Gradient methods with FEM-Subspaces

Roland Becker (UPPA)

* Switch the loops : view from optimization

* Complexity estimates are crucial in optimization
  (Linear programming, "large-scale"-problems)

$$e_k \leq \frac{C}{1+k} \quad \text{vs.} \quad e_k \leq \frac{C}{1+k^2}$$

$$e_k \leq C_0\left(1-\frac{1}{C_1\kappa}\right)^k \quad \text{vs.} \quad e_k \leq C_0\left(1-\frac{1}{C_1\kappa^{1/2}}\right)^k \qquad \kappa : \text{condition no.}$$

<u>Ref.</u> : AFEM (Amsterdam, Berlin, Paris, Wien, ...)

Optimization (Nesterov, ...)

Motivation : Semi-linear PDEs

$$\begin{cases} -\Delta u = f(u) & \Omega \\ u = 0 & \partial\Omega \end{cases}$$

Stationary points of

$$E : X \to \mathbb{R} \qquad X = H_0^1(\Omega)$$

$$E(u) := \frac{1}{2}\int_\Omega |\nabla u|^2 - \int_\Omega F(u), \quad F' = f$$

In the simplest case, E is strictly convex, we can use steepest-descent

$$\begin{cases} u_{k+1} = u_k - t_k g_k, \quad u_0 \text{ given.} \\ \int_\Omega \nabla g_k \cdot \nabla v = \int_\Omega \nabla u_k \cdot \nabla v - \int_\Omega f(u_k) v \quad \forall v \in X \end{cases} \qquad g_k = \nabla E(u_k)$$

We want to use FEM-spaces $X_k \subset X$

$$g_k \in X_k : \quad \int \nabla g_k \cdot \nabla v = E'(u_k)(v) \quad \forall v \in X_k \qquad (\text{Riesz})$$

from now : "u" $\longrightarrow$ "x" "E" $\longrightarrow$ "f"

Gradient method

$f : X \to \mathbb{R}$ smooth strictly convex $\langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \mu \|x-y\|^2$

$$Q_t(x,y) := f(y) + \langle \nabla f(y), x-y \rangle + \frac{1}{2t}\|x-y\|^2$$

$$G_t(y) := \operatorname*{argmin}_{x \in X} Q_t(x,y) = y - t\nabla f(y)$$

$$Q_t^*(y) := \min_{x \in X} Q_t(x,y) = f(y) - \frac{1}{2t}\|G_t(y) - y\|^2 = f(y) - \frac{t}{2}\|\nabla f(y)\|^2$$

INPUT : $t_0 > 0$, $x_0 \in X$ $k = 0$

1. while $Q_t^*(x_k) < f(G_{t_k}(x_k))$: $t_k = t_k/2$ (Armijo-Goldstein)
2. $x_{k+1} = G_{t_k}(x_k)$
3. $t_k = 2t_k$, $k = k+1$

Rq: If ∇f L-Lipschitz we have $t_k \geq \frac{1}{L}$

We have: $f(x_{k+1}) \leq \min_{x \in X} Q_{t_k}(x, x_k) = Q_{t_k}(x_{k+1}, x_k) = f(x_k) - \frac{1}{2t_k}\|x_{k+1} - x_k\|^2$

Let $\Delta f_k := f(x_k) - f^*$ then $\boxed{\Delta f_{k+1} \leq \Delta f_k - \frac{1}{2t_k}\|x_{k+1} - x_k\|^2}$

If $t_k \geq t_L > 0$ (and...) this gives $x_k \to x_\infty$ & $\nabla f(x_\infty) = 0$... (DIY)

<u>We want more:</u>

suppose $x^* \in \operatorname{argmin}_{x \in X} f(x)$ $\quad (\nabla f(x^*) = 0)$

$$\left.\begin{array}{l} f(x_{k+1}) \leq f(x_k) - \frac{t_k}{2}\|\nabla f(x_k)\|^2 \\ f(x^*) \geq f(x_k) + \langle \nabla f(x_k), x^* - x_k \rangle \end{array}\right\} \Rightarrow \boxed{\underbrace{f(x_{k+1}) - f(x^*)}_{\Delta f_{k+1}} \leq \langle \nabla f(x_k), x_k - x^* \rangle - \frac{t_k}{2}\|\nabla f(x_k)\|^2}$$

$$\Delta f_{k+1} \leq \langle \nabla f(x_k), x_k - x^* \rangle - \frac{t_k}{2} \|\nabla f(x_k)\|^2$$

$$= \frac{t_k}{2} \left(2\langle \nabla f(x_k), t_k^{-1}(x_k - x^*) \rangle - \|\nabla f(x_k)\|^2 \right)$$

$$\left(2\langle a,b\rangle - \|a\|^2 = \|b\|^2 - \|b-a\|^2\right)$$

$$= \frac{t_k}{2} \left(\|t_k^{-1}(x_k - x^*)\|^2 - \|\underbrace{t_k^{-1}(x_k - x^*) - \nabla f(x_k)}_{= t_k^{-1}(x_{k+1} - x^*)}\|^2 \right)$$

suppose $\boxed{t_k \geq t_L > 0}$: $\boxed{\Delta f_{k+1} \leq \frac{1}{2t_L} \left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \right)}$

