Parameter choice rules for convex variational regularization

Dirk Lorenz
(joint work with Bangti Jin)

IP-TA 2010, Warsaw

Setup: Inverse problem in a Banach space

- X Banach space, Y Hilbert space
- $K : X \to Y$ linear, bounded
- $x^\dagger \in X$, $Kx^\dagger = y^\dagger$
- $\|y^\dagger - y^\delta\| \leq \delta$

Given: K, y^δ, (δ)

Searched: Approximation to $x^\dagger$

Setup: Inverse problem in a Banach space

- X Banach space, Y Hilbert space
- $K : X \to Y$ linear, bounded
- $x^\dagger \in X$, $Kx^\dagger = y^\dagger$
- $\|y^\dagger - y^\delta\| \leq \delta$

Given: K, y^δ, (δ)

Searched: Approximation to $x^\dagger$

"Solve $Kx = y^\delta$"

Variational regularization

- K does not have a bounded inverse.
- Even solving $Kx = y^\dagger$ is unstable.

Regularizing functional: $R : X \to [0, \infty]$

- convex
- coercive
- lower semi-continuous

Tikhonov functional: $\mathcal{J}_\alpha(x) = \frac{1}{2}\|Kx - y^\delta\|^2 + \alpha R(x)$

$$x_\alpha^\delta \in \operatorname*{argmin}_x \mathcal{J}_\alpha(x)$$

Gives a non-linear procedure, i.e. x_α^δ depends non-linear on y^δ.

Variational regularization

- K does not have a bounded inverse.
- Even solving $Kx = y^\dagger$ is unstable.

Regularizing functional: $R : X \to [0, \infty]$

- convex
- coercive
- lower semi-continuous

Tikhonov functional: $\mathcal{J}_\alpha(x) = \frac{1}{2}\|Kx - y^\delta\|^2 + \alpha R(x)$

$$x_\alpha^\delta \in \underset{x}{\operatorname{argmin}}\, \mathcal{J}_\alpha(x)$$

Gives a non-linear procedure, i.e. x_α^δ depends non-linear on y^δ.

Examples:

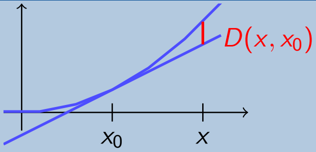
$$R(x) = |x|_{BV}, \qquad R(x) = \|x\|_p^p = \sum |x_i|^p, \qquad R(x) = \|x\|_1 + \tfrac{\beta}{2}\|x\|_2^2.$$

Various errors

- $x_\alpha^\delta \in \operatorname{argmin}_x \frac{1}{2}\|Kx - y^\delta\|^2 + \alpha R(x)$
- $x_\alpha \in \operatorname{argmin}_x \frac{1}{2}\|Kx - y^\dagger\|^2 + \alpha R(x)$

Appropriate distance measure: Bregman distance

$\xi \in \partial R(x_0)$ subgradient
$D_\xi(x, x_0) = R(x) - R(x_0) - \langle \xi, x - x_0 \rangle$
$D \geq 0$ for convex R

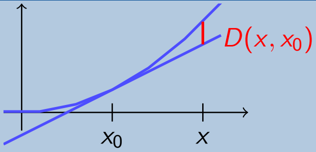


Various errors

- $x_\alpha^\delta \in \operatorname{argmin}_x \frac{1}{2}\|Kx - y^\delta\|^2 + \alpha R(x)$
- $x_\alpha \in \operatorname{argmin}_x \frac{1}{2}\|Kx - y^\dagger\|^2 + \alpha R(x)$

Appropriate distance measure: Bregman distance

$\xi \in \partial R(x_0)$ subgradient
$D_\xi(x, x_0) = R(x) - R(x_0) - \langle \xi, x - x_0\rangle$
$D \geq 0$ for convex R



Total error: $D(x_\alpha^\delta, x^\dagger)$
Approximation error: $D(x_\alpha, x^\dagger)$
Data error: $D(x_\alpha^\delta, x_\alpha)$

