束上の劣モジュラ関数最大化

前原　貴憲

機械学習分野において単調劣モジュラ集合関数最大化問題が盛んに研究されている．これは単調劣モジュラ関数が情報量の数学的モデルとなっているため，データ要約を単調劣モジュラ関数最大化問題に定式化できるからである．ところで，集合関数を考えるというのは，暗黙にデータは構造をもたないただの集合である，と考えていることになる．しかし，実際は集合よりも複雑な構造をもつデータを扱うこともあり，その際には複雑な構造上の単調劣モジュラ関数最大化理論が必要となる．本記事ではデータが束の構造をもつ問題に注目し，束上の単調劣モジュラ関数を最大化する問題を紹介する．

キーワード：劣モジュラ関数最大化，束，機械学習

1. はじめに

劣モジュラ集合関数 (submodular set function) とは集合 E の部分集合を受け取って実数を返す関数 $f\colon 2^E \to \mathbb{R}$ であって，任意の $X, Y \subseteq E$ について以下の不等式を満たすもののことをいう：

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y). \quad (1)$$

この条件は**劣モジュラ性** (submodularity) と呼ばれている．劣モジュラ性は**限界効用逓減性** (diminishing return property) と呼ばれる以下の条件と同値となる：任意の $X \subseteq Y$ $(\subseteq E)$ および $e \in E$ について以下の不等式を満たす：

$$f(X \cup \{e\}) - f(X) \geq f(Y \cup \{e\}) - f(Y). \quad (2)$$

劣モジュラ性および限界効用逓減性は組合せ最適化・ゲーム理論・経済学などの幅広い分野において自然に出現する性質であり，これまでにさまざまな観点から研究が行われてきた．

機械学習分野では 2010 年ごろから劣モジュラ集合関数を**最大化**する問題が活発に研究されている．機械学習では「手元にあるデータを使ってモデルを学習することにより，まだ手元にないデータに対する性能（汎化性能）を向上させること」を目指す．汎化性能を向上させるための基本的な方針に「与えられたデータを要約して使う」というものがある．手元のデータを直接使ってしまうと手元のデータに特化（過学習）したモデルが得られてしまうかもしれないが，一度要約を経由すればその危険性が減らせるため，より汎化性能の高いモデルが得られるだろう，というアイデアである．最も素朴なデータ要約法は「データの部分集合を選び出す」ことであろう．$f(X)$ でデータ X の「情報量」をあらわすことにすると，データ要約問題は適当な条件，たとえば要素数が一定以下といった条件のもとで $f(X)$ を最大化する問題に定式化できる．さて，情報量関数 f はどのような性質をもつ関数だろうか．典型的には以下の性質を期待できそうである．

- データが増えると情報も増える：$X \subseteq Y$ ならば $f(X) \leq f(Y)$.
- データ量が少ないところに新データが加わったときの情報量増分は，データ量が多いところに新データが加わったときの情報量増分よりも大きい：$X \subseteq Y$ ならば $f(X \cup \{e\}) - f(X) \geq f(Y \cup \{e\}) - f(Y)$.

前者は**単調性** (monotonicity) であり，後者はまさに劣モジュラ性・限界効用逓減性である．すなわち，機械学習分野では単調劣モジュラ関数最大化問題は「データを要約する問題」の数理モデルと認識されており，現在でも活発に研究が行われている．

データ要約の問題を単調劣モジュラ関数最大化問題にモデル化する最大の恩恵は近似アルゴリズムの存在である．一般に，単調劣モジュラ関数最大化問題は NP 困難であるが，さまざまな制約下で近似アルゴリズムが構築できる．たとえば要素数制約であれば単純な貪欲法によって $1 - 1/e \approx 0.632$ 近似が達成できるし [1]，ナップサック制約であれば部分列挙と貪欲法を組み合わせた手法で $1 - 1/e$ 近似が達成できる [2]．マトロイド制約であれば貪欲法を連続化した手法で $1 - 1/e$ 近似が達成できる [3]．

ところで，データ要約問題を劣モジュラ**集合**関数最大化問題で定式化するということは，暗黙に「データはただの集合であって特別な構造をもっていない」と仮定していることになる．しかし，機械学習で扱うデータは何かしらの構造をもっていることがあり，そのよ

