\mathbb{R}^d)$. Smooth compactly supported functions are dense in $L^2(\mathbb{R}^d)$, and the action of $\delta_{\vec{\xi}}$ on such functions is well-defined.
- $\mathcal{X} = H^r(\mathbb{R}^d)$ with $r > d/2$, $\delta_{\vec{\xi}} \in H^{-r}(\mathbb{R}^d)$ is a bounded linear functional on $H^r(\mathbb{R}^d)$.
-

$$\mathbb{E}\ell(x^\delta(t^*)) = \left\langle e^{-A^*At^*}x_0 + \int_0^{t^*} e^{-A^*A(t^*-s)}A^*y^\delta ds\,,\,\ell \right\rangle. \quad (17)$$

$$\text{Var}(\ell(x^\delta)(t^*)) = \left\langle \int_0^{t^*} e^{-A^*A(t^*-s)}Qe^{-A^*A(t^*-s)}[f(s)]^2 ds\,\ell\,,\,\ell \right\rangle. \quad (18)$$

Moments

- A set of points $\Theta = \{\vec{r}_i\}_{i=1}^m \subset \mathbb{R}^d$.
- Uncertainty quantification of quantities $\left\{x^\delta(t^*, \vec{r}_i)\right\}_{i=1}^m$.

$$\bar{\mathbf{x}}_i \equiv \mathbb{E}x^\delta(t^*,\vec{r}_i) = \left[e^{-A^*At}x_0 + \int_0^t e^{-A^*A(t-s)}A^*y^\delta ds\right](\vec{r}_i) \tag{19}$$

$$\sigma_i \equiv \mathrm{Var}(x^\delta(t^*, \vec{r}_i)) = \sum_{j=1}^{\infty} q_j u_j^2(\vec{r}_i) \int_0^t e^{-2\lambda_j^2(t-s)}[f(s)]^2 ds. \tag{20}$$

$$\mathbb{E}|x^\delta(t^*,\vec{r}_i)|^p = 2^{\frac{p}{2}}\pi^{-\frac{1}{2}}\Gamma\left(\frac{1+p}{2}\right)\sigma_i^p\, {}_1F_1\left(-\frac{p}{2}, \frac{1}{2}, -\frac{\bar{\mathbf{x}}_i^2}{2\sigma_i^2}\right), \tag{21}$$

where $\Gamma(\cdot)$ and ${}_1F_1(\cdot,\cdot,\cdot)$ are gamma function and Kummer's function of the first kind.

- Application in Biosensor Tomography.

Confidence interval I

- $\hat{\mathbf{x}}_i := \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_{i,j}$ and $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_{i,j} - \hat{\mathbf{x}}_i)^2$.
- $100(1-\alpha)\%$ confidence interval of $\mathbb{E}x^{\delta}(t^*,\vec{r}_i)$:

$$\mathbb{E}x^{\delta}(t^*,\vec{r}_i) \in \left[\hat{\mathbf{x}}_i - t_{n-1,1-\frac{\alpha}{2}}\frac{s_i}{\sqrt{n}}\ ,\ \hat{\mathbf{x}}_i + t_{n-1,1-\frac{\alpha}{2}}\frac{s_i}{\sqrt{n}}\right], \tag{22}$$

  where $t_{n-1,1-\frac{\alpha}{2}}$ represents the $\left(1-\frac{\alpha}{2}\right)$-th quantile of the t-distribution.
- By the asymptotic distributions of $\hat{\mathbf{x}}_i$, the approximate formula for (22):

$$\mathbb{E}x^{\delta}(t^*,\vec{r}_i) \in \left[\hat{\mathbf{x}}_i - |z_{\alpha/2}|\frac{s_i}{\sqrt{n}}\ ,\ \hat{\mathbf{x}}_i + |z_{\alpha/2}|\frac{s_i}{\sqrt{n}}\right], \tag{23}$$

  where $z_{\alpha/2}$ is the standard normal quantile. For $\alpha = 0.05$, $|z_{\alpha/2}| \approx 1.96$.

Confidence interval II

Definition

$\varphi : (0,\infty) \to (0,\infty)$ is called an index function if it is continuous and strictly increasing, and $\lim_{\lambda\to 0+}\varphi(\lambda) = 0$. Let $\mathcal{I}$ denote the set of all index functions.

