Automatic FISTA restart

Hippolyte Labarrière
Joint work with Jean-François Aujol, Charles Dossal and Aude Rondepierre

Financed by the French Agence Nationale de la Recherche (ANR) under reference ANR-PRC-CE23 MaSDOL

Institut de Mathématiques de Toulouse, INSA Toulouse, Institut de Mathématiques de Bordeaux

SIAM OP 23, Seattle, 2nd June 2023

Plan

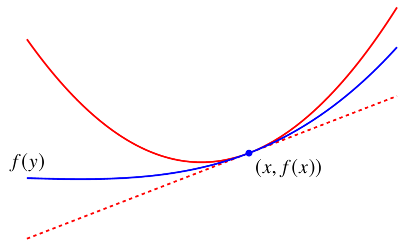
Framework

Minimization problem

$$\min_{x\in\mathbb{R}^N} F(x) = f(x) + h(x),$$

where:

- f is a convex differentiable function having a L-Lipschitz gradient,



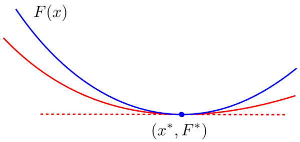
- h is a convex proper lower semicontinuous function,
- F has a non-empty set of minimizers X^*.

Framework

Assumption Q_μ:

F has a quadratic growth around its set of minimizers i.e:

$$\exists \mu > 0,\ \forall x \in \mathbb{R}^N,\ \frac{\mu}{2} d(x, X^*)^2 \leqslant F(x) - F^*.$$



Example: LASSO function:

$$F(x) = \frac{1}{2}\|Ax - y\|^2 + \lambda \|x\|_1.$$

State of the art

Forward-Backward:

$$\forall k > 0, \quad x_k = \text{prox}_{sh}(x_{k-1} - s\nabla f(x_{k-1}))$$

$$\text{Convex setting:} \quad F(x_k) - F^* = O\left(k^{-1}\right).$$

$$Q_\mu : \quad F(x_k) - F^* = O\left(e^{-\frac{\mu}{L}k}\right).$$

State of the art

Forward-Backward:

$$\forall k > 0, \quad x_k = \text{prox}_{sh}(x_{k-1} - s\nabla f(x_{k-1}))$$

$$\text{Convex setting:} \quad F(x_k) - F^* = O\left(k^{-1}\right).$$

$$Q_\mu: \quad F(x_k) - F^* = O\left(e^{-\frac{\mu}{L}k}\right).$$

Inertial methods:

$$\forall k > 0, \quad \begin{cases} x_k = \text{prox}_{sh}(y_{k-1} - s\nabla f(y_{k-1})) \\ y_k = x_k + \alpha_k(x_k - x_{k-1}) \end{cases}$$

State of the art

Inertial methods:

$$\forall k > 0, \quad \begin{cases} x_k = \text{prox}_{sh}(y_{k-1} - s\nabla f(y_{k-1})) \\ y_k = x_k + \alpha_k(x_k - x_{k-1}) \end{cases}$$

State of the art

Inertial methods:

$$\forall k > 0, \quad \begin{cases} x_k = \text{prox}_{sh}(y_{k-1} - s\nabla f(y_{k-1})) \\ y_k = x_k + \alpha_k(x_k - x_{k-1}) \end{cases}$$

FISTA (Beck and Teboulle, '09, Nesterov, '83): $\alpha_k = \frac{k-1}{k+2}$

$$\text{Convex setting:} \quad F(x_k) - F^* = O\left(k^{-2}\right).$$

$$Q_\mu : \quad F(x_k) - F^* = O\left(k^{-2}\right).$$

V-FISTA (Beck, '17, Nesterov, '03): $\alpha_k = \alpha$

Q_μ: if $\alpha = 1 - \omega\sqrt{\frac{\mu}{L}}$ and $\frac{L}{\mu} \geqslant 100$:

$$F(x_k) - F^* = O\left(e^{-K\sqrt{\frac{\mu}{L}}k}\right)$$

where $\omega = 1.46$ and $K = 0.45$.$\rightarrow$ Aujol, Dossal, L, Rondepierre, '23, forthcoming preprint.

State of the art

Restarting FISTA, why?

- to take advantage of inertia,
- to avoid oscillations.

