# Restarting accelerated gradient methods with a rough strong convexity estimate

Olivier Fercoq

Joint work with Zheng Qu

20 March 2017

# Minimisation of composite functions

Minimise the "strongly" convex composite function $F$

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

- $f : \mathbb{R}^N \to \mathbb{R}$, convex, differentiable, with $L$-Lipschitz gradient
- $\psi : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, convex, with simple proximal operator

$$\text{prox}_{\psi}(x) = \arg\min_{y \in \mathbb{R}^N} \psi(y) + \frac{1}{2}\|x - y\|_L^2$$

- $F = f + g$ features some kind of strong convexity

# The local error bound property

Let $\mathcal{X}_*$ be the set of optimal solutions such that
$\forall x_* \in \mathcal{X}_*$, $\forall x \in \mathbb{R}^n$, $F_* = F(x_*) \leq F(x)$.

## Assumption

There exists $s > 0$ and $\mu_F(s) > 0$ such that if $\mathrm{dist}_L(x, \mathcal{X}_*) \leq s$,

$$F(x) \geq F_* + \frac{\mu_F(s)}{2} \mathrm{dist}_L(x, \mathcal{X}_*)^2$$

Examples:

- $F(x) = \phi(Ax)$ with $\nabla^2 \phi(x) > 0$, $\forall x$
- $F(x) = \frac{1}{2}\|Ax - b\|^2 + \lambda \|x\|_1$

Local error bound for $s > 0 \Rightarrow$ local error bound $\forall$ compact set

# Algorithms: FISTA

Choose $x_0 \in \mathrm{dom}\,\psi$. Set $\theta_0 = 1$ and $z_0 = x_0$.

**for** $k \geq 0$ **do**

$y_k = (1-\theta_k)x_k + \theta_k z_k$

$x_{k+1} = \arg\min_{x \in \mathbb{R}^N} \left\{ \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2}\|x - y_k\|_L^2 + \psi(x) \right\}$

$z_{k+1} = z_k + \frac{1}{\theta_k}(x_{k+1} - y_k)$

$\theta_{k+1} = \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}$

**end for**

# Algorithms: APG

Choose $x_0 \in \operatorname{dom} \psi$. Set $\theta_0 = 1$ and $z_0 = x_0$.

**for** $k \geq 0$ **do**

$y_k = (1-\theta_k)x_k + \theta_k z_k$

$z_{k+1} = \arg\min_{z \in \mathbb{R}^N} \left\{ \langle \nabla f(y_k), z - y_k \rangle + \frac{\theta_k}{2} \| z - z_k \|_L^2 + \psi(z) \right\}$

$x_{k+1} = y_k + \theta_k (z_{k+1} - z_k)$

$\theta_{k+1} = \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}$

**end for**

# Algorithms: APPROX

Choose $x_0 \in \mathrm{dom}\,\psi$. Set $\theta_0 = \frac{\tau}{n}$ and $z_0 = x_0$.

**for** $k \geq 0$ **do**

$y_k = (1-\theta_k)x_k + \theta_k z_k$

Randomly generate $S_k \sim \hat{S}$

**for** $i \in S_k$ **do**

$$z^i_{k+1} = \arg\min_{z \in \mathbb{R}^{n_i}} \left\{ \langle \nabla_i f(y_k), z - y^i_k \rangle + \frac{\theta_k n v_i}{2\tau} |z - z^i_k|^2 + \psi^i(z) \right\}$$

**end for**

$x_{k+1} = y_k + \frac{n}{\tau}\theta_k (z_{k+1} - z_k)$

$\theta_{k+1} = \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}$

**end for**

# Accelerated gradient methods

| | $\mu_F = 0$ | $\mu_F > 0$ is known |
|---|---|---|
| FISTA | Beck & Teboulle | Vandenberghe |
| APG | Nesterov | Nesterov |
| dual APG | Nesterov | Nesterov |
| APPROX | Fercoq & Richtárik | Lin, Lu & Xiao |
| | $O(1/k^2)$ | $O(1-\sqrt{\mu_F})^k)$ |

The algorithms that guarantee linear convergence depend explicitly on $\mu_F$ (e.g. $\theta_k = \sqrt{\mu_F}, \forall k$)

# Restart when $\mu_F$ is known

## Proposition (Nesterov: Conditional restarting at $x_k$)

*Let $(x_k, z_k)$ be the iterates of FISTA. We have*

$$F(x_k) - F(x_*) \leq \frac{\theta_{k-1}^2}{\mu_F}(F(x_0) - F(x_*)).$$

*Moreover, given $\alpha < 1$, if*

$$k \geq 2\sqrt{\frac{1}{\alpha\mu_F}} - 1,$$

*then $F(x_k) - F(x_*) \leq \alpha(F(x_0) - F(x_*))$.*

# FISTA with restart

Choose $x_0 \in \mathrm{dom}\,\psi$. Set $\theta_0 = 1$ and $z_0 = x_0$.

