1 Subgradient Projection Method

In this section, we consider a standard subgradient projection method as applied to a convex optimization problem subject to a simple set of constraints. In particular, we focus on the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array} \tag{1}$$

where $X \subseteq \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. In a more general setting, one would require that X is contained in an open set U, with $U \subseteq \text{relint}(dom\, f)$, in order to have $\partial f(x)$ nonempty for every $x \in X$ and f to be continuous over X. This is because the subgradient method with projections that we consider here would require the existence of a subgradient of f at any $x \in X$. In addition, the continuity of f is indispensable in the analysis presented here.

Throughout this section, we use the following assumption for the set X and the function f.

Assumption 1 *The set X is convex and closed. The function f is convex over $\mathbb{R}^n$. The optimal value f^* of the problem (1) is finite.*

The subgradient projection method is an iterative method that starts with some initial feasible vector $x_0 \in X$, and generates the next iterate by taking a step along the negative subgradient direction $-s_k$ of f at x_k and then, by projecting on the set X to maintain feasibility. Formally, a typical iteration of the subgradient projection method is given by

$$x_{k+1} = P_X\left[x_k - \alpha_k s_k\right], \tag{2}$$

where the scalar $\alpha_k > 0$ is a stepsize, x_k is the current iterate, and $P_X[y]$ is the projection of a vector y on the set X. By the Projection Theorem, the projection of any vector y on X exists, and it is unique since X is closed and convex.

We assume that the set X is simple so that the projection on X is easy. Examples of such sets X include a nonnegative orthant, a box, and a ball. When X is the nonnegative orthant $\mathbb{R}^n_+$, then the projection on X decomposes into projections per coordinate. In particular, in this case, the projection $P_X[x]$ is the vector x^+ with components $x_i^+ = \max\{x_i, 0\}$. When X is the box,

$$X = \{x \in \mathbb{R}^n \mid a_i \le x_i \le b_i \;\text{ for all } i\},$$

again the projection on X decomposes into projections per coordinate, and the components of the projection $P_X[x]$ vector are given by

$$[P_X[x]]_i = \begin{cases} a_i & \text{if } x_i < a_i \\ x_i & \text{if } a_i \le x_i \le b_i \\ b_i & \text{if } b_i < x_i. \end{cases}$$

We consider the subgradient projection method with several stepsize rules:

- *Constant stepsize*, where for some $\alpha > 0$, we have $\alpha_k = \alpha$ for all k.

- *Diminishing stepsize*, where $\alpha_k \to 0$ and $\sum_k \alpha_k = \infty$.
- *Polyak's stepsize*, where $\alpha_k = \frac{f(x_k)-f^*}{\|s_k\|^2}$.
- *Modified Polyak's stepsize*, where $\alpha_k = \frac{f(x_k)-\hat{f}_k}{\|s_k\|^2}$ and $\hat{f}_k = \min_{0\le j\le k} f(x_j) - \delta$ for some scalar $\delta > 0$.

The constant stepsize rule is suitable when we are interested in finding an approximate solution to the problem (1). Diminishing stepsize rule is an off-line rule and is typically used with $\alpha_k = \frac{c}{k+1}$ or $\frac{c}{\sqrt{k+1}}$ for some $c > 0$. The constant and the diminishing stepsize are also well suited for some distributed implementations of the method.

Let us mention that Polyak's stepsize in general form involves a parameter γ_k. In particular, the stepsize is given by

$$\alpha_k = \gamma_k \, \frac{f(x_k) - f^*}{\|s_k\|^2},$$

where γ_k is bounded away from zero and away from 2, i.e., $0 < \underline{\gamma} \le \gamma_k \le \overline{\gamma} < 2$ for all k. Here, however, we simply use $\gamma_k = 1$ for all k, since in general it is not clear what should guide the choice for γ_k at each iteration k. Furthermore, all the results that we establish here, also hold for the Polyak's rule in general form.