Summing up: $\sum_{k=n+1}^{\infty} \Delta f_k \leq \frac{1}{2t_L} \|x_n - x^*\|^2 \leq \frac{1}{2t_L \mu} \Delta f_n \Rightarrow \boxed{\Delta f_n \leq C \rho^n \Delta f_0}$

$\sum_{k=n+1}^{\infty} a_k \leq C a_n \Rightarrow a_{n+k} \leq (C+1)\rho^k a_n \quad \rho = \frac{C}{C+1} \Rightarrow \boxed{\rho = 1 - \frac{1}{6\kappa}}$

$\kappa = L/\mu$

if $t_L \gtrsim 1/L$

If $\boxed{t_k \leq \frac{2}{\mu + L}}$ we have a simpler proof:

$$\|x^* - x_{k+1}\|^2 = \|x^* - x_k + t_k \nabla f(x_k)\|^2 = \|x^* - x_k\|^2 + 2t_k \langle \nabla f(x_k), x^* - x_k \rangle + t_k^2 \|\nabla f(x_k)\|^2$$

$$\langle \nabla f(x_k), x^* - x_k \rangle = \langle \nabla f(x_k) - \nabla f(x^*), x^* - x_k \rangle \leq -\frac{\mu}{L+\mu}\|x^* - x_k\|^2 - \frac{1}{L+\mu}\|\nabla f(x_k)\|^2$$

$$\leq \left(1 - \frac{2t_k \mu}{L+\mu}\right)\|x^* - x_k\|^2 + t_k \underbrace{\left(t_k - \frac{2}{L+\mu}\right)}_{\leq 0}\|\nabla f(x_k)\|^2$$

$$\Rightarrow \boxed{\|x^* - x_{k+1}\|^2 \leq \rho_S \|x^* - x_k\|^2 \qquad \rho_S = 1 - \frac{1}{6K}}$$

With the same argument: $\|x_{k+1} - x_k\|^2 \leq \rho_S \|x_k - x_{k-1}\|^2$

Gradient method on sub-spaces

$$X_0 \subset \dots X_k \subset X_{k+1} \subset \dots X, \quad x_{k+1} \in X_k$$

$$\boxed{x_{k+1} = x_k - t_k P_k \nabla f(x_k)}$$

$$x_{k+1} \in X_k : \langle x_{k+1}, y \rangle = \langle x_k, y \rangle - f'(x_k)(y) \quad \forall y \in X_k$$

We still have: $\boxed{\Delta f_{k+1} \le \Delta f_k - \frac{1}{2t_k}\|x_{k+1} - x_k\|^2}$ since

$$x_{k+1} = \operatorname*{argmin}_{x \in X_k} Q_{t_k}(x, x_k) = G^k_{t_k}(x_k)$$

$$\min_{x \in X_k} Q_{t_k}(x, x_k) = f(x_k) - \frac{t_k}{2}\|P_k \nabla f(x_k)\|^2$$

$$\forall x \in X_k : \quad Q_t(x, x_k) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2t}\|x - x_k\|^2$$

$$= f(x_k) + \langle P_k \nabla f(x_k), x - x_k \rangle + \frac{1}{2t}\|x - x_k\|^2$$

BUT NOW:

(convexity) (step-length rule)

$$f(x_{k+1}) - f(x^*) \le \langle \nabla f(x_k), x^* - x_k \rangle - \frac{1}{2t_k} \| P_k \nabla f(x_k) \|^2$$

$$= \langle P_k \nabla f(x_k), x^* - x_k \rangle - \frac{1}{2t_k} \| P_k \nabla f(x_k) \|^2 + D_k$$

$$= \frac{1}{2t_k} \left(\| x^* - x_k \|^2 - \| x^* - x_{k+1} \|^2 \right) + D_k \quad \text{(as before)}$$

with $\boxed{D_k := \langle (I - P_k) \nabla f(x_k), x^* - x_k \rangle}$

First idea! if $|D_k| \le \lambda \Delta f_k$ $(0 < \lambda < 1)$ then $(t_k \ge t_L > 0)$