まえはら　たかのり
理化学研究所革新知能統合研究センター
takanori.maehara@riken.jp

うな場合にはこの仮定が妥当でないことになる．たとえば，同じデータが繰り返し現れるような状況では集合の代わりに多重集合を考えてデータの出現回数を情報として取り入れたほうがよいだろうし，文字列や時系列といったデータでは順序構造に着目するのが自然であろう．このような構造をもったデータの要約問題を数理的に扱う場合に，単純に劣モジュラ集合関数の理論を適用するのではなく，もっと構造に寄り添った新しい関数クラスを導入し，それに対して近似アルゴリズムを導出する，というものが最近の機械学習分野における劣モジュラ最大化の研究の一つのスタイルとなっている．

このスタイルの研究の一番面白い（そして難しい）部分は，「良い関数クラス」を定義する部分にある．理想的な関数クラスは以下の三つの条件をすべて満たすものである．

1. 具体的な応用問題を表現できる．
2. 近似アルゴリズムを導出できる．
3. 数学的な広がりがある．

実際に手を動かしてみるとわかるが，一つ目の条件と二つ目の条件のどちらか一方だけを満たす関数クラスを作るのはそれほど難しくない．一方で，両方を満たすような定義を作るのはそれほど簡単ではなく，多くの試行錯誤が必要となる．三つ目の条件を満たすのはさらに難しいが，具体的な問題を一つ解くだけならわざわざ関数クラスを導入する必要はないので可能な限り追求したい条件といえる．

さて，最近われわれの研究グループは**束** (lattice) 上の劣モジュラ関数最大化を研究している．機械学習で扱うデータの中には順序構造をもつものが少なからず存在しており，それらを数学的にうまく表現しようとすると自然な形で束が出現する．われわれはこれまで分配束およびモジュラ束に対する適当な関数クラスを考えて近似アルゴリズムの構築に成功した．また，この理論はより深い数学的理論につながることが期待できる状況にある．本稿ではこの理論を簡単に紹介したい．

2. 束と束劣モジュラ性

束は順序集合 $(\mathcal{L}, \leq)$ であって，任意の 2 元 $X, Y \in \mathcal{L}$ について最小上界 $X \vee Y = \inf\{Z \in \mathcal{L} : Z \geq X, Z \geq Y\}$ および最大下界 $X \wedge Y = \sup\{Z \in \mathcal{L} : Z \leq X, Z \leq Y\}$ が存在するもののことをいう．以下では束は常に最小元・最大元をもつとする．最小元を $\bot$ であらわし，最大限を $\top$ であらわすことにする．

最も基本的な束は**ブール束** (boolean lattice) と呼ばれる，集合 E の部分集合全体が包含関係についてなす束 $(2^E, \subseteq)$ であろう．ブール束では最小上界 $X \vee Y$ と最大下界 $X \wedge Y$ はそれぞれ集合の合併 $X \cup Y$ と共通部分 $X \cap Y$ に対応する．最小元は空集合であり，最大元は全体集合である．その他の重要な束の例に**線型空間束** (vector space lattice) と呼ばれる，線型空間 V の部分空間全体が包含関係についてなす束 $(\mathcal{L}, \subseteq)$ がある．線型空間束では最小上界 $X \vee Y$ は線型包 $\mathrm{span}(X \cup Y)$ であり，最大下界 $X \wedge Y$ は線型空間の共通部分空間 $X \cap Y$ に対応する．最小元はゼロだけからなる部分空間であり，最大元は全体空間である．

実は「束上の劣モジュラ関数」は古くから研究が行われてきたコンセプトであり，特に新しいものではない．束 $\mathcal{L}$ 上の関数 $f\colon \mathcal{L} \to \mathbb{R}$ が**束劣モジュラ関数** (lattice submodular function) であるとは，任意の $X, Y \in \mathcal{L}$ について以下の不等式が成り立つことをいう [4].

$$f(X) + f(Y) \geq f(X \vee Y) + f(X \wedge Y). \quad (3)$$

なるほど，この関数クラスに対して近似アルゴリズムを導くのがこの研究の目標なのか，と思うかもしれないが，実はそれはあまりうまくいかない．多くの機械学習的応用で必要となるのは劣モジュラ性そのものでなく限界効用逓減性である．これらは集合関数においては同値となるが，一般の束ではそもそも限界効用逓減性をどう定義するかが自明でなく，素朴に定義すると限界効用逓減性と同値にならない．