Proposition

Suppose that $x^\dagger, x_0 \in H^r(\mathbb{R}^d)$ with $r > d/2$. Let x^δ_{as} the solution of (50) without random term (i.e. $f(t) \equiv 0$) [a]. If

$$\left\| x^\delta_{as}(t^*) - x^\dagger \right\|_{H^r(\mathbb{R}^d)} \le C_{as} \cdot \varphi(\delta), \tag{24}$$

where C_{as} is a constant and $\varphi \in \mathcal{I}$. Then, there holds

$$\mathbf{P}\left(\left| \hat{\mathbf{x}}_i - x^\dagger(\vec{r}_i) \right| \le C_i \cdot \left(\frac{1}{\sqrt{n}} + \varphi(\delta)\right) \right) \ge 1 - \alpha, \tag{25}$$

where $C_i := \max\left(t_{n-1,1-\frac{\alpha}{2}} \cdot s_i, \|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)} C_{as} \right)$.

[a] x^δ_{as} coincides with the conventional asymptotical regularization solution with the regularity in $H^r(\mathbb{R}^d)$.

Confidence interval

Proposition

Consequently, if $n = c_1[\varphi(\delta)]^{-2}$ *with a fixed* $c_1 > 0$*, then*

$$\mathbf{P}\left(\left|\hat{\mathbf{x}}_i - x^\dagger(\vec{r}_i)\right| \leq \tilde{C}_i \cdot \varphi(\delta)\right) \geq 1 - \alpha \tag{26}$$

with $\tilde{C}_i := \max\left(2c_1^{-1/2} s_i, \|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)} C_{as}\right)$.
Furthermore, if $\alpha = \psi(\delta)$ *with* $\psi \in \mathcal{I}$*, and* $n = [\varphi(\delta)]^{-2} \ln\left([\psi(\delta)]^{-1}\right)$*, then*

$$\mathbf{P}\left(\left|\hat{\mathbf{x}}_i - x^\dagger(\vec{r}_i)\right| \leq \tilde{\tilde{C}}_i \cdot \varphi(\delta)\right) \geq 1 - \psi(\delta), \tag{27}$$

where $\tilde{\tilde{C}}_i := \max\left(2\sqrt{2} s_i, \|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)} C_{as}\right)$.

Confidence interval

Proposition

As a consequence, $\hat{\mathbf{x}}_i \to x^\dagger(\vec{r}_i)$ *in probability when* $\delta \to 0$.

Example

- $100(1-\alpha)\%$ confidence interval (with $\alpha = \delta^q$) of the estimate $\hat{\mathbf{x}}_i$ for IP (1) under Hölder-type source conditions $x^\dagger \in \mathcal{R}((A^*A)^p)$.
- By the standard argument, the inequality (24) holds with $\varphi(\delta) = \delta^p$. Hence, if we set $n = \delta^{-2p}\ln(\delta^{-q})$, for small enough δ it holds

$$\mathbf{P}\left(\left|\hat{\mathbf{x}}_i - x^\dagger(\vec{r}_i)\right| \leq \tilde{\tilde{C}}_i \cdot \delta^p\right) \geq 1 - \delta^q.$$

Error estimation I

- Approximate lower/upper estimator.

$$\mathbf{x}_i^l := \hat{\mathbf{x}}_i - |z_{\alpha/2}|\frac{s_i}{\sqrt{n}} - C_{as}\|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)}\varphi(\delta), \ \mathbf{x}_i^u := \hat{\mathbf{x}}_i + |z_{\alpha/2}|\frac{s_i}{\sqrt{n}} + C_{as}\|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)}\varphi(\delta). \quad (28)$$

For fixed δ the probability that $x^\dagger(\vec{r}_i)$ fall outside of the interval

$$\left[\hat{\mathbf{x}}_i - C_{as}\|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)}\varphi(\delta)\ ,\ \hat{\mathbf{x}}_i + C_{as}\|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)}\varphi(\delta)\right] \subset \left[\mathbf{x}_i^l, \mathbf{x}_i^u\right]$$

converges to 0 as $n \to \infty$.