Total discrepancy: $\|Kx_\alpha^\delta - y^\dagger\|$
Approximation discrepancy: $\|Kx_\alpha - y^\dagger\|$
Data discrepancy: $\|K(x_\alpha^\delta - x_\alpha)\|$

Various errors

- The approximation error $D(x_\alpha, x^\dagger)$ says, how diffucult it is to approximate $x^\dagger$.
  Its behavior may point to helpful conditions on $x^\dagger$.
- The data error $D(x_\alpha^\delta, x_\alpha)$ says, how noise influences the accuracy of the reconstruction.
- Usual procedure: Split total error into approximation and data error and estimate both separately.
- Not possible here, since the Bregman distance does not fulfill the triangle inequality.

Various error estimates

Theorem (Jin, L. 2010)

With $\xi_\alpha = K^*(Kx_\alpha - y^\dagger)/alpha$ it holds that

$$D_{\xi_\alpha}(x_\alpha^\delta, x_\alpha) \leq \frac{\delta^2}{2\alpha}, \qquad \|K(x_\alpha^\delta - x_\alpha)\| \leq 2\delta.$$

If there exists $\xi \in \partial R(x^\dagger)$ and w such that $K^* w = \xi$ then

$$D_\xi(x_\alpha, x^\dagger) \leq \frac{\|w\|^2}{2}\alpha, \qquad \|Kx_\alpha - y^\dagger\| \leq 2\|w\|\alpha.$$

Theorem (Burger, Osher, 2004)

If there exists $\xi \in \partial R(x^\dagger)$ and w such that $K^* w = \xi$ then

$$D_\xi(x_\alpha^\delta, x^\dagger) \leq \tfrac{1}{2}\Big(\frac{\delta}{\sqrt{\alpha}} + \|w\|\sqrt{\alpha}\Big)^2, \qquad \|Kx_\alpha^\delta - y^\delta\| \leq \delta + 2\alpha\|w\|$$

A kind of triangle inequality

$$D_{\xi_\alpha}(x_\alpha^\delta, x_\alpha) \leq \frac{\delta^2}{2\alpha}, \qquad D_\xi(x_\alpha, x^\dagger) \leq \frac{\|w\|^2}{2}\alpha$$

$$D_\xi(x_\alpha^\delta, x^\dagger) \leq \tfrac{1}{2}\Big(\frac{\delta}{\sqrt{\alpha}} + \|w\|\sqrt{\alpha}\Big)^2$$

$\rightsquigarrow$ Sum of approximation and data error behaves like the total error.

Even more:

Lemma (Jin, L. 2010)

$$|D_\xi(x_\alpha^\delta, x^\dagger) - (D_{\xi_\alpha}(x_\alpha^\delta, x_\alpha) + D_\xi(x_\alpha^\delta, x^\dagger))| \leq 6\|w\|\delta.$$

The sum of approximation and data error differs from the total error only in magnitude δ.

The Hanke-Raus rule

$$\frac{\|Kx_\alpha^\delta - y^\delta\|^2}{\alpha} \leq \frac{(\delta + 2\alpha\|w\|)^2}{\alpha} = \Big(\frac{\delta}{\sqrt{\alpha}} + 2\sqrt{\alpha}\|w\|\Big)^2$$
$$\approx \frac{1}{2}\Big(\frac{\delta}{\sqrt{\alpha}} + \sqrt{\alpha}\|w\|\Big)^2 \geq D(x_\alpha^\delta, x^\dagger).$$

- Idea: Take
$$\phi(\alpha) = \frac{\|Kx_\alpha^\delta - y^\delta\|^2}{\alpha}$$
as an approximation to the total error $D(x_\alpha^\delta, x^\dagger)$.
- Hence: Choose α such that $\phi(\alpha)$ is minimal.
- Can be realized without any knowledge on δ.
- Resembles the Hanke-Raus rule [Hanke, Raus 1996]

An error estimate for Hanke-Raus

Theorem (Jin, L. 2010)

