# Large noise in variational regularization

Tapio Helin

Department of Mathematics and Statistics

University of Helsinki

Warwick, February 23, 2016

# Joint work with

**Martin Burger**
University of Münster

**Hanne Kekkonen**
University of Helsinki

preprint: Burger M, Helin T and Kekkonen H, *Large noise in variational regularization*, arXiv: 1602.00520.

# Problem setting

We consider a linear ill-posed problem

$$f = Ku$$

for a continuous linear operator $K : X \to Y$, where $X$ and $Y$ are separable Banach and Hilbert spaces, respectively. Suppose we are given noisy data by

$$f^{\delta} = Ku^{*} + \delta n,$$

where $u^{*}$ is the true solution and $\delta \cdot n$ is the noise vector with parameter $\delta > 0$ describing the noise level.

We would like to understand **convergence rates** in Tikhonov regularization for general **convex** regularization terms...

# ...with a little twist!

**Example.** Suppose $\mathrm{Range}(K) \subset L^2(\mathbb{T})$ and our data is given by

$$\{\langle f^\delta, e_j \rangle\}_{j=1}^J$$

in some basis $\{e_j\}_{j\in\mathbb{N}} \subset L^2(\mathbb{T})$. In inverse problem literature involving practical statistical inference the corresponding noise vector

$$n_J = \{\langle n, e_j \rangle\}_{j=1}^J \in \mathbb{R}^J$$

is assumed to have white noise statistics, i.e., $\langle n, e_j \rangle \sim \mathcal{N}(0,1)$ i.i.d. Hence

$$\mathbb{E}\, \|n_J\|_2^2 = \sum_{j=1}^J \mathbb{E}\langle n, e_j \rangle^2 = J \to \infty$$

as $J$ grows and consequently $n$ cannot be asymptotically modelled in $L^2(\mathbb{T})$!

# Large noise and background

Earlier work towards large noise in regularization:

- Egger 2008, Mathé and Tautenhahn, 2011,
- Eggermont, LaRiccia and Nashed 2009,
- Kekkonen, Lassas and Siltanen, 2014

Other connections:

- Frequentist cost: N. Bissantz, A. Munk, L. Cavalier, S. Agapiou and many others
- Schuster, Kaltenbacher, Hofmann and Kazimierski: *Regularization methods in Banach spaces*, de Gruyter, 2012.

# Our setup

Let $(Z, Y, Z^*)$ be a triplet such that $Z \subset Y$ is a dense subspace with Banach structure and assume

$$\langle u, v \rangle_{Z \times Z^*} = \langle u, v \rangle_Y$$

whenever $u \in Z$ and $v \in Y = Y^* \subset Z^*$.

**Two assumptions:**

(1) noise can be modelled in $Z^*$, i.e. $n \in Z^*$ and

(2) $K : X \to Z$ is continuous.

# Solution

We take the regularized solution $u_\alpha^\delta$ to be the minimizer of

$$J_\alpha^\delta(u) = \frac{1}{2}\|Ku\|_Y^2 - \langle Ku, f^\delta\rangle_{Z\times Z^*} + \alpha R(u)$$

with a convex regularization functional $R : X \to \mathbb{R} \cup \{\infty\}$.

Our main assumptions on $R$ are

(R1) $R$ is lower semicontinuous in some topology $\tau$ on $X$,

(R2) the sub-level sets $\{R \leq \rho\}$, $\rho > 0$, are compact in the topology $\tau$ on $X$ and

(R3) the convex conjugate $R^*$ is finite on a ball in $X^*$ centered at zero.

Moreover, we employ a symmetry condition $R(-u) = R(u)$ for all $u \in X$ for convenience.

# Optimality condition

The functional $J_\alpha^\delta$ is minimized by $u_\alpha^\delta$ that satisfies

$$K^*(Ku_\alpha^\delta - f^\delta) + \alpha \xi_\alpha^\delta = 0$$

for some $\xi_\alpha^\delta \in \partial R(u_\alpha^\delta)$, where the subdifferential $\partial R$ is defined by

$$\partial R(u) = \{\xi \in X^* \mid R(u) - R(v) \leq \langle \xi, u - v\rangle_{X^* \times X} \text{ for all } v \in X\}$$

Assumptions on $R$ guarantee

- existence of $u_\alpha^\delta$ and
- an a priori bound to $R(u_\alpha^\delta)$.

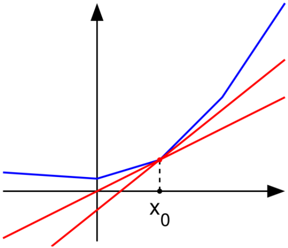


Figure: Subdifferential set at $x_0$.

