

単調 Lipschitz 連続勾配デノイザに基づく近接分離アルゴリズム — Lipschitz 制約なしアプローチと説明可能性

Plugging Monotone Lipschitz-Gradient Denoiser into Proximal Splitting
Algorithms: A Lipschitz Control Free Approach and Explainability

湯川 正裕^{1*}

Masahiro Yukawa

概要 単調で Lipschitz 連続（定数が 1 より大）な勾配写像を MoL-Grad デノイザと呼ぶ。本稿では、第一に、MoL-Grad デノイザが弱凸関数の（一価の）近接写像を特徴付けることを述べ、Moreau 分解の弱凸関数への拡張を与える。第二に、近接分離アルゴリズムの近接写像を MoL-Grad デノイザで代用した AI アシスト型連続最適化戦略に基づくアルゴリズムによって生成される点列が、デノイザから誘導される暗黙的正則化関数を含む目的関数の最小解に弱収束することを述べる。第三に、集合値写像（多価写像）としての近接写像に関する近接包含定理と近接変換定理を与え、これらに基づく「不連続 shrinkage 作用素から MoL-Grad デノイザへの変換法（連続緩和の方法）」とその応用例を紹介する。

キーワード 近接分離アルゴリズム、弱凸関数、連続緩和、機械学習、説明可能性、Plug-and-Play

1. 単調 Lipschitz 連続勾配（MoL-Grad）デノイザ — AI アシスト型連続最適化戦略

AI 革命とも言われる機械学習技術の目覚ましい進展は、社会に大きな変革を生みだしている。本稿では、その中核を担う深層学習に基づくデータ駆動型アプローチとモデルベースアプローチの架け橋となる、筆者の最近の研究 [1, 2] を紹介する^{*1}。本稿で紹介する内容は、弱凸関数を含む最適化問題を解く手法を扱う点において、最適化コミュニティで活発に研究されている DC 計画問題や、RAMP シンポジウム 2024 における山田功先生（科学大）のご講演 [3] とも関連のあるものであることを冒頭で述べておきたい。

主要ポイント.

1. MoL-Grad デノイザと一価の近接写像の等価性定理 [1] — 弱凸関数の近接写像に関する基礎的結果（2 章・定理 1）

1 慶應義塾大学理工学部電気情報工学科, 〒223-8522 横浜市港北区日吉 3-14-1

Department of Electronics and Electrical Engineering, Keio University, Hiyoshi 3-14-1, Kohoku-ku, Yokohama, Kanagawa, 223-8522 JAPAN

* E-mail address: yukawa@elec.keio.ac.jp

*1 初めからこのような方向性を目指していたというより、「非凸関数の近接写像の中で、最適化の観点から有用なものは、どのような数理的性質を持つか」という素朴な疑問に対して、好奇心に任せて研究し、最後に応用価値について熟慮した結果、幸運にも、データ駆動型アプローチとモデルベースアプローチを繋ぐ成果としてまとめることができた点を付記する。

2. AI アシスト型連続最適化戦略に関する収束定理 [1] — 近接分離アルゴリズムの近接写像を MoL-Grad デノイザに置き換えた手法の収束解析 (3 章・定理 3, 4)
3. 近接包含変換法による連続緩和 [2] — 不連続 shrinkage 作用素から (AI アシスト型連続最適化戦略に利用可能な) MoL-Grad デノイザを生成する方法 (5 章・定理 5, 6)

1.1. ニューラルネットワークを活用した最適化アプローチ

データ駆動型アプローチは、最適化アルゴリズムに基づく従来のモデルベースアプローチを凌駕する性能が達成される例も多く、潜在的な可能性を持つ一方で、解釈性・説明可能性^{*2}が限定的であることや、大量のデータが必要であることなど、解決すべき問題を抱えている [4, 5]。このようなデータ駆動型アプローチの問題点の抜本的解決を目指して、データ駆動型とモデルベースの両アプローチを融合する研究が盛んに行われてきた。本稿では、そのような試みの一つである Plug-and-Play (PnP) 法 [6–8] という技術に焦点を当て、その中で、後述する「単調 Lipschitz 連続勾配デノイザ (定義 2) [1] (以下、簡単のため、MoL-Grad デノイザと呼ぶ)」が最適化という観点から本質的な役割を演じることを説明していく。

本章では、主に画像処理における「デノイザ (denoiser)」を念頭に動機を述べる。デノイザという用語を選んだのは、研究結果の応用価値を分かりやすく伝えるための筆者なりの工夫であり、一般的な数学用語である写像 (あるいは、作用素) に読み替えても差し支えない。つまり、主役は、「単調性と Lipschitz 連続性を持つ勾配作用素」である。画像から雑音 (noise) を除去するのがデノイザの本来の役割であるが、それ以外にも、(特に、画像処理における) 逆問題を解くために有用である [9]。PnP 法は、デノイザを活用して逆問題を解く方法の一つである。逆問題を解くソルバーとして近接分離アルゴリズムが広く用いられているが、PnP 法では、更新式の近接写像をデノイザ (ニューラルネットワーク) に置き換えて (近接写像のところにデノイザを差し込んで (plug)), アルゴリズムを実行する (play)。近接分離アルゴリズムで広く用いられる soft shrinkage (ℓ_1 ノルムの近接写像) [10] は、ノイズ除去のために提案された一種のデノイザであるから、これをもっと高性能なデノイザに置き換えれば、一段と良い性能が達成できるに違いないというわけである。

1.2. Lipschitz 制約なしアプローチを考える意義

Soft shrinkage とともに提案されたのが、hard shrinkage [10] である。これらに対して、その間を取ったような firm shrinkage が初めの二つより高い性能を示すことが知られている (図 1)。hard shrinkage は ℓ_0 擬似ノルムの近接写像^{*3}であり、soft shrinkage は、それぞれ ℓ_1 ノルムの近接写像である。また、firm shrinkage は、ある意味で ℓ_0 擬似ノルムと ℓ_1 ノルムをつなぐ弱凸関数の近接写像である [11] から、これらは、作用を受ける対象のスパース性を促進す

*2 本稿では、「ニューラルネットワーク (デノイザ) を組み込んだアルゴリズムがどういう最適化問題を解いているか」という率直な問いに対して、何らかの明確な形で説明できることを「説明可能性」と定義する。

*3 正確には、 ℓ_0 擬似ノルムの近接写像は集合値写像であり、hard shrinkage は集合の中から一つの要素を選んだものである。このようにして定義される一価写像は、集合値写像のセレクションと呼ばれる。

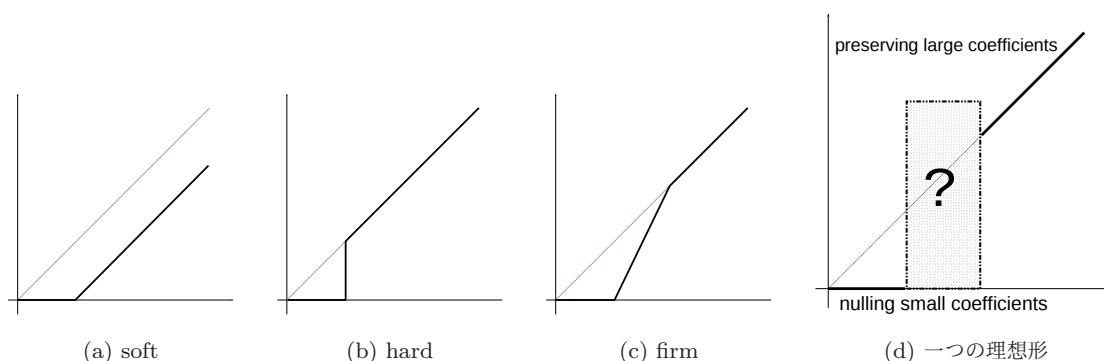


図1 スパース性を促進するために提案された作用素と一つの理想形

る作用素である。スパース性を促進することを考えた場合、理想的には、(i) ゼロに近い成分はゼロにする^{*4}、(ii) ゼロから遠い成分はそのまま残すことが望ましいだろう。これを図示すると、図1(d)のようになる。凸関数の近接写像であり、非拡大性（1-Lipschitz 連続性）を有する soft shrinkage は、条件 (i), (ii) を同時に満足できないことがわかる。このことは、非拡大性を持つ shrinkage のみを考えた場合の一種の限界を示唆すると言えるだろう。一方、hard shrinkage は、条件 (i), (ii) を満たしているが、不連続写像であるため、例えば、同 shrinkage を近接分離アルゴリズムの近接写像として採用した場合、最適解への収束を保証するのは容易でない。firm shrinkage は、条件 (i), (ii) を満たしており、さらに、Lipschitz 連続写像であるため、近接分離アルゴリズムに用いた際、最適解への収束を保証することが可能である。これらの例から分かるように、条件 (i), (ii) を同時に満足する連続写像を定義しようとすると、必然的に傾きが1を超える部分が出てきてしまう。すなわち、理想的なスパース性促進効果を得るためには、写像の Lipschitz 定数が1を超える以外に方法はないのである。

深層学習で得られるデノイザについて直ちに同様の結論を得ることは容易でないかもしれない。しかし、これに類する性質があってもおかしくない、少なくとも、その可能性を簡単に捨てたくない、というのが筆者の信念であり、これが Lipschitz 定数に制約のないアプローチを提唱する意義であると考えている。

1.3. MoL-Grad デノイザを採用した PnP 型最適化戦略

本稿で議論していく話の中心になるリサーチクエスションは、以下の二つである。

1. PnP 法で得られるアルゴリズムが何らかの目的関数を最適化していると解釈できるのは、どのようなクラスのデノイザ（ニューラルネットワーク）を採用したときか。→ 本稿の答えは、MoL-Grad デノイザ（MoL-Grad 型ニューラルネットワーク）である。

*4 ここでの議論は、広義スパースベクトル [3] に拡張しても成立する。ここで、広義スパースベクトルとは、多くの成分値がゼロ付近に集中しているベクトルのことである（信号処理応用では、広義スパースベクトルが所望されることも少なくない）。この場合、ゼロに近い成分は、ゼロにより近い値にすれば十分であり、必ずしも、ぴったりゼロにする必要性はない。

2. そのとき、最適化される目的関数の中で、デノイザに対応する正則化関数は、どのように特徴づけられるか。→ デノイザ $\nabla\psi$ には、正則化関数 $\varphi = \psi^* - (1/2) \|\cdot\|^2$ が対応する（共役関数 ψ^* の定義は、2 章参照）。この正則化関数 φ を、特に、暗黙的正則化関数（implicit regularizer）と呼ぶ。