**for** $k \geq 0$ **do**

$y_k = (1-\theta_k)x_k + \theta_k z_k$

$x_{k+1} = \arg\min_{x \in \mathbb{R}^N} \left\{ \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2}\|x - y_k\|_L^2 + \psi(x) \right\}$

$z_{k+1} = z_k + \frac{1}{\theta_k}(x_{k+1} - y_k)$

$\theta_{k+1} = \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}$

**if** $k \equiv 0 \mod \left\lceil 2\sqrt{\frac{1}{\alpha\mu_F}} - 1 \right\rceil$ **then**

$\theta_{k+1} = \theta_0$

$z_{k+1} = x_{k+1}$

**end if**

**end for**

Issue: the algorithm still depends on $\mu_F$

# Methods when $\mu_F$ is not known

- Dual APG with adaptive restart [Nesterov]
  1. Start with $x_0$ and an estimate $\mu$ of $\mu_F$.
  2. Perform periodic restart as if $\mu$ were smaller than $\mu_F$
  3. If the "gradient" is not small enough at the time of restart, decrease $\mu$ and go back to step 1.

  $\rightarrow$ Annoying transient phase (go back to $x_0$)

- Heuristic adaptive restart [O'Donoghue & Candes]
  - If $F(x_{k+1}) > F(x_k)$, then restart

  $\rightarrow$ Works well in practice but no guarantee

# Our goal

- Perform periodic restart with an arbitrary frequency
- Show convergence at a linear rate
- Result for FISTA, APG and APPROX

# Complexity without restart

## Proposition

*The iterates of FISTA and APG satisfy for all $k \geq 1$,*

$$\frac{1}{\theta_{k-1}^2}(F(x_k) - F_*) + \frac{1}{2}\,\mathrm{dist}_L(z_k, \mathcal{X}_*)^2 \leq \frac{1}{2}\,\mathrm{dist}_L(x_0, \mathcal{X}_*)^2$$

$$\frac{1}{2}\,\mathrm{dist}_L(x_k, \mathcal{X}_*)^2 \leq \frac{1}{2}\,\mathrm{dist}_L(x_0, \mathcal{X}_*)^2$$

→ First inequality is a direct consequence of classical results using $\mathrm{dist}_L(z_k, \mathcal{X}_*) \leq \|z_k - x_*\|_L$

→ The second is a stability result

# Unconditional restarting

## Theorem (Restarting for FISTA and APG)

*Let* $(x_k, z_k)$ *be the iterates of FISTA or APG.*
*Let* $\sigma \in [0, 1]$ *and* $\bar{x}_k = (1 - \sigma)x_k + \sigma z_k$*. We have for* $\mu_F = \mu_F(\mathrm{dist}_L(x_0, \mathcal{X}_*))$,

$$\frac{1}{2}\,\mathrm{dist}_L(\bar{x}_k, \mathcal{X}_*)^2 \leq \frac{1}{2}\max\left(\sigma, 1 - \frac{\sigma\mu_F}{\theta_{k-1}^2}\right)\mathrm{dist}_L(x_0, \mathcal{X}_*)^2$$

# Proof

$$\frac{1}{2}\,\mathrm{dist}_L(\bar{x}_k, \mathcal{X}_*)^2 \leq \frac{1-\sigma}{2}\,\mathrm{dist}_L(x_k, \mathcal{X}_*)^2 + \frac{\sigma}{2}\,\mathrm{dist}_L(z_k, \mathcal{X}_*)^2$$

definition of $\bar{x}_k = (1-\sigma)x_k + \sigma z_k$

# Proof

$$\frac{1}{2}\,\mathrm{dist}_L(\bar{x}_k,\mathcal{X}_*)^2 \le \frac{1-\sigma}{2}\,\mathrm{dist}_L(x_k,\mathcal{X}_*)^2 + \frac{\sigma}{2}\,\mathrm{dist}_L(z_k,\mathcal{X}_*)^2$$

$$= \Big(1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\mathrm{dist}(x_k,\mathcal{X}_*)^2 + \frac{\sigma}{\theta_{k-1}^2}\Big(\frac{\mu_F}{2}\,\mathrm{dist}(x_k,\mathcal{X}_*)^2 + \frac{\theta_{k-1}^2}{2}\mathrm{dist}(z_k,\mathcal{X}_*)^2\Big)$$