# Bregman distance

**Definition**

For $\xi_u \in \partial R(u)$ we define symmetric Bregman distance between $u$ and $v$ as

$$D_R^{\xi_u,\xi_v}(u,v) = \langle \xi_u - \xi_v, u - v \rangle_{X^* \times X}.$$

Example. Suppose $R(u) = \frac{1}{2}\|u\|_X^2$ with $X$ Hilbert. Then $\partial R(u) = \{u\}$ and

$$D_R(u,v) = \|u - v\|_X^2 .$$

Example. Negative Shannon entropy $R(u) = \int_{\mathbb{R}} (u \ln u - u) dx$ on $L^1_+(\mathbb{R})$ yields "Kullback–Leibler -like" divergence

$$D_R(u,v) = \int_{\mathbb{R}} (u - v) \ln\left(\frac{u}{v}\right) dx.$$

# How to obtain traditional error estimates

By writing out the optimality condition for $f^\delta = Ku^* + n$ we obtain

$$K^*K(u_\alpha^\delta - u^*) + \alpha(\xi_\alpha^\delta - \xi^*) = K^*n - \alpha\xi^*,$$

where $\xi^* \in \partial R(u^*)$ was added on both sides. Taking a duality product with $u_\alpha - u^*$ we get

$$\|K(u_\alpha^\delta - u^*)\|_Y^2 + \alpha D_R^{\xi_\alpha^\delta,\xi^*}(u_\alpha^\delta, u^*) \leq \langle \delta K^*n - \alpha\xi^*, u_\alpha^\delta - u^*\rangle_{X^*\times X}.$$

The nice case leading directly to estimates if $n \in Y$ and the ideal source condition $\xi^* = K^*w^* \in X^*$ for $w^* \in Y$. Then

$$\langle \delta K^*n - \alpha\xi^*, u_\alpha^\delta - u^*\rangle_{X^*\times X} = \langle \delta n - \alpha w^*, K(u_\alpha^\delta - u^*)\rangle_Y,$$

and Young's inequality implies

$$\frac{1}{2}\|K(u_\alpha^\delta - u^*)\|_Y^2 + \alpha D_R^{\xi_\alpha^\delta,\xi^*}(u_\alpha^\delta, u^*) \leq \frac{1}{2}\|\delta n - \alpha w^*\|_Y^2.$$

# Convex conjugate

The convex conjugate $R^* : X^* \to \mathbb{R} \cup \{\infty\}$ defined via

$$R^*(q) = \sup_{u \in X} \left( \langle q, u \rangle_{X^* \times X} - R(u) \right).$$

Generalized Young's inequality: $\langle q, u \rangle_{X^* \times X} \le R(u) + R^*(q)$.

Important example: Let $R$ be one-homogeneous and let

$$S(q) = \sup_{R(u) \le 1} \langle q, u \rangle_{X^* \times X}$$

. Then we have

$$R^*(q) = \begin{cases} 0 & \text{if } S(q) \le 1 \\ +\infty & \text{else} \end{cases}$$

# Approximated source conditions to rescue

The key idea is to consider how well you are able to approximate $\xi^*$ and $K^*n$ with elements $K^*w_1$ and $K^*w_2$ for $w_j \in Y$.

$$
\begin{aligned}
\langle \delta K^* n - \alpha \xi^*, u_\alpha^\delta - u^* \rangle_{X^* \times X} \\
&= \delta \langle K^* n - K^* w_2, u_\alpha^\delta - u^* \rangle_{X^* \times X} \\
&\quad + \alpha \langle \xi^* - K^* w_1, u_\alpha^\delta - u^* \rangle_{X^* \times X} \\
&\quad + \langle \delta w_2 - \alpha w_1, K(u_\alpha^\delta - u^*) \rangle_Y,
\end{aligned}
$$

For the case $R(u) = \|u\|_X^r$ with $r > 1$ this approximation is quantified in literature by distance function

$$
d_\rho(\eta) := \inf_{w \in Y} \{ \|K^* w - \eta\|_{X^*} \mid \|w\|_Y \le \rho \}
$$

and its asymptotics as $\rho \to \infty$.

# First bound

**Theorem (BHK16)**

*We have a bound*

$$D_R^{\xi_\alpha^\delta,\xi^*}(u_\alpha^\delta, u^*) \leq (\zeta_1 + \frac{\delta}{\alpha}\zeta_2)R(u_\alpha^\delta - u^*) + e_{\alpha,\zeta_1}(\xi^*) + \frac{\delta}{\alpha} e_{\delta,\zeta_2}(K^* n).$$

*where* $\zeta_1, \zeta_2 > 0$ *are arbitrary and*

$$\begin{aligned} e_{\beta,\zeta}(\eta) &= \inf_{w\in Y}\left(\zeta R^*\left(\frac{K^* w - \eta}{\zeta}\right) + \frac{\beta}{2}\|w\|_Y^2\right) \\ &= -\inf_{v\in X}\left(\frac{1}{2\beta}\|Kv\|_Y^2 - \langle \eta, v\rangle_{X^*\times X} + \zeta R(v)\right) \end{aligned}$$

# Some added structure to move forward

Assume there exists $\theta \in [0,1]$ such that

$$R(u-v) \leq C_\theta(u,v)\left(D_R^{\xi_u,\xi_v}(u,v)\right)^\theta$$

for all $u, v \in X$, $\xi_u \in \partial R(u)$ and $\xi_v \in \partial R(v)$. Above the constant $C_\theta$ is bounded on sets where $R(u)$ and $R(v)$ are bounded.

