On Fractional Tikhonov Regularization

Daniel Gerth
joint work with E. Klann, R. Ramlau and L. Reichel

Technical University Chemnitz

October 27, 2015

- Introduction
- Convergence Theory
- fractional vs. standard Tikhonov regularization

Overview

- Introduction
- Convergence Theory
- fractional vs. standard Tikhonov regularization

- We look for the solution x of a linear ill-posed problem $Ax = b$, $A \in \mathcal{L}(X, Y)$ compact, X, Y Hilbert spaces.
- only noisy data $b^\delta = b + \epsilon$ is available:

  (deterministic) $||\epsilon|| = ||b - b^\delta|| \leq \delta$, $\delta > 0$

  (stochastic) $\mathbb{E}(||\epsilon||) = f(\eta) < \infty$, $f(\eta) \to 0$ as $\eta \to 0$
- Recall Tikhonov regularization:

$$x_\mu^\delta = \min_{x \in X} ||Ax - b^\delta||^2 + \mu ||x||^2,$$

where x is typically obtained via

$$(A^* A + \mu I) x_\mu^\delta = A^* b.$$

Tikhonov regularization is known to be oversmoothing due to A^*. The idea of Fractional Tikhonov methods is to lessen this influence.

- Hochstenbach, Reichel (2011): use a weighted residual
- similar concept in Louis (1989) and Mathé–Tautenhahn (2011)

$$x_\mu^\delta = \min_{x \in X} \|Ax - b^\delta\|_W^2 + \mu\|x\|_X^2$$

with $\|y\|_W := \|W^{1/2}y\|_Y$, $W = (AA^*)^{(\alpha-1)/2}$, $0 \le \alpha \le 1$.
One obtains x_μ^δ via

$$((A^*A)^{(\alpha+1)/2} + \mu I)x = (A^*A)^{(\alpha-1)/2} A^* b^\delta$$

or equivalently

$$(A^*A + \mu(A^*A)^{\frac{1-\alpha}{2}})x_\mu^\delta = A^*b^\delta$$

which corresponds to

$$x_\mu^\delta = \min_{x \in X} \|Ax - b^\delta\|^2 + \mu\|Bx\|_X^2$$

with $B^*B = (A^*A)^{\frac{1-\alpha}{2}}$ a.k.a. *generalized Tikhonov regularization*.

The second approach is due to Klann and Ramlau (2008). Find x_μ^δ as solution of

$$(A^*A + \mu I)^\alpha x = (A^*A)^{\alpha-1} A^* b^\delta,$$

for $0 < \alpha \leq 1$. There is no (known) associated minimization problem.

For $\alpha = 1$ both approaches coincide with Tikhonov regularization.

Let us compare the filter functions:

- Tikhonov regularization:

$$F_\mu(\sigma) = \frac{\sigma^2}{\sigma^2 + \mu}$$

- fractional Tikhonov after Hoechstenbach-Reichel, later referred to as method (1.7)-(1.8):

$$F_\mu(\sigma) = \frac{\sigma^{\alpha+1}}{\sigma^{\alpha+1} + \mu}$$

- fractional Tikhonov after Klann-Ramlau, later referred to as method (1.10):

$$F_\mu(\sigma) = \left(\frac{\sigma^2}{\sigma^2 + \mu}\right)^\alpha$$

Overview

- Introduction
- Convergence Theory
- fractional vs. standard Tikhonov regularization

The following Theorems can be extended to all regularizing filters[1] with

$$\sup_{0<\sigma\leq\sigma_1} \sigma^{-1}|F_{\mu,\alpha}(\sigma)| \leq c\mu^{-\beta} \tag{1}$$

$$\sup_{0<\sigma\leq\sigma_1} |1-F_{\mu,\alpha}(\sigma)|\sigma^{\nu^*} \leq c_{\nu^*}\mu^{\beta\nu^*}. \tag{2}$$

Assume $x^\dagger = A^\dagger b$ fulfils, with $\nu > 0$ and constant ρ,

$$x^\dagger \in \text{range}((A^*A)^{\nu/2}) \quad \text{and} \quad \|x^\dagger\|_\nu := \left(\sum_{n\geq 1} \sigma_n^{-2\nu}|\langle x^\dagger, u_n\rangle|^2\right)^{1/2} \leq \rho$$

[1] A filter $F_{\mu,\alpha}$ is said to be regularizing, if $\sup_n |F_{\mu,\alpha}(\sigma_n)\sigma_n^{-1}| = c(\mu) < \infty$, $\lim_{\mu\to 0} F_{\mu,\alpha} = 1$ point wise in σ_n and $|F_{\mu,\alpha}(\sigma_n)| \leq c$ for all μ and σ_n

A-priori parameter choice

Assume $\mathbb{E}(||b - b^\delta||) = f(\eta) < \infty$.

- For all $-1 < \alpha \leq 1$ and $0 \leq \nu \leq \alpha + 1$ the fractional Tikhonov method (1.7)-(1.8) of Hoechstenbach and Reichel fulfills

$$\mathbb{E}\|x^\dagger - x_\mu^\delta\|_X \leq c\, f(\eta)^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}} \quad \text{with} \quad \mu = C \left(\frac{f(\eta)}{\rho} \right)^{(\alpha+1)/(\nu+1)}$$

- For $\alpha \in (1/2, 1]$, the fractional Tikhonov method (1.10) of Klann and Ramlau fulfills

$$\mathbb{E}\|x^\dagger - x_\mu^\delta\|_X \leq c\, f(\eta)^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}} \quad \text{with} \quad \mu = C \left(\frac{f(\eta)}{\rho} \right)^{1/2(\nu+1)}$$

  for all $0 < \nu < 2$. The result can be extended to $0 < \alpha < \frac{1}{2}$ with presmoothing of the data.

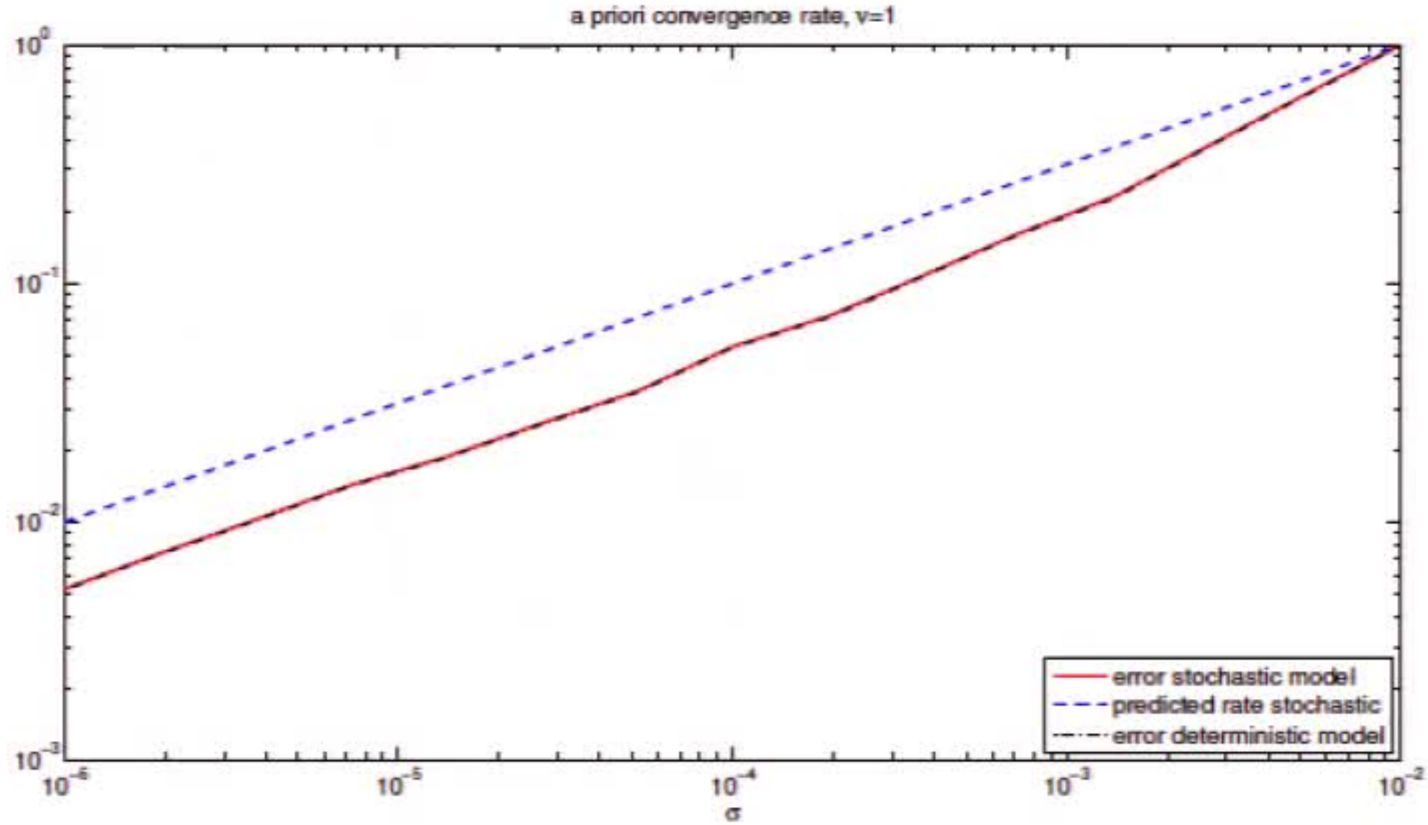
Numerical experiment:

