

近接分離アルゴリズムとその応用

—信号処理・画像处理的観点から—

小野 峻佑

情報系分野において、凸最適化技術の一つである近接分離アルゴリズムが活発に応用されるようになった。その理由はさまざまであるが、誤解を恐れず一言で述べると、微分できない目的関数や多様な制約条件を含む大規模な最適化問題を効率的に解けるからである。本稿では、代表的な近接分離アルゴリズムに関して、信号処理・画像処理の応用家の観点から解説を試みるとともに、最近の動向や発展的な話題についても触れる。

キーワード：凸最適化、近接分離アルゴリズム、信号処理、画像処理、スパース性

1. 序論

信号処理・画像処理・機械学習などの情報科学・工学分野において、凸最適化技術は問題解決のための基礎ツールとして中心的な役割を担ってきた。特に、関数の勾配情報のみを用いて最適化を行う一次法と呼ばれるクラスのアルゴリズムがこれらの分野では重宝される。その理由としては、1. 変数のサイズ（例：画像処理であれば画素数）が非常に大きく、ヘッセ行列などの情報を利用することが計算量的に厳しいケースが多い、2. アプリケーションによっては、精度の高い解を必要としない（例：画像として見た目があまり変わらなければOK）ことなどが挙げられる。

2005年以降、当該分野において「凸関数ではあるものの微分不可能な関数（例： l_1 ノルム）を含む大規模な最適化問題」の効率的な求解を要請する技術が爆発的に増加した。圧縮センシング、スパース符号化、画像復元、ロバスト推定などがその代表例である。これらの問題に対しては、当然ながら目的関数の微分可能性を前提としたアルゴリズム全般が利用できない。このような「応用上有用ではあるが微分不可能な凸関数」はほかにも多数存在する。また、それらの関数と行列の合成関数を含む目的関数を制約条件とともに最適化するような難度の高い問題を解かねばならない応用も多い。

この困難な状況を一変させたのが、近接分離と呼ばれる原理に基づく最適化アルゴリズム群である（以降、近接分離アルゴリズムと呼ぶ）。近接分離アルゴリズムでは共通して“近接写像 (proximity operator)”と呼

ばれる手続きが鍵となる。定義は2節に譲るが、近接写像は微分できない凸関数に対してもその名のとおり“写像”として一意に定義でき、（微分可能な凸関数に対する）勾配降下操作に代わる役割を果たすことができる。近接分離アルゴリズムは今や当該分野では必要不可欠なツールとなってきており、その理由は以下の2点に集約される。一つ目は、 l_1 ノルムなどの微分不可能な凸関数の近接写像が（多少の誤差を許容しつつ）効率的に計算可能であること、二つ目は、複数の凸関数の和を最適化するような複雑な問題であっても、適切な式変形を経由することで個々の関数の近接写像の計算から構成されるアルゴリズムによって解ける問題に帰着できることである。

代表的な近接分離アルゴリズムとしては、近接勾配法 (proximal gradient method, forward-backward splitting method) [1–3]、交互方向乗数法 (alternating direction method of multipliers, ADMM) [4–6]、主-双対近接分離法 (primal-dual splitting method) [7–10] などが挙げられる。その歴史は古く、アルゴリズムの原型は1970年代から研究されているが、その工学的有用性が明らかとなったのは、これらの手法がスパース推定や画像復元などに応用され始めた2005年以降である（たとえば、[3, 11–14]）。これらの研究を皮切りに、近接分離アルゴリズムの応用は爆発的な広がりを見せ、信号処理・画像処理・機械学習以外にもAI・通信・制御など、多種多様な分野において現在も増え続けている（詳しくは、近接分離アルゴリズムに関するレビュー論文 [15–17] などを参照されたい）。

以上を踏まえ、本稿では、前述した3種類の近接分離アルゴリズムに関して、それがどのように応用されているかという観点も踏まえつつ解説を試みる。まず次節で、凸関数や近接写像などの基本的な道具を導入

おの しゅんすけ
東京工業大学情報理工学院
〒226-8503 神奈川県横浜市緑区長津田町 4259 G3-52

する。次に 3~5 節で、近接勾配法、交互方向乗数法、主-双対近接分離法についてそれぞれ詳細に説明し、その応用についても述べる。さらに、より発展的な内容や最近の動向について 6 節で紹介し、最後に、7 節で本稿をまとめる。

2. 準備

本節では、近接分離アルゴリズムについて解説するうえで必要となる最小限の数理的道具を紹介する。より詳しく知りたい読者は文献 [18] などを参照されたい。

2.1 基本的な道具

関数 $f: \mathbb{R}^N \rightarrow (-\infty, \infty]$ が¹、任意の $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ と $\lambda \in [0, 1]$ に対して $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$ を満たすとき、関数 f を凸関数という。凸関数 f の実効定義域 $\text{dom}(f) := \{\mathbf{x} \in \mathbb{R}^N | f(\mathbf{x}) < \infty\}$ が空集合でないとき、 f を真凸関数と呼ぶ。真凸関数 f のレベル集合 $\text{lev}_{\leq \alpha}(f) := \{\mathbf{x} \in \mathbb{R}^N | f(\mathbf{x}) \leq \alpha\}$ が任意の $\alpha \in \mathbb{R}$ について閉集合となると f を下半連続な真凸関数という。以降、 \mathbb{R}^N 上のすべての下半連続な真凸関数の集合を $\Gamma_0(\mathbb{R}^N)$ で表す。関数 $f \in \Gamma_0(\mathbb{R}^N)$ に対して定義される $f^*(\mathbf{y}) := \sup_{\mathbf{x} \in \mathbb{R}^N} \{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})\}$ ($\langle \cdot, \cdot \rangle$ はユークリッド内積) を f の凸共役関数といい、 $f^* \in \Gamma_0(\mathbb{R}^N)$ となる。

集合 $C \subset \mathbb{R}^N$ が任意の $\mathbf{x}, \mathbf{y} \in C$ と $\lambda \in [0, 1]$ に対して $\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in C$ を満たすとき、集合 C を凸集合という。また、凸集合が閉集合であるとき、閉凸集合という。空でない閉凸集合 C に対して指示関数 (indicator function) $\iota_C: \mathbb{R}^N \rightarrow (-\infty, \infty]$ を

$$\iota_C(\mathbf{x}) := \begin{cases} 0, & \text{if } \mathbf{x} \in C, \\ \infty, & \text{otherwise} \end{cases} \quad (1)$$

のように定義すると、 $\iota_C \in \Gamma_0(\mathbb{R}^N)$ となる。

任意の $\gamma > 0$ に対し、 $f \in \Gamma_0(\mathbb{R}^N)$ の近接写像 (proximity operator/proximal mapping) は

$$\text{prox}_{\gamma f}(\mathbf{x}) := \underset{\mathbf{y} \in \mathbb{R}^N}{\text{argmin}} f(\mathbf{y}) + \frac{1}{2\gamma} \|\mathbf{x} - \mathbf{y}\|^2 \quad (2)$$