このような考えのもと，機械学習分野では限界効用逓減性に注目した劣モジュラ関数風の関数クラスである **DR 劣モジュラ関数** (DR submodular function) が導入された．ここで DR は限界効用逓減 (diminishing return) の頭文字である．この研究の目標は，一言で言えば「うまく束上で DR 劣モジュラ関数を定義すること」になる．

3. 整数格子上の DR 劣モジュラ関数

一般の束上の DR 劣モジュラ関数に進む前に，まずは具体的な束である**整数格子** (integer lattice) での定義を見ておこう．DR 劣モジュラ関数はもともとこの空間で導入されたものである．

非負整数の集合を $\mathbb{N}$ であらわす．有限集合 E で添字づけられた非負整数ベクトル全体の集合 $\mathbb{N}^E$ を整数格子という．整数格子上の関数 $f\colon \mathbb{N}^E \to \mathbb{R}$ が DR 劣モジュラ [5] であるとは，任意の $X \leq Y$ なる $X, Y \in \mathbb{N}^E$ および任意の $i \in E$ について

$$f(X+e_i)-f(X) \geq f(Y+e_i)-f(Y) \tag{4}$$

が成り立つことをいう．ここで e_i は i 成分のみが 1 の単位ベクトルである．関数 f が DR 劣モジュラであれば f は束劣モジュラになるが，しかしこの逆は成立しない．DR 劣モジュラ関数は比較的扱いやすい関数クラスであり，成分ごとの不等式制約のもとで $1-1/e$ 近似が達成できる [5].

さて，次節から分配束およびモジュラ束に対する DR 劣モジュラ関数最大化の理論を紹介する．これらの束に DR 劣モジュラ関数を定義する際にはもちろん整数格子における DR 劣モジュラ関数の定義を参考にするのだが，整数格子に対する DR 劣モジュラ性の定義にはベクトルの足し算という整数格子特有の演算が利用されているので，そのままでは拡張できない．この部分をどのように一般化するかが最初の課題となる．

4. 分配束上の DR 劣モジュラ関数

文書要約問題 (document summarization problem) とは，複数の文から構成されている文書が与えられたとき，いくつかの文を選び出すことで文書の要約を作る問題である．文書中の文全体を D であらわし，文の集合 $X \subseteq D$ の「情報量」を $f(X)$ であらわせば，この問題は次の形で記述できる．

$$\begin{array}{ll} \text{maximize} & f(X) \\ \text{subject to} & |X| \leq k, \\ & X \subseteq D. \end{array} \tag{5}$$

f としてさまざまなモデルが考えられており，それらのいくつかは単調劣モジュラ集合関数となる [6]．そのような場合について文章要約問題は単調劣モジュラ集合関数最大化問題となる．

さて，現実の文書ではしばしば文同士が論理的な関係をもっており，「文 b がなければ文 a は理解できない」といったことが生じる．このような関係に違反しないように要約を作ることを考えよう．この問題は有向非巡回グラフ (DAG) を用いると綺麗に表現できる．文 b がなければ文 a が理解できないとき，a から b に枝を引いて D を DAG とみなす．すると，論理関係に違反しない文の集合 X は「$a \in X$ のとき a から到達可能な頂点はすべて X に含まれる」という性質をもつものとして特徴づけられる．このような集合を**下側閉集合** (downward closed set) という．つまり，われわれの解くべき問題は以下となる．

$$\begin{array}{ll} \text{maximize} & f(X) \\ \text{subject to} & |X| \leq k, \quad X \text{ は下側閉集合}, \\ & X \subseteq D. \end{array} \tag{6}$$

この問題を解きたいのだが，下側閉集合という制約をどう扱えばよいかはあまり自明でない．そこで，ここでは下側閉集合全体が束をなすことに注目する．下側閉集合 X, Y について $X \cup Y$ および $X \cap Y$ もまた下側閉集合となる．すなわち下側閉集合全体は包含関係で束をなす．この束を $\mathcal{L}(D)$ と書けば，上の問題の下側閉集合制約を見た目上外すことができて

$$\begin{array}{ll} \text{maximize} & f(X) \\ \text{subject to} & |X| \leq k, \\ & X \in \mathcal{L}(D) \end{array} \tag{7}$$