Tikhonov regularization for the inverse heat equation

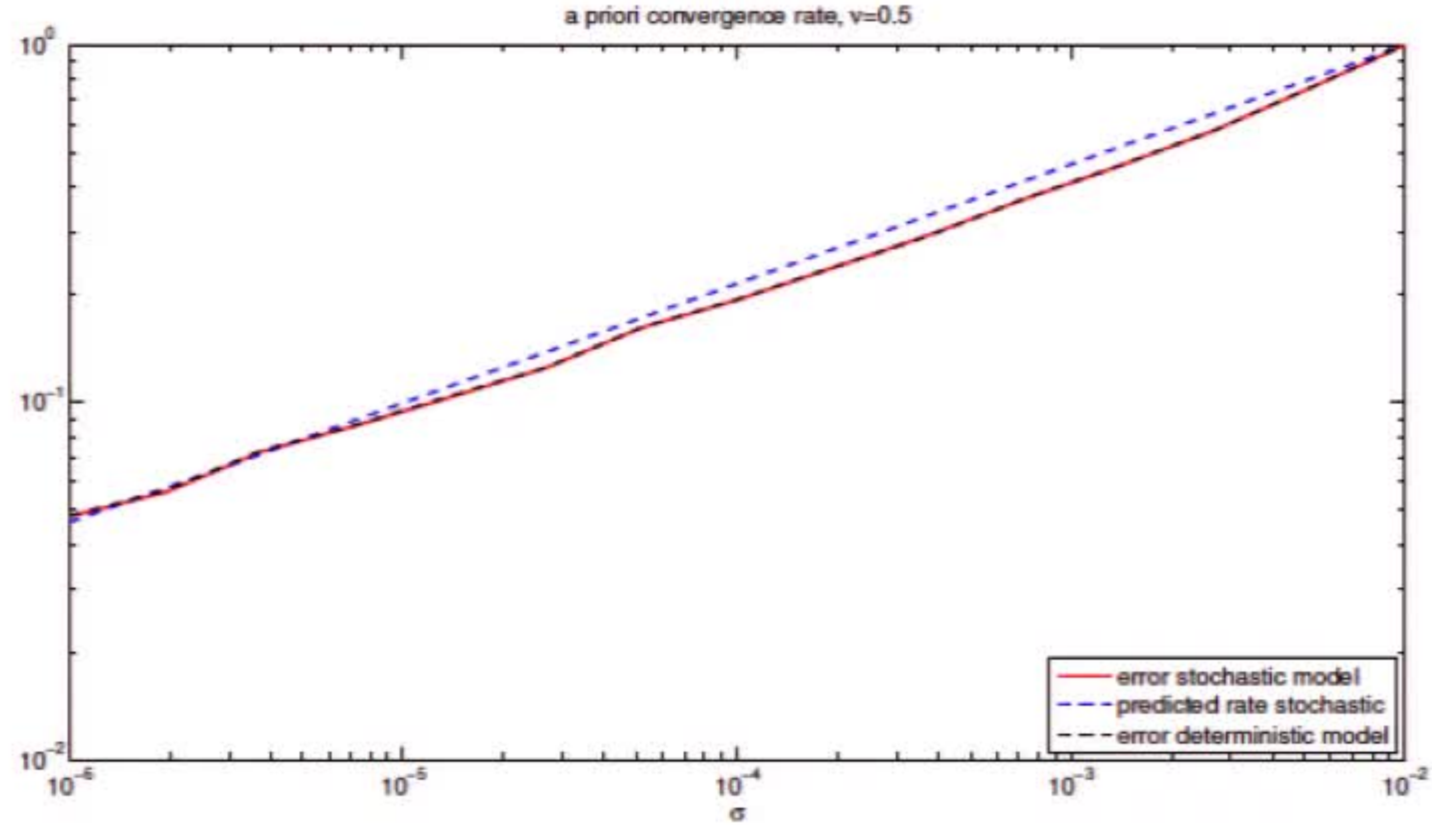
$$y(s) = \int_0^s \frac{t^{-3/2}}{2\sqrt{\pi}} \exp\left(-\frac{1}{4t^2}\right) x(t)\, dt$$

with Gaussian noise $y - y^\delta \sim \mathcal{N}(0, \eta^2 I_n)$ and various values of ν

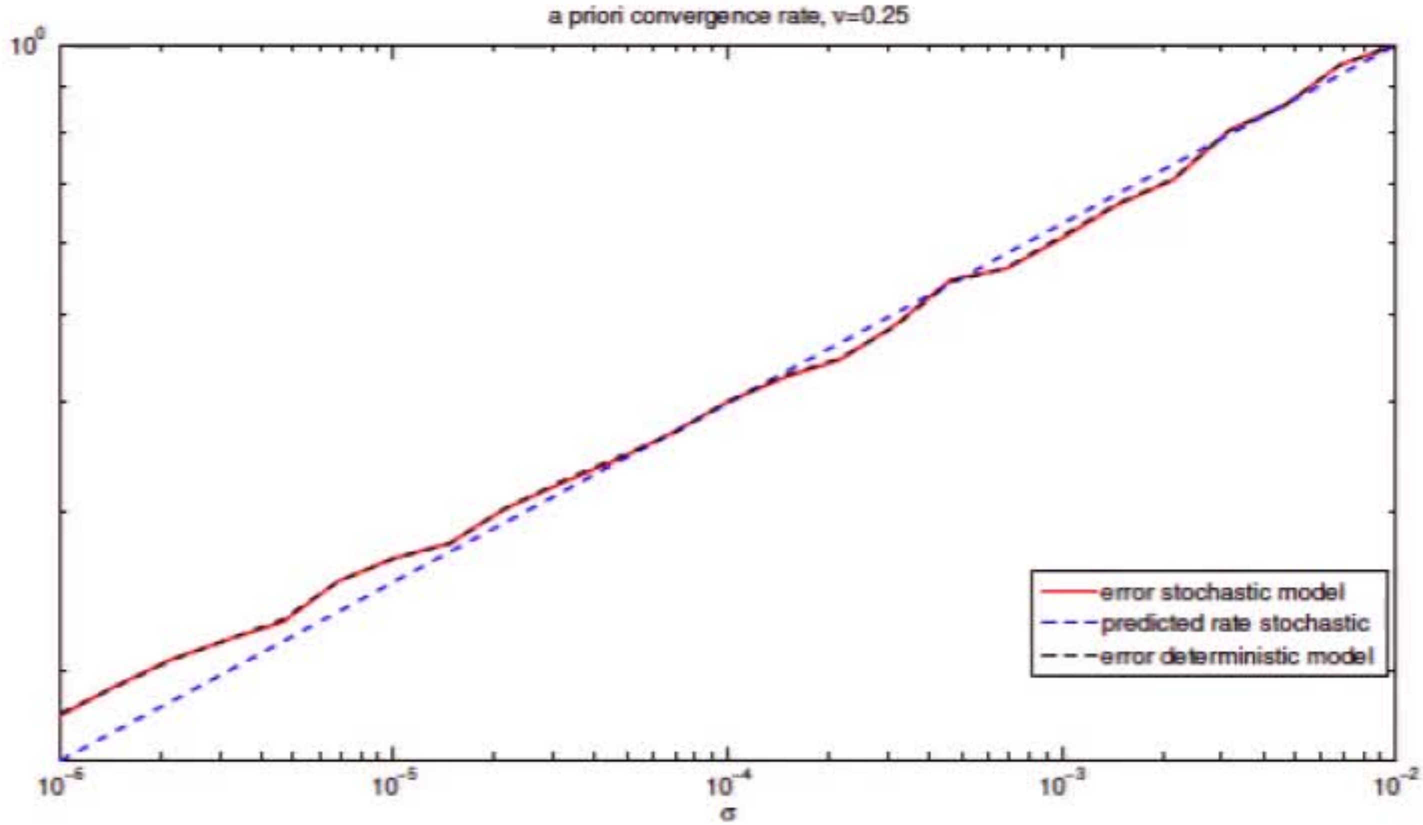
$\nu = 1$


a priori convergence rate, v=1
10^0
10^-1
10^-2
10^-3
10^-6
10^-5
10^-4
10^-3
10^-2
σ
error stochastic model
predicted rate stochastic
error deterministic model

$\nu = 0.5$


a priori convergence rate, v=0.5
10^0
10^-1
10^-2
10^-6
10^-5
10^-4
10^-3
10^-2
σ
error stochastic model
predicted rate stochastic
error deterministic model