Polyak's stepsize is suitable when the optimal function value f^* is known and the evaluation of the function f is possible and "easy". There are some optimization problems for which f^* is known and the algorithm is applied to generate an optimal solution. This is the case, for example, in the feasibility problem of finding a point $\tilde{x}$ satisfying a given system of convex inequalities $g_j(x) \le 0$ for $j = 1, \ldots, m$ and a set X constraint. Such a problem can be equivalently casted as an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & \max_{1\le j\le m} g_j(x) \\ \text{subject to} & x \in X, \end{array}$$

or

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{m} \max\{0, g_j(x)\} \\ \text{subject to} & x \in X. \end{array}$$

Evidently, in feasibility problems, we have $f^* = 0$ when the set defined with inequalities and the set X is nonempty.

The modified Polyak's stepsize is a simple adaptation of the Polyak's stepsize to accommodate the situations when f^* is not known. Note that the modified Polyak's stepsize is using a fixed positive parameter δ and, as we will see, the method with this stepsize exhibits behavior similar to that of the method with the constant stepsize, thus only generating an approximate solution. There exist other more complex modifications of the Polyak's rule for which the method generates optimal solutions.

In what follows, we study the convergence properties of the method under the stepsize rules described above. We note that the subgradient projection method does not necessarily generate iterates with decreasing cost. What makes the method work is the property that at each new iteration, the method either decreases the function value or decreases the distance to the optimal set. This property is captured in the following lemma, providing a basis for the analysis of the method.

Lemma 1 *Let Assumption 1 hold. Let $y \in X$ be arbitrary but fixed. Then, for the subgradient projection method with any stepsize rule, we have for all k,*

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k(f(x_k) - f(y)) + \alpha_k^2 \|s_k\|^2.$$

Proof. Since the projection is a nonexpansive mapping it follows that for any k and any $y \in X$,

$$\|x_{k+1} - y\|^2 \leq \|x_k - \alpha_k s_k - y\|^2 = \|x_k - y\|^2 - 2\alpha_k s_k^T(x_k - y) + \alpha_k^2 \|s_k\|^2.$$

By the subgradient property, we have for any k and any $y \in X$,

$$f(x_k) - f(y) \leq s_k^T(x_k - y).$$

The desired relation follows by combining the preceding two inequalities. ■

1.1 Convergence for Constant and Diminishing Rule

In our analysis in this section, we use an additional assumption on subgradients of f, namely, we assume that the subgradients are bounded uniformly over X.

Assumption 2 *There exists a constant $L > 0$ such that*

$$\|s_k\| \leq L \qquad \textit{for all } x \in X.$$

This assumption is satisfied for example, when X is a compact set and f is continuously differentiable over X. Under this assumption, for the method with the constant stepsize, we have the following result.

Theorem 1 *Let Assumptions 1 and 2 hold. Then, for the subgradient projection method with the constant stepsize α, we have*

$$\liminf_{k \to \infty} f(x_k) \leq f^* + \frac{\alpha L^2}{2}.$$

Proof. To arrive at a contradiction, assume that the given relation does not hold, i.e., assume that

$$\liminf_{k \to \infty} f(x_k) > f^* + \frac{\alpha L^2}{2}.$$

Then, for some sufficiently small $\epsilon > 0$ and some $K \geq 1$, we have

$$f(x_k) \geq f^* + \frac{\alpha L^2}{2} + 2\epsilon \qquad \text{for all } k \geq K.$$

The function f is continuous over X, so that there exists $\hat{y} \in X$ such that $f(\hat{y}) = f^* + \epsilon$, implying that

$$f(x_k) - f(\hat{y}) \geq \frac{\alpha L^2}{2} + \epsilon \qquad \text{for all } k \geq K.$$

By using the relation of Lemma 1 with $\alpha_k = \alpha$ and $y = \hat{y}$, we obtain for all $k \geq K$,

$$\begin{aligned}\|x_{k+1} - \hat{y}\|^2 &\leq \|x_k - \hat{y}\|^2 - 2\alpha(f(x_k) - f(\hat{y})) + \alpha^2 \|s_k\|^2 \\ &\leq \|x_k - \hat{y}\|^2 - 2\alpha\left(\frac{\alpha L^2}{2} + \epsilon\right) + \alpha^2 L^2,\end{aligned}$$

where in the last inequality we use the uniform boundedness of the subgradients. Hence

$$\|x_{k+1} - \hat{y}\|^2 \leq \|x_k - \hat{y}\|^2 - 2\alpha\epsilon \qquad \text{for all } k \geq K,$$

and by summing the preceding inequalities over $k = K+1, \cdots, t$, we obtain for $t \geq K$,

$$\|x_t - \hat{y}\|^2 \leq \|x_K - \hat{y}\|^2 - 2(t-K)\alpha\epsilon. \tag{3}$$

However, the preceding relation fails to hold for sufficiently large t. Therefore, we must have

