不動点制約付き非平滑凸最適化

飯塚　秀明

有限個の閉凸集合の共通部分や凸関数の最小解集合といった，複雑な制約集合上での凸最適化は，ネットワーク資源割り当てや機械学習などに現れる実問題解決の強力な手段となっている．この凸最適化を可能とする方法の一つは，複雑な制約集合を準非拡大写像と呼ばれる計算可能な非線形写像の不動点集合として表現し，不動点を見つけるためのアルゴリズムを利用する方法である．本稿では，準非拡大写像の不動点制約付き非平滑凸最適化問題を解くための不動点最適化アルゴリズムとアンサンブル学習への応用に関する筆者の最近の取り組みを紹介する．

キーワード：準非拡大写像，非拡大写像，不動点，不動点最適化アルゴリズム，アンサンブル学習

1. はじめに

凸最適化理論は, オペレーションズ・リサーチや工学に現れる実最適化問題を解決できる強力な手段の一つである．特に, この理論により, 複雑な制約集合上の凸最小化問題を解くことが可能となる．複雑な制約集合とは，たとえば, 有限個の閉凸集合の共通部分, 凸関数の最小解集合, 変分不等式の解集合のような, 距離射影を計算することが一般に困難な集合のことを言う．このような複雑な制約集合の多くは，ある計算可能な**非拡大写像の不動点集合**として表現ができる．たとえば, 信号処理 [1, 2], 逆問題 [3], 最適制御 [4], ネットワーク資源割り当て [5, 6], 機械学習 [7] などに現れる実最適化問題の制約集合は, 非拡大写像の不動点集合として表現できる．これらの例が示すように, 不動点集合は豊富な表現力をもつと言える．

非拡大写像の不動点集合上で微分可能な凸関数を最小化する問題（平滑凸最適化問題と呼ぶ）に関する有用な手法は，ハイブリッド最急降下法 [8] である．制約なし平滑最適化のための最急降下法と不動点を見つけるための手法 [9, 10] に基づいたこの手法は, 信号処理分野に現れる複雑な最適化問題 [2, 3] の解決に成功している．その後, ハイブリッド最急降下法の加速法 [1, 11, 12] がいくつか提案された．その一方で, 非拡大写像よりも広いクラスに分類される**準非拡大写像**の不動点集合上での平滑凸最適化に関する結果 [13] が報告されている．これは, 準非拡大写像の不動点集合の枠組みでなければ表現ができない制約集合，たとえば，非平滑凸関数のレベル集合 [14, 15] のような複雑な制約下での平滑凸最適化を可能にする．その他の興味ある関心事は, 目的関数が**非平滑凸**であるときに適用可能な最適化手法を構築できるかどうか, であろう．なぜならば, 非平滑凸最適化問題は, 信号復元 [16], 帯域幅割り当て [17], 機械学習 [18, Chapter 4] といった重要な応用例を含んでいるからである．

本稿では，準非拡大写像の不動点集合上の非平滑凸最適化問題 [19]（問題 2.2）について議論をする．そして，この問題を解くための**不動点最適化**アルゴリズム [19]（アルゴリズム 3.1）を紹介する．提案アルゴリズムは，凸目的関数の劣勾配[1]と増分型アルゴリズム [20] のアイデアを利用することで非平滑凸最適化に適用可能となる．加えて，提案アルゴリズムの収束解析についても紹介する [19]（定理 3.1）．さらに，提案アルゴリズムをアンサンブル学習（例題 2.2）に適用し，その収束性や有用性について検証する．

2. 不動点問題

以下の準非拡大写像に関する不動点問題 [21, Chapter 4]，[22, Chapter 3]，[23, Chapter 3] を考察する．

問題 2.1（不動点問題）．$\mathbb{R}^N$ を内積 $\langle\cdot,\cdot\rangle$ とノルム $\|\cdot\|$ をもつ N 次元ユークリッド空間とし, 写像 $T\colon \mathbb{R}^N \to \mathbb{R}^N$ は**準非拡大** (quasi-nonexpansive)，すなわち，任意の

[1] 凸関数 g_i の $x \in \mathbb{R}^N$ での劣微分 [24, Section 23] とは，$\partial g_i(x) := \{u \in \mathbb{R}^N\colon g_i(y) \geq g_i(x) + \langle y-x, u\rangle\ (y \in \mathbb{R}^N)\}$ のことを言う．$u_i(x) \in \partial g_i(x)$ を g_i の $x \in \mathbb{R}^N$ での劣勾配と呼ぶ．たとえば，$g_i(x) := \|x\|\ (x \in \mathbb{R}^N)$ の劣微分は，

$$\partial\|\cdot\|(x) = \begin{cases} \left\{\dfrac{x}{\|x\|}\right\} & (x \neq 0), \\ B(0;1) & (x = 0) \end{cases}$$

のように書ける [21, Example 16.25]．ただし，$B(0;1) \subset \mathbb{R}^N$ は中心 0，半径 1 の閉球を表す．その他の計算可能な劣勾配を有する凸関数の例については，[21, Chapter 16] を参照せよ．