$\nu = 0.25$


a priori convergence rate, v=0.25
10^0
10^-6
10^-5
10^-4
10^-3
10^-2
σ
error stochastic model
predicted rate stochastic
error deterministic model

For the rest of the talk we consider the discrepancy principle, i.e., for $||b - b^\delta|| \leq \delta$ find μ s.t.

$$||Ax_\mu^\delta - b^\delta|| = \tau\delta, \qquad \tau > 1$$

- For all exponents $\alpha > 0$ and $0 < \nu \leq \alpha$, the fractional Tikhonov method (1.7)-(1.8) of Hoechstenbach and Reichel fulfills
$$\|x^\dagger - x_\mu^\delta\|_X \leq c\,\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}$$
the regularization parameter μ determined by the discrepancy principle
- For all exponents $\alpha \in (1/2, 1]$ and $0 < \nu \leq 1$, the fractional Tikhonov method of (1.10) of Klann and Ramlau fulfills
$$\|x^\dagger - x_\mu^\delta\|_X \leq c\,\delta^{\frac{\nu}{\nu+1}}\rho^{\frac{1}{\nu+1}}$$
with the regularization parameter μ given by the discrepancy principle

We want to lift the results to the equation

$$||Ax_\mu^\delta - b^\delta|| = \tau \mathbb{E}(||b - b^\delta||), \qquad \tau > 1 \tag{3}$$

given **one** realization of b^δ, μ becomes a random variable!
Let us alter (3) slightly:

$$\mathbb{E}(||Ax_\mu^\delta - b^\delta||) = \tau \mathbb{E}(||b - b^\delta||), \qquad \tau > 1 \tag{4}$$

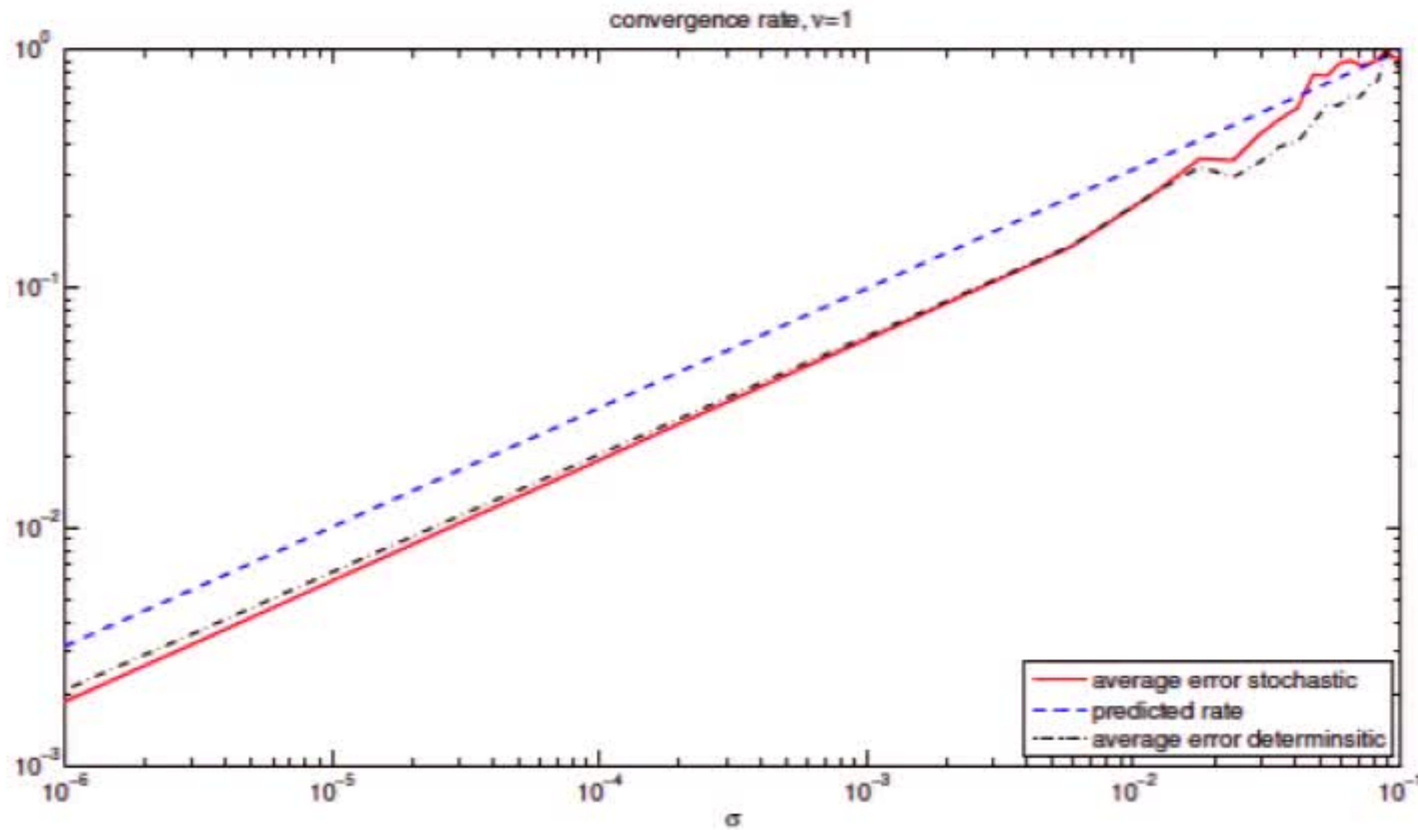
Then μ is not a stochastic quantity anymore and for all Filter methods it holds

$$\mathbb{E}\|x^\dagger - x_\mu^\delta\|_X \leq c\,\mathbb{E}(||\epsilon||)^{\frac{\nu}{\nu+1}} \rho^{\frac{1}{\nu+1}}$$

In practice, approximate $\mathbb{E}(||Ax_\mu^\delta - b^\delta||)$ with several measurements, $\mathbb{E}(\left\|Ax_\mu^\delta - b^\delta\right\|) \approx \frac{1}{N}\sum_{i=1}^N \left\|Ax_\mu^\delta - b^{\delta_i}\right\|$

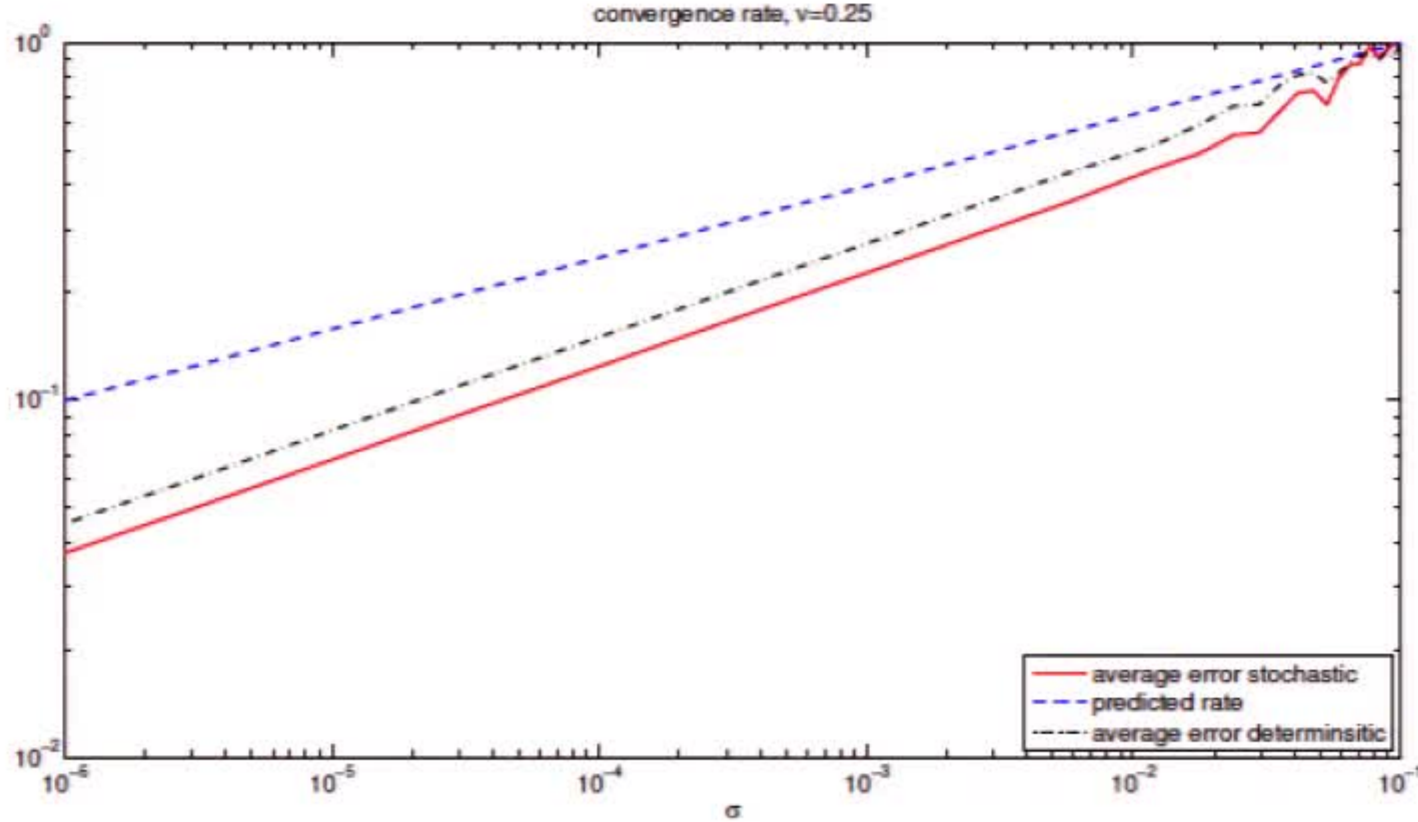
Numerical experiment

standard Tikhonov regularization for the inverse heat equation with Gaussian noise $b - b^\delta \sim \mathcal{N}(0, \eta^2 I_n)$, $\nu = 1$



