

最小カットによる構造情報を用いた主成分分析

Principal Component Analysis using Structured Sparsity via Graph Cuts

杉本和正*¹ 河原吉伸*^{1*2} 鷲尾隆*^{1*2}
 Kazumasa Sugimoto Yoshinobu Kawahara Takashi Washio

*¹大阪大学 産業科学研究所 *²JST 科学技術振興機構
 The Institute of Scientific and Industrial Research, Osaka University Japan Science and Technology Agency

In this paper, we propose a *minimum-cut* approach to *Structured Sparse Principal Component Analysis* (SSPCA), which can be calculated much faster than the existing one. PCA is a method widely used in data analysis to explore low-dimensional feature spaces for a given data. In recent years, *structured sparsity*, such as hierarchical structures and group structures among variables in original data, has been widely studied to employ prior information. In this paper, we develop an efficient SSPCA formulation based on minimum-cut using *submodularity* in structured norms.

1. はじめに

ビッグデータに対する解析は、情報化社会となっている昨今、非常に重要視されている課題の一つである。情報化社会となり、世の中には大量のデータが存在し、そのデータを解析する事で例えばマーケティング等に役立っている。しかし、データが大きくなるにつれて、計算量が大きくなるという問題や、解析計算自体が複雑になってしまうという問題により、データの解析は難しくなる。そのため、ビッグデータに対していかにして効率的、効果的な解析を行えるかという事が盛んに研究されている。

近年、データに含まれている構造情報、グループ情報を利用した**構造正則化** [2] という研究が行われている。これは、データに含まれている変数がグループ構造や階層構造を持っている場合に、その構造情報を利用して解析を行うというものである。例えば機械学習における画像解析において、顔認識は古くから研究されてきたテーマの一つである。顔認識では、色彩値をベースにした解析や明度をベースにした解析などを用いて顔を表すピクセルを選択する。この顔認識において、顔の一部を表す一つのピクセルを選択した場合、そのピクセルと隣り合うピクセルもまた顔の一部である可能性が高い。このように、機械学習において解析対象となるデータには、隣り合った変数同士での関連性や、グループ性が存在する事が多々ある。構造正則化においては、このような変数同士の関連性、つまり構造の情報をを用いる事で、従来より効率的、効果的に解析を行う事が出来る。

データ解析に一般的に用いられている手法として**主成分分析** (*Principal Component Analysis*) がある [8]。この主成分分析に上記の構造正則化を組み合わせる事で構造情報を利用して主成分分析を行う**構造正則化主成分分析** (*Structured Sparsity Principal Component Analysis*) が提案されている [7]。本稿ではこの構造正則化主成分分析に対し、グラフ理論における**最小カット**を用いた計算を適用する事で、より大きなデータに対応出来る事を述べる。最小カットとは、グラフ理論の流量問題においてそのグラフの最小可流量を求める問題であるが、機械学習の正則化計算においてこの最小カットを用いる事が出来る事が知られている。最小カットを用いる事により、従来の勾

配法を元にした正則化計算よりも高速に計算が可能となる。そのため、例えば画像解析では、従来の数十倍、数百倍のサイズの画像に対して同程度の計算時間で計算を行う事が出来ると期待される。

本稿の構成は以下である。2. で構造正則化主成分分析の基本的な定式化について述べる。3. では、2. で定式化した構造正則化主成分分析を最小カットを用いた計算が適用出来る形に再度定式化を行い、4. で最小カットを用いて最適化計算を行う方法を述べる。5. で本稿のまとめとする。

2. 構造正則化主成分分析

本章では、主成分分析に変数間の構造情報を取り込んだ構造正則化主成分分析 [7] の基本的な定式化を行う。

2.1 主成分分析

主成分分析 (*Principal Component Analysis*, 以下 PCA) はデータ解析において広く用いられているツールである。本節では最小二乗誤差推定に基づく主成分分析の定式化を行う。

