Two Generalizations of the Projected Gradient Method for Convexly Constrained Inverse Problems: Hybrid steepest descent method, Adaptive projected subgradient method

Author(s): Yamada, Isao; Ogura, Nobuhiko

Citation: 数理解析研究所講究録 (2004), 1362: 88-94

Issue Date: 2004-04

URL: http://hdl.handle.net/2433/25282

Type: Departmental Bulletin Paper

Textversion: publisher

Kyoto University
Two Generalizations of the Projected Gradient Method for Convexly Constrained Inverse Problems —
Hybrid steepest descent method, Adaptive projected subgradient method

山田 功 (Isao Yamada) † 小倉信彦 (Nobuhiko Ogura) ‡
† Dept. of Communications & Integrated Systems, Tokyo Institute of Technology
‡ Precision and Intelligence Laboratory, Tokyo Institute of Technology

Abstract In this paper, we present a brief review on the central results of two generalizations of a classical convex optimization technique named the projected gradient method [1, 2]. The 1st generalization has been made by extending the convex projection operator, used in the projected gradient method, to the (quasi-)nonexpansive mapping in a real Hilbert space. By this generalization, we deduce the hybrid steepest descent method [3–10] (see also [11]) that can minimize the convex cost function over the fixed point set of nonexpansive mapping [3–9, 11] (these results can also be interpreted as generalizations of fixed point iterations found for example in [12–15]) or, more generally, over the fixed point set of quasi-nonexpansive mapping [10]. Since (i) the solution set of wide range of convexly constrained inverse problems, for example in signal processing and image reconstruction, can be characterized as the fixed point set of certain nonexpansive mapping [5, 6, 9, 16–18], and (ii) subgradient projection operator and its variations are typical examples of quasi-nonexpansive mapping [10, 19], the hybrid steepest descent method has rich applications in broad range of mathematical sciences and engineering. The 2nd generalization has been made for the Polyak’s subgradient algorithm [20] that was originally developed as a version, of the projected gradient method, for unsmooth convex optimization problem with a fixed target value. By extending the Polyak’s subgradient algorithm to the case where the convex cost function itself keeps changing in the whole process, we deduce the adaptive projected subgradient method [21–23] that can minimize asymptotically the sequence of unsmooth nonnegative convex cost functions. The adaptive projected subgradient method can serve as a unified guiding principle of a wide range of set theoretic adaptive filtering schemes [24–30] for nonstationary random processes. The great flexibilities in the choice of (quasi-)nonexpansive mapping as well as unsmooth convex cost functions in the proposed methods yield naturally inherently parallel structures (in the sense of [31]).

1 Preliminaries

Let $\mathcal{H}$ be a real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. For a continuous convex function $\Phi : \mathcal{H} \to \mathbb{R}$, the subdifferential of $\Phi$ at $y \in \mathcal{H}$, the set of all subgradients of $\Phi$ at $y$: $\partial \Phi(y) := \{ g \in \mathcal{H} | \langle x-y, g \rangle + \Phi(y) \leq \Phi(x), \forall x \in \mathcal{H} \}$ is nonempty. The convex function $\Phi : \mathcal{H} \to \mathbb{R}$ has a unique subgradient at $y \in \mathcal{H}$ if $\Phi$ is Gâteaux differentiable at $y$. This unique subgradient is nothing but the Gâteaux differential $\Phi'(y)$. A fixed point of a mapping $T : \mathcal{H} \to \mathcal{H}$ is a point $x \in \mathcal{H}$ such that $T(x) = x$. $Fix(T) := \{ x \in \mathcal{H} | T(x) = x \}$ denotes the fixed point set of $T$. A mapping $T : \mathcal{H} \to \mathcal{H}$ is called (i) strictly contractive if $\| T(x) - T(y) \| \leq \kappa \| x - y \|$ for some $\kappa \in (0, 1)$ and all $x, y \in \mathcal{H}$ [The Banach-Picard fixed point theorem guarantees the unique existence of the fixed point, say $x_* \in Fix(T)$, of $T$ and the strong convergence of $(T^n(x_0))_{n \geq 0}$ to