となる．

(7) に現れた束 $\mathcal{L}(D)$ は**分配束** (distributive lattice) をなす．ここで分配束とは束であって，任意の $X, Y \in \mathcal{L}$ について

$$(X \vee Y) \wedge Z = (X \wedge Z) \vee (Y \wedge Z) \tag{8}$$

が成立するもののことをいう．上で述べたように任意の DAG について下方閉集合全体は分配束をなす．逆に，任意の分配束は適当な DAG D について $\mathcal{L}(D)$ と同型になる．そこで，ここからは文書要約問題を離れて一般の分配束に対する最大化問題を考えることにしよう．

まず，いくつかの用語を導入しよう．$X \in \mathcal{L}$ の**高さ** (height) とは，下限 $\perp \in \mathcal{L}$ から X までの（束を DAG と思ったときの）距離のことをいう．分配束ではこの距離は経路に依存しないことが証明できる．X の高さを（記号を濫用して）$|X|$ であらわす．$E \in \mathcal{L}$ が **join 既約** (join irreducible) であるとは $E = E_1 \vee E_2$ と書いたとき $E = E_1$ または $E = E_2$ が成り立つことをいう．ブール束におけるシングルトン $\{e\}$ は join 既約元である．また，整数格子における単位ベクトルの定数倍 ke_i も join 既約元である．以下では可読性のため束上の一般の元は大文字 X であらわし，join 既約元を小文字 e であらわすことにする．join 既約元 e が $X \in \mathcal{L}$ に対して**許容的** (admissible) であるとは $|X \vee e| = |X| + 1$ が成り立つことをいう．X に対して許容的な既約元全体を $\text{adm}(X)$ であらわす．

以上の準備のもと，分配束上の DR 劣モジュラ関数は次のように定義される．

定義 1 (Gottschalk and Peis [7])**.** 関数 $f\colon \mathcal{L} \to \mathbb{R}$ が

束 DR 劣モジュラ関数 (*lattice DR submodular function*) であるとは，任意の $X \leq Y$ なる $X, Y \in \mathcal{L}$ および任意の $a \leq b$ なる $a \in \mathrm{adm}(X)$, $b \in \mathrm{adm}(Y)$ について

$$f(X \vee a) - f(X) \geq f(Y \vee b) - f(Y) \tag{9}$$

が成り立つことをいう．

ブール束において束 DR 劣モジュラ性は通常の集合関数に対する劣モジュラ性・限界効用逓減性と一致する．また，整数格子を成分ごとの不等式で束と思ったものにおいて束 DR 劣モジュラ性は整数格子に対する通常の DR 劣モジュラ性と一致する．

束 DR 劣モジュラ関数を高さ制約のもとで最大化するのは難しくない．Gottschalk and Peis [7] は以下の結果を得た．

定理 2 (Gottschalk and Peis [7])**.** 分配束において高さ制約つき単調束 DR 劣モジュラ関数最大化問題は $1 - 1/e$ 近似可能．

この定理の証明は劣モジュラ集合関数に対する貪欲法の近似率評価を「束に一般化」することで得られる．証明を見るとこの気持ちが伝わると思うので，以下に簡単に証明を示す．

証明. 初期解 $X_0 = \bot$ から初めて毎ステップで許容的な要素のうち最も関数値が大きくなるものを追加していくアルゴリズム（貪欲法）を考える．このアルゴリズムの近似率を評価しよう．

最適解を X^* とし，貪欲法の i ステップ目の解を X_i とする．$X_i \leq X^* \vee X_i$ なので X_i から $X^* \vee X_i$ に至る鎖 $X_i = X_i^0 < X_i^1 < \cdots < X_i^l = X^* \vee X_i$ がとれる．ここで各 p について $X_i^p = X_i^{p-1} \vee b_p$ なる $b_p \in \mathrm{adm}(X_i^{p-1})$ が存在する．最適解と i ステップ目の解のギャップを望遠鏡和に展開して

$$\begin{aligned} f(X^*) - f(X_i) &\leq f(X^* \vee X_i) - f(X_i) \\ &\leq \sum_{p=1}^{l} \left(f(X_i^p) - f(X_i^{p-1}) \right) \end{aligned} \tag{10}$$