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \frac{\alpha L^2}{2}.$$

■

When the set X is compact, and we know the maximal distance d of the set X, i.e., $d = \max_{x,y\in X} \|x - y\|$, we can use relation (3) to compute an upper bound of the minimal number of iterations N needed to guarantee that the error level αL^2 is achieved, i.e., to guarantee that

$$\min_{0\leq k\leq N} f(x_k) \leq f^* + \alpha L^2.$$

In particular, by choosing $\epsilon = \frac{\alpha L^2}{2}$ and using relation (3), we can see that the number N is given by

$$N = \left\lfloor \frac{d^2}{\alpha^2 L^2} \right\rfloor.$$

Theorem 2 *Let Assumptions 1 and 2 hold. Then, for the subgradient projection method with the diminishing*[1] *stepsize α_k, we have*

$$\liminf_{k\to\infty} f(x_k) = f^*.$$

Proof. To obtain a contradiction, assume that the given relation does not hold. Then, for some sufficiently small $\epsilon > 0$ a large enough $K \geq 1$, we have

$$f(x_k) \geq f^* + 2\epsilon \qquad \text{for all } k \geq K.$$

Since the function f is continuous over X, there exists $\hat{y} \in X$ such that $f(\hat{y}) = f^* + \epsilon$, implying that

$$f(x_k) - f(\hat{y}) \geq \epsilon \qquad \text{for all } k \geq K.$$

By using the relation of Lemma 1 with $y = \hat{y}$, we obtain for all k,

$$\|x_{k+1} - \hat{y}\|^2 \leq \|x_k - \hat{y}\|^2 - 2\alpha_k(f(x_k) - f(\hat{y})) + \alpha_k^2 \|s_k\|^2$$

[1]Diminishing stepsize α_k is assumed to satisfy $\alpha_k \to 0$ and $\sum_k \alpha_k = +\infty$.

$$\leq \|x_k - \hat{y}\|^2 - 2\alpha_k \epsilon + \alpha_k^2 L^2,$$

where in the last inequality we also use the uniform boundedness of the subgradients. Hence

$$\|x_{k+1} - \hat{y}\|^2 \leq \|x_k - \hat{y}\|^2 - \alpha_k \left(2\epsilon - \alpha_k L^2\right) \qquad \text{for all } k \geq K.$$

Since $\alpha_k \to 0$, there exists $\hat{k} \geq K$ such that $\epsilon \geq \alpha_k L^2$, implying that

$$\|x_{k+1} - \hat{y}\|^2 \leq \|x_k - \hat{y}\|^2 - \alpha_k \epsilon \qquad \text{for all } k \geq \hat{k}.$$

By summing the preceding inequalities over $k = \hat{k}, \ldots, t$, we obtain

$$\|x_t - \hat{y}\|^2 \leq \|x_{\hat{k}} - \hat{y}\|^2 - \epsilon \sum_{i=\hat{k}}^{t-1} \alpha_i. \tag{4}$$

By letting $t \to \infty$ and using the fact $\sum_k \alpha_k = +\infty$, we see that the left hand side of relation (4) tends to $-\infty$, while its right hand side is nonnegative - a contradiction. Therefore, we must have $\liminf_{k\to\infty} f(x_k) = f^*$. ■

When the optimal set X^* of the problem in Eq. (1) is nonempty and we impose a stronger condition on the diminishing stepsize, namely $\sum_k \alpha_k^2 < \infty$, we can show that the whole sequence $\{x_k\}$ generated by the subgradient projection method converges to some optimal solution x^*. This result is formally given in the following theorem.