Let $\alpha^{HR} \in \operatorname{argmin}_{\alpha\in[0,\|K\|^2]} \phi(\alpha)$ *and define* $\delta^* = \|Kx^\delta_{\alpha^{HR}} - y^\delta\|$*. If there exists* $\xi \in \partial R(x^\dagger)$ *and* w *such that* $K^*w = \xi$

$$D(x^\delta_{\alpha^{HR}}, x^\dagger) \leq C\Big(1 + \big(\tfrac{\delta}{\delta^*}\big)^2\Big)\max(\delta,\delta^*).$$

- Problem: δ^* may be much smaller than δ (even 0).
- Condition on the noise gives convergence.

Theorem (Jin, L. 2010)

Additionally: $\epsilon > 0$ *such that for every* $z \in \overline{K(\operatorname{dom}\partial R)}$ *it holds that*

$$\langle y^\delta - y^\dagger, z\rangle \leq (1-\epsilon)\|y^\delta - y^\dagger\|\|z\|.$$

Then $$D(x^\delta_{\alpha^{HR}}, x^\dagger) \to 0, \quad \text{for} \quad \delta \to 0.$$

Yet another error estimate

Theorem

For $q \in]0,1[$ and $\xi_\alpha^\delta = -K^*(Kx_\alpha^\delta - y^\delta)/\alpha$ there holds

$$D_{\xi_\alpha^\delta}(x_{q\alpha}^\delta, x_\alpha^\delta) \leq \frac{(1-q)^2\|Kx_\alpha^\delta - y^\delta\|^2}{2\alpha q} = \frac{(1-q)^2}{2q}\phi(\alpha).$$

The discrete quasi-optimality principle

Let $q \in]0,1[$ and $\alpha_0 > 0$. Choose k such that

$$\mu_k = D(x_{q^k\alpha}^\delta, x_{q^{k-1}\alpha}^\delta)$$

is minimal.

Yet another error estimate

Theorem

For $q \in]0,1[$ and $\xi_\alpha^\delta = -K^*(Kx_\alpha^\delta - y^\delta)/\alpha$ there holds

$$D_{\xi_\alpha^\delta}(x_{q\alpha}^\delta, x_\alpha^\delta) \leq \frac{(1-q)^2\|Kx_\alpha^\delta - y^\delta\|^2}{2\alpha q} = \frac{(1-q)^2}{2q}\phi(\alpha).$$

The discrete quasi-optimality principle

Let $q \in]0,1[$ and $\alpha_0 > 0$. Choose k such that

$$\mu_k = D(x_{q^k\alpha}^\delta, x_{q^{k-1}\alpha}^\delta)$$

is minimal.

Continuous quasi-optimality principle: $\alpha > 0$ such that $\|\alpha \frac{dx_\alpha^\delta}{d\alpha}\|$ is minimal.

Here $\alpha \mapsto x_\alpha^\delta$ may be non-differentiable.

Convergence on autoregularizable sets

Theorem (Jin, L. 2010)

Let R be p-convex and let $\langle \xi, x\rangle$ be independent of $\xi \in \partial R(x)$. Then $\alpha^{qo}(y^\delta)$ chosen by the quasi-optimality principle is positive and fulfills

$$\alpha^{qo}(y^\delta) \to 0 \quad \textit{for } \delta \to 0.$$

Convergence on autoregularizable sets

Theorem (Jin, L. 2010)

Let R be p-convex and let $\langle \xi, x\rangle$ be independent of $\xi \in \partial R(x)$. Then $\alpha^{qo}(y^\delta)$ chosen by the quasi-optimality principle is positive and fulfills

$$\alpha^{qo}(y^\delta) \to 0 \quad \text{for } \delta \to 0.$$

Definition (similar to Glasko, Kriksin 1984)

For $r > 0$ and $q \in]0,1[$ we define the autoregularizable set by

$$\mathcal{D}_r = \{y^\delta \in Y \mid \left|D(x^\delta_{q^k}, x^\delta_{q^{k-1}}) - D(x_{q^k}, x_{q^{k-1}})\right| \geq rD(x^\delta_{q^k}, x_{q^k})\}.$$