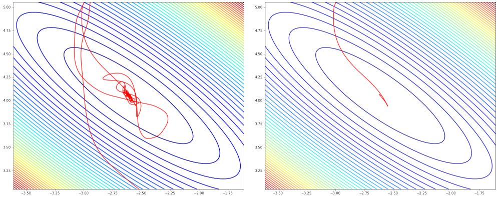


Figure: Trajectory of the iterates of FISTA (left) and FISTA restart (right) for a least-squares problem ($N = 20$).

State of the art

Restarting FISTA, how?

Algorithm 1 : FISTA restart

Require: $x_0 \in \mathbb{R}^N, y_0 = x_0 k = 0, i = 0.$
  repeat
    $k = k+1, i = i+1$
    $x_k = \text{prox}_{sh}(y_{k-1} - s\nabla f(y_{k-1}))$
    if Restart condition is *True* **then**
      $i = 1$
    end if
    $y_k = x_k + \frac{i-1}{i+2}(x_k - x_{k-1})$
  until Exit condition is *True*

$\rightarrow$ Cutting inertia is equivalent to restarting the algorithm from the last iterate.

State of the art

Objective: get a restart condition that

- does not require to know the growth parameter μ,
- ensures a fast convergence of the method: $F(x_k) - F^* = O(e^{-K\sqrt{\frac{\mu}{L}}k})$,
- is not computationnaly expensive,
- is easy to implement.

State of the art

Empiric FISTA restart (O'Donoghue and Candès, '15, Beck and Teboulle, '09)

Restart under some exit condition

- on F:

$$F(x_k) > F(x_{k-1}),$$

- on ∇F:

$$\langle \nabla F(x_k), x_k - x_{k-1} \rangle > 0.$$

State of the art

Empiric FISTA restart (O'Donoghue and Candès, '15, Beck and Teboulle, '09)

Restart under some exit condition

- on F:

$$F(x_k) > F(x_{k-1}),$$

- on ∇F:

$$\langle \nabla F(x_k), x_k - x_{k-1} \rangle > 0.$$

Fixed FISTA restart (Nesterov '13, O'Donoghue and Candès '15...)

Restart every k^* iterations where k^* is defined according to the growth parameter μ. If $k^* = \left\lfloor 2e\sqrt{\frac{L}{\mu}} \right\rfloor$:

$$F(x_k) - F^* = O\left(e^{-\frac{1}{e}\sqrt{\frac{\mu}{L}}k}\right).$$

Generalization: Scheduled restarts, Roulet and D'Aspremont '17.

State of the art

Adaptive FISTA restart

Restart according to the geometry of F and previous iterations.

- Fercoq and Qu '19: $F(x_k) - F^* = o\left(e^{-\frac{\sqrt{2}-1}{2\sqrt{e}(2-\sqrt{\frac{\mu}{\mu_0}})}\sqrt{\frac{\mu}{L}}k}\right)$.
- Alamo et al. '19: $F(x_k) - F^* = O\left(e^{-\frac{1}{16}\sqrt{\frac{\mu}{L}}k}\right)$.
- Alamo et al. '22: $F(x_k) - F^* = O\left(e^{-\frac{\ln(15)}{4e}\sqrt{\frac{\mu}{L}}k}\right)$, where $\frac{\ln(15)}{4e} \approx \frac{1}{4}$.
- Renegar and Grimmer '22: $F(x_k) - F^* = O\left(e^{-\frac{1}{2\sqrt{2}}\sqrt{\frac{\mu}{L}}k}\right)$.

Contribution

Strategy of the scheme:

- to estimate the growth parameter μ at each restart,
- to adapt the number of iterations of the following restart according to this estimation.
- to stop the algorithm when the exit condition $\|g(r_j)\| \leqslant \varepsilon$ is satisfied where:

$$g(y) = L(y - \text{prox}_{sh}(y - \frac{1}{L}\nabla f(y))).$$

Contribution

Algorithm 2 : Automatic FISTA restart

Require: $r_0 \in \mathbb{R}^N, j = 1$

$n_0 = \lfloor 2C \rfloor$

$r_1 = \text{FISTA}(r_0, n_0)$

$n_1 = \lfloor 2C \rfloor$

repeat

$j = j + 1$

$r_j = \text{FISTA}(r_{j-1}, n_{j-1})$

$$\tilde{\mu}_j = \min_{\substack{i \in \mathbb{N}^* \\ i < j}} \frac{4L}{(n_{i-1}+1)^2} \frac{F(r_{i-1}) - F(r_j)}{F(r_i) - F(r_j)}$$