Theorem 3 *Let Assumptions 1 and 2 hold. Also, assume that the optimal set X^* of the problem (1) is nonempty. Suppose that the stepsize is such that $\sum_k \alpha_k = \infty$ and $\sum_k \alpha_k^2 < \infty$. Then, for the iterate sequence $\{x_k\}$ generated by the subgradient projection method with such a stepsize α_k, we have*

$$\lim_{k\to\infty} \|x_k - x^*\| = 0 \qquad \textit{for some } x^* \in X^*.$$

Proof. By using the relation of Lemma 1 with $y = x^*$, we obtain for all k,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\alpha_k (f(x_k) - f^*) + \alpha_k^2 \|s_k\|^2 \\ &\leq \|x_k - x^*\|^2 + \alpha_k^2 L^2, \end{aligned}$$

where in the last inequality we use $f(x_k) \geq f^*$ and the uniform boundedness of the subgradients. By summing the preceding inequalities over $k = \hat{k}, \ldots, K$ for some arbitrary $\hat{k}$ and K with $\hat{k} < K$, we obtain

$$\|x_K - x^*\|^2 \leq \|x_{\hat{k}} - x^*\|^2 + L^2 \sum_{k=\hat{k}}^{K-1} \alpha_k^2.$$

Therefore,

$$\limsup_{K\to\infty} \|x_K - x^*\|^2 \leq \|x_{\hat{k}} - x^*\|^2 + L^2 \sum_{k=\hat{k}}^{\infty} \alpha_k^2. \tag{5}$$

By letting $\hat{k} = 0$, we see that the sequence $\{\|x_k\|\}$ is bounded and hence, it has at least one accumulation point. Since $\sum_k \alpha_k^2 < \infty$, it follows that $\alpha_k \to 0$. By Theorem 2 we have $\liminf_{k\to\infty} f(x_k) = f^*$. Thus, one of the accumulation points of $\{x_k\}$ must belong to the optimal set X^*. Let $\{x_{k_i}\}$ be a subsequence such that $x_{k_i} \to \hat{x}^*$ with $\hat{x}^* \in X^*$.

By setting $x^* = \hat{x}^*$ and $\hat{k} = k_i$ in Eq. (5), and by letting $i \to \infty$, we obtain

$$\limsup_{K\to\infty} \|x_K - \hat{x}^*\|^2 \leq \lim_{i\to\infty} \|x_{k_i} - \hat{x}^*\|^2 + \lim_{i\to\infty} \sum_{k=k_i}^{\infty} \alpha_k^2 = 0,$$

where we have used $\lim_{i\to\infty} \|x_{k_i} - \hat{x}^*\| = 0$ and

$$\lim_{i\to\infty} \sum_{k=k_i}^{\infty} \alpha_k^2 = 0,$$

which follows from $\sum_k \alpha_k^2 < \infty$. Hence, the whole sequence $\{x_k\}$ converges to $\hat{x}^* \in X^*$. ■

1.2 Convergence for Polyak's Stepsize and its Modification

For the case when the optimal value f^* is known Polyak (1969) had suggested the stepsize rule of the form

$$\alpha_k = \frac{f(x_k) - f^*}{\|s_k\|^2}.$$

For the convergence analysis of the Polyak's stepize, we do not need the assumption on uniform boundedness of the subgradients. This is because, the stepize "normalizes" the directions that the method is using. In particular, the direction that the method is using is given by

$$\alpha_k s_k = \frac{f(x_k) - f^*}{\|s_k\|} \frac{s_k}{\|s_k\|}.$$

We have the following convergence result.

Theorem 4 *Let Assumption 1 hold. Also, assume that the optimal set X^* of the problem (1) is nonempty. Then, for the iterate sequence $\{x_k\}$ generated by the subgradient projection method with Polyak's stepsize α_k, we have*

$$\lim_{k\to\infty} \|x_k - x^*\| = 0 \qquad \text{for some } x^* \in X^*.$$

Proof. By using the relation of Lemma 1 with $y = x^*$, we obtain for all k and any $x^* \in X^*$,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\alpha_k(f(x_k) - f^*) + \alpha_k^2\|s_k\|^2 \\ &= \|x_k - x^*\|^2 - \frac{(f(x_k) - f^*)^2}{\|s_k\|^2}, \end{aligned}$$

where in the last inequality we use the stepsize expression $\alpha_k = \frac{f(x_k)-f^*}{\|s_k\|^2}$. Therefore, it follows that for any $x^* \in X^*$, and any k and s with $k > s$,

$$\|x_{k+1} - x^*\|^2 \leq \|x_s - x^*\|^2 - \sum_{i=s}^{k} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2}. \tag{6}$$