入力データとして $\mathbf{X} \in \mathbb{R}^{n \times p}$ が与えられた時、PCA では \mathbf{X} を主成分行列 $\mathbf{U} \in \mathbb{R}^{n \times r}$ と係数行列 $\mathbf{V}^{p \times r}$ に分解する。 $r < p$ である事から、元のデータ行列 \mathbf{X} に対し、より低次元の主成分行列 \mathbf{U} が得られる事がわかる。PCA では、この行列 \mathbf{U}, \mathbf{V} を求める際に、元のデータ行列 \mathbf{X} の持つ情報量の損失を最小にするように求める。最小二乗誤差を用いた PCA の定式化は式 (1) で表される。

$$\min_{\mathbf{U}, \mathbf{V}} \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \quad (1)$$

元のデータ \mathbf{X} と、分解後のデータ \mathbf{UV}^T の損失情報量を最小化する事で主成分を導出する。

2.2 構造正則化主成分分析

前節で最小二乗誤差推定を用いて定式化した PCA は回帰型の最適化問題に類する事が知られており、lasso [12] や elastic net [14] 等のノルム制約付き回帰問題と同様にスパースノルムを用いた**スパース正則化**を行う事が出来る [15]。更に、Jenatton らは変数間にグループ構造や階層構造等の何らかの構造がある場合、構造正則化を用いる事で、構造情報を元にスパース正則化を行う**構造正則化主成分分析** (*Structured Sparse Principal*

連絡先: 杉本和正, 大阪大学産業科学研究所, 567-0047 大阪府茨木市美穂が丘 8-1-N502-2, sugimoto@ar.sanken.osaka-u.ac.jp

Component Analysis, 以下 SSPCA) を提唱した [7]. 本節では SSPCA の定式化について述べる.

SSPCA は *Dictionary Learning* を基礎として定式化される. 行列 $\mathbf{X} \in \mathbb{R}^{n \times p}$ が与えられた時, Dictionary Learning では行列分解によって辞書行列 $\mathbf{V} \in \mathbb{R}^{p \times r}$ と, 分解係数行列 $\mathbf{U} \in \mathbb{R}^{n \times r}$ を算出する. 行列積 \mathbf{UV}^T が \mathbf{X} の分解と呼ばれる. 行列の分解学習を行う際, 事前情報を利用するために \mathbf{U}, \mathbf{V} の (擬似) ノルム Ω_u, Ω_v を用いる事が出来る [3, 13]. SSPCA の一般式は以下の形となる.

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \quad & \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 + \lambda \sum_{k=1}^r \Omega_v(V^k) \\ \text{s.t.} \quad & \forall k, \Omega_u(U^k) \leq 1 \end{aligned} \quad (2)$$

$\|\bullet\|_F$ はフロベニウスノルムと呼ばれ, 行列 $\mathbf{Y} \in \mathbb{R}^{n \times p}$ に対し, $\|\mathbf{Y}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p Y_{ij}^2$ で定義される. また, U^k, V^k は行列 \mathbf{U}, \mathbf{V} の第 k 列を取り出した n 次元ベクトルである. 正則化パラメータ $\lambda \geq 0$ は正則化の度合いを調節する.

2.3 劣モジュラ関数と Lovász 拡張

3. 及び 4. において前節で定式化した SSPCA の最適化法を述べるが, そこで関数の劣モジュラ性という性質を用いる. 劣モジュラ性とは, 任意の有限集合 $V = \{1, \dots, d\}$ と部分集合 $S \subseteq V, T \subseteq V, S \subset T$ に対し, 集合関数 $F: 2^V \rightarrow \mathbb{R}$ が式 (3) の様な性質を持つ事を指し, 劣モジュラ性を持つ集合関数を劣モジュラ関数という.

$$F(S) + F(T) \geq F(S \cap T) + F(S \cup T) \quad (3)$$

劣モジュラ性は連続領域における凸性に対応する離散領域での概念であり [4], 劣モジュラ性を利用した最適化は連続領域における最大化や最小化と同様に考える事が出来る. しかし, SSPCA(式 (2)) の正則化ノルム Ω_v は連続関数であり, 集合関数の性質である劣モジュラ性は持たない. そこで, Lovász 拡張 [9, 10] を導入する. Lovász 拡張は集合関数の連続緩和の一種であり, 集合関数である劣モジュラ関数と連続領域の性質である凸性を関係付ける概念である. 上記の集合 V, S, T に対し, f の Lovász 拡張 $\hat{f}: \mathbb{R}^d \rightarrow \mathbb{R}$ は式 (4) で定義される.