で定義される。ただし、 $\|\cdot\|$ は ℓ_2 ノルムである。関数 $\frac{1}{2\gamma} \|\mathbf{x} - \mathbf{y}\|^2$ は強圧的¹かつ狭義凸 (凸関数の定義中の不等号 \geq が $>$ で成立) であるため、写像先の存在性と一意性がそれぞれ保証される。

近接写像に関する有用な性質を二つ紹介する。

¹ 関数 f に対し、 $\|\mathbf{x}\| \rightarrow \infty$ ならば $f(\mathbf{x}) \rightarrow \infty$ であるとき、 f は強圧的 (coercive) であるという。

・ (凸共役関数の近接写像) 関数 $f \in \Gamma_0(\mathbb{R}^N)$ の凸共役関数 f^* の近接写像は

$$\text{prox}_{\gamma f^*}(\mathbf{x}) = \mathbf{x} - \gamma \text{prox}_{\frac{1}{\gamma} f}\left(\frac{1}{\gamma} \mathbf{x}\right) \quad (3)$$

のように f の近接写像を用いて表現できる。

・ (近接写像の分離性) 関数 $\varphi_k \in \Gamma_0(\mathbb{R}^{N_k})$ に対し直積空間 $\mathcal{X} := \mathbb{R}^{N_1} \times \dots \times \mathbb{R}^{N_K}$ 上で関数 $\varphi(\mathbf{x}_1, \dots, \mathbf{x}_K) := \sum_{k=1}^K \varphi_k(\mathbf{x}_k) \in \Gamma_0(\mathcal{X})$ を定義すると、以下が成立する。

$$\begin{aligned} \text{prox}_{\gamma \varphi}(\mathbf{x}_1, \dots, \mathbf{x}_K) \\ = (\text{prox}_{\gamma \varphi_1}(\mathbf{x}_1), \dots, \text{prox}_{\gamma \varphi_K}(\mathbf{x}_K)) \end{aligned} \quad (4)$$

2.2 計算可能な近接写像の例

近接写像が (現実的な時間で) 計算できるかどうかは、当然ながら f に依存する。幸いなことに、応用上有用な非可微分凸関数のうち、近接写像が効率的に計算可能なもの (prox 可能 (proximable) と呼ぶ) が数多く存在することが知られている。本稿で取り上げる応用で重要な役割を果たすいくつかの prox 可能な凸関数について、その計算方法とともに紹介しよう。

まずは、冒頭でも登場した ℓ_1 ノルムである。定義は $\|\mathbf{x}\|_1 := \sum_{i=1}^N |x_i|$ (x_i は $\mathbf{x} \in \mathbb{R}^N$ の i 番目の要素) であり、スパース性の評価指標²として幅広く採用されている (たとえば、[20, 21])。 ℓ_1 ノルムの近接写像は、 $i = 1, \dots, N$ に対し、次の形で与えられる。

$$[\text{prox}_{\gamma \|\cdot\|_1}(\mathbf{x})]_i = \text{sgn}(x_i) \max\{|x_i| - \gamma, 0\} \quad (5)$$

ここで $\text{sgn}(\cdot)$ は (\cdot) の符号を表す。これは、入力ベクトル $\mathbf{x} \in \mathbb{R}^N$ の各要素の絶対値を削るソフト閾値処理 (soft-thresholding) と等価である [22]。

次に、グループ構造をもつスパース性を評価する際によく用いられる (たとえば、[23, 24]) 混合 $\ell_{1,2}$ ノルム (mixed $\ell_{1,2}$ norm) を紹介しよう³。定義は $\|\mathbf{x}\|_{1,2} := \sum_{\mathfrak{g} \in \mathfrak{G}} \|\mathbf{x}_{\mathfrak{g}}\|$ であり、ここで $\mathbf{x}_{\mathfrak{g}}$ は $\mathbf{x} \in \mathbb{R}^N$ の要素を重複のないグループに分割した際の各グループベクトルを表す (\mathfrak{G} はグループのインデックス集合)。近接写像の計算は以下のようなグループごとにスケー

² ℓ_1 ノルムは、スパース性 (ごく一部の要素しか値をもたない) の理想的な評価指標である ℓ_0 擬ノルム (ベクトルの非ゼロ要素の個数として定義される非凸関数) の下界の中で最大の凸関数になっているため、最適化が困難である ℓ_0 擬ノルムの代わりに用いられることが多い。実際に、スパースな信号を少数の観測データから再構成 (圧縮センシング) する際に、観測過程を表す行列に対するある条件のもとで、 ℓ_0 擬ノルム最小化と ℓ_1 ノルム最小化の解が一致することが知られている (たとえば、[19] などを参照)。

³ $\ell_{2,1}$ と表記する場合もある。

リングされたソフト閾値処理になり、各 $\mathbf{g} \in \mathfrak{G}$ に対し、 ℓ_1 ノルムのとときと同様に効率的に計算できる [3].

$$[\text{prox}_{\gamma \|\cdot\|_{1,2}}(\mathbf{x})]_{\mathbf{g}} = \max \left\{ 1 - \frac{\gamma}{\|\mathbf{x}_{\mathbf{g}}\|}, 0 \right\} \mathbf{x}_{\mathbf{g}} \quad (6)$$

今度は、入力ベクトルではなく行列の場合を考える。行列 $\mathbf{X} \in \mathbb{R}^{M \times N}$ ($M \leq N$ とする) の i 番目に大きい特異値を $\sigma_i(\mathbf{X})$ で表したとき、 $\|\mathbf{X}\|_* := \sum_{i=1}^M \sigma_i(\mathbf{X})$ で定義される核ノルム/核型ノルム (nuclear norm) が有用な非可微分凸関数として知られている (たとえば, [25]). その理由は、核型ノルムが行列の低ランク性を評価する際に妥当な指標となるからである⁴. その近接写像は特異値に対するソフト閾値処理に帰着される [27]. すなわち、行列 \mathbf{X} の特異値分解を $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$ とすると、

$$\text{prox}_{\gamma \|\cdot\|_*}(\mathbf{X}) = \mathbf{U}\tilde{\Sigma}_{\gamma}\mathbf{V}^T \quad (7)$$

となる。ここで、 $\tilde{\Sigma}_{\gamma}$ は $\max\{\sigma_i(\mathbf{X}) - \gamma, 0\}$ ($i = 1, \dots, M$) を要素としてもつ対角行列である。

最後に、空でない閉凸集合 $C \subset \mathbb{R}^N$ の指示関数 ι_C の近接写像について説明しよう。定義 (2) において $f := \iota_C$ とすると、入力ベクトル $\mathbf{x} \in \mathbb{R}^N$ と任意の $\gamma > 0$ に対して、

$$\begin{aligned} \text{prox}_{\gamma \iota_C}(\mathbf{x}) &= \underset{\mathbf{y} \in C}{\text{argmin}} \frac{1}{2\gamma} \|\mathbf{x} - \mathbf{y}\|^2 \\ &= \underset{\mathbf{y} \in C}{\text{argmin}} \|\mathbf{x} - \mathbf{y}\| =: P_C(\mathbf{x}) \end{aligned} \quad (8)$$

