基礎技術としての劣モジュラ最適化

永野　清仁

連続最適化において凸性は問題が扱いやすいことを示唆する．一方，離散的な対象を扱う離散最適化においては凸関数と対応する概念が劣モジュラ関数であると現在広く認識されつつある．また，劣モジュラ関数は応用数学の様々な場面で現れる基本関数であり，この意味でも興味深い研究対象である．本稿では，離散版の凸最適化とでもいうべき，劣モジュラ最適化に関する理論研究・応用研究について近年の動向を含め解説する．

キーワード：劣モジュラ関数，組合せ最適化，機械学習，離散凸解析

1.　はじめに

関数の凸性は連続変数を扱う連続最適化において重要な概念である．その理由の1つは，極小性が大域的最小性を保証するという便利な性質を持つためである．凸関数の理論，凸解析は1970年代に確立され[21]，現在では実際の計算に役立つ理論となっている．例えば，データ集合の解析を通じその中から有用な規則や判断基準などの抽出を目指す機械学習の分野では，サポートベクターマシン（SVM）やカーネル法などで凸最適化手法が成功している．

離散変数を扱う離散最適化において，凸関数と対応する概念として劣モジュラ関数がある．劣モジュラ関数の研究は1970年のEdmonds[1]の結果に端を発している．1983年にLovász[13]は劣モジュラ性がn次元単位超立方体の端点集合$\{0,1\}^n$上での凸性とある意味で等価であることを示した．このLovászの結果は現在では室田[14]の離散凸解析の理論へと発展している．また，劣モジュラ関数は凸性の文脈のみから重要というわけではなく，応用数学における様々な場面で現れる基本的な関数であることもその特長である．本稿では，基礎技術としての劣モジュラ最適化の重要性について解説する．

1.1　劣モジュラ最適化研究の動向

凸関数最小化に対応する劣モジュラ関数の最小化問題に対しては，1999年に組合せ的な多項式時間アルゴリズムが初めて発見された[6][22]．これを一つの契機として，近年離散アルゴリズムの分野において，様々なタイプの劣モジュラ最適化問題に対するアルゴリズムが発達してきている．またここ数年，特に2005年以降，機械学習の分野では劣モジュラ関数の重要性が強く認識されるようになってきている．これは，劣モジュラ最適化手法が近年整備されつつあるという背景とともに，グラフ理論のカット関数や情報理論のエントロピー関数など，機械学習と関連の深い関数が劣モジュラ性を持つことが要因である．

2010年現在，離散アルゴリズムを扱う世界最高の国際会議SODA，STOC，FOCSや，機械学習に関する世界最高の国際会議NIPS，ICMLなどにおいて，劣モジュラ最適化に関する研究成果がコンスタントに相当数発表されている[1]．これは劣モジュラ最適化手法に対する国際的な注目度の高さを示している．

1.2　劣モジュラ最適化問題について

本稿では，劣モジュラ関数に関する基本事項について整理し，続いて劣モジュラ最適化手法に関する研究についていくつかトピックを選んでその内容を概観する．具体的に扱う内容は次の通りである．

- 劣モジュラ関数最大化とその応用（3節）
- 劣モジュラ最適化とクラスタリング（4節）
- 組合せ的制約下の劣モジュラ関数最小化（5節）

劣モジュラ関数最大化は2007年以降の離散アルゴリズムの分野で最も精力的に研究されているテーマの一つであり，応用上も重要な問題である．クラスタリングは劣モジュラ関数最小化などの最適化手法の代表的な応用例である．また，組合せ的な制約の下での劣モ

ながの　きよひと
東京大学 生産技術研究所
〒153-8505 目黒区駒場 4-6-1

[1] 離散最適化や機械学習の分野などでは，注目度の高い国際会議で研究成果を発表することは非常に重要である．その一方，連続最適化分野などに関して現在これは当てはまらない．

ジュラ関数最小化問題については理論的に興味深い結果が明らかになってきている.

2. 劣モジュラ関数とは

劣モジュラ関数は離散世界の凸関数とみなせるという点で理論上重要であると同時に，応用数学の様々な場面で現れる基本的な関数である．本節では基本事項を定義し，基本的な劣モジュラ関数の具体例をいくつか挙げる．さらに凸性と劣モジュラ関数の等価性に関するLovász[13]の結果について述べる．劣モジュラ関数に関する理論については藤重[3]を参照されたい.

