Fedor S. Stonyakin

MIRROR DESCENT IN CONSTRAINED OPTIMIZATION FOR CONVEX FUNCTIONALS WITH NONSTANDARD GROWTH PROPERTIES

Mirror Descent and Convex Optimization Problems With Non-Smooth Inequality Constraints
(Anastasia Bayandina, Pavel Dvurechensky, Alexander Gasnikov, Fedor Stonyakin, Alexander Titov)
https://arxiv.org/abs/1710.06612

MIPT, October 27, 2017

1. Problem Classes

Problem formulation: $f^* = \min_{x \in Q} f(x)$, $g(x) \leqslant 0$, where

- f and g are convex function.
- Q is a simple closed convex set.
- We assume existence of $x^* \in Q$ such that $f(x^*) = f^*$.

Adaptive Mirror Descent algorithms:

1. $\|\nabla f(x)\|_* \leqslant M \; \forall x \in Q$ (*A. Bayandina, P. Dvurechensky, A. Gasnikov*).
2. Our aim (*F. Stonyakin, A. Titov*): non-bounded $\|\nabla f(x)\|_*$, but

$$\|\nabla f(x) - \nabla f(y)\|_* \leqslant L\|x - y\| \quad \forall x, y \in Q$$

Nesterov, Y.E. Subgradient methods for convex functions with nonstandard growth properties

Workshop: «Three oracles», Moscow, 2016

Let $A_i \succ 0$, $i = 1, \dots, m$. Consider the convex functions

$$f_i(x) = \frac{1}{2}\langle A_i x, x\rangle - \langle b_i, x\rangle + \alpha_i, \; i = 1, \dots, m.$$

Define the objective function $f(x) = \max\limits_{1 \leqslant i \leqslant m} f_i(x)$.

Note:

- f is a nonsmooth convex function.
- The gradients of f_i are Lipschitz continuous. However, it is not true for subgradients of f.
- Subgradients of f are not bounded.

$$\|\nabla f(x)\| \leqslant \max_{1 \leqslant i \leqslant m} \|\nabla f(x^*)\| + \max_{1 \leqslant i \leqslant m} L_i \|x - x^*\|.$$

Application: Truss Topology Design problem with weights of the bars.

2. Mirror Descent Basics

Let E be a finite-dimensional real vector space and E^* be its dual. We denote the value of a linear function $g \in E^*$ at $x \in E$ by $\langle g, x\rangle$. Let $\|\cdot\|_E$ be some norm on E, $\|\cdot\|_{E,*}$ be its dual, defined by

$$\|g\|_{E,*} = \max_x \big\{\langle g, x\rangle, \|x\|_E \leqslant 1\big\}$$

We use $\nabla f(x)$ to denote any subgradient of a function f at a point $x \in \operatorname{dom} f$. We choose a *prox-function* $d(x)$, which is continuous, convex on X and

1. admits a continuous gradient $d'(x)$, where $x \in X$;
2. $d(x)$ is 1-strongly convex on X with respect to $\|\cdot\|_E$, i.e., for any $x, y \in X$

$$d(y) - d(x) - \langle d'(x), y - x\rangle \geqslant \frac{1}{2}\|y - x\|_E^2.$$

Without loss of generality, we assume that $\min_{x \in X} d(x) = 0$.

We define also the corresponding *Bregman divergence*

$$V[z](x) = d(x) - d(z) - \langle d'(z), x - z\rangle,\ x \in X, z \in X^{\cdot}$$

Given a vector $x \in X$, and a vector $g \in E^*$, the Mirror Descent step is defined as

$$x_+ = \mathrm{Mirr}[x](g) := \arg\min_{u \in X} \big\{\langle g, u\rangle + V[x](u)\big\}. \tag{2.1}$$

Assumption: $\mathrm{Mirr}[x](g)$ is easily computable.

3. Problem Statement

Consider the following convex constrained minimization problem

$$\min\{f(x) :\ x \in X \subset E, \quad g(x) \leqslant 0\}, \tag{3.1}$$

where X is a convex subset of a finite-dimensional real vector space E, $f : X \to \mathbb{R}$, $g : E \to \mathbb{R}$ are convex functions.

Assumptions:

1. g is non-smooth and is Lipschitz-continuous

$$|\ g(x) - g(y)\ | \leqslant M_g \|x - y\|_E, \quad x, y \in X. \tag{3.2}$$

2. There exist a point $\bar{x} \in riX$, such that $g(\bar{x}) < 0$.

Definition 3.1. *We assume that Let x_* be a solution to (3.1). We say that a point $\tilde{x} \in X$ is an ε-solution to (3.1) if*

$$f(\tilde{x}) - f(x_*) \leqslant \varepsilon, \quad g(\tilde{x}) \leqslant \varepsilon. \tag{3.3}$$

4. The case of Lipschitz-continuous Objective Function

A. Bayandina, P. Dvurechensky, A. Gasnikov

Let f be a non-smooth Lipschitz-continuous function

$$|\, f(x) - f(y) \,| \leqslant M_f \|x - y\|_E, \quad x, y \in X. \tag{4.1}$$

Let x_* be a solution to (3.1) and assume that we know a constant $\Theta_0 > 0$ such that

$$d(x_*) \leqslant \Theta_0^2. \tag{4.2}$$

For example, if X is a compact set, one can choose

$$\Theta_0^2 = \max_{x \in X} d(x).$$

Algorithm 1. Adaptive Mirror Descent (Non-Smooth Lipschitz-continuous Objective)
Input: accuracy $\varepsilon > 0$; Θ_0 s.t. $d(x_*) \leqslant \Theta_0^2$.

