Bregman iterations and higher order variational inequalities

Benjamin Sprung,
joint work with Thorsten Hohage

Institut für Numerische und Angewandte Mathematik
Georg-August Universität Göttingen

Malta, May 21, 2018

SFB
755

inv
opt
RTG
2088
stat

Outline

Outline

1. Overview: convergence rates and source conditions
    - Convergence rates and source conditions
    - Grasmair's second order VSC
    - Bregman iteration
2. Higher order variational source conditions
    - Hilbert space setting
    - Banach space setting
3. Numerical example

Convergence rates

Let $F\colon \mathbb{X} \to \mathbb{Y}$ be our forward operator and $R_{\alpha(\delta,g^\delta)}$ some regularization method.

- worst case error:

$$\mathcal{E}(\delta, f^\dagger) := \sup\left\{\|R_{\alpha(\delta,g^\delta)}(g^\delta) - f^\dagger\| \,:\, \|g^\delta - F(f^\dagger)\| \leq \delta\right\}$$

- explicit error bounds/convergence rates:

$$\mathcal{E}(\delta, f^\dagger) \leq \psi(\delta) \qquad \text{for all } f^\dagger \in K,$$

where K is some smoothness class and $\psi : [0, \infty) \to [0, \infty)$ is an *index function*, i.e. non-decreasing and vanishing at 0.

Source conditions

spectral source conditions (SSC): (used since the 1980s)

$$f^\dagger = \left(F'[f^\dagger]^* F'[f^\dagger]\right)^{\nu/2} \omega \quad \text{for some } \nu > 0 \qquad (\mathrm{SSC}(f^\dagger, \nu))$$

corresponding convergence rate: $\mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{\nu/(\nu+1)}\right)$

variational source conditions (VSC):

$$\forall f: \qquad 2\left\langle f^\dagger, f^\dagger - f \right\rangle \leq \frac{1}{2}\|f - f^\dagger\|^2 + \psi\left(\left\|F(f) - F(f^\dagger)\right\|_{\mathbb{Y}}^2\right).$$

Here $\psi : [0, \infty) \to [0, \infty)$ is an *index function*.
Corresponding rate: $\mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\sqrt{\psi(\delta^2)}\right)$

First used (with $\psi(t) = c\sqrt{t}$) in

B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*. **Inverse Problems** 23:987–1010, 2007.

SSC versus VSC

advantages of variational source conditions

1. VSCs work in Banach spaces and for general penalty terms $\mathcal{R}$ and data fidelity terms $\mathcal{S}$.
2. VSCs lead to simpler proofs and better results for nonlinear F. One of the reasons is that no Fréchet derivative $F'[f^\dagger]$ at the unknown solution is involved.
3. VSCs are not only sufficient, but even necessary for certain rates of convergence.

disadvantages of variational source conditions

1. VSCs are difficult to interpret.
2. VSCs only work for low smoothness.

Disadvantage 1: difficult to interpret

For many years, VSCs could be interpreted only for some particular examples. Recent progress:

- If $\exists a > 0 \forall s \in \mathbb{R} \quad F : W_2^s(\mathcal{M}) \to W_2^{s+a}(\mathcal{M})$ is a linear norm isomorphism and $\mathbb{X} = \mathbb{Y} = L^2(\mathcal{M})$, then for $u \in (0, a)$

  $$\begin{array}{lll} f^\dagger \in B^u_{2,\infty}(\mathcal{M}) & \Leftrightarrow \ \text{VSC}\left(f^\dagger, ct^{\frac{u}{u+a}}\right) & \Leftrightarrow \ \mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{\frac{u}{u+a}}\right) \\ f^\dagger \in W^u_2(\mathcal{M}) & \Leftrightarrow \ \text{SSC}\left(f^\dagger, \frac{u}{a}\right) & \Rightarrow \ \mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{\frac{u}{u+a}}\right) \end{array}$$

  Typical functions in $B^{1/2}_{2,\infty}(\mathbb{R}) \setminus W^{1/2}_2(\mathbb{R})$: piecewise smooth with jumps
- Also for nonlinear inverse medium scattering problems VSC could be interpreted in terms of Sobolev smoothness.