2.1 基本事項の定義

自然数 n について，$[n]=\{1,2,\cdots,n\}$ とし，$[n]$ のべき集合を $2^{[n]}$ で表す．すなわち $2^{[n]}=\{S:S\subseteq[n]\}$ である．定義域を $2^{[n]}$ とする実数値関数 f を，$[n]$ を台集合とする**集合関数**とよぶ．任意の部分集合 $S\subseteq[n]$ について，n 次元 0-1 ベクトル $I_S\in\{0,1\}^n$ を

$$(I_S\text{の第}\ i\ \text{成分})=\begin{cases}1 & \text{if}\quad i\in S\\ 0 & \text{if}\quad i\notin S\end{cases}$$

と定義することで $[n]$ の部分集合と n 次元 0-1 ベクトルの1対1対応が得られる．よって，集合関数 $f:2^{[n]}\to\mathbb{R}$ は $\{0,1\}^n$ 上の実数値関数と等価である.

集合関数 $f:2^{[n]}\to\mathbb{R}$ が任意の $S,T\subseteq[n]$ で

$$f(S)+f(T)\geq f(S\cup T)+f(S\cap T) \qquad (2.1)$$

を満たすとき，f を**劣モジュラ関数** (submodular function) とよぶ．劣モジュラ関数について，式(2.1)とは異なる特徴づけを与えておくと理解を深める上でも役に立つ．関数 f が劣モジュラ関数であることは，任意の $S,T\subseteq[n]$ と要素 $i\in V$ で $S\subseteq T$ かつ $i\in[n]-T$ を満たすものについて

$$f(S\cup\{i\})-f(S)\geq f(T\cup\{i\})-f(T) \qquad (2.2)$$

となることと等価となる．式(2.2)は，集合として小さいほど，新しく要素を加えることで生じる関数値の変化の量が大きいことを意味する．この性質から，劣モジュラ関数が「規模の経済性」あるいは「限界効用逓減の法則」を自然にモデル化していると解釈できる.

集合関数 $f:2^{[n]}\to\mathbb{R}$ が任意の $S\subseteq T\subseteq[n]$ で $f(S)\leq f(T)$ となるとき，f は**単調**であるという．任意の $S\subseteq[n]$ で $f([n]-S)=f(S)$ が成立するとき，f は**対称**であるという.

本稿では，劣モジュラ関数 $f:2^{[n]}\to\mathbb{R}$ の現れる最適化問題において，f の情報は関数値オラクルとして与えられるものとする．つまり，与えられた $S\subseteq[n]$ に対し $f(S)$ の値の計算を1回の基本操作とみなす.

2.2 劣モジュラ関数の例

応用上重要と思われる劣モジュラ関数 $f:2^{[n]}\to\mathbb{R}$ の具体例を挙げる.

線形な関数. n 次元ベクトル $w\in\mathbb{R}^n$ が与えられたとき，関数 f を $f(S)=w(S)=\sum_{i\in S}w_i\ (S\subseteq[n])$ と定義すれば f と $-f$ はともに劣モジュラ関数となる.

カット関数. 頂点集合を $[n]$，辺集合を E とする無向グラフ $G=([n],E)$ が与えられ，各辺 $e\in E$ には正の辺重み c_e が定まっているとする．頂点部分集合 $S\subseteq[n]$ について，端点の一方が S，もう一方が $[n]-S$ に含まれる辺集合を $E(S,[n]-S)$ で表すと

$$f(S)=\sum\{c_e:e\in E(S,V-S)\}\ (S\subseteq[n]) \qquad (2.3)$$

で定義される関数 f は**カット関数**と呼ばれ，対称な劣モジュラ関数となる．図1の左側はすべての辺重みが1の場合の例である.

被覆関数. 有限個の点集合 U と，そのいくつかの部分集合 $U_1,\cdots,U_n\subseteq U$ があり，各点 $u\in U$ には正の点重み c_u が定まっているとする．このとき

$$f(S)=\sum\{c_u:u\in\textstyle\bigcup_{i\in S}U_i\}\ (S\subseteq[n]) \qquad (2.4)$$

で定義される関数は**被覆関数**とよばれ，単調な劣モジュラ関数となる．単調性は自明であり，劣モジュラ性も式(2.2)の成立から容易に確認できる．図1の右側は $n=3$ かつ全点重みが1の場合の例である.