By letting $s = 0$, we see that the iterate sequence $\{x_k\}$ is bounded and therefore, it has an accumulation point. Furthermore, note that Eq. (6) implies that

$$\sum_{i=0}^{\infty} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \leq \|x_0 - x^*\|^2 < \infty. \tag{7}$$

Suppose that none of the accumulation points of $\{x_k\}$ belongs to the optimal set X^*. Then, for some small scalar $\epsilon > 0$, we have

$$f(x_k) > f^* \qquad \text{for all } k.$$

Since $\{x_k\}$ is bounded, so is the subgradient sequence $\{s_k\}$, i.e., there is a scalar $c > 0$ such that

$$\|s_k\| \leq c \qquad \text{for all } k.$$

Therefore,

$$\frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \geq \frac{\epsilon^2}{c^2} \qquad \text{for all } k.$$

By summing the preceding relations over k, we obtain

$$\sum_{i=0}^{\infty} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \geq \sum_{i=0}^{\infty} \frac{\epsilon^2}{c^2} = \infty,$$

thus contradicting the relation in Eq. (7). Hence, every accumulation point of $\{x_k\}$ must belong to the set X^*.

Let $\hat{x}^*$ be an accumulation point of the sequence $\{x_k\}$, and let $\{x_{k_j}\}$ be a subsequence of $\{x_k\}$. By letting $x^* = \hat{x}^*$ and $s = k_j$ in Eq. (6), we obtain for all $k > k_j$.

$$\|x_{k+1} - \hat{x}^*\|^2 \leq \|x_{k_j} - \hat{x}^*\|^2 - \sum_{i=k_j}^{k} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2}.$$

Therefore,

$$\limsup_{k\to\infty} \|x_{k+1} - \hat{x}^*\|^2 \leq \|x_{k_j} - \hat{x}^*\|^2 - \sum_{i=k_j}^{\infty} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2}.$$

By letting $j \to \infty$, and by using the relation $\|x_{k_j} - \hat{x}^*\| \to 0$ and Eq. (7), we obtain

$$\limsup_{k\to\infty} \|x_{k+1} - \hat{x}^*\|^2 \leq \lim_{j\to\infty} \left(\|x_{k_j} - \hat{x}^*\|^2 - \sum_{i=k_j}^{\infty} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \right) = 0,$$

thus showing that the entire sequence converges to $\hat{x}^* \in X^*$. ■

The relation shown at the beginning of the proof of Theorem 4, namely,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{(f(x_k) - f^*)^2}{\|s_k\|^2} \qquad \text{for all } k \text{ and any } x^* \in X^* \tag{8}$$

will be important in our assessment of the convergence rate of the subgradient projection method in Section 1.3.

We now provide a convergence result for the modified Polyak's stepsize. Note that for this stepsize, we have

$$\alpha_k = \frac{f(x_k) - \hat{f}_k}{\|s_k\|^2} = \frac{f(x_k) - \min_{0\le j\le k} f(x_j) + \delta}{\|s_k\|^2} \ge \frac{\delta}{\|s_k\|^2}.$$

As we will see, this stepsize remains bounded away from zero, so the convergence result is similar to that of Theorem 1.

