UniGe | MaLGa

Iterative regularization for low complexity regularizers

Silvia Villa
MaLGa- Machine learning Genova Center
DIMA, Università di Genova

Data Science Seminar Series
London School of Economics and Political Science

February, 21^{st}, 2022

Outline

Collaborators

Joint project with **Lorenzo Rosasco**

and: **Guillaume Garrigos, Mathurin Massias, Cesare Molinari, Luca Calatroni, Cristian Vega, Simon Matet, Bang Cong Vu**.

Underdetermined linear systems

Given:

- $\mathcal{X}$, $\mathcal{Y}$ Hilbert spaces
- $A\colon \mathcal{X} \to \mathcal{Y}$ linear and bounded, $b \in R(A)$
- $R\colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ convex and lsc

Solve:

$$\min R(x) \ : \ Ax = b$$

Underdetermined linear systems

Given:

- $\mathcal{X}$, $\mathcal{Y}$ Hilbert spaces
- $A \colon \mathcal{X} \to \mathcal{Y}$ linear and bounded, $b \in R(A)$
- $R \colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ convex and lsc

Solve:

$$\min R(x) \ : \ Ax = b$$

If R is strongly convex then there exists a unique solution $x^\dagger$.

Inverse problems and learning — Choice of R

- $\|x\|^2$
- $\|x\|_1$
- $TV(x)$
- $\|x\|_*$

Inverse problems and learning — Choice of R

- $\|x\|^2$
- $\|x\|_1$
- $TV(x)$
- $\|x\|_*$

Convex and possibly nonsmooth

Inverse problems and learning — stability

Solve:

$$\min R(x) \ : \ Ax = b$$

knowing only b^δ such that $\|b - b^\delta\| \le \delta$.

Outline

Regularization

Strongly convex regularization

Iterative (implicit) regularization

Convex regularization

Special cases

Experiments

Explicit regularization a.k.a Tikhonov regularization

Given $D\colon \mathcal{Y} \times \mathcal{Y} \to [0, +\infty[$

$$\text{minimize} \quad D(Ax, b^\delta) + \lambda R(x)$$

Theorem

If:

- R *is strongly convex*
- *Im*$(A^*) \cap \partial R(x^\dagger) \neq \varnothing$
- $x^{\delta,\lambda}$ *is the unique solution of the regularized problem.*

Then

$$\|x^{\delta,\lambda} - x^\dagger\| \leq C\left(\frac{\delta}{\sqrt{\lambda}} + \sqrt{\delta} + \sqrt{\lambda}\right)$$

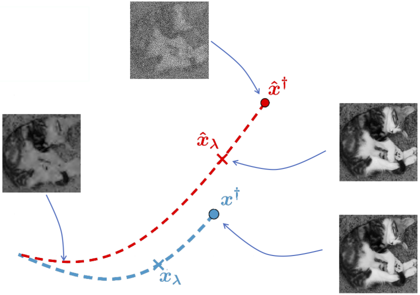
Choosing $\lambda_\delta \sim \delta$*:*

$$\|x^{\delta,\lambda_\delta} - x^\dagger\| \leq C\sqrt{\delta}.$$

[Burger-Osher, Convergence rates of convex variational regularization, 2004], [Benning-Burger, Error estimates for general fidelities, 2011]

What about computations?

- choose an interval $[\lambda_{\min}, \lambda_{\max}]$
- approximately solve the regularized problem for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$
- select the best λ according to a validation criterion



Outline

Regularization

Strongly convex regularization

Iterative (implicit) regularization

Convex regularization

Special cases

Experiments

Iterative regularization: an optimization point of view

1. Choose a convergent algorithm to solve

$$\min R(x) \; : \; Ax = b$$

   Call the iterates $(x_k)_{k \in \mathbb{N}}$.

Iterative regularization: an optimization point of view

1. Choose a convergent algorithm to solve

$$\min R(x) \; : \; Ax = b$$

   Call the iterates $(x_k)_{k\in\mathbb{N}}$.

2. Apply the same algorithm to

$$\min R(x) \; : \; Ax = b^\delta$$

   Call the iterates $(x_k^\delta)_{k\in\mathbb{N}}$.

