About Using Dynamical Systems as Regularisation Methods and Their Optimal Convergence Rates

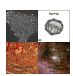
Peter Elbau

joint work with

Radu Boţ, Guozhi Dong, and Otmar Scherzer

Reconstruction Methods for Inverse Problems
Banff International Research Station
28th of June, 2019

universität wien



FWF Special Research Program F68
Tomography Across the Scales
tomography.csc.univie.ac.at

FWF Der Wissenschaftsfonds.

Introduction

- Throughout this talk, let $\mathcal{X}$ and $\mathcal{Y}$ be two real Hilbert spaces and $L : \mathcal{X} \to \mathcal{Y}$ be a bounded, linear operator.

- We consider the inverse problem

  $$Lx = y$$

  to determine for a given approximation $y \in \bar{B}_\delta(y^*)$ of the exact data $y^* \in \mathcal{Y}$ a good approximation of the minimum norm solution $x^* \in L^{-1}(\{y^*\})$ characterised by

  $$\|x^*\| = \inf\{\|x\| \mid Lx = y^*\}.$$

- A classical way to solve this is by Tikhonov regularisation where regularised solutions $X_\alpha(y)$, $y \in \mathcal{Y}$, $\alpha > 0$, are defined via

  $$\mathcal{T}_{y,\alpha}(X_\alpha(y)) = \inf_{x \in \mathcal{X}} \mathcal{T}_{y,\alpha}(x),$$

  with the functionals

  $$\mathcal{T}_{y,\alpha} : \mathcal{X} \to [0,\infty),\ \mathcal{T}_{y,\alpha}(x) = \|Lx - y\|^2 + \alpha\|x\|^2.$$

- These regularised solutions are well-defined and we can write $X_\alpha(y)$ explicitly in the form

  $$X_\alpha(y) = (L^*L + \alpha)^{-1}L^*y.$$

Regularisation by Bounded Approximations of the Inverse

- In the following, we want to focus on regularisation methods of the form

$$X_\alpha(y) = R_\alpha(L^*L)L^*y$$

for some family $(R_\alpha)_{\alpha>0}$ of continuous functions $R_\alpha : [0,\infty) \to [0,\infty)$.

- Hereby, $R_\alpha(L^*L)$ should be considered as a bounded approximation of the inverse of $L^*L|_{\mathcal{N}(L)^\perp}$, which converges to it for $\alpha \to 0$.

- For this kind of methods, we want to study how fast the regularised solution converges to the exact solution as the error in the data disappears.

- If we choose the regularisation parameter α optimal, this means to estimate the best worst case error

$$\sup_{y\in\bar{B}_\delta(y^*)} \inf_{\alpha>0} \|X_\alpha(y) - x^*\|^2.$$

- As an intermediate step, it is helpful to consider the regularisation error which is introduced by the regularisation for exact data y^*:

$$\|X_\alpha(y^*) - x^*\|^2.$$

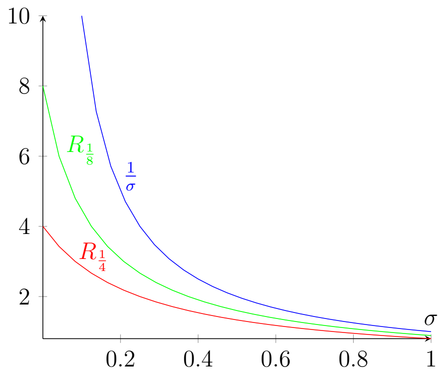
H. W. Engl, M. Hanke, and A. Neubauer
Regularization of Inverse Problems
Kluwer Academic Publishers, 1996

A. Neubauer
On Converse and Saturation Results for Tikhonov Regularization of Linear Ill-Posed Problems
SIAM Journal on Numerical Analysis 34.2. 1997

Example: Regularisation Error for Tikhonov Regularisation

- Tikhonov regularisation corresponds to the choice

$$R_\alpha(\sigma) = \frac{1}{\sigma + \alpha}.$$



- Consider as an example the operator

$$L : \ell^2 \to \ell^2, \quad L(x_k)_{k=1}^\infty = (\lambda_k x_k)_{k=1}^\infty,$$

with a monotonically decreasing sequence $\lambda_k \downarrow 0$.

- Then, the singular values of L are $\sigma_k = \lambda_k^2$ and

$$\|X_\alpha(y^*) - x^*\|^2 = \|R_\alpha(L^*L)L^*y^* - x^*\|^2 = \left\|\left(R_\alpha(L^*L)L^*L - \mathbb{1}\right)x^*\right\|^2 = \sum_{k=1}^\infty \left[\frac{\alpha x_k^*}{\sigma_k + \alpha}\right]^2.$$

- For those k where $\sigma_k \gg \alpha$, we get $\frac{\alpha}{\sigma_k+\alpha} = \mathcal{O}(\alpha)$, that is, as the regularisation parameter α vanishes, the components $X_{\alpha,k}(y^*)$ belonging to large singular values converge linearly to those of the exact solution.

- For those k with $\sigma_k \ll \alpha$, we have on the other hand $\frac{\alpha}{\sigma_k+\alpha} = \mathcal{O}(1)$. Therefore, the error from the components corresponding to the small singular values behaves as the values x_k^*.

Example: Regularisation Error for Tikhonov Regularisation and Spectral Tail

We thus define the spectral tail

$$e : (0,\infty) \to \mathbb{R},\ e(\sigma) = \sum_{\{k|\sigma_k \le \sigma\}} (x_k^*)^2.$$

Proposition: Regularisation Error for Tikhonov Regularisation

Let $\mu < 2$. Then, there exists a constant $C_1 > 0$ with

$$e(\sigma) \le C_1 \sigma^\mu \text{ for all } \sigma > 0 \text{ [spectral tail]}$$

if and only if there exists a constant $C_2 > 0$ with

$$\|X_\alpha(y^*) - x^*\|^2 \le C_2 \alpha^\mu \text{ for all } \alpha > 0 \text{ [regularisation error]}.$$

That the spectral tail estimate follows from the convergence rate part can be seen from

$$\frac{\alpha}{\sigma_k + \alpha} \ge \frac{1}{2} \text{ for all } \sigma_k \le \alpha,$$

which implies

$$\frac{1}{4} e(\alpha) \le \sum_{\{k|\sigma_k \le \alpha\}} \frac{\alpha^2}{(\sigma_k + \alpha)^2} {x_k^*}^2 + \sum_{\{k|\sigma_k > \alpha\}} \frac{\alpha^2}{(\sigma_k + \alpha)^2} {x_k^*}^2 = \|X_\alpha(y^*) - x^*\|^2.$$

Example: Error for Noisy Data for Tikhonov Regularisation

- To get from this an expression for the best worst case error, we use the triangular inequality to estimate

$$\|X_\alpha(y) - x^*\|^2 \le \big(\|X_\alpha(y) - X_\alpha(y^*)\| + \|X_\alpha(y^*) - x^*\|\big)^2.$$