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^{m-1} (\hat{x}_j - \hat{x}_{j+1}) f(U_j) + \hat{x}_m f(U_m) \quad (4)$$

$\hat{x}_1 \geq \dots \geq \hat{x}_m$ は学習に用いる観測データベクトル \mathbf{x} 中の異なる値を降順に並べたものであり, $U_j = \{i \in V : x_i \geq \hat{x}_j\}$ と定義される.

2.4 構造正則化ノルム

式 (2) で用いた構造正則化ノルム Ω_v は前節で述べた Lovász 拡張の一種であり, 劣モジュラ性を持つ集合関数を連続緩和したものと見なす事が出来る. 例えば, 以下に述べるラプラシアン正則化等は Lovász 拡張に属している事が知られている.

2.4.1 ラプラシアン正則化

グラフ構造 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ と, 各エッジの重み $w_{ij}(i, j \in \mathcal{V})$ が与えられたとする. ただし, \mathcal{V}, \mathcal{E} はそれぞれグラフ中のノードとエッジを表す. また, 学習対象となる各変数が各ノードに属する. グラフ \mathcal{G} 上のカット関数 F_{cut} を正則化ノルム Ω とする事で, グラフ上で連結している変数同士がグループとして同時に選択されやすくなる. このように定めた正則化ノルムをラプ

ラシアン正則化 [1] と呼ぶ.

$$\begin{aligned} \Omega_{\mathcal{L}} &= \sum \{w_{ij} : i \in U, j \in \mathcal{V} \setminus U\} \\ &= \mathbf{e}_U^T \mathbf{L} \mathbf{e}_U \end{aligned} \quad (5)$$

ただし, \mathbf{L} はグラフ・ラプラシアンであり, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ と定義される. \mathbf{D} は対角行列で, グラフ中の各ノードのランクを表す. \mathbf{W} はグラフの隣接行列で各エッジの重みを格納, つまり $W_{ij} = w_{ij}$ とする事でグラフ構造を行列化したものである.

2.4.2 l_1/l_∞ -正則化

変数間のグループ構造を元に構造正則化を議論してきたが, 場合によってはグループの重なりが生じるケースも多々存在する. 例えばグループとして G_1, G_2 の 2 つがあり, ある要素 x_j がそのどちらにも属しているというケースではグループの重なりが生じている. このような場合には, 重なりを許した集合の組 \mathcal{G} を考え, 構造正則化ノルムを式 (6) のように定義する.

$$\Omega_{l_1/l_\infty} = \sum_{g \in \mathcal{G} \cap S \neq \emptyset} d(g) \quad (6)$$

$d: 2^V \rightarrow \mathbb{R}_+$ は任意の非負の集合関数である. このように定義した構造正則化ノルムは, いわゆる l_1/l_∞ 正則化ノルムと等しくなる.

3. 定式化

2. では SSPCA の基本的な定式化を述べた. 本章では 4. で述べる最小カットによる最適化の準備として, SSPCA が線形の最小化問題である事を述べる.

3.1 PCA の行列微分による変形

PCA の一般式を単純化すると以下のように表す事が出来る.

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \quad (7)$$

解を一意的に定めるため, 直行条件 $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ を制約として追加する.

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \quad & \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I} \end{aligned} \quad (8)$$

\mathbf{V} に関して偏微分し, その導関数を 0 とする事で式 (10) を得る.

$$\frac{\partial}{\partial \mathbf{V}} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 = 0 \quad (9)$$

$$\mathbf{V}^T = \mathbf{U}^T \mathbf{X} \quad (10)$$

目的関数 (7) に代入して整理すると

$$\begin{aligned} \min_{\mathbf{U}} \quad & \|\mathbf{X} - \mathbf{U} \mathbf{U}^T \mathbf{X}\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}^T \mathbf{U} = \mathbf{I} \end{aligned} \quad (11)$$

となる.