いいづか　ひであき
明治大学理工学部情報科学科
〒 214–8571 神奈川県川崎市多摩区東三田 1–1–1
iiduka@cs.meiji.ac.jp

$x \in \mathbb{R}^N$ と $y \in \mathrm{Fix}(T)$ に対して，

$$\|T(x) - y\| \le \|x - y\|$$

を満たすものとする．ただし，$\mathrm{Fix}(T)$ は T の不動点集合

$$\mathrm{Fix}(T) := \left\{x^\star \in \mathbb{R}^N : x^\star = T(x^\star)\right\}$$

を表す．このとき, T の不動点 $x^\star = T(x^\star)$ を見つけよ．

任意の $x, y \in \mathbb{R}^N$ に対して, $\|T(x) - T(y)\| \le \|x - y\|$ を満たす写像 T を**非拡大** (nonexpansive) 写像と呼ぶ．一般に，非拡大写像は準非拡大性を満たす．また，T が**準堅非拡大** (quasi-firmly nonexpansive) であるとは，任意の $x \in \mathbb{R}^N$ と $y \in \mathrm{Fix}(T)$ に対して，

$$\|T(x) - y\|^2 \le \|x - y\|^2 - \|(\mathrm{Id} - T)(x)\|^2$$

を満たすときを言う．ただし，Id は $\mathbb{R}^N$ 上の恒等写像とする．準非拡大写像 T が与えられるとき，$S := (1/2)(\mathrm{Id} + T)$ と定義される写像 S は準堅非拡大性 [21, Proposition 4.2] を満たし，$\mathrm{Fix}(S) = \mathrm{Fix}(T)$ が成立する．すなわち，計算可能な準非拡大写像から，計算可能な準堅非拡大写像を容易に構成することができる．ここで，計算可能な準非拡大写像に関する問題 2.1 の例を紹介する．

例題 2.1（非平滑凸関数のレベル集合問題 [14, 15]）**.** $g_i : \mathbb{R}^N \to \mathbb{R}$ $(i = 1, 2, \ldots, m)$ を（必ずしも微分可能とは限らない）凸関数とし，g_i の $x \in \mathbb{R}^N$ での劣勾配は効率的に計算可能であるとする．g_i のレベル集合を $\mathrm{lev}_{\le 0} g_i := \{x \in \mathbb{R}^N : g_i(x) \le 0\}$ とおく．このとき，

$$\bigcap_{i=1}^{m} \mathrm{lev}_{\le 0} g_i \ (\neq \emptyset) \tag{1}$$

に属する点 $x^\star \in \mathbb{R}^N$ を見つける問題を考察する．(1) で定義された集合への距離射影は一般には計算が困難である[2]．$u_i(x)$ を g_i の $x \in \mathbb{R}^N$ での劣勾配とし, 写像 T_i を

$$T_i(x) := \begin{cases} x - \dfrac{g_i(x)}{\|u_i(x)\|^2} u_i(x) & (g_i(x) > 0), \\ x & (g_i(x) \le 0) \end{cases} \tag{2}$$

で定義する（(2) で定義される T_i は，g_i に関する劣勾配射影 [15, Proposition 2.3] と呼ばれる）．このとき，T_i は計算可能な準非拡大写像となり，$\mathrm{Fix}(T_i) = \mathrm{lev}_{\le 0} g_i$ を満たす [14, Lemma 3.1]．ここで，$(w_i)_{i=1}^m \subset (0, 1)$ は $\sum_{i=1}^m w_i = 1$ を満たす点列とし，任意の $x \in \mathbb{R}^N$ に対して，

$$T(x) := \sum_{i=1}^{m} w_i T_i(x) \tag{3}$$

とすると，(3) で定義される T は計算可能な準非拡大写像となり，以下を満足する[3]．

$$\bigcap_{i=1}^{m} \mathrm{lev}_{\le 0} g_i = \mathrm{Fix}\left(\sum_{i=1}^{m} w_i T_i\right) = \mathrm{Fix}(T).$$

例題 2.1 は，ある複雑な集合に属する点を見つける問題が，計算可能な準非拡大写像に関する不動点問題（問題 2.1）として定式化できることを示している（その他の例題については，たとえば，[25, 例題 2.1(i),(ii)] を参照されたい）．本稿では，準非拡大写像 T が直接，利用できるものと仮定し，T を利用することにより構成されるアルゴリズムについて考察する．

問題 2.1 を解くための不動点アルゴリズムは数多く開発されている [9, 10, 26–28] が，以下で構成される Krasnosel'skiĭ–Mann 不動点アルゴリズム [26, 27] は最も単純で有用な手法の一つであろう．$x_0 \in \mathbb{R}^N$, $\alpha \in (0, 1)$ が与えられたとき，任意の $n \in \mathbb{N}$ に対して，

$$\begin{aligned} x_{n+1} = T_\alpha(x_n) &:= [\alpha \mathrm{Id} + (1 - \alpha) T](x_n) \\ &= \alpha x_n + (1 - \alpha) T(x_n). \end{aligned} \tag{4}$$