- The first term can be bounded by a multiple of $\|y - y^*\|^2$ because of the boundedness of the operator $R_\alpha(L^*L)L^*$:

$$\|X_\alpha(y) - X_\alpha(y^*)\|^2 = \|R_\alpha(L^*L)L^*(y - y^*)\|^2 = \left\langle y - y^*, R_\alpha^2(LL^*)LL^*(y - y^*)\right\rangle.$$

  With

$$\sigma R_\alpha^2(\sigma) = \frac{\sigma}{(\sigma+\alpha)^2} \le \frac{1}{4\alpha} \text{ for all } \sigma \ge 0,$$

  we explicitly get

$$\|X_\alpha(y) - X_\alpha(y^*)\|^2 \le \frac{1}{4}\frac{\delta^2}{\alpha} \text{ for all } y \in \bar{B}_\delta(y^*).$$

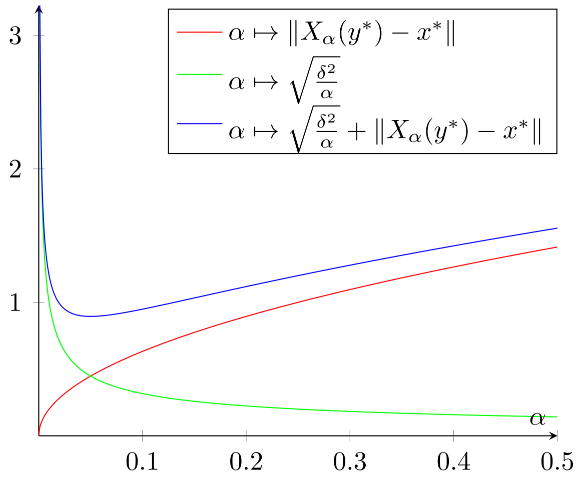
- The second term is the regularisation error which we estimated before.

Example: Best Worst Case Error for Tikhonov Regularisation

- Thus, the best worst case error can be estimated by

$$\sup_{y\in\bar{B}_\delta(y^*)}\inf_{\alpha>0}\|X_\alpha(y)-x^*\|^2 \le \inf_{\alpha>0}\left(\frac{1}{2}\sqrt{\frac{\delta^2}{\alpha}}+\|X_\alpha(y^*)-x^*\|\right)^2.$$

- An order optimal choice for the parameter α in this infimum can be achieved by balancing both terms $\frac{\delta^2}{\alpha}$ and $\|X_\alpha(y^*)-x^*\|^2$.



The error $\|X_\alpha(y)-x^*\|$ is estimated by the sum of the monotonically increasing regularisation error and the monotonically decreasing term $\sqrt{\frac{\delta^2}{\alpha}}$ from the noise.

Noise-free to Noisy Transform and the Best Worst Case Error for Tikhonov Regularisation

This leads us to choose α such that $\delta^2 = \alpha\|X_\alpha(y^*) - x^*\|^2$.

Definition: Noise-free to Noisy Transform

Let $\varphi : (0,\infty) \to [0,\infty)$ be a monotonically increasing function which is not everywhere zero. We define the noise-free to noisy transform

$$\Phi[\varphi] : (0,\infty) \to (0,\infty),\ \Phi[\varphi](\delta) = \frac{\delta^2}{\hat{\varphi}^{-1}(\delta)},$$

where

$$\hat{\varphi}(\alpha) = \sqrt{\alpha\varphi(\alpha)} \text{ and } \hat{\varphi}^{-1}(\delta) = \inf\{\alpha > 0 \mid \hat{\varphi}(\alpha) \geq \delta\}.$$

- We remark that the definition is monotone in the sense that $\varphi \leq \psi$ implies $\Phi[\varphi] \leq \Phi[\psi]$.
- If we have that

  $$\|X_\alpha(y^*) - x^*\|^2 \leq \varphi(\alpha) \text{ for all } \alpha > 0$$

  for some non-trivial, monotonically increasing function $\varphi : (0,\infty) \to [0,\infty)$, then we can thus estimate the best worst case error by

  $$\sup_{y\in\bar{B}_\delta(y^*)} \inf_{\alpha>0} \|X_\alpha(y) - x^*\|^2 \leq 4\Phi\left[\alpha \mapsto \|X_\alpha(y^*) - x^*\|^2\right](\delta) \leq \Phi[\varphi](\delta) \text{ for all } \delta > 0.$$

Generalisation to Monotonic Regularisation Methods

Definition: Monotonic Regularisation Method

A family of continuous functions $R_\alpha : [0, \infty) \to [0, \infty)$ is a monotonic regularisation method if

(i) we have $R_\alpha(\sigma) \leq \frac{1}{\sigma}$ [no overshooting];

(ii) the error function

$$\tilde{R}_\alpha : (0, \infty) \to [0, 1],\ \tilde{R}_\alpha(\sigma) = 1 - \sigma R_\alpha(\sigma),$$

is monotonically decreasing [the error is smaller for larger singular values];

(iii) the function

$$(0, \infty) \to [0, 1],\ \alpha \mapsto \tilde{R}_\alpha(\sigma)$$

is for every $\sigma > 0$ continuous and monotonically increasing [the error decreases as $\alpha \downarrow 0$];

(iv) there exist constants $\beta, \tilde{\beta} \in (0, 1)$ such that $R_\alpha(\sigma) \leq \frac{\beta}{\sqrt{\alpha\sigma}}$ and $1 - \beta \leq \tilde{R}_\alpha(\alpha) < \tilde{\beta}$ [normalisation of the regularisation parameter α].

The error function $\tilde{R}_\alpha$ is hereby introduced such that

$$x^* - X_\alpha(y^*) = (\mathbb{1} - R_\alpha(L^*L)L^*L)x^* = \tilde{R}_\alpha(L^*L)x^*.$$

Saturation of Convergence Rates

For this type of regularisation methods, we can characterise the convergence rates precisely in terms of the spectral tail

$$e : (0,\infty) \to [0,\infty),\ \ e(\sigma) = \|\mathbf{E}_{(0,\sigma]}x^*\|^2$$

of the exact solution x^*. Here, $A \mapsto \mathbf{E}_A$ denotes the spectral measure of the operator L^*L.

Since the convergence rates cannot exceed the rate at which the components of the large singular values are restored to its exact values, we only consider convergence functions which do not decay too fast.