デノイザと正則化関数の対応は、定理 1 で詳述し、PnP 法ベースのアルゴリズムと最適化については、定理 3、4 で述べる。前述したように、MoL-Grad デノイザは、「単調性と Lipschitz 連続性を持つ勾配作用素」である。以下、これらの性質に着目する理由を簡潔に述べる。まず初めに、「(十分に一般的な仮定の下で) ある点 x が関数 f の最小解であるための必要十分条件は、 $0 \in \partial f(x)$ である (Fermat's rule)*5」ことから、デノイザと最小化問題を結びつけるためには、デノイザと劣勾配を関係付ける必要がある。そこで、デノイザがある関数の劣勾配で表現できるという前提を受け入れることにすると、劣勾配が単調作用素であることから、「単調性」が必要になる。

デノイザの「連続性」の条件を外すのは簡単ではないだろう。デノイザに連続性を仮定すると、自動的に「Fréchet 微分可能な凸関数の勾配」として表現できることになる [12, Proposition 17.41]（つまり、劣勾配は、単集合 (singleton) になる）。MoL-Grad デノイザに課される「Lipschitz 連続性」は、通常の連続性より強い仮定であるが、最適化アルゴリズムの収束性を保証する際に広く用いられる一般的な仮定である（この条件を外した場合、ステップサイズを減少させていくなどの工夫が必要になり、収束速度の低下につながるなど、実用上のデメリットも存在する）。実は、Lipschitz 連続性は、デノイザを近接写像（定義 1）として表現する上で本質的である（定理 1）。

MoL-Grad デノイザを採用した PnP 法ベースの近接分離アルゴリズムとして、前方後方近接分離アルゴリズム (Forward Backward Splitting 法 ; FBS 法) [13, 14] に基づく PnP-FBS 法と主双対近接分離アルゴリズム (Primal Dual Splitting 法 ; PDS 法) [15, 16] に基づく PnP-PDS 法の収束について議論する。近接写像からデノイザへの置き換えによって得られたこれらの手法は、目的関数から設計されたものでないため、一般に、何らかの最適化問題に関係する保証はない（関係していたとしても、その問題を特定するのは容易でないだろう）。しかし、幸いにも、MoL-Grad デノイザを採用した PnP-FBS 法と PnP-PDS 法は、デノイザから誘導される「暗黙的正則化関数」を含むそれぞれの目的関数を最適化する。特に、PnP-PDS 法において、MoL-Grad デノイザの Lipschitz 定数を制限することなく、最適解への収束が保証されることを述べる。MoL-Grad デノイザの例として、既存の非凸関数の近接写像から、Tied-Weight ニューラルネットワーク、バイアス低減型 OSCAR、そして、不連続 shrinkage 作用素の連続緩和（次節）までを紹介する。

*5 ∂f は、凸解析で一般的に用いられる劣勾配である。実ヒルベルト空間 $\mathcal{H}$ 上で定義される関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ の点 $x \in \mathcal{H}$ における劣勾配は、 $\partial f(x) := \{z \in \mathcal{H} \mid \langle y - x, z \rangle + f(x) \leq f(y), \forall y \in \mathcal{H}\}$ で定義される。

1.4. 不連続 shrinkage 作用素の連続緩和 — 近接包含変換法

不連続 shrinkage 作用素のグラフを含む極大単調作用素（厳密には、極大巡回単調作用素）は、不連続 shrinkage を与える非凸正則化関数の「下半連続 1-弱凸包」の近接写像に対応する。これを「近接包含定理（定理 5）」と呼ぶ。そして、この下半連続 1-弱凸包を（1 未満の正定数で）スケーリングした下半連続弱凸関数の近接写像（一価写像）が、極大単調作用素から「簡単な四則演算と逆写像」を通して得られる。これが「近接変換定理（定理 6）」であり、その結果として得られる一価の近接写像が連続緩和を与える。この不連続 shrinkage 作用素から一価の近接写像（連続緩和）への変換法を「近接包含変換法（proximal inclusion-conversion method）」と名付けたい。この変換法の応用として、trimmed ℓ_1 ノルム [17] の近接写像（一般に、集合値写像）の連続緩和を扱った例も紹介する。

2. MoL-Grad デノイザと一価の近接写像

実ヒルベルト空間 $\mathcal{H}$ 上で定義された非凸関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty] := \mathbb{R} \cup \{+\infty\}$ を考えよう*6。 $\mathcal{H}$ 上の内積とノルムを、それぞれ $\langle \cdot, \cdot \rangle$ 、 $\|\cdot\|$ で表す。本稿全体を通して、 $\text{dom } f := \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \emptyset$ （関数 f が真（proper）であること）を仮定する。

共役関数（定義と性質）。関数 $f^* : \mathcal{H} \rightarrow (-\infty, \infty] : x \mapsto \sup_{y \in \mathcal{H}} (\langle x, y \rangle - f(y))$ を凸共役（convex conjugate）という。共役関数 f^* が真となる必要十分条件は、 f 自身が真であり、さらに、 f の下界となる連続なアフィン関数（ $f \geq \langle a, \cdot \rangle + b$ となる $(a, b) \in \mathcal{H} \times \mathbb{R}$ ）が存在することである（[12, Propositions 13.10, 13.12] 参照）。一般に、連続な凸関数 $f : \mathcal{H} \rightarrow \mathbb{R}$ と正定数 $\rho > 0$ について、(i) f が ρ -平滑であることと (ii) f^* が ρ^{-1} -強凸であることは等価である [12, Theorem 18.15]（本稿で登場する関数の関係性を理解する上で重要な性質である）。

$y \in \mathcal{H}$ に関する関数 $f(y) + (1/2) \|x - y\|^2$ ($x \in \mathcal{H}$ は固定) は、一般に、最小値を持たないこともある。最小値は存在するけれど、最小値を達成する y （最小解）が二つ以上存在することもある。そのため、近接写像は、一般に、集合値写像（多価写像）になる（5 章参照）。それでは、近接写像が多価であるとき、複数の最小解 y の中から適当なものを一つ選んで一価写像を定義したとき、この一価写像が連続写像になることはあるだろうか。

命題 1 ([1]). $\mathcal{H}$ 上で定義された関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ に対して、 $T(x) \in \text{argmin}_{y \in \mathcal{H}} [f(y) + (1/2) \|x - y\|^2] \neq \emptyset$ ($x \in \mathcal{H}$) で定義される $T : \mathcal{H} \rightarrow \mathcal{H}$ が連続写像であるとする。このとき、全ての $x \in \mathcal{H}$ に対して、関数 $f + (1/2) \|x - \cdot\|^2$ は、唯一つの最小解を持つ。

命題 1 より、近接写像から連続写像を定義するためには、近接写像は一価である必要性があることが分かった。そのため、一価の近接写像を定義する。

*6 非線形解析に不慣れな読者のための参考書として、[12, 18] をお奨めする。本稿で扱う用語の多くは、[19] でも確認できる。信号処理で用いられる非線形解析の基礎事項について日本語で解説した入門レベルの拙著 [20] も紹介しておく。

定義 1 (s-prox 写像 (一価の近接写像) [1]). $\mathcal{H}$ 上で定義された関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ を考える。任意の $x \in \mathcal{H}$ に対して、 $f + \frac{1}{2\gamma} \|x - \cdot\|^2$ ($\gamma > 0$) が唯一つの最小解を持つものとする。このとき、

$$\text{s-Prox}_{\gamma f} : \mathcal{H} \rightarrow \mathcal{H} : x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2\gamma} \|x - y\|^2 \right) \quad (1)$$

を f の (インデックス γ の) s-prox 写像という。

s-prox 写像は、以下に定義される MoL-Grad デノイザを用いて特徴づけられる。

定義 2 (MoL-Grad デノイザ). 写像 $T : \mathcal{H} \rightarrow \mathcal{H}$ が、 β^{-1} -平滑凸関数^{*7} ψ ($\beta \in (0, 1)$) の勾配として $T = \nabla \psi$ の形で表現できるとき、 T は MoL-Grad デノイザであるという^{*8}。

微分可能な関数の凸性は、勾配の単調性によって特徴づけられる。このことから、MoL-Grad デノイザを特徴付ける鍵となる性質は、(i) 単調性、(ii) Lipschitz 連続性 (厳密には、Lipschitz 定数が 1 より大)、(iii) (Fréchet 微分可能な関数の) 勾配として表現できること^{*9}であり、これらが定義 2 に記した写像の本質的な性質であると考えて、MoL-Grad と命名した。下半連続な真凸関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ 全体の集合を $\Gamma_0(\mathcal{H})$ と表す。話の中核へ入っていく前に、弱凸関数と強凸関数を定義しておく。特に、弱凸関数は、本稿で中心的な役割を担う。

定義 3 (弱凸関数・強凸関数). $\eta \in (-\infty, +\infty]$ に対して、 $h + (\eta/2) \|\cdot\|^2$ が凸関数であるとする。このとき、 $\eta \in (0, +\infty)$ であれば、 h は弱凸関数であり、 $\eta \in (-\infty, 0)$ であれば、 h は強凸関数である。 $\eta = +\infty$ は、 h が弱凸関数でないことを意味する。特に、弱凸 (強凸) の程度を明示したいとき、 η -弱凸関数 ($|\eta|$ -強凸関数) という用語を用いる^{*10}。定義より、 η -弱凸関数 ($\eta > 0$) は、 γ -弱凸 ($\forall \gamma > \eta$) である。

さて、s-prox 写像と MoL-Grad デノイザの関係性について考えよう。 φ を $(1 - \beta)$ -弱凸関数であるとしよう。このとき、 $\check{\varphi} := \varphi + ((1 - \beta)/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ とすると、 $(\varphi + (1/2) \|\cdot\|^2)^* = (\check{\varphi} + (\beta/2) \|\cdot\|^2)^*$ であり、 $\check{\varphi} \in \Gamma_0(\mathcal{H})$ であることに注意して [12, Proposition 14.1] を適用すると、 $(\check{\varphi} + (\beta/2) \|\cdot\|^2)^* = \beta(\check{\varphi}^*)$ が成り立つ。ここで、 $\beta(\cdot)$ は、(インデックス β の) Moreau 包 (Moreau envelope)^{*11}を表す。なお、(凸関数の) Moreau 包は、Fréchet 微分可能な凸関数であり、その勾配は、 $\nabla \beta(\check{\varphi}^*) = \beta^{-1}(\text{Id} - \text{s-Prox}_{\check{\varphi}^*}) = \beta^{-1} \text{s-Prox}_{\check{\varphi}}$ で与えられるため、 $\nabla \beta(\check{\varphi}^*)$ は β^{-1} -Lipschitz 連続、すなわち、 $\beta(\check{\varphi}^*)$ が β^{-1} -平滑凸関数であることが直ちに分かる (凸関数の近接写像 $\text{s-Prox}_{\check{\varphi}}$ の非拡大性に注意)。したがって、 $\psi := \beta(\check{\varphi}^*)$ と

*7 ある $\kappa > 0$ に対して、Fréchet 微分可能な関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ の勾配が Lipschitz 連続性 $\|\nabla f(x) - \nabla f(y)\| \leq \kappa \|x - y\|$ ($\forall x, y \in \mathcal{H}$) を満たすとき、 f は κ -平滑関数であるという。