このアルゴリズムは，各反復について，T で構成される $T_\alpha := \alpha \mathrm{Id} + (1 - \alpha) T$ を一回利用することで点列を構成していることがわかる．Krasnosel'skiĭ–Mann 不動点アルゴリズム (4) で生成される点列 $(x_n)_{n \in \mathbb{N}}$ は T の不動点に収束することが保証されている [21, Theorem 5.14]．その他の有用なアルゴリズムとしては，任意の $n \in \mathbb{N}$ に対して，

$$\begin{aligned} x_{n+1} &:= T(x_n) - \alpha_n (T(x_n) - x_0) \\ &= \alpha_n x_0 + (1 - \alpha_n) T(x_n) \end{aligned} \tag{5}$$

[2] 空でない閉凸集合 C への距離射影 P は，任意の $x \in \mathbb{R}^N$ に対して，$P(x) \in C$，$\|x - P(x)\| = \inf_{y \in C} \|x - y\|$ で定義される．たとえば，半空間 $C := \{x \in \mathbb{R}^N : \langle a, x \rangle \le r\}$（ただし，$a\ (\neq 0) \in \mathbb{R}^N$，$r \in \mathbb{R}$）への距離射影 P は以下のように計算可能である．

$$P(x) = \begin{cases} x + \dfrac{r - \langle a, x \rangle}{\|a\|^2} a & (x \notin C), \\ x & (x \in C). \end{cases}$$

一方で，(1) のような複雑な形状をもつ集合への距離射影の計算は，一般には容易ではない．

[3] 一つ目の等式は [21, Proposition 4.34] に依る．

で構成される Halpern 不動点アルゴリズム [9, 10] が挙げられる．ただし，$(\alpha_n)_{n\in\mathbb{N}} \subset [0,1]$ は $\lim_{n\to\infty}\alpha_n = 0$, $\sum_{n=0}^{\infty}\alpha_n = \infty$ を満たす点列とする．Halpern 不動点アルゴリズム (5) で生成される点列 $(x_n)_{n\in\mathbb{N}}$ は T の不動点集合上の $f_0(x) := (1/2)\|x - x_0\|^2$ $(x \in \mathbb{R}^N)$ の最小解に収束する [29, Theorem 6.19]．アルゴリズム (5) は，初期点 x_0 と $T(x_n)$ との凸結合で点列を構成しているのに対し，アルゴリズム (4) は x_n と $T(x_n)$ との凸結合から成る．さらに，アルゴリズム (5) は，$\nabla f_0(x) = x - x_0$ から，以下のように表現できる．

$$x_{n+1} = T(x_n) - \alpha_n \nabla f_0\left(T(x_n)\right). \tag{6}$$

例題 2.1 にも見られるように，ある複雑な凸制約集合は，計算可能な準非拡大写像の不動点集合として表現することができる．このことから，準非拡大写像の不動点集合上の凸最適化問題の導入（以下，問題 2.2）は，複雑な凸制約集合上の凸最適化の考察を可能とする[4]．

問題 2.2（不動点集合上の非平滑凸最適化問題）**.** I 個の写像 $T_i\colon \mathbb{R}^N \to \mathbb{R}^N$ $(i \in \mathcal{I} := \{1, 2, \ldots, I\})$ と関数 $f_i\colon \mathbb{R}^N \to \mathbb{R}$ $(i \in \mathcal{I})$ が以下の条件を満たすとする．

仮定 2.1.

(A1) $T_i\colon \mathbb{R}^N \to \mathbb{R}^N$ は $X := \bigcap_{i\in\mathcal{I}} \mathrm{Fix}(T_i) \neq \emptyset$ を満たす計算可能な準堅非拡大写像である．

(A2) $f_i\colon \mathbb{R}^N \to \mathbb{R}$ は（必ずしも微分可能とは限らない）凸関数である．

(A3) f_i の劣勾配は計算可能とする．

このとき，

$$\begin{aligned} x^\star &\in \operatorname*{argmin}_{x\in X} \sum_{i\in\mathcal{I}} f_i(x) \\ &:= \left\{ x^\star \in X\colon \sum_{i\in\mathcal{I}} f_i(x^\star) = \min_{x\in X} \sum_{i\in\mathcal{I}} f_i(x) \right\} \end{aligned}$$

を見つけよ．

ここで，問題 2.2 の例を紹介する．

例題 2.2（アンサンブル学習 [32, 33]）**.** クラス分類に対するアンサンブル学習に関して，各事例 a は y にラベル付けられているとし，a を K クラスに分類するための N 個の分類器 $(h^n)_{n=1}^N$ が存在するとする．事例 a に対して，分類器 h^n $(n = 1, 2, \ldots, N)$ は判別測定量 $x^n := h^n(a)$，すなわち，$x := (x^n)_{n=1}^N \in \mathbb{R}^N$ を出力する．a に対する分類は，N 個の分類器で得られる出力結果が結合された後，分類器アンサンブルにより決定される．重み付け測定量 [32, (1)], [33, (1)] は，各事例 a に対して，$H(a) := \sum_{n=1}^N w^n x^n = \langle w, x\rangle$ で計算される．ただし，$w^n \in \mathbb{R}$ $(n = 1, 2, \ldots, N)$ は，n 番目の分類器における重みを表し，$w := (w^n)_{n=1}^N \in \mathbb{R}^N$ とする．標本集合 $\{(a_m, y_m)\}_{m=1}^M$ に対して，分類器により，訓練集合 $S := \{(x_m, y_m)\}_{m=1}^M$ が与えられる．ただし，$x_m := (x_m^n)_{n=1}^N \in \mathbb{R}^N$ $(m = 1, 2, \ldots, M)$ であり，各成分 $x_m^n \in \mathbb{R}$ $(n = 1, 2, \ldots, N, m = 1, 2, \ldots, M)$ は m 番目の標本と n 番目の分類器に対する測定量を表す．分類器アンサンブルにおける学習アルゴリズムの主目的 [32, (2), (3)], [33, (2), (4)] は，以下を満たす分類器重み $w^\star \in C_+ := \mathbb{R}_+^N = \{w := (w^n)_{n=1}^N \in \mathbb{R}^N\colon w^n \geq 0\ (n = 1, 2, \ldots, N)\}$ を見つけることである．