Definition: Compatibility of Convergence Rates

A monotonically increasing function $\varphi : (0,\infty) \to (0,\infty)$ is called compatible with the regularisation method $(R_\alpha)_{\alpha>0}$ if there exists for arbitrary $\Lambda > 0$ a monotonically decreasing, integrable function $F : [1,\infty) \to \mathbb{R}$ such that

$$\tilde{R}_\alpha^2(\sigma) \le F\left(\frac{\varphi(\sigma)}{\varphi(\alpha)}\right) \text{ for all } 0 < \alpha \le \sigma \le \Lambda.$$

For the Tikhonov example, we may consider convergence rates of the form $\varphi(\alpha) = \alpha^\mu$. Then, with the error function $\tilde{R}_\alpha(\sigma) = \frac{\alpha}{\sigma+\alpha}$, this condition reads

$$\frac{\alpha^2}{(\sigma+\alpha)^2} \le F\left((\tfrac{\sigma}{\alpha})^\mu\right) \text{ for all } 0 < \alpha \le \sigma \le \Lambda, \text{ that is, } \frac{1}{(z^{\frac{1}{\mu}}+1)^2} \le F(z) \text{ for all } z \in [1,\infty),$$

which is only possible for $\mu < 2$ (with $F(z) = z^{-\frac{2}{\mu}}$, for example).

Optimal Convergence Rates

Theorem: Characterisation of Convergence Rates

Let $\varphi : (0,\infty) \to (0,\infty)$ be a monotonically increasing function which is compatible with $(R_\alpha)_{\alpha>0}$ and there exists a continuous and monotonically increasing function $G : (0,\infty) \to (0,\infty)$ with $\varphi(\gamma\alpha) \le G(\gamma)\varphi(\alpha)$ for all $\gamma \ge 1,\ \alpha > 0$. Then, the following statements are equivalent:

(i) There exists a constant $C_1 > 0$ with

$$\|\mathbf{E}_{(0,\sigma]}x^*\|^2 \le C_1\varphi(\sigma) \text{ for all } \sigma > 0$$ [spectral tail].

(ii) There exists a constant $C_2 > 0$ with

$$\|X_\alpha(y^*) - x^*\|^2 \le C_2\varphi(\alpha) \text{ for all } \alpha > 0$$ [regularisation error].

(iii) There exists a constant $C_3 > 0$ with

$$\sup_{y\in\bar{B}_\delta(y^*)} \inf_{\alpha>0} \|X_\alpha(y) - x^*\|^2 \le C_3\Phi[\varphi](\delta) \text{ for all } \delta > 0$$ [best worst case error].

A. Neubauer
On Converse and Saturation Results for Tikhonov Regularization of Linear Ill-Posed Problems
SIAM Journal on Numerical Analysis 34.2. 1997

R. Boţ, G. Dong, P. Elbau, and O. Scherzer
Convergence Rates of First and Higher Order Dynamics for Solving Linear Ill-posed Problems
arXiv: 1812.09343, 2018

Choice of Convergence Rate Functions

- The most common choices for the convergence rate function φ are the Hölder rates ϕ_μ,

$$\phi_\mu(\alpha) = \alpha^\mu,\ \mu > 0,$$

for which we have with $\hat{\phi}_\mu(\alpha) = \sqrt{\alpha\phi_\mu(\alpha)} = \alpha^{\frac{1+\mu}{2}}$ that

$$\Phi[\phi_\mu](\delta) = \frac{\delta^2}{\delta^{\frac{2}{1+\mu}}} = \delta^{\frac{2\mu}{1+\mu}}.$$

- But also, for example, logarithmic rates $\psi_{\mu,\nu}$,

$$\psi_{\mu,\nu}(\alpha) = \begin{cases} |\log(\alpha)|^{-\mu} & \text{if } \alpha < \mathrm{e}^{-\frac{\mu}{\nu}}, \\ (\frac{\nu}{\mu})^\mu & \text{if } \alpha \geq \mathrm{e}^{-\frac{\mu}{\nu}}, \end{cases} \text{ with } \mu > 0,\ \nu > 0,$$

are a valid choice for φ.

Because of the estimate

$$\frac{\psi_{\mu,\nu}(\sigma)}{\psi_{\mu,\nu}(\alpha)} \leq \left(\frac{\sigma}{\alpha}\right)^\nu \text{ for all } 0 < \alpha \leq \sigma,$$

the compatibility condition for $\psi_{\mu,\nu}$ can be easily derived from the ones for ϕ_μ.

And for its noise-free to noisy transform, we get with some constants $C, c > 0$:

$$c\psi_{\mu,\nu}(\delta) \leq \Phi[\psi_{\mu,\nu}](\delta) \leq C\psi_{\mu,\nu}(\delta) \text{ for all } \delta > 0.$$

The Standard Source Condition

- One of the first conditions on the source x^* to obtain convergence rates was the range condition

$$x^* \in \mathcal{R}(\varphi^{\frac{1}{2}}(L^*L)).$$

- This condition implies (by plugging in $x^* = \varphi^{\frac{1}{2}}(L^*L)\xi$ for some $\xi \in \mathcal{X}$) that

$$\|\mathbf{E}_{(0,\sigma]}x^*\|^2 \leq C_1\varphi(\sigma) \text{ for all } \sigma > 0$$

  for some constant $C_1 > 0$.

- Thus, this standard source condition is sufficient to guarantee with some constants $C_2 > 0$ and $C_3 > 0$ the convergence rates

$$\|X_\alpha(y^*) - x^*\|^2 \leq C_2\varphi(\alpha) \text{ for all } \alpha > 0$$

  and

$$\sup_{y \in \bar{B}_\delta(y^*)} \inf_{\alpha > 0} \|X_\alpha(y) - x^*\|^2 \leq C_3\Phi[\varphi](\delta) \text{ for all } \delta > 0.$$

- However, the condition is not optimal in the sense that the method can converge at these rates without x^* fulfilling the standard source condition.

C. W. Groetsch
The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind
Pitman Publishing, 1984

A. Neubauer
On Converse and Saturation Results for Tikhonov Regularization of Linear Ill-Posed Problems
SIAM Journal on Numerical Analysis 34.2. 1997

Variational Source Conditions

A way to optimal source conditions was found in variational inequalities, originally introduced in the general setting of non-linear operators on Banach spaces. For linear operators, we consider here the homogeneous version.

Theorem: Variational Inequalities and Spectral Tail

Let $\eta \in (0,1)$ and $\varphi : (0,\infty) \to (0,\infty)$ be a monotonically increasing function.Then, there exists a constant $C_1 > 0$ with

$$\|\mathbf{E}_{(0,\sigma]} x^*\|^2 \leq C_1 \varphi(\sigma) \text{ for all } \sigma > 0 \text{ [spectral tail]}$$

if and only if there exists a constant C_η with

$$\langle x^*, x \rangle \leq C_\eta \|\varphi^{\frac{1}{2\eta}}(L^*L)x\|^\eta \, \|x\|^{1-\eta} \text{ for all } x \in \mathcal{X} \text{ [variational source condition].}$$

That the variational inequality is sufficient, is easy to see by evaluating it at $x = \mathbf{E}_{(0,\sigma]} x^*$.