情報理論のエントロピー関数. 確率変数 $X_1,\cdots,X_n$ が与えられているとき，$[n]$ の部分集合 S について

$$f(S)=(\{X_i:i\in S\}\text{の同時エントロピー})$$

と定義すると，関数 f は単調な劣モジュラ関数となる．このとき $f(S)$ の値は複数の確率変数 $\{X_i:i\in S\}$ がどれだけ独立にふるまうかを表す量である．詳細については文献[12]などを参照されたい.

さらに，マトロイドのランク関数が劣モジュラであることは組合せ最適化理論において重要である．また容易に確認できるが，劣モジュラ関数は非負のスカラー倍や，和に関して閉じている.

2.3 凸性と劣モジュラ関数の関係

ここでは劣モジュラ性と凸性のある種の等価性について説明する．このために，任意の劣モジュラとは限

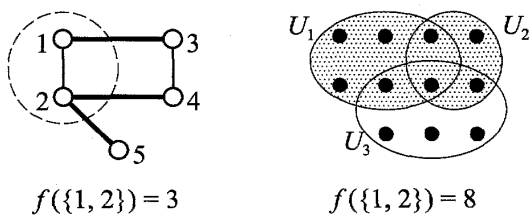

図1 カット関数と被覆関数

らない集合関数 $f : 2^{[n]} \to \mathbb{R}$（ただし $f(\emptyset)=0$）について，その定義域の離散領域 $2^{[n]} \cong \{0,1\}^n$ から連続領域である非負象限 $\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_i \geq 0, \forall i \in [n]\}$ への拡張である **Lovász 拡張** $\hat{f} : \mathbb{R}_+^n \to \mathbb{R}$ を導入する．

任意の $x \in \mathbb{R}_+^n$ を固定し，x の成分の異なる値を大きい順に $\xi_1 > \cdots > \xi_l$ とし，さらに $\xi_{l+1}=0$ とおく．各 $j=1,\cdots,l$ で，$S_j = \{i \in [n] : x_i \geq \xi_j\}$，$\mu_j = \xi_j - \xi_{j+1}$ と定めれば $x = \sum_{j=1}^{l} \mu_j I_{S_j}$ のように分解される．例えば $x=(0.4, 1, 0.4, 0.1)^\top$ なら $x = 0.6 \cdot I_{\{2\}} + 0.3 \cdot I_{\{1,2,3\}} + 0.1 \cdot I_{\{1,2,3,4\}}$ となる．この分解を用いて，Lovász 拡張の関数値 $\hat{f}(x)$ を

$$\hat{f}(x) = \sum_{j=1}^{l} \mu_j f(S_j) \tag{2.5}$$

と定める．定義から各 $S \subseteq [n]$ で $\hat{f}(I_S) = f(S)$ が成立するので，確かに $\hat{f} : \mathbb{R}_+^n \to \mathbb{R}$ は関数 $f : 2^{[n]} \to \mathbb{R}$ の自然な拡張になっている．

Lovász 拡張を用いることで，劣モジュラ関数と凸関数の簡潔な対応関係が得られる．

定理 2.1（Lovász[13]） 集合関数 $f : 2^{[n]} \to \mathbb{R}$ が劣モジュラであるための必要十分条件は，f の Lovász 拡張 $\hat{f} : \mathbb{R}_+^n \to \mathbb{R}$ が凸関数となることである．

劣モジュラ性と凸性の共通する性質が他にもいくつか知られている．これについては文献[14]において詳しく論じられている．

3. 劣モジュラ関数最大化とその応用

関数 $f : 2^{[n]} \to \mathbb{R}$ を単調な劣モジュラ関数とし，k を n 以下の自然数とする．部分集合 S のサイズ $|S|$ に関する制約付きの f の最大化問題を考える：

(MAX) 目的関数：$f(S) \to$ 最大
　　　制約条件：$S \subseteq [n]$，$|S| \leq k$.