Iterative regularization: an optimization point of view

1. Choose a convergent algorithm to solve

$$\min R(x) \;:\; Ax = b$$

   Call the iterates $(x_k)_{k\in\mathbb{N}}$.

2. Apply the same algorithm to

$$\min R(x) \;:\; Ax = b^\delta$$

   Call the iterates $(x_k^\delta)_{k\in\mathbb{N}}$.

3. $\|x_k^\delta - x^\dagger\| \leq \|x_k^\delta - x_k\| + \|x_k - x^\dagger\|$

UniGe | MaLGa

The algorithm in the strongly convex case

$$\min_{Ax=b} R(x) \qquad \longleftrightarrow \qquad \min_{x\in\mathcal{X}} R(x) + \iota_{\{b\}}(Ax),$$

where $\iota_{\{b\}}(x) = 0$ if $x = b$ and $\iota_{\{b\}}(x) = +\infty$ otherwise.

The algorithm in the strongly convex case

$$\min_{Ax=b} R(x) \quad \longleftrightarrow \quad \min_{x\in\mathcal{X}} R(x) + \iota_{\{b\}}(Ax),$$

where $\iota_{\{b\}}(x) = 0$ if $x = b$ and $\iota_{\{b\}}(x) = +\infty$ otherwise.

Dual problem

$$\min_{v\in\mathcal{Y}} d(v), \quad d(v) = R^*(-A^*v) + \langle b, v\rangle.$$

Im$(A^*) \cap \partial R(x^\dagger) \neq \varnothing \implies d$ **has a solution**
R strongly convex $\Rightarrow d$ **is smooth**

Let (v_k) be generated by an (accelerated) gradient method and

$$x_k = \nabla R^*(-\gamma A^* v_k).$$

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]
- a.k.a. mirror descent [Nemirovski-Yudin 1983, Teboulle-Beck 2003, Gunasekar et al 2018] or linearized Bregman iteration [Osher et al. 2005, Burger-Resmerita-He, 2007...]

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]
- a.k.a. mirror descent [Nemirovski-Yudin 1983, Teboulle-Beck 2003, Gunasekar et al 2018] or linearized Bregman iteration [Osher et al. 2005, Burger-Resmerita-He, 2007...]
- Theorem[Matet-Rosasco-V.-Vu, 2017] If x_k^δ is generated by Gradient Descent on the noisy dual, then

$$\|x_k^\delta - x^\dagger\| \leq \sqrt{k}\delta + \frac{1}{\sqrt{k}}, \qquad k_\delta \sim \delta^{-1} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

UniGe | MaLGa

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]
- a.k.a. mirror descent [Nemirovski-Yudin 1983, Teboulle-Beck 2003, Gunasekar et al 2018] or linearized Bregman iteration [Osher et al. 2005, Burger-Resmerita-He, 2007...]
- Theorem[Matet-Rosasco-V.-Vu, 2017] If x_k^δ is generated by Gradient Descent on the noisy dual, then

$$\|x_k^\delta - x^\dagger\| \leq \sqrt{k}\delta + \frac{1}{\sqrt{k}}, \qquad k_\delta \sim \delta^{-1} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

If x_k^δ is generated by Accelerated Gradient Descent on the noisy dual, then

$$\|x_k^\delta - x^\dagger\| \leq k\delta + \frac{1}{k}, \qquad k_\delta \sim \delta^{-1/2} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]
- a.k.a. mirror descent [Nemirovski-Yudin 1983, Teboulle-Beck 2003, Gunasekar et al 2018] or linearized Bregman iteration [Osher et al. 2005, Burger-Resmerita-He, 2007...]
- Theorem[Matet-Rosasco-V.-Vu, 2017] If x_k^δ is generated by Gradient Descent on the noisy dual, then

  $$\|x_k^\delta - x^\dagger\| \leq \sqrt{k}\delta + \frac{1}{\sqrt{k}}, \qquad k_\delta \sim \delta^{-1} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