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 23.3. 2007

T. Hein and B. Hofmann
Approximate source conditions for nonlinear ill-posed problems – chances and limitations
Inverse Problems 25, 035003, 2009

R. Andreev, P. Elbau, M. V. de Hoop, L. Qiu, and O. Scherzer
Generalized Convergence Rates Results for Linear Inverse Problems in Hilbert Spaces
Numerical Functional Analysis and Optimization 36.5. 2015

V. Albani, P. Elbau, M. V. de Hoop, and O. Scherzer
Optimal Convergence Rates Results for Linear Inverse Problems in Hilbert Spaces
Numerical Functional Analysis and Optimization 37.5. 2016

Proof of the Necessity

- We estimate the left hand side by choosing a $\Lambda > 0$ so that

$$\frac{1}{2}\left|\langle x^*, x\rangle\right| \le \left|\left\langle \mathbf{E}_{(0,\Lambda]} x^*, x\right\rangle\right| \text{ and } \frac{1}{2}\left|\langle x^*, x\rangle\right| \le \left|\left\langle \mathbf{E}_{[\Lambda,\infty)} x^*, x\right\rangle\right|.$$

- The first term is directly estimated by

$$\left|\left\langle \mathbf{E}_{(0,\Lambda]} x^*, x\right\rangle\right| \le \|\mathbf{E}_{(0,\Lambda]} x^*\|\,\|x\| \le C_1 \|x\| \varphi^{\frac{1}{2}}(\Lambda).$$

- For the second term, we consider the bounded, invertible operator $T = \varphi^{\frac{1}{2\eta}}(L^*L)|_{\mathcal{R}(\mathbf{E}_{[\Lambda,\infty)})}$. Then, we have with some $C > 0$ that

$$\left|\left\langle \mathbf{E}_{[\Lambda,\infty)} x^*, x\right\rangle\right| = \left|\left\langle T^{-1}\mathbf{E}_{[\Lambda,\infty)} x^*, T\mathbf{E}_{[\Lambda,\infty)} x\right\rangle\right| \le \sqrt{C}\frac{\|\varphi^{\frac{1}{2\eta}}(L^*L)x\|}{\varphi^{\frac{1-\eta}{2\eta}}(\Lambda)},$$

since we have a $C > 0$ so that

$$\|T^{-1}\mathbf{E}_{[\Lambda,\infty)} x^*\|^2 \le \lim_{\varepsilon \downarrow 0} \int_{\Lambda-\varepsilon}^{\|L\|^2} \frac{1}{\varphi^{\frac{1}{\eta}}(\sigma)}\,\mathrm{d}e(\sigma) \le C\frac{1}{\varphi^{\frac{1}{\eta}-1}(\Lambda)}.$$

- Thus, with some constant $C_\eta > 0$ (independent of Λ)

$$\frac{1}{2}\left|\langle x^*, x\rangle\right| \le \left|\left\langle \mathbf{E}_{(0,\Lambda]} x^*, x\right\rangle\right|^{1-\eta}\left|\left\langle \mathbf{E}_{[\Lambda,\infty)} x^*, x\right\rangle\right|^{\eta} \le C_\eta \|\varphi^{\frac{1}{2\eta}}(L^*L)x\|^\eta \|x\|^{1-\eta}.$$

Approximative Source Conditions

Another equivalent source condition can be obtained by considering how fast the distance of the solution x^* to the set $\{\psi(L^*L)\xi \mid \xi \in \bar{B}_\rho(0)\}$ converges to zero as $\rho \to \infty$.

Theorem: Approximative Source Condition and Spectral Tail

Let $\eta \in (0,1)$ and $\varphi : (0,\infty) \to (0,\infty)$ be a monotonically increasing function which is compatible with $(R_\alpha)_{\alpha>0}$. Then, there exists a constant $C_1 > 0$ with

$$\|\mathbf{E}_{(0,\sigma]}x^*\|^2 \le C_1\varphi(\sigma) \text{ for all } \sigma > 0 \text{ [spectral tail]}$$

if and only if there exists a constant C_η with

$$\inf_{\xi\in\bar{B}_\rho(0)} \|x^* - \varphi^{\frac{1}{2\eta}}(L^*L)\xi\| \le C_\eta\rho^{-\frac{\eta}{1-\eta}} \text{ for all } \rho > 0 \text{ [approximative source condition].}$$

That the spectral tail implies the approximative source condition can be directly checked by inserting $\xi = T^{-1}\mathbf{E}_{(\alpha,\infty)}x^*$, where $T = \varphi^{\frac{1}{2\eta}}(L^*L)|_{\mathcal{R}(\mathbf{E}_{(\alpha,\infty)})}$. Then, we have

$$\|x^* - \varphi^{\frac{1}{2\eta}}(L^*L)\xi\|^2 = \|\mathbf{E}_{(0,\alpha]}x^*\|^2 \le \varphi(\alpha) \text{ and } \|\xi\|^2 \le \int_\alpha^{\|L\|^2} \frac{1}{\varphi^{\frac{1}{\eta}}(\sigma)}\,\mathrm{d}\|\mathbf{E}_{(0,\sigma]}x_*\|^2 \le C\varphi^{1-\frac{1}{\eta}}(\alpha).$$

B. Hofmann and P. Mathé
Analysis of Profile Functions for General Linear Regularization Methods
SIAM Journal on Numerical Analysis 45.3. 2007

J. Flemming, B. Hofmann, and P. Mathé
Sharp converse results for the regularization error using distance functions
Inverse Problems 27, 025006, 2011

Iterative Regularisation Methods and their Continuous Limits

- A classical example of an interative regularisation algorithm is Landweber's method, which can be seen as a gradient descent method for minimising the convex energy $\mathcal{S}_y(x) = \frac{1}{2}\|Lx - y\|^2$:

$$x^{(k+1)}(y) = x^{(k)}(y) - \tau\nabla\mathcal{S}_y(x^{(k)}(y)) = x^{(k)}(y) - \tau L^*(Lx^{(k)}(y) - y),\ k \in \mathbb{N}_0,$$

  with appropriately chosen $\tau \in (0, \frac{2}{\|L\|^2})$ and $x^{(0)}(y) = 0$.

- If we stop the iteration at a certain iteration k, then we may consider $x^{(k)}(y)$ as a regularised solution for the data y, where k takes the role of the regularisation parameter.

- This can be interpreted as an explicit Euler discretisation of the differential equation

$$\partial_t\xi(t; y) + \nabla\mathcal{S}_y(\xi(t; y)) = 0,$$

  which with the initial condition $\xi(0; y) = 0$ is known as Showalter's method.

- Then, $\xi(t; y)$ is a regularised solution for the data y and t takes the role of the regularisation parameter.

