1 Deterministic Constrained Problems

Formally speaking, we consider the following convex constrained minimization problem

$$\min\{f(x): \quad x \in X \subset E, \quad g(x) \leq 0\}, \tag{1}$$

In this section, we consider problem (1) in two different settings, namely, non-smooth Lipschitz-continuous objective function f and general objective function f, which is not necessarily Lipschitz-continuous, e.g. a quadratic function. In both cases, we assume that g is non-smooth and is Lipschitz-continuous

$$|g(x) - g(y)| \leq M_g \|x - y\|_2, \quad x, y \in X. \tag{2}$$

Let x_* be a solution to (1). We say that a point $\tilde{x} \in X$ is an ε-*solution* to (1) if

$$f(\tilde{x}) - f(x_*) \leq \varepsilon, \quad g(\tilde{x}) \leq \varepsilon. \tag{3}$$

The methods we describe are based on the of Polyak's switching subgradient method [4] for constrained convex problems, also analyzed in [3], and Mirror Descent method originated in [2]; see also [1].

1.1 Convex Non-Smooth Objective Function

In this subsection, we assume that f is a non-smooth Lipschitz-continuous function

$$|f(x) - f(y)| \leq M_f \|x - y\|_2, \quad x, y \in X. \tag{4}$$

Let x_* be a solution to (1) and assume that we know a constant $\Theta_0 > 0$ such that

$$\frac{1}{2}\|x_0 - x_*\|_2^2 \leq \Theta_0^2. \tag{5}$$

Theorem 1. *Assume that inequalities* (2) *and* (4) *hold and a known constant* $\Theta_0 > 0$ *is such that* $\frac{1}{2}\|x_0 - x_*\|_2^2 \leq \Theta_0^2$. *Then, Algorithm 1 stops after not more than*

$$k = \left\lceil \frac{2\max\{M_f^2, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil \tag{6}$$

iterations and $\bar{x}^k$ *is an* ε*-solution to* (1) *in the sense of* (3).

Proof. First, let us prove that the inequality in the stopping criterion holds for k defined in (6). By (2) and (4), we have that, for any $i \in \{0, ..., k-1\}$, $M_i \leq \max\{M_f, M_g\}$. Hence, by (6), $\sum\limits_{j=0}^{k-1} \frac{1}{M_j^2} \geq \frac{k}{\max\{M_f^2, M_g^2\}} \geq \frac{2\Theta_0^2}{\varepsilon^2}$.

Algorithm 1 Adaptive Subgradient Descent (Non-Smooth Objective)

Input: accuracy $\varepsilon > 0$; Θ_0 s.t. $\frac{1}{2}\|x_0 - x_*\|_2^2 \le \Theta_0^2$.

```
x^0 = x_0.
Initialize the set I as empty set.
Set k = 0.
repeat
    if g(x^k) ≤ ε then
        M_k = ||∇f(x^k)||_2,
        h_k = ε / M_k^2
        x^{k+1} = π_X(x^k − h_k ∇f(x^k)) ("productive step")
        Add k to I.
    else
        M_k = ||∇g(x^k)||_2
        h_k = ε / M_k^2
        x^{k+1} = π_X(x^k − h_k ∇g(x^k)) ("non-productive step")
    end if
    Set k = k + 1.
until sum_{j=0}^{k-1} 1/M_j^2 ≥ 2Θ_0^2 / ε^2
```

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

Denote $[k] = \{i \in \{0,...,k-1\}\}$, $J = [k] \setminus I$. From main Lemma for subgradient descent, we have, for all $i \in I$ and all $u \in X$,

$$h_i \cdot \big(f(x^i) - f(u)\big) \le \frac{h_i^2}{2}\|\nabla f(x^i)\|_2^2 + \frac{1}{2}\|x^i - u\|_2^2 - \frac{1}{2}\|x^{i+1} - u\|_2^2$$

and, for all $i \in J$ and all $u \in X$,

$$h_i \cdot \big(g(x^i) - g(u)\big) \le \frac{h_i^2}{2}\|\nabla g(x^i)\|_2^2 + \frac{1}{2}\|x^i - u\|_2^2 - \frac{1}{2}\|x^{i+1} - u\|_2^2.$$

Summing up these inequalities for i from 0 to $k-1$, using the definition of h_i, $i \in \{0,...,k-1\}$, and taking $u = x_*$, we obtain

$$\begin{aligned} &\sum_{i\in I} h_i\big(f(x^i) - f(x_*)\big) + \sum_{i\in J} h_i\big(g(x^i) - g(x_*)\big) \\ &\le \sum_{i\in I} \frac{h_i^2 M_i^2}{2} + \sum_{i\in J} \frac{h_i^2 M_i^2}{2} + \sum_{i\in[k]} \big(\frac{1}{2}\|x^i - x_*\|_2^2 - \frac{1}{2}\|x^{i+1} - x_*\|_2^2\big) \\ &\le \frac{\varepsilon}{2}\sum_{i\in[k]} h_i + \Theta_0^2. \end{aligned} \quad (7)$$