$$(1-\lambda) \sum_{k=n+1}^{\infty} \Delta f_k \le \left(\frac{1}{2 t_L \mu} + \lambda \right) \Delta f_n$$

$$\Rightarrow \quad \Delta f_n \le c_0 \, \rho^n \, \Delta f_0 \qquad \rho = (1 - 1/c_1 \kappa)$$

$$D_k = \langle (I - P_k) \nabla f(x_k), x^* - x_k \rangle$$

$$= \langle \nabla f(x_k), (I - P_k)(x^* - x_k) \rangle \qquad x_k^* :\in \underset{x \in X_k}{\operatorname{argmin}} f(x)$$

$$= \langle \nabla f(x_k), x^* - x_k - y \rangle \qquad y \in X_k$$

then $|D_k| \leq C_{Rd}\, \eta_k(x_k)\, \| x^* - x_k \|$

$$\leq \frac{C_{Rd}^2}{\mu} \eta_k^2(x_k) + \frac{1}{2}\left(f(x_k) - f(x^*)\right) \qquad \text{(strict convexity)}$$

So

$$\boxed{\Delta f_{k+1} \leq \frac{1}{2} \Delta f_k + \frac{1}{2t_k}\left(\|x^* - x_k\|^2 - \|x^* - x_{k+1}\|^2\right) + \frac{C_R^2}{\mu} \eta_k^2(x_k)}$$

We would need $\eta_k^2(x_k) \leq \lambda \left(f(x_k) - f(x_k^*)\right)$ (λ small)

Second idea: Use estimator reduction

If REF : $\eta^2_{k+1}(x_{k+1}) \leq q_M \, \eta^2_k(x_k) + C_S^2 \|x_{k+1} - x_k\|^2$ (Dörfler marking)

But we have : $f(x_{k+1}) - f(x_k) \leq - \frac{1}{2 t_k} \|x_{k+1} - x_k\|^2$

So $$\boxed{\eta^2_k(x_{k+1}) \leq q_M \, \eta^2_k(x_k) + \lambda t_k \big(f(x_k) - f(x_{k+1})\big)}$$

$$\lambda = 2 C_S^2$$

If no REF : Use the red inequality as refinement criterion !?

We have :

$$f(x_{k+1}) - f(x^*) \leq \frac{1}{t}\left(\|x^* - x_k\|^2 - \|x^* - x_{k+1}\|^2\right) + \frac{2C_k^2}{\mu}\, \eta_k^2(x_k)$$

$$\eta_{k+1}^2(x_{k+1}) \leq q_M\, \eta_k^2(x_k) + \lambda\, t_u \left(f(x_k) - f(x_{k+1})\right)$$

with $0 < t_L \leq t_k \leq t_u \quad \forall k$

$$\boxed{e_k := f(x_k) - f(x^*) + c_1\, \eta_k^2(x_k) + c_2 \|x_{k+1} - x_k\|^2}$$ (c_1, c_2 big enough)

$$e_{k+1} \leq \tilde{q}\, e_k + \frac{1}{2t_k}\left(\|x^* - x_k\|^2 - \|x^* - x_{k+1}\|^2\right)$$

Summing up : ...

INPUT : $t_0 > 0$, $x_0 \in X$, $\lambda > 0$, $k = 0$

1. while $\min\limits_{x \in X} Q_{t_k}(x, x_k) < f(G_{t_k}(x_k))$: $t_k = t_k/2$
2. $x_{k+1} = G_{t_k}(x_k)$
3. If $\eta_k^2(x_{k+1}) \geq \frac{q}{M}\eta_k^2(x_k) + \lambda t_k (f(x_k) - f(x_{k+1}))$: REFINE
4. $t_k = 2t_k$, $k = k+1$

<u>thm:</u> For any $\lambda > 0$, $\bar{t} \geq t_k \geq \underline{t} > 0, \dots,$ we have

$$\boxed{e_{k+n} \leq C \rho^n e_k \qquad \rho = 1 - \frac{1}{c_3 K}}$$

$$e_k := f(x_k) - f(x^*) + c_1 \eta_k^2(x_k) + c_2 \|x_{k+1} - x_k\|^2$$

Special case : linear equation

$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ with appropriate A, b.

Then $f(x_k) - f(x^*) = \frac{1}{2}\|x^* - x_k\|^2$ etc. ...

Complexity estimate

We want $\sum_{k=1}^{n} \dim X_k \leq C\, \varepsilon^{-1/s}$ if $e_n \leq \varepsilon$

and "s" : approximation speed $\sup_{N \in \mathbb{N}} e_N^* N^s < +\infty$

$$e_N^* = \min \left\{ e(x_Y^*) \,,\ x^* \in \underset{x \in Y}{\operatorname{argmin}} f(x) \;\middle|\; \dim Y \leq N,\ Y \text{ admissible} \right\}$$

...