D. W. Showalter
Representation and Computation of the Pseudoinverse
Proceedings of the American Mathematical Society 18. 1967

D. W. Showalter and A. Ben-Israel
Representation and Computation of the Generalized Inverse of a Bounded Linear Operator Between Hilbert Spaces
Atti della Accademia Nazionale dei Lincei. Rendiconti della Classe di Scienze Fisiche, Matematiche e Naturali 48. 1970

Nesterov's Algorithm

- An accelerated version of this first order method was suggested by Nesterov:

$$x^{(k+1)}(y) = \tilde{x}^{(k)}(y) - \hat{\tau}^2 \nabla \mathcal{S}_y(\tilde{x}^{(k)}(y)),$$
$$\tilde{x}^{(k)}(y) \quad = x^{(k)}(y) + \frac{k-1}{k+2}(x^{(k)}(y) - x^{(k-1)}(y)).$$

- This has the increased rate of

$$\mathcal{S}_y(x^{(k)}(y)) - \min_{x \in \mathcal{X}} \mathcal{S}_y(x) = \mathcal{O}(\tfrac{1}{k^2})$$

in the image domain.

- Writing $x^{(k)}(y) = \xi(k\hat{\tau}; y)$ and doing a Taylor approximation

$$x^{(k+1)}(y) - x^{(k)}(y) = \hat{\tau}\partial_t \xi(k\hat{\tau}; y) + \frac{\hat{\tau}^2}{2}\partial_{tt}\xi(k\hat{\tau}; y) + o(\hat{\tau}^2),$$

we find by setting $t = k\hat{\tau}$ and plugging this in the algorithm that

$$\hat{\tau}\partial_t \xi(k\hat{\tau}; y) + \frac{\hat{\tau}^2}{2}\partial_{tt}\xi(k\hat{\tau}; y) = \left(1 - \frac{3\hat{\tau}}{t + 2\hat{\tau}}\right)\left(\hat{\tau}\partial_t \xi(k\hat{\tau}; y) - \frac{\hat{\tau}^2}{2}\partial_{tt}\xi(k\hat{\tau}; y)\right)$$
$$- \hat{\tau}^2 \nabla \mathcal{S}_y\big(\xi(k\hat{\tau}; y) + \mathcal{O}(\hat{\tau})\big) + o(\hat{\tau}^2).$$

Y. Nesterov
A method of solving a convex programming problem with convergence rate $\mathcal{O}(\frac{1}{k^2})$
Soviet Mathematics. Doklady 27.2. 1983

Vanishing Viscosity Flow

Taking the formal limit $\hat{\tau} \to 0$ of this equation

$$\hat{\tau}\partial_t\xi(k\hat{\tau};y) + \frac{\hat{\tau}^2}{2}\partial_{tt}\xi(k\hat{\tau};y) = \left(1 - \frac{3\hat{\tau}}{t+2\hat{\tau}}\right)\left(\hat{\tau}\partial_t\xi(k\hat{\tau};y) - \frac{\hat{\tau}^2}{2}\partial_{tt}\xi(k\hat{\tau};y)\right) - \hat{\tau}^2\nabla\mathcal{S}_y\big(\xi(k\hat{\tau};y) + \mathcal{O}(\hat{\tau})\big) + o(\hat{\tau}^2),$$

the terms of order $\hat{\tau}^2$ give us the vanishing viscosity flow equation

$$\partial_{tt}\xi(t;y) + \frac{3}{t}\partial_t\xi(t;y) + \nabla\mathcal{S}_y(\xi(t;y)) = 0.$$

Theorem: Continuous Version of Nesterov's Algorithm

Let ξ be the solution of this differential equation with $\xi(0;y) = x^{(0)}(y)$ and $\partial_t\xi(0;y) = 0$, and $x^{(k)}(y)$ be the iterates of Nesterov's algorithm.
Then, we have for arbitrary $T > 0$ that

$$\lim_{\hat{\tau}\to 0} \max_{k \le \frac{T}{\hat{\tau}}} \|x^{(k)}(y) - \xi(k\hat{\tau};y)\| = 0.$$

W. Su, S. Boyd, and E. J. Candès
A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights
Journal of Machine Learning Research 17.153. 2016

Heavy Ball Method

- A physical interpretation of the vanishing viscosity flow

$$\partial_{tt}\xi(t;y) + \frac{b}{t}\partial_t\xi(t;y) + \nabla\mathcal{S}_y(\xi(t;y)) = 0,\ t > 0,$$
$$\xi(0;y) = 0,$$

  for arbitrary coefficient $b > 0$ in the case $\mathcal{X} = \mathbb{R}^3$ is that of a particle moving along the curve $t \mapsto \xi(t;y)$ in the potential field $\mathcal{S}_y$ suffering from the viscous resistance $\frac{b}{t}\partial_t\xi(t;y)$, which is linear in the velocity and tends to zero as the time t increases.

- An interesting comparision can be the corresponding flow for a non-vanishing resistance term, that is, considering the heavy ball equation

$$\partial_{tt}\xi(t;y) + b\partial_t\xi(t;y) + \nabla\mathcal{S}_y(\xi(t;y)) = 0,\ t > 0,$$
$$\partial_t\xi(0;y) = 0,$$
$$\xi(0;y) = 0$$

  for $b > 0$.

- In contrast to the vanishing viscosity equation, where $\partial_t\xi(0;y) = 0$ is enforced by the singularity at $t = 0$, we have to specify two initial conditions for the heavy ball equation.

Writing Iterative Methods in the Form of Regularisation Methods

- We thus consider a differential equation of the form

$$\partial_t^N \xi(t;y) + \sum_{k=1}^{N} a_k(t)\partial_t^k \xi(t;y) + L^*(L\xi(t;y) - y) = 0 \text{ for all } t > 0,$$
$$\partial_t^k \xi(0;y) = 0 \text{ for all } k = 0,\dots,N-1.$$

- To find the solution, we write

$$\xi(t;y) = \int_{(0,\|L\|^2]} \rho(t;\sigma)\,\mathrm{d}\mathbf{E}_\sigma L^* y.$$

- Then, the equation reduces to the initial value problem

$$\partial_t^N \rho(t;\sigma) + \sum_{k=1}^{N} a_k(t)\partial_t^k \rho(t;\sigma) + \sigma\rho(t;\sigma) = 1 \text{ for all } \sigma \geq 0,\ t > 0,$$
$$\partial_t^k \rho(0;\sigma) = 0 \text{ for all } \sigma \geq 0,\ k = 0,\dots,N-1.$$

- As before for the regularisation methods, let us introduce the error function

$$\tilde\rho(t;\sigma) = 1 - \sigma\rho(t;\sigma) \text{ for all } \sigma \geq 0,\ t > 0.$$