$$w^\star \in \operatorname*{argmin}_{w\in C_+} \frac{1}{2}\sum_{m=1}^M \left(\langle x_m, w\rangle - y_m\right)^2. \tag{7}$$

一方で，アンサンブル学習における疎性学習 [32, Subsection 2.2.2], [33, Subsection 3.2.2] のための最適化問題は，目的関数 $f(w) := (1/2)\sum_{m=1}^M (\langle x_m, w\rangle - y_m)^2$ を $C_+ := \mathbb{R}_+^N$ と

$$C_1 := \left\{ w := (w^n)_{n=1}^N\colon \|w\|_1 := \sum_{n=1}^N |w^n| \leq t_1 \right\} \tag{8}$$

との共通部分 $C_+ \cap C_1$ 上で最適化する問題として定式化される．ただし，t_1 は疎性制御パラメータを表す．疎性学習に関する分類器重みは，l_1 ノルム $\|\cdot\|_1$ を組み込むことで学習されることを (8) は示している．ここで，$g_1(w) := \|w\|_1 - t_1$ $(w \in \mathbb{R}^N)$ と定義すると，g_1 は凸関数であり，$C_1 = \mathrm{lev}_{\leq 0}\, g_1 \neq \emptyset$ を満たす．さらに，$g_1(w) := \|w\|_1 - t_1$ $(w \in \mathbb{R}^N)$ のときの (2) で定義される T_1，すなわち，

$$T_1(w) := \begin{cases} w - \dfrac{g_1(w)}{\|u_1(w)\|^2} u_1(w) & (g_1(w) > 0), \\ w & (g_1(w) \leq 0) \end{cases} \tag{9}$$

は $\mathrm{Fix}(T_1) = C_1$ を満たす準非拡大写像となる（例題 2.1 を参照せよ）．$\mathbb{R}^N$ の各点での $g_1(\cdot) := \|\cdot\|_1 - t_1$ の劣勾配は計算可能 [21, Example 16.25] なので，(9) で定義される写像 T_1 は容易に計算できる．ここで，

[4] 非拡大写像の不動点集合上の平滑凸最適化問題とその解法については，[4, 5, 8, 11, 12, 30, 31] を参照されたい．

$\omega_1+\omega_2=1$ を満たす $\omega_1,\omega_2\in(0,1)$ が与えられたとき，写像 T を，任意の $w\in\mathbb{R}^N$ に対して，

$$T(w):=\omega_1P_+(w)+\omega_2T_1(w) \tag{10}$$

と定義する．ただし，P_+ は C_+ への距離射影を表す．このとき，P_+ と (9) で定義される T_1 の準非拡大性から，(10) で定義される写像 T も準非拡大となる [21, Exercise 4.11]．さらに，[21, Proposition 4.34] から，

$$\begin{aligned}\mathrm{Fix}(T)&=\mathrm{Fix}\left(\omega_1P_++\omega_2T_1\right)=\mathrm{Fix}(P_+)\cap\mathrm{Fix}(T_1)\\&=C_+\cap C_1\end{aligned}$$

が成り立つ．以上のことから，疎性学習を考慮した分類器アンサンブル問題は，以下で表現することができる．

$$w^\star\in\underset{w\in\mathrm{Fix}(T)}{\mathrm{argmin}}\ \frac{1}{2}\sum_{m=1}^{M}\left(\langle x_m,w\rangle-y_m\right)^2.$$

3. 不動点最適化アルゴリズム

問題 2.2 を解くために，I ユーザから成るネットワークシステムを考察し，ユーザ i $(\in\mathcal{I})$ は固有の写像 T_i と目的関数 f_i をもつものとする．これにより，ユーザ i と異なるユーザ j は T_i と f_i の陽な形状を知ることができない．非集中型最適化アルゴリズム[5]の構築のために，以下を仮定する．

仮定 3.1 (A4)**.** 各ユーザは隣接ユーザと通信できる．

仮定 3.1 の下での有用な非集中型最適化アルゴリズムの一つに，**増分型アルゴリズム** [20] がある．増分型アルゴリズムとは，ある反復回数 n に対して，ユーザ $(i-1)$ が保有する $x_{n,i-1}\in\mathbb{R}^N$ を隣接ユーザ i に送り，ユーザ i は自身のもつ T_i,f_i と $x_{n,i-1}$ を用いて，$x_{n,i}=x_{n,i}(T_i,f_i,x_{n,i-1})$ により，点列を更新するアルゴリズムを指す．以下は，問題 2.2 を解くための増分型劣勾配アルゴリズム [19, Algorithm 4.1] である．

アルゴリズム 3.1.