となる。これは、入力ベクトル \mathbf{x} とのユークリッド距離が最小になる C 内の一点を見つけることに相当し、距離射影/凸射影 (metric/convex projection) と呼ばれる操作 (P_C で表される) にほかならない。距離射影が計算可能かどうかは C の形状に依存する。以下に P_C が計算可能な C の例を紹介する。

- ・ 範囲制約 (box constraint) : $C := [a, b]^N$ ($a < b$) で定義される。距離射影はこの範囲外の要素を a か b どちらか近いほうの値に置き換えるだけである。
- ・ ℓ_2 ノルム球 (ℓ_2 -norm ball) : $C := \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x} - \mathbf{b}\| \leq \varepsilon\}$ ($\varepsilon > 0, \mathbf{b} \in \mathbb{R}^N$) で定義される。距離射影は球の外にあるベクトル \mathbf{x} に対し $P_C(\mathbf{x}) = \mathbf{b} + \frac{\varepsilon(\mathbf{x} - \mathbf{b})}{\|\mathbf{x} - \mathbf{b}\|}$ で与えられる。
- ・ ℓ_1 ノルム球 (ℓ_1 -norm ball) : $C := \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x} - \mathbf{b}\|_1 \leq \varepsilon\}$ ($\varepsilon > 0, \mathbf{b} \in \mathbb{R}^N$) で定義される。距離射影は $O(N)$ の手続きで終了するアルゴリズム [28] によって効率的に計算可能である。

⁴ ℓ_1 ノルムと ℓ_0 擬ノルムの関係と同様に、核型ノルムは行列のランク関数の下界の中で最大の凸関数となる [26].

3. 近接勾配法

まずは近接分離アルゴリズムの中で最も基本的なものである近接勾配法 (proximal gradient method, forward-backward splitting method) [1–3] について紹介する。なお、以降の節で登場する凸最適化問題はすべて解が存在すると仮定する⁵。

次のような問題を考えよう。

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}) \quad (9)$$

ただし、 $f, g \in \Gamma_0(\mathbb{R}^N)$ であり、 f は微分可能かつその勾配 ∇f が β -リプシッツ連続であるとする、つまり

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\| \quad (\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^N) \quad (10)$$

を満たす $\beta > 0$ が存在する。

近接勾配法はこの問題の最適解を求めるための反復アルゴリズムであり、Algorithm 1 で与えられる。

Algorithm 1. 近接勾配法

```

input :  $\mathbf{x}^{(0)}, \gamma \in (0, \frac{2}{\beta})$ 
1 while 停止条件を満たすまで do
2    $\mathbf{x}^{(n+1)} = \text{prox}_{\gamma g}(\mathbf{x}^{(n)} - \gamma \nabla f(\mathbf{x}^{(n)}))$ ;
3    $n \leftarrow n + 1$ ;
output:  $\mathbf{x}^{(n)}$ 

```

その手続きは至ってシンプルである。まず、微分可能な関数 f の勾配方向に更新し、その後、関数 g の近接写像を計算する。その際のステップサイズ幅は ∇f のリプシッツ定数により定まる⁶。以下に近接勾配法の収束定理を示す。

定理 1 (Bauschke and Combettes [18]). Algorithm 1 により生成される点列 $(\mathbf{x}^{(n)})_{n \geq 1}$ は問題 (9) の最適解の 1 点に収束する。

証明やより厳密な議論は [3, 18] を参照されたい。一見、非常に単純な構造に見えるこのアルゴリズムがなぜブレイクスルーをもたらしたのか、応用を介して次節で説明する。

⁵ 解の存在性に関しては問題ごとに厳密な条件が存在するが、議論が煩雑になることを避けるため割愛する。詳しく知りたい読者は [18] などを参照されたい。

⁶ 一般にリプシッツ定数を正確に求めることは難しいが、応用で用いられる f のほとんどは ℓ_2 ノルムの二乗と行列の合成関数として表現されるため、その行列の特異値からリプシッツ定数を見積もることが可能である。

3.1 スパース推定への応用

以下のような線形モデルを考える.

$$\mathbf{v} = \Phi \bar{\mathbf{u}} + \mathbf{n} \quad (11)$$

ただし, $\bar{\mathbf{u}} \in \mathbb{R}^N$ はスパース, $\mathbf{n} \in \mathbb{R}^N$ は加法的白色ガウス雑音であると仮定する. このようなモデルは信号処理・統計的推定・機械学習などで頻繁に登場する. たとえば, スパース信号復元問題 (圧縮センシング [21, 29]) の場合は, $\bar{\mathbf{u}}$ が復元すべき信号ベクトル, $\mathbf{v} \in \mathbb{R}^M$ が観測情報, $\Phi \in \mathbb{R}^{M \times N}$ が物理的な観測過程を表す行列になる. 他方, スパース回帰問題 (LASSO [20]) の場合は, $\bar{\mathbf{u}}$ が学習すべきパラメータベクトル, \mathbf{v} が出力データ, Φ が K 個の訓練サンプルベクトル $\phi_k \in \mathbb{R}^N$ を転置し縦に並べた行列に対応する. 上記は, \mathbf{v} (と Φ) からスパースな $\bar{\mathbf{u}}$ を推定する問題とみなすことができ, 以下のような凸最適化問題として定式化される.

$$\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|\Phi \mathbf{u} - \mathbf{v}\|^2 + \lambda \|\mathbf{u}\|_1 \quad (12)$$

ここで, 第一項はデータ項 (\mathbf{v} との整合性を担保), 第二項はスパース項 (\mathbf{u} のスパース性を促進) と呼ばれ, $\lambda > 0$ はスパース項の重み (ハイパーパラメータ) である.

一見単純な構造に見える問題 (12) であるが, ℓ_1 ノルムが微分不可能であるため, 微分可能性を仮定したアルゴリズムでは解くことができない. また, 推定するベクトルの次元 N が 10^4 を超えるようなケースが多いため, 複雑な手続きを要するアルゴリズムの利用が制限される. この難しい状況を打開したのが ISTA (Iterative Shrinkage Thresholding Algorithm) [30] と呼ばれるアルゴリズムであり, 問題 (9) において $f(\mathbf{u}) := \frac{1}{2} \|\Phi \mathbf{u} - \mathbf{v}\|^2$, $g(\mathbf{u}) := \|\mathbf{u}\|_1$ とすることで, 近接勾配法の一実現例として以下のように与えられる.

$$\mathbf{u}^{(n+1)} = \text{prox}_{\gamma \lambda \|\cdot\|_1} (\mathbf{u}^{(n)} - \gamma (\Phi^T \Phi \mathbf{u}^{(n+1)} - \Phi^T \mathbf{v})) \quad (13)$$

式 (5) で紹介したように, ℓ_1 ノルムの近接写像は簡単に計算可能なソフト閾値処理で済み, データ項の勾配も計算可能⁷であるため, 非常に簡潔な手続きを繰り返すだけで問題 (12) を解くことができる.