Estimation of the parameter μ.

if $n_{j-1} \leqslant C\sqrt{\frac{L}{\tilde{\mu}_j}}$ **then**

$n_j = 2n_{j-1}$

Update of the number of iterations per restart.

end if

until $\|g(r_j)\| \leqslant \varepsilon$

Contribution

Theorem (Aujol, Dossal, L., Rondepierre, '21)

If F satisfies the assumptions stated before and $C > 4$, then

$$F(r_j^+) - F^* = O\left(e^{-\frac{\log\left(\frac{C^2}{4}-1\right)}{4C}\sqrt{\frac{\mu}{L}}\sum_{i=0}^{j} n_i} \right).$$

Let $C = 6.38$, then

$$F(r_j^+) - F^* = O\left(e^{-\frac{1}{12}\sqrt{\frac{\mu}{L}}\sum_{i=0}^{j} n_i} \right).$$

Numerical experiments

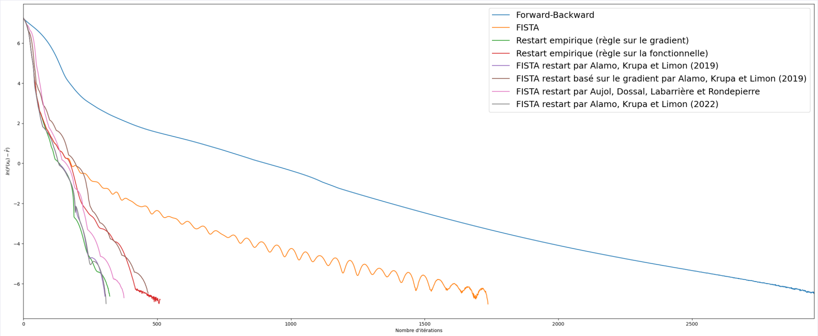
Image inpainting:

$$\min_x F(x) := \frac{1}{2}\|Mx - y\|^2 + \lambda\|Tx\|_1,$$

where M is a mask operator and T is an orthogonal transformation ensuring that Tx^0 is sparse.

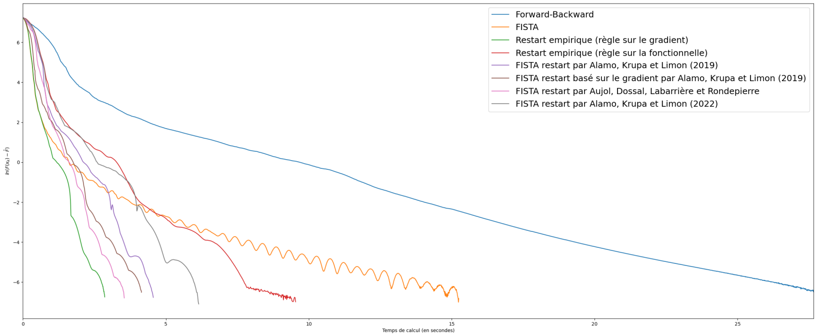
Numerical experiments

Image inpainting:



Numerical experiments

Image inpainting:



Conclusion

Summary:

Algorithm	Convergence rate
Forward-Backward	$O\left(e^{-\frac{\mu}{L}k}\right)$
V-FISTA	$O\left(e^{-\frac{9}{20}\sqrt{\frac{\mu}{L}}k}\right)$
Optimal FISTA restart	$O\left(e^{-\frac{1}{e}\sqrt{\frac{\mu}{L}}k}\right)$
Empirical FISTA restart	$O(k^{-2})$
Fercoq and Qu '19	$O\left(e^{-\frac{\sqrt{2}-1}{2\sqrt{e}(2-\sqrt{\frac{\mu}{\mu_0}})}\sqrt{\frac{\mu}{L}}k}\right)$
Alamo et al. '19	$O\left(e^{-\frac{1}{16}\sqrt{\frac{\mu}{L}}k}\right)$
Alamo et al. '22	$O\left(e^{-\frac{\ln(15)}{4e}\sqrt{\frac{\mu}{L}}k}\right)$
Renegar and Grimmer '22	$O\left(e^{-\frac{1}{2\sqrt{2}}\sqrt{\frac{\mu}{L}}k}\right)$
Automatic FISTA restart	$O\left(e^{-\frac{1}{12}\sqrt{\frac{\mu}{L}}k}\right)$