Require: $n\in\mathbb{N}$, $\alpha_i\in(0,1)$, $T_{\alpha_i}:=\alpha_i\mathrm{Id}+(1-\alpha_i)T_i$, $(\lambda_n)_{n\in\mathbb{N}}\subset(0,\infty)$

1: $n\leftarrow 0$, $x_0=x_{0,0}\in\mathbb{R}^N$
2: **loop**
3:   **for** $i=1$ to $i=I$ **do**
4:     $g_{n,i}\in\partial f_i\left(T_{\alpha_i}(x_{n,i-1})\right)$
5:     $x_{n,i}:=T_{\alpha_i}(x_{n,i-1})-\lambda_n g_{n,i}$
6:   **end for**
7:   $x_{n+1}=x_{n,I}=x_{n+1,0}$
8:   $n\leftarrow n+1$
9: **end loop**

アルゴリズム 3.1 のステップ 4，および，ステップ 5 は，Halpern 不動点アルゴリズム (5), (6) に基づいた計算法となっている．

アルゴリズム 3.1 の収束解析は以下のとおりである．

定理 3.1 (Theorem 4.1, Theorem 4.2 [19])**.** 仮定 (A1)–(A4) を満たすとし，(A1)' T_i $(i\in\mathcal{I})$ は不動点閉とする．また，アルゴリズム 3.1 で生成される点列 $(x_{n,i})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ は有界[6]であるとする．このとき，以下の (i), (ii) が成立する．

(i) $\lambda_n:=\lambda\in(0,\infty)$ ならば，

$$\begin{aligned}\liminf_{n\to\infty}f\left(x_n\right)\leq f^\star+I\left(\frac{I\sqrt{N_1}N_2}{2}+N_1\right)\lambda\\+N_2\sum_{i\in\mathcal{I}}\sum_{j=1}^{i-1}\sqrt{\frac{\left(1-\alpha_j\right)N_\lambda\lambda}{\alpha_j}}\\+\left(\sqrt{N_1}+N_2\right)\sum_{i\in\mathcal{I}}\sqrt{\frac{\left(1-\alpha_i\right)N_\lambda\lambda}{\alpha_i}}.\end{aligned}$$

ただし，$f:=\sum_{i\in\mathcal{I}}f_i$, $f^\star$ は問題 2.2 の最適値，N_i $(i=1,2)$ は定数であり，N_λ は λ に依存する定数である．

(ii) $(\lambda_n)_{n\in\mathbb{N}}\subset(0,\infty)$ が $\lim_{n\to\infty}\lambda_n=0$, $\sum_{n=0}^{\infty}\lambda_n=\infty$ を満たす[7]ならば，問題 2.2 の解集合に属する $(x_{n,i})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ の集積点が存在する．さらに，f_i $(i\in\mathcal{I})$ のうち一つが狭義凸であれば，$(x_{n,i})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ は問題 2.2 の一意解に収束する．

定理 3.1(ii) の仮定のもとでは，アルゴリズム 3.1 は

---

[5] すべての T_i と f_i の形状を事前に知ることができるときに実装可能なアルゴリズムを集中型最適化アルゴリズムと呼ぶことにする．たとえば，$x_{n+1}:=x_{n+1}(T_1,T_2,\ldots,T_I,f_1,f_2,\ldots,f_I,x_n)$ によって構成されるアルゴリズムは集中型最適化アルゴリズムとなる．

[6] $\mathrm{Fix}(T_i)\subset X_i$ となる有界閉凸集合 X_i が存在するならば，アルゴリズム 3.1 のステップ 5 を $x_{n,i}:=P_{X_i}[T_{\alpha_i}(x_{n,i-1})-\lambda_n g_{n,i}]$ で置き換えることで，$(x_{n,i})_{n\in\mathbb{N}}$ $(i\in\mathcal{I})$ の有界性が保証される．

[7] $\lambda_n:=1/(n+1)^a$ $(a\in(0,1])$ はこれらの条件を満たす．

$$f(x_n) \le f^\star + I\left\{ \left(\sqrt{N_1} + \frac{(I+1)N_2}{2}\right)\sqrt{\frac{N_3}{n+1}} + \left(\frac{(I-1)\sqrt{N_1}N_2}{2} + N_1\right)\lambda_n \right\}$$

を満たす（ただし，N_i $(i = 1, 2, 3)$ は定数）[19, Corollary 4.2]．すなわち，$\lambda_n := 1/(n+1)$ のとき，

$$f(x_n) = f^\star + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$$

を満たす．

4. アンサンブル学習への応用

疎性学習（例題 2.2）と多様性学習 [33, Subsections 3.2.3 and 4.1] を考慮した分類器アンサンブル問題は，目的関数 $f(w) := (1/2)\sum_{m=1}^{M}(\langle x_m, w\rangle - y_m)^2$ を，$C_+ := \mathbb{R}_+^N$, (8) で定義された集合 C_1，そして，

$$C_2 := \left\{ w\colon \sum_{m=1}^{M} \left\{\langle [x_m], w\rangle - \langle x_m, w\rangle^2\right\} \ge t_2 \right\} \tag{11}$$

