Universal Intermediate Gradient Method with Inexact Oracle

Dvurechensky Pavel[1,2], Gasnikov Alexander[1], and Kamzolov Dmitry[1] ⋆

[1] Moscow Institute of Physics and Technology, Moscow, Russia
[2] Weierstrass Institute for Applied Analysis and Stochastics
Berlin, German

In this paper, we consider first-order methods for convex optimization. The recent renaissance of these methods was mostly motivated by large-scale problems in data analysis, imaging, machine learning. In 2013 Nesterov proposed a Universal Fast Gradient Method (UFGM) which can solve a broad class of convex problems with Hölder-continuous subgradient and is uniformly optimal over this class. In 2013, Devolder, Glineur and Nesterov proposed an Intermediate Gradient Method (IGM) for problems with inexact oracle, which interpolates between GM and FGM to exploit the trade-off between the rate of convergence and the rate of error accumulation.

In this paper, we present Universal Intermediate Gradient Method (UIGM) for problems with deterministic inexact oracle. Our method both enjoys the universality with respect to Hölder class of the problem and interpolates between Universal Gradient Method and UFGM, thus, allowing to balance the rate of convergence of the method and rate of the error accumulation. First, we consider a composite convex optimization problem $\min_{x\in Q}\left[F(x)=f(x)+h(x)\right]$ where Q is a closed convex set, h is a simple convex function and function f is convex and subdifferentiable on Q with inexact Hölder-continuous subgradient. We assume that problem is solvable with optimal solution x^*. For such problems we construct our method and prove the theorem on its convergence rate. If the error δ of the oracle satisfies $\mathrm{O}\left(\frac{\varepsilon}{N^{p-1}}\right)$, where N is the number of algorithm steps, our method generates a point y s.t. $F(y)-F_*\le\varepsilon$ in $\mathrm{O}\left(\inf\limits_{\nu\in[0,1]}\left(\frac{L_\nu R^{1+\nu}}{\varepsilon}\right)^{\frac{2}{1+2p\nu-\nu}}\right)$ interations, where $\nu\in[0{,}1]$ is the Hölder parameter, L_ν is the Hölder constant, $p\in[0{,}1]$ is the method parameter, R is an estimate for the distance from the starting point to the solution.

Then, under additional assumption, that F is strongly convex function with known constant μ, we use restart technique to obtain an algorithm with faster rate of convergence. We restart when $\frac{4}{\mu A_k}\omega_n\le 1$. In our method A_k and ω_n is easy computed parameteres. If the error δ of the oracle satisfies $\mathrm{O}\left(\frac{\varepsilon}{N^{p-1}}\right)$, our method generates a point y s.t. $F(y)-F_*\le\varepsilon$ in

$$N=\mathrm{O}\left(\inf_{\nu\in[0,1]}\left(\frac{L_\nu^{\frac{2}{1+\nu}}}{\mu\varepsilon^{\frac{1-\nu}{1+\nu}}}\omega_n\right)^{\frac{1+\nu}{1+2p\nu-\nu}}\cdot\left\lceil\ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil\right)$$

⋆ This research was supported by a grant of the RF President MK-1806.2017.9.