一般に, 凸関数 f と g が両方 prox 可能であっても, $f+g$ の近接写像は prox 可能と “ならない”. ゆえに, 手続きをいかにして prox 可能な関数ごとに “分離” す

⁷ 行列 Φ が大きく, かつ密な行列である場合は, 行列-ベクトル積自体が重い処理となるため工夫が必要である. これについては 6 節の議論を参照されたい.

るかが肝要であり, その意味で近接勾配法もこのスピリットを体現したアルゴリズムになっている. また, 近接勾配法の収束レートを $O(1/n)$ から $O(1/n^2)$ に大幅に改善するテクニックとして FISTA [13] が知られている. 詳しくは文献 [13, 31, 32]などを参照されたい.

4. 交互方向乗数法

今度は次のような凸最適化問題を考えよう.

$$\min_{\mathbf{x} \in \mathbb{R}^N, \mathbf{z} \in \mathbb{R}^L} g(\mathbf{x}) + h(\mathbf{z}) \quad \text{s.t.} \quad \mathbf{z} = \mathbf{G}\mathbf{x} \quad (14)$$

ただし, $g \in \Gamma_0(\mathbb{R}^N)$, $h \in \Gamma_0(\mathbb{R}^L)$, $\mathbf{G} \in \mathbb{R}^{L \times N}$ である. 任意の初期値 $\mathbf{z}^{(0)}, \mathbf{y}^{(0)} \in \mathbb{R}^L$ と $\gamma > 0$ に対し, 交互方向乗数法 (alternating direction method of multipliers, ADMM) [4–6] は, 問題 (14) の解を Algorithm 2 により求める (\mathbf{y} は双対変数に相当する).

Algorithm 2. 交互方向乗数法 (ADMM)

```

input :  $\mathbf{z}^{(0)}, \mathbf{y}^{(0)}, \gamma > 0$ 
1 while 停止条件を満たすまで do
2    $\mathbf{x}^{(n+1)} = \underset{\mathbf{x}}{\text{argmin}} g(\mathbf{x}) + \frac{1}{2\gamma} \|\mathbf{z}^{(n)} - \mathbf{G}\mathbf{x} - \mathbf{y}^{(n)}\|^2;$ 
3    $\mathbf{z}^{(n+1)} = \text{prox}_{\gamma h} (\mathbf{G}\mathbf{x}^{(n+1)} + \mathbf{y}^{(n)});$ 
4    $\mathbf{y}^{(n+1)} = \mathbf{y}^{(n)} + \mathbf{G}\mathbf{x}^{(n+1)} - \mathbf{z}^{(n+1)};$ 
5    $n \leftarrow n + 1;$ 
output:  $\mathbf{x}^{(n)}$ 

```

交互方向乗数法の代表的な収束定理を紹介する.

定理 2 (Eckstein and Bertsekas [5]). 問題 (14) のラグランジュ関数 $\mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{y}) := g(\mathbf{x}) + h(\mathbf{z}) + \langle \mathbf{y}, \mathbf{G}\mathbf{x} - \mathbf{z} \rangle$ の鞍点が存在し, \mathbf{G} が列フルランク (full column rank), すなわち, $\mathbf{G}^T \mathbf{G}$ が可逆であると仮定する. このとき, Algorithm 2 で生成される点列 $(\mathbf{x}^{(n)}, \mathbf{z}^{(n)})_{n \geq 1}$ は問題 (14) の最適解の 1 点へ収束する.

Algorithm 2 において, ステップ 2 は関数 g と二次関数の和の最小化となっている. ここで注意すべきは二次関数中の目的変数 \mathbf{x} に行列 \mathbf{G} がかかっていることである. これが単位行列であれば, 定義 (2) から, ステップ 2 は関数 g の近接写像になるが, \mathbf{G} の存在のため近接写像よりも最適化が困難になる. ゆえに, 関数 g として許容できる (= このステップが計算できる) ものは同じ二次関数程度となる. 一方, ステップ 3 は関数 h の近接写像そのものとなっているため, h が prox 可能であればよい. この条件を踏まえたうえで, 応用上非常に重要な式変形のテクニックを紹介する.

4.1 変数分離を介した式変形

次のような凸最適化問題を考える.

$$\min_{\mathbf{x} \in \mathbb{R}^N} \sum_{i=1}^I \varphi_i(\mathbf{A}_i \mathbf{x}) \text{ s.t. } \begin{cases} \mathbf{B}_1 \mathbf{x} \in C_1 \\ \vdots \\ \mathbf{B}_J \mathbf{x} \in C_J \end{cases} \quad (15)$$

ただし, $\varphi_i \in \Gamma_0(\mathbb{R}^{N_i})$ は prox 可能, $\mathbf{A}_i \in \mathbb{R}^{N_i \times N}$, $\mathbf{B}_j \in \mathbb{R}^{M_j \times N}$, $C_j \subset \mathbb{R}^{M_j}$ は距離射影が計算可能な閉凸集合である. まず, 各 C_j の指示関数 ι_{C_j} を導入し, 問題 (15) を以下のように書き換える

$$\min_{\mathbf{x} \in \mathbb{R}^N} \sum_{i=1}^I \varphi_i(\mathbf{A}_i \mathbf{x}) + \sum_{j=1}^J \iota_{C_j}(\mathbf{B}_j \mathbf{x}) \quad (16)$$

次に, 補助変数 $\mathbf{z}_1, \dots, \mathbf{z}_{I+J}$ を用意し, 各関数の変数を置き換えることで, すべての項で共有されていた変数を分離する. その際, 元の変数同士の対応関係を線形等式制約により記述する.

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}_1, \dots, \mathbf{z}_{I+J}} \sum_{i=1}^I \varphi_i(\mathbf{z}_i) + \sum_{j=1}^J \iota_{C_j}(\mathbf{z}_{I+j}) \\ \text{s.t. } \begin{cases} \mathbf{A}_1 \mathbf{x} = \mathbf{z}_1, \dots, \mathbf{A}_I \mathbf{x} = \mathbf{z}_I \\ \mathbf{B}_1 \mathbf{x} = \mathbf{z}_{I+1}, \dots, \mathbf{B}_J \mathbf{x} = \mathbf{z}_{I+J} \end{cases} \end{aligned} \quad (17)$$

ここで, 変数 $\mathbf{z} := (\mathbf{z}_1^\top \dots \mathbf{z}_{I+J}^\top)^\top$ と行列 $\mathbf{G} := (\mathbf{A}_1^\top \dots \mathbf{A}_I^\top \quad \mathbf{B}_1^\top \dots \mathbf{B}_J^\top)^\top$ を用意し, $h(\mathbf{z}) := \sum_{i=1}^I \varphi_i(\mathbf{z}_i) + \sum_{j=1}^J \iota_{C_j}(\mathbf{z}_{I+j})$ と定義すると, 問題 (17) は交互方向乗数法が対象とする問題の標準形 (14) と同一視できる (ただし $g(\mathbf{x}) := 0$). さらにこのとき, Algorithm 2 の各ステップの計算は,

- ・ (ステップ 2) $g(\mathbf{x}) = 0$ であることから, 二次関数の最小化に帰着される.
- ・ (ステップ 3) $h(\mathbf{z})$ の定義から, 性質 (4) が使え, 各 φ_i と ι_{C_j} の近接写像の計算に分解できる.