**Example**

Let $R(u) = \frac{1}{2}\|u\|_X^2$. Then $D_R^{\xi_u,\xi_v}(u,v) = \|u-v\|_X^2 = 2R(u-v)$ and above $\theta = 1$ and $C_\theta(u,v) \equiv \frac{1}{2}$.

**Example**

Let $R$ be one-homogeneous, symmetric around zero, and convex. By triangle inequality

$$R(u-v) \leq R(u) + R(v),$$

and hence $\theta = 0$ and $C_0(u,v) = R(u) + R(v)$.

# Application: One-homogeneous problem

A priori bound:

$$R(u_\alpha^\delta) \lesssim R(u^*) + \frac{\delta}{\alpha} e_{\delta,\frac{\alpha}{\delta}}(K^*n) \lesssim R(u^*)$$

Together with triangle inequality it follows for the one-homogeneous case that

$$\begin{aligned} & D_R^{\xi_\alpha^\delta,\xi^*}(u_\alpha^\delta, u^*) \\ & \lesssim \inf_{\zeta_1>0} \left(\zeta_1 R(u^*) + e_{\alpha,\zeta_1}(\xi^*)\right) + \frac{\delta}{\alpha} \inf_{\zeta_2>0} \left(\zeta_2 R(u^*) + e_{\delta,\zeta_2}(K^*n)\right). \end{aligned}$$

# Application: One-homogeneous problem

Suppose that $R$ is one-homogeneous and recall

$$R^*(cq) = \begin{cases} 0 & \text{if } S(q) = \sup_{R(u)\leq 1} \langle q, u\rangle_{X^*\times X} \leq \frac{1}{c} \\ +\infty & \text{else} \end{cases}$$

We have

$$\begin{aligned} e_{\beta,\zeta}(\eta) &= \inf_{w\in Y} \left( \zeta R^* \left( \frac{K^* w - \eta}{\zeta} \right) + \frac{\beta}{2} \|w\|_Y^2 \right) \\ &= \frac{\beta}{2} \underbrace{\inf_{w\in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(\eta - K^* w) \leq \zeta \right\}}_{\text{assumption on decay } \lesssim \zeta^{-r}} \end{aligned}$$

# One-homogeneous problem

Quantification of the approximative source condition:

$$e_{\alpha,\zeta_1}(\xi^*) = \frac{\alpha}{2} \underbrace{\inf_{w \in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(\xi^* - K^* w) \leq \zeta_1 \right\}}_{\lesssim \zeta_1^{-r_1}} \lesssim \alpha \zeta_1^{-r_1}$$

and

$$e_{\delta,\zeta_2}(K^* n) = \frac{\delta}{2} \underbrace{\inf_{w \in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(K^* n - K^* w) \leq \zeta_2 \right\}}_{\lesssim \zeta_2^{-r_2}} \lesssim \delta \zeta_2^{-r_2}$$

# One-homogeneous problem

**Theorem**

*Let $X$ be a Banach space and $R$ one-homogeneous. Suppose that decay on $\xi^*$ and $n$ is described by $r_1$ and $r_2$, respectively. Optimal convergence rate is obtained by choice $\alpha \simeq \delta^\kappa$ where*

$$\kappa = \begin{cases} \frac{(1+r_1)(2+r_2)}{(2+r_1)(1+r_2)} & \text{for } r_1 \leq r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1, \end{cases}$$

*we have that*

$$D_R^{\xi_\alpha^\delta, \xi^*}(u_\alpha^\delta, u^*) \lesssim \begin{cases} \delta^{\frac{2+r_2}{(2+r_1)(1+r_2)}} & \text{for } r_1 \leq r_2 \text{ and} \\ \delta^{\frac{1}{1+r_1}} & \text{for } r_2 < r_1. \end{cases}$$

# Frequentist cost

The pointwise theory can be applied to obtain estimates on Bregman-distance based frequentist cost of

$$f = Ku + N,$$

where $N$ is random.

- Take $X = Y = L^2(\mathbb{T})$ and $Z = H^s(\mathbb{T})$, $s > \frac{1}{2}$ and
- assume $N$ is Gaussian white noise $\Rightarrow N \in Z^*$ a.s.