Conclusion

Perspectives:

- **Free-FISTA**: a parameter-free first order method ensuring fast convergence of the error (Aujol, Calatroni, Dossal, L, Rondepierre '23, forthcoming preprint)
  $\rightarrow$ Automatic estimation of both L and μ using restart and backtracking.

Conclusion

Perspectives:

- **Free-FISTA**: a parameter-free first order method ensuring fast convergence of the error (Aujol, Calatroni, Dossal, L, Rondepierre '23, forthcoming preprint)
  $\rightarrow$ Automatic estimation of both L and μ using restart and backtracking.
- Could we restart V-FISTA to get better decay rates without knowing μ?

Conclusion

Perspectives:

- **Free-FISTA**: a parameter-free first order method ensuring fast convergence of the error (Aujol, Calatroni, Dossal, L, Rondepierre ’23, forthcoming preprint)
  $\rightarrow$ Automatic estimation of both L and μ using restart and backtracking.
- Could we restart V-FISTA to get better decay rates without knowing μ?
- Could we combine restart with other strategies aimed at damping oscillations?
  Ex: Hessian-driven damping (Maulen, Peypouquet ’23)

Conclusion

Perspectives:

- **Free-FISTA**: a parameter-free first order method ensuring fast convergence of the error (Aujol, Calatroni, Dossal, L, Rondepierre '23, forthcoming preprint)
  $\rightarrow$ Automatic estimation of both L and μ using restart and backtracking.
- Could we restart V-FISTA to get better decay rates without knowing μ?
- Could we combine restart with other strategies aimed at damping oscillations?
  Ex: Hessian-driven damping (Maulen, Peypouquet '23)

Preprint:
Jean-François Aujol, Charles Dossal, Hippolyte Labarrière, Aude Rondepierre. FISTA restart using an automatic estimation of the growth parameter. 2021. ⟨hal-03153525v4⟩

Website:
`https://www.math.univ-toulouse.fr/~hlabarri/`

Thank you for your attention!

References I

T. Alamo, P. Krupa, and D. Limon.
Gradient based restart fista.
In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3936–3941. IEEE, 2019.

T. Alamo, P. Krupa, and D. Limon.
Restart of accelerated first order methods with linear convergence under a quadratic functional growth condition.
IEEE Transactions on Automatic Control, 2022.

T. Alamo, D. Limon, and P. Krupa.
Restart FISTA with global linear convergence.
pages 1969–1974, 2019.

A. Beck.
First-order methods in optimization.
SIAM, 2017.

A. Beck and M. Teboulle.
A fast iterative shrinkage-thresholding algorithm for linear inverse problems.
SIAM journal on imaging sciences, 2(1):183–202, 2009.

O. Fercoq and Z. Qu.
Adaptive restart of accelerated gradient methods under local quadratic growth condition.
IMA Journal of Numerical Analysis, 39(4):2069–2095, 2019.

References II

J. J. Maulen and J. Peypouquet.
A speed restart scheme for a dynamics with hessian driven damping.
arXiv preprint arXiv:2301.12240, 2023.

I. Necoara, Y. Nesterov, and F. Glineur.
Linear convergence of first order methods for non-strongly convex optimization.
Mathematical Programming, 175(1):69–107, 2019.

Y. Nesterov.
A method of solving a convex programming problem with convergence rate $o(1/k^2)$.
In *Sov. Math. Dokl*, volume 27, 1983.

B. O'donoghue and E. Candes.
Adaptive restart for accelerated gradient schemes.
Foundations of computational mathematics, 15(3):715–732, 2015.

J. Renegar and B. Grimmer.
A simple nearly optimal restart scheme for speeding up first-order methods.
Foundations of Computational Mathematics, 22(1):211–256, Feb 2022.