📖 T. Hohage, F. Weidling. *Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods.* **SIAM J. Numer. Anal.** 55:598–620, 2017.

📖 T. Hohage, F. Weidling. *Verification of a variational source condition for acoustic inverse medium scattering problems.* **Inverse Problems** 31:075006 , 2015.

Disadvantage 2: VSCs only work for low order smoothness.

Lemma

If F is Fréchet differentiable and $f^\dagger$ satisfies a VSC with ψ such that $\lim_{t\searrow 0}\psi(t)/\sqrt{t}=0$, then $f^\dagger\in\operatorname{argmin}_f\mathcal{R}(f)$.

In other words: $\psi(t)=c\sqrt{t}$ corresponding to $\mathcal{E}(\delta,f^\dagger)=\mathcal{O}\left(\sqrt{\delta}\right)$ or $\nu=1$ in SSCs is the fastest rate achievable with VSCs.

topic of this talk:

new types of VSCs corresponding to SSC$(f^\dagger,\nu)$ with $\nu\in(1,\infty)$

Outline

The dual problem

setting:

- $\mathbb{X}$, $\mathbb{Y}$ Banach spaces
- $T : \mathbb{X} \to \mathbb{Y}$ bounded, linear
- $\mathcal{R} : \mathbb{X} \to \overline{\mathbb{R}}$, $\mathcal{S}_{g^\delta} : \mathbb{Y} \to \overline{\mathbb{R}}$ proper, convex, l.s.c.

Tikhonov minimization problem:

$$\inf_{f \in \mathbb{X}} \left[\frac{1}{\alpha} \mathcal{S}_{g^\delta}(Tf) + \mathcal{R}(f) \right]$$

dual problem:

$$\sup_{p \in \mathbb{Y}^*} \left[-\frac{1}{\alpha} \mathcal{S}^*_{g^\delta}(-\alpha p) - \mathcal{R}^*(T^* p) \right]$$

Grasmair's second order VSC

idea: Impose a VSC on $\overline{p}$, with $T^*\overline{p} \in \partial\mathcal{R}(f^\dagger)$ based on dual problem:

- Let $\mathbb{Y}$ be a q-smooth Banach space with $q > 1$, $\frac{1}{q} + \frac{1}{q_*} = 1$, $\mathcal{S}_{\overline{g}}(g) = \frac{1}{q}\|g - \overline{g}\|_{\mathbb{Y}}^q$ and $\mathcal{S}(g) = \frac{1}{q}\|g\|_{\mathbb{Y}}^q$.
- Let $p^* \in \partial\mathcal{S}(\overline{p})$.

second order VSC with concave index function ψ ($\mathrm{VSC}_{\mathrm{G}}(f^\dagger, \psi)$):

$$\forall p \in \mathbb{Y}^* : \ \langle p^*, \overline{p} - p\rangle \leq \psi\left(\Delta_{\mathcal{R}^*}^{f^\dagger}(T^*p, T^*\overline{p})\right)$$

Theorem

Let ϕ denote the conjugate of the (convex) inverse function ψ^{-1}. Then $\mathrm{VSC}^2(f^\dagger, \psi)$ *implies*

$$\Delta_{\mathcal{R}}^{f^*}(f_\alpha^\delta, f^\dagger) = \mathcal{O}\left(\phi\left(\alpha^{q_*-1}\right) + \frac{\delta^q}{\alpha}\right).$$

M. Grasmair. *Variational inequalities and higher order convergence rates for Tikhonov regularization on Banach spaces* **J. Inverse Ill-Posed Probl.**

Outline

1. Overview: convergence rates and source conditions
   - Convergence rates and source conditions
   - Grasmair's second order VSC
   - Bregman iteration

2. Higher order variational source conditions
   - Hilbert space setting
   - Banach space setting

3. Numerical example

Saturation of Tikhonov regularization

Theorem

For Tikhonov regularization in Hilbert spaces

$$\mathcal{E}(\delta, f^\dagger) = o\left(\delta^{2/3}\right) \quad \Rightarrow \quad f^\dagger = 0.$$

📖 C.W. Groetsch. *The Theory of Tikhonov regularization for Fredholm equations of the first kind.* Pitman, 1984.