Numerical experiment

$\nu = 0.25$



Back to (3), we need another alteration

$$\left\|Ax_\mu^\delta - b^\delta\right\| = \tau(\eta)E(\|\epsilon\|) \tag{5}$$

with $\tau(\eta) > 1 \ \forall\eta$, $\tau(\eta) \to \infty$ and $\tau(\eta)\mathbb{E}(||\epsilon||) \to 0$ for $\eta \to 0$.
Then under the previous assumptions

$$\mathbb{P}\left(\|x^\dagger - x_\mu^\delta\|_X \geq c(\tau(\eta)\mathbb{E}(\|\epsilon\|))^{\nu/(\nu+1)}\rho^{1/(\nu+1)}\right) \leq \mathbb{P}\left(||\epsilon|| > \tau(\eta)\mathbb{E}(||\epsilon||)\right)$$

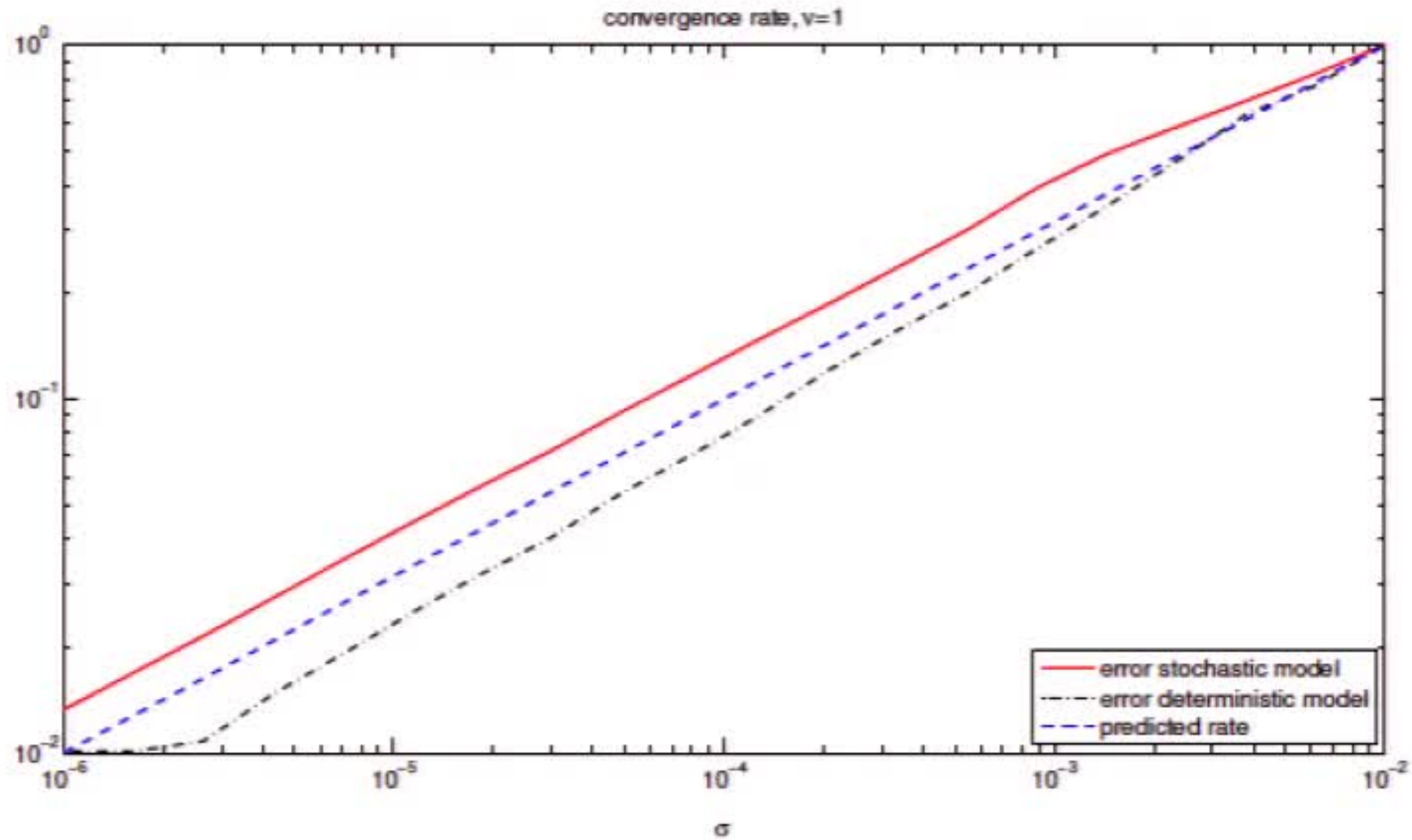
In general

$$\mathbb{P}\left(||\epsilon|| > \tau(\eta)\mathbb{E}(||\epsilon||)\right) \leq \frac{\mathbb{E}(||\epsilon||)}{\tau(\eta)\mathbb{E}(||\epsilon||)} = \frac{1}{\tau(\eta)\tau_0}.$$

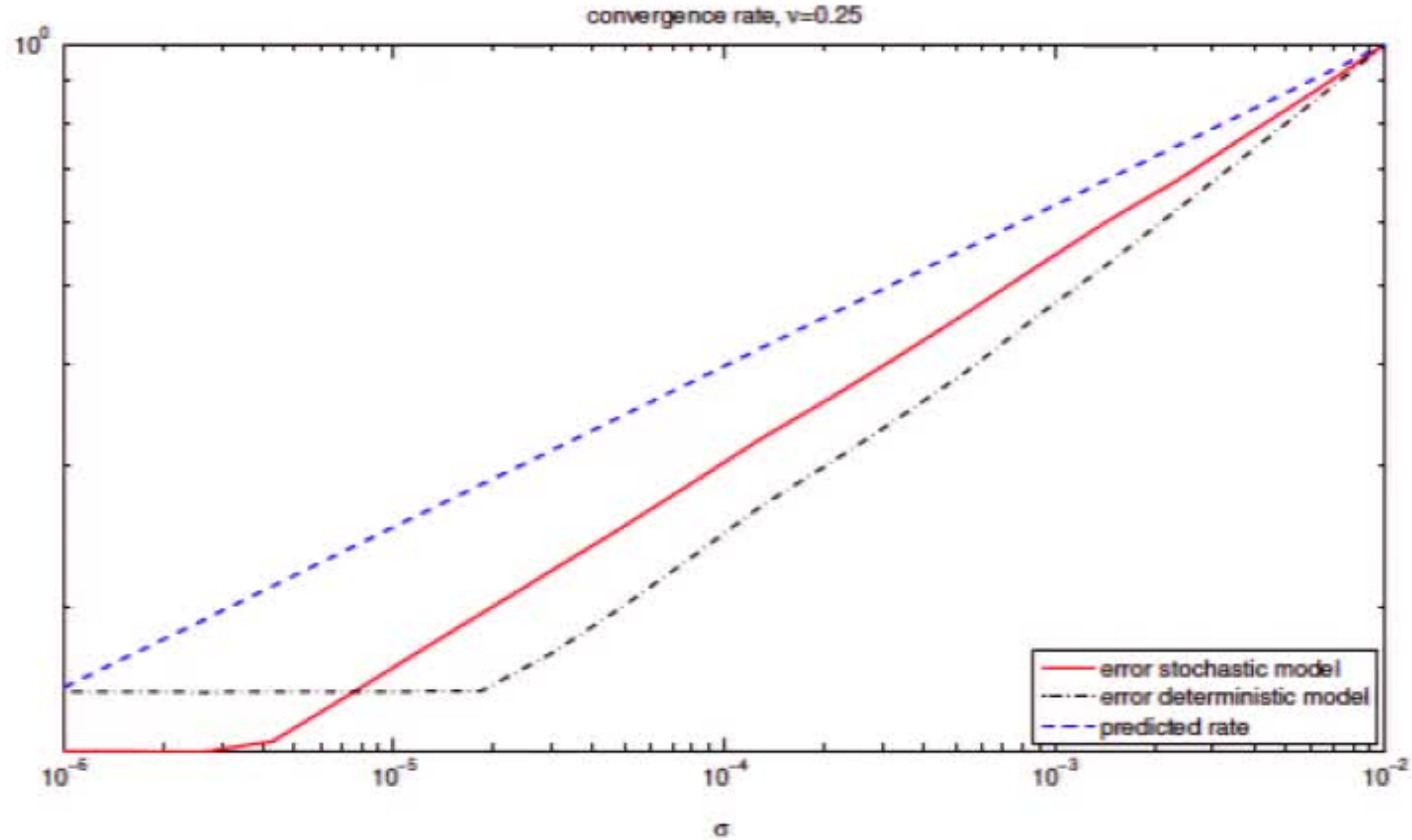
If additional we enforce $\left\|x_\mu^\delta\right\| \leq C_1$ and $|x_\mu^\delta(t)| \leq C_2$, $C_1, C_2 < \infty$, then

$$\mathbb{E}\|x^\dagger - x_\mu^\delta\|_X \to 0 \quad \text{as} \quad \eta \to 0$$