以上のような手続きにより, 行列が合成された多数の関数と制約条件から構成される問題を交互方向乗数法によって求解できる形までもっていくことが可能となる. 結果として, 前節で紹介した近接勾配法に比べて非常に幅広いクラスの問題を解くことができる. ただし, ステップ 2 の計算において線型方程式系を解く必要があるため, \mathbf{G} の構造によってはこのステップがボトルネックとなることに留意されたい.

4.2 ロバスト主成分分析への応用

交互方向乗数法は, 画像復元 [14, 33], 無線センサネットワーク [34, 35] などへの応用を皮切りに, 信号

処理・コンピュータビジョン・機械学習など, 幅広い分野で活用されている [6, 15, 17]. 本節では汎用性の高い技術であるロバスト主成分分析 [36] を題材にする.

高次元データに内在する本質的な低次元構造を抽出・解析する方法として, 主成分分析 (principal component analysis, PCA) [37] は長らく中心的な役割を果たしてきた. 主成分分析では, M 次元データベクトル N 個を横に並べて構成したデータ行列 $\mathbf{M} \in \mathbb{R}^{M \times N}$ が, ある低ランク行列 (主成分) によって良好に近似できると仮定しデータを分析する. 主成分分析は汎用的な手法であるが, 誤差がガウス分布に従うことを想定しているため, データに外れ値や欠損値が混入している場合に性能が一気に悪化するという欠点があった.

この課題を解決するため, 外れ値・欠損値に対して頑健なロバスト主成分分析 (robust PCA, RPCA) [36, 38] が提案された. ロバスト主成分分析では, 与えられたデータ行列 \mathbf{M} に対し, 低ランク行列とスパースな誤差行列を次の凸最適化問題の解として特徴づける.

$$\min_{\mathbf{L}, \mathbf{S} \in \mathbb{R}^{M \times N}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \text{ subject to } \mathbf{M} = \mathbf{L} + \mathbf{S} \quad (18)$$

2.2 節で触れたように, 核型ノルムと l_1 ノルムはそれぞれランク関数と l_0 擬ノルムの下界の中で最大の凸関数となるため, 結果として問題 (18) は低ランク行列とスパース行列の同時推定問題として自然な定式化となっていることがわかる.

問題 (18) は微分不可能な凸関数を複数含む制約付き最適化問題であるため簡単には解けないが, 4.1 節で紹介したテクニックにより変形することで, 交互方向乗数法によって解ける形に帰着できる. まず, 制約条件 $\mathbf{M} = \mathbf{L} + \mathbf{S}$ を指示関数を用いて目的関数に加える.

$$\min_{\mathbf{L}, \mathbf{S} \in \mathbb{R}^{M \times N}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \iota_{\{\mathbf{M}\}}(\mathbf{L} + \mathbf{S}) \quad (19)$$

ここで $\{\mathbf{M}\}$ は行列 \mathbf{M} 一点からなる閉凸集合を表しており, 元の制約条件と等価であることがわかる. あとは, 補助変数 $\mathbf{Z}_1 := \mathbf{L}$, $\mathbf{Z}_2 := \mathbf{S}$, $\mathbf{Z}_3 := \mathbf{L} + \mathbf{S}$ (行列変数であるため大文字で表記している) を導入し, 4.1 節の手順に従って問題 (14) に帰着させればよい.

実際のアロリズムの計算がどうなるか見ていこう. 今回の場合, 行列 \mathbf{G} が単位行列のみから構成されるシンプルなものになるため, ステップ 2 は以下のように簡単な計算で済むことがわかる.



(a) 入力画像 (b) 低ランク \mathbf{L} (c) スパース \mathbf{S}

図 1 ロバスト主成分分析による画像分離

$$\begin{aligned} \mathbf{L}^{(n+1)} &= \\ \frac{1}{3}(2(\mathbf{Z}_1^{(n)} - \mathbf{Y}_1^{(n)}) - \mathbf{Z}_2^{(n)} + \mathbf{Y}_2^{(n)}) + \mathbf{Z}_3^{(n)} - \mathbf{Y}_3^{(n)} \\ \mathbf{S}^{(n+1)} &= \mathbf{Z}_1^{(n)} - \mathbf{Y}_1^{(n)} + \mathbf{Z}_3^{(n)} - \mathbf{Y}_3^{(n)} - 2\mathbf{L}^{(n+1)} \end{aligned}$$

ステップ 3 は各関数の近接写像を独立に計算すればよい。集合 $\{\mathbf{M}\}$ への距離射影はそのまま \mathbf{M} を代入するだけである。核型ノルムと ℓ_1 ノルムの近接写像に関しては 2.2 節を参照されたい。

ロバスト主成分分析を利用し、れんが壁画像の上に重畳された文字を分離した結果を示す (図 1)。画像を行列とみると、れんが壁は同じパターンの繰り返し = 低ランクになり、文字列は画像全体の一部 = スパースになるため、結果としてそれぞれの成分を綺麗に分離することができる。

5. 主-双対近接分離法

本節では、前節で紹介した交互方向乗数法よりさらに汎用的なアルゴリズムを紹介する⁸。次のような問題を考える。

$$\min_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{G}\mathbf{x}) \quad (20)$$

ただし、 $f, g \in \Gamma_0(\mathbb{R}^N)$, $h \in \Gamma_0(\mathbb{R}^L)$ であり、 f は微分可能かつその勾配 ∇f が β -リプシッツ連続であるとする ($\beta > 0$)。問題 (20) を解くために Condat [9] と Vu [10] により独立に開発されたのが主-双対近接分離法 (primal-dual splitting method) であり、Algorithm 3 で与えられる⁹。

主-双対近接分離法の収束定理を紹介する。

定理 3 (Condat [9])。パラメータ $\gamma_1, \gamma_2 > 0$ が次の不等式を満たすとすする。

⁸ もちろん、近接勾配法や交互方向乗数法にも利点は存在する。アルゴリズムの使い分けに関しては 7 節で触れる。

⁹ 正確には Vu のアルゴリズムのほうがより広いクラスの最適化問題を対象としているが、実用上は問題 (20) で十分なことがほとんどである。また、これらのアルゴリズムは $f = 0$ のケースを解く Chambolle-Pock アルゴリズム [7] を一般化したものである。

Algorithm 3. 主-双対近接分離法

input : $\mathbf{x}^{(0)}, \mathbf{y}^{(0)}, \gamma_1, \gamma_2 > 0$
1 while 停止条件を満たすまで **do**
2 $\mathbf{x}^{(n+1)} =$
 $\text{prox}_{\gamma_1 g}(\mathbf{x}^{(n)} - \gamma_1(\nabla f(\mathbf{x}^{(n)}) + \mathbf{G}^\top \mathbf{y}^{(n)}));$
3 $\mathbf{y}^{(n+1)} =$
 $\text{prox}_{\gamma_2 h^*}(\mathbf{y}^{(n)} + \gamma_2 \mathbf{G}(2\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}));$
4 $n \leftarrow n + 1;$
output: $\mathbf{x}^{(n)}$

$$\gamma_1 \left(\frac{\beta}{2} + \gamma_2 \sigma_1(\mathbf{G}^\top \mathbf{G}) \right) < 1 \quad (21)$$