関数 f が式(2.4)で定義される被覆関数の場合，問題（MAX）は**最大被覆問題**とよばれる．簡単のため U の点の重みをすべて 1 とすれば，最大被覆問題は，$U_1, \cdots, U_n \subseteq U$ の中から高々 k 個選んだとき，それらにより U の点をできるだけ多く覆う選び方を見つける問題である．図 2 は $n=6$，$|U|=21$，$k=3$ かつ U の全点重みが 1 の場合の最大被覆問題の例とその最適解 $S=\{1,3,5\}$ を示している．

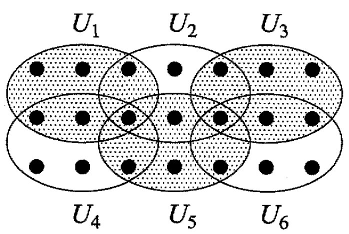

図 2　最大被覆問題とその最適解

本節では，問題（MAX）の難しさやこの問題に対するアルゴリズムについて述べ，その後でどのような問題が劣モジュラ関数最大化に帰着されるかを眺める．

3.1 最大化問題の近似アルゴリズム

問題（MAX）について，関数 f の単調性から $|S| = k$ となる $S \subseteq V$ をすべて調べれば最適解は得られるが，これは，k が大きければ現実的ではない．実は問題（MAX）の特殊ケースである最大被覆問題ですら NP 困難である．Nemhauser ら[19]による問題（MAX）に対する下記の「貪欲」に要素を増やしていく方法が古くから知られており，近似率 $1-1/\mathrm{e} \approx 0.632$ を達成する．つまり問題（MAX）の最適値を OPT とすると，Algorithm 1（表 1）で最終的に得られる $S \subseteq V$ について，$f(S) \geq (1-1/\mathrm{e}) \cdot \mathrm{OPT}$ が成立する．

この近似率 $1-1/\mathrm{e}$ について，Feige[2]は最大被覆問題に対してさえも本質的に最良であることを示している．その一方で，河原-永野-津田-Bilmes[8]は Lovász 拡張を用いたアルゴリズム設計により，問題（MAX）に対する多項式時間解法ではないが厳密に現実的な時間で解く手法を提案している．Sviridenko[23]，Vondrak[25]は問題（MAX）のサイズ制約をそれぞれナップサック制約，マトロイド制約に置き換えても近似率 $1-1/\mathrm{e}$ を達成可能であることを示している．

3.2 劣モジュラ関数最大化の応用

問題（MAX）の応用は数多く研究されており，センサ最適配置問題はその代表的な例である．また，Krause-Gusetrin[12]は単純ベイズモデルにおいて，ある確率変数 Y に関する不確実性を小さくする最適な k 個の確率変数を選択する問題が劣モジュラ関数最大化に帰着されることを示している．さらに，興味深い応用として Kempe ら[10]の社会的ネットワークにおける伝播モデルについて述べる．無向グラフ G

表 1

Algorithm 1（問題 (MAX) の近似解法 [19]）

Input: 単調な劣モジュラ関数 f, k
S を空集合とする．
repeat
　$f(S \cup \{i\})$ を最大化する $i \in [n] - S$ を 1 つ選ぶ．
　$S \cup \{i\}$ を S と置きなおす．
until $|S| = k$

$=([n], E)$ で，$[n]$ は人の集合に対応し，その中の k 人にある同一の商品を渡すことを考える．$i\in[n]$ が商品を持っているとき，i に隣接する人が i の影響を受けその商品を買うかどうかは確率的に決まり，その影響は伝播していく．商品購入者数の期待値を最大化する最初の k 人の選び方を求める問題は，いくつかの伝播モデルで問題（MAX）に帰着される．

4. 劣モジュラ最適化とクラスタリング

関数 $f: 2^{[n]}\to\mathbb{R}$ を劣モジュラ関数とする．劣モジュラ最小化問題は次のように記述される：

(MIN) 目的関数：$f(S)\to$最小
　　　制約条件：$S\subseteq[n]$.