Numerical verification:
Landweber iteration for the inverse heat equation with Gaussian noise $b - b^\delta \sim \mathcal{N}(0, \eta^2 I_n)$, $\nu = 1$



$nu = 0.25$


convergence rate, ν=0.25
10^0
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
σ
error stochastic model
error deterministic model
predicted rate

Overview

- Introduction
- Convergence Theory
- fractional vs. standard Tikhonov regularization

Initial question of the research: How to choose α optimally?

Still not known, but we have the following results:

- with the discrepancy principle, it is $\frac{d\mu(\alpha)}{d\alpha} < 0$
- The reconstruction error is

$$x^\dagger - x_\mu^\delta = \sum_{\sigma_n > 0} (1 - F_{\mu,\alpha}(\sigma_n)) \langle x^\dagger, u_n \rangle u_n + \sum_{\sigma_n > 0} F_{\mu,\alpha}(\sigma_n) \frac{1}{\sigma_n} \langle -\epsilon, v_n \rangle u_n.$$

- the sign of $\frac{dF_{\mu,\alpha}(\sigma)}{d\alpha} = -\frac{d}{d\alpha}(1 - F_{\mu,\alpha}(\sigma))$ changes at some $\sigma_0 > 0$.
- We identified 2 cases where fractional Tikhonov outperforms standard Tikhonov regularization:
  a) the problem is severely ill-posed, i.e., the singular values of A decrease rapidly to zero, and
  b) the error in b^δ is concentrated to low frequencies.

We consider two test problems:

1) a severely ill-posed Fredholm integral equation of the first kind given by

$$b_1(s) = [A_1 x](s) = \int_0^1 \sqrt{s^2+t^2}\, x(t)dt, \quad 0 \le s \le 1, \qquad (6)$$

with error-free data $b_1(s) = \frac{1}{3}\left((1+s^2)^{3/2} - s^3\right)$ and solution $x_1^\dagger(t) = t$, introduced by Fox and Goodwin

2) a mildly ill-posed Volterra integral equation of the first kind

$$b_2(s) = [A_2 x](s) = \int_0^s x(t)dt, \quad 0 \le s \le 1, \qquad (7)$$

with error-free data

$$b_2(s) = \begin{cases} -s & 0 \le s \le 0.5, \\ s-1 & 0.5 < s \le 1, \end{cases}$$

and solution

$$x_2^\dagger(t) = \begin{cases} -1 & 0 \le t \le 0.5, \\ 1 & 0.5 < t < 1. \end{cases}$$

Let us have a look at some reconstructions:

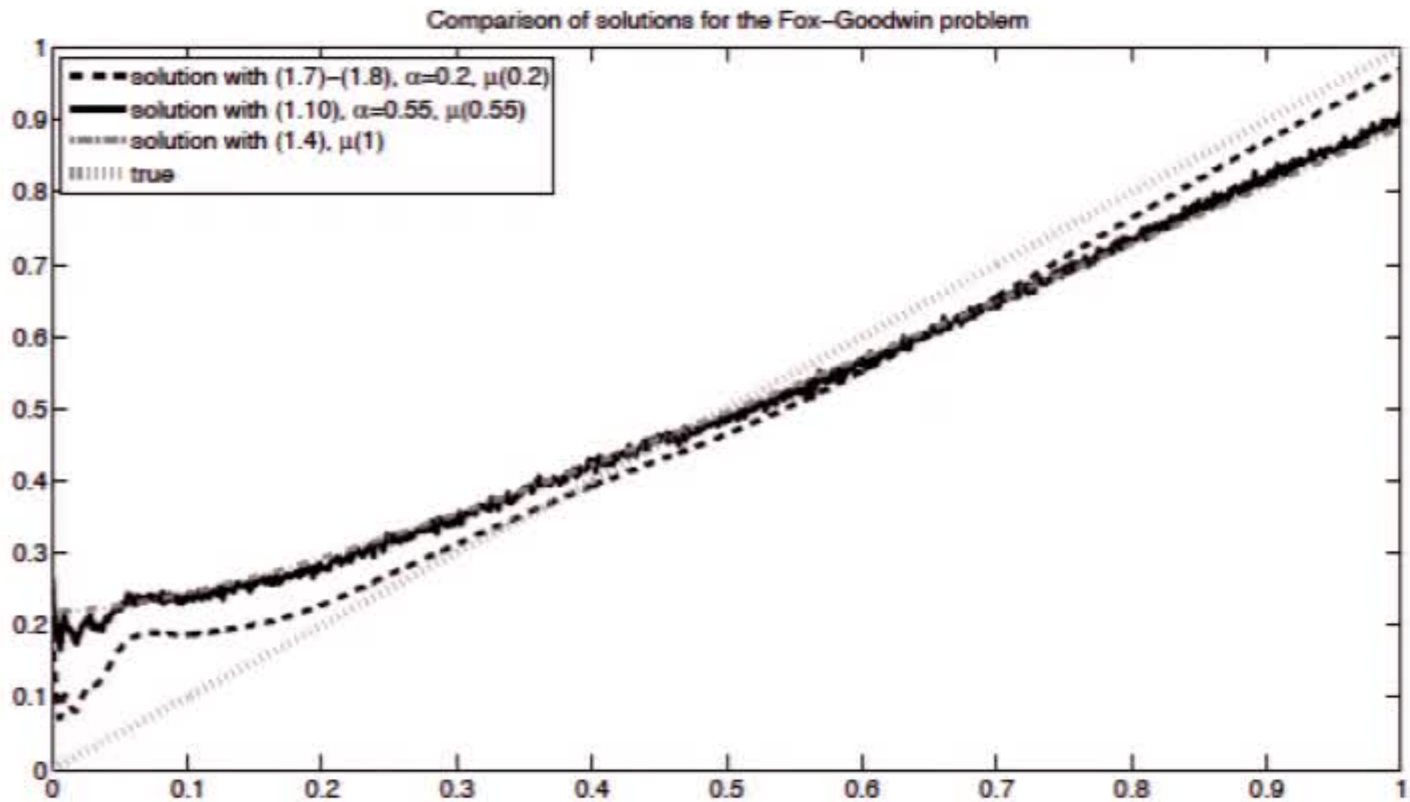


Figure: Comparison of solutions for the Fox–Goodwin problem (6), 5% Gaussian noise, "optimal" α, $\tau = 1.1$.