*8 MoL-Grad は、Monotone Lipschitz-Gradient の略称である。

*9 ベクトル解析の概念を用いれば、 $\mathbb{R}^n$ で定義される C^1 級写像 $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ が保存場である (ある関数の勾配として表現できる) ための必要十分条件は、 $\mathbb{R}^n$ 上の全ての点において、 T の Jacobi 行列が対称性を持つことである。

*10 η -弱凸性は、 η -weakly convex [21]、 η -hypoconvex [22]、 $1/\eta$ -proximal [23, 24] と呼ばれる。

*11 凸関数 $f \in \Gamma_0(\mathcal{H})$ の (インデックス $\gamma > 0$ の) Moreau 包は、 $\gamma f : \mathcal{H} \rightarrow \mathbb{R} : x \mapsto \min_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2\gamma} \|x - y\|^2 \right)$ で定義される。Moreau 包の性質については、[14]、[19, Fact 17.17]などを参照されたい。

すれば、 $T = \nabla\psi$ は、MoL-Grad デノイザであることが分かる。

逆に、 $T = \nabla\psi$ が MoL-Grad デノイザであるとする、 ψ が β^{-1} -平滑凸関数であることから、 $\psi^* \in \Gamma_0(\mathcal{H})$ は β -強凸関数である [12, Theorem 18.15]。したがって、 $\varphi := \psi^* - (1/2) \|\cdot\|^2 = (\psi^* - (\beta/2) \|\cdot\|^2) - ((1 - \beta)/2) \|\cdot\|^2$ は、 $(1 - \beta)$ -弱凸関数である。このとき、 β -強凸関数 $\varphi + (1/2) \|\cdot\|^2 = \psi^*$ は唯一つの最小解を持つため、写像 s-Prox_φ を定義できる。

以上の議論の要点をまとめてみよう。

- (a) $(1 - \beta)$ -弱凸関数 φ から MoL-Grad デノイザ $T = \nabla\psi$ を定義できる。ここで、 $\psi := {}^\beta(\check{\varphi}^*) = (\varphi + (1/2) \|\cdot\|^2)^*$ は、 β^{-1} -平滑凸関数である。
- (b) 逆に、MoL-Grad デノイザ $T = \nabla\psi$ (ψ は β^{-1} -平滑凸関数) から、 $(1 - \beta)$ -弱凸関数 $\varphi := \psi^* - (1/2) \|\cdot\|^2$ を定義できる。

実は、(a), (b) のように定義した φ と ψ の間に、関係式

$$\nabla\psi = \text{s-Prox}_\varphi \quad (2)$$

が成り立つ。

定理 1 (s-prox 写像と MoL-Grad デノイザの等価性定理 [1]). 写像 $T : \mathcal{H} \rightarrow \mathcal{H}$ と定数 $\beta \in (0, 1)$ について、以下の二条件は等価である。

- (C1) $\varphi + ((1 - \beta)/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ を満たす弱凸関数 $\varphi : \mathcal{H} \rightarrow (-\infty, +\infty]$ を用いて、 $T = \text{s-Prox}_\varphi$ と表現できる^{*12}。
- (C2) MoL-Grad デノイザ $T = \nabla\psi$ として表現できる。

特に、(C1) が成り立つとき、 $\check{\varphi} := \varphi + ((1 - \beta)/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ となると、 β^{-1} -平滑凸関数 $\psi := \left(\varphi + (1/2) \|\cdot\|^2\right)^* = {}^\beta(\check{\varphi}^*) \in \Gamma_0(\mathcal{H})$ によって $T = \nabla\psi$ が成り立つ。逆に、(C2) が成り立つとき、 $(1 - \beta)$ -弱凸関数 $\varphi := \psi^* - (1/2) \|\cdot\|^2$ によって $T = \text{s-Prox}_\varphi$ が成り立つ。

注意 1. β が区間 $(0, 1)$ 外の場合について、簡単に述べておきたい。

1. MoL-Grad デノイザ (定義 2) で $\beta \geq 1$ とした場合: φ が凸関数 ($\beta > 1$ のときは強凸関数) であるから、Moreau による以下の定理 [25, Corollary 10.c] が成り立つ。

定理 2 (Moreau 1965 [25]). 実ヒルベルト空間 $\mathcal{H}$ 上で定義された写像 $T : \mathcal{H} \rightarrow \mathcal{H}$ に対して、以下の二条件は等価である。

- (a) 凸関数 $g \in \Gamma_0(\mathcal{H})$ を用いて $T = \text{s-Prox}_g$ と表現できる。
- (b) T が非拡大 (1-Lipschitz 連続) であり、さらに、ある凸関数 $\psi \in \Gamma_0(\mathcal{H})$ を用いて、 $T(x) \in \partial\psi(x)$ ($\forall x \in \mathcal{H}$) と表現できる。このとき、 ψ は微分可能 (Fréchet 微分可能) であり、 $\partial\psi(x) = \{\nabla\psi(x)\}$ である (すなわち、 $T = \nabla\psi$)。

定理 1 と定理 2 を比較すると、弱凸関数 φ と凸関数 g が対応することが分かる。したがって、定理 1 は、定理 2 の (凸関数から弱凸関数への) 拡張になっている。

^{*12} この場合、 $\varphi + \|\cdot - x\|^2$ が (全ての $x \in \mathcal{H}$ に対して) 強凸関数であることから、最小解の存在性と一意性が保証される。

2. 定理 1(C1) で $\varphi + ((1 - \beta)/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ となる「最小の」 $(1 - \beta)$ が $1 - \beta \geq 1 (\Leftrightarrow \beta \leq 0)$ の場合：この場合、Lipschitz 連続写像 s-Prox_φ を定義することはできない。このことは、MoL-Grad デノイザの Lipschitz 定数 β^{-1} が、 β を 0 に近づけた極限を取ると $+\infty$ になってしまう、あるいは、 $\beta < 0$ のとき、 β^{-1} が負値を取ってしまうことから予想できる（厳密な議論は、系 2（5 章）を参照）。

定理 1 の β の範囲は、正の実数全体に拡張できる^{*13}。ただし、本稿の主役である MoL-Grad デノイザに対応するのは、あくまで定理 1 の範囲 $\beta \in (0, 1)$ であることを強調しておく。

系 1 (定理 1 の一般形). 写像 $T : \mathcal{H} \rightarrow \mathcal{H}$ と定数 $\beta \in \mathbb{R}_{++}$ について、以下の二条件は等価である。

(C1) $\check{\varphi} := \varphi + ((1 - \beta)/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ を満たす関数 $\varphi : \mathcal{H} \rightarrow (-\infty, +\infty]$ を用いて、 $T = \text{s-Prox}_\varphi$ と表現できる。

(C2) β^{-1} -平滑凸関数 ψ の勾配として $T = \nabla \psi$ の形で表現できる。

この等価な条件が満たされるとき、 ψ^* は β -強凸関数であり、関係式 $\varphi = \psi^* - (1/2) \|\cdot\|^2$ と $\psi = \beta(\check{\varphi}^*) \in \Gamma_0(\mathcal{H})$ が成り立つ。

$\mathcal{H}$ 上の恒等写像を $\text{Id} : \mathcal{H} \rightarrow \mathcal{H} : x \mapsto x$ で表す。弱凸関数の近接写像（s-prox 写像）は、二次関数を加えて凸化した関数の近接写像を用いて表現することができる。

命題 2 ([1]). $\phi : \mathcal{H} \rightarrow (-\infty, +\infty]$ が $\check{\phi} := \phi + (\rho/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ ($\rho \in \mathbb{R}_{++}$) を満たす弱凸関数であるとする。このとき、全ての正定数 $\sigma \in \mathbb{R}_{++}$ に対して、

1. (ϕ と $\check{\phi}$ の近接写像の関係性) $\text{s-Prox}_{\check{\phi}/\sigma} = \text{s-Prox}_{\phi/(\sigma+\rho)} \circ \left(\frac{\sigma}{\sigma+\rho} \text{Id} \right)$
2. (拡張 Moreau 分解) $\text{Id} = \text{s-Prox}_{\sigma\check{\phi}^*} + \sigma \text{s-Prox}_{\phi/(\sigma+\rho)} \circ (\sigma + \rho)^{-1} \text{Id}$

が成り立つ。

3. MoL-Grad デノイザを採用した PnP 型近接分離アルゴリズムと連続最適化

本章では、PnP 法に基づく二つの近接分離アルゴリズム（PnP-FBS 法と PnP-PDS 法）を紹介する。近接写像の代わりに MoL-Grad デノイザを採用することで、これらの手法によって生成される点列は、MoL-Grad デノイザから導かれる「暗黙的正則化関数」を含む目的関数の最小解に収束する。PnP-FBS 法で扱えるタイプの問題に対して、同手法で収束が保証される Lipschitz 定数の範囲を超える状況であっても、PnP-PDS 法を用いることで最適解への収束が保証される。

^{*13} 真である下半連続関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ と定数 $\lambda > 0$ を考える。このとき、(i) $f + (1/(2\lambda)) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ と (ii) 全ての $\mu \in (0, \lambda)$ に対して、 $((\lambda - \mu)/\lambda) \text{s-Prox}_{\mu f}$ が堅非拡大（すなわち、 $2((\lambda - \mu)/\lambda) \text{s-Prox}_{\mu f} - \text{Id}$ が非拡大）であることが等価である [22, Theorem 6.5]。この事実は、定理 1 と類似点があるが、関数 f の存在とその下半連続性を初めから仮定している点や定理の主張する内容が、定理 1 とは本質的に異なる。また、[22, Theorem 6.5] の主張する範囲を強凸まで拡張した、より一般的な内容を含む結果（ただし、追加で近接有界性（prox-boundedness）を仮定）も知られている [26] が、同様の理由により、系 1 とは本質的に異なるものである。

3.1. 前方後方近接分離アルゴリズム (PnP-FBS アルゴリズム)

定理 1 と命題 2 から、1 章で提示した二つの問いに関する定理が得られる^{*14}。

定理 3 (MoL-Grad デノイザを採用した PnP-FBS 法の収束定理 [1]). κ -平滑な ρ -強凸関数 $f \in \Gamma_0(\mathcal{H})$ ($\kappa > \rho > 0$) と定数 $\beta \in ((\kappa - \rho)/(\kappa + \rho), 1) \subsetneq (0, 1)$ に対する MoL-Grad デノイザ $T = \nabla\psi : \mathcal{H} \rightarrow \mathcal{H}$ を考える。定数 $\mu \in [(1 - \beta)/\rho, (1 + \beta)/\kappa]$ に対して、以下が成り立つ。

1. $\mu\hat{f} + \check{\varphi} = \mu f + \varphi$ ($\varphi := \psi^* - (1/2) \|\cdot\|^2$) となる二つの凸関数 $\hat{f} := f - [(1 - \beta)/(2\mu)] \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ と $\check{\varphi} := \varphi + [(1 - \beta)/2] \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ を定義する^{*15}。このとき、

$$T \circ (\text{Id} - \mu\nabla f) = \text{s-Prox}_{\beta^{-1}\check{\varphi}} \circ (\text{Id} - \beta^{-1}\mu\nabla\hat{f}) \quad (3)$$