- $100(1-\alpha)\%$ confidence L^2-error as

$$\Delta^2 := \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i^u - \mathbf{x}_i^l)^2. \quad (29)$$

- For $\alpha = \psi(\delta)$ and $n = [\varphi(\delta)]^{-2}\ln\left([\psi(\delta)]^{-1}\right)$ with $\psi, \varphi \in \mathcal{I}$,
$\Delta \leq 2\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\sqrt{2}s_i + C_{as}\|\delta_{\vec{r}_i}\|_{H^{-r}(\mathbb{R}^d)}\right)^2}\varphi(\delta) \to 0$ as $\delta \to 0$.

Error estimation II

- Pointwise error estimate

$$x^{l,u}(r) := \frac{\mathbf{x}^{l,u}_{i+1} - \mathbf{x}^{l,u}_{i}}{r_{i+1} - r_i} r + \frac{\mathbf{x}^{l,u}_{i} r_{i+1} - \mathbf{x}^{l,u}_{i+1} r_i}{r_{i+1} - r_i}, \quad r \in [r_i, r_{i+1}], \quad i = 1, \cdots, m.$$

$$x^l(r) \leq x(r) \leq x^u(r), \quad \forall\, r \in \mathbb{R}^d,$$

$$\Delta(r) := x^u(r) - x^l(r) = \mathcal{O}\left(\varphi(\delta)\right), \quad \forall\, r \in \mathbb{R}^d.$$

- for $\alpha = \psi(\delta)$ and $n = [\varphi(\delta)]^{-2} \ln\left([\psi(\delta)]^{-1}\right)$ with $\psi, \varphi \in \mathcal{I}$, for smoothing $x^\dagger$ and a small enough interval $[r_i, r_{i+1}]$, for all $r \in [r_i, r_{i+1}]$ it holds

$$\Delta(r) \leq \max\left(\mathbf{x}^u_i - \mathbf{x}^l_i, \mathbf{x}^u_{i+1} - \mathbf{x}^l_{i+1}\right) \leq 2 \max_{j \in \{i, i+1\}} \left(\sqrt{2} s_j + C_{as} \|\delta_{\vec{r}_j}\|_{H^{-r}(\mathbb{R}^d)}\right) \cdot \varphi(\delta) \to 0$$

Outline

Spectral analysis

- $x_0 \in N(A)^\perp$.
- $\{\lambda_j; u_j, v_j\}_{j=1}^{\infty}$: singular system for A: $Au_j = \lambda_j v_j$, $A^* v_j = \lambda_j u_j$, $\|A\| = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_j \geq \lambda_{j+1} \geq \cdots \to 0$ as $j \to \infty$.
- $\tilde{x}^\delta(t) = \sum_j \xi_j(t) u_j + \hat{x}(t)$.
- $\mathbb{E}\|x^\delta(t) - x_0\|^2 \leq \mathbb{E}\|\tilde{x}^\delta(t) - x_0\|^2$ for $x_0 \in N(A)^\perp$ and $\hat{x}(t) \in N(A)$.
- $$\langle dx^\delta, u_j \rangle = \langle A^*(y^\delta - Ax^\delta) dt, u_j \rangle + \langle f(t) dB_t, u_j \rangle, \ j = 1, 2, \cdots . \tag{30}$$
- $$d\xi_j(t) = \left(\lambda_j \langle y^\delta, v_j \rangle - \lambda_j^2 \xi_j(t)\right) dt + f(t) d\beta_j(t), \ \xi_j(0) = \langle x_0, u_j \rangle. \tag{31}$$

Spectral analysis

Proposition

The stochastic differential equation (31) has a unique solution

$$\xi_j(t) = e^{-\lambda_j^2 t}\langle x_0, u_j\rangle + \frac{1-e^{-\lambda_j^2 t}}{\lambda_j}\langle y^\delta, v_j\rangle + \int_0^t e^{-\lambda_j^2(t-s)} f(s) d\beta_j(s),$$

where $\int_0^t e^{-\lambda_j^2(t-s)} f(s) d\beta_j(s)$ is Gaussian $\mathcal{N}\left(0, \int_0^t e^{-2\lambda_j^2(t-s)}[f(s)]^2 ds\right)$. $\xi_j(t)$ is also Gaussian with mean

$$\mathbb{E}\xi_j(t) = e^{-\lambda_j^2 t}\langle x_0, u_j\rangle + \frac{1-e^{-\lambda_j^2 t}}{\lambda_j}\langle y^\delta, v_j\rangle \tag{32}$$

and variance

$$\mathbb{E}\left(\xi_j(t) - \mathbb{E}\xi_j(t)\right)^2 = \int_0^t e^{-2\lambda_j^2(t-s)}[f(s)]^2 ds. \tag{33}$$