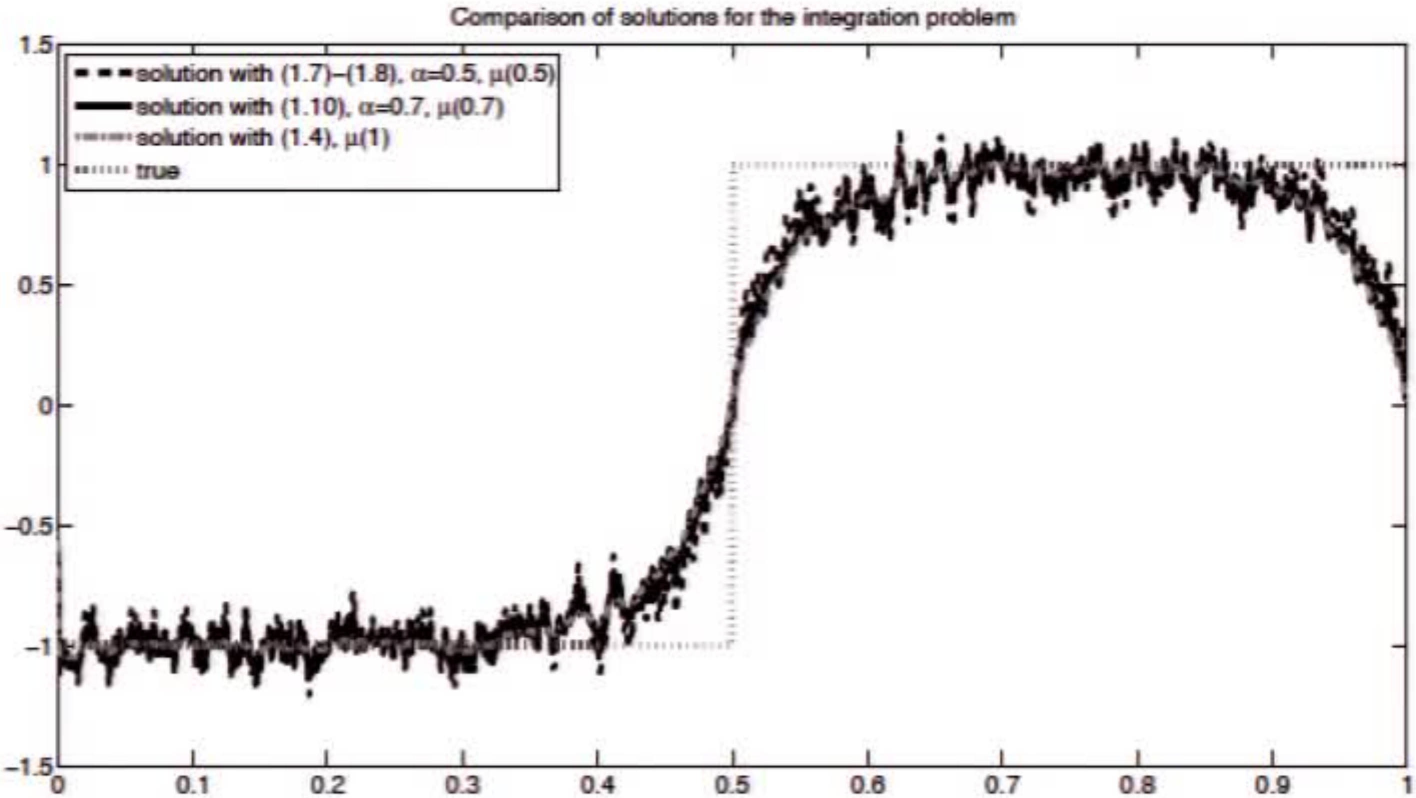


Figure: Comparison of solutions for the integration problem (7) with 5% Gaussian noise, "optimal" α, $\tau = 1.1$.

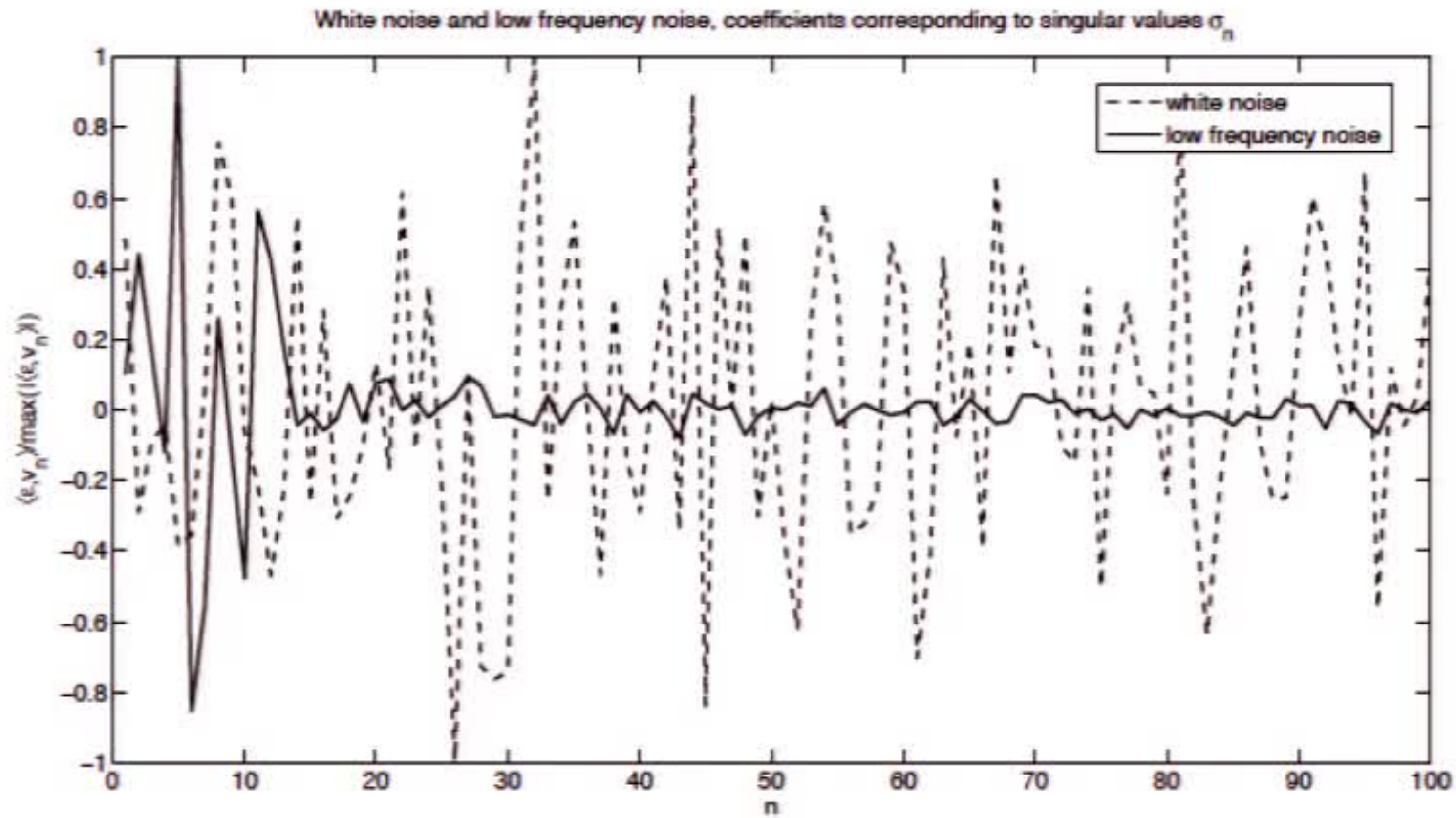


Figure: Low frequency noise vs. Gaussian noise

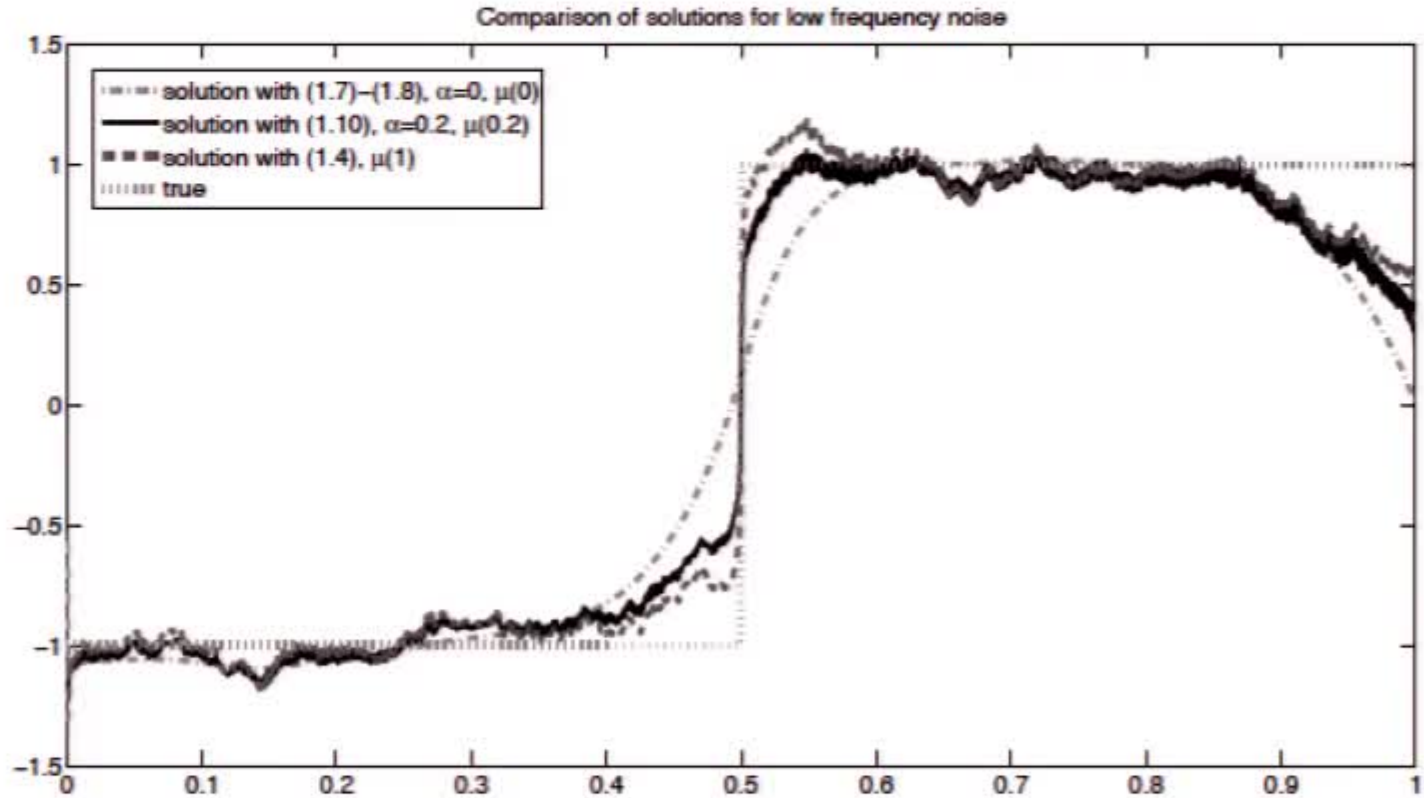


Figure: Comparison of solutions for the integration problem with low-frequency noise, "optimal" α, $\tau = 1.1$.