- The best possible rate $\mathcal{O}\left(\delta^{2/3}\right)$ corresponds to a SSC($f^\dagger, 2$).
- Under SSC($f^\dagger, \nu$) with $\nu > 2$ we also only get $\mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{2/3}\right)$.
- 2 is called the qualification of Tikhonov regularization.

Iterated Tikhonov regularization

Higher order rates can be obtained by iterating standard Tikhonov regularization

$$\hat{f}_\alpha = \hat{f}_\alpha^{(1)} \in \mathrm{argmin}_{f \in \mathbb{X}} \left[\| Tf - g^\delta \|^2 + \alpha \| f - f_0 \|^2 \right]$$

by replacing the initial guess f_0 with the last iterate:

$$\hat{f}_\alpha^{(n)} \in \mathrm{argmin}_{f \in \mathbb{X}} \left[\| Tf - g^\delta \|^2 + \alpha \| f - \hat{f}_\alpha^{(n-1)} \|^2 \right]$$

Proposition

n-times iterated Tikhonov regularization has ”qualification” $2n$:

$$\mathrm{SSC}(f^\dagger, \nu) \quad \Rightarrow \quad \mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{\frac{\nu}{\nu+1}}\right) \qquad \textit{for } \nu \leq 2n.$$

Bregman iterations

For $f_0 \in \mathbb{X}$ and $f^* \in \partial\mathcal{R}(f_0)$ define the Bregman distance by

$$\Delta_{\mathcal{R}}^{f^*}(f, f_0) = \mathcal{R}(f) - \mathcal{R}(f_0) - \langle f^*, f - f_0 \rangle.$$

In Banach spaces we can replace the squared norm by the Bregman distance:

$$\hat{f}_{\alpha}^{(n+1)} \in \mathrm{argmin}_{f \in X} \left[\frac{1}{\alpha} \mathcal{S}_{g^\delta}\left(Tf\right) + \Delta_{\mathcal{R}}^{f_n^*}(f, \hat{f}_{\alpha}^{(n)}) \right]$$

with $f_n^* \in \partial\mathcal{R}(\hat{f}_{\alpha}^{(n)})$.

- Y. Censor, S. Zenios. *Proximal minimization algorithm with D-functions.* **J. Optim. Theory Appl.** 73:451–464, 1992.
- J. Eckstein. *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming.* **Math. Oper. Res.** 18:202–226, 1993.
- S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin. *An Iterative Regularization Method for Total Variation-Based Image Restoration.*, **Multiscale Modeling & Simulation** 4:460–489, 2005.

Outline

Outline

1. Overview: convergence rates and source conditions
    - Convergence rates and source conditions
    - Grasmair's second order VSC
    - Bregman iteration

2. Higher order variational source conditions
    - Hilbert space setting
    - Banach space setting

3. Numerical example

VSCs of order $n \in \mathbb{N}$ in Hilbert spaces

Assumption (VSC$^{2n-1}(f^\dagger, \psi)$)

$f^\dagger = (T^*T)^{n-1}\overline{\omega}^{(n-1)}$ for some $\overline{\omega}^{(n-1)} \in \mathbb{X}$ and

$$\forall f \in \mathbb{X}: \; 2\langle \overline{\omega}^{(n-1)}, f\rangle_{\mathbb{X}} \le \frac{1}{2}\|f\|_{\mathbb{X}}^2 + \psi\left(\|Tf\|_{\mathbb{Y}}^2\right)$$

Assumption (VSC$^{2n}(f^\dagger, \psi)$)

$f^\dagger = (T^*T)^{n-1}T^*\overline{p}^{(n)}$ for some $\overline{p}^{(n)} \in \mathbb{Y}$ and

$$\forall p \in \mathbb{Y}: \; 2\langle p, \overline{p}^{(n)}\rangle_{\mathbb{Y}} \le \frac{1}{2}\|p\|_{\mathbb{Y}}^2 + \psi\left(\|T^*p\|_{\mathbb{X}}^2\right)$$

Proposition

- $\mathrm{VSC}(f^\dagger, \psi) = \mathrm{VSC}^1(f^\dagger, \psi)$
- $\mathrm{SSC}(f^\dagger, \nu + n - 1) \;\Rightarrow\; \mathrm{VSC}^n\left(f^\dagger, ct^{\frac{\nu}{\nu+1}}\right), \quad n \in \mathbb{N}, \nu \in (0, 1]$