Consequently, if $f \in \mathcal{S}$, then $\xi_j(t) \sim \mathcal{N}\left(\frac{\langle y^\delta, v_j\rangle}{\lambda_j}, 0\right)$ as $t \to \infty$.

Regularization

Regularized solution

$x^\delta(t) = \sum_j \xi_j(t) u_j \Rightarrow$

$$x^\delta(t) = (1 - A^*Ag(t, A^*A))x_0 + g(t, A^*A)A^*y^\delta + \int_0^t e^{-A^*A(t-s)} f(s) dB_s, \tag{34}$$

$$g(t, \lambda) = \frac{1 - e^{-\lambda t}}{\lambda}. \tag{35}$$

Theorem on Regularization

If the terminating time $t^* = t^*(\delta, y^\delta)$ is chosen so that

$$\lim_{\delta \to 0} t^* = \infty, \quad \lim_{\delta \to 0} \delta \cdot t^* = 0, \tag{36}$$

the $x^\delta(t^*)$ converges to $x^\dagger$ in the sense of mean square as $\delta \to 0$.

Proof

- $\mathbb{E}\|x^\delta(t) - x^\dagger\|^2 = \|\mathbb{E}x^\delta(t) - x^\dagger\|^2 + \mathbb{E}\|x^\delta(t) - \mathbb{E}x^\delta(t)\|^2$.
- $$\begin{aligned}\|\mathbb{E}x^\delta(t) - x^\dagger\| &= \|r(t, A^*A)(x_0 - x^\dagger) + g(t, A^*A)A^*(y^\delta - y)\| \\ &\leq \|e^{-tA^*A}(x_0 - x^\dagger)\| + \vartheta t^{1/2}\delta,\end{aligned} \tag{37}$$

  where we have used

  $$\|g(t, A^*A)A^*(y^\delta - y)\| \leq \sup_{\lambda \in (0, \|A\|^2)} \sqrt{\lambda} g(t, \lambda)\|y^\delta - y\| \leq \delta \sup_{\lambda \in (0, \|A\|^2)} \frac{1 - e^{-\lambda t}}{\sqrt{\lambda}} \leq \vartheta t^{1/2}\delta,$$

  where $\vartheta = \sup_{\lambda \in \mathbf{R}_+} \sqrt{\lambda}(\lambda - e^{-\lambda}) \approx 0.6382$. Hence,

  $$\|\mathbb{E}x^\delta(t) - x^\dagger\| \to 0 \quad \text{as} \quad \delta \to 0. \tag{38}$$

- Ito-isometry: $\mathbb{E}\left\|\int_0^t g(s)dB_s\right\|^2 = \mathbb{E}\int_0^t \|g(s)\|^2 ds + f(t) \in \mathcal{S} \Rightarrow$

  $$\begin{aligned}\mathbb{E}\|x^\delta(t) - \mathbb{E}x^\delta(t)\|^2 &= \mathbb{E}\left\|\int_0^t e^{-A^*A(t-s)} f(s)dB_s\right\|^2 \\ &= \mathbb{E}\int_0^t \left\|e^{-A^*A(t-s)} f(s)\right\|^2 ds = \mathbb{E}\int_0^t tr(e^{-2A^*A(t-s)}[f(s)]^2)ds \to 0\end{aligned} \tag{39}$$

  as $t \to \infty$.