Since, for $i \in J$, $g(x^i) - g(x_*) \ge g(x^i) > \varepsilon$, by convexity of f and the definition of $\bar{x}^k$, we have

$$\left(\sum_{i\in I} h_i\right)\left(f(\bar{x}^k)-f(x_*)\right) \leq \sum_{i\in I} h_i\left(f(x^i)-f(x_*)\right) < \frac{\varepsilon}{2}\sum_{i\in[k]} h_i - \varepsilon\sum_{i\in J} h_i + \Theta_0^2$$
$$= \varepsilon\sum_{i\in I} h_i - \frac{\varepsilon^2}{2}\sum_{i\in[k]}\frac{1}{M_i^2} + \Theta_0^2 \leq \varepsilon\sum_{i\in I} h_i, \qquad (8)$$

where in the last inequality, the stopping criterion is used. As long as the inequality is strict, the case of the empty I is impossible. Thus, the point $\bar{x}^k$ is correctly defined. Dividing both parts of the inequality by $\sum\limits_{i\in I} h_i$, we obtain the left inequality in (3).

For $i \in I$, it holds that $g(x^i) \leq \varepsilon$. Then, by the definition of $\bar{x}^k$ and the convexity of g,

$$g(\bar{x}^k) \leq \left(\sum_{i\in I} h_i\right)^{-1}\sum_{i\in I} h_i g(x^i) \leq \varepsilon.$$

□

Let us now show that Algorithm 1 allows to reconstruct an approximate solution to the problem, which is dual to (1). We consider a special type of problem (1) with g given by

$$g(x) = \max_{i\in\{1,\ldots,m\}}\left\{g_i(x)\right\}. \qquad (9)$$

Then, the dual problem to (1) is

$$\varphi(\lambda) = \min_{x\in X}\left\{f(x) + \sum_{i=1}^{m}\lambda_i g_i(x)\right\} \to \max_{\lambda_i\geq 0, i=1,\ldots,m}\varphi(\lambda), \qquad (10)$$

where $\lambda_i \geq 0, i = 1,\ldots,m$ are Lagrange multipliers.

We slightly modify the assumption (5) and assume that the set X is bounded and that we know a constant $\Theta_0 > 0$ such that

$$\max_{x\in X}\frac{1}{2}\|x_0 - x\|_2^2 \leq \Theta_0^2.$$

As before, denote $[k] = \{j \in \{0,\ldots,k-1\}\}$, $J = [k]\setminus I$. Let $j \in J$. Then a subgradient of $g(x)$ is used to make the j-th step of Algorithm 1. To find this subgradient, it is natural to find an active constraint $i \in 1,\ldots,m$ such that $g(x^j) = g_i(x^j)$ and use $\nabla g(x^j) = \nabla g_i(x^j)$ to make a step. Denote $i(j) \in 1,\ldots,m$ the number of active constraint, whose subgradient is used to make a non-productive step at iteration $j \in J$. In other words, $g(x^j) = g_{i(j)}(x^j)$ and $\nabla g(x^j) = \nabla g_{i(j)}(x^j)$. We define an approximate dual solution on a step $k \geq 0$ as

$$\bar{\lambda}_i^k = \frac{1}{\sum\limits_{j\in I} h_j}\sum_{j\in J, i(j)=i} h_j, \quad i \in \{1,\ldots,m\}. \qquad (11)$$

and modify Algorithm 1 to return a pair $(\bar{x}^k, \bar{\lambda}^k)$.

Theorem 2. *Assume that the set X is bounded, the inequalities* (2) *and* (4) *hold and a known constant $\Theta_0 > 0$ is such that $d(x_*) \leq \Theta_0^2$. Then, modified Algorithm 1 stops after not more than*

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

iterations and the pair $(\bar{x}^k, \bar{\lambda}^k)$ returned by this algorithm satisfies

$$f(\bar{x}^k) - \varphi(\bar{\lambda}^k) \leq \varepsilon, \quad g(\bar{x}^k) \leq \varepsilon. \tag{12}$$