- The function $\tilde\rho$ solves the homogeneous equation with the initial conditions

$$\tilde\rho(0;\sigma) = 1 \text{ and } \partial_t^k \tilde\rho(0;\sigma) = 0 \text{ for } k = 1,\dots,N-1 \text{ for all } \sigma \geq 0.$$

Solutions for Showalter's, Heavy Ball, and Vanishing Viscosity Methods

Solving the corresponding equations, we have

- for Showalter's method the equation $\partial_t\tilde{\rho}(t;\sigma)+\sigma\tilde{\rho}(t;\sigma)=0$ with the solution

  $$\tilde{\rho}(t;\sigma)=\mathrm{e}^{-\sigma t};$$

- for the heavy ball method the equation $\partial_{tt}\tilde{\rho}(t;\sigma)+b\partial_t\tilde{\rho}(t;\sigma)+\sigma\tilde{\rho}(t;\sigma)=0$ with the solution

  $$\tilde{\rho}(t;\sigma)=\begin{cases}\mathrm{e}^{-\frac{bt}{2}}\left(\cosh\left(\beta_-(\sigma)\frac{bt}{2}\right)+\frac{1}{\beta_-(\sigma)}\sinh\left(\beta_-(\sigma)\frac{bt}{2}\right)\right) & \text{if } \sigma\in(0,\frac{b^2}{4}),\\ \mathrm{e}^{-\frac{bt}{2}}\left(\cos\left(\beta_+(\sigma)\frac{bt}{2}\right)+\frac{1}{\beta_+(\sigma)}\sin\left(\beta_+(\sigma)\frac{bt}{2}\right)\right) & \text{if } \sigma\in(\frac{b^2}{4},\infty),\\ \mathrm{e}^{-\frac{bt}{2}}(1+\frac{bt}{2}) & \text{if } \sigma=\frac{b^2}{4},\end{cases}$$

  where

  $$\beta_-(\sigma)=\sqrt{1-\frac{4\sigma}{b^2}}\text{ and }\beta_+(\sigma)=\sqrt{\frac{4\sigma}{b^2}-1};$$

- and for the vanishing viscosity method the equation $\partial_{tt}\tilde{\rho}(t;\sigma)+\frac{b}{t}\partial_t\tilde{\rho}(t;\sigma)+\sigma\tilde{\rho}(t;\sigma)=0$ with the solution

  $$\tilde{\rho}(t;\sigma)=u(t\sqrt{\sigma})\text{ with }u(\tau)=\left(\frac{2}{\tau}\right)^{\frac{1}{2}(b-1)}\Gamma(\tfrac{1}{2}(b+1))J_{\frac{1}{2}(b-1)}(\tau).$$

Showalter's Method

If we thus identify

$$R_\alpha(\sigma) = \rho(\tfrac{1}{\alpha};\sigma) = \frac{1-\mathrm{e}^{-\frac{\sigma}{\alpha}}}{\sigma}, \text{ and correspondingly, } \tilde{R}_\alpha(\sigma) = \tilde{\rho}(\tfrac{1}{\alpha};\sigma) = \mathrm{e}^{-\frac{\sigma}{\alpha}},$$

we find that $(R_\alpha)_{\alpha>0}$ fulfils all requirements of a monotonic regularisation method.

Corollary: Convergence Rates of Showalter's Method

Let $\eta \in (0,1)$, $\mu > 0$. Then, the following statements are equivalent ($X_\alpha(y) = R_\alpha(L^*L)L^*y$)

(i) There exists a constant $C_1 > 0$ with

$$\|\xi(t;y^*) - x^*\|^2 = \|X_{\frac{1}{t}}(y^*) - x^*\|^2 \le C_1 t^{-\mu} \text{ for all } t > 0.$$

(ii) There exists a constant $C_2 > 0$ with

$$\inf_{t>0} \|\xi(t;y) - x^*\|^2 = \inf_{t>0} \|X_{\frac{1}{t}}(y) - x^*\|^2 \le C_2 \|y - y^*\|^{\frac{2\mu}{1+\mu}} \text{ for all } y \in \mathcal{Y}.$$

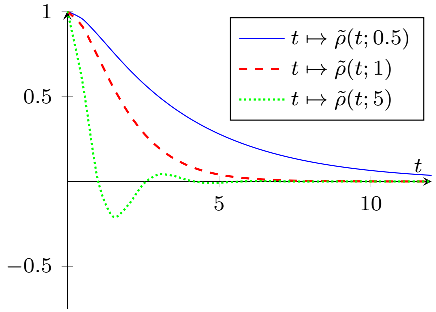
(iii) There exists a constant $C_\eta > 0$ with $\langle x^*, x\rangle \le C_\eta \|(L^*L)^{\frac{\mu}{2\eta}} x\|^\eta \, \|x\|^{1-\eta}$ for all $x \in \mathcal{X}$.

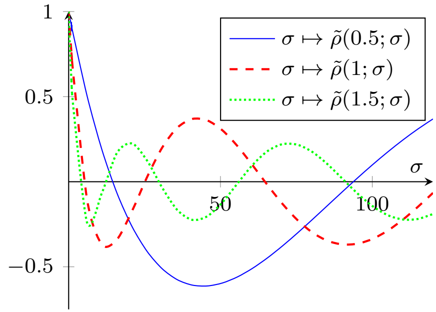
H. W. Engl, M. Hanke, and A. Neubauer
Regularization of Inverse Problems
Kluwer Academic Publishers, 1996

R. Boţ, G. Dong, P. Elbau, and O. Scherzer
Convergence Rates of First and Higher Order Dynamics for Solving Linear Ill-posed Problems
arXiv: 1812.09343, 2018

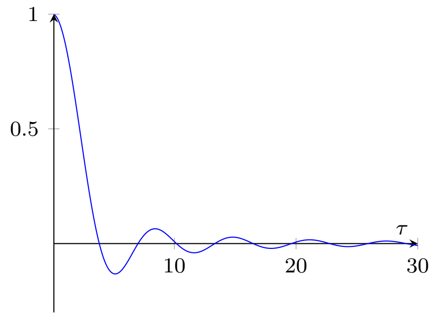
Heavy Ball and Vanishing Viscosity Methods as Regularisation Method

- The solution for the heavy ball method has the following form





- Similarly, the function u, and thus $\tilde{\rho}(\cdot; \sigma)$ and $\tilde{\rho}(t; \cdot)$, for the vanishing viscosity flow are oscillating functions.



- Therefore, these methods do not fit our definition of a monotonic regularisation method.

Non-Monotonic Regularisation Methods

We thus generalise our convergence rates results to regularisation methods with non-monotonic error functions.