  If x_k^δ is generated by Accelerated Gradient Descent on the noisy dual, then

  $$\|x_k^\delta - x^\dagger\| \leq k\delta + \frac{1}{k}, \qquad k_\delta \sim \delta^{-1/2} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

- ADMM, a.k.a. Bregman iteration [Bachmayr-Burger 2009, Burger et. al. 2007]

Regularization results

- $R(x) = ||x||^2$, Landweber method (1950), see [Engl-Hanke-Neubauer, Regularization of inverse problems, 1996], accelerated version [Neubauer, 2017]
- a.k.a. mirror descent [Nemirovski-Yudin 1983, Teboulle-Beck 2003, Gunasekar et al 2018] or linearized Bregman iteration [Osher et al. 2005, Burger-Resmerita-He, 2007...]
- Theorem[Matet-Rosasco-V.-Vu, 2017] If x_k^δ is generated by Gradient Descent on the noisy dual, then

$$\|x_k^\delta - x^\dagger\| \leq \sqrt{k}\delta + \frac{1}{\sqrt{k}}, \qquad k_\delta \sim \delta^{-1} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

  If x_k^δ is generated by Accelerated Gradient Descent on the noisy dual, then

$$\|x_k^\delta - x^\dagger\| \leq k\delta + \frac{1}{k}, \qquad k_\delta \sim \delta^{-1/2} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \sqrt{\delta}$$

- ADMM, a.k.a. Bregman iteration [Bachmayr-Burger 2009, Burger et. al. 2007]
- Also: nonlinear inverse problems [Kaltenbacher-Neubauer-Scherzer, Iterative Regularization for nonlinear inverse problems, 2008], learning [Yao-Rosasco-Caponnetto 2005, Rosasco-V. 2015]

Other discrepancies

$$\begin{array}{ccc} \begin{array}{c} \min R(x) \\ \text{s.t. } D(Ax,b)=0 \end{array} & \longrightarrow & \dfrac{1}{\lambda} D(Ax,b) + R(x) \\ \uparrow & & \downarrow \\ \min_{v\in\mathcal{Y}} \underbrace{\langle v,b\rangle + R^*(-A^*v)}_{=d(v)} & \longleftarrow & \underbrace{\dfrac{1}{\lambda} D^*(\lambda v, y) + R^*(-A^*v)}_{=d_\lambda(v)}. \end{array}$$

Other discrepancies

$$
\begin{array}{ccc}
\begin{array}{c} \min R(x) \\ \text{s.t. } D(Ax,b)=0 \end{array} & \longrightarrow & \frac{1}{\lambda} D(Ax,b) + R(x) \\
\uparrow & & \downarrow \\
\min_{v \in \mathcal{Y}} \underbrace{\langle v, b\rangle + R^*(-A^*v)}_{=d(v)} & \longleftarrow & \underbrace{\frac{1}{\lambda} D^*(\lambda v, y) + R^*(-A^*v)}_{=d_\lambda(v)}\,.
\end{array}
$$

A diagonal approach[Lemaire 80s-90s]

$$
x_{k+1} = \text{Algo}(x_k, \lambda_k), \qquad \text{with } \lambda_k \to 0.
$$

UniGe | MaLGa

Regularization results

- Assumptions on D
- Im$A^* \cap \partial R(x^\dagger) \neq \varnothing$
- $\lambda_k \to 0$ (at a suitable rate, depending on D)

If Algo = forward-backward on the noisy dual, then [Garrigos-Rosasco-V. 2017]

$$\|x_k^\delta - x^\dagger\| \leq \frac{1}{\sqrt{k}} + k\delta, \qquad k_\delta \sim \delta^{-2/3} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \delta^{1/3}$$

If Algo = accelerated forward-backward on the noisy dual, then [Calatroni-Garrigos-Rosasco-V. 2021]

$$\|x_k^\delta - x^\dagger\| \leq \frac{1}{k^2} + k^2\delta^2, \qquad k_\delta \sim \delta^{-1/2} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \delta^{1/2}$$

Regularization results

- Assumptions on D
- Im$A^* \cap \partial R(x^\dagger) \neq \varnothing$
- $\lambda_k \to 0$ (at a suitable rate, depending on D)