との共通部分 $C_+ \cap C_1 \cap C_2$ 上で最小化する問題として定式化される．ただし，$[x_m] := ((x_m^n)^2)_{n=1}^N = ((x_m^1)^2, (x_m^2)^2, \dots, (x_m^N)^2)^\top \in \mathbb{R}^N$ であり，t_2 は多様性制御パラメータである．関数 $g_2\colon \mathbb{R}^N \to \mathbb{R}$ を任意の $w \in \mathbb{R}^N$ に対して，$g_2(w) := t_2 - f_{\mathrm{div}}(w) = t_2 - \sum_{m=1}^{M}\{\langle [x_m], w\rangle - \langle x_m, w\rangle^2\}$ と定義する．このとき，g_2 は $C_2 = \mathrm{lev}_{\le 0} g_2$ となる凸関数であり，$g_2(w) := t_2 - f_{\mathrm{div}}(w)$ $(w \in \mathbb{R}^N)$ のときの (2) で定義される写像 T_2 は計算可能な準非拡大写像となる（(9) で定義される T_1 と同等の性質をもつ）．例題 2.2 の議論と T_1, T_2 の不動点閉性[8]から，

$$T := \frac{1}{2}\left[\mathrm{Id} + (\omega_1 P_+ + \omega_2 T_1 + \omega_3 T_2)\right] \tag{12}$$

で定義される写像 T は不動点閉な準堅非拡大となり，$\mathrm{Fix}(T) = C_+ \cap C_1 \cap C_2$ を満たす．定理 3.1(ii) により，$T_{\alpha_i} := \alpha_i \mathrm{Id} + (1-\alpha_i)T$, $f_i(w) := (1/2)(\langle x_i, w\rangle - y_i)^2$ $(i \in \mathcal{I} := \{1, 2, \dots, M\})$[9]に関するアルゴリズム 3.1 で生成される点列 $(x_n)_{n\in\mathbb{N}}$ は分類器アンサンブル問題の解

[8] T_i が不動点閉 (fixed-point closed) [14, Lemma 2.1] であるとは，任意の $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^N$ に対して，$(x_n)_{n\in\mathbb{N}}$ が $x \in \mathbb{R}^N$ に収束し，かつ，$\lim_{n\to\infty}\|x_n - T_i(x_n)\| = 0$ ならば，$x \in \mathrm{Fix}(T_i)$ を満たすときを言う．(2) で定義される劣勾配射影は不動点閉な準非拡大写像である [14, Lemma 3.1]．

[9] $x_i \ne 0$ ならば f_i は狭義凸性を満たす [7, (2.3)]．

表 1 分類問題に用いたデータセット [35]（[34, Table 4.1] からの抜粋）

データセット	事例数	属性	クラス数
breast-cancer	683	10	2
australian	690	14	2
ionosphere	351	34	2
iris	150	4	3
segment	2310	19	7

$$w^\star \in \operatorname*{argmin}_{w\in\mathrm{Fix}(T)} f(w) = \operatorname*{argmin}_{w\in\bigcap_{i=1}^3 C_i} \sum_{i=1}^{M} f_i(w) \tag{13}$$

に収束することが保証される．アルゴリズム 3.1 [10]について，$\omega_i = 1/3$（(12) で使用する），$\alpha_i := 1/2$，初期点 $x_0 = 0$ とし，ステップ幅 $(\lambda_n)_{n\in\mathbb{N}}$ は，定理 3.1 の結果を考慮した

$$\lambda_n := 10^{-2},\ \frac{10^{-2}}{n+1}\ (n \in \mathbb{N})$$

に加えて，Armijo 条件を満たすステップ幅 [34, Subsection 3.3] も用いた[11]．反復回数 n が重みを学習するのに必要な訓練データ数と一致としたときに，アルゴリズム 3.1 を停止した．

疎性，および，多様性学習を考慮した既存アンサンブル法 [33] は，問題 (13) を $C_+ := \mathbb{R}_+^N$ 上で

$$f(w) + \bar{\alpha}\|w\|_1 - \bar{\beta} f_{\mathrm{div}}(w)$$

を最小化する問題（ただし，$\bar{\alpha}, \bar{\beta}$ は [33, Figure 2] で計算される制御パラメータ）に緩和し，この緩和問題の閉形式解

$$w^{*\top} := \frac{1}{1+2\bar{\beta}}\left(\sum_{m=1}^{M}\left(y_m x_m + \bar{\beta}[x_m]\right) - \bar{\alpha} I\right)^\top \times \left(\sum_{m=1}^{M}\left(x_m x_m^\top\right)\right)^{-1} \tag{14}$$