†This work was supported in part by JSPS grants-in-Aid (C15500129).
A mapping $f : \mathcal{H} \rightarrow \mathcal{H}$ is called (i) monotone over $S \subset \mathcal{H}$ if $\langle f(u) - f(v), u - v \rangle \geq 0$, $\forall u, v \in S$. In particular, a mapping $\Phi$ which is monotone over $S \subset \mathcal{H}$ is called (ii) $\alpha$-monotone over $S$ if $\langle \Phi(u) - \Phi(v), u - v \rangle = 0 \Leftrightarrow \Phi(u) = \Phi(v)$, $\forall u, v \in S$ [34]; (iii) uniformly monotone over $S$ if there exists a strictly monotone increasing continuous function $\alpha : [0, \infty) \rightarrow [0, \infty)$, with $\alpha(0) = 0$ and $\alpha(t) \rightarrow \infty$ as $t \rightarrow \infty$, satisfying $\langle \Phi(u) - \Phi(v), u - v \rangle \geq \alpha(||u - v||)||u - v||$ for all $u, v \in S$ [38]; (iv) $\eta$-strongly monotone over $S$ if there exists $\eta > 0$ such that $\langle \Phi(u) - \Phi(v), u - v \rangle \geq \eta||u - v||^2$ for all $u, v \in S$ [38].

The variational inequality problem $VIP(\mathcal{F}, C)$ is defined as follows: given $\mathcal{F} : \mathcal{H} \rightarrow \mathcal{H}$ which is monotone over a nonempty closed convex set $C \subset \mathcal{H}$, find $u^* \in C$ such that $\langle v - u^*, \mathcal{F}(u^*) \rangle \geq 0$, $\forall v \in C$. If a function $\Theta : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ is convex over a closed
convex set $C$ and Gâteaux differentiable with derivative $\Theta'$ over an open set $U \supset C$, then $\Theta'$ is paramonotone over $C$. For such a $\Theta$, the set $\Gamma := \{u \in C \mid \Theta'(u) = \inf \Theta'(C)\}$ is nothing but the solution set of $VIP(\Theta', C)$ [33]. Given $\mathcal{F} : \mathcal{H} \to \mathcal{H}$ which is monotone over a nonempty closed convex set $C$, $u^* \in C$ is a solution of $VIP(\mathcal{F}, C)$ if and only if $u^* \in Fix \{ FC (I - \mu \mathcal{F}) \}$ for an arbitrarily fixed $\mu > 0$ (For related mathematical discussion in this section, the readers should consult, e.g., [6, 9, 19, 31–38]).

2 Hybrid Steepest Descent Method

Theorem 1 (Strong convergence for nonexpansive mapping [6, 9]) Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. Suppose that a mapping $\mathcal{F} : \mathcal{H} \to \mathcal{H}$ is $\kappa$-Lipschitzian and $\eta$-strongly monotone over $T(\mathcal{H})$. Then, by using any sequence $(\lambda_n)_{n \geq 1} \subset [0, \infty)$ satisfying (W1) $\lim_{n \to +\infty} \lambda_n = 0$, (W2) $\sum_{n \geq 1} \lambda_n = +\infty$, (W3) $\sum_{n \geq 1} |\lambda_n - \lambda_{n+1}| < +\infty$ or $(\lambda_n)_{n \geq 1} \subset (0, \infty)$ satisfying (L1) $\lim_{n \to +\infty} \lambda_n = 0$, (L2) $\sum_{n \geq 1} \lambda_n = +\infty$, (L3) $\lim_{n \to +\infty} (\lambda_n - \lambda) \lambda_{n+1} = 0$, the sequence $(u_n)_{n \geq 0}$ generated, with arbitrary $u_0 \in \mathcal{H}$, by

$$u_{n+1} := T(u_n) - \lambda_{n+1} \mathcal{F}(T(u_n))$$

converges strongly to the uniquely existing solution of the VIP: find $u^* \in Fix(T)$ such that $(u - u^*, \mathcal{F}(u^*)) \geq 0$, for all $u \in Fix(T)$. (Note: The condition (L3) was relaxed recently to $\lim_{n \to +\infty} \lambda_n = 1$ [11].)

Theorem 1 is a generalization of a fixed point iteration [12–15] so called the anchor method:

$$u_{n+1} := \lambda_{n+1} a + (1 - \lambda_{n+1}) T(u_n),$$

which converges strongly to $P_{Fix(T)}(a)$.

The hybrid steepest descent method (2) can be applied to more general monotone operators [7, 8] if $\dim(\mathcal{H}) < \infty$. Moreover, by the use of slowly changing sequence of nonexpansive mappings having same fixed point sets, a variation of the hybrid steepest descent method is gifted with notable robustness to the numerical errors possibly unavoidable in the iterative computations [9].

The next theorem shows that the hybrid steepest descent method can also be applied to the variational inequality problem over the fixed point set of quasi-nonexpansive mappings.