Outline

Convergence rates with noisy data

A priori stopping rule

Let $x^\delta(t)$ be solution of (50) with $f(t) \in \mathcal{S}_{C\cdot\phi}$. Then, under the source condition $x_0 - x^\dagger = \varphi(A^*A)v, \|v\| \leq \rho, \varphi \in \mathcal{S}_{C_1}$, if $t^* = \Theta^{-1}(\delta)$ with $\Theta(t) = t^{-1/2}\varphi(t^{-1})$, we have

$$\mathbb{E}\|x^\delta(t^*) - x^\dagger\|^2 = \mathcal{O}([\varphi([\Theta^{-1}(\delta)]^{-1})]^2) \quad \text{as} \quad \delta \to 0.$$

If $\varphi = \varphi_p(\lambda) = \lambda^p$, we have $\mathbb{E}\|x^\delta(t^*) - x^\dagger\|^2 = \mathcal{O}(\delta^{\frac{4p}{2p+1}})$
If $\varphi = \varphi_\mu(\lambda) = \log^{-\mu}(1/\lambda)$, we have $\mathbb{E}\|x^\delta(t^*) - x^\dagger\|^2 = \mathcal{O}(\log^{-2\mu}(\delta^{-1}))$.

Proof.

$$\begin{aligned} \mathbb{E}\|x^\delta(t) - x^\dagger\|^2 &= \|\mathbb{E}x^\delta(t) - x^\dagger\|^2 + \mathbb{E}\|x^\delta(t) - \mathbb{E}x^\delta(t)\|^2 \\ &\leq \left(\|e^{-tA^*A}(x_0 - x^\dagger)\| + \vartheta t^{1/2}\delta\right)^2 + \mathbb{E}\int_0^t tr(e^{-2A^*A(t-s)}[f(s)]^2)ds \\ &\leq 2C_1^2\rho^2[\varphi(1/t)]^2 + 2\vartheta^2 t\delta^2 + C^2[\varphi(1/t)]^2. \end{aligned} \tag{40}$$

□

Convergence rates with noisy data

- Discrepancy principle:

$$t_i^* := \inf\{t > 0 : \chi_i(t) < 0\}, \quad i = 1, 2, \tag{41}$$

$$\chi_1(t) := \|A\mathbb{E}x^\delta(t) - y^\delta\| - \tau\delta, \quad \chi_2(t) := \mathbb{E}\|Ax^\delta(t) - y^\delta\|^2 - \tau\delta^2, \quad f(t) \in \mathcal{S}, \tau > 1.$$

Existence

If $\|Ax_0 - y^\delta\| > \tau\delta$, there always exists a unique t_i^* in (41).

A posteriori stopping rule, $f(t) \in \mathcal{S}_{C\cdot\phi}$

(i) Under the Hölder-type source conditions φ_p:

$$t^* = \mathcal{O}\left(\delta^{-\frac{2}{2p+1}}\right) \quad \text{and} \quad \mathbb{E}\|x^\delta(t^*) - x^\dagger\|^2 = \mathcal{O}\left(\delta^{\frac{4p}{2p+1}}\right). \tag{42}$$

(ii) Under the logarithmic source conditions φ_μ:

$$t^* = o\left(\delta^{-\frac{2}{2p+1}} \log^{-\frac{2}{2p+1}}(\delta^{-1})\right) \quad \text{and} \quad \mathbb{E}\|x^\delta(t^*) - x^\dagger\|^2 = \mathcal{O}\left(\log^{-2\mu}(\delta^{-1})\right). \tag{43}$$

Outline

Converse results

- Spectral tail:

$$\omega(\lambda) = \sum_{j:\ \lambda_j^2 < \lambda} \langle x_0 - x^\dagger, u_j \rangle u_j. \tag{44}$$

$$\|\mathbb{E}x(t) - x^\dagger\|^2 = \int_0^{\|A\|^2} e^{-\lambda t} d\omega(\lambda), \tag{45}$$

Convergence rates with exact data and converse results

Let $\varphi \in \mathcal{S}_{C_2}^{\sigma}$. Then, the following two statements are equivalent:

(i) $\left\|\mathbb{E}x(t) - x^\dagger\right\|^2 \leq C_3 \varphi(1/t)$ for all $t > 0$.

(ii) $\omega(\lambda) \leq C_4 \varphi(\lambda)$ for all $\lambda > 0$.

If $f \in \mathcal{S}_{C \cdot \sqrt{\varphi}}$, every one of above statements is also equivalent to:

(iii) $\mathbb{E}\left\|x(t) - x^\dagger\right\|^2 \leq C_5 \varphi(1/t)$ for all $t > 0$.