Proof. From the main Lemma for one step of the subgradient descent, we have, for all $j \in I$ and all $u \in X$,

$$h_j\big(f(x^j) - f(u)\big) \leq \frac{h_j^2}{2}\|\nabla f(x^j)\|_2^2 + \frac{1}{2}\|x^j - u\|_2^2 - \frac{1}{2}\|x^{j+1} - u\|_2^2$$

and, for all $j \in J$ and all $u \in X$,

$$\begin{aligned} h_j\big(g_{i(j)}(x^j) - g_{i(j)}(u)\big) &\leq h_j\langle \nabla g_{i(j)}(x^j), x^j - u\rangle \\ &= h_j\langle \nabla g(x^j), x^j - u\rangle \\ &\leq \frac{h_j^2}{2}\|\nabla g(x^j)\|_2^2 + \frac{1}{2}\|x^j - u\|_2^2 - \frac{1}{2}\|x^{j+1} - u\|_2^2. \end{aligned}$$

Summing up these inequalities for j from 0 to $k-1$, using the definition of h_j, $j \in \{0, ..., k-1\}$, we obtain, for all $u \in X$,

$$\begin{aligned} &\sum_{j\in I} h_j\big(f(x^j) - f(u)\big) + \sum_{j\in J} h_j\big(g_{i(j)}(x^j) - g_{i(j)}(u)\big) \\ &\qquad \leq \sum_{i\in I} \frac{h_j^2 M_j^2}{2} + \sum_{j\in J} \frac{h_j^2 M_j^2}{2} + \sum_{j\in[k]} \big(\frac{1}{2}\|x^j - u\|_2^2 - \frac{1}{2}\|x^{j+1} - u\|_2^2\big) \\ &\qquad \leq \frac{\varepsilon}{2}\sum_{j\in[k]} h_j + \Theta_0^2. \end{aligned}$$

Since, for $j \in J$, $g_{i(j)}(x^j) = g(x^j) > \varepsilon$, by convexity of f and the definition of $\bar{x}^k$, we have, for all $u \in X$,

$$
\begin{aligned}
\left(\sum_{j\in I} h_j\right)\left(f(\bar{x}^k) - f(u)\right) &\le \sum_{j\in I} h_j\left(f(x^j) - f(u)\right) \\
&\le \frac{\varepsilon}{2}\sum_{j\in[k]} h_j + \Theta_0^2 - \sum_{j\in J} h_j\left(g_{i(j)}(x^j) - g_{i(j)}(u)\right) \\
&< \frac{\varepsilon}{2}\sum_{j\in[k]} h_i + \Theta_0^2 - \varepsilon\sum_{j\in J} h_i + \sum_{j\in J} h_j g_{i(j)}(u) \\
&= \varepsilon\sum_{j\in I} h_j - \frac{\varepsilon^2}{2}\sum_{j\in[k]}\frac{1}{M_j^2} + \Theta_0^2 + \sum_{j\in J} h_j g_{i(j)}(u) \\
&\le \varepsilon\sum_{j\in I} h_j + \sum_{j\in J} h_j g_{i(j)}(u), \qquad (13)
\end{aligned}
$$

where in the last inequality, the stopping criterion is used. At the same time, by (11), for all $u \in X$,

$$
\sum_{j\in J} h_j g_{i(j)}(u) = \sum_{i=1}^{m}\sum_{j\in J, i(j)=i} h_j g_{i(j)}(u) = \left(\sum_{j\in I} h_j\right)\sum_{i=1}^{m}\bar{\lambda}_i^k g_i(u).
$$

This and (13) give, for all $u \in X$,

$$
\left(\sum_{j\in I} h_j\right) f(\bar{x}^k) < \left(\sum_{j\in I} h_j\right)\left(f(u) + \varepsilon + \sum_{i=1}^{m}\bar{\lambda}_i^k g_i(u)\right).
$$

Since the inequality is strict and holds for all $u \in X$, we have $\left(\sum\limits_{j\in I} h_j\right) \neq 0$ and

$$
\begin{aligned}
f(\bar{x}^k) &< \varepsilon + \min_{u\in X}\left\{f(u) + \sum_{i=1}^{m}\bar{\lambda}_i^k g_i(u)\right\} \\
&= \varepsilon + \varphi(\bar{\lambda}^k). \qquad (14)
\end{aligned}
$$

Second inequality in (12) follows from Theorem 1. □

References

1. A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization.* Society for Industrial and Applied Mathematics, 2001.
2. A. Nemirovsky and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. J. Wiley & Sons, New York, 1983.
3. Y. Nesterov. *Introductory Lectures on Convex Optimization: a basic course*. Kluwer Academic Publishers, Massachusetts, 2004.
4. B. Polyak. A general method of solving extremum problems. *Soviet Mathematics Doklady*, 8(3):593–597, 1967.