Higher order convergence rates

Theorem (Hohage, S 2017)

For $m \geq n/2$ and an index function ψ, $\mathrm{VSC}^n(\psi)$ implies

$$\|\hat{f}_\alpha^{(m)} - f^\dagger\|_{\mathbb{X}}^2 = \mathcal{O}\left(\frac{\delta^2}{\alpha} + \alpha^{n-1}(-\psi)^*\left(-\frac{1}{\alpha}\right)\right).$$

Theorem (Hohage, S)

For $m \geq \frac{n}{2}$, $\nu \in (0,1)$, $f^\dagger \neq 0$ and the estimator $f_{\alpha_*}^{(m)}$ with optimally chosen α_* we have

$$\exists A > 0 \,:\, \mathrm{VSC}^n\left(f^\dagger, At^{\frac{\nu}{\nu+1}}\right) \quad \Leftrightarrow \quad \mathcal{E}(\delta, f^\dagger) = \mathcal{O}\left(\delta^{\frac{\nu+n-1}{\nu+n}}\right)$$

Characterization of VSC^n

Theorem

Let $n \in \mathbb{N}$, $\mathbb{X} = \mathbb{Y} = L^2(\mathcal{M})$, *and assume there exists* $a > 0$ *such that for all* $s \in \mathbb{R}$ *the operator*

$$T : W_2^s(\mathcal{M}) \to W_2^{s+a}(\mathcal{M})$$

is bounded and boundedly invertible. Then for all $n \in \mathbb{N}$ *and* $\nu \in (0,1)$

$$\exists A > 0 \, : \, \mathrm{VSC}^n\left(f^\dagger, At^{\frac{\nu}{\nu+1}}\right) \quad \Leftrightarrow \quad f^\dagger \in B_{2,\infty}^{a(n-1+\nu)}(\mathcal{M}).$$

Based on

T. Hohage, F. Weidling. *Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods.* **SIAM J. Numer. Anal.** 55:598–620, 2017.

Outline

1. Overview: convergence rates and source conditions
   - Convergence rates and source conditions
   - Grasmair's second order VSC
   - Bregman iteration
2. Higher order variational source conditions
   - Hilbert space setting
   - Banach space setting
3. Numerical example

Variational source condition of third order

From now on let $\mathcal{S}_{g^\delta}(g) = \|g - g^\delta\|_{\mathbb{Y}}^q$, for some $1 < q < \infty$.

Assumption (VSC^3)

Let $T^*\overline{p} \in \partial\mathcal{R}(f^\dagger)$, $T\overline{\omega} \in \partial\mathcal{S}^*(\overline{p})$. *Assume there exist constants* $\beta \geq 0, \overline{t} > 0$, $\mu > 1$, *an index function* ψ *and for all* $0 < t \leq \overline{t}$ $f_t^* \in \partial\mathcal{R}(f^\dagger - t\overline{\omega})$ *such that*

$$\forall f \in \mathbb{X}\ \forall t \in (0, \overline{t}]:$$

$$\begin{aligned}\left\langle f_t^* - T^*\overline{p}, f^\dagger - t\overline{\omega} - f \right\rangle \leq{}& \Delta_{\mathcal{R}}\left(f, f^\dagger - t\overline{\omega}\right) \\ &+ t^2\psi\left(t^{-q}\left\| Tf - g^\dagger + tT\overline{\omega}\right\|^q\right) + \beta t^{2\mu}.\end{aligned}$$

Convergence rates under VSC^3

Theorem (Hohage, S)

Let $\mathbb{Y}$ be q-smooth and r-convex and choose α such that $\delta = o(\alpha^{q^*-1})$. Then VSC^3 *implies*

$$\Delta_{\mathcal{R}}\left(\hat{f}_\alpha^{(2)}, f^\dagger\right) = \mathcal{O}\left(\frac{\delta^q}{\alpha} + \alpha^{2(q^*-1)}(-\tilde{\psi})^*\left(\frac{-C}{\alpha^{q^*-1}}\right) + \beta\alpha^{2\mu(q^*-1)}\right),$$

where $\tilde{\psi}(x) = \psi\left(x^{\frac{q}{r}}\right)$.