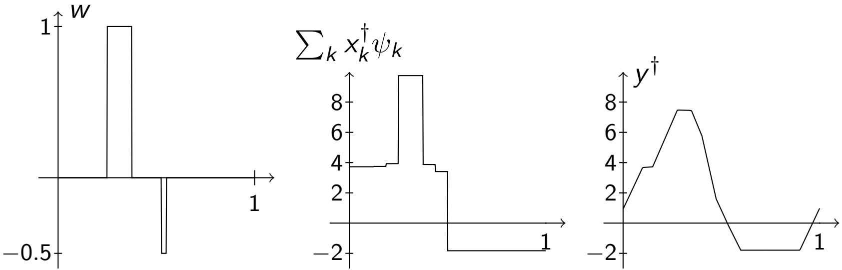
Theorem (Jin, L. 2010)

Let $\delta_n \to 0$, $y^{\delta_n} \to y^\dagger$, $y^{\delta_n} \in \mathcal{D}_r$. Then

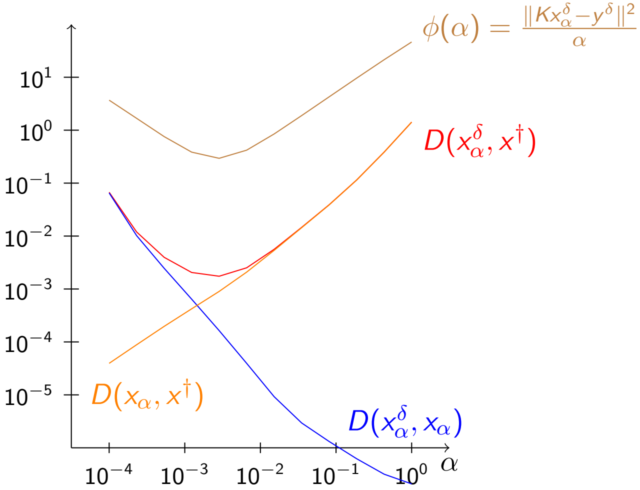
$$D(x^{\delta_n}_{\alpha^{qo}(y^{\delta_n})}, x^\dagger) \to 0, \quad \text{for} \quad n \to \infty.$$

Example setup

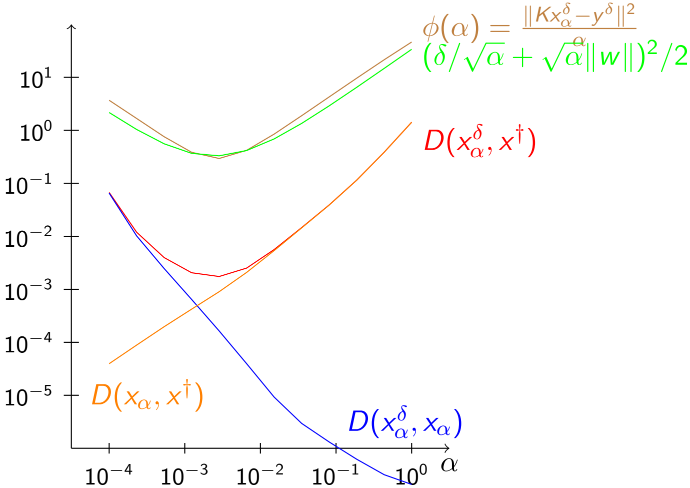
- $K : \ell^2 \to L^2[0,1]$
- $Kx = \sum_k x_k \psi_k * h$
- (ψ_k) Haar-Wavelet Basis
- $h = \chi_{[-0.1,0.1]}$
- $R(x) = \|x\|_p^p$, $p = 1.2$
- $K^* w = \partial R(x^\dagger)$