Definition: Non-Monotonic Regularisation Method

A family of continuous functions $r_\alpha : [0,\infty) \to [0,\infty)$ is a non-monotonic regularisation method if

(i) we have $r_\alpha(\sigma) \leq \frac{2}{\sigma}$ [overshooting allowed but not more than 100% error];

(ii) the error function

$$\tilde{r}_\alpha : (0,\infty) \to [-1,1],\ \tilde{r}_\alpha(\sigma) = 1 - \sigma r_\alpha(\sigma),$$

is non-negative and monotonically decreasing on $(0,\alpha)$ [monotonic for small singular values at least up to the regularisation parameter];

(iii) there exists a monotonic regularisation method $(R_\alpha)_{\alpha>0}$ with

$$|\tilde{r}_\alpha| \leq \tilde{R}_\alpha$$

[there exists a monotonic envelope for the error];

(iv) there exists a constant $\beta \in (0,1)$ such that $r_\alpha(\sigma) \leq \frac{\beta}{\sqrt{\alpha\sigma}}$ [normalisation of the regularisation parameter].

Convergence Rates of Non-Monotonic Regularisation Methods

Let $(r_\alpha)_{\alpha>0}$ be a regularisation method and $(R_\alpha)_{\alpha>0}$ be a monotonic regularisation method with $|\tilde{r}_\alpha| \leq \tilde{R}_\alpha$. We denote the regularised solution by $x_\alpha(y) = r_\alpha(L^*L)L^*y$.

Theorem: Convergence Rates of Non-Monotonic Regularisation Methods

Let $\varphi : (0,\infty) \to (0,\infty)$ be a monotonically increasing, G-homogeneous function which is compatible with $(R_\alpha)_{\alpha>0}$. Then, the following statements are equivalent:

(i) There exists a constant $C_1 > 0$ with

$$\|\mathbf{E}_{(0,\sigma]}x^*\|^2 \leq C_1\varphi(\sigma) \text{ for all } \sigma > 0$$ [spectral tail].

(ii) There exists a constant $C_2 > 0$ with

$$\|x_\alpha(y^*) - x^*\|^2 \leq C_2\varphi(\alpha) \text{ for all } \alpha > 0$$ [regularisation error].

(iii) There exists a constant $C_3 > 0$ with

$$\sup_{y\in\bar{B}_\delta(y^*)} \inf_{\alpha>0} \|x_\alpha(y) - x^*\|^2 \leq C_3\Phi[\varphi](\delta) \text{ for all } \delta > 0$$ [best worst case error].

R. Boţ, G. Dong, P. Elbau, and O. Scherzer
Convergence Rates of First and Higher Order Dynamics for Solving Linear Ill-posed Problems
arXiv: 1812.09343, 2018

Proof for Exact Data

- We write $e(\sigma) = \|\mathbf{E}_{(0,\sigma]} x^*\|^2$.
- Since $\tilde{r}_\alpha$ is monotonically decreasing for small singular values, we have that

$$\tilde{r}_\alpha^2(\alpha) e(\alpha) \leq \int_0^\alpha \tilde{r}_\alpha^2(\sigma)\, \mathrm{d}e(\sigma) \leq \|\tilde{r}_\alpha(L^*L)x^*\|^2 = \|x^* - x_\alpha(y^*)\|^2 \leq \|x^* - X_\alpha(y^*)\|^2.$$

- On the other hand, we have with $0 \leq \tilde{R}_\alpha \leq 1$ that

$$\int_0^\alpha \tilde{R}_\alpha^2(\sigma)\, \mathrm{d}e(\sigma) \leq \int_0^\alpha \mathrm{d}e(\sigma) = e(\alpha)$$

  and thanks to the compatibility of φ to (R_α):

$$\int_\alpha^{\|L\|^2} \tilde{R}_\alpha^2(\sigma)\, \mathrm{d}e(\sigma) \leq \int_\alpha^{\|L\|^2} F\left(\frac{\varphi(\sigma)}{\varphi(\alpha)}\right) \mathrm{d}e(\sigma) \leq \varphi(\alpha)(1 + \|F\|_{L^1}).$$

- Putting this together, we find

$$C_1 e(\alpha) \leq \|x^* - x_\alpha(y^*)\|^2 \leq \|x^* - X_\alpha(y^*)\|^2 \leq C_2 e(\alpha) + C_3 \varphi(\alpha).$$

- This shows that for exact data, the convergence rates of the non-monotonic and its monotonic envelope as well as the spectral tail are equivalent to each other.

Proof for Noisy Data

- For the upper bound, we use (exactly as in the introductory example) the boundedness of the operator r_α:

$$\|X_\alpha(y) - X_\alpha(y^*)\|^2 \le \|x_\alpha(y) - x_\alpha(y^*)\|^2 = \left\langle y - y^*, r_\alpha^2(LL^*)LL^*(y - y^*)\right\rangle$$
$$\le \delta^2 \sup_\sigma \sigma r_\alpha^2(\sigma) \le \beta \frac{\delta^2}{\alpha},$$

  which gives us from convergence rates φ for the exact data that

$$\sup_{y \in \bar{B}_\delta(y^*)} \inf_{\alpha > 0} \|X_\alpha(y) - x^*\|^2 \le \Phi[\varphi](\delta) \text{ and } \sup_{y \in \bar{B}_\delta(y^*)} \inf_{\alpha > 0} \|x_\alpha(y) - x^*\|^2 \le \Phi[\varphi](\delta).$$

- For the lower bound, we write

$$\|x_\alpha(y) - x^*\|^2 = \|x_\alpha(y^*) - x^*\|^2 + \left\langle y - y^*, r_\alpha^2(LL^*)LL^*(y - y^*)\right\rangle$$
$$+ 2\left\langle r_\alpha(LL^*)(y - y^*), r_\alpha(LL^*)LL^*y^* - y^*\right\rangle.$$

- We pick the special value $\alpha_\delta = \hat{e}^{-1}(\delta)$, $\hat{e}(\alpha) = \sqrt{\alpha e(\alpha)}$, and assume that $\alpha_\delta \in \boldsymbol{\sigma}(L^*L) \setminus \{0\}$. (We can later do an interpolation argument to get the result also for values δ without this property.)

Proof for Noisy Data

- Then, for the interval $[a_\delta, 2\alpha_\delta]$ with $\tilde{R}_\alpha(a_\delta) < \tilde{\beta}$, we pick an element

$$y = y^* + \delta \frac{z_{\alpha,\delta}}{\|z_{\alpha,\delta}\|} \text{ with } z_{\alpha,\delta} \in \mathcal{R}(\mathbf{F}_{[a_\delta, 2\alpha_\delta]}),$$

where $\mathbf{F}$ is the spectral measure of LL^*, in such a way that the last term in the previous equation becomes non-negative, and we find

$$\|x_\alpha(y) - x^*\|^2 \geq \|x_\alpha(y^*) - x^*\|^2 + \delta^2 \min_{\sigma \in [a_\delta, 2\alpha_\delta]} \sigma r_\alpha^2(\sigma).$$

- By using the equivalence of the regularisation error to the spectral tail and $\tilde{R}_\alpha(a_\delta) < \tilde{\beta}$, we can estimate this further by

$$\inf_{\alpha > 0} \|x_\alpha(y) - x^*\|^2 \geq \min \left\{ c_1 e(\alpha_\delta), c_2 \frac{\delta^2}{\alpha_\delta} \right\}.$$

- However, since $\alpha_\delta = \hat{e}^{-1}(\delta)$, we have by definition of the noise-free to noisy transform Φ that

$$\inf_{\alpha > 0} \|x_\alpha(y) - x^*\|^2 \geq c_0 \Phi[e](\delta).$$

- Analogously, we get the lower bound for $\inf_{\alpha > 0} \|X_\alpha(y) - x^*\|^2$.