が成り立つ。ここで、平滑項 $\mu\hat{f}$ の勾配に対するステップサイズは、 $\beta^{-1} \in (1, 2/L_{\mu\nabla\hat{f}})$ を満たす ($L_{\mu\nabla\hat{f}} := \mu\kappa - (1 - \beta) > 0$ は勾配 $\mu\nabla\hat{f}$ の Lipschitz 定数)。

2. 関数 $\mu f + \varphi$ が $\mathcal{H}$ で最小値を取るとする。このとき、任意の初期点 $x_0 \in \mathcal{H}$ から

$$x_{k+1} := T(x_k - \mu\nabla f(x_k)), \quad k \in \mathbb{N} \quad (\text{PnP-FBS 法}) \quad (4)$$

によって生成される点列 $(x_k)_{k \in \mathbb{N}} \subset \mathcal{H}$ は、(暗黙的正則化関数 φ を含む) 目的関数 $\mu f + \varphi$ の (一つの) 最小解 $\hat{x} \in \mathcal{H}$ に弱収束する。

PnP-FBS 法の更新式 (4) では、本来の FBS 法において、(下半連続な真凸関数の) 近接写像があるべき箇所に、デノイザ T が置かれている。このアルゴリズムは、 T に付随する先験知識整合性関数^{*16} φ とデータ整合性関数 f をバランスよく小さくしたいというタスクを扱っていると解釈できる。 T は画像らしいものに近づける働きを持ち、最急降下ステップは観測画像 (劣化画像) から離れすぎないようにする働きを持つ。これによって、観測画像の情報を程よく残しながら、(観測画像より劣化の少ない) クリーンな画像らしいものができると期待するというわけである。

デノイザ T による先験知識を活用した推定戦略は、直感的には、巧くいきそうに思えるものであるし、実際、既存アプローチを凌ぐ性能が得られることが報告されている [6–8]。また、人の手で設計された先験知識整合性関数の近接写像 (デノイザ) と比べて、機械が大量のデータから学習したニューラルネットワーク (デノイザ) の方が、統計量をより正確に反映していて信頼性が高いという風潮も広まってきているように感じられる。ところが、一般に、 T に何も

*14 弱凸関数の近接写像を用いた場合の PnP-FBS 法の収束については、[21] で詳しく議論されている。ただし、[21] では、ユークリッド空間に限定して議論していることに加えて、 φ の値域が $\mathbb{R}$ に制限されている ($+\infty$ が含まれていない)。この制限は、例えば、ReLU やそれを途中で飽和させた活性化関数を用いたい状況で、大きな障壁となる。

*15 関数に付したアクセント記号 $\hat{(\cdot)}$ と $\check{(\cdot)}$ は、それぞれ、 f と φ に加えられる凹関数 $-[(1 - \beta)/(2\mu)] \|\cdot\|^2$ と凸関数 $[(1 - \beta)/2] \|\cdot\|^2$ の形状になぞらえて使用されている。

*16 [3] に倣い、先験知識への整合性が高いほど小さな値を取るよう設計された関数を先験知識整合性関数 (一般に、正則化関数と呼ばれることが多い)、推定値と観測データの整合性が高いほど小さな値を取るよう設計された関数をデータ整合性関数と呼ぶことにする。

条件を課さない場合、 T が何らかの先験知識に付随するものである保証はなく、また、式 (4) がなんらかの目的関数を最適化する保証もない。定理 3 は、 T が MoL-Grad デノイザの条件を満たしさえすれば、(適切な仮定とパラメータ設定の下で) PnP-FBS 法が、関数 $\mu f + \varphi$ の最小解へ弱収束する点列を生成することを明確に述べている。このとき、 φ が T に付随する先験知識整合性関数 (暗黙的正則化関数) である。

注意 2 (定理 3 におけるパラメータの範囲について). $\mu f + \varphi$ の凸性と Lipschitz 定数 $L_{\mu \nabla \hat{f}}$ の正值性を保証する条件が $\mu \geq (1 - \beta)/\rho$ である。一方、式 (3) の右辺で $\mu \hat{f}$ の勾配にかかるステップサイズ β^{-1} が $2/L_{\mu \nabla \hat{f}}$ 未満になることを保証する条件が $\mu < (1 + \beta)/\kappa$ である。この 2 条件を満たす μ の存在条件が $\beta > (\kappa - \rho)/(\kappa + \rho)$ である。

3.2. 主双対近接分離アルゴリズム (PnP-PDS アルゴリズム)

本節のタスク設定は、以下を使用する。

- $\mathcal{H}, \mathcal{U}$: 実ヒルベルト空間 ($\mathcal{U}$ 上のノルムを $\|\cdot\|_{\mathcal{U}}$ で表す)
- $L: \mathcal{H} \rightarrow \mathcal{U}$: $\mathcal{H}$ から $\mathcal{U}$ への有界線形写像
- $f \in \Gamma_0(\mathcal{H})$: 平滑 ρ -強凸関数 ($\rho \in \mathbb{R}_{++}$)
- $T := \nabla \psi: \mathcal{U} \rightarrow \mathcal{U}$: β^{-1} -Lipschitz 連続な MoL-Grad デノイザ ($\psi \in \Gamma_0(\mathcal{H})$, $\beta \in (0, 1)$)

タスク. 「 Lx に関する先験知識 (T 由来) との整合性」と「 f で評価されるデータ整合性」をよい塩梅で高くする $x \in \mathcal{H}$ を求めよ。

例えば、 $\mathcal{H} = \mathcal{U}$, $L = \text{Id}$ とすると、前章の問題設定に一致する。また、 L を第一種差分作用素に選び、 T をデノイザ (スパース性を促進する作用素) に選べば、全変動に関する先験知識を活用することができる。さて、今回も、PnP-FBS 法の導出に倣い、PDS 法の近接写像をそのまま MoL-Grad デノイザに置き換えるだけで良いだろうか。残念ながら、PDS 法は、近接写像と勾配降下ステップの合成写像が含まれる形になっていないため、この単純な置き換え戦略では、最適解への収束を保証することが容易でない。

そこで、上記タスクを扱うためのアルゴリズムを導出するために、まず初めに、関数^{*17}

$$\hat{f}: x \mapsto f(x) - \frac{\rho}{2\|L\|^2} \|Lx\|_{\mathcal{U}}^2 \in \Gamma_0(\mathcal{H})$$

を定義する。このとき、勾配

$$\nabla \hat{f}(x_k) = \nabla f(x_k) - (\rho/\|L\|^2) L^* Lx_k \quad (5)$$

は、 κ -Lipschitz 連続 ($\kappa \in \mathbb{R}_{++}$) である。次に、暗黙的正則化関数

$$g := \left(\sigma + \frac{\rho}{\|L\|^2} \right) \underbrace{\left(\psi^* - \frac{1}{2} \|\cdot\|_{\mathcal{U}}^2 \right)}_{=\varphi}$$

を定義する。ここで、 φ のスケール係数 $\sigma + \frac{\rho}{\|L\|^2}$ は、PnP-PDS 法において、定数 $\sigma > 0$

^{*17} $\|L\| := \sup_{x \neq 0} \|Lx\|_{\mathcal{U}} / \|x\|$ は、 L の作用素ノルムである。

が双対変数のステップサイズに対応するように設定されている。このとき、

$$\check{g} := g + \frac{\rho}{2\|L\|^2} \|\cdot\|_{\mathcal{U}}^2 \in \Gamma_0(\mathcal{U})$$

とすると、 $f + g \circ L = \hat{f} + \check{g} \circ L \in \Gamma_0(\mathcal{H})$ が成り立つ。そして、定理 1 と命題 2 を用いると、

$$\text{s-Prox}_{\sigma\check{g}^*} = \text{Id} - \sigma T \circ \left(\sigma + \frac{\rho}{\|L\|^2} \right)^{-1} \text{Id} \quad (6)$$

を示すことができる。 $\hat{f}$ と $\check{g}$ は凸関数であり、さらに、 $\hat{f}$ は平滑であるから、式 (5), (6) を Condat–Vũ の主双対近接分離アルゴリズム [15, 16] (正確には、Condat–Vũ iterates (form II) [27]) に代入すると、アルゴリズム 1 を得る^{*18}。

アルゴリズム 1 (PnP-PDS アルゴリズム).

Initialization: $x_0 \in \mathcal{H}$, $u_0 \in \mathcal{U}$

Requirements: $\sigma > 0$, $\tau > 0$, $\rho > 0$

$$\tilde{u}_{k+1} := u_k + \sigma L x_k$$

$$u_{k+1} := \tilde{u}_{k+1} - \sigma T \left(\left(\sigma + \frac{\rho}{\|L\|^2} \right)^{-1} \tilde{u}_{k+1} \right)$$

$$x_{k+1} := \left(\text{Id} + \frac{\tau\rho}{\|L\|^2} L^* L \right) x_k - \tau \nabla f(x_k) - \tau L^*(2u_{k+1} - u_k)$$

アルゴリズム 1 には、 g の文字は現れず、その情報は T にエンコードされている。これが g を暗黙的正則化関数と呼ぶ理由である。以下、アルゴリズム 1 に関する収束定理を述べる。

定理 4 (MoL-Grad デノイザを採用した PnP-PDS アルゴリズムの収束定理 [1]). ステップサイズ σ, τ に関する条件

- (i) $\sigma \leq \frac{\beta\rho}{(1-\beta)\|L\|^2}$
- (ii) $\tau \left(\sigma\|L\|^2 + \frac{\kappa}{2} \right) < 1$

を仮定する。さらに、主問題と双対問題

$$\min_{x \in \mathcal{H}} f(x) + g(Lx) \quad (\text{主問題}) \quad (7)$$

$$\min_{u \in \mathcal{U}} \hat{f}^*(-L^*u) + \check{g}^*(u) \quad (\text{双対問題}) \quad (8)$$

の解が存在すると仮定する。このとき、任意の初期点 $(x_0, u_0) \in \mathcal{H} \times \mathcal{U}$ からアルゴリズム 1 によって生成される 2 つの点列 $(x_k)_{k \in \mathbb{N}} \subset \mathcal{H}$ と $(u_k)_{k \in \mathbb{N}} \subset \mathcal{U}$ は、それぞれ、暗黙的正則化関数 g から決まる主問題 (7) の解 $\hat{x} \in \mathcal{H}$ と双対問題 (8) の解 $\hat{u} \in \mathcal{U}$ に弱収束する。

^{*18} L^* は、有界線形作用素 L の随伴作用素である。アルゴリズム 1 において、 x_{k+1} の定義式に含まれる項 $(\tau\rho/\|L\|^2)L^*Lx_k$ は、Condat–Vũ iterates (form II) [27] には含まれず、実は、この僅かな修正が、収束保証の鍵になっている。なお、アルゴリズム 1 の導出過程は、他の近接分離アルゴリズムにも拡張できる (詳細な手順は、[1, Section IV.C] を参照)。

