Universal Line-Search Method

Sergey Guminov, MIPT
Alexander Gasnikov, MIPT

June 28, 2018

Problem setting

Consider the global minimization problem

$$f(x) \to \min_{x \in \mathbb{R}^n},$$

where f is convex, differentiable and has Hölder continuous gradient:

$$\|\nabla f(x) - \nabla f(y)\| \leqslant M_\nu \|x - y\|^\nu \quad \forall\, x, y \in \mathbb{R}^n.$$

It is assumed that a solution exists. It is denoted

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^n} f(x).$$

Key Lemma

Lemma

Let function $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and have Hölder continuous gradient. Then for any $\delta > 0$ we have

$$f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{M}{2}\|y - x\|^2 + \frac{\delta}{2},$$

where

$$M = M(\delta, \nu, M_\nu) = \left[\frac{1-\nu}{1+\nu}\frac{M_\nu}{\delta}\right]^{\frac{1-\nu}{1+\nu}} M_\nu.$$

Universal Fast Gradient Method

Algorithm 1: UFGM(f, L_0, x_0, ε, T)

$y_0 \leftarrow x_0$, $z_0 \leftarrow x_0$, $\alpha_0 \leftarrow 0$, $\psi_0(x) \leftarrow V_{x_0}(x)$
for $k = 0$ to $T - 1$ do
  $L_{k+1} \leftarrow \frac{L_k}{2}$
  while True do
    $v_k = \operatorname*{argmin}_{x \in \mathbb{R}^n} \psi_k(x)$
    $\alpha_{k+1} \leftarrow \frac{1}{2L_{k+1}} + \sqrt{\frac{1}{4L_{k+1}^2} + \alpha_k^2 \frac{L_k}{L_{k+1}}}$
    $\tau_k \leftarrow \frac{1}{\alpha_{k+1} L_{k+1}}$
    $x_{k+1} \leftarrow \tau_k v_k + (1 - \tau_k) y_k$
    $z_{k+1} \leftarrow \operatorname*{argmin}_{z \in \mathbb{R}^n} \alpha_{k+1} \langle \nabla f(x_{k+1}), z - v_k \rangle + V_{v_k}(z)$
    $y_{k+1} \leftarrow \tau_k z_{k+1} + (1 - \tau_k) y_k$
    if $f(y_{k+1}) \leqslant f(x_{k+1}) + \langle \nabla f(x_{k+1}), y_{k+1} - x_{k+1} \rangle + \frac{L_{k+1}}{2} \|y_{k+1} - x_{k+1}\|^2 + \frac{\tau_k \varepsilon}{2}$ then
      break
    end
    else
      $L_{k+1} \leftarrow 2L_{k+1}$
    end
  end
  $\psi_{k+1}(x) \leftarrow \psi_k(x) + \alpha_{k+1} [f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle]$
end
return y_T

Theorem

Let f be a differentiable convex function with Hölder continuous gradient with some exponent ν and $M_\nu < \infty$. Let $L_0 \leqslant M(\varepsilon, \nu, M_\nu)$. Then

$$f(y_k) - f(x^*) \leq \left[\frac{2^{2+4\nu} M_\nu^2}{\varepsilon^{1-\nu} k^{1+3\nu}}\right]^{\frac{1}{1+\nu}} + \frac{\varepsilon}{2}.$$

What follows is that one may obtain an ε-accurate solution in

$$k \leqslant \inf_{\nu \in [0,1]} \left[2^{\frac{2+4\nu}{1+3\nu}} \left(\frac{M_\nu}{\varepsilon}\right)^{\frac{2}{1+3\nu}} R^{\frac{2+2\nu}{1+3\nu}} \right]$$

iterations.

Algorithm 2: Linear Coupling(f, L, x_0, T)

$y_0 \leftarrow x_0$, $z_0 \leftarrow x_0$, $\alpha_0 \leftarrow 0$

for $k = 0$ **to** $T - 1$ **do**

$\quad \alpha_{k+1} \leftarrow \frac{1}{2L} + \sqrt{\frac{1}{4L^2} + \alpha_k^2}$

$\quad \tau_k \leftarrow \frac{1}{\alpha_{k+1} L}$

$\quad x_{k+1} \leftarrow \tau_k z_k + (1 - \tau_k) y_k$

$\quad y_{k+1} \leftarrow x_{k+1} - \frac{1}{L} \nabla f(x_{k+1})$

$\quad z_{k+1} \leftarrow z_k - \alpha_{k+1} \nabla f(x_{k+1})$

end

return y_T

Algorithm 3: ULCM(f, L_0, x_0, ε, T)