Example

- $\mathbb{Y} = L^q([0,1])$ for $1 < q \leq 2$ such that $\mathbb{Y}$ is q-smooth and 2-convex
- $\mathbb{X} = L^{\mu^*}([0,1])$ with $\mu^* = 2$ or $\mu^* \geq 3$ and $\mathcal{R} = \frac{1}{\mu^*}\|\cdot\|_{\mathbb{X}}^{\mu^*}$
- assume $\mathcal{R}'(f^\dagger) = T^*\overline{p}, (\mathcal{S}^*)'(\overline{p}) = T\overline{\omega}, \mathcal{R}''(f^\dagger, \overline{\omega}) = T^*\overline{p}^{(2)}$

Then VSC^3 holds and we have by the above theorem that

$$\Delta_{\mathcal{R}}\left(\hat{f}_\alpha^{(2)}, f^\dagger\right) = \mathcal{O}\left(\frac{\delta^q}{\alpha} + \alpha^{3(q^*-1)} + \alpha^{2\mu(q^*-1)}\right),$$

where $\frac{1}{\mu^*} + \frac{1}{\mu} = 1$. The best known error bound for Tikhonov regularization (requiring existence of $\overline{p}, \overline{\omega}$ as above) is

$$\Delta_{\mathcal{R}}\left(\hat{f}_\alpha, f^\dagger\right) = \mathcal{O}\left(\frac{\delta^q}{\alpha} + \alpha^{2(q^*-1)}\right).$$

A. Neubauer, T. Hein, B. Hofmann, S. Kindermann and U. Tautenhahn. *Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces.* **Applicable Analysis** 89(11), 2010.

Outline

1. Overview: convergence rates and source conditions
   - Convergence rates and source conditions
   - Grasmair's second order VSC
   - Bregman iteration
2. Higher order variational source conditions
   - Hilbert space setting
   - Banach space setting
3. Numerical example

Maximum entropy regularization

Let

$$T\colon L^1([0,1]) \to L^2([0,1])$$

be bounded and linear. We want to approximate $f^\dagger \in L^1, f^\dagger > 0$ (e.g. a probability density) from noisy data $g^\delta \in L^2$ with

$$\|g^\dagger - g^\delta\|_{L^2} \leq \delta$$

and some a-priori guess $f_0 \in L^1$ of $f^\dagger$.
To this end we apply generalized Tikhonov regularization in the form of maximum entropy regularization

$$\hat{f}_\alpha \in \operatorname{argmin}_{f \in \mathcal{C}} \left[\|Tf - g^\delta\|_{L^2}^2 + \alpha \mathcal{R}_{f_0}(f) \right],$$

where actually $\mathcal{R}_{f_0}(f) = \mathrm{KL}(f, f_0)$.

Bregman iteration for maximum entropy regularization

Bregman iteration is then given by

$$\hat{f}_\alpha^{(n)} \in \operatorname{argmin}_{f\in\mathcal{C}} \left[\|Tf - g^\delta\|_{L^2}^2 + \alpha \Delta_{\mathcal{R}_{f_0}}(f, \hat{f}_\alpha^{(n-1)}) \right],$$

and as we have $\Delta_{\mathcal{R}_{f_0}}(f, \hat{f}_\alpha^{(n-1)}) = \mathrm{KL}(f, \hat{f}_\alpha^{(n-1)})$ this is the same as

$$\hat{f}_\alpha^{(n)} \in \operatorname{argmin}_{f\in\mathcal{C}} \left[\|Tf - g^\delta\|_{L^2}^2 + \alpha \mathrm{KL}(f, \hat{f}_\alpha^{(n-1)}) \right].$$

Y. Censor, S. Zenios. *Proximal minimization algorithm with D-functions.* **J. Optim. Theory Appl.** 73:451–464, 1992.

Convergence rates

Theorem (Hohage, S)

Suppose that T and its L^2-adjoint T^* are $a > 0$ times smoothing. Moreover, suppose there exists $\rho > 0$ such that

$$\rho \leq f^\dagger, f_0 \leq \rho^{-1} \qquad \textit{a.e. in } [0,1]$$

as well, as $\sup_{f \in \mathcal{C}} \|f\|_{L^\infty} < \infty$ *and* $a > d/2$. *Finally, assume that*