Remark

- The following two statements are equivalent:
  (i) There exists a constant $C > 0$ with

$$\omega(\lambda) \leq C_a \varphi^{2\nu}(\lambda) \quad \text{for all } \lambda > 0.$$

  (ii) There exists a constant $C_b > 0$ such that

$$\left|\langle x_0 - x^\dagger, x\rangle\right| \leq C_b \|\varphi(L^*L)x\|^\nu \|x\|^{1-\nu} \quad \text{for all } x \in X. \tag{46}$$

- $x_0 - x^\dagger \in \mathcal{R}(\psi^\nu(A^*A))$ implies the variational inequality.
- Conversely the variational inequality implies that the relation $x_0 - x^\dagger \in \mathcal{R}(\psi^\nu(L^*L))$ holds for every continuous function φ with $\psi \geq c\varphi^\mu$ for $c > 0$ and $\mu \in (0, \nu)$.

The best worst case mean square error

- $B_\delta(y) := \{\tilde{y} \in \mathcal{Y} : \ \|\tilde{y} - y\| \le \delta\}$.
- $x(t;\tilde{y})$: solution of (50) with y^δ replaced $\tilde{y} \in B_\delta(y)$.

Convergence rates

Let $\phi(1/\cdot) \equiv \varphi(\cdot) \in \mathcal{S}^g_\zeta$ and denote by $\tilde{\phi}(t) = \sqrt{t^{-1}\phi(t)}$ and $\psi(\delta) = \delta^2 \tilde{\phi}^{-1}(\delta)$.
Then, the following two statements are equivalent:

(a) There exists a constant $c > 0$ such that

$$\sup_{\tilde{y} \in \tilde{B}_\delta(y)} \inf_{t>0} \mathbb{E}\|x(t,\tilde{y}) - x^\dagger\|^2 \le c\psi(\delta) \quad \text{for all } \delta > 0. \tag{47}$$

(b) There exists a constant $\tilde{c} > 0$ such that

$$\mathbb{E}\|x(t) - x^\dagger\|^2 \le \tilde{c}\phi(t) \quad \text{for all } t > 0. \tag{48}$$

Outline

Discrepancy principle

Stochastic discrepancy principle for two stochastic regularization methods, namely

- Optimal stopping time

$$t^* = \min\left\{t \geq 0 : \|Ax(t) - y^\delta\| \leq \kappa\delta\right\}, \quad \kappa > 1, \tag{49}$$

for generalized stochastic asymptotical regularization:

$$dx^\delta = \phi(A^*A)A^*(y^\delta - Ax^\delta)dt + f(t)dB(t), \quad x^\delta(0) = x_0. \tag{50}$$

- Optimal stopping iteration

$$k^* = \min\left\{k \geq 0 : \|Ax_k - y^\delta\| \leq \kappa\delta\right\}, \quad \kappa > 1, \tag{51}$$

for generalized stochastic Landerweber iteration:

$$x^\delta_{k+1} = x^\delta_k + \phi(A^*A)A^*(y^\delta - Ax^\delta_k) + w_{k+1}, \quad x^\delta(0) = x_0, \tag{52}$$

Discrepancy principle

Theorem (Hint: k^* is a martingale)

Assume that

$$\phi(\lambda) \in [\beta', \beta], \sqrt{\lambda}\phi(\lambda) \leq \gamma, \lambda\phi(\lambda) \leq 1, \inf_{\lambda \in (\alpha, \|A\|]} \lambda\phi(\lambda) > 0, \quad \varepsilon \leq \eta\delta. \tag{53}$$

Then,

$$\mathbb{E}(k^*) \leq \frac{\|x_0 - x^\dagger\|^2}{\mu\delta^2}, \qquad \lim_{\delta \to 0} \delta^2 \mathbb{E}(k^*(\delta)) = 0. \tag{54}$$

Assume further that

$$\varepsilon \leq \eta_2 \delta^2. \tag{55}$$

Then, we have

$$\lim_{\delta \to 0} \mathbb{E}(\|x_{k^*} - x^\dagger\|) = 0. \tag{56}$$

Proof.

Doob's optional stopping theorem + inequalities. □

Thank you for your attention!

Questions?