注意 3 (定理 4 で Lipschitz 定数 β^{-1} の上限がないことについて). 定理 3 では、PnP-FBS アルゴリズムの収束を保証するために、 β に下限があった (この理由は、注意 2 を参照)。これは、MoL-Grad デノイザ T の Lipschitz 定数 β^{-1} に上限が存在することを意味している。一方、 $\mathcal{H} = \mathcal{U}$ 、 $L = \text{Id}$ とし、さらに、定理 3 の μ との対応のために、 $\sigma + \rho / \|L\|^2 = \sigma + \rho := \mu_{\text{PDS}}^{-1}$ と設定して、アルゴリズム 1 を適用すれば、 $\mu_{\text{PDS}} f + \varphi$ を最小化することができる。この場合、 β は 0 と 1 の間の任意の実数値を取ることができる、すなわち、デノイザ T の Lipschitz 定数 β^{-1} に上界はない*19。したがって、PnP-PDS 法は、MoL-Grad デノイザの Lipschitz 定数に制約を課す必要のないアプローチ (Lipschitz 制約なしアプローチ) である。素朴な疑問として、なぜ PnP-PDS 法は、このような性質を持つのだろうか。鍵になるのは、アルゴリズム 1 において、デノイザ T にステップサイズ σ が乗じられている点である。実際、 σT の Lipschitz 定数は、不等式 $\sigma\beta^{-1} \leq \rho / [(1 - \beta) \|L\|^2] \rightarrow \rho / \|L\|^2$ ($\beta \rightarrow 0$) で与えられる上界を持つため、 T が推定解へ与える影響が、適切に調整されるのである。

さて、前段落で定義した $\mu_{\text{PDS}} = (\sigma + \rho)^{-1}$ は、定理 3 で定義した μ と同じ範囲の値を取ることができるだろうか。簡単な考察から、 β が小さな値を取るとき、 μ_{PDS} の上限が (μ と比べて) 大きくなることが分かる。実際、定理 4 条件 (i) と $\sigma > 0$ より、 $\mu_{\text{PDS}} = (\sigma + \rho)^{-1} \in [(1 - \beta)/\rho, 1/\rho]$ であるから、定理 3 の $\mu \in [(1 - \beta)/\rho, (1 + \beta)/\kappa]$ と上限が異なる。 $\kappa > \rho$ であることから、十分小さな $\beta \approx 0$ に対して、 $\kappa > (1 + \beta)\rho$ が成立する。厳密には、 $\beta < (\kappa - \rho)/\rho$ のとき、 $\kappa > (1 + \beta)\rho$ となり、このとき、 $1/\rho > (1 + \beta)/\kappa$ であるから、 μ_{PDS} の上限の方が μ の上限より大きくなる、つまり、 μ_{PDS} の選択範囲が広がる ($\beta < 1$ より、 $(\kappa - \rho)/\rho \geq 1$ の場合、 μ_{PDS} の選択範囲の方が必ず広がる)。このことから、 $(\kappa - \rho)/\rho > (\kappa - \rho)/(\kappa + \rho)$ であることを踏まえ、以下のことが言える。

- $\beta \leq (\kappa - \rho)/(\kappa + \rho)$ の場合 : PnP-PDS アルゴリズムを使用するとよい。定理 3 で与えられている β の範囲外のため、PnP-FBS アルゴリズムは使用できない。
- $\beta > (\kappa - \rho)/(\kappa + \rho)$ の場合 : (双対変数が不要な) PnP-FBS アルゴリズムが使用できる。特に、 $(\kappa - \rho)/\rho < \beta < 1$ のとき、PnP-FBS アルゴリズムの方が μ の選択範囲という点で考えても、理に叶っている。一方、 $(\kappa - \rho)/\rho > \beta$ のとき、PnP-PDS アルゴリズムの方が μ_{PDS} の上限が大きく取れるので、もし、先験知識整合性関数のウェイトを相対的に小さくしたい場合は、PnP-PDS アルゴリズムを選択すると良い。経験上、先験知識整合性関数のウェイトをできるだけ大きく (μ をできるだけ小さく) した方が良い推定解が得られやすい傾向にあり、そのような場合、(双対変数が不要な) PnP-FBS アルゴリズムを選ぶとよいだろう。

*19 MoL-Grad デノイザでは、Lipschitz 定数 β^{-1} が 1 より大きい場合が重要であるという考えから $\beta < 1$ としており、これが原因で T の Lipschitz 定数 β^{-1} の下限が存在している。しかし、注意 2 で述べたように、 $\beta \geq 1$ のとき、 s-Prox_ρ が Moreau の近接写像 (凸関数の近接写像) となり、凸最適化で標準的に用いられる FBS アルゴリズムの議論に帰着する。この意味で、 $\beta \geq 1$ の場合への拡張は自明である。

4. MoL-Grad デノイザの具体例

例 1 (Firm shrinkage). 正定数 $\tau > 0$ に対して、minimax concave (MC) 罰則関数 [21, 28,

29] $\varphi_\tau^{\text{MC}} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} |x| - \frac{1}{2\tau}x^2 & \text{if } |x| \leq \tau \\ \frac{1}{2}\tau & \text{if } |x| > \tau \end{cases}$ を定義する。このとき、二つのしきい値 $(0 <) \lambda_1 < \lambda_2$ に対する firm shrinkage [30]

$$\text{firm}_{\lambda_1, \lambda_2} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 0 & \text{if } |x| \leq \lambda_1 \\ \text{sign}(x) \frac{\lambda_2(|x| - \lambda_1)}{\lambda_2 - \lambda_1} & \text{if } \lambda_1 < |x| \leq \lambda_2 \\ x & \text{if } |x| > \lambda_2 \end{cases} \quad (9)$$

は、 λ_1/λ_2 -弱凸関数 $\varphi := \lambda_1 \varphi_{\lambda_2}^{\text{MC}}$ の近接写像 s-Prox_φ に一致する。

MC 罰則関数の他にも、SCAD [31] などの弱凸罰則関数の近接写像や garotte shrinkage [32] など、MoL-Grad デノイザである。

例 2 (Tied Weight ニューラルネットワーク). ユークリッド空間 $\mathbb{R}^m$ 上で定義される 1-平滑凸関数 $\psi \in \Gamma_0(\mathbb{R}^m)$ の勾配 $\nabla\psi : \mathbb{R}^m \rightarrow \mathbb{R}^m$ で表現される活性化関数（非拡大写像）を考える^{*20}。重み行列 $(\mathbf{O} \neq) \mathbf{W} \in \mathbb{R}^{m \times n}$ と活性化関数 $\nabla\psi$ で定義される tied weight ニューラルネットワーク $T := \mathbf{W}^\top \circ \nabla\psi \circ \mathbf{W} = \nabla(\psi \circ \mathbf{W})$ は、関数 $\varphi := (\psi \circ \mathbf{W})^* - (1/2) \|\cdot\|_2^2$ の s-prox 写像 $T = \text{s-Prox}_\varphi$ である。特に、重み行列が $\kappa := \|\mathbf{W}^\top \mathbf{W}\| > 1$ を満たすとき^{*21}、 T は MoL-Grad デノイザであり、このとき、 φ は $(1 - \kappa^{-1})$ -弱凸である。（もし $\kappa \in (0, 1]$ であれば、 T は、凸関数 φ の近接写像である。）昨年、北海道大学で開催された信号処理シンポジウムにおいて、MoL-Grad デノイザとして表現可能な tied weight 多層ニューラルネットワークアーキテクチャを提案し、画像処理への適用例を示している²²ので、実応用に興味のある読者は [34] を参照されたい。

例 3 (バイアス低減型 OSCAR[35]). 類似した特徴ベクトルを漏れなく抽出（グルーピング）する目的^{*22}で、OSCAR 正則化関数 [39–41]

$$\Omega_{\lambda_1, \lambda_2}^{\text{OSCAR}} : \mathbb{R}^n \rightarrow \mathbb{R} : \mathbf{x} \mapsto \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \sum_{i < j} \max\{|x_i|, |x_j|\}$$

が提案されている（正定数 $\lambda_1, \lambda_2 > 0$ ）。凸関数である OSCAR を用いた場合に生じる推定バ

^{*20} 実際、ReLU 関数、シグモイド関数、softmax 関数など、深層学習などで広く用いられる活性化関数の多くが、凸関数 $f \in \Gamma_0(\mathbb{R}^m)$ の近接写像 s-Prox_f （したがって、勾配 $\nabla(f + (1/2)\|\cdot\|^2)^*$ ）として表現できる [33]。

^{*21} $\|\mathbf{W}^\top \mathbf{W}\|$ は、行列 $\mathbf{W}^\top \mathbf{W}$ のスペクトルノルム（最大固有値）である。

^{*22} 同目的のために提案された手法として、elastic net [36], fused lasso [37], clustered lasso [38] などがある。Elastic net は、強い相関を持つ特徴ベクトルの係数を等しくする作用を持たないため、グループ構造の抽出が困難である可能性が指摘されている [39, 40]。一方、fused lasso は、強い相関を持つ特徴ベクトルが連続する場合に限り、係数を等しくする作用を持つ。clustered lasso は、負の相関を持つ特徴ベクトルに対して、グルーピング作用を持たない。

イアスを低減する目的で提案された Debiased OSCAR (DOSCAR)^{*23} [35]

$$T_{\lambda_1, \lambda_2, \omega, \eta}^{\text{DOSCAR}} := \omega \text{s-Prox}_{\Omega_{\lambda_1, \lambda_2}^{\text{OSCAR}}} - (\omega - 1) \text{s-Prox}_{\eta \Omega_{\lambda_1, \lambda_2}^{\text{OSCAR}}} \quad (10)$$

は、MoL-Grad デノイザである ($\omega, \eta > 1$)。

例 4 (Moreau 強化関数の s-Prox 写像). 凸関数 $f \in \Gamma_0(\mathcal{H})$ を用いて定義される [11, 29]

$$f_\lambda : \mathcal{H} \rightarrow (-\infty, +\infty] : x \mapsto f(x) - \lambda f(x) \quad (11)$$

を Moreau 強化関数と呼ぶ ($\lambda > 0$)。Moreau 強化関数は弱凸関数であり、その近接写像 (s-prox 作用素) は、MoL-Grad デノイザである。具体的な例として、[1, Examples 1, 2] では、ベクトルに対する firm shrinkage とグループ型 firm shrinkage を紹介しているので、参照されたい。

例 5 (不連続 shrinkage 作用素の連続緩和 [2]). Hard shrinkage [10] に代表される不連続 shrinkage 作用素 (定義 5) に対して、連続緩和を定義することができる (5 章)。この連続緩和は、MoL-Grad デノイザである。6 章で、連続緩和の具体例として、hard shrinkage に加えて、trimmed ℓ_1 shrinkage (trimmed ℓ_1 ノルム [17] の近接写像から定義される shrinkage 作用素) の連続緩和について言及する。