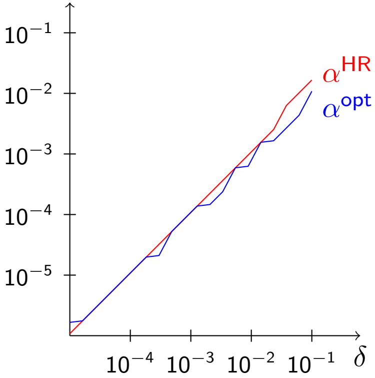
Example 1: Accuracy of the estimates

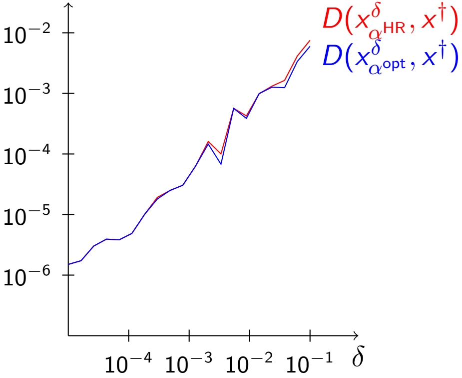


Example 1: Accuracy of the estimates

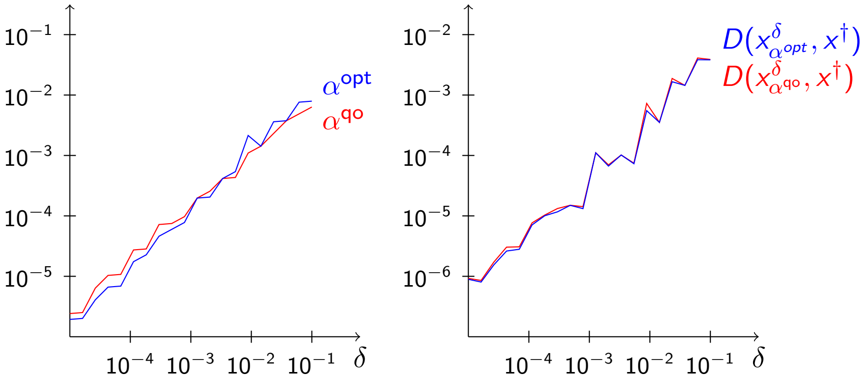


Example 2: The Hanke-Raus rule

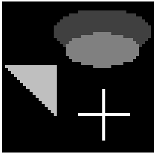




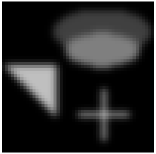
Example 3: The quasi-optimality principle



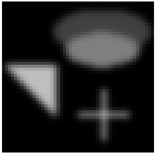
Example 4: Deblurring with the elastic net



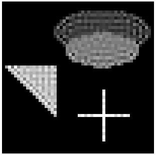
$x^\dagger$



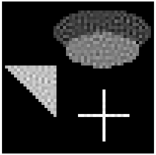
$y^\dagger$



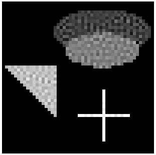
y^δ



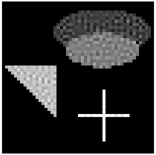
smallest Bregman distance



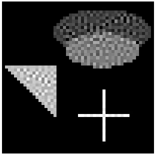
smallest norm



HR-rule



quasi-optimality



discrepancy principle

Example 4: Deblurring with the elastic net

	α	$D(x_\alpha^\delta, x^\dagger)$	$\|x_\alpha^\delta - x^\dagger\|_2$
smallest Bregman distance	1.10e-02	5.54e-02	1.02e+01
smallest norm	3.20e-03	7.39e-02	7.36e+00
Hanke-Raus	2.61e-03	9.03e-02	7.47e+00
quasi-optimality	3.02e-03	7.51e-02	7.38e+00
discrepancy	9.29e-04	7.16e-01	1.01e+01

Conclusion

- Approximation error and data error are useful for general convex regularization.
- Convergence can be guaranteed for the Hanke-Raus rule and for the quasi-optimality principle under additional assumptions.
- These assumption seem strong for Hanke-Raus and strange form quasi-optimality. (Some analysis like Kindermann, Neubauer possible?)
- Both rules perform remarkably well in practice.