$y_0 \leftarrow x_0$, $z_0 \leftarrow x_0$, $\alpha_0 \leftarrow 0$
for $k = 0$ to $T - 1$ do
- $L_{k+1} \leftarrow \frac{L_k}{2}$
- while True do
  - $\alpha_{k+1} \leftarrow \frac{1}{2L_{k+1}} + \sqrt{\frac{1}{4L_{k+1}^2} + \alpha_k^2 \frac{L_k}{L_{k+1}}}$
  - $\tau_k \leftarrow \frac{1}{\alpha_{k+1} L_{k+1}}$
  - $x_{k+1} \leftarrow \tau_k z_k + (1 - \tau_k) y_k$
  - $h_{k+1} \leftarrow \operatorname*{argmin}_{h \geqslant 0} f(x_{k+1} - h \nabla f(x_{k+1}))$
  - $y_{k+1} \leftarrow x_{k+1} - h_{k+1} \nabla f(x_{k+1})$
  - $z_{k+1} \leftarrow z_k - \alpha_{k+1} \nabla f(x_{k+1})$
  - if $\langle \alpha_{k+1} \nabla f(x_{k+1}), z_k - z_{k+1} \rangle - \frac{1}{2} \|z_k - z_{k+1}\|^2 \leq \alpha_{k+1}^2 L_{k+1} (f(x_{k+1}) - f(y_{k+1}) + \frac{\tau_k \varepsilon}{2})$ then
    - break
  - end
  - else
    - $L_{k+1} \leftarrow 2L_{k+1}$
  - end
- end

end
return y_T

Stopping criterion

The condition

$$f(y_m) - \hat{f}_m \leq \varepsilon,$$

where

$$\hat{f}_m = \min_{u:\ \|z_0 - u\| \leq R} \frac{1}{\alpha_m L_m^2} \sum_{i=1}^{m} \left[\alpha_i \left(f(x_i) + \langle \nabla f(x_i), u - x_i \rangle\right)\right],$$

is a calculable stopping criterion.

Numerical Experiments

$$f(x) = \sum_{i=1}^{n} i x_i^2.$$

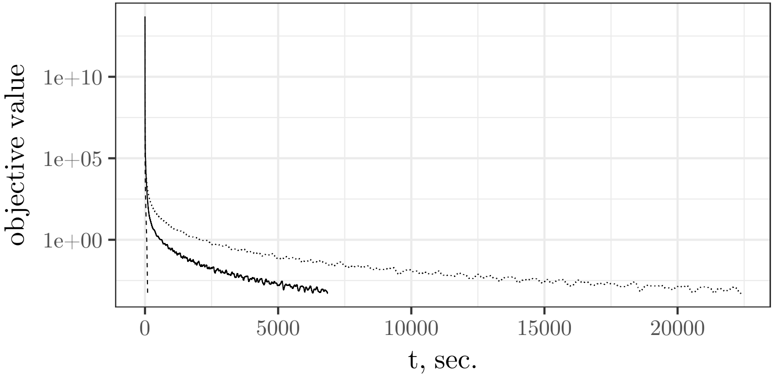

Figure 1: Methods convergence for the smooth problems with $n = 10^3$ (top) and $n = 10^6$ (bottom). The solid line stands for the UFGM method, the dotted line stands for the ULCM method, the dashed line stands for the NCG method.

$$f(x) = \max_{i=1,\dots,n} x_i + \frac{\mu}{2}\|x\|_2^2.$$

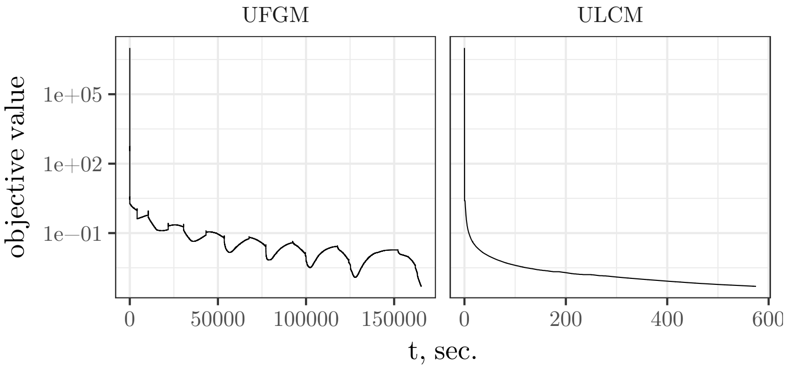

Figure 2: Methods convergence for the non-smooth problems with $n = 10^3$ (top) and $n = 10^6$ (bottom).