Rearrange

# Proof

$$\frac{1}{2}\operatorname{dist}_L(\bar{x}_k, \mathcal{X}_*)^2 \le \frac{1-\sigma}{2}\operatorname{dist}_L(x_k, \mathcal{X}_*)^2 + \frac{\sigma}{2}\operatorname{dist}_L(z_k, \mathcal{X}_*)^2$$

$$= \Big(1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\operatorname{dist}(x_k, \mathcal{X}_*)^2 + \frac{\sigma}{\theta_{k-1}^2}\Big(\frac{\mu_F}{2}\operatorname{dist}(x_k, \mathcal{X}_*)^2 + \frac{\theta_{k-1}^2}{2}\operatorname{dist}(z_k, \mathcal{X}_*)^2\Big)$$

$$\le \max\Big(0, 1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\operatorname{dist}(x_k, \mathcal{X}_*)^2 + \frac{\sigma}{\theta_{k-1}^2}\Big(F(x_k)-F_* + \frac{\theta_{k-1}^2}{2}\operatorname{dist}(z_k, \mathcal{X}_*)^2\Big)$$

$\max(0, x) \ge x$ and local error bound

# Proof

$$\frac{1}{2}\,\mathrm{dist}_L(\bar{x}_k,\mathcal{X}_*)^2 \le \frac{1-\sigma}{2}\,\mathrm{dist}_L(x_k,\mathcal{X}_*)^2 + \frac{\sigma}{2}\,\mathrm{dist}_L(z_k,\mathcal{X}_*)^2$$

$$= \Big(1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\mathrm{dist}(x_k,\mathcal{X}_*)^2 + \frac{\sigma}{\theta_{k-1}^2}\Big(\frac{\mu_F}{2}\,\mathrm{dist}(x_k,\mathcal{X}_*)^2 + \frac{\theta_{k-1}^2}{2}\mathrm{dist}(z_k,\mathcal{X}_*)^2\Big)$$

$$\le \max\Big(0, 1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\mathrm{dist}(x_k,\mathcal{X}_*)^2 + \frac{\sigma}{\theta_{k-1}^2}\Big(F(x_k)-F_* + \frac{\theta_{k-1}^2}{2}\mathrm{dist}(z_k,\mathcal{X}_*)^2\Big)$$

$$\frac{1}{2}\,\mathrm{dist}_L(\bar{x}_k,\mathcal{X}_*)^2 \le \max\Big(0, 1-\sigma-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\,\mathrm{dist}_L(x_0,\mathcal{X}_*)^2 + \frac{\sigma}{2}\,\mathrm{dist}_L(x_0,\mathcal{X}_*)^2$$

$$= \max\Big(\sigma, 1-\frac{\sigma\mu_F}{\theta_{k-1}^2}\Big)\frac{1}{2}\,\mathrm{dist}_L(x_0,\mathcal{X}_*)^2$$

Complexity of FISTA/APG + stability

# Nb iters to reach $F(x_k) - F(x_*) \leq 10^{-10}$

$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1$, $N = 4$ (iris dataset)

| $\mu_{\text{est}}$ | 1 | 0.1 | 0.01 | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-8}$ |
|---|---|---|---|---|---|---|---|---|
| Dual APG with adaptive restart | 447 | 398 | 265 | **156** | 162 | 163 | 163 | 163 |
| FISTA-$\mu$ | 751 | 352 | **170** | 173 | 264 | 291 | 277 | 277 |
| FISTA restarted: | | | | | | | | |
| at $x$, Proposition | 751 | 687 | 297 | **160** | 198 | *278* | *278* | *278* |
| at $\bar{x}$, Theorem | 633 | 274 | **168** | 211 | *278* | *278* | *278* | *278* |
| if $F(x_{k+1}) > F(x_k)$ | **121** | | | | | | | |
| APG-$\mu$ | 751 | 351 | **340** | 882 | 2580 | 7453 | >1e4 | >1e4 |
| APG restarted: | | | | | | | | |
| at $x$, Proposition | 751 | 684 | 297 | **189** | 311 | 894 | 1471 | 4488 |
| at $\bar{x}$, Theorem | 632 | 275 | **173** | 281 | 794 | 1310 | 3977 | *>1e4* |
| if $F(x_{k+1}) > F(x_k)$ | *>1e4* | | | | | | | |