ただし $\sigma_1(\cdot)$ は最大特異値を表す。このとき、Algorithm 3 で生成される点列 $(\mathbf{x}^{(n)})_{n \geq 1}$ は問題 (20) の最適解の一点へ収束する。

このアルゴリズムの特筆すべき点は、交互方向乗数法と異なり、行列 \mathbf{G} に起因する線型方程式系の求解が完全に回避されている点にある。各ステップをみればわかるように、 f の勾配降下、 g, h の近接写像 (h^* の近接写像は (3) により h のそれを用いて計算可能)、および \mathbf{G} と \mathbf{G}^\top のみでアルゴリズムが構成されている。また、4.1 節で紹介した変数分離のテクニックを関数 $h(\mathbf{G}\mathbf{x})$ に対して用いることで、問題 (15) のような多数の関数・制約条件から成る問題を解くこともできる。

5.1 全変動最小化に基づく画像復元への応用

撮像時に重畳されるセンサノイズ、焦点ぼけ・手ブレ、画素欠損などの劣化を伴う観測データから観測モデルに基づき未知の元画像を推定することを画像復元という。画像復元は、リモートセンシング、医用生体工学、材料工学、天文学などさまざまな分野において重要な課題であり、数多くの研究が行われている。

本節では、数式上の利便性のため、サイズが $N_1 \times N_2$ のグレースケール画像を $N = N_1 N_2$ 次元ベクトルとして扱う (N は画素数に相当する)。観測データ $\mathbf{v} \in \mathbb{R}^M$ が以下のような観測モデルに従って取得された状況を考えよう。

$$\mathbf{v} = \Phi \bar{\mathbf{u}} + \mathbf{n} \quad (22)$$

ここで $\bar{\mathbf{u}} \in \mathbb{R}^N$ は元画像、 $\Phi \in \mathbb{R}^{M \times N}$ は劣化過程を記述する行列 (ぼけ、欠損など)、 $\mathbf{n} \in \mathbb{R}^M$ は加法性白色ガウス雑音である。観測モデル (22) の逆問題である画像復元問題は本質的に不良設定/悪条件であるため、推定対象 (元画像) に関する何らかの“事前知識”を解法に取り入れる必要がある。このようなアプローチを正則化と呼ぶ (その意味で 3.1 節で登場したスパー

ス項も正則化の一種である)。

画像に対する正則化としてポピュラーな手法が、本節で扱う全変動 (total variation, TV) [39] である。画像 $\mathbf{u} \in \mathbb{R}^N$ に対し、縦／横の隣接画素間差分を計算する行列をそれぞれ $\mathbf{D}_v, \mathbf{D}_h \in \mathbb{R}^{N \times N}$ とし、これらを縦に並べた行列を $\mathbf{D} := (\mathbf{D}_v^\top \mathbf{D}_h^\top)^\top \in \mathbb{R}^{2N \times N}$ とすると、全変動は次のように定義される。

$$\text{TV}(\mathbf{u}) := \|\mathbf{D}\mathbf{u}\|_{1,2} = \sum_{i=1}^N \sqrt{d_{v,i}^2 + d_{h,i}^2} \quad (23)$$

ここで混合 $\ell_{1,2}$ ノルムは各画素 ($i = 1, \dots, N$) における縦差分 $d_{v,i}$ と横差分 $d_{h,i}$ の二乗和の平方根 (= ℓ_2 ノルム) をとった後にその総和をとる (= ℓ_1 ノルム) 操作をまとめて表すために用いられている。2.2 節で混合 $\ell_{1,2}$ ノルムを導入した際に言及したように、当該ノルムはグループ構造をもつスパース性を評価する凸関数として妥当なものであった。自然画像は一般に隣接画素間に強い相関があるため、ほとんどの差分値はゼロに近い値をとるが、ある画素の周辺にエッジが存在する場合のみ縦・横差分共に大きな値をとる。つまり、自然画像の縦・横差分画像 (エッジ画像) はグループスパース性 (各グループベクトル = 各画素の縦横差分値) を有すると解釈できるため、これを事前知識として活用するうえで自然な定義となっている。

観測モデル (22) のもと、全変動最小化に基づく画像復元は次のように定式化される。

$$\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|\Phi\mathbf{u} - \mathbf{v}\|^2 + \lambda \|\mathbf{D}\mathbf{u}\|_{1,2} \quad \text{s.t.} \quad \mathbf{u} \in [0, 255]^N \quad (24)$$

ここで第二項はデータ項、制約条件は復元画像のダイナミックレンジ (8bit) を表している。パラメータ $\lambda > 0$ は全変動項とデータ項のバランスを調整する重みであり、ノイズの強さに応じて決定する。

全変動は、prox 可能ではあるが微分不可能な関数 $\|\cdot\|_{1,2}$ に行列 \mathbf{D} が合成された関数であり、そのままでは近接写像を計算できない。これに加えて制約条件 $\mathbf{u} \in [0, 255]^N$ も存在するが、主-双対近接分離法を用いることで効率的な求解が可能となる。

まず、指示関数を用いて問題 (24) 中の制約条件を目的関数に加える。

$$\min_{\mathbf{u} \in \mathbb{R}^N} \frac{1}{2} \|\Phi\mathbf{u} - \mathbf{v}\|^2 + \lambda \|\mathbf{D}\mathbf{u}\|_{1,2} + \iota_{[0,255]^N}(\mathbf{u}) \quad (25)$$

すると、第一項を $h(\mathbf{G}\mathbf{u})$ 、第二項を $f(\mathbf{u})$ 、第 3 項を $g(\mathbf{u})$ とすることで、問題 (20) に帰着でき、そのまま



(a) テスト画像 (b) 劣化画像 (c) 復元結果

図 2 全変動最小化による画像復元

主-双対近接分離法が適用可能となる。必要となる計算は、データ項に関する勾配降下、混合 $\ell_{1,2}$ ノルムの近接写像、および範囲制約への距離射影であり、すべて効率的に計算可能である。

実際にこのアルゴリズムを用いて画像を復元した結果を図 2 に示す。綺麗なテスト画像 (a) に人工的な欠損 (90%) とノイズを加えたものが (b) であり、(b) から全変動最小化により復元した結果が (c) である。エッジを保存しながら良好な復元が実現できている。

本節で紹介した全変動は最も基本的な定義のものであったが、これにはグラデーション領域に階段状の模様が出てしまう、カラー画像にそのまま適用すると色ムラが発生するといった欠点も存在する。これを解決するため、全変動自体の拡張・一般化が数多く提案されている。代表的なものとして、カラー画像用の全変動 [40–43]、二階差分以上を考慮した一般化 [44]、近傍や非局所的な情報を取り入れた一般化 [45–48] などが挙げられる。これらもすべて prox 可能な凸関数と差分を計算する行列の合成関数として表現されるため、近接分離アルゴリズムで扱うことができる。また、全変動に加えて別の正則化を利用することで、画像を区分的平滑な構成成分と細かいテクスチャ成分に分離する手法も研究されている [49–51]。