Theorem 5 *Let Assumption 1 hold. Then, for the iterate sequence $\{x_k\}$ generated by the subgradient projection method with modified Polyak's stepsize α_k, we have*

$$\liminf_{k\to\infty} f(x_k) \le f^* + \delta.$$

Proof. To obtain a contradiction, assume that the given relation does not hold, i.e., assume that for some sufficiently small $\epsilon > 0$,

$$\liminf_{k\to\infty} f(x_k) > f^* + \delta + \epsilon.$$

The function f is continuous over X, so there exists $\hat{y} \in X$ such that $f(\hat{y}) = f^* + \epsilon$, implying that

$$f(x_k) \ge f(\hat{y}) + \delta \qquad \text{for all } k.$$

Therefore

$$\hat{f}_k = \min_{0\le j\le k} f(x_j) - \delta \ge f(\hat{y}) \qquad \text{for all } k,$$

implying that

$$f(x_k) - f(\hat{y}) \ge f(x_k) - \hat{f}_k \qquad \text{for all } k. \tag{9}$$

By using the relation of Lemma 1 with $y = \hat{y}$, we obtain for all k,

$$\begin{aligned} \|x_{k+1} - \hat{y}\|^2 &\le \|x_k - \hat{y}\|^2 - 2\alpha_k(f(x_k) - f(\hat{y})) + \alpha_k^2\|s_k\|^2 \\ &\le \|x_k - \hat{y}\|^2 - 2\alpha_k(f(x_k) - \hat{f}_k) + \alpha_k^2\|s_k\|^2, \end{aligned}$$

where the last inequality is obtained by using the relation (9). Using the stepsize expression $\alpha_k = \frac{f(x_k)-\hat{f}_k}{\|s_k\|^2}$, we see that for all k,

$$\|x_{k+1} - \hat{y}\|^2 \le \|x_k - \hat{y}\|^2 - \frac{(f(x_k) - \hat{f}_k)^2}{\|s_k\|^2}. \tag{10}$$

By summing the preceding inequalities, we obtain for all k,

$$\|x_{k+1} - x^*\|^2 \le \|x_0 - x^*\|^2 - \sum_{i=0}^{k} \frac{(f(x_i) - \hat{f}_k)^2}{\|s_i\|^2}. \tag{11}$$

In particular, from the relation (11) it follows that the iterate sequence $\{x_k\}$ is bounded and that

$$\sum_{i=0}^{\infty} \frac{(f(x_i) - \hat{f}_k)^2}{\|s_i\|^2} \leq \|x_0 - x^*\|^2 < \infty. \tag{12}$$

Since $\{x_k\}$ is bounded, so is the subgradient sequence $\{s_k\}$, i.e., there is a scalar $c > 0$ such that

$$\|s_k\| \leq c \qquad \text{for all } k.$$

Therefore,

$$\frac{(f(x_i) - \hat{f}_k)^2}{\|s_i\|^2} \geq \frac{\delta^2}{c^2} \qquad \text{for all } k,$$

implying that

$$\sum_{i=0}^{\infty} \frac{(f(x_i) - f^*)^2}{\|s_i\|^2} \geq \sum_{i=0}^{\infty} \frac{\delta^2}{c^2} = \infty,$$

contradicting the relation in Eq. (12). Hence, we must have $\liminf_{k\to\infty} f(x_k) \leq f^* + \delta$. ■

When the set X is compact, and we know the maximal distance d of the set X, i.e., $d = \max_{x,y\in X} \|x - y\|$, we can use relation (10) to compute an upper bound of the minimal number of iterations N needed to guarantee that the error level δ is achieved, i.e., to guarantee that

$$\min_{0\leq k\leq N} f(x_k) \leq f^* + \delta.$$

In particular, by using relation (3), we can see that the number N is given by

$$N = \left\lfloor \frac{d^2 L^2}{\delta^2} \right\rfloor,$$

where L is an upper bound on the subgradient norms $\|s_k\|$.

1.3 Convergence Rate

The convergence rate of the projection subgradient method is at best linear. The linear convergence is attained with Polyak's stepsize and for an objective function with *a sharp set of minima*. In particular, we say that f has *a sharp set of minima over* X when for some scalar $\mu > 0$,

$$f(x) - f^* \geq \mu \, \mathrm{dist}(x, X^*) \qquad \text{for all } x \in X, \tag{13}$$

where $\mathrm{dist}(x, Y)$ is the distance from the vector x to the set Y. It can be seen that a (polyhedral) function has a sharp set of minima. Let f be *a polyhedral function*, i.e., a function of the form

$$f(x) = \max_{1\leq i\leq m} \{a_i^T x + b\},$$

where $b_i \in \mathbb{R}$ and $a_i \in \mathbb{R}^n$ with $a_i \neq 0$ for all i. Then, it can be seen that f satisfies the relation for a sharp set of minima with $\mu = \min_{1\leq i\leq m} \|a_i\|$, provided that X^* is not empty.