を得る．いま，b_p 以下にある $a_p \in \mathrm{adm}(X_i)$ を任意にとる．分配束ではそのようなものは常に存在することに注意する（b_p 以下の join 既約元であって X_i に含まれておらず，一番高さが低いものをとればよい）．すると，束 DR 劣モジュラ性より

$$\begin{aligned} f(X_i^p) - f(X_i^{p-1}) &\leq f(X_i \vee a_p) - f(X_i) \\ &\leq f(X_{i+1}) - f(X_i) \end{aligned} \tag{11}$$

となるため，$l \leq k$ とあわせて

$$f(X^*) - f(X_i) \leq k \left(f(X_{i+1}) - f(X_i) \right) \tag{12}$$

となる．以下この不等式を整理すれば $f(X_i) \geq (1 - 1/e) f(X^*)$ を得る． □

この証明は数式だけを見ると単調劣モジュラ集合関数に対する証明 [1] と全く一致している．集合の場合に素朴に行っていた操作を「分配束だからこういうことができる」と逐一確認していくのがこの証明でやっていることの本質といえる．

高さ制約以外の制約としてナップサック制約がよく調べられている．$c \colon \mathcal{L} \to \mathbb{R}$ が**束モジュラ関数** (lattice modular function) であるとは c と $-c$ がともに束劣モジュラ関数であることをいう．$c \geq 0$ なる単調束モジュラ関数 c および $B \geq 0$ について $c(X) \leq B$ と書ける制約をナップサック制約と呼ぶことにする．このとき，以下が証明できる．

定理 3 (Gottschalk and Peis [7])**.** $\mathrm{P} \neq \mathrm{NP}$ の仮定のもと，分配束上のナップサック制約つき単調束 DR 劣モジュラ関数最大化問題は定数近似不可能．

一方，c に「整合的」という条件を貸すと近似が可能となる（整合的の定義は省略する）．劣モジュラ集合関数に対する部分列挙・貪欲法 [2] を拡張することで以下の結果が得られる．

定理 4 (Nakashima and Maehara [8])**.** 分配束上の整合的なナップサック制約つき単調束 DR 劣モジュラ関数最大化問題は 1/3 近似可能．

ごく最近われわれはこの結果を改善し，複数ナップサック制約に対するより良い近似率のアルゴリズムの構築に成功した．

定理 5 (Maehara et al. [9])**.** $O(1)$ 個の整合的なナップサック制約つき単調束 DR 劣モジュラ関数最大化問題は $1 - 1/e - o(1)$ 近似可能．

この結果は複数ナップサック制約に対する連続貪欲法 [10] を分配束に拡張することで得られたものである．分配束は局所的にブール束を張り合わせたものとして表現できる．各ブール束に対する連続拡張として集合の場合と同様に超立方体をとれば，分配束全体の連続拡張は超立方体を複数張り合わせたものとなる．この構造を**メディアン複体** (median complex) という（図 1）．メディアン複体上で連続貪欲法が動くことを示すには，任意の 2 点 $x \leq y$ について x から y への「良い経路」の存在を示す必要がある．この経路はメディアン複体上に一意に存在する特殊な L_1 測地線であり，メディアン複体が「束 DR 劣モジュラ関数最大化と相性の良い測地的構造」をもつことを示している．

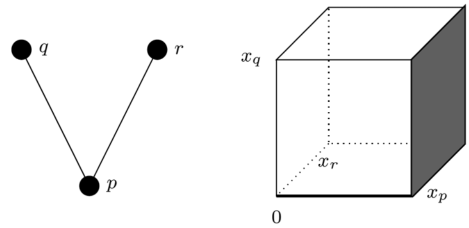

図 1　分配束の連続拡張．左図が分配束の DAG 表現，右図の太線部・斜線部が対応するメディアン複体．メディアン複体では 1 次元複体と 2 次元複体が一点で接している．