劣モジュラ関数最小化に対しては多項式時間アルゴリズムが知られている[5]．また実用上は藤重-Wolfe のアルゴリズム[3]が高速であることが知られている．劣モジュラ関数最小化の代表的な応用例としてクラスタリングがある．データ点集合 $[n]=\{1, 2, \cdots, n\}$ が与えられたとき，それらを k 個 $(k<n)$ のクラスタ（同種なデータの集まり），$S_1, \cdots, S_k$ に分ける作業をクラスタリングとよぶ．機械学習では教師なし学習とよばれるタイプの問題である．

4.1 グラフ分割の場合

データが無向グラフ $G=([n], E)$ の形で与えられており，関連する 2 点間に辺 $e\in E$ がある場合を考える．データ集合 $[n]$ の k 分割 $S_1, \cdots, S_k$ について，異なるブロックにまたがる辺数は式(2.3)で定義されるカット関数 f を用いて $(1/2)\cdot\sum_{j=1}^{k} f(S_j)$ となる．分割数 k の最適なクラスタリングを求めるには，$\sum_{j=1}^{k} f(S_j)$ の最小化を考えるのが自然である．この最小化問題は**最小 k カット問題**とよばれる NP 困難な問題である．図 3 は最小 3 カットの例を示している．以下では，より一般の場合について議論する．

4.2 一般のクラスタリング問題

一般の劣モジュラ関数 $f: 2^{[n]}\to\mathbb{R}$ を用いて，最小 k カット問題を次の形に一般化する：

(SCP) 目的関数：$\sum_{j=1}^{k} f(S_j)\to$最小
　　　制約条件：$S_1, \cdots, S_k$ は $[n]$ の分割,
　　　　　　　　$S_j\neq\emptyset, \quad \forall j\in\{1, \cdots, k\}$.

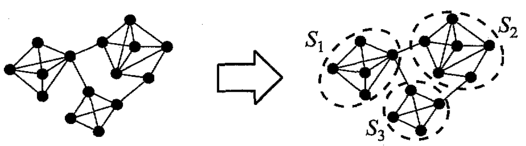

図 3　最小 k カット $(k=3)$

問題（SCP）は，f が対称な場合に限っても NP 困難である．Narasimhan ら[18]は f が対称な場合を扱い，f がカット関数の場合ばかりではなく，機械学習において用いられるいくつかの規準に従うクラスタリングが問題（SCP）に帰着されることを示している．趙-永持-茨木[26]は対称劣モジュラ関数最小化アルゴリズム[20]を用いて，f が対称な場合の問題（SCP）に対し，近似率 $2(1-1/k)$ を達成するアルゴリズムを与えている．

また，問題（SCP）をクラスタリング問題としてとらえたとき，分割数 k はあらかじめ決定していない場合も多い．永野-河原-岩田[15]は一般の f について，クラスタ数 k の決定と $[n]$ の分割の決定を同時に決定する問題を扱い，交差劣モジュラ関数の理論を用いることで，この問題を多項式時間で厳密に最適化するクラスタリング手法を提案している．

問題（SCP）のような単純な最適化問題では，外れ値の存在などにより，最適解がクラスタ間のサイズが偏ったクラスタリングとなり得る．Narasimhan-Bilmes[17]は，問題（SCP）と同様の設定で，さらにクラスタサイズのバランスを考慮に入れたクラスタリング問題に対する局所探索アルゴリズムを提案している．しかしこの手法は，実験的にも理論的にも良くない局所最適解に陥りやすい．河原-永野-岡本[9]はより一般的な問題に対して，2 つの劣モジュラ関数の差の最小化を部分問題として解くような手法を提案しており，計算機実験的にもかなり良いクラスタリングが得られることを報告している．

5. 組合せ的な制約下の最小化問題

前節で扱った劣モジュラ関数最小化問題において，制約条件を付け加えるのは自然である．関数 $f: 2^{[n]}\to\mathbb{R}$ を劣モジュラ関数とし，次の形の最小化問題を考える：

(C-MIN) 目的関数：$f(S)\to$最小
　　　　制約条件：$S\in\mathcal{F}\subseteq 2^{[n]}$.

ここで，許容領域 $\mathcal{F}$ は $[n]$ のある部分集合族である．許容領域 $\mathcal{F}$ が分配束を成す場合，つまり $\mathcal{F}$ が和集合演算と積集合演算に関して閉じているとき，問題（C-MIN）は制約のない場合と同様に多項式時間で解ける[6][22]．その一方で，Svitkina-Fleisher[24]は $\mathcal{F}=\{S\subseteq[n]: |S|\geq k\}$ のように単純なサイズ制約下の最小化問題ですら，近似率が定数となる多項式時間