If Algo = forward-backward on the noisy dual, then [Garrigos-Rosasco-V. 2017]

$$\|x_k^\delta - x^\dagger\| \leq \frac{1}{\sqrt{k}} + k\delta, \qquad k_\delta \sim \delta^{-2/3} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \delta^{1/3}$$

If Algo = accelerated forward-backward on the noisy dual, then [Calatroni-Garrigos-Rosasco-V. 2021]

$$\|x_k^\delta - x^\dagger\| \leq \frac{1}{k^2} + k^2\delta^2, \qquad k_\delta \sim \delta^{-1/2} \implies \|x_{k_\delta}^\delta - x^\dagger\| \leq \delta^{1/2}$$

See also [Benning-Burger, 2011].

Outline

Regularization

Strongly convex regularization

Iterative (implicit) regularization

Convex regularization

Special cases

Experiments

Convex regularizers

What if R is not strongly convex?

Convex regularizers

What if R is not strongly convex?

Steps:

- identify an algorithm
- derive convergence rates on suitable quantities (the solution is not unique in general)
- special case: ℓ^1 norm
- unfeasible case

Based on: [Massias,Molinari,Rosasco, V., Iterative regularization for low complexity regularizers, 2022]

Primal-dual framework

Consider $R = F + G$, F L-smooth, and G convex

$$\min F(x) + G(x) + \iota_b(Ax) \qquad (P)$$

UniGe | MaLGa

Primal-dual framework

Consider $R = F + G$, F L-smooth, and G convex

$$\min F(x) + G(x) + \iota_b(Ax) \qquad (P)$$

Langrangian:

$$\mathcal{L}(x, y) = F(x) + G(x) + \langle y, Ax - b \rangle$$

Primal-dual framework

Consider $R = F + G$, F L-smooth, and G convex

$$\min F(x) + G(x) + \iota_b(Ax) \qquad (P)$$

Langrangian:

$$\mathcal{L}(x, y) = F(x) + G(x) + \langle y, Ax - b \rangle$$

Assumption: There exists a saddle point of $\mathcal{L}$ (primal-dual solution) (x_*, y_*)

$$\forall (x, y) \qquad \mathcal{L}(x_*, y) - \mathcal{L}(x, y_*) \leq 0$$

Primal-dual framework

Consider $R = F + G$, F L-smooth, and G convex

$$\min F(x) + G(x) + \iota_b(Ax) \qquad (P)$$

Langrangian:

$$\mathcal{L}(x, y) = F(x) + G(x) + \langle y, Ax - b\rangle$$

Assumption: There exists a saddle point of $\mathcal{L}$ (primal-dual solution) (x_*, y_*)

$$\forall (x, y) \qquad \mathcal{L}(x_*, y) - \mathcal{L}(x, y_*) \leq 0$$

(x_*, y_*) saddle point of $\mathcal{L}$ $\implies \begin{cases} -A^* y_* \in \partial R(x_*) \\ Ax_* = b \end{cases}$, $\quad x_*$ is a solution of (P)

UniGe | MaLGa

Condat-Vu algorithm

$$\begin{cases} \tilde{y}_k = 2y_k - y_{k-1} \\ x_{k+1} = \mathrm{prox}_{\tau G}(x_k - \tau(\nabla F(x_k) + A^* \tilde{y}_k)) \\ y_{k+1} = y_k + \sigma(Ax_{k+1} - b) \end{cases}$$

Condat-Vu algorithm

$$\begin{cases} \tilde{y}_k = 2y_k - y_{k-1} \\ x_{k+1} = \mathrm{prox}_{\tau G}(x_k - \tau(\nabla F(x_k) + A^* \tilde{y}_k)) \\ y_{k+1} = y_k + \sigma(A x_{k+1} - b) \end{cases}$$

- weak convergence of the iterates (x_k, y_k) to a primal-dual solution (if $\tau \leq (L + \sigma\|A\|^2)^{-1}$)