Convergence Rates for the Heavy Ball Method

If we set with the solutions ρ and $\tilde{\rho}$ for heavy ball method

$$r_\alpha(\sigma) = \rho(\tfrac{b}{2\alpha};\sigma), \text{ and correspondingly, } \tilde{r}_\alpha(\sigma) = \tilde{\rho}(\tfrac{b}{2\alpha};\sigma),$$

then it can be shown that $(r_\alpha)_{\alpha>0}$ fulfils all requirements of a non-monotonic regularisation method.

Corollary: Convergence of the Heavy Ball Method

Let $\eta \in (0,1)$ and $\mu > 0$. Then, the following statements are equivalent:

(i) There exists a constant $C_1 > 0$ with

$$\|\xi(t;y^*) - x^*\|^2 \le C_1 t^{-\mu} \text{ for all } t > 0.$$

(ii) There exists a constant $C_2 > 0$ with

$$\inf_{t>0} \|\xi(t;y) - x^*\|^2 \le C_2 \|y - y^*\|^{\frac{2\mu}{1+\mu}} \text{ for all } y \in \mathcal{Y}.$$

(iii) There exists a constant $C_\eta > 0$ with $\langle x^*, x\rangle \le C_\eta \|(L^*L)^{\frac{\mu}{2\eta}} x\|^\eta \|x\|^{1-\eta}$ for all $x \in \mathcal{X}$.

Y. Zhang and B. Hofmann
On the second-order asymptotical regularization of linear ill-posed inverse problems
Applicable Analysis: An International Journal. 2018

R. Boţ, G. Dong, P. Elbau, and O. Scherzer
Convergence Rates of First and Higher Order Dynamics for Solving Linear Ill-posed Problems
arXiv: 1812.09343, 2018

Convergence Rates for the Vanishing Viscosity Flow

Let $\tilde{\rho}(t;\sigma) = u(t\sqrt{\sigma})$ be the solution for the vanishing viscosity method. Setting $\tilde{r}_\alpha(\sigma) = \tilde{\rho}(\frac{c}{\sqrt{\alpha}};\sigma)$, for a suitable constant $c > 0$, $(r_\alpha)_{\alpha>0}$ fulfils all requirements of a non-monotonic regularisation method.

In contrast to Showalter's and the heavy ball method, the convergence for high singular values is not exponentially fast and the compatibility condition only allows for rates up to some Hölder rate.

Corollary: Convergence of the Vanishing Viscosity Solution

Let $\eta \in (0,1)$ and $\mu \in (0, \frac{b}{2})$. Then, the following statements are equivalent

(i) There exists a constant $C_1 > 0$ with

$$\|\xi(t;y^*) - x^*\|^2 \le C_1 t^{-2\mu} \text{ for all } t > 0.$$

(ii) There exists a constant $C_2 > 0$ with

$$\inf_{t>0} \|\xi(t;y) - x^*\|^2 \le C_2 \|y - y^*\|^{\frac{2\mu}{1+\mu}} \text{ for all } y \in \mathcal{Y}.$$

(iii) There exists a constant $C_\eta > 0$ with $\langle x^*, x\rangle \le C_\eta \|(L^*L)^{\frac{\mu}{2\eta}} x\|^\eta \, \|x\|^{1-\eta}$ for all $x \in \mathcal{X}$.

Convergence Rates of the Residuum without Source Conditions

- Without a source condition on the solution x^*, we do not get any convergence rates for the regularisation error $\|x_\alpha(y^*) - x^*\|^2$.

- However, we have for the convergence in the image domain:

$$\|Lx_\alpha(y^*)-y^*\|^2 = \|Lr_\alpha(L^*L)L^*Lx^* - Lx^*\|^2 = \|r_\alpha(L^*L)L^*L\bar{x}^* - \bar{x}^*\|^2 = \|x_\alpha(\bar{y}^*)-\bar{x}^*\|^2$$

  with $\bar{x}^* = \sqrt{L^*L}\,x^*$ and $\bar{y}^* = L\bar{x}^*$.

  Therefore, the convergence in the image domain corresponds to the convergence to the minimum-norm solution $\bar{x}^* \in \mathcal{R}((L^*L)^{\frac{1}{2}})$, which allows us to apply the previous results, even in the case without source conditions on x^*.

- In particular, we find for the vanishing viscosity flow that

$$\|L\xi(t;y^*) - y^*\|^2 = o(t^{-b}) + o(t^{-2}),$$

  slightly improving the previously known rates of $\mathcal{O}(t^{-\frac{2b}{3}})$ for $b \in (0,3)$ in the setting of general convex data terms $\mathcal{S}_y$.

H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont
Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity
Mathematical Programming. A Publication of the Mathematical Optimization Society 168.1–2. 2018

J.-F. Aujol, C. Dossal, and A. Rondepierre
Optimal convergence rates for Nesterov acceleration
arXiv: 1805.05719, 2018

Summary

- We have introduced a class of non-monotonic regularisation methods of the form

$$x_\alpha(y) = r_\alpha(L^*L)L^*y$$

for a linear inverse problem $Lx = y$ for which the convergence rates of the regularisation error

$$\|x_\alpha(y^*) - x^*\|^2$$

and the best worst case error

$$\sup_{y \in \bar{B}_\delta(y^*)} \inf_{\alpha>0} \|x_\alpha(y) - x^*\|^2$$

can be uniquely characterised by the behaviour of the spectral tail $\|\mathbf{E}_{(0,\sigma]}x^*\|^2$ of the minimum norm solution x^*, by a variational source condition, or by an approximative source condition.

- We applied this result in particular to the vanishing viscosity flow, which is the second order dynamical system

$$\partial_{tt}\xi(t;y) + \frac{b}{t}\partial_t\xi(t;y) + L^*(L\xi(t;y) - y) = 0,$$

which corresponds for $b = 3$ to the continuum limit of Nesterov's accelerated gradient descent method, and the resulting convergence rates of the residual error show the same quadratic convergence as Nestrov's algorithm.

Thank you very much for your attention!