3.2 線形式への変形

本節では式 (11) が線形式で表せる事を示す. $f(\mathbf{U}) = \|\mathbf{X} - \mathbf{U} \mathbf{U}^T \mathbf{X}\|_F^2$ とおく. 2.2 で述べたように, フロベニウスノルム $\|\bullet\|_F$ は行列 $\mathbf{Y} \in \mathbb{R}^{n \times p}$ に対して $\|\mathbf{Y}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p Y_{ij}^2$ と

定義される。フロベニウスノルムの定義、行列のトレースの性質、及び式 (8) で追加した制約 $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ を用いて、 $f(\mathbf{U})$ は以下のように変形出来る。

$$\begin{aligned}
 f(\mathbf{U}) &= \|\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X}\|_F^2 \\
 &= \text{tr}[(\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X})^T (\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X})] \\
 &= \text{tr}[\mathbf{X}^T \mathbf{X}] - \text{tr}[\mathbf{X}^T \mathbf{U}\mathbf{U}^T \mathbf{X}] \\
 &= \|\mathbf{X}\|_F^2 - \|\mathbf{U}^T \mathbf{X}\|_F^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^p X_{ij}^2 - \sum_{k=1}^r \sum_{l=1}^p (\mathbf{U}^T \mathbf{X})_{kl}^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^p X_{ij}^2 - \sum_{k=1}^r \sum_{l=1}^p \left(\sum_{m=1}^n U_{mk} X_{ml} \right)^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^p X_{ij}^2 - \sum_{k=1}^r \sum_{l=1}^p \sum_{m=1}^n (U_{mk} X_{ml}) \sum_{h=1}^n (U_{hk} X_{hl})
 \end{aligned}$$

ここで、 \mathbf{X} は観測データの行列であるため、第一項の $\sum_{i=1}^n \sum_{j=1}^p X_{ij}^2$ は定数となる。これを C で表すと、 $f(\mathbf{U})$ は以下のように表す事が出来る。

$$f(\mathbf{U}) = - \sum_{l=1}^p \sum_{k=1}^r \sum_{m=1}^n (U_{mk} X_{ml}) \sum_{h=1}^n (U_{hk} X_{hl}) + C \quad (12)$$

今回考えている SSPCA では、構造正則化によるスパース性を、次元 p 方向に適用している。つまり、全体集合 $V = \{1, \dots, p\}$ から構造正則化による特徴選択 (変数選択) によって部分集合 \mathcal{S} を選択すると、その選択結果は p に関する部分にのみ作用する。つまり、 $f(\mathbf{U})$ から部分集合 \mathcal{S} に関する部分のみを選択した関数 $f_{\mathcal{S}}(\mathbf{U})$ は以下のように表される。

$$f_{\mathcal{S}}(\mathbf{U}) = - \sum_{l \in \mathcal{S}} \left\{ \sum_{k=1}^r \sum_{m=1}^n (U_{mk} X_{ml}) \sum_{h=1}^n (U_{hk} X_{hl}) \right\} + C \quad (13)$$

よって、

$$\begin{aligned}
 f(\mathbf{U}) &= - \sum_{l=1}^p \sum_{k=1}^r \sum_{m=1}^n (U_{mk} X_{ml}) \sum_{h=1}^n (U_{hk} X_{hl}) + C \\
 &= -\mathbf{c} \mathbf{e}_{\mathcal{S}} + C
 \end{aligned} \quad (14)$$

と線形に表す事が出来る。ただし、 \mathbf{c} は係数を表す実数ベクトルである。また、 $\mathbf{e}_{\mathcal{S}}$ は例えば $[1, 0, \dots, 1, 1, 0, 0, 1]^T$ のように、 \mathcal{S} の選択結果に依存する $|\mathcal{S}|$ 次元 $0-1$ ベクトルである。よって、SSPCA は、以下のような線形式 + 正則化項 (+ 定数項) という形で表す事が出来る。

$$\min_{\mathcal{S} \subseteq V} -\mathbf{c} \mathbf{e}_{\mathcal{S}} + \lambda \Omega(\mathcal{S}) + C \quad (15)$$