Condat-Vu algorithm

$$\begin{cases} \tilde{y}_k = 2y_k - y_{k-1} \\ x_{k+1} = \operatorname{prox}_{\tau G}(x_k - \tau(\nabla F(x_k) + A^* \tilde{y}_k)) \\ y_{k+1} = y_k + \sigma(A x_{k+1} - b) \end{cases}$$

- weak convergence of the iterates (x_k, y_k) to a primal-dual solution (if $\tau \leq (L + \sigma \|A\|^2)^{-1}$)
- preconditioning

Condat-Vu algorithm

$$\begin{cases} \tilde{y}_k = 2y_k - y_{k-1} \\ x_{k+1} = \operatorname{prox}_{\tau G}(x_k - \tau(\nabla F(x_k) + A^* \tilde{y}_k)) \\ y_{k+1} = y_k + \sigma(Ax_{k+1} - b) \end{cases}$$

- weak convergence of the iterates (x_k, y_k) to a primal-dual solution (if $\tau \leq (L + \sigma\|A\|^2)^{-1}$)
- preconditioning
- inexact computations of prox

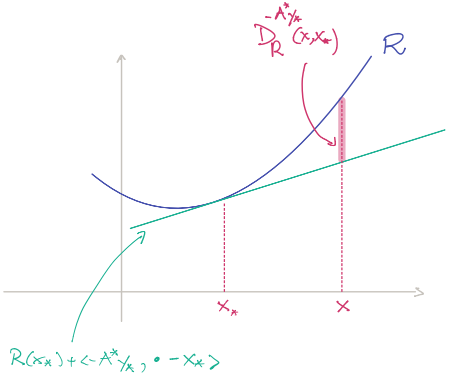
Condat-Vu algorithm

$$\begin{cases} \tilde{y}_k = 2y_k - y_{k-1} \\ x_{k+1} = \operatorname{prox}_{\tau G}(x_k - \tau(\nabla F(x_k) + A^* \tilde{y}_k)) \\ y_{k+1} = y_k + \sigma(Ax_{k+1} - b^\delta) \end{cases}$$

- weak convergence of the iterates (x_k, y_k) to a primal-dual solution (if $\tau \leq (L + \sigma \|A\|^2)^{-1}$)
- preconditioning
- inexact computations of prox
- in our analysis: view b^δ as another source of errors

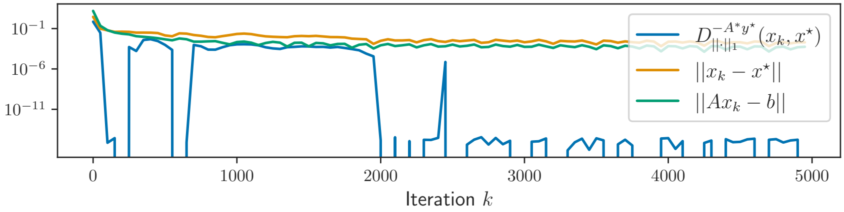
How to measure distance from optimality?

$$
\begin{aligned}
\mathcal{L}(x, y_*) - \mathcal{L}(x_*, y) &= R(x) - R(x_*) + \langle y_*, Ax - b\rangle - \cancel{\langle y, Ax_* - b\rangle} \\
&= R(x) - R(x_*) - \langle -A^* y_*, x - x_*\rangle \\
&= D_R^{-A^* y_*}(x, x_*)
\end{aligned}
$$



UniGe | MaLGa

Bregman distance is not enough for ℓ^1



Optimality condition

Theorem

If

- (x_*, y_*) *is a saddle point of* $\mathcal{L}$ *and* $(x, y) \in \mathcal{X} \times \mathcal{Y}$
- $\mathcal{L}(x, y_*) - \mathcal{L}(x_*, y) = 0$
- $Ax = b$

Then (x, y_*) *is a primal-dual solution*

Regularization properties of the Condat-Vu algorithm

Theorem (Stability and early stopping)