6. 発展的な話題

本稿で紹介した近接分離アルゴリズムは非常に柔軟なフレームワークであるが、一方で当然ながら扱えないクラスの問題もあり、それらの中には応用的観点から重要なものも存在する。たとえば、グループ型ノルムや非ガウス分布の対数尤度関数を用いて表現される制約条件は、ハイパーパラメータの設定を格段に簡易化するという大きな恩恵をもたらすが、距離射影の計算が困難であるため従来の近接分離アルゴリズムでは扱えなかった。このジレンマを解決するため、制約条件のエピグラフ分解を利用したテクニック [52, 53] や制約条件の不動点表現による特徴づけを利用したアルゴリズム [54] が提案されている。ほかにも、「最適解が

無数に存在する凸最適化問題」の解集合上で別の指標を最適化する階層型の凸最適化問題も、近接分離アルゴリズムでは扱えないが重要な問題のクラスである。これに対しては、ハイブリッド最急降下法と呼ばれる不動点アルゴリズムに基づく解法が提案されている [55–57].

近接分離アルゴリズムに限らずほとんどの反復アルゴリズムは、各反復で問題に含まれる行列と変数ベクトルの積の計算が必要となる。しかし、画像処理などの応用では、行列のサイズ自体が非常に大きくなる（画素数の二乗オーダー）ため、行列の構造によってはこの演算自体が計算量的ボトルネックになる。これに対し、近接分離アルゴリズムを確率化/乱択化/分散化したもの [58–61] を応用することで、計算量を削減しつつ高速化するアプローチが研究されている [62–65]. 確率的最適化についてより詳しく知りたい読者は、同特集の別記事 [66] を参照されたい。

また最近では、近接分離アルゴリズムを交互最適化の部分問題ソルバーとして用いた、柔軟かつ効率的な行列・テンソル因子分解アルゴリズムも提案されている [67–69]. 少ない反復数でそれなりの解が得られる近接分離アルゴリズムの性質ともマッチしている。

7. 結び

情報科学・工学分野において活発に応用されている近接分離アルゴリズムについて解説した。特に、近接勾配法、交互方向乗数法、主-双対近接分離法について詳述したが、最後にこれらのアルゴリズムの使い分けについて簡単に述べておきたい。まず、近接勾配法で解けるクラスの問題（微分可能な凸関数+ prox 可能な凸関数の最小化）であれば、近接勾配法の加速版である FISTA [13] を用いるのが最も効率的である。そうでない場合、基本的には 4.1 節で説明した式変形を行った後、交互方向乗数法が主-双対近接分離法を用いることになる。その際、ステップ 2 の線型方程式系が効率的に解ける（例： \mathbf{G} が高速な変換で対角化可能など）のであれば、交互方向乗数法を採用するとよい。理由は、交互方向乗数法のほうが（経験的に）収束が速く、かつステップサイズの調整が容易だからである。

より発展的な画像処理応用に興味のある読者は、手前味噌ではあるが解説記事 [70] などを参照されたい。また、数学的な厳密性を欠くが、近接分離アルゴリズムを非凸最適化問題に“形式的に”応用した研究 [71–74] や、畳み込みニューラルネットワークや非局所フィルタリングのような最適化に基づかないノイズ除去手法を近接写像と“解釈”することでその適用範囲を広げる

研究 [75–77] なども行われている。今後ますます応用範囲が拡大していきであろう近接分離アルゴリズムを理解するうえで、本稿がその一助となれば幸いである。

参考文献

- [1] G. B. Passty, “Ergodic convergence to a zero of the sum of monotone operators in Hilbert space,” *Journal of Mathematical Analysis and Applications*, **72**, pp. 383–390, 1979.
- [2] P. Tseng, “Applications of a splitting algorithm to decomposition in convex programming and variational inequalities,” *SIAM Journal on Control and Optimization*, **29**, pp. 119–138, 1991.
- [3] P. L. Combettes and V. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling and Simulation*, **4**, pp. 1168–1200, 2005.
- [4] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite elements approximations,” *Computers & Mathematics with Applications*, **2**, pp. 17–40, 1976.
- [5] J. Eckstein and D. Bertsekas, “On the Douglas–Rachford splitting method and proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, **55**, pp. 293–318, 1992.
- [6] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, **3**, pp. 1–122, 2011.
- [7] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging Vision*, **40**, pp. 120–145, 2010.
- [8] P. L. Combettes and J.-C. Pesquet, “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,” *Set-Valued and Variational Analysis*, **20**, pp. 307–330, 2012.
- [9] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *Journal of Optimization Theory and Applications*, **158**, pp. 460–479, 2013.
- [10] B. C. Vu, “A splitting algorithm for dual monotone inclusions involving cocoercive operators,” *Advances in Computational Mathematics*, **38**, pp. 667–681, 2013.
- [11] P. L. Combettes and J.-C. Pesquet, “A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE Journal of Selected Topics in Signal Processing*, **1**, pp. 564–574, 2007.
- [12] F. X. Dupe, J. M. Fadili and J. L. Starck, “A proximal iteration for deconvolving Poisson noisy images using sparse representations,” *IEEE Transactions on Image Processing*, **18**, pp. 310–321, 2009.
- [13] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, **2**, pp. 183–202, 2009.
- [14] T. Goldstein and S. Osher, “The split Bregman method for L1-regularized problems,” *SIAM Journal on Imaging Sciences*, **2**, pp. 323–343, 2009.