751: Proximal gradient        > *1e4*: APG

# Restarting Accelerated coordinate descent

Expected separable overapproximation ($\mathbf{E}[|\hat{S}|] = \tau$):

$$\mathbf{E}[F(x + h_{[\hat{S}]})] \leq F(x_k) + \frac{\tau}{n}\Big(\langle \nabla f(x_k), h\rangle + \frac{1}{2}\|h\|_v^2\Big)$$

Choose $x_0 \in \mathrm{dom}\,\psi$. Set $\theta_0 = \frac{\tau}{n}$ and $z_0 = x_0$.

**for** $k \geq 0$ **do**

$y_k = (1-\theta_k)x_k + \theta_k z_k$

Randomly generate $S_k \sim \hat{S}$

**for** $i \in S_k$ **do**

$$z_{k+1}^i = \arg\min_{z \in \mathbb{R}^{n_i}} \Big\{\langle \nabla_i f(y_k), z - y_k^i\rangle + \frac{\theta_k n v_i}{2\tau}|z - z_k^i|^2 + \psi^i(z)\Big\}$$

**end for**

$x_{k+1} = y_k + \frac{n}{\tau}\theta_k(z_{k+1} - z_k)$

$\theta_{k+1} = \frac{\sqrt{\theta_k^4 + 4\theta_k^2} - \theta_k^2}{2}$

**end for**

# Complexity of APPROX without restart

$$\Delta(x) = \frac{1-\theta_0}{\theta_0^2}(F(x) - F_*) + \frac{1}{2\theta_0^2}\,\mathrm{dist}_v(x, \mathcal{X}_*)^2$$

## Proposition

*The iterates of APPROX satisfy for all $k \geq 1$,*

$$\mathbf{E}\Big[\frac{1}{\theta_{k-1}^2}(F(x_k) - F_*) + \frac{1}{2\theta_0^2}\,\mathrm{dist}_v(z_k, \mathcal{X}_*)^2\Big] \leq \Delta(x_0)$$

$$\mathbf{E}[\Delta(x_k)] \leq \Delta(x_0) - \sum_{i=0}^{k} \frac{\gamma_k^i}{\theta_{i-1}^2}\mathbf{E}[F(x_i) - F_*] + \frac{1-\theta_0}{\theta_0^2}\mathbf{E}[F(x_k) - F_*]$$

*where* $\gamma_k^i \geq 0$, $\sum_i \gamma_k^i = 1$ *and* $x_k = \sum_i \gamma_k^i z_i$

# Contraction result

## Notation

$$\mathring{x}_k = \frac{1}{Z}\left(\sum_{i=0}^{k} \frac{\gamma_k^i}{\theta_{i-1}^2} x_i + \left(\frac{1}{\theta_0\theta_{k-1}} - \frac{1-\theta_0}{\theta_0^2}\right) x_k\right)$$

$$m_k(\mu) = \frac{\mu\theta_0^2}{1+\mu(1-\theta_0)}\left(\sum_{i=0}^{k-1} \frac{\gamma_k^i}{\theta_{i-1}^2} + \frac{1}{\theta_0\theta_{k-1}} - \frac{1-\theta_0}{\theta_0^2}\right)$$

$$\Delta(x) = \frac{1-\theta_0}{\theta_0^2}(F(x) - F_*) + \frac{1}{2\theta_0^2}\,\mathrm{dist}_v(x, \mathcal{X}_*)^2$$

## Theorem (Restarting for APPROX)

*Let* $\sigma \in [0,1]$, $\bar{x}_k = \sigma x_k + (1-\sigma)\mathring{x}_k$. *If* $\mu_F = \mu_F(+\infty) > 0$,

$$\mathbf{E}\Big[\Delta(\bar{x}_k)\Big] \leq \max\big(\sigma, 1 - \sigma m_k(\mu_F)\big)\Delta(x_0)$$