Definition 2 (Quasi-shrinking mapping [10]) Suppose that $T : \mathcal{H} \to \mathcal{H}$ is quasi- nonexpansive with $Fix(T) \cap C \neq \emptyset$ for some closed convex set $C \subset \mathcal{H}$. Then $T : \mathcal{H} \to \mathcal{H}$ is called quasi-shrinking on $C(\subset \mathcal{H})$ if

$$D : r \in [0, \infty) \mapsto \inf_{u \in Fix(T), r \cap C \neq \emptyset} d(u, Fix(T)) - d(T(u), Fix(T))$$

satisfies $D(r) = 0 \iff r = 0$, where $\triangleright (Fix(T), r \cap C \neq \emptyset) := \{x \in \mathcal{H} \mid d(x, Fix(T)) \geq r\}$. □

Proposition 3 [10] Suppose that a continuous convex function $\Phi : \mathcal{H} \to \mathbb{R}$ has $lev_{\leq 0} \neq \emptyset$ and bounded subdifferential $\partial \Phi : \mathcal{H} \to 2^\mathcal{H}$, i.e., $\partial \Phi$ maps bounded sets to bounded sets. Define $T_\alpha := (1 - \alpha) I + \alpha T_{sp}(\Phi)$ for $\alpha \in (0, 2]$ [hence $Fix(T_\alpha) = lev_{\leq 0} \Phi$; see (1) for the definition of $T_{sp}(\Phi)$]. Then, we have the followings:

(a) If a selection of subgradient of $\Phi$, say $\Phi' : \mathcal{H} \to \mathcal{H}$, is uniformly monotone over $\mathcal{H}$, then $T_\alpha$ is quasi-shrinking on any nonempty bounded closed convex set $C$ satisfying $C \cap lev_{\leq 0} \Phi \neq \emptyset$. 
(b) Assume \( \dim(\mathcal{H}) < \infty \). Then \( T_\alpha \) is quasi-shrinking on any nonempty bounded closed convex set \( C(\subset \mathcal{H}) \) satisfying \( C \cap \text{lev}_0 \Phi \neq \emptyset \). \( \square \)

**Theorem 4** *(Strong convergence for quasi-shrinking mapping [10])* Suppose that \( T : \mathcal{H} \rightarrow \mathcal{H} \) is a quasi-nonexpansive mapping with \( \text{Fix}(T) \neq \emptyset \). Let \( \mathcal{F} : \mathcal{H} \rightarrow \mathcal{H} \) be \( \kappa \)-Lipschitzian and \( \eta \)-strongly monotone over \( T(\mathcal{H}) \) [Hence VIP(\( \mathcal{F} \), Fix(T)) has its unique solution \( u^* \in \text{Fix}(T) \)]. Suppose also that there exists some \((f, u_0) \in \text{Fix}(T) \times \mathcal{H} \) for which \( T \) is quasi-shrinking on \( C_{f}(u_0) := \{x \in \mathcal{H} | \|x - f\| \leq R_f := \max(\|u_0 - f\|, \frac{\|\mu \mathcal{F}(f)\|}{1 - \sqrt{1 - \mu(2\eta - \mu\kappa^2)})} \} \).

Then for any \( \mu \in (0, \frac{2\eta}{\kappa^2}) \) and any \( (\lambda_n)_{n \geq 1} \subset [0, 1] \) satisfying \( (H1) \lim_{n \rightarrow \infty} \lambda_n = 0 \), and \( (H2) \sum_{n \geq 1} \lambda_n = \infty \), the sequence \((u_n)_{n \geq 0} \), generated by \( u_{n+1} := T(u_n) - \lambda_{n+1}\mu \mathcal{F}(T(u_n)) \), converges strongly to \( u^* \). \( \square \)

If \( \dim(\mathcal{H}) < \infty \), in a way similar to the discussions in [7-9], we can generalize Theorem 4 for application to more general monotone operators [10].

### 3 Adaptive Projected Subgradient Method

**Theorem 5** *(Adaptive Projected Subgradient Method [21, 22])* Let \( \Theta_n : \mathcal{H} \rightarrow [0, \infty) \) \((\forall n \in \mathbb{N})\) be a sequence of continuous convex functions and \( K \subset \mathcal{H} \) a nonempty closed convex set. For an arbitrarily given \( u_0 \in K \), the adaptive projected subgradient method produces a sequence \((u_n)_{n \in \mathbb{N}} \subset K \) by

\[
\begin{align*}
 u_{n+1} := \begin{cases} 
 P_K \left( u_n - \lambda_n \frac{\Theta_n[u_n]}{\|\Theta_n(u_n)\|^2} \Theta_n'(u_n) \right) & \text{if } \Theta_n'(u_n) \neq 0, \\
 u_n & \text{otherwise,}
\end{cases}
\end{align*}
\]

where \( \Theta_n'(u_n) \in \partial \Theta_n(u_n) \) and \( 0 \leq \lambda_n \leq 2 \). Then the sequence \((u_n)_{n \in \mathbb{N}} \) satisfies the followings.