なお，ナップサック制約が整合的という条件は，制約領域がこの測地的構造のもとで測地的凸になるための十分条件である．

束の連続拡張を考えるのは劣モジュラ関数最小化の文脈でも行われてきた [11]．しかし，最小化と最大化では異なる連続拡張が行われるため，それに応じた異なる幾何的構造が出現する．最大化で現れる幾何的構造にはわかっていない部分が多く，今後の重要な研究課題だと考えている．

5.　モジュラ束上の劣モジュラ関数

主成分分析 (principal component analysis) はデータ分析の最も基本的な手法の一つである．ベクトル空間 V の元 $v_1, \ldots, v_N \in V$ が与えられたとき，V の低次元部分空間 X であってこれらを最もよく近似するもの，すなわち，

$$f(X) := \sum_{i=1}^{N} \|\Pi_X v_i\|^2 \tag{13}$$

を最大化する X を求める問題が主成分分析である．ここで $\|\cdot\|$ はユークリッドノルム，Π_X は部分空間 X への射影をあらわす．

この問題は自然に束上の最適化問題として表現できる．$\mathcal{L}(V)$ を V 上の部分空間全体とすると $\mathcal{L}(V)$ は部分空間の包含関係によって束をなす．この束における $X \in \mathcal{L}(V)$ の高さは X のベクトル空間としての次元と一致するので，主成分分析は以下の形で記述できる．

$$\begin{aligned} &\text{maximum} && f(X) \\ &\text{subject to} && |X| \le k, \\ &&& X \in \mathcal{L}(V). \end{aligned} \tag{14}$$

この問題は特異値分解もしくは貪欲法によって最適に解くことができる．

主成分分析は劣モジュラ性と無関係に思えるかもしれないが，われわれは主成分分析が貪欲法で解けることに注目して劣モジュラ関数との関係を見出そうと考えた．もし主成分分析を束上の劣モジュラ最大化問題と見なすことができれば，より複雑な目的関数や束に対しても近似アルゴリズムを構築できるかもしれない，という期待があった．

まず，束構造に着目しよう．部分空間全体の束 $\mathcal{L}(V)$ は**モジュラ束** (modular lattice) をなす．ここでモジュラ束とは束であって，任意の $X \le Y$ なる $X, Y \in \mathcal{L}$ および $W \in \mathcal{L}$ について

$$(X \vee W) \wedge Y = X \vee (W \wedge Y) \tag{15}$$

が成立するもののことをいう．分配束はモジュラ束である．

モジュラ束に対しても束 DR 劣モジュラ性はそのまま定義できるが，この定義はわれわれの用途には不十分である．$f(X)$ を主成分分析の目的関数とし，v, X, Y, a, b を次のようにおく．

- $v_1 = (1, 0)$; $f(X) = \|\Pi_X v_1\|^2$,
- $X = 0$,
- $Y = \mathrm{span}\{(0, 1)\}$,
- $a = b = (\epsilon/\sqrt{1+\epsilon^2}, 1/\sqrt{1+\epsilon^2})$.

ただし $\epsilon > 0$ は非常に小さな数とする．このとき

$$f(X \vee a) - f(X) = \Theta(\epsilon), \tag{16}$$
$$f(Y \vee b) - f(Y) = 1 \tag{17}$$

となり，束 DR 劣モジュラ性は（近似の意味ですら）成り立たず，主成分分析はこの枠組みでは表現できない．このようなことが起きるのは次の理由による：a はほぼ v_1 に直交しているため，上の第一式のように限界効用は非常に小さな値となる．一方，二つ目の不等式においても Y と b は v_1 にほぼ直交しているので限界効用は小さくなって欲しいと思うのだが，$Y \vee b$ は線型包のせいで全空間を張ってしまうため，元のほぼ直交していたという性質は失われ，限界効用が大きくなってしまう．すなわち，この問題は $Y \vee b$ に Y でも b でもない要素が入ってくる，というモジュラ束の閉包的な性質に由来するものといえる．

この問題を解決するにはモジュラ束を念頭に置いた新しい DR 劣モジュラ性を定義することが必要となる．$a, a' \in \mathrm{adm}(X)$ が $X \vee a = X \vee a'$ を満たすとき $a = a' \mod X$ と書く．この関係は同値類をなす．$[a]_X$ で a を含む同値類をあらわすことにする．