4. 最適化計算

本稿における SSPCA の最適化計算には最小カットを用いるが、その計算方法として *parametric max-flow* [6] を用いる。本章では、劣モジュラ関数の性質及び Lovász 拡張を用いて、SSPCA に *parametric max-flow* を適用する事を考える。

4.1 劣モジュラ最小化問題への変形

簡単のために、以下のようなベクトルでの正則化項つき最適化問題を考える。

$$\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{w} - \mathbf{z}\|_2^2 + \lambda \Omega(\mathbf{w}) \quad (16)$$

これは係数ベクトル \mathbf{w} に関する最小化問題であり、観測ベクトル \mathbf{z} との差を最小化する。 λ は正則化パラメータで、 Ω は劣モジュラ性を持つ構造正則化ノルムである。凸双対性により及び Lovász 拡張に基づいて式 (16) は以下のように式変形する事が出来る [2]。

$$\begin{aligned}
 &\min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{w} - \mathbf{z}\|_2^2 + \lambda \Omega(\mathbf{w}) \\
 &= \min_{\mathbf{w} \in \mathbb{R}^p} \max_{\Omega^*(\mathbf{s}) \leq 1} \frac{1}{2} \|\mathbf{w} - \mathbf{z}\|_2^2 + \lambda \mathbf{s}^T \mathbf{w} \\
 &= \max_{\Omega^*(\mathbf{s}) \leq 1} \min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{w} - \mathbf{z}\|_2^2 + \lambda \mathbf{s}^T \mathbf{w} \\
 &= \max_{\Omega^*(\mathbf{s}) \leq 1} \frac{1}{2} \|\mathbf{z}\|_2^2 - \frac{1}{2} \|\lambda \mathbf{s} - \mathbf{z}\|_2^2
 \end{aligned} \quad (17)$$

劣モジュラ関数の性質を利用し、式 (17) は以下のように書き換えられる。

$$\begin{aligned}
 &\min_{\substack{\forall ACV, \\ s(A) \subset F(A)}} \left\| \mathbf{s} - \frac{1}{\lambda} \mathbf{z} \right\|_2^2 \\
 &= \min_{\substack{\forall ACV, \\ t(A) \subset F(A) - \lambda^{-1} \mathbf{z}(A)}} \|\mathbf{t}\|_2^2
 \end{aligned} \quad (18)$$

ここで、 $\mathbf{t} = \mathbf{s} - 1/\lambda \mathbf{z}$ 、 $\mathbf{w} = \mathbf{z} - \lambda \mathbf{s} = -\lambda \mathbf{t}$ である。このように、式 (16) のような \mathbf{w} に関する正則化項つき最適化問題は、式 (18) のような \mathbf{t} に関する最小化問題に帰着する事ができる。同様の計算により、式 (15) は劣モジュラ最小化問題に帰着する事が出来る。

4.2 分割可能な劣モジュラ関数

劣モジュラ関数の最小化アルゴリズムは多数提唱されており、多項式時間で解く事が可能である。更に、劣モジュラ関数が**分解可能な劣モジュラ関数** [11] である場合、最大流 (最小カット) アルゴリズムを用いて多項式時間アルゴリズムよりも効率的に計算が可能となる。ここで、分解可能な劣モジュラ関数とは、元の問題を複数の小問題に分割して解く事が出来る劣モジュラ関数を指す。式 (15) において、第一項が線形項で第二項 (正則化項) が劣モジュラ項でありである事から全体として劣モジュラ関数になっており、劣モジュラ最小化の多項式時間アルゴリズムが適用可能である。しかし、構造正則化においては構造正則化ノルムは分割可能な劣モジュラ関数の Lovász 拡張となる事が知られており、SSPCA は分解可能な劣モジュラ関数の Lovász 拡張であって最大流アルゴリズムによって多項式時間アルゴリズムよりも高速に計算が可能となる。