Let (x_k^δ, y_k^δ) be the (averaged) sequence obtained with b^δ instead of b. Assume $\tau \leq \xi(\xi L + \sigma\|A\|^2)^{-1}$ for some $0 < \xi < 1$. Then

$$\mathcal{L}(x_k^\delta, y_*) - \mathcal{L}(x_*, y_k^\delta) \leq \frac{1}{k} + \delta + \delta^2 k$$

$$\|Ax_k^\delta - b\|^2 \leq \frac{1}{k} + \delta + \delta^2 k$$

UniGe | MaLGa

Regularization properties of the Condat-Vu algorithm

Theorem (Stability and early stopping)

Let (x_k^δ, y_k^δ) be the (averaged) sequence obtained with b^δ instead of b. Assume $\tau \leq \xi(\xi L + \sigma\|A\|^2)^{-1}$ for some $0 < \xi < 1$. Then

$$\mathcal{L}(x_k^\delta, y_*) - \mathcal{L}(x_*, y_k^\delta) \leq \frac{1}{k} + \delta + \delta^2 k$$

$$\|Ax_k^\delta - b\|^2 \leq \frac{1}{k} + \delta + \delta^2 k$$

If $k \sim \delta^{-1}$, there exists $c > 0$ such that

$$\mathcal{L}(x_k^\delta, y_*) - \mathcal{L}(x_*, y_k^\delta) \leq \delta \qquad \|Ax_k^\delta - b\|^2 \leq \delta(1+\delta)$$

UniGe | MaLGa

Regularization properties of the Condat-Vu algorithm

Theorem (Stability and early stopping)

Let (x_k^δ, y_k^δ) be the (averaged) sequence obtained with b^δ instead of b. Assume $\tau \leq \xi(\xi L + \sigma\|A\|^2)^{-1}$ for some $0 < \xi < 1$. Then

$$\mathcal{L}(x_k^\delta, y_*) - \mathcal{L}(x_*, y_k^\delta) \leq \frac{1}{k} + \delta + \delta^2 k$$

$$\|Ax_k^\delta - b\|^2 \leq \frac{1}{k} + \delta + \delta^2 k$$

If $k \sim \delta^{-1}$, there exists $c > 0$ such that

$$\mathcal{L}(x_k^\delta, y_*) - \mathcal{L}(x_*, y_k^\delta) \leq \delta \qquad \|Ax_k^\delta - b\|^2 \leq \delta(1+\delta)$$

See also [Rasch-Chambolle, 2021]

Outline

Sparse recovery

Theorem

Assume that $Ax = b$ has an s-sparse solution x_*, and that the s-RIP holds. Then

$$\min \|x\|_1 \quad \text{s.t. } Ax = b$$

has a unique solution and

$$\|x_k^\delta - x_*\|^2 \leq Q'_s(\frac{1}{k} + \delta + \delta^2 k) + Q_s(\frac{1}{k} + \delta + \delta^2 k)$$

UniGe | MaLGa

Sparse recovery

Theorem

Assume that $Ax = b$ has an s-sparse solution x_*, and that the s-RIP holds. Then

$$\min \|x\|_1 \quad \text{s.t. } Ax = b$$

has a unique solution and

$$\|x_k^\delta - x_*\|^2 \leq Q_s'(\frac{1}{k} + \delta + \delta^2 k) + Q_s(\frac{1}{k} + \delta + \delta^2 k)$$

Based on [Grasmair-Scherzer-Haltmeier, Necessary and sufficient conditions for linear convergence of ℓ^1 regularization 2011]

Sparse recovery

Theorem

Assume that $Ax = b$ has an s-sparse solution x_*, and that the s-RIP holds. Then

$$\min \|x\|_1 \quad \text{s.t. } Ax = b$$

has a unique solution and

$$\|x_k^\delta - x_*\|^2 \leq Q_s'(\frac{1}{k} + \delta + \delta^2 k) + Q_s(\frac{1}{k} + \delta + \delta^2 k)$$

Based on [Grasmair-Scherzer-Haltmeier, Necessary and sufficient conditions for linear convergence of ℓ^1 regularization 2011] The constants in the bound depend on s (as for Tikhonov) and on $\|b^\delta\|$ (differently from Tikhonov).

UniGe | MaLGa

Unfeasible case: $\{Ax = b\} = \varnothing$

Remark: The (averaged) Condat-Vu sequence (x_k, y_k) satisfies $\|(x_k, y_k)\| \to +\infty$.