定義 6 (Nakashima and Maehara [8])**.**　モジュラ束

$\mathcal{L}$ 上の関数 $f\colon \mathcal{L} \to \mathbb{R}$ が下向き DR 劣モジュラ関数 (downward DR submodular function) であるとは，任意の $X \leq Y$ および任意の $b \in \mathrm{adm}(Y)$ についてある $b' \in [b]_Y$ が存在し，任意の $a \leq b'$ なる a について

$$f(X \vee a) - f(X) \leq f(Y \vee b) - f(Y) \qquad (18)$$

が成立することをいう．

定義 7 (Nakashima and Maehara [8])**.** モジュラ束 $\mathcal{L}$ 上の関数 $f\colon \mathcal{L} \to \mathbb{R}$ が上向き DR 劣モジュラ関数 (upward DR submodular function) であるとは，任意の $X \leq Y$ および任意の $a \in \mathrm{adm}(X)$ についてある $a' \in [a]_X$ が存在し，任意の $a' \leq b$ なる b について

$$f(X \vee a) - f(X) \leq f(Y \vee b) - f(Y) \qquad (19)$$

が成立することをいう．

上向きと下向きを特に区別せずに言及するとき，単に**方向付き DR 劣モジュラ** (directional DR submodular) と呼ぶ．下向き DR 劣モジュラかつ上向き DR 劣モジュラであっても束 DR 劣モジュラになるとは限らない．分配束では DR 劣モジュラ性，下向き DR 劣モジュラ性，上向き DR 劣モジュラ性はすべて一致するので，これらの新しい概念は分配束における DR 劣モジュラ性の一般化になっている．

主成分分析の目的関数は下向き DR 劣モジュラかつ上向き DR 劣モジュラとなることが証明できる [8]. すなわち，主成分分析はモジュラ束における方向付き DR 劣モジュラ関数最大化問題である．より一般に，任意の非負単調凹関数 ρ について

$$f(X) := \sum_{i=1}^{N} \rho(\|\Pi_X v_i\|^2) \qquad (20)$$

は方向付き DR 劣モジュラ関数となる．この関数は大きなノルムの寄与を軽減するように修正した主成分分析の目的関数と見ることができ，外れ値に強いロバスト主成分分析に利用できる．

さて，方向付き DR 劣モジュラ関数に対しても次の近似アルゴリズムが得られる．どちらも証明は難しくなく，分配束のときに示したのと同じように集合版の証明をモジュラ束の性質を使いながら一行一行追いかけていけば得られる．

定理 8 (Nakashima and Maehara [8])**.** モジュラ束において高さ制約つき単調方向付き DR 劣モジュラ関数最大化問題は $1 - 1/e$ 近似可能．

定理 9 (Nakashima and Maehara [8])**.** モジュラ束において整合的なナップサック制約つき単調方向付き DR 劣モジュラ関数最大化問題は 1/3 近似可能.

モジュラ束に対する連続拡張は未だ得られていない．どのような幾何的構造を考えればよいかを明らかにするのはこの分野の重要な未解決問題といえる．

方向付き DR 劣モジュラ関数は主成分分析を劣モジュラ最大化の文脈で捉えるために導入した新しいクラスであるが，最近，この関数クラスと束上のマトロイドの関係がわかってきた．**マトロイド** (matroid) はベクトルの線型独立性を抽象化した数学的構造であり，増加公理・交換公理・ランク公理などさまざまな同値な公理によって特徴づけできる [12]．マトロイドを束に一般化したものは**スーパーマトロイド** (supermatroid) と呼ばれている [13]．束 $\mathcal{L}$ の部分集合 $\mathcal{I}$ がスーパーマトロイドであるとは，任意の $X \in \mathcal{L}$ について $\{I \in \mathcal{I} : I \leq X\}$ の極大元がすべて同じ高さをもつことをいう．スーパーマトロイドがどのような公理で特徴づけられるかを明らかにするのは重要な問題と認識されており [14]，分配束上では Barnabei et al. [15] によって，凸幾何（下側局所分配束）においては Fujishige et al. [16] および Sano [17] によって特徴づけが与えられてきた．