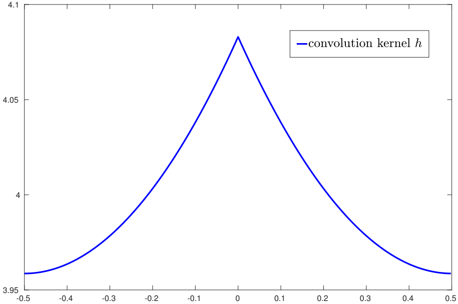
$$f^\dagger, f_0 \in B^s_{2,\infty}([0,1]) \qquad \textit{for some } s \in \left(2a + \frac{1}{2}, 3a\right].$$

Then VSC^3 *holds true with* $\mu = 2$ *and* $\psi(\tau) = C\tau^{\frac{s-2a}{s-a}}$ *for some* $C > 0$. *Hence, for the parameter choice* $\alpha \sim \delta^{\frac{2a}{s+a}}$ *we obtain the convergence rate*

$$\mathrm{KL}\left(f^\dagger, \hat{f}_\alpha^{(2)}\right) \leq \mathcal{O}\left(\delta^{\frac{2s}{s+a}}\right), \qquad \delta \searrow 0.$$

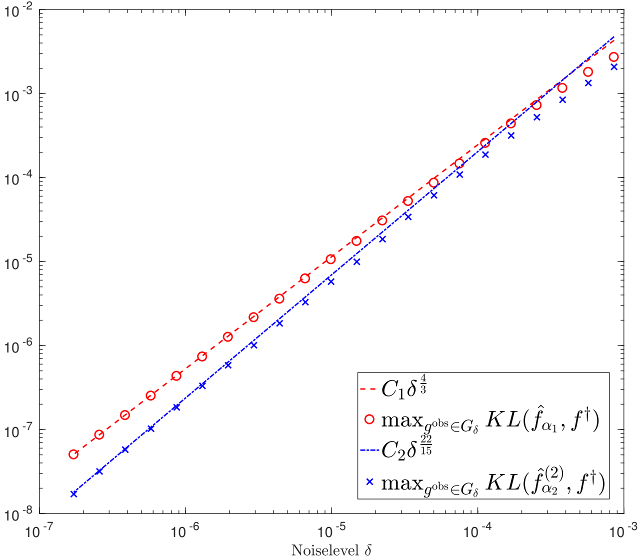
Test example

Let T be given as the convolution operator
$$Tf = f * h = \left(\tfrac{1}{4} - \partial_x^2\right)^{-1} f.$$

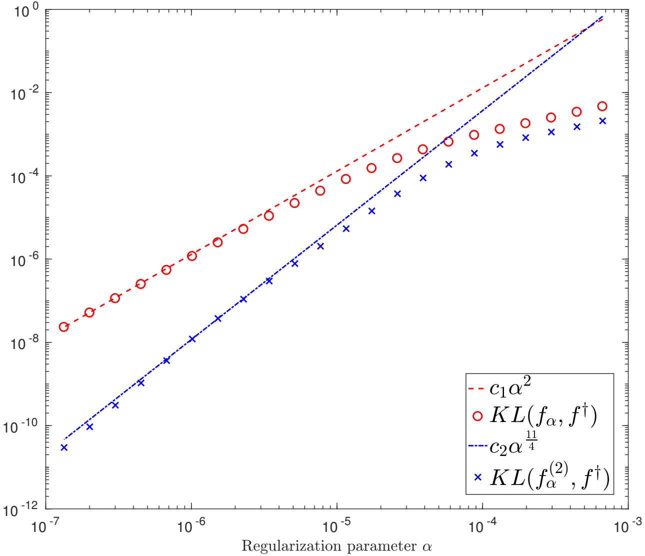


Let $f^\dagger = 1 + B_5$, where $B_5 \in B^{5.5}_{2,\infty}$, $f_0 = 1$.

Convergence rates



Approximation Error



Open questions

- VSCs of order > 3 in Banach spaces
- higher order convergence rates for non-linear forward operator
- higher order convergence rates for other regularization method, e.g. alternating split Bregman, Landweber, ...

Thank you for your attention!