を利用する方法である．ただし，$I := (1, 1, \dots, 1)^\top \in \mathbb{R}^N$ とする．行列 $\sum_{m=1}^{M}(x_m x_m^\top)$ が非正則ならば，$(\sum_{m=1}^{M}(x_m x_m^\top))^{-1}$ を $\sum_{m=1}^{M}(x_m x_m^\top)$ の疑似逆行列で

[10] この実験では，アルゴリズム 3.1 の拡張アルゴリズムである確率的不動点最適化アルゴリズム [34, Algorithm 1] を用いた．

[11] 定理 3.1(ii) の条件を満たす最も単純なステップ幅は，$\lambda_n = c/(n+1)$ $(n \in \mathbb{N})$（ただし，$c > 0$）であるが，高速に収束するように前もって適切な c を設定することはとても難しい [7, 34]．なぜならば，適切な c の値は，問題の次元数，目的関数や制約集合の形状，劣勾配の選択などといったさまざまな要因に依存しているからである．この問題点を解決する一つの方法が直線探索法を用いてステップ幅を決定することである．

表 2　分類精度と実行時間（[34, Table 4.4] からの抜粋）

	Alg.(14) [33]		Alg.3.1（定数幅）		Alg.3.1（減少幅）		Alg.3.1（直線探索）	
	精度（%）	時間（秒）	精度（%）	時間（秒）	精度（%）	時間（秒）	精度（%）	時間（秒）
breast-cancer	74.29	0.017	94.62	0.749	84.38	0.710	95.16	0.879
australian	59.95	0.018	86.06	0.777	79.53	0.751	85.86	0.936
ionosphere	69.07	0.038	73.66	0.337	69.93	0.341	73.66	0.486
iris	69.58	0.027	70.92	0.391	66.72	0.420	77.56	0.520
segment	15.24	0.166	72.91	29.16	45.04	29.49	71.16	32.24
...	...	...	...	...	...	...	...	...
平均	53.82	1.076	69.46	5.353	61.25	5.433	70.12	6.087
t 検定	–	–	0.055	–	0.123	–	0.039	–

置き換えることにする．

評価指標は，アンサンブル法の分類精度と実行時間とした．加えて，マイクロソフト・エクセルで実装されている関数 T.TEST を用いて，アンサンブル法の平均精度を評価した．関数 T.TEST はスチューデントの t 検定における確率を返し，有意水準は 5%である．すなわち，関数の値が 0.05 よりも小さいならば，(14) を用いた既存アンサンブル法 [33] とアルゴリズム 3.1 を用いたアンサンブル法との間には有意差があると言える．実験で使用した計算機は，Intel(R) Core(TM) i7-5650U CPU，および，二つの 4 GB 1600 MHz DDR3 メモリ（RAM は 8 GB），Mac OS X El Capitan (Version 10.11.6) オペレーティング・システムを備えた 13 インチ MacBook Air である．この実験で用いたアルゴリズムは Python 3.5.3 で実装された．また，実験で使用したデータセットは，LIBSVM [35] データセットから選んだ（表 1）．データセットに関する検証については，10-分割交差検証を用いた．基本学習器はサポートベクターマシンであり，アンサンブルは 100 個のサポートベクターマシン分類器を Bagging により構成した．

表 2 は，実験で使用したアンサンブル法に関する分類精度と実行時間を示している．アルゴリズム 3.1 を用いたアンサンブル法は既存法と比べ実行時間が多くなるが，分類精度は向上していることがわかる．ステップ幅 $\lambda_n = 10^{-2}$ を用いた手法（Alg.3.1（定数幅））やステップ幅 $\lambda_n = 10^{-2}/(n+1)$ を用いた手法（Alg.3.1（減少幅））の性能は良いときもあれば悪いときもある．これは，適切なステップ幅の選び方がデータセットの情報に依存しているからである．一方で，Armijo 直線探索法によるステップ幅を用いた手法（Alg.3.1（直線探索））は，全体的に高い精度を持ち，かつ，既存法と比べて有意差がある（関数 T.TEST の値が 0.039）と言える．

5. まとめ

本稿では，準非拡大写像の不動点制約下での非平滑凸最適化問題を考察し，アンサンブル学習に現れる実問題がこの最適化問題の実例になることを紹介した．次に，この最適化問題を解くための不動点最適化アルゴリズムとその収束解析を与えた．ある条件下では，提案アルゴリズムで生成される点列は問題の解に収束することが保証される．提案アルゴリズムの性能を評価するために，アンサンブル学習に関する既存法との数値比較を行った．数値結果から，提案アルゴリズムの有用性を示すことができた．

次世代ネットワークやビッグデータ解析などに登場する確率的最適化の研究が急速に発展していることから，不動点理論と確率的最適化とを融合する新解法とその収束解析に関する研究 [34] は，今後の進展が期待される研究の一つと言えよう．

謝辞　本稿を発表する機会を与えていただいた専修大学の高野先生，筑波大学の吉瀬先生に深く感謝申し上げます．なお，本研究は，日本学術振興会 科学研究費補助金 基盤研究 (C) (15K04763) の補助を受けている．

参考文献

[1] P. L. Combettes, “A block-iterative surrogate constraint splitting methodfor quadratic signal recovery,” *IEEE Transactions on Signal Processing*, **51**, pp. 1771–1782, 2003.
[2] K. Slavakis and I. Yamada, “Robust wideband beamforming by the hybrid steepest descent method,” *IEEE Transactions on Signal Processing*, **55**, pp. 4511–4522, 2007.
[3] M. Yamagishi and I. Yamada, “Nonexpansiveness of a linearized augmented Lagrangian operator for hierarchical convex optimization,” *Inverse Problems*, **33**, No. 044003, 2017.