We now want to investigate the role of τ of the discrepancy principle.
Denote with $\bar{x}_\mu(\tau)^\delta$ the solution obtained with standard Tikhonov regularization. We are interested in the comparison of fractional Tikhonov regularization with the standard method. We define

$$\widetilde{\mathrm{re}}(\tau) = \frac{\left\|\tilde{x}^\delta_{\mu(\tau),\tilde{\alpha}^*} - x^\dagger\right\|}{\left\|\bar{x}^\delta_{\mu(\tau)} - x^\dagger\right\|} \quad \text{and} \quad \widehat{\mathrm{re}}(\tau) = \frac{\left\|\hat{x}^\delta_{\mu(\tau),\hat{\alpha}^*} - x^\dagger\right\|}{\left\|\bar{x}^\delta_{\mu(\tau)} - x^\dagger\right\|}.$$

where α is chosen optimal from a discrete predefined set. Here, $\tilde{x}^\delta_{\mu(\tau),\tilde{\alpha}^*}$ is the solution of (1.7)-(1.8) [Hoechstenbach Reichel] and $\hat{x}^\delta_{\mu(\tau),\tilde{\alpha}^*}$ the one with (1.10) [Klann Ramlau].

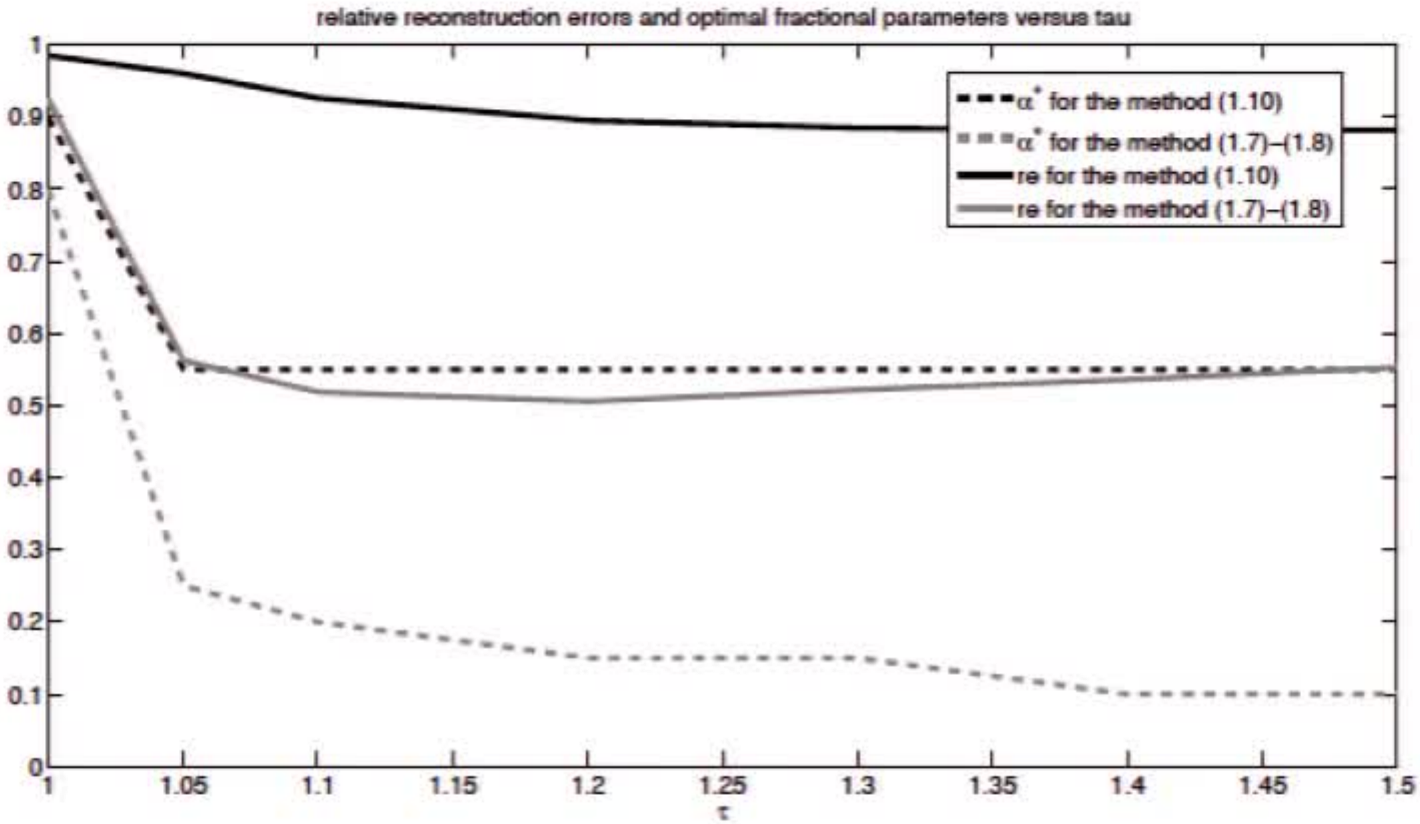


Figure: Relative errors and optimal fractional parameters α^* as functions of τ for the Fox–Goodwin problem with 5% Gaussian white noise.

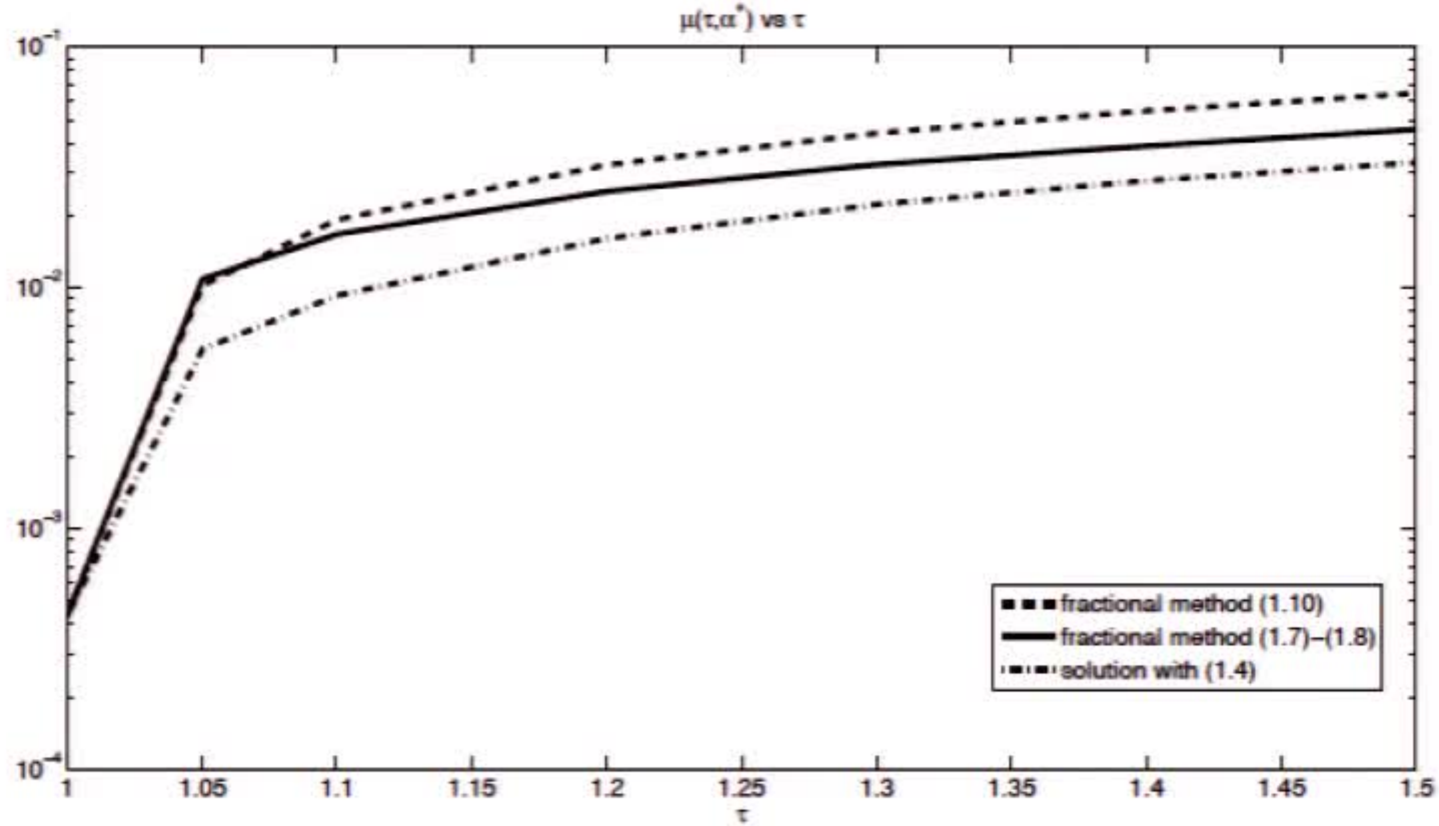


Figure: Regularization parameter under the previous setting.