4.3 parametric max-flow

図 1 のようなソースノード s 、シンクノード t からなる $s-t$ グラフを考える。図 1 中のシンクノード t に繋がるノードの重みを t_1, \dots, t_d とし、この $s-t$ グラフ中の最大流を *parametric max-flow* を用いて求める。最小カット最大流定理 [5] により、最大流を求める事と最小カットを求める事は等価となる。最大流によって選択されたノード t_1, \dots, t_d の値が、式 (18) における最適解となる。

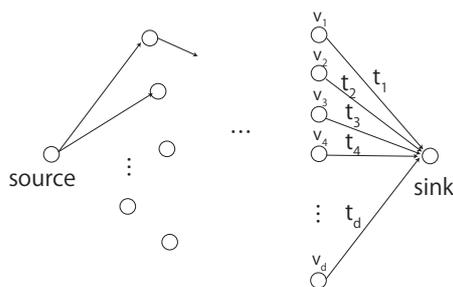


図1 s-t グラフ

このように parametric max-flow を用いる事で、 λ の変化に対して、最適解がどのように変化するかを追跡し、いわゆる正則化パスのように求める事が出来る。従来の勾配ベースの最適化手法においては、最適化パラメータ λ の決定が非常に重要であり、交差検定等を用いて λ を決定する必要があった。parametric max-flow を用いた最適化においては、上述の用に λ の変化に対する解の軌跡が得られるため、交差検定が省略でき、全体としての高速化が実現できる。

5. まとめ

近年盛んに研究されている変数間の構造情報を利用した最適化問題、いわゆる構造正則化に対して、parametric max-flow による最小カットを用いて最適化を行う手法について述べた。本稿では主成分分析に構造正則化を適用する手法を元にし、構造正則化ノルムが分解可能な劣モジュラ関数の Lovász 拡張であるという性質を持つ事を利用して線形問題に定式化する事で最小カットを用いて解ける事を述べた。これにより、従来よりも大規模なデータに対する解析を行う事ができると期待される。また、従来の正則化問題においては正則化パラメータ λ の決定という重要な問題があったが、parametric max-flow を用いる事でその計算を簡略化する事ができるという点について述べた。

参考文献

- [1] R.K. Ando and T. Zhang. Learning on graph with laplacian regularization. *Advances in neural information processing systems*, Vol. 19, p. 25, 2007.
- [2] F. Bach. Structured sparsity-inducing norms through submodular functions. *arXiv preprint arXiv:1008.4220*, 2010.
- [3] F. Bach, J. Mairal, and J. Ponce. Convex sparse matrix factorizations. *arXiv preprint arXiv:0812.1869*, 2008.
- [4] Jack Edmonds. Submodular functions, matroids, and certain polyhedra. *Combinatorial Optimization—Eureka, You Shrink!*, pp. 11–26. Springer, 2003.
- [5] P. Elias, A. Feinstein, and C. Shannon. A note on the maximum flow through a network. *Information Theory, IRE Transactions on*, Vol. 2, No. 4, pp. 117–119, 1956.
- [6] G. Gallo, M.D. Grigoriadis, and R.E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, Vol. 18, No. 1, pp. 30–55, 1989.
- [7] R. Jenatton, G. Obozinski, and F. Bach. Structured sparse principal component analysis. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.

- [8] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.
- [9] László Lovász. Submodular functions and convexity. In *Mathematical Programming The State of the Art*, pp. 235–257. Springer, 1983.
- [10] Kazuo Murota. *Discrete convex analysis*, Vol. 10. Society for Industrial and Applied Mathematics, 1987.
- [11] Peter Stobbe and Andreas Krause. Efficient minimization of decomposable submodular functions. *arXiv preprint arXiv:1010.5511*, 2010.
- [12] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [13] D.M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, Vol. 10, No. 3, pp. 515–534, 2009.
- [14] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 67, No. 2, pp. 301–320, 2005.
- [15] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, Vol. 15, No. 2, pp. 265–286, 2006.