Convergence of the iterates and rates for an inertial forward-backward descent method

Antonin Chambolle, Charles Dossal

CMAP, Ecole Polytechnique, CNRS, Palaiseau
Institut de Mathématiques de Bordeaux, Univ. Bordeaux 1
France

Outline

- Setting;
- Inertial Forward-Backward splitting: a variation on a theme of Alvarez-Attouch, 99 (Oliny-Moudafi, 03, Pock-Lorenz, 14);
- Conditions for global convergence;
- Rates and convergence for accelerated descent algorithms.

Setting

Minimize $F(x) = f(x) + g(x)$, $x \in X$ Hilbert space, with ∇f L-Lipschitz and g "simple", meaning that

$$\text{prox}_{\tau g}(x) := \arg\min_x g(y) + \frac{1}{2\tau}\|x - y\|^2$$

can be evaluated. Method based on Forward-Backward splitting

$$x \mapsto Tx := (I + \tau\partial g)^{-1}(I - \tau\nabla f)(x)$$

for $\tau \leq 1/L$. (Or $2/L$.)

Inertial Method

- $x_0 = y_0$;
- $x_n = T(y_{n-1}),\ n \geq 1$;
- $y_n = x_n + \alpha_n(x_n - x_{n-1}),\ n \geq 1$,

where $\alpha_n \in [0, 1]$ will be specified.

Convergence analysis

• This is a bit technical, we sketch the arguments (from Alvarez-Attouch, originally)

• First, if $\Phi_n = \frac{1}{2}\|x_n - x^*\|^2$, where x^* is a solution, then a computation shows that

$$\Phi_n - \Phi_{n+1} = \frac{1}{2}\|x_n - x_{n+1}\|^2 + \langle y_n - x_{n+1}, x_{n+1} - x^* \rangle - \alpha_n \langle x_n - x_{n-1}, x_{n+1} - x^* \rangle \quad (1)$$

Then, using that $-\nabla f(x^*) \in \partial g(x^*)$ and $y_n - x_{n+1} - \tau \nabla f(y_n) \in \tau \partial g(x_{n+1})$, and the monotonicity of ∂g, we find

$$\langle y_n - x_{n+1}, x_{n+1} - x^* \rangle + \tau \langle \nabla f(x^*) - \nabla f(y_n), x_{n+1} - x^* \rangle \geq 0$$

which is combined with the previous to give

$$\Phi_n - \Phi_{n+1} \geq \delta_{n+1} + \tau \langle \nabla f(y_n) - \nabla f(x^*), x_{n+1} - x^* \rangle \\ - \alpha_n \langle x_n - x_{n-1}, x_{n+1} - x^* \rangle.$$

where $\delta_{n+1} := \|x_n - x_{n+1}\|^2/2$.

Then, the co-coercivity of ∇f yields

$$\begin{aligned} &\langle \nabla f(y_n) - \nabla f(x^*), x_{n+1} - x^* \rangle \\ &\quad \geq \frac{1}{L}\|\nabla f(y_n) - \nabla f(x^*)\|^2 + \langle \nabla f(y_n) - \nabla f(x^*), x_{n+1} - y_n \rangle \\ &\qquad\qquad \geq -\frac{L}{4}\|x_{n+1} - y_n\|^2. \end{aligned}$$

and it follows

$$\Phi_n - \Phi_{n+1} \geqslant \delta_{n+1} - \frac{\tau L}{4}\|x_{n+1} - y_n\|^2 - \alpha_n \langle x_n - x_{n-1}, x_{n+1} - x^* \rangle.$$

Then, simple computations which only use $0 \leq \alpha_n \leq 1$ eventually lead to

$$\Phi_{n+1} - \Phi_n - \alpha_n(\Phi_n - \Phi_{n-1}) \leq -\left(1 - \frac{\tau L}{2}\right) \frac{\|x_{n+1} - y_n\|^2}{2} + 2\alpha_n\delta_n$$

so that if $\tau \leq 2/L$,

$$\Phi_{n+1} - \Phi_n - \alpha_n(\Phi_n - \Phi_{n-1}) \leq 2\alpha_n\delta_n.$$

By induction, we can deduce that

$$\Phi_{n+1} - \Phi_n \leq 2\sum_{j=2}^{n}\left(\prod_{l=j}^{n}\alpha_l\right)\delta_n$$

If we denote $\beta_{j,n} = \prod_{l=j}^{n} \alpha_l$, we deduce

$$\Phi_{n+1} - \Phi_n \leq 2 \sum_{j=2}^{n} \beta_{j,n} \delta_j$$

Hence (summing from 2 to $+\infty$)

$$\lim_{n \to \infty} \Phi_n - \Phi_2 \leq \sum_{n=1}^{+\infty} \sum_{j=2}^{n} \beta_{j,n} \delta_j = 2 \sum_{j=2}^{+\infty} \delta_j \sum_{n=j}^{+\infty} \beta_{j,n}$$

(where we recall $\Phi_n = \|x_n - x^*\|^2/2$). In fact, what we really can show is that Φ_n converges as soon as

$$\sum_{j=2}^{+\infty} \delta_j \sum_{n=j}^{+\infty} \beta_{j,n} < +\infty.$$

Convergence of $(x_n)_n$

One can show (cf Opial's thm) that if Φ_n converges, then also x_n converges weakly to some x^* solution of the problem (because then subsequences converge to fixed points).

A classical assumption is $\alpha_n \leq \alpha < 1$, in which case $\beta_{j,n} \leq \alpha^{n-j}$, $\sum_{n \geq j} \beta_{j,n} \leq 1/(1-\alpha)$, and the condition for convergence boils down to $\sum_{j \geq 2} \delta_j < +\infty$, which needs to be shown separately (for instance, it is true if $\alpha \leq 1/3$, cf Lorenz-Pock 2014).

We want however to consider more general cases with $\alpha_n \to 1$. Assume now that

$$\alpha_n \leq \frac{n-1}{n+2}.$$

Then (if $k-j \geq 2$)

$$\beta_{j,k} \leq \prod_{n=j}^{k} \frac{n-1}{n+2} \leq \left(\frac{j+1}{k}\right)^3$$

and one can show

$$\sum_{k \geq j} \beta_{j,k} \leq \frac{j+5}{2}.$$

Condition for convergence

In this case, we obtain the convergence of the iterates whenever

$$\sum_{j=2}^{+\infty} \delta_j \frac{j+5}{2} < +\infty,$$

that is, if

$$\sum_{j \geq 2} j \|x_{j+1} - x_j\|^2 < +\infty.$$

Remark: up to now, we only used the monotonicity of ∂g and co-coercivity of ∇f, as well as $\tau \leq 2/L$.

Descent rule for FB descent

Now, we use the fact that we are minimizing and recall the descent rule for the FB descent (cf Tseng'08, Beck-Teboulle'09, Nesterov'05). The point $T\bar{x}$ is by definition a minimizer of the strongly convex function

$$f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + g(x) + \frac{\|x - \bar{x}\|^2}{2\tau}$$

Hence it follows that for all $x \in X$ (using also the convexity of f),

$$
\begin{aligned}
& g(x) + f(x) + \frac{\|x - \bar{x}\|^2}{2\tau} \geq \\
& \qquad g(x) + f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{\|x - \bar{x}\|^2}{2\tau} \geq \\
& g(T\bar{x}) + f(\bar{x}) + \langle \nabla f(\bar{x}), T\bar{x} - \bar{x} \rangle + \frac{\|T\bar{x} - \bar{x}\|^2}{2\tau} + \frac{\|x - T\bar{x}\|^2}{2\tau} \\
& \qquad \geq g(T\bar{x}) + f(T\bar{x}) + \left(\frac{1}{\tau} - L \right) \frac{\|T\bar{x} - \bar{x}\|^2}{2} + \frac{\|x - T\bar{x}\|^2}{2\tau} .
\end{aligned}
$$

FB Descent rule

Hence if $\tau \leq 1/L$ (here we loose something),

$$g(x) + f(x) + \frac{\|x - \bar{x}\|^2}{2\tau} \geq g(T\bar{x}) + f(T\bar{x}) + \frac{\|x - T\bar{x}\|^2}{2\tau}. \quad (*)$$

for any $x, \bar{x}$.

Inertial gradient descent

Choose in this inequality $\bar{x} = y_n$, $T\bar{x} = Ty_n = x_{n+1}$, and $x = ((t_{n+1} - 1)x_n + x^*)/t_{n+1}$ where α_n and t_n are related: $\alpha_n = (t_n - 1)/t_{n+1}$. Then calculations show that

$$
\begin{aligned}
(t_{n+1}^2 - t_{n+1})(g(x_n) + f(x_n) - (g(x^*) + f(x^*)) \\
+ \frac{1}{2\tau} \left\| -t_n x_n + x^* + (t_n - 1)x_{n-1} \right\|^2 \\
\geq t_{n+1}^2 (g(x_{n+1}) + f(x_{n+1}) - (g(x^*) + f(x^*)) \\
+ \frac{1}{2\tau} \left\| (t_{n+1} - 1)x_n + x^* - t_{n+1}x_{n+1} \right\|^2
\end{aligned}
$$

So that if $t_{n+1}^2 - t_{n+1} \leq t_n^2$, one deduces the recursion:

Recursion for inertial gradient descent

$$
\begin{aligned}
t_n^2(F(x_n) - F(x^*)) &- (t_n^2 - t_{n+1}^2 + t_{n+1})(F(x_n) - F(x^*)) \\
&+ \frac{1}{2\tau}\|x_{n-1} - x^* + t_n(x_n - x_{n-1})\|^2 \\
&\geq t_{n+1}^2(F(x_{n+1}) - F(x^*)) \\
&\qquad + \frac{1}{2\tau}\|x_n - x^* + t_{n+1}(x_{n+1} - x_n)\|^2
\end{aligned}
$$

Hence (choosing $\tau = 1/L$):

$$
F(x_n) - F(x^*) \leq \frac{L}{2t_n^2}\|x_0 - x^*\|^2
$$

FISTA acceleration (Beck and Teboulle, 2008/09)

It consists in cancelling the red terms: $t_{n+1} = \sqrt{1+4t_n^2}/2$. Similar behaviour (a bit less "optimal") with $t_n = (n+1)/2$:

$$F(x_n) - F(x^*) \leq \frac{2L}{(n+1)^2}\|x_0 - x^*\|^2$$

In this case, observe that $\alpha_n = (n-1)/(n+2)$, so that as mentioned earlier, we obtain weak convergence of the iterates as soon as

$$\sum_{j\geq 2} j\|x_{j+1} - x_j\|^2 < +\infty. \qquad (**)$$

However, we do not know if $(**)$ holds.

A variant of FISTA

Assume that t_n increases a little slower, and α_n goes slower to 1: For instance, choose $t_n = 1 + (n-1)/a$, with $a > 2$, so that $\alpha_n = (n-1)/(n+a) < (n-1)/(n+2)$. Then one still has weak convergence as soon as $(**)$ holds. Observe that now

$$t_n^2 - t_{n+1}^2 + t_{n+1} = \frac{(a-2)n + (a-1)^2}{a^2}$$

so that from the recursion, we get in addition (from the red terms in the recursion) that

$$\sum_{n \geq 0} n(F(x_n) - F(x^*)) < +\infty$$

(which is still not $(**)$).

Then, consider the inequality $(*)$ again, with as before $\bar{x} = y_n$, $T\bar{x} = Ty_n = x_{n+1}$, but now $x = x_n$. It follows (with still $\tau = 1/L$)

$$F(x_n) + L\frac{\|x_n - y_n\|^2}{2} = F(x_n) + L\alpha_n^2\frac{\|x_n - x_{n-1}\|^2}{2} \geq F(x_{n+1}) + L\frac{\|x_n - x_{n+1}\|^2}{2},$$

hence, if $\alpha_n = (n-1)/(n+a)$ and recalling $\delta_n = \|x_n - x_{n-1}\|^2/2$,

$$(n+a)^2\delta_{n+1} - (n-1)^2\delta_n \leq \frac{1}{L}(n+a)^2(F(x_n) - F(x^*) - (F(x_{n+1}) - F(x^*))).$$

Summing, we deduce that

$$\begin{aligned} & a\sum_{n\geq 2}(2n+a-2)\delta_n \\ & \leq \frac{1}{L}\left((a+1)^2(F(x_1)-F(x^*))+\sum_{n\geq 2}(2n+2a-1)(F(x_n)-F(x^*))\right) \\ & \qquad < +\infty \end{aligned}$$

showing $(**)$. The (weak) convergence of the sequence $(x_n)_n$ follows.

Conclusion

- We still don't know if the iterates of the "standard" FISTA algorithm converge.
- An "infinitesimal" modification yields a convergent algorithm with roughly the same estimate:

$$F(x_n) - F(x^*) \leq \frac{La^2}{2(n+a-1)^2} \|x_0 - x^*\|^2$$

- (In practice we observe a very similar behaviour.) (Including for the iterates!)

Conclusion

- We still don't know if the iterates of the "standard" FISTA algorithm converge.
- An "infinitesimal" modification yields a convergent algorithm with roughly the same estimate:

$$F(x_n) - F(x^*) \leq \frac{La^2}{2(n+a-1)^2} \|x_0 - x^*\|^2$$

- (In practice we observe a very similar behaviour.) (Including for the iterates!)

Thank you for your attention