[4] H. Iiduka and I. Yamada, "Computational method for solving a stochastic linear-quadratic control problem given an unsolvable stochastic algebraic Riccati equation," *SIAM Journal on Control and Optimization*, **50**, pp. 2173–2192, 2012.
[5] H. Iiduka, "Fixed point optimization algorithm and its application to power control in CDMA data networks," *Mathematical Programming*, **133**, pp. 227–242, 2012.
[6] H. Iiduka, "Fixed point optimization algorithms for distributed optimization in networked systems," *SIAM Journal on Optimization*, **23**, pp. 1–26, 2013.
[7] Y. Hayashi and H. Iiduka, "Optimality and convergence for convex ensemble learning with sparsity and diversity based on fixed point optimization," *Neurocomputing*, **273**, pp. 367–372, 2018.
[8] I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," *Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications*, D. Butnariu, Y. Censor and S. Reich (eds.), Elsevier, pp. 473–504, 2001.
[9] B. Halpern, "Fixed points of nonexpanding maps," *Bulletin of the American Mathematical Society*, **73**, pp. 957–961,1967.
[10] R. Wittmann, "Approximation of fixed points of nonexpansive mappings," *Archiv der Mathematik*, **58**, pp. 486–491, 1992.
[11] H. Iiduka, "Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping," *Mathematical Programming*, **149**, pp. 131–165, 2015.
[12] H. Iiduka and I. Yamada, "A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping," *SIAM Journal on Optimization*, **19**, pp. 1881–1893, 2009.
[13] I. Yamada and N. Ogura, "Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings," *Numerical Functional Analysis and Optimization*, **25**, pp. 619–655, 2004.
[14] H. H. Bauschke and J. Chen, "A projection method for approximating fixed points of quasinonexpansive mappings without the usual demiclosedness condition," *Journal of Nonlinear and Convex Analysis*, **15**, pp. 129–135, 2014.
[15] H. H. Bauschke and P. L. Combettes, "A weak-to-strong convergence principle for Fejér monotone methods in Hilbert space," *Mathematics of Operations Research*, **26**, pp. 248–264, 2001.
[16] P. L. Combettes and J. C. Pesquet, "A Douglas-Rachford splitting approach tononsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, **1**, pp. 564–574, 2007.
[17] H. Iiduka and K. Sakurai, "Decentralized optimization for network resource allocation with nonsmooth utility functions," submitted.
[18] T. Hastie, R. Tibshirani and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press, Boca Raton, 2015.
[19] H. Iiduka, "Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings," *Mathematical Programming*, **159**, pp. 509–538, 2016.
[20] A. Nedić and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, **12**, pp. 109–138, 2001.
[21] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
[22] K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory*, *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, 1990.
[23] W. Takahashi, *Nonlinear Functional Analysis*, Yokohama Publishers, 2000.
[24] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
[25] 飯塚秀明,"不動点制約付き非平滑凸最適化とその応用,"日本オペレーションズ・リサーチ学会　数理計画研究部会第 29 回 RAMP シンポジウム論文集, pp. 125–142, 2017.
[26] M. A. Krasnosel'skiĭ, "Two remarks on the method of successive approximations," *Uspekhi Matematicheskikh Nauk*, **10**, pp. 123–127, 1955.
[27] W. R. Mann, "Mean value methods in iteration," In *Proceedings of American Mathematical Society*, **4**, pp. 506–510, 1953.
[28] K. Nakajo and W. Takahashi, "Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups," *Journal of Mathematical Analysis and Applications*, **279**, pp. 372–379, 2003.
[29] V. Berinde, *Iterative Approximation of Fixed Points*, Springer, 2007.
[30] P. L. Combettes and P. Bondon, "Hard-constrained inconsistent signal feasibility problems," *IEEE Transactions on Signal Processing*, **47**, pp. 2460–2468, 1999.
[31] H. Iiduka, "Iterative algorithm for triple-hierarchical constrained nonconvex optimization problem and its application to network bandwidth allocation," *SIAM Journal on Optimization*, **22**, pp. 862–878, 2012.
[32] X. C. Yin, K. Huang, H. W. Hao, K. Iqbal and Z. B. Wang, "A novel classifier ensemble method with sparsity and diversity," *Neurocomputing*, **134**, pp. 214–221, 2014.
[33] X. C. Yin, K. Huang, C. Yang and H. W. Hao, "Convex ensemble learning with sparsity and diversity," *Information Fusion*, **20**, pp. 49–58, 2014.
[34] H. Iiduka, "Stochastic fixed point optimization algorithm and its applications to classifier ensemble," submitted.
[35] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, **2**, 27:1–27:27, 2011. https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/（2017 年 5 月 1 日閲覧）