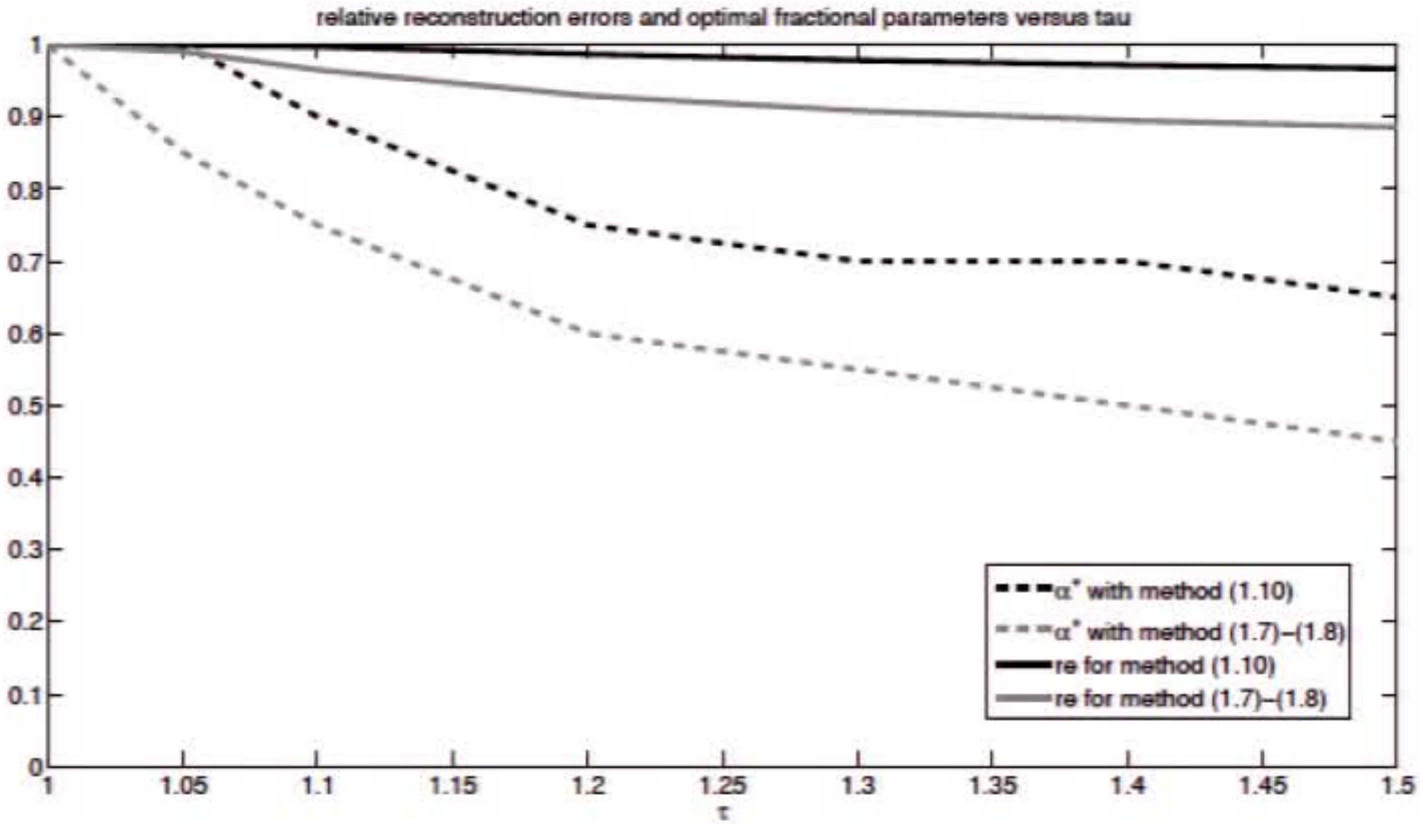


Figure: Relative errors and optimal fractional parameters α^* as functions of τ for the integration problem 5% Gaussian white noise.

A slightly changed experiment: we compare fractional Tikhonov methods and varying values of τ with standard Tikhonov regularization and $\tau = 1$.

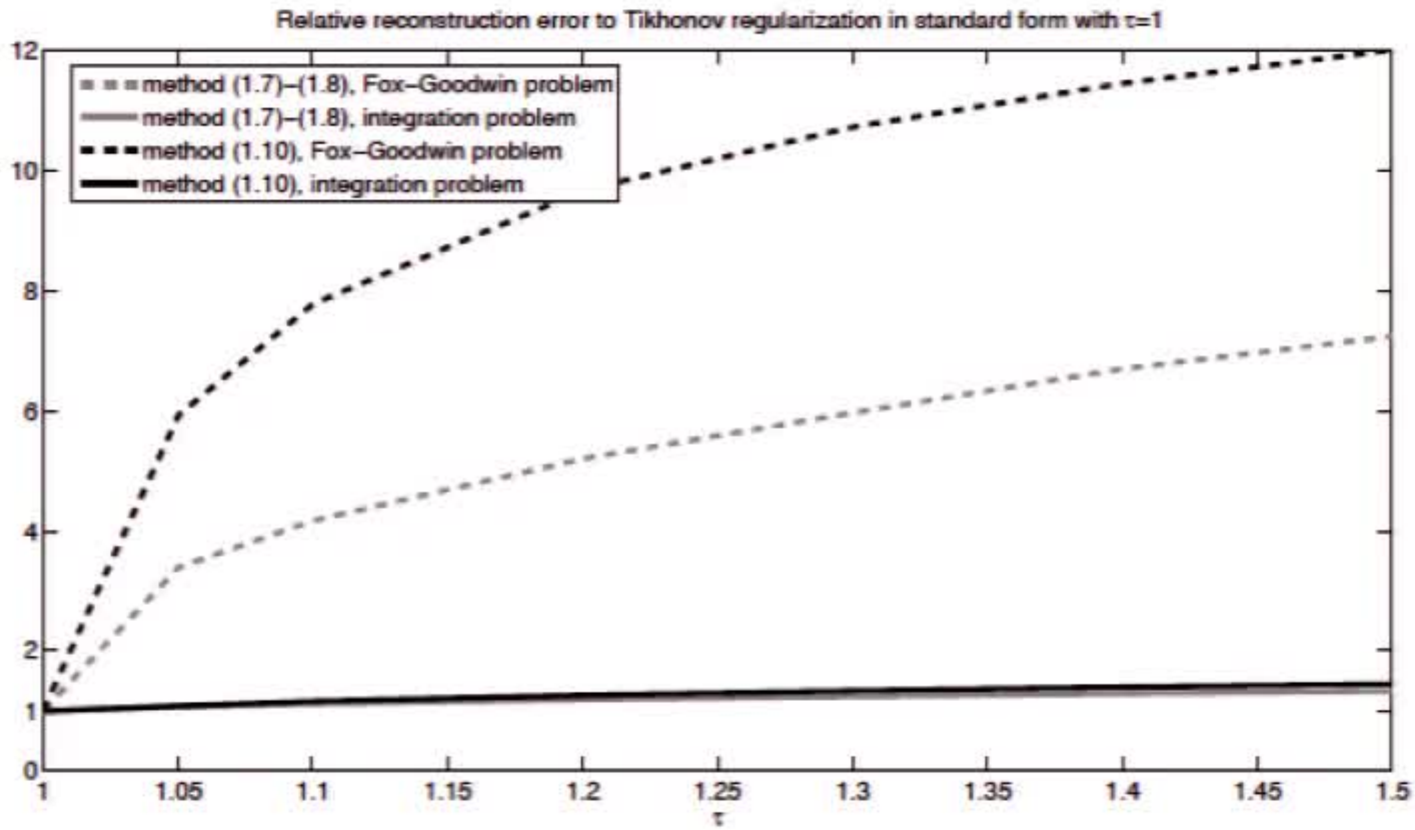


Figure: Solutions for a 2D deconvolution problem with low frequency noise. The relative reconstruction error to Tikhonov solution in standard form are $\widehat{\mathrm{re}} = 0.36$ and $\widetilde{\mathrm{re}} = 0.48$, respectively.