We want to find converge rates for $\mathbb{E} D_R^{\xi_\alpha^\delta, \xi^*}(U_\alpha^\delta, u^*)$.

# Frequentist cost

For one-homogeneous $R$ the Bregman-based frequentist cost can be estimated by

$$
\begin{aligned}
&\mathbb{E} D_R^{\xi_\alpha^\delta,\xi^*}(u_\alpha^\delta, u^*) \\
&\lesssim \inf_{\zeta_1>0} (\zeta_1 R(u^*) + e_{\alpha,\zeta_1}(\xi^*)) + \mathbb{E} \inf_{\zeta_2>0} \left( \frac{\delta}{\alpha} \zeta_2 R(u^*) + \frac{\delta}{\alpha} e_{\delta,\zeta_2}(K^* N) \right) \\
&\lesssim \inf_{\zeta_1>0} (\zeta_1 R(u^*) + e_{\alpha,\zeta_1}(\xi^*)) + \inf_{\zeta_2>0} \left( \frac{\delta}{\alpha} \zeta_2 R(u^*) + \frac{\delta}{\alpha} \mathbb{E} e_{\delta,\zeta_2}(K^* N) \right)
\end{aligned}
$$

Probabilistic source condition $\approx$ decay rate of $\mathbb{E} e_{\delta,\zeta_2}(K^* N)$.

# Quadratic regularization

For $R(u) = \frac{1}{2}\|u\|^2_{L^2(\mathbb{T})}$, we have

$$\mathbb{E} e_{\alpha,\zeta}(K^*N) = \frac{\alpha}{2}\mathrm{Tr}_{L^2(\mathbb{T})}(K(K^*K + \alpha\zeta I)^{-1}K^*)$$

### Theorem

*Suppose $R(u) = \frac{1}{2}\|u\|^2_{L^2(\mathbb{T})}$ and one has an exact source condition on $\xi^*$. Moreover, we assume that $\{\lambda_j\}_{j=1}^\infty$ are eigenvalues of $KK^* : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ and there exists $0 < \epsilon \leq 1$ such that $\sum_{j=1}^\infty \lambda_j^\epsilon < \infty$. It follows that for $\alpha \simeq \delta^\kappa$ for $\kappa = \frac{2}{2+\epsilon}$ we obtain*

$$\mathbb{E} D_R^{\xi_\alpha^\delta,\xi^*}(U_\alpha^\delta, u^*) = \mathbb{E}\|U_\alpha^\delta - u^*\|^2_{L^2(\mathbb{T})} \lesssim \delta^{\frac{2}{2+\epsilon}}.$$

# Besov norm regularization

## Theorem

*Let us assume that* $K : B_1^s(\mathbb{T}) \to L^2(\mathbb{T})$, *R is defined by*

$$R(u) = \|u\|_{B_1^s(\mathbb{T})} = \sum_{\ell=1}^{\infty} \ell^{s-1/2}|u_\ell|,$$

*where* $u = \sum_\ell u_\ell \psi_\ell$ *in some smooth wavelet basis* $\{\psi_\ell\}$ *and* $\xi^*$ *satisfies the approximate source condition of order* $r_1 \geq 0$. *Then for the choice*

$$\alpha \simeq \delta^\kappa \quad \text{for} \quad \kappa = (1+t) \cdot \frac{1+r_1}{2+r_1},$$

*where* $t > 0$ *describes the smoothness of* $K$. *Then*

$$\mathbb{E} D_R^{\xi_\alpha^\delta, \xi^*}(U_\alpha^\delta, u^*) \lesssim \delta^{\frac{1+t}{2+r_1}}.$$

# Total variation regularization

**Theorem**

*Let us assume that $K$ is of order $s+t$, $t>0$, smoothing pseudodifferential operator, $R(u)=\int_{\mathbb{T}}|\nabla u|\,dx$, and $\xi^*$ satisfies the approximate source condition of order $r_1 \geq 0$. Then for the choice*

$$\alpha \simeq \delta^{\kappa} \quad \text{for} \quad \kappa = \frac{1+r_1}{(2+r_1)(1-\mu)}$$

*we obtain the convergence rate*

$$\mathbb{E} D_R^{\xi_\alpha^\delta,\xi^*}(U_\alpha^\delta, u^*) \lesssim \delta^{\frac{1}{(2+r_1)(1-\mu)}} \leq \delta^{\frac{1}{2+r_1}}$$

*where $\mu = \frac{t}{2(s+t)}$.*

# Conclusions

- Convergence rates (or consistency estimates) are possible for large noise and general convex regularization terms
- Infinite-dimensional frequentist cost for penalties like Besov and TV
- Bayesian cost - see earlier work by Kekkonen

preprint: Burger M, Helin T and Kekkonen H, *Large noise in variational regularization*, arXiv: 1602.00520.