Accelerated gradient method (Nesterov)

$(X_k = X)$

GM: $x_{k+1} = \operatorname{argmin}_{x \in X} Q_{t_k}(x, x_k)$ AGM: $x_{k+1} = \operatorname{argmin}_{x \in X} Q_{t_k}(x, y_k)$

(SL-rule) (convexity)
$$\begin{cases} f(x_{k+1}) \leq Q_{t_k}(x_{k+1}, y_k) = f(y_k) - \frac{t_k}{2}\|\nabla f(y_k)\|^2 \\ f(x) \geq f(y_k) + \langle \nabla f(y_k), x - y_k \rangle \quad \forall x \in X \end{cases}$$

$$\Rightarrow \quad f(x_{k+1}) - f(x) \leq \langle \nabla f(y_k), y_k - x \rangle - \frac{t_k}{2}\|\nabla f(y_k)\|^2$$

$$\Rightarrow \begin{cases} \Delta f_{k+1} \leq \langle \nabla f(y_k), y_k - x^* \rangle - \frac{t_k}{2}\|\nabla f(y_k)\|^2 & (x = x^*) \\ \Delta f_{k+1} - \Delta f_k \leq \langle \nabla f(y_k), y_k - x_k \rangle - \frac{t_k}{2}\|\nabla f(y_k)\|^2 & (x = x_k) \end{cases}$$

$$\Rightarrow \quad \Delta f(x_{k+1}) - \alpha \Delta f_k \leq \langle \nabla f(y_k), y_k - \alpha x_k - (1-\alpha) x^* \rangle - \frac{t_k}{2}\|\nabla f(y_k)\|^2$$

$(0 < \alpha < 1)$

$$\Delta f(x_{k+1}) - \alpha \Delta f_k \leq \langle \nabla f(y_k), y_k - \alpha x_k - (1-\alpha)x^* \rangle - \frac{t_k}{2}\|\nabla f(y_k)\|^2$$

$$= \frac{1}{2t_k}\left(2\langle t_k \nabla f(y_k), y_k - \alpha x_k - (1-\alpha)x^* \rangle - \|t_k \nabla f(y_k)\|^2 \right)$$

$$2ab - a^2 = b^2 - (b-a)^2$$

$$= \frac{1}{2t_k}\left(\|y_k - \alpha x_k - (1-\alpha)x^*\|^2 - \|\underbrace{y_k - \alpha x_k - t_k \nabla f(y_k)}_{x_{k+1} - \alpha x_k} - (1-\alpha)x^*\|^2 \right)$$

So we want $x_{k+1} - \alpha x_k = y_{k+1} - \alpha x_{k+1}$

or $y_{k+1} = x_{k+1} + \alpha(x_{k+1} - x_k)$ let $\tilde{x}_k := x_k + \frac{\alpha}{1-\alpha}(x_k - x_{k-1})$

then $$\boxed{\Delta f(x_{k+1}) \leq \alpha \Delta f(x_k) + \frac{(1-\alpha)^2}{2t_k}\left(\|\tilde{x}_k - x^*\|^2 - \|\tilde{x}_{k+1} - x^*\|^2 \right)}$$

Suppose $t_k = \frac{1}{L}$ & $\|\tilde{x}_k - x^*\|^2 \leq \frac{1}{\mu} \Delta f_k$ (?)

then $$\sum_{k=n+1}^{\infty} \Delta f_k \leq \alpha \sum_{k=n}^{\infty} \Delta f_k + \frac{(1-\alpha)^2}{2} K \, \Delta f_n$$

$$\Rightarrow \sum_{k=n+1}^{\infty} \Delta f_k \leq \underbrace{\left(\frac{\alpha}{1-\alpha} + \frac{1-\alpha}{2} K \right)}_{\varphi(\alpha) \text{ strictly convex}} \Delta f_n \qquad \varphi(\alpha_*) \leq 2 K^{1/2}$$

so we end up with

$$\boxed{\Delta f_n \leq c_0 \rho^n \Delta f_0 \qquad \rho = 1 - \frac{1}{c_1 \sqrt{K}}}$$

AGM on FEM-subspaces:

In order to deal with $X_k \subset X$ we have to think about:

$$\langle (I - P_k)\nabla f(y_k), y_k - x^* \rangle = \langle \nabla f(y_k), x_k^* - x^* - y \rangle \quad y \in X_k$$

$\Rightarrow$ Evaluate the estimator at y_k! ...

THANK YOU!