5. 近接包含変換法による連続緩和 — 不連続 shrinkage から MoL-Grad デノイザへ

前章まで、一価の近接写像である s-prox 写像について議論してきた。本章では、近接写像が一価であるとは限らない一般の場合について、集合値写像としての近接写像を定義し、同写像を簡単な四則演算と逆写像によって連続写像 (厳密には、MoL-Grad デノイザ) に緩和する手法を与える。ここで、 $\mathcal{H}$ 上の点を $\mathcal{H}$ の部分集合に対応させる写像 T を集合値写像 (または、多値写像) といい、 $T : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ と表記する^{*24}。

具体的には、まず、非凸関数の近接写像 (集合値写像) とある種の包含関係にある「1-弱凸関数の近接写像 (十分に一般的な条件下で存在する)」を考える。この 1-弱凸関数は、元の非凸関数のある種のエンベロープ (近接包 (proximal hull) [23] と等価) に相当する。そして、その 1-弱凸関数をスケーリングした関数の近接写像 (s-prox 写像) が元の近接写像の連続緩和を与える。この連続緩和は、「1-弱凸関数の (集合値写像としての) 近接写像」の吉田近似が、同関数をスケーリングした $1/(1 + \delta)$ -弱凸関数 ($\delta > 0$) の一価の近接写像と $(1/(1 + \delta))\text{Id}$ の合成写像で表現できることを示すものになっている。

^{*23} DOSCAR [35] は、二つの近接写像の外分 (正負係数によるアフィン結合) を与える。同作用素は、firm shrinkage が二つの soft shrinkage (ℓ_1 ノルム (凸関数) の近接写像) の外分として表現できるという事実に着目し、「一般の凸関数の場合においても、二つの近接写像の外分を用いれば、バイアス低減効果を得ることができるのでは」という発想から提案されている。実際、DOSCAR は、著しいバイアス低減効果を示すことが確認されている。

^{*24} $2^{\mathcal{H}}$ は、 $\mathcal{H}$ の部分集合を全て含む集合族 (部分集合を要素として持つ集合) であり、べき集合 (power set) と呼ばれる。つまり、 $A \in 2^{\mathcal{H}}$ は、 $\mathcal{H}$ の部分集合である。なお、集合値写像は、 $T : \mathcal{H} \rightrightarrows \mathcal{H}$ と表記されることもある [23]。

定義 4 (近接写像 [22, 23]). 関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ と正定数 $\gamma \in \mathbb{R}_{++}$ に対して、

$$\mathbf{Prox}_{\gamma f} : \mathcal{H} \rightarrow 2^{\mathcal{H}} : x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} \left(f(y) + \frac{1}{2\gamma} \|x - y\|^2 \right) \quad (12)$$

を f のインデックス $\gamma \in \mathbb{R}_{++}$ の近接写像という。一般に、 $\mathbf{Prox}_{\gamma f}$ は集合値写像であり、 $\mathbf{Prox}_{\gamma f}(x)$ は空集合になることもある。

定義 5 (不連続 shrinkage 作用素). 関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ と正定数 $\gamma \in \mathbb{R}_{++}$ に対して、ある点 $y \in \mathcal{H}$ で $\mathbf{Prox}_{\gamma f}(y)$ が二つ以上の要素を持つとする。このとき、 $T(x) \in \mathbf{Prox}_{\gamma f}(x)$ ($x \in \mathcal{H}$) となる不連続写像 $T : \mathcal{H} \rightarrow \mathcal{H}$ (命題 1 参照) を不連続 shrinkage 作用素と呼ぶ。

1-弱凸関数^{*25}の近接写像は、対応する関数の劣勾配によって特徴付けられる。

命題 3 (1-弱凸関数の近接写像 — 劣勾配による特徴付け [2]). 集合値写像 $T : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ に対して、以下の二条件は等価である。

1. ある凸関数 $\psi \in \Gamma_0(\mathcal{H})$ に対して、 $T = \partial\psi$ が成り立つ。
2. $\phi + (1/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ を満たす 1-弱凸関数 ϕ に対して、 $T = \mathbf{Prox}_{\phi}$ が成り立つ。

写像 T が等価な条件 1, 2 を満たすとき、 ψ と ϕ の間に関係式 $\phi = \psi^* - (1/2) \|\cdot\|^2$ ($\Leftrightarrow \psi = (\phi + (1/2) \|\cdot\|^2)^*$) が成り立つ。

定義 6 (下半連続 1-弱凸包 [2] (近接包 [23])). 関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ に対して、 $f + (1/2) \|\cdot\|^2$ の下界となる連続なアフィン関数が存在すると仮定する^{*26}。このとき、 $\tilde{f} := (f + (1/2) \|\cdot\|^2)^* - (1/2) \|\cdot\|^2$ を f の下半連続 1-弱凸包 (近接包; proximal hull) という^{*27}。 $\tilde{f}$ は、 $f(x) \geq \phi(x)$ ($\forall x \in \mathcal{H}$) を満たす下半連続な 1-弱凸関数 $\phi : \mathcal{H} \rightarrow (-\infty, +\infty]$ の中で最大のものである。

一般に、非凸関数の近接写像は、下半連続 1-弱凸包 (もし存在すれば) の近接写像の部分集合になっている。

定理 5 (近接包含定理 [2]). 関数 $f : \mathcal{H} \rightarrow (-\infty, +\infty]$ に対して、 $f + (1/2) \|\cdot\|^2$ の下界となる連続なアフィン関数が存在すると仮定する。このとき、

$$\mathbf{Prox}_f(x) \subset \mathbf{Prox}_{\tilde{f}}(x), \quad \forall x \in \mathcal{H} \quad (13)$$

が成り立つ。

^{*25} 凸関数や ρ -弱凸関数 ($\rho \leq 1$) は、定義より、全て 1-弱凸関数である。

^{*26} この仮定は、関数 $(f + (1/2) \|\cdot\|^2)^*$ が真であることを保証する条件である (2 章参照)。また、同仮定は、 f の (インデックス 1 の) Moreau 包について、 ${}^1f(x) = \inf_{y \in \mathcal{H}} (f(y) + (1/2) \|x - y\|^2) > -\infty$ となる $x \in \mathcal{H}$ が存在することと等価な条件である。ここで、 f の凸性や下半連続性を仮定していないため、 $f + (1/2) \|x - \cdot\|^2$ の最小値が存在するとは限らないことに注意されたい (2 章で扱った Moreau 包では、下半連続な真凸関数 f を対象としていたので、 $f + (1/2) \|x - \cdot\|^2$ の狭義凸性 (strict convexity) と強圧性 (coercivity) から、最小解の存在性が保証されていた)。後者の条件は、(関数 f の) 近接有界性 (prox-boundedness) と呼ばれる [23]。例えば、 f が下に有界であれば、条件を満たす。

^{*27} 定義の仮定と $\tilde{f} + (1/2) \|\cdot\|^2 = (f + (1/2) \|\cdot\|^2)^*$ という関係より、 $\tilde{f} + (1/2) \|\cdot\|^2$ は、 $f + (1/2) \|\cdot\|^2$ の下半連続凸包である [12, Proposition 13.45]。また、 $\tilde{f}$ は、近接包に一致する ([23, Exercise 11.26] 参照)。厳密には、近接包は、下半連続 1-弱凸包の前提となる仮定を満たさない場合にも定義されるが、このとき、 $(f + (1/2) \|\cdot\|^2)^*$ が真でないため、 $\tilde{f} + (1/2) \|\cdot\|^2$ は、 $f + (1/2) \|\cdot\|^2$ の下半連続凸包ではないことに注意されたい。

命題 4 ([24]). 下半連続関数 $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ に対して、 $f + (1/2) \|\cdot\|^2$ の下界となる連続なアフィン関数が存在すると仮定する。このとき、

$$\text{conv } \mathbf{Prox}_f(\mathbf{x}) = \mathbf{Prox}_{\tilde{f}}(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n \quad (14)$$

が成り立つ。ここで、conv は集合の凸包を表す。

系 2 ($\mathbf{Prox}_\varphi$ が Lipschitz 連続でないケース). $\varphi + (\eta/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ となる最小の η が $\eta \geq 1$ であるとする。このとき、写像 $\mathbf{Prox}_\varphi$ は Lipschitz 連続でない。

(証明) $\eta = 1$ の場合 : 命題 3.2 のケースに該当するので、ある凸関数 $\psi \in \Gamma_0(\mathcal{H})$ に対して、 $\mathbf{Prox}_\varphi = \partial\psi$ が成り立つ。 $\mathbf{Prox}_\varphi (= \partial\psi)$ が κ -Lipschitz 連続 ($\kappa > 0$) であると仮定する。このとき、 ψ は Fréchet 微分可能であり、 $\partial\psi = \{\nabla\psi\}$ であるから、 $\mathbf{Prox}_\varphi = \nabla\psi$ である。すると、 $\kappa \leq 1$ の場合、定理 2 より、 φ は凸関数であるから、仮定に矛盾する ($\kappa < 1$ のとき、(κ -Lipschitz 連続であると同時に) 1-Lipschitz 連続でもあることに注意)。また、 $\kappa > 1$ の場合、定理 1 より、 φ は $(1 - \kappa^{-1})$ -弱凸関数であるから、この場合も、仮定に矛盾する。したがって、 $\mathbf{Prox}_\varphi$ は Lipschitz 連続でない。

$\eta > 1$ の場合 : $\mathbf{Prox}_\varphi(x) = \emptyset$ となる $x \in \mathcal{H}$ が存在すれば、明らかに、 $\mathbf{Prox}_\varphi$ は不連続写像である。以下、 $\mathbf{Prox}_\varphi(x) \neq \emptyset$ ($\forall x \in \mathcal{H}$) とする。このとき、 $(\varphi + (1/2) \|\cdot\|^2)^*$ が真であることは容易に示せる。したがって、定理 5 より、 $(\emptyset \neq) \mathbf{Prox}_\varphi(x) \subset \mathbf{Prox}_{\tilde{\varphi}}(x)$ ($\forall x \in \mathcal{H}$) が成り立つ。 $\mathbf{Prox}_{\tilde{\varphi}}$ が一価写像であるとき、 $\mathbf{Prox}_\varphi = \mathbf{Prox}_{\tilde{\varphi}}$ であるから、1-弱凸関数 $\tilde{\varphi}$ は $\eta = 1$ の場合に該当するので、 $\mathbf{Prox}_{\tilde{\varphi}}$ は (よって、 $\mathbf{Prox}_\varphi$ も) Lipschitz 連続でない。 $\mathbf{Prox}_{\tilde{\varphi}}$ が多価写像であるとき、命題 1 より、 $\mathbf{Prox}_\varphi$ は不連続であるから、Lipschitz 連続でない。

系 3 (弱凸定数 η による近接写像の分類 [2]). $\eta \in (-\infty, +\infty]$ に対して、 η -弱凸関数 f を考える。特に、 $\eta < +\infty$ のとき、 $f + (\eta/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ が成り立つとする。

1. f が 1-弱凸の場合 ($\eta \leq 1$). $\mathbf{Prox}_f$ は極大巡回単調^{*28}である。特に、 $\eta < 1$ のとき、近接写像は、(一価の) $(1 - \eta)^{-1}$ -Lipschitz 連続写像である。
2. f が 1-弱凸でない場合 ($\eta > 1$).