- [15] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke and H. Wolkowicz (eds.), Springer-Verlag, pp. 185–212, 2011.
- [16] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, **1**, pp. 127–239, 2014.
- [17] 小野峻佑, "近接分離による分散凸最適化—交互方向乗数法に基づくアプローチを中心として—," *計測と制御*, **55**, pp. 954–959, 2016.
- [18] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [19] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, **25**, pp. 21–30, 2008.
- [20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, pp. 267–288, 1996.
- [21] E. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, **52**, pp. 489–509, 2006.
- [22] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, **41**, pp. 613–627, 1995.
- [23] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, pp. 49–67, 2006.
- [24] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, **55**, pp. 5302–5316, 2009.
- [25] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, **9**, pp. 717–772, 2009.
- [26] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.
- [27] J.-F. Cai, E. Candès and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, **20**, pp. 1956–1982, 2010.
- [28] J. Duchi, S. S-Shwartz, Y. Singer and T. Chandra, "Efficient projections onto the ℓ_1 -ball for learning in high dimensions," In *Proceedings of International Conference on Machine Learning*, pp. 272–279, 2008.
- [29] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, **52**, pp. 1289–1306, 2006.
- [30] I. Daubechies, M. D. Frieese and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, **57**, pp. 1413–1457, 2004.
- [31] M. Yamagishi and I. Yamada, "Over-relaxation of the fast iterative shrinkage-thresholding algorithm with variable stepsize," *Inverse Problems*, **27**, article number: 105008, 2011.
- [32] A. Chambolle and C. Dossal, "On the convergence of the iterates of the 'fast iterative shrinkage/thresholding algorithm'," *Journal of Optimization Theory and Applications*, **166**, pp. 968–982, 2015.
- [33] M. Afonso, J. Bioucas-Dias and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, **19**, pp. 2345–2356, 2010.
- [34] I. D. Schizas, A. Ribeiro and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Transactions on Signal Processing*, **56**, pp. 350–364, 2008.
- [35] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *IEEE Transactions on Signal Processing*, **58**, pp. 3816–3827, 2010.
- [36] E. J. Candès, X. Li, Y. Ma and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, **58**, article number: 11, 2011.
- [37] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.
- [38] J. Wright, A. Ganesh, S. Rao, Y. Peng and Y. Ma "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," In *Proceedings of Neural Information Processing Systems*, pp. 2080–2088, 2009.
- [39] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, **60**, pp. 259–268, 1992.
- [40] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, **2**, pp. 455–484, 2008.
- [41] B. Goldluecke, E. Strelakovsky and D. Cremers, "The natural vectorial total variation which arises from geometric measure theory," *SIAM Journal on Imaging Sciences*, **5**, pp. 537–563, 2012.
- [42] T. Miyata and Y. Sakai, "Total variation defined by weighted L infinity norm for utilizing inter channel dependency," In *Proceedings of IEEE International Conference on Image Processing*, pp. 3057–3060, 2012.
- [43] S. Ono and I. Yamada, "Decorrelated vectorial total variation," In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4090–4097, 2014.
- [44] K. Bredies, K. Kunisch and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, **3**, pp. 492–526, 2010.
- [45] S. Lefkimiatis, A. R. P. Maragos and M. Unser, "Structure tensor total variation," *SIAM Journal on Imaging Sciences*, **8**, pp. 1090–1122, 2015.
- [46] S. Ono, K. Shirai and M. Okuda, "Vectorial total variation based on arranged structure tensor for multichannel image restoration," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 4528–4532, 2016.
- [47] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, **7**, pp. 1005–1028, 2009.
- [48] X. Zhang, M. Burger, X. Bresson and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, **3**, pp. 253–276, 2010.

- [49] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, **19**, pp. 553–572, 2003.
- [50] J.-F. Aujol, G. Gilboa, T. Chan and S. Osher, "Structure-texture image decomposition—modeling, algorithms, and parameter selection," *International Journal of Computer Vision*, **67**, pp. 111–136, 2006.
- [51] S. Ono, T. Miyata and I. Yamada, "Cartoon-texture image decomposition using blockwise low-rank texture characterization," *IEEE Transactions on Image Processing*, **23**, pp. 1128–1142, 2014.
- [52] G. Chierchia, N. Pustelnik, J.-C. Pesquet and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal, Image and Video Processing*, **9**, pp. 1737–1749, 2014.
- [53] S. Ono and I. Yamada, "Second-order total generalized variation constraint," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 4938–4942, 2014.
- [54] S. Ono and I. Yamada, "Signal recovery with certain involved convex data-fidelity constraints," *IEEE Transactions on Signal Processing*, **63**, pp. 6149–6163, 2015.
- [55] I. Yamada, M. Yukawa and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasiconvex mappings," *Fixed-Point Algorithm for Inverse Problems in Science and Engineering*, H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke and H. Wolkowicz (eds.), Springer-Verlag, pp. 345–390, 2011.
- [56] S. Ono and I. Yamada, "Hierarchical convex optimization with primal-dual splitting," *IEEE Transactions on Signal Processing*, **63**, pp. 373–388, 2015.
- [57] M. Yamagishi and I. Yamada, "Nonexpansiveness of a linearized augmented lagrangian operator for hierarchical convex optimization," *Inverse Problems*, **33**, article number: 044003, 2017.
- [58] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *Journal of Nonlinear Convex Analysis*, **16**, pp. 2453–2490, 2015.
- [59] T. Suzuki, "Stochastic dual coordinate ascent with alternating direction method of multipliers," In *Proceedings of International Conference on Machine Learning*, pp. 736–744, 2014.
- [60] Y. Zhang and L. Xiao, "Stochastic primal-dual coordinate method for regularized empirical risk minimization," *The Journal of Machine Learning Research*, **18**, pp. 2939–2980, 2017.
- [61] P. L. Combettes and J. Eckstein, "Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions," *Mathematical Programming*, **168**, pp. 645–672, 2018.
- [62] S. Ono, M. Yamagishi, T. Miyata and I. Kumazawa, "Image restoration using a stochastic variant of the alternating direction method of multipliers," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 4523–4527, 2016.
- [63] P. L. Combettes and J.-C. Pesquet, "Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration," In *Proceedings of European Signal Processing Conference*, pp. 1813–1817, 2016.
- [64] A. Chambolle, M. J. Ehrhardt, P. Richtárik and C.-B. Schonlieb, "Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications," *SIAM Journal on Optimization*, **28**, pp. 2783–2808, 2018.
- [65] S. Ono, "Efficient constrained signal reconstruction by randomized epigraphical projection," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, 2019, to appear.
- [66] 鈴木大慈, "機械学習問題における確率的最適化技法," *オペレーションズ・リサーチ: 経営の科学*, **64**(6), pp. 360–367, 2019.
- [67] K. Huang, N. D. Sidiropoulos and A. P. Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Transactions on Signal Processing*, **64**, pp. 5052–5065, 2016.
- [68] A. P. Liavas, G. Kostoulas, G. Lourakis, K. Huang and N. D. Sidiropoulos, "Nesterov-based alternating optimization for nonnegative tensor factorization: Algorithm and parallel implementation," *IEEE Transactions on Signal Processing*, **66**, pp. 944–953, 2018.
- [69] S. Ono, "Efficient constrained tensor factorization by alternating optimization with primal-dual splitting," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 3379–3383, 2018.
- [70] 白井啓一郎, 馬場達也, 小野峻佑, 奥田正浩, "局所特徴に基づく正則化を用いた画像処理と最適化の役割," *Fundamentals Review*, **11**, pp. 40–53, 2017.
- [71] Z. Wen, C. Yang, X. Liu and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems*, **28**, article number: 115010, 2012.
- [72] D. L. Sun and C. Fevotte, "Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence," In *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 6201–6205, 2014.
- [73] S. Ono and I. Yamada, "Color-line regularization for color artifact removal," *IEEE Transactions on Computational Imaging*, **2**, pp. 204–217, 2016.
- [74] S. Ono, " L_0 gradient projection," *IEEE Transactions on Image Processing*, **26**, pp. 1554–1564, 2017.
- [75] S. V. Venkatakrishnan, C. A. Bouman and B. Wohlberg, "Plug-and-play priors for model based reconstruction," In *Proceedings of IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013.
- [76] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Transactions on Computational Imaging*, **2**, pp. 408–423, 2016.
- [77] S. Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Processing Letters*, **24**, pp. 1108–1112, 2017.