1. $x^0 = \arg\min_{x \in X} d(x)$.
2. Initialize the set I as empty set.
3. Set $k = 0$.
4. **repeat**
5. **if** $g(x^k) \leqslant \varepsilon$ **then** $M_k = \|\nabla f(x^k)\|_{E,*}$, $h_k = \frac{\varepsilon}{M_k^2}$
6. $x^{k+1} = \mathrm{Mirr}[x^k](h_k \nabla f(x^k))$ ("productive step")
7. Add k to I.
8. **else** $M_k = \|\nabla g(x^k)\|_{E,*}$, $h_k = \frac{\varepsilon}{M_k^2}$
9. $x^{k+1} = \mathrm{Mirr}[x^k](h_k \nabla g(x^k))$ ("non-productive step")
10. **end if**
11. Set $k = k + 1$.
12. **until** $\sum_{j=0}^{k-1} \frac{1}{M_j^2} \geqslant \frac{2\Theta_0^2}{\varepsilon^2}$

Output: $\bar{x}^k := \frac{\sum_{i \in I} h_i x^i}{\sum_{i \in I} h_i}$

Theorem 4.1. *For*

$$k = \left\lceil \frac{2 \max\{M_f^2, M_g^2\} \Theta_0^2}{\varepsilon^2} \right\rceil, \tag{4.3}$$

$\bar{x}^k$ *is an* ε*-solution to* (3.1) *in the sense of* (3.3)*.*

5. General Convex Objective Function

Algorithm 1. Adaptive Mirror Descent (General Convex Objective)
Input: accuracy $\varepsilon > 0$, Θ_0 $s.t.$ $d(x_*) \leqslant \Theta_0^2$.

1. $x^0 = arg \min\limits_{x \in X} d(x)$.
2. Initialize the set I as empty set.
3. Set $k = 0$.
4. **repeat**
5. **if** $g(x^k) \leqslant \varepsilon$ **then** $h_k = \frac{\varepsilon}{\|\nabla f(x^k)\|_{E,*}}$
6. $x^{k+1} = \mathrm{Mirr}[x^k](h_k \nabla f(x^k))$ ("productive step")
7. Add k to I.
8. **else** $h_k = \frac{\varepsilon}{\|\nabla g(x^k)\|_{E,*}^2}$
9. $x^{k+1} = \mathrm{Mirr}[x^k](h_k \nabla g(x^k))$ ("non-productive step")
10. **end if**
11. Set $k = k + 1$.
12. **until** $\mid I \mid + \sum\limits_{j \in J} \frac{1}{\|\nabla g(x^k)\|_{E,*}^2} \geqslant \frac{2\Theta_0^2}{\varepsilon^2}$

Output: $\bar{x}^k := arg \min\limits_{x^j, j \in I} f(x^j)$

Given a function f and a point $y \in X$, we define for $x \in X$

$$v_f[y](x) = \begin{cases} \left\langle \frac{\nabla f(x)}{\|\nabla f(x)\|_{E,*}}, x - y \right\rangle, & \nabla f(x) \neq 0 \\ 0 & \nabla f(x) = 0 \end{cases} \tag{5.1}$$

Lemma 1. *Assume that f is a convex function. Then, for any $x \in X$,*

$$f(x) - f(x_*) \leqslant \omega(v_f[x_*](x)),$$

where

$$\omega(\tau) = \begin{cases} \max_{x \in X} \{f(x) - f(x_*) : \|x - x_*\| \leq \tau\} & \tau \geq 0, \\ 0 & \tau < 0. \end{cases} \tag{5.2}$$

Theorem 5.1. *If in Algorithm 2*

$$k = \left\lceil \frac{2max\{1, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil$$

then $\min_{i \in I} v_f[x_*](x^i) \leqslant \varepsilon$ *and* $g(\bar{x}^k) \leqslant \varepsilon$ *for all* $i \in I$.

Corollary 1. *Assume that the objective function* f *in*

$$\min\{f(x) : x \in X \subset E, g(x) \leqslant 0\} \tag{5.3}$$

is differentiable and its gradient is Lipschitz continuous

$$\|\nabla f(x) - \nabla f(y)\|_{E,*} \leqslant L\|x - y\| \; \forall x, y \in X.$$

Then

$$f(\bar{x}^k) - f(x_*) \leqslant \varepsilon \|\nabla f(x_*)\|_{E,*} + \frac{L\varepsilon^2}{2}, \quad g(\bar{x}^k) \leqslant \varepsilon. \tag{5.4}$$

Remark 5.1. The previous result is useful for the special class of non-smooth convex objective function. Assume in Corollary 1 the objective function $f = \max\limits_{1\leqslant m\leqslant M} f_k$, where f_k are convex and differentiable functions with Lipschitz continuous gradients

$$\|\nabla f_m(x) - \nabla f_m(y)\|_{E,*} \leqslant L_m\|x-y\| \;\forall x, y \in X,\; m = \overline{1, M},$$

$$L = \max_{1\leqslant m\leqslant M} L_m.$$

Then $\bar{x}^k$ is $\max\{\varepsilon, \varepsilon\|\nabla f(x_*)\|_{E,*} + \max \frac{L\varepsilon^2}{2}\}$-solution to (5.3) in the sense of (5.4).

Thank you for attention!