(a) *(Monotone approximation)* Suppose that

\[
\forall u_n \notin \Omega_n := \{u \in K | \Theta_n(u) = \Theta_n^*\} \neq \emptyset,
\]

where \( \Theta_n^* := \inf_{u \in K} \Theta_n(u). \)

Then, by using \( \forall \lambda_n \in \left(0, 2 \left(1 - \frac{\Theta_n^*}{\Theta_n(u_n)}\right)\right) \), we have

\[
\forall u^{*(n)} \in \Omega_n, \|u_{n+1} - u^{*(n)}\| < \|u_n - u^{*(n)}\|.
\]

(b) *(Boundedness, Asymptotic optimality)* Suppose

\[
\exists N_0 \in \mathbb{N} \text{ s.t. } \begin{cases} 
 \Theta_n^* = 0, \forall n \geq N_0 \text{ and } \\
 \Omega := \bigcap_{n \geq N_0} \Omega_n \neq \emptyset.
\end{cases}
\]

Then \((u_n)_{n \in \mathbb{N}} \) is bounded. Moreover if we specially use \( \forall \lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2) \), we have \( \lim_{n \rightarrow \infty} \Theta_n(u_n) = 0 \) provided that \((\Theta_n'(u_n))_{n \in \mathbb{N}} \) is bounded.

\( ^1 \)In this case, \( \Theta_n(u_n) > \Theta_n^* \geq 0. \)
(c) (Strong convergence) Assume (3) and \( \Omega \) has some relative interior w.r.t. a hyperplane \( \Pi(\subset H) \), i.e., there exist \( \tilde{u} \in \Pi \cap \Omega \) and \( \varepsilon > 0 \) satisfying \( \{ v \in \Pi \mid \| v - \tilde{u} \| \leq \varepsilon \} \subset \Omega \). Then, by using \( \forall \lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2) \), \( (u_n)_{n \in \mathbb{N}} \) converges strongly to some \( \tilde{u} \in K \), i.e., \( \lim_{n \rightarrow \infty} \| u_n - \tilde{u} \| = 0 \). Moreover \( \lim_{n \rightarrow \infty} \Theta_n(\tilde{u}) = 0 \) if (i) \( (\Theta'_n(u_n))_{n \in \mathbb{N}} \) is bounded and (ii) there exists bounded \( (\Theta'_n(\tilde{u}))_{n \in \mathbb{N}} \) where \( \Theta'_n(\tilde{u}) \in \partial \Theta_n(\tilde{u}), \forall n \in \mathbb{N} \).

(d) (A characterization of \( \tilde{u} \)) Assume the existence of some interior \( \tilde{u} \) of \( \Omega \), i.e., there exists \( \rho > 0 \) satisfying \( \{ v \in H \mid \| v - \tilde{u} \| \leq \rho \} \subset \Omega \). In addition to the conditions (i) and (ii) in (c), assume that there exists \( \delta > 0 \) satisfying

\[
\forall n \geq N_0, \forall u \in \Gamma \setminus (\text{lev}_{\leq 0} \Theta_n), \exists \Theta'_n(u) \in \partial \Theta_n(u), \| \Theta'_n(u) \| \geq \delta,
\]

where \( \Gamma := \{(1-s)\tilde{u} + s \bar{u} \in H \mid s \in (0, 1)\} \). Then, by using \( \forall \lambda_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2) \),

\[
\lim_{n \rightarrow \infty} u_n =: \tilde{u} \in \liminf_{n \rightarrow \infty} \Omega_n, \text{ where } \liminf_{n \rightarrow \infty} \Omega_n \text{ stands for the closure of } \liminf_{n \rightarrow \infty} \Omega_n := \bigcup_{n \geq 0} \bigcap_{k \geq n} \Omega_k.
\]

4 Concluding Remarks

In this paper, we briefly present central results on the hybrid steepest descent method and the adaptive projected subgradient method recently developed by our research group. For detailed mathematical discussions of the methods and their applications to inverse problems and signal processing problems, see [3–10, 16, 17, 21–23, 30] and references therein.

References