Unfeasible case: $\{Ax = b\} = \varnothing$

Remark: The (averaged) Condat-Vu sequence (x_k, y_k) satisfies $\|(x_k, y_k)\| \to +\infty$.
Consider

$$(P') \quad \min R(x) \quad \text{s. t. } A^*Ax = A^*b \qquad \text{vs } (P) \quad \min R(x) \quad \text{s. t. } Ax = b$$

Unfeasible case: $\{Ax = b\} = \varnothing$

Remark: The (averaged) Condat-Vu sequence (x_k, y_k) satisfies $\|(x_k, y_k)\| \to +\infty$.
Consider

$$(P') \quad \min R(x) \quad \text{s. t. } A^*Ax = A^*b \qquad \text{vs } (P) \quad \min R(x) \quad \text{s. t. } Ax = b$$

Always feasible if $\dim \mathcal{X}, \dim \mathcal{Y} < +\infty$.

Unfeasible case: $\{Ax = b\} = \varnothing$

Remark: The (averaged) Condat-Vu sequence (x_k, y_k) satisfies $\|(x_k, y_k)\| \to +\infty$. Consider

$$(P') \quad \min R(x) \quad \text{s. t. } A^*Ax = A^*b \quad \text{vs } (P) \quad \min R(x) \quad \text{s. t. } Ax = b$$

Always feasible if $\dim \mathcal{X}, \dim \mathcal{Y} < +\infty$. Assume that (P') has a primal-dual solution (x^*, y^*).

Unfeasible case: $\{Ax = b\} = \varnothing$

Remark: The (averaged) Condat-Vu sequence (x_k, y_k) satisfies $\|(x_k, y_k)\| \to +\infty$. Consider

$$(P') \quad \min R(x) \quad \text{s. t. } A^*Ax = A^*b \qquad \text{vs } (P) \quad \min R(x) \quad \text{s. t. } Ax = b$$

Always feasible if $\dim \mathcal{X}, \dim \mathcal{Y} < +\infty$. Assume that (P') has a primal-dual solution (x^*, y^*).

Theorem

Consider the "original" averaged Condat-Vu algorithm for $\delta = 0$. Then x_k weakly converges to some solution of (P'). If $\delta > 0$ and $A^*Ax = A^*b^\delta$ has a solution, then

$$D^{-A^*Ay^*}(x_k^\delta, x^*) \leq \frac{1}{k} + \delta + \delta^2 k$$

and

$$\|A^*Ax_k^\delta - A^*b\| \leq \frac{1}{k} + \delta + \delta^2 k + \delta^2$$

Outline

Sparse recovery

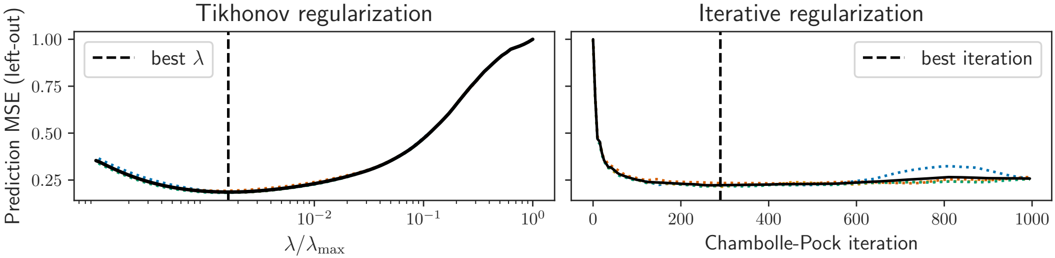


Figure: Comparison of Tikhonov regularization and iterative regularization on rcv1 (LIBSVM package). Both methods reach similar lowest prediction errors (left:0.195, right: 0.21) while the iterative approach is much faster (2.5 s vs. 125 s).