In following theorem, we establish a linear convergence rate for the subgradient projection method. In the proof of the theorem, we use relation (8) mentioned after the proof of Theorem 4.

Theorem 6 *Let Assumption 1 hold and let the optimal set X^* be nonempty. Also, assume that f has a sharp set of minima over X [cf. Eq. (13)]. Then, the subgradient projection method with Polyak's stepsize α_k converges linearly, i.e., we have*

$$\operatorname{dist}(x_k, X^*)^2 \leq \left(1 - \frac{\mu^2}{L^2}\right)^k \operatorname{dist}(x_0, X^*)^2 \qquad \textit{for all } k,$$

where L is the subgradient norm bound uniform over X [$\|\nabla f(x)\| \leq L$ for all $x \in X$].

Proof. We start with the relation (8), i.e.,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{(f(x_k) - f^*)^2}{\|s_k\|^2} \qquad \text{for all } k \text{ and any } x^* \in X^*.$$

It follows that the sequence $\{x_k\}$ is bounded, so that the subgradient sequence $\{s_k\}$ is also bounded. Let L be an upper bound on the subgradient norms. By taking the minimum with respect to $x^* \in X^*$ in both sides of the preceding relation, and also using the subgradient boundedness, we obtain

$$\operatorname{dist}(x_{k+1}, X^*)^2 \leq \operatorname{dist}(x_k, X^*)^2 - \frac{(f(x_k) - f^*)^2}{L^2} \qquad \text{for all } k.$$

Using the sharp minima relation (13), we obtain

$$\operatorname{dist}(x_{k+1}, X^*)^2 \leq \operatorname{dist}(x_k, X^*)^2 - \frac{\mu^2}{L^2} \operatorname{dist}(x_k, X^*)^2 = \left(1 - \frac{\mu^2}{L^2}\right) \operatorname{dist}(x_k, X^*)^2 \qquad \text{for all } k,$$

which implies the desired result. ■

The potential drawback of the projection subgradient method is the same as for the gradient projection method: it can be very slow when the subgradient directions are almost perpendicular to the directions pointing toward the optimal set X^*, corresponding to

$$s_k^T (x_k - x^*) \approx 0.$$

In this case, the method may exhibit zig-zag behavior, depending on the initial iterate x_0. This may happen in particular when the level sets of the function f are prolonged in certain directions (known as ill-posedness), as illustrated in Figure 1 [for $X = \mathbb{R}^n$].

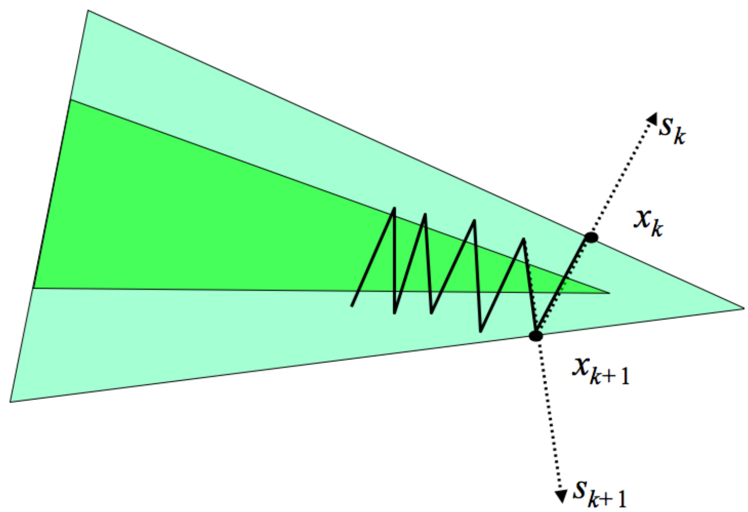

Figure 1: A zig-zagging behavior of the subgradient method when the level sets of the function f are prolonged.