われわれはモジュラ束上のスーパーマトロイドについて，**ランク関数** (rank function) を用いた次の特徴づけを与えた．$\mathcal{I}$ のランク関数 $r(X)$ を $r(X) = \max\{|I| : I \in \mathcal{I} : I \leq X\}$ で定義する．

定理 10 (Maehara and Nakashima [18])**.** モジュラ束において $\mathcal{I}$ がスーパーマトロイドであることと，ランク関数が以下の 3 条件を満たすことは同値．

1. $r(\bot) = 0$,
2. $X' > X$ かつこれらの間の距離が 1 のとき $r(X') - r(X) \in \{0, 1\}$,
3. $r(X)$ は下向き DR 劣モジュラ．

ここで条件 3 は上向き DR 劣モジュラに置き換えても成立する．

この結果は方向付き DR 劣モジュラ関数が単に最適化しやすいクラスの関数というだけではなく，数学的に深い背景をもつ関数クラスであることを示唆している．

6. おわりに

本稿では機械学習の問題を動機として，束上の劣モジュラ関数最大化の理論を概説した．この分野はごく最近研究が始まったばかりのものであり，まだまだ多くの未解決問題が残されている分野である．また，束だけでなく「集合よりも複雑な構造をもつ空間での劣

モジュラ関数最大化」は実用的にも重要な分野と認識されており，より多くの研究者の参入が待たれている．この記事をきっかけに一人でも多くの方がこれらの分野に挑戦していただけると幸いである．

参考文献

[1] G. Nemhauser, L. Wolsey and M. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Mathematical Programming*, **14**, pp. 265–294, 1978.

[2] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operations Research Letters*, **32**, pp. 41–43, 2004.

[3] J. Vondrák, "Optimal approximation for the submodular welfare problem in the value oracle model," In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC'08)*, pp. 67–74, 2008.

[4] D. Topkis, "Minimizing a submodular function on a lattice," *Operations Research*, **26**, pp. 305–321, 1978.

[5] T. Soma and Y. Yoshida, "Maximizing monotone submodular functions over the integer lattice," *Mathematical Programming*, **172**, pp. 539–563, 2018.

[6] H. Lin and J. Bilmes, "A class of submodular functions for document summarization," In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL'11)*, pp. 510–520, 2011.

[7] C. Gottschalk and B. Peis, "Submodular function maximization over distributive and integer lattices," *arXiv:1505.05423*, 2015.

[8] S. Nakashima and T. Maehara, "Subspace selection via DR-submodular maximization on lattices," In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI'19)*, pp. 4618–4625, 2019.

[9] T. Maehara, S. Nakashima and Y. Yamaguchi, "Multiple knapsack-constrained monotone DR-submodular maximization on distributive lattice," *Mathematical Programming*, 2021.

[10] A. Kulik, H. Shachnai and T. Tamir, "Maximizing submodular set functions subject to multiple linear constraints," In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*, pp. 545–554, 2009.

[11] H. Hirai, "L-convexity on graph structures," *Journal of the Operations Research Society of Japan*, **61**, pp. 71–109, 2018.

[12] J. Oxley, *Matroid Theory*, Oxford University Press, 2006.

[13] F. Dunstan, A. Ingleton and D. Welsh, "Supermatroids," In *Proceedings of the Conference on Combinatorial Mathematics*, pp. 72–122, 1972.

[14] M. Wild, "Weakly submodular rank functions, supermatroids, and the flat lattice of a distributive supermatroid," *Discrete Mathematics*, **308**, pp. 999–1017, 2008.

[15] M. Barnabei, G. Nicoletti and L. Pezzoli, "Matroids on partially ordered sets," *Advances in Applied Mathematics*, **21**, pp. 78–112, 1998.

[16] S. Fujishige, G. Koshevoy and Y. Sano, "Matroids on convex geometries (cg-matroids)," *Discrete Mathematics*, **307**, pp. 1936–1950, 2007.

[17] Y. Sano, "Rank functions of strict cg-matroids," *Discrete Mathematics*, **308**, pp. 4734–4744, 2008.

[18] T. Maehara and S. Nakashima, "Rank axiom of modular supermatroids: A connection with directional DR submodular functions," *arXiv:2009.00200*, 2020.