アルゴリズムの設計が，f が単調な場合でさえも，($P \neq NP$ 予想とは無関係に) 不可能であることを示している．この結果はサイズ制約下の最大化問題について定数近似アルゴリズムが存在するという事実 (3節参照) と対照的である．

最短路問題，巡回セールスマン問題，集合被覆問題などの組合せ最適化問題は，$2^{[n]}$ 上の線形な関数 w を用いて

$$\min\{w(S) : S \in \mathcal{F}\}$$

のような形で記述される．よって，多くの組合せ最適化問題は，線形な関数 w を劣モジュラ関数 f に置き換えることで自然に一般化され，制約付きの劣モジュラ関数最小化問題 (C-MIN) として解釈できる．離散アルゴリズム理論の立場からすれば，線形関数の場合と劣モジュラ関数の場合で，許容領域 $\mathcal{F}$ がどのようなときに難しさが変わるのかに興味がある．

岩田-永野[7]は問題 (C-MIN) において f が非負，$\mathcal{F}$ が集合被覆制約の場合に，貪欲法はうまく動かないが，Lovász 拡張を用いた緩和問題に対する主双対法やラウンディング法では，目的関数が線形の場合，つまり通常の集合被覆問題の場合と同じ近似率が達成されることを示している．また，Goel-Karande-Tripathi-Wang[4]や Koufogiannakis-Young[11]も組合せ的な制約下の劣モジュラ関数最小化問題を扱っている．

最大化問題の場合と同様，制約付きの最小化問題について，理論的な側面ばかりでなく応用を視野に入れた研究が活発になることで，さらなる広がりを持つ研究対象となっていくことが期待される．

6. おわりに

本稿では劣モジュラ最適化の理論・応用研究についていくつかトピックを選んで解説してきた．劣モジュラ最適化においては，対象を抽象化して扱うことで複数の問題に対する共通の一般化を実現している．興味深いことに，劣モジュラ関数の理論あるいはマトロイド理論では，少ない仮定の下で問題を扱うことによって，ときには個別の問題を考えるよりも見通しの良い議論が可能となる．また，初めて出会った最適化問題であっても，問題の劣モジュラ性が示されれば，計算量や問題の難しさに対する最低ラインの保証が可能となる．しかしその一方で，劣モジュラ最適化アルゴリズムの計算量は必ずしも実用的ではない．個別の問題に対する高速化は劣モジュラ最適化手法と同時に議論されるべき内容である．

劣モジュラ最適化が基礎技術として重要であると，分野を超えて広く認識される日が来るであろうか．答えは我々の手の中にある．

参考文献

[1] J. Edmonds: Submodular functions, matroids, and certain polyhedra, In R. Guy, H. Hanani, N. Sauer and J. Schönheim, eds., *Combinatorial Structures and Their Applications*, Gordon and Breach, New York, 1970, 69-87.

[2] U. Feige: A threshold of ln n for approximating set cover, *Journal of the ACM*, **45** (1998), 634-652.

[3] S. Fujishige: *Submodular Functions and Optimization* (Second Edition), Elsevier, Amsterdam, 2005.

[4] G. Goel, C. Karande, P. Tripathi and L. Wang: Approximability of combinatorial problems with multi-agent submodular cost functions, *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science* (FOCS 2009), 755-764.

[5] S. Iwata: Submodular function minimization, *Mathematical Programming*, **112** (2008), 45-64.

[6] S. Iwata, L. Fleischer and S. Fujishige: A combinatorial strongly polynomial algorithm for minimizing submodular functions, *Journal of the ACM*, **48** (2001), 761-777.

[7] S. Iwata and K. Nagano: Submodular function minimization under covering constraints, *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science* (FOCS 2009), 671-680.

[8] Y. Kawahara, K. Nagano, K. Tsuda and J. A. Bilmes: Submodularity cuts and applications, In *Advances in Neural Information Processing Systems* 22 (NIPS 2009), 916-924.

[9] Y. Kawahara, K. Nagano and Y. Okamoto: Submodular fractional programming for balanced clustering, Pattern Recognition Letters, to appear.

[10] D. Kempe, J. Kleinberg and E. Tardos: Maximizing the spread of influence through a social network, In *Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD 2003), 137-146.

[11] C. Koufogiannakis and N. Young: Greedy Δ-approximation algorithm for covering with arbitrary constraints and submodular cost, In: S. Albers *et al*. (eds.) ICALP 2009, