Aalto University
School of Electrical Engineering

A robust FISTA-like algorithm

Mihai I. Florea and Sergiy A. Vorobyov
Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

Motivation

- **Scope:** all optimization problems with composite objective functions
- **Problem** with line-search FISTA [1]: Lipschitz constant estimate can only **increase** while the algorithm is running
- FISTA converges slowly when
  - L_0, the initial Lipschitz constant estimate exceeds the actual value
  - the local curvature of the smooth part of the objective is large near the starting point but decreases around the optimum
- We formulate a FISTA-like method with **fully adaptive** line-search

Idea

- Nesterov's FGM [2] is derived using the **estimate sequence** framework
- **Augmented estimate sequence** [3]
  - can construct z-FISTA, equivalent to FISTA
  - endows z-FISTA with full adaptive line-search
  - generalizes to strong convexity [3]

Results

Robust line-search FISTA

Algorithm 1 A robust FISTA-like algorithm

$\mathbf{z}_0 = \mathbf{x}_0$, $T_0 = 0$
for k = 0,...,K-1 **do**
  $\hat{L} := \gamma_d L_k$
  loop
    $\hat{t} := \frac{1+\sqrt{1+4\hat{L}T_k}}{2\hat{L}}$ ▷ Step size validation
    $\hat{T} := T_k + \hat{t}$
    $\hat{\mathbf{y}} := \frac{1}{\hat{T}}(T_k \mathbf{x}_k + \hat{t}\mathbf{z}_k)$
    $\hat{\mathbf{x}} := p_{\hat{L}}(\hat{\mathbf{y}})$
    if $f(\hat{\mathbf{x}}) \le U_{\hat{L},\hat{\mathbf{y}}}(\hat{\mathbf{x}})$ **then** stop search
    else $\hat{L} := \gamma_u \hat{L}$
  end loop
  $\mathbf{z}_{k+1} = \mathbf{z}_k + \hat{t}\hat{L}(\hat{\mathbf{x}} - \hat{\mathbf{y}})$ ▷ State update
  $L_{k+1} := \hat{L}$, $\mathbf{x}_{k+1} := \hat{\mathbf{x}}$, $T_{k+1} := \hat{T}$
end for

Type	Symbol	Domain	Description	z-FISTA
Input	$\mathbf{x}_0$	$\mathbb{R}^n$	initial estimate of $\mathbf{x}^*$	same
Input	L_0	$(0,\infty)$	initial estimate of L_f	same
Input	γ_u	$(1,\infty)$	increase rate of $\hat{L}$	same
Input	γ_d	$(0,1)$	decrease rate of $\hat{L}$	none
Internal	$\hat{L}$	$(0,\infty)$	estimate of L_f	same
Internal	$\hat{\mathbf{x}}$	Q	estimate of $\mathbf{x}_{k+1}$	same
Internal	$\hat{\mathbf{y}}$	$\mathbb{R}^n$	estimate of $\mathbf{y}_{k+1}$	$\mathbf{y}_{k+1}$
Internal	$\hat{t}$	$(0,\infty)$	weight of $\mathbf{z}_{k+1}$	t_{k+1} / L_{k+1}
Internal	$\hat{T}$	$(0,\infty)$	estimate of T_{k+1}	T_{k+1} / L_{k+1}
Output	$\mathbf{x}_K$	Q	final estimate of $\mathbf{x}^*$	same

Convergence analysis

- Provable convergence rate is automatically updated as the algorithm is running

$$F(\mathbf{x}_k) - F^* \le \frac{1}{2T_k}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \qquad \forall k \ge 1 \tag{1}$$

- Worst case rate is same as FISTA

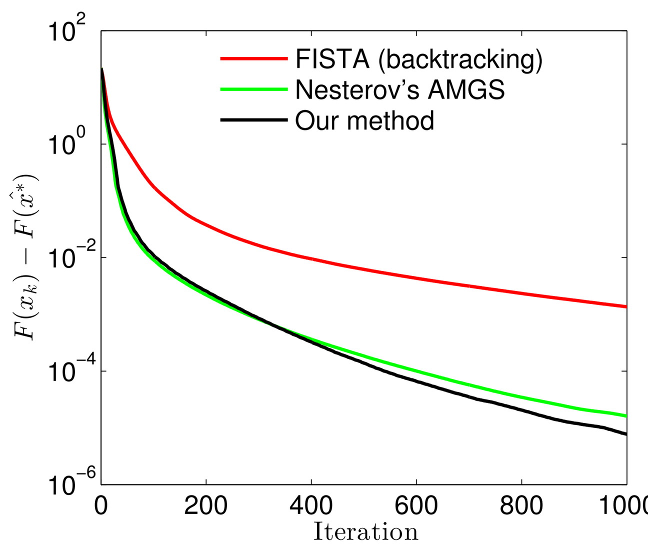
$$F(\mathbf{x}_k) - F^* \le \frac{2\gamma_u L_f}{(k+1)^2}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \qquad \forall k \ge 1 \tag{2}$$

- The rate T_k is better than the corresponding value t_k^2 / L_k in FISTA

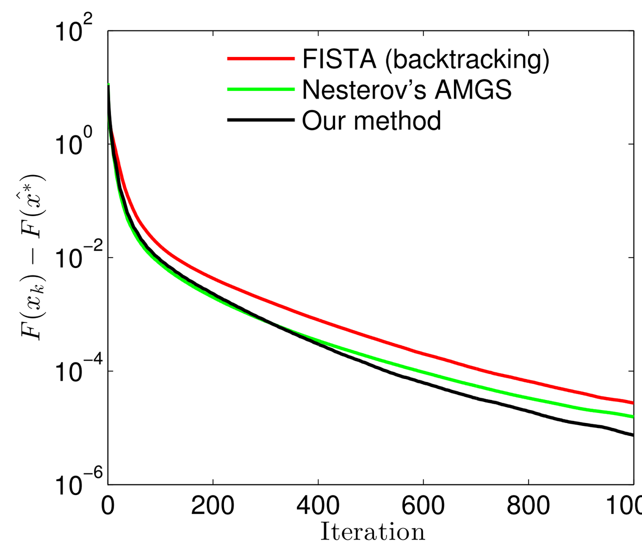
Complexity analysis

Iteration complexity (oracle function calls)	FISTA f	FISTA ∇f	FISTA $\text{prox}_{\tau\Psi}$	Nesterov's AMGS f	Nesterov's AMGS ∇f	Nesterov's AMGS $\text{prox}_{\tau\Psi}$	Our method f	Our method ∇f	Our method $\text{prox}_{\tau\Psi}$
Step size validation	2	1	1	0	2	1	2	1	1
Backtrack	1	0	1	0	2	1	2	1	1
State update	0	0	0	0	0	1	0	0	0
Iteration without backtrack	2	1	1	0	2	2	2	1	1

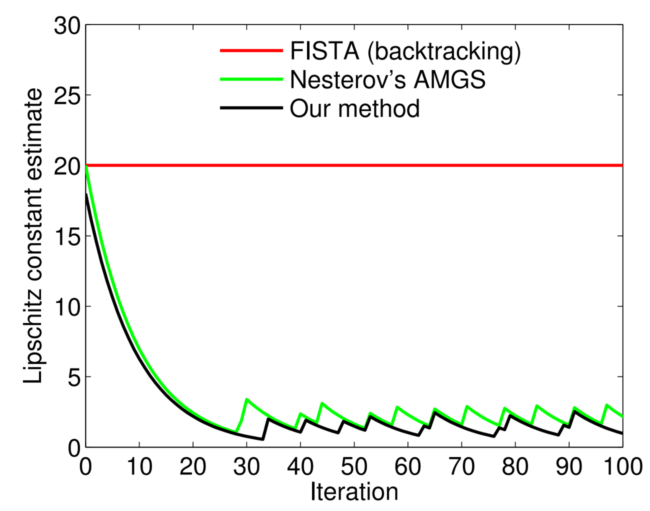
Numerical analysis



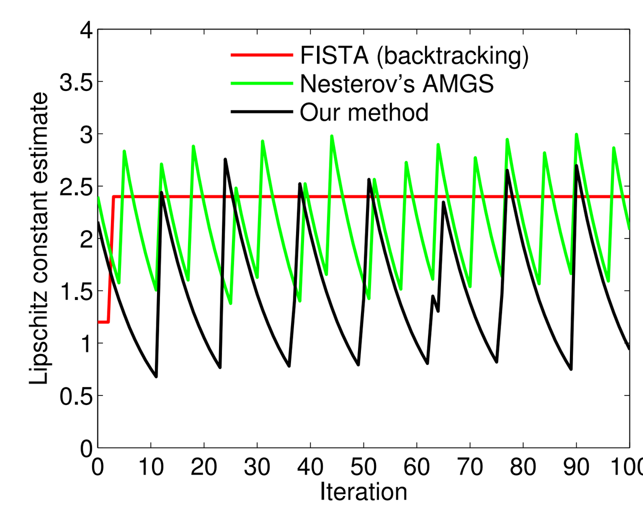
Convergence rate ($L_0 = 10L_f$)



Convergence rate ($L_0 = 0.3L_f$)



Lipschitz constant estimate ($L_0 = 10L_f$)



Lipschitz constant estimate ($L_0 = 0.3L_f$)

Conclusions

- Backtracks rarely occur
- Cheaper iterations than Nesterov's AMGS [4]
- Faster than **both** FISTA and Nesterov's AMGS
- **Extents to strongly convex case**
  - particular case of the **Accelerate Composite Gradient Method [3]**

References

[1] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[2] Y. Nesterov, *Introductory lectures on convex optimization. Applied optimization, vol. 87.* Kluwer Academic Publishers, Boston, 2004.

[3] **M. I. Florea and S. A. Vorobyov, "An accelerated composite gradient method for large-scale composite objective problems,"** ***arXiv preprint*** **arXiv:1612.02352 [math.OC], Dec. 2016.**

[4] Y. Nesterov, "Gradient methods for minimizing composite objective function," CORE, Université Catholique de Louvain, Tech. Rep. 76, Sep. 2007.

Contact

Feel free to contact us with questions and comments at:

Mihai I. Florea mihai.florea@aalto.fi
Sergiy A. Vorobyov sergiy.vorobyov@aalto.fi

Department of Signal Processing and Acoustics
School of Electrical Engineering
Aalto University, Espoo, Finland