(a) ユークリッド空間 $\mathcal{H} := \mathbb{R}^n$ の場合. $\eta < +\infty$ であれば、近接写像 $\mathbf{Prox}_f$ は極大単調でない [23, Proposition 12.19]。

(b) 一般の場合. (i) $f + (1/2) \|\cdot\|^2$ の下界となる連続なアフィン関数が存在し、さらに、(ii) $\hat{x} \in \text{dom } \partial(f + (1/2) \|\cdot\|^2)^{**} \setminus \text{dom } \partial(f + (1/2) \|\cdot\|^2)$ が存在すると仮定する。このとき、点 $u \in \partial(f + (1/2) \|\cdot\|^2)^{**}(\hat{x}) \neq \emptyset$ において、狭義の包含関係 $\mathbf{Prox}_f(u) \subsetneq \mathbf{Prox}_{\tilde{f}}(u)$ が成り立つ、すなわち、近接写像 $\mathbf{Prox}_f$ は極大単調でない。

系 3 の下で定理 5 を眺めると、 $\mathbf{Prox}_{\tilde{f}}$ は、 $\mathbf{Prox}_f$ の (一つの) 極大単調拡張 (極大巡回単調拡張) を与えることが分かる。

^{*28} 命題 3 と Rockafeller の巡回単調定理 [12, Theorem 22.18] より、1-弱凸関数の近接写像は、極大巡回単調写像によって特徴付けられることが直ちに分かる (詳細に興味のある読者は、[2, Remark 1] をご参照いただきたい)。

定理 6 (近接変換定理 [2]). $\phi + (1/2) \|\cdot\|^2 \in \Gamma_0(\mathcal{H})$ を満たす 1-弱凸関数 ϕ に対して、緩和係数 $\delta \in \mathbb{R}_{++}$ でスケーリングした $(\delta + 1)^{-1}$ -弱凸関数 $\phi/(\delta + 1)$ の近接写像は、

$$\begin{aligned} \text{s-Prox}_{\phi/(\delta+1)} &= \underbrace{\left[\mathbf{Prox}_{\phi}^{-1} + \delta \text{Id} \right]^{-1}}_{\mathbf{Prox}_{\phi} \text{ の吉田近似}} \circ (\delta + 1) \text{Id} \\ &= \delta^{-1} \left[\text{Id} - \underbrace{(\text{Id} + \delta \mathbf{Prox}_{\phi})^{-1}}_{\delta \mathbf{Prox}_{\phi} \text{ のレゾルベント}} \right] \circ (\delta + 1) \text{Id} \end{aligned} \quad (15)$$

で与えられる^{*29}。また、 $\text{s-Prox}_{\phi/(\delta+1)}$ は、 $(1 + 1/\delta)$ -Lipschitz 連続な MoL-Grad デノイザである。

6. 連続緩和から導出される MoL-Grad デノイザの具体例

例 6 (Hard shrinkage の連続緩和). ℓ_0 擬似ノルムの下半連続 1-弱凸包は、MC 罰則関数である。MC 罰則関数の近接写像は、インデックスを適当に選べば、firm shrinkage に一致する [28] (例 1) が、ちょうど 1-弱凸になるようにインデックスを選んだ場合、hard shrinkage を与えることが知られている [42, 43]。実際、正定数 $\tau \in \mathbb{R}_{++}$ に対して、 ℓ_0 擬

似ノルム (の定数倍) $g_{\tau} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \frac{\tau^2}{2} \|x\|_0 = \begin{cases} 0 & \text{if } x = 0 \\ \frac{\tau^2}{2} & \text{if } x \neq 0 \end{cases}$ の下半連続 1-弱凸包

$\tilde{g}_{\tau} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \tau \varphi_{\tau}^{\text{MC}}(x) = \begin{cases} \tau |x| - \frac{1}{2} x^2 & \text{if } |x| \leq \tau \\ \frac{\tau^2}{2} & \text{if } |x| > \tau \end{cases}$ の近接写像は、セレクション

$$\text{hard}_{\tau}(x) \in \mathbf{Prox}_{g_{\tau}}(x) \subset \mathbf{Prox}_{\tilde{g}_{\tau}}(x) = \begin{cases} \{0\} & \text{if } |x| < \tau \\ [0, x] & \text{if } |x| = \tau \\ \{x\} & \text{if } |x| > \tau \end{cases} \quad (x \in \mathbb{R}) \quad (16)$$

の形で $\text{hard}_{\tau} : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} 0 & \text{if } |x| \leq \tau \\ x & \text{if } |x| > \tau \end{cases}$ を与える。ここで、式 (16) の包含関係は、定理

5 に合致することに注意されたい^{*30}。この場合、 $\delta > 0$ に対して、式 (15) を用いて得られる $\text{s-Prox}_{\tilde{g}_{\tau}/(1+\delta)}$ は、firm shrinkage である。したがって、hard shrinkage の連続緩和は、firm shrinkage である。

^{*29} 二つ目の等号は、[12, Proposition 23.7] ($\mathcal{H} = \mathbb{R}^n$ の場合は、[23, Lemma 12.14]) による。定理は、 $\mathbf{Prox}_{\phi}$ の (インデックス δ) の吉田近似が、 $\delta \mathbf{Prox}_{\phi} = \text{s-Prox}_{\phi/(\delta+1)} \circ (\delta + 1)^{-1} \text{Id}$ という形で表現できることを示唆している。また、 $\psi := (\phi + (1/2) \|\cdot\|^2)^*$ を定義すれば、 $\mathbf{Prox}_{\phi} = (\partial \psi^*)^{-1} = \partial \psi$ であり (初めの等号は、[1, Lemma 1] 参照)、その吉田近似と ψ の Moreau 包の間に、関係式 $\delta \mathbf{Prox}_{\phi} = \delta (\partial \psi) = \nabla \delta \psi$ が成り立つ [12, Proposition 12.30]。

^{*30} 厳密には、 $\mathbf{Prox}_{\tilde{g}_{\tau}}$ は、 $\mathbf{Prox}_{g_{\tau}}$ の極大単調拡張 (maximally monotone extension) を与える [43]。

例 7 (Trimmed ℓ_1 ノルムの近接写像の連続緩和). N 次元ユークリッド空間上のベクトル $\mathbf{x} \in \mathbb{R}^N$ に対して、成分毎に絶対値を取った $|\mathbf{x}| \in \mathbb{R}_+^N$ を降順に並べ替えたベクトル $|\mathbf{x}|_{\downarrow} \in \mathbb{R}_{+\downarrow}^N := \{\mathbf{x} \in \mathbb{R}_+^N \mid x_1 \geq x_2 \geq \dots \geq x_N\}$ を定義する。非負の重みベクトル $\mathbf{w} \in \mathbb{R}_+$ に対して、関数 $\Omega_{\mathbf{w}}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}_+ : \mathbf{x} \mapsto \mathbf{w}^\top |\mathbf{x}|_{\downarrow}$ を定義する。このとき、 $\mathbf{w}$ が降順であれば、 $\Omega_{\mathbf{w}}(\mathbf{x})$ は、Ordered Weighted ℓ_1 (OWL) ノルム [41] と呼ばれるノルムになる。逆に、 $\mathbf{w}$ が昇順であれば、 $\Omega_{\mathbf{w}}(\mathbf{x})$ は Reverse Ordered Weighted ℓ_1 (ROWL) 罰則関数 [44] と呼ばれる非凸関数となり、その近接写像 (ROWL shrinkage と呼ぶ) は、集合値写像となる。ROWL において、特に、 $\mathbf{w}$ が 0 とある正值重み $w > 0$ のみを含むベクトルの場合、Trimmed ℓ_1 ノルム^{*31}[17] に一致する。Trimmed ℓ_1 ノルムの近接写像 [45] (ROWL shrinkage の特別な場合に相当する不連続写像) の連続緩和について、(3 次元以上の場合に対する) 閉形式表現を求めることは困難であるが、[46] で、同写像を有限回の反復で求めるアルゴリズムを与えている。不連続 shrinkage は、間違った方向に推定解を誘導してしまうことがあるのに対して、連続緩和は、この問題を回避できることが数値実験で実証されている [2, Section V.B-1]。

7. 先行研究との関係性

MoL-Grad デノイザと先行研究の関係性について、簡潔に述べておきたい (詳しい議論は [1, Section V.C, V.D] 参照)。

勾配ステップデノイザ [7, 8, 47]. 勾配ステップデノイザは、 $T_{\text{GS}} = \text{Id} - \nabla g$ の形で与えられる。MoL-Grad デノイザと最も関連の深い [8, Proposition 3.1] では、 $\text{Id} - \nabla g$ が縮小写像 (Lipschitz 定数が 1 未満) であることが仮定されているため、勾配ステップデノイザは、以下の性質を満たすものに限定される。第一に、全ての点において、 T_{GS} の Jacobi 行列は、正定値性を持つ。第二に、MoL-Grad デノイザの φ に相当する関数が平滑かつ弱凸 ($\Leftrightarrow \psi := (1/2) \|\cdot\|^2 - g$ が強凸かつ平滑 $\Leftrightarrow T = \nabla \psi$ が強単調かつ Lipschitz 連続) である。第三に、 φ の弱凸定数が $\eta_{\text{GS}} \in (0, 1/2)$ の範囲に限定される。MC 関数など、知られている正則化関数の多くは平滑でないため、平滑性の制限は、応用価値の高いデノイザを排除してしまう可能性が危惧される。なお、[8, Proposition 3.1] は、定理 1 の特別な場合として、容易に得ることができる (最小平均二乗誤差 (Minimum Mean Squared Error ; MMSE) デノイザ [48] についても同様のことが言える)。

Regularization by Denoising (RED) [49]. RED では、デノイザ $T_{\text{RED}} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ について、(i) 非拡大性、(ii) $JT_{\text{RED}}(\mathbf{x}) (\forall \mathbf{x} \in \mathbb{R}^n)$ (Jacobi 行列の対称性) (iii) $T_{\text{RED}}(c\mathbf{x}) = cT_{\text{RED}}(\mathbf{x}) (\mathbf{x} \in \mathbb{R}^n, \forall c \in [1 - \epsilon, 1 + \epsilon], \exists \epsilon > 0)$ (局所均質性) を仮定する。これらの仮定の下、RED 正則化関数 $h_{\text{RED}}(\mathbf{x}) := (1/2)\mathbf{x}^\top (\mathbf{x} - T_{\text{RED}}(\mathbf{x}))$ の勾配が $\nabla h_{\text{RED}}(\mathbf{x}) = \mathbf{x} - T_{\text{RED}}(\mathbf{x})$ で与えら