→ Possible to deal with local error bound too

# APPROX with periodic restart

Choose $x_0 \in \mathrm{dom}\,\psi$, set $z_0 = x_0$ and $\theta_0 = \frac{\tau}{n}$.
Choose $\sigma \in (0,1)$ and $K \in \mathbb{N}$.
**for** $r \geq 0$ **do**
  $k(r) = K \times r$
  $(x_{k(r+1)}, \mathring{x}_{k(r+1)}) = \mathrm{APPROX}(\bar{x}_{k(r)}, \theta_0, K)$
  $\bar{x}_{k(r+1)} = \sigma x_{k(r)} + (1-\sigma)\mathring{x}_{k(r)}$
**end for**

## Corollary

$$\mathbf{E}\Big[\Delta(\bar{x}_{k(r)})\Big] \leq \big(\max(\sigma, 1 - \sigma m_k(\mu_F))\big)^r \Delta(x_0)$$

$$\leq \Big(\big(\max(\sigma, 1 - \sigma m_k(\mu_F))\big)^{1/K}\Big)^{k(r)} \Delta(x_0)$$

# How good is this rate?

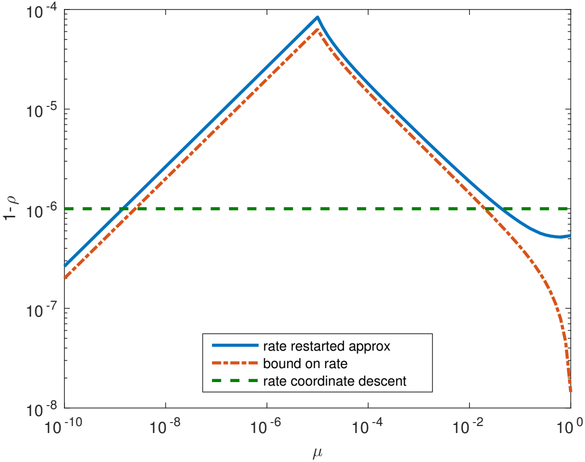
- Given an estimate $\mu_{\text{est}}$ of $\mu_F$, take $\sigma = \frac{1}{1+m_K(\mu_{\text{est}})}$.
- $m_K(\mu) \in O(\mu\theta_0^2\, K^2)$
- Take $K = \left\lceil \frac{2\sqrt{3}}{\theta_0}\sqrt{1+\frac{1}{\mu_{\text{est}}}} - \frac{2}{\theta_0} + 1 \right\rceil$

## Corollary

*If* $k \geq \frac{n}{\tau}\left(6\sqrt{6}\max\left(\frac{1}{\sqrt{\mu_{\text{est}}}}, \frac{\sqrt{\mu_{\text{est}}}}{\mu_F}\right)\log\left(\frac{\theta_0^2\Delta(x_0)}{\epsilon}\right) + \frac{4\sqrt{3}}{\sqrt{\mu_{\text{est}}}}\right)$,

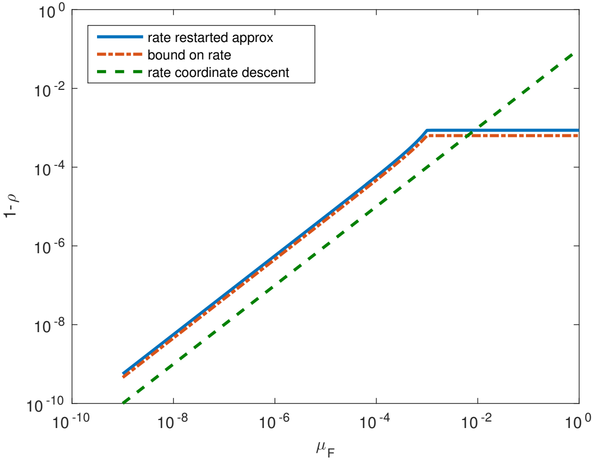
$$(1-\theta_0)(F(x_k)-F_*) + \frac{1}{2}\|x_k - x_*\|_v^2 \leq \epsilon.$$

# Rate of APPROX with periodic restart




Rate as a function of the estimate $\mu$ ($\mu_F = 10^{-5}$, $n = 10$)

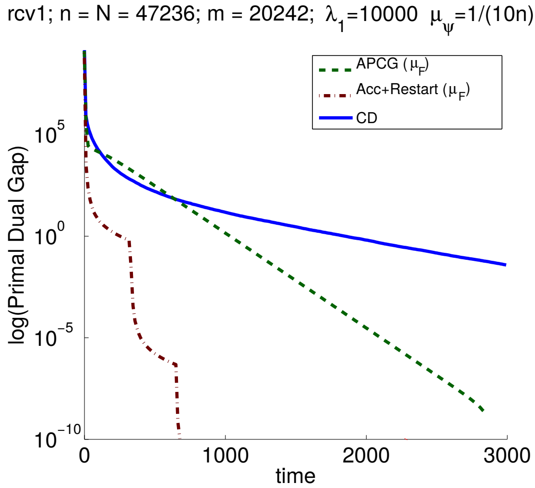
# Rate of APPROX with restart every 107n it.




Rate as a function of the true $\mu_F$ ($\mu = 10^{-3}$, $n = 10$)

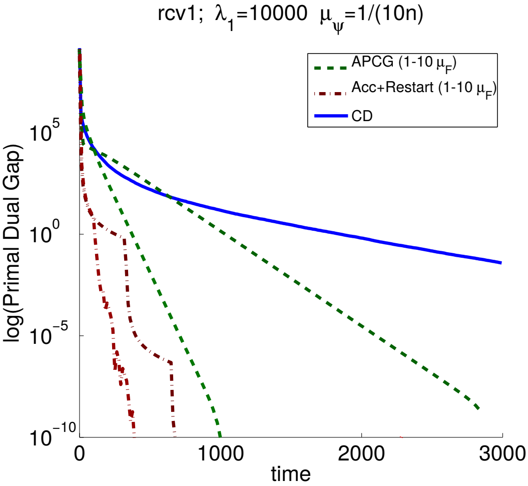
# Logistic regression problem ($\mu_{\text{est}} = \mu_\Psi$)

$$\min_{x\in\mathbb{R}^N} \frac{\lambda_1}{2\|A^\top b\|_\infty} \sum_{j=1}^m \log(1+\exp(b_j a_j^\top x)) + \|x\|_1 + \frac{\mu_\Psi}{2}\|x\|^2$$



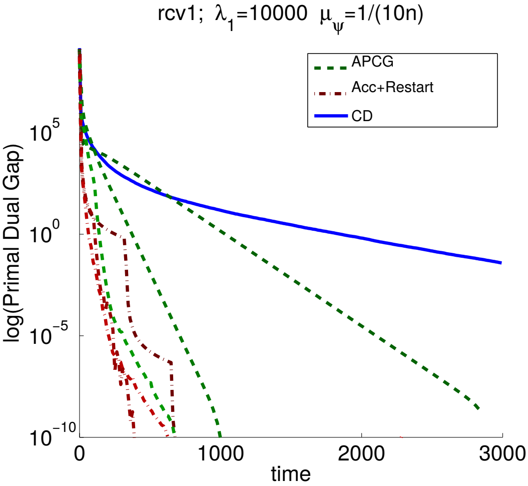
# Logistic regression problem ($\mu_{\text{est}} = 10\mu_\Psi$)

$$\min_{x\in\mathbb{R}^N} \frac{\lambda_1}{2\|A^\top b\|_\infty} \sum_{j=1}^m \log(1+\exp(b_j a_j^\top x)) + \|x\|_1 + \frac{\mu_\Psi}{2}\|x\|^2$$



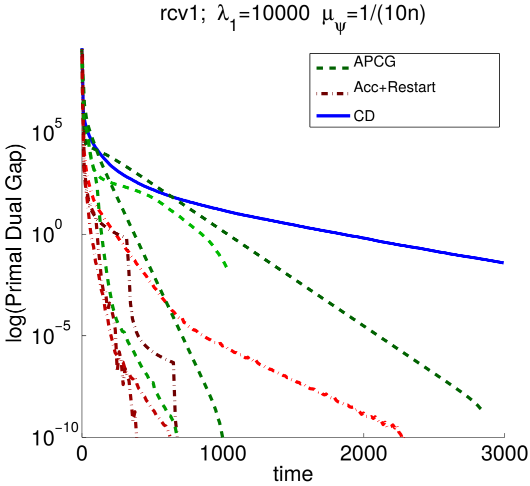
# Logistic regression problem ($\mu_{\text{est}} = 100\mu_\Psi$)

$$\min_{x\in\mathbb{R}^N} \frac{\lambda_1}{2\|A^\top b\|_\infty} \sum_{j=1}^m \log(1+\exp(b_j a_j^\top x)) + \|x\|_1 + \frac{\mu_\Psi}{2}\|x\|^2$$



# Logistic regression problem ($\mu_{\text{est}} = 1000\mu_\Psi$)

$$\min_{x\in\mathbb{R}^N} \frac{\lambda_1}{2\|A^\top b\|_\infty} \sum_{j=1}^m \log(1+\exp(b_j a_j^\top x)) + \|x\|_1 + \frac{\mu_\Psi}{2}\|x\|^2$$



# Nesterov's adaptive restart

Here $F$ is $\mu_F$-strongly convex

## Proposition

*For $L \geq L(\nabla(f))$, define* $T_L(x) = \mathrm{prox}_{\frac{1}{L}\Psi}\left(x - \frac{1}{L}\nabla f(x)\right)$
*For all $x$, we have*

$$\frac{L}{2}\|x - T_L(x)\|^2 \leq F(x) - F_*$$
$$4L^2\|x - T_L(x)\|^2 \geq \mu_F^2\|T_L(x) - x_*\|^2$$

# Checking the estimate $\mu_{\text{est}}$

## Corollary

*If* $F(x_k) - F_* \leq \rho(\mu_{\text{est}}, \mu_F)^{k-1} \frac{L}{2} \|x_1 - x_*\|^2$, *and* $x_1 = T_L(x_0)$,

$$\|x_k - T_L(x_k)\|^2 \leq \frac{2}{L}(F(x_k) - F_*) \leq \rho(\mu_{\text{est}}, \mu_F)^{k-1} \|x_1 - x_*\|^2$$

$$\leq \rho(\mu_{\text{est}}, \mu_F)^{k-1} \|T_L(x_0) - x_*\|^2 \leq \rho(\mu_{\text{est}}, \mu_F)^{k-1} \frac{4L^2}{\mu_F^2} \|x_0 - T_L(x_0)\|^2$$

## Algorithm

Choose $\mu_{\text{est}}$. Run as if we had $\mu_{\text{est}} \leq \mu_F$.
Check if $\|x_k - T_L(x_k)\|^2 \leq \rho(\mu_{\text{est}}, \mu_{\text{est}})^{k-1} \frac{4L^2}{\mu_{\text{est}}^2} \|x_0 - T_L(x_0)\|^2$
If not: reduce $\mu_{\text{est}}$ and restart from $x_0$.

# Improvement

Denote $a_r(\mu) = \prod_{i=0}^{r} \max\left(\sigma_i, 1 - \sigma_i \frac{\mu}{\theta_{K_i}^2}\right)$

Choose $\bar{x}_0 \in \mathrm{dom}\,\Psi$ and $\mu_0 \in (0, 1)$.
**for** $r \geq 0$ **do**
Choose $K_r = \frac{4}{\sqrt{\mu_r}}$ and $\sigma_r = \frac{1}{1+\mu_r/\theta_{K_r-1}^2}$
$x_{k(r+1)}, z_{k(r+1)} = \mathrm{FISTA}(\bar{x}_{k(r)}, K_r)$
$\bar{x}_{k(r+1)} = (1 - \sigma_r)x_{k(r+1)} + \sigma_r z_{k(r+1)}$
Choose $\mu_{r+1}$ to be the largest $\mu \leq \mu_r$ such that

$$\|\bar{x}_{k(r+1)} - T_L(\bar{x}_{k(r+1)})\|^2 \leq \frac{4L^2}{\mu_{r+1}^2} a_r(\mu) \|\bar{x}_0 - T_L(\bar{x}_0)\|^2$$

**end for**

$\rightarrow$ If we detect that $\mu_r$ is too big, we decrease it and go on

# Theoretical results

- $\lim \mu_r = \mu_\infty \geq \min(\mu_0, \mu_F)$
- The number of iteration to get an $\epsilon$-solution is at most

$$O\left(\sqrt{L/\mu_\infty}\log(L/\mu_\infty) + \sqrt{L/\mu_\infty}\log(1/\epsilon)\right)$$

Open question:
Same development for randomized coordinate descent

# Conclusion

## Summary

- Linear convergence rate for accelerated gradient restarted at any frequency
- Restarted accelerated coordinate descent
- Good behaviour in practice

## Future research

- Nesterov's adaptive restart for coordinate descent
- Restart smoothed duality gap primal-dual methods