Sparse recovery

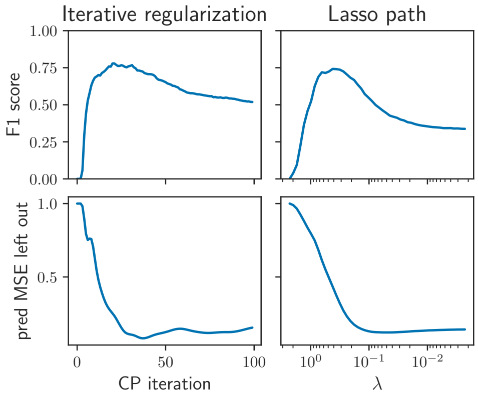


Figure: Comparison of estimation and prediction performances of iterative and Tykhonov regularization for sparse recovery. Iterative regularization attains similar performances to explicit regularization, but in few iterations.

UniGe | MaLGa

Sparse recovery

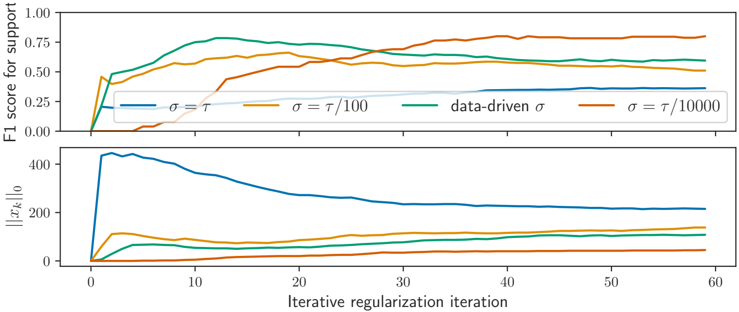


Figure: To maintain sparsity in the early iterates, it is important to set σ correctly. Our datadriven choice behaves well: the iterates sparsity increases steadily, and they reach the highest $F1$ score.

Matrix recovery

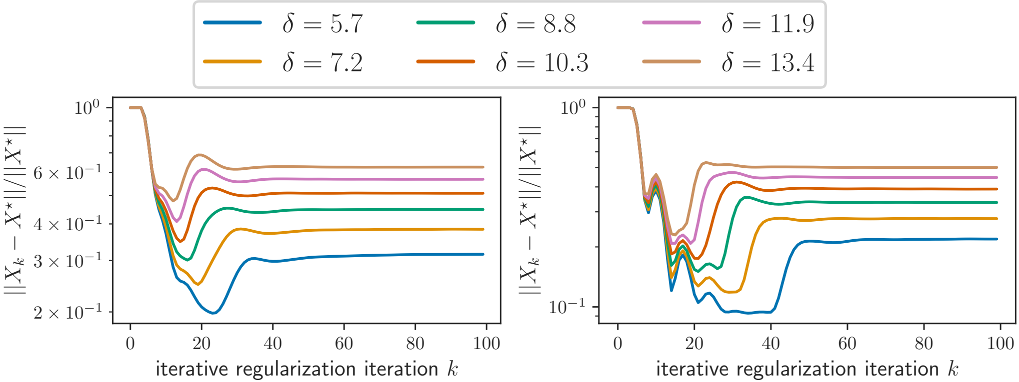


Figure: Semiconvergence of iterates for the low rank matrix completion problem, in dimension 200×200 (left) and 500×500 (right)

Conclusions and future work

- computational regularization (inverse problems)

Conclusions and future work

- computational regularization (inverse problems)
- optimization viewpoint, based on:
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for convex regularizers, PMLR 130:1684-1692, 2021
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for low complexity regularizers, arxiv 2022

Conclusions and future work

- computational regularization (inverse problems)
- optimization viewpoint, based on:
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for convex regularizers, PMLR 130:1684-1692, 2021
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for low complexity regularizers, arxiv 2022
- Code available here: https://lcsl.github.io/iterreg

Conclusions and future work

- computational regularization (inverse problems)
- optimization viewpoint, based on:
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for convex regularizers, PMLR 130:1684-1692, 2021
  M. Massias, C. Molinari, L. Rosasco, S. Villa, Iterative regularization for low complexity regularizers, arxiv 2022
- Code available here: https://lcsl.github.io/iterreg
- finite vs infinite dimensional
- unfeasible case
- stochastic variants/ learning setting