^{*31} Trimmed ℓ_1 ノルムを罰則関数として使用したときの解が、ある条件下で、非ゼロ成分数に関する制約 (スパース制約) を満たすことが知られている [17]。

れることを利用し、勾配法ベースのアルゴリズムで最適化するというのが RED の基本指針である。RED は、 $f_{\text{RED}} := (1/2)\mathbf{x}^\top T_{\text{RED}}(\mathbf{x})$ を用いて、形式的に $T_{\text{RED}} = \nabla f_{\text{RED}}$ と表現できるが、非拡大性などの強い制限がかかることに加え、本稿で述べてきた MoL-Grad デノイザとは、デノイザの使われ方が根本的に異なる^{*32}。

Maximally Monotone Operator (MMO) [51]. 極大単調作用素 (MMO) となるニューラルネットワークを採用した PnP-FBS 法の収束先を変分不等式問題の解として特徴づける枠組みが提案されている [51, Proposition 3.1]。この枠組みは、収束先を最適化問題の解として特徴づけるアプローチを特別な場合として含む、より一般的なものである一方、ニューラルネットワーク (デノイザ) が非拡大写像^{*33}である場合に制限される。この点において、同枠組みは、本稿で扱った枠組みを包含するものではない (Lipschitz 定数のみの観点から大雑把に言えば、互いに補集合の関係にある)。さらに、PnP-PDS 法を扱っている点も [51] との差異であり、特に、MoL-Grad デノイザの場合、注意 3 で述べたように、PnP-PDS 法は、Lipschitz 制約なしアプローチであるという大きな意味を持つ。

8. 結び

MoL-Grad デノイザに関する二編の論文 [1, 2] の本質を解説することに主眼をおいた。定理 1, 2 の関係性、系 1、注意 3、命題 4、近接変換定理 (定理 6) と吉田近似・レゾルベントとの関係性、下半連続 1-弱凸包 (定義 6) と近接包の等価性、下半連続 1-弱凸包の前提条件と近接有界性の等価性など、論文 [1, 2] に含まれない内容にも触れた。

謝辞. 本発表の機会をいただいた成島康史教授 (慶應大) に感謝申し上げます。本研究は、JSPS 科研費 23K22762 の助成を受けたものです。

参考文献

- [1] M. Yukawa and I. Yamada, “Monotone Lipschitz-gradient denoiser: Explainability of operator regularization approaches free from Lipschitz constant control,” *IEEE Trans. Signal Process.*, 2024, doi: 10.1109/TSP.2025.3580667.
- [2] M. Yukawa, “Continuous relaxation of discontinuous shrinkage operator: Proximal inclusion and conversion,” *IEEE Open J. Signal Process.*, vol. 6, pp. 753–767, 2025.
- [3] 山田功、久米啓太、張毅, “私家版：スパース先験知識の利用効果を高める連続最適化の新展開,” 第 36 回 RAMP 数理最適化シンポジウム (RAMP 2024) 予稿集, 2024.
- [4] W. Samek and K.-R. Müller, *Towards Explainable Artificial Intelligence*. Cham: Springer International Publishing, 2019, pp. 5–22.
- [5] X. Li, H. Xiong, X. Li, and *et al.*, “Interpretable deep learning: interpretation, interpretability,

^{*32} RED は、目的関数を最小化する保証がない (合意均衡点が求まる) ことも指摘されている [50]。

^{*33} 正確には、デノイザ T_{MMO} には、 $N_{\text{MMO}} := 2T_{\text{MMO}} - \text{Id}$ が非拡大であるという、通常の非拡大性より強い条件 (堅非拡大性) が課される。そのため、[51] では、 N_{MMO} をニューラルネットワークで学習するアプローチが提案されている。

- trustworthiness, and beyond,” *Knowl. Inf. Syst.*, vol. 64, pp. 3197–3234, 2022.
- [6] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in *Proc. IEEE Global Conf. Signal Inf. Process*, 2013, pp. 945–948.
 - [7] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin, “It has potential: Gradient-driven denoisers for convergent solutions to inverse problems,” in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 18 152–18 164.
 - [8] S. Hurault, A. Leclaire, and N. Papadakis, “Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization,” in *Proc. ICML*, vol. 162, 2022, pp. 9483–9505.
 - [9] M. Elad, B. Kawar, and G. Vaksman, “Image denoising: The deep learning revolution and beyond — a survey paper,” *SIAM J. Imaging Sciences*, vol. 16, no. 3, pp. 1594–1654, 2023.
 - [10] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
 - [11] J. Abe, M. Yamagishi, and I. Yamada, “Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition,” *Inverse Problems*, vol. 36, no. 3, pp. 1–36, Feb. 2020.
 - [12] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. New York: NY: Springer, 2017.
 - [13] P. L. Lions and B. Mercier, “Splitting algorithms for the sum of two nonlinear operators,” *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, 1979.
 - [14] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
 - [15] L. Condat, “A primal dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *J. Optim. Theory Appl.*, vol. 158, pp. 460–479, Aug. 2013.
 - [16] B. C. Vũ, “A splitting algorithm for dual monotone inclusions involving cocoercive operators,” *Adv. Comput. Math.*, vol. 38, pp. 667–681, Apr. 2013.
 - [17] J. Gotoh, A. Takeda, and K. Tono, “DC formulations and algorithms for sparse optimization problems,” *Math. Program.*, vol. 169, pp. 141—176, 2018.
 - [18] 山田功, *工学のための関数解析*. 数理工学社, 2009.
 - [19] I. Yamada, M. Yukawa, and M. Yamagishi, “Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York: Springer, 2011, vol. 49, pp. 345–390.
 - [20] 湯川正裕, “連載講座 — 非線形適応信号処理への凸解析的アプローチ (第 1 回~ 第 5 回),” *日本音響学会誌*, vol. 78, no. 8~ 12, 2022.
 - [21] I. Bayram, “On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty,” *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1597–1608, 2016.
 - [22] H. H. Bauschke, W. M. Moursi, and X. Wang, “Generalized monotone operators and their averaged resolvents,” *Math. Program.*, vol. 189, no. 55–74, Sep. 2021.
 - [23] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 3rd ed. Berlin Heidelberg: Springer, 2010.
 - [24] J. Chen, X. Wang, and C. Planiden, “A proximal average for prox-bounded functions,” *SIAM J. Optim.*, vol. 30, no. 2, pp. 1366–1390, 2020.
 - [25] J. J. Moreau, “Proximité et dualité dans un espace hilbertien,” *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.
 - [26] H. Luo, X. Wang, and X. Yang, “Various notions of nonexpansiveness coincide for proximal mappings of functions,” *SIAM J. Optim.*, vol. 34, no. 1, pp. 642–653, 2024.

- [27] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, “Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists,” *SIAM Review*, vol. 65, no. 2, pp. 375–435, 2023.
- [28] C. H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, Apr. 2010.
- [29] I. Selesnick, “Sparse regularization via convex analysis,” *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.
- [30] H.-Y. Gao and A. G. Bruce, “Waveshrink with firm shrinkage,” *Statistica Sinica*, vol. 7, no. 4, pp. 855–874, 1997.
- [31] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” *J. American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
- [32] H.-Y. Gao, “Wavelet shrinkage denoising using the non-negative garrote,” *J. Comput. Graph. Statist.*, vol. 7, no. 4, pp. 469–488, 1998.
- [33] P. L. Combettes and J. C. Pesquet, “Deep neural network structures solving variational inequalities,” *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, Feb. 2020.
- [34] H. Shimizu and M. Yukawa, “Implicit regularizer associated with tied-nonnegative-weight neural network: Plug-and-play for image restruction and convergence to optimal point,” in *Proc. IEICE Signal Processing Symposium*, Sapporo: Japan, 2024.
- [35] K. Suzuki and M. Yukawa, “External division of two proximity operators: An application to signal recovery with structured sparsity,” in *Proc. IEEE ICASSP*, 2024, pp. 9471–9475.
- [36] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. Roy. Stat. Soc. Ser. B*, vol. 67, no. 2, pp. 301–320, Mar. 2005.
- [37] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, “Sparsity and smoothness via the fused lasso,” *J. Roy. Stat. Soc. Ser. B*, vol. 67, no. 1, pp. 91–108, Dec. 2004.
- [38] Y. She, “Sparse regression with exact clustering,” *Electron. J. Stat.*, vol. 4, pp. 1055–1096, 2010.
- [39] H. D. Bondell and B. J. Reich, “Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR,” *Biometrics*, vol. 64, no. 1, pp. 115–123, 2008.
- [40] L. W. Zhong and J. T. Kwok, “Efficient sparse modeling with automatic feature grouping,” *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1436–1447, 2012.
- [41] X. Zeng and M. A. T. Figueiredo, “The ordered weighted ℓ_1 norm: Atomic formulation, projections, and algorithms,” *arXiv*, 2015, arXiv:1409.4271 [cs.DS].
- [42] M. Kowalski, “Thresholding rules and iterative shrinkage/thresholding algorithm: A convergence study,” in *Proc. IEEE ICIP*, 2014, pp. 4151–4155.
- [43] I. Bayram, “Penalty functions derived from monotone mappings,” *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 265–269, 2015.
- [44] T. Sasaki, Y. Bandoh, and M. Kitahara, “Sparse regularization based on reverse ordered weighted ℓ_1 -norm and its application to edge-preserving smoothing,” in *Proc. IEEE ICASSP*, 2024, pp. 9531–9535.
- [45] Y. Wang, Z. Luo, and X. Zhang, “New improved penalty methods for sparse reconstruction based on difference of two norms,” *Tech. Rep.*, Aug. 2013, 10.13140/RG.2.1.3256.3369.
- [46] T. Okuda, K. Suzuki, and M. Yukawa, “Sparse signal recovery based on continuous relaxation of reversely ordered weighted ℓ_1 shrinkage operator,” in *Proc. IEICE Signal Processing Symposium*, Sapporo: Japan, Dec. 2024.
- [47] S. Hurault, A. Leclaire, and N. Papadakis, “Gradient step denoiser for convergent plug-and-play,” in *Proc. ICLR*, 2022.
- [48] X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov, “Provable convergence of plug-and-play

- priors with MMSE denoisers,” *IEEE Signal Process. Lett.*, vol. 27, pp. 1280–1284, 2020.
- [49] Y. Romano, M. Elad, and P. Milanfar, “The little engine that could: Regularization by denoising (RED),” *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [50] E. T. Reehorst and P. Schniter, “Regularization by denoising: Clarifications and new interpretations,” *IEEE Trans. Computat. Imag.*, vol. 5, no. 1, pp. 52–67, 2019.
- [51] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, “Learning maximally monotone operators for image recovery,” *SIAM J. Imag. Sci.*, vol. 14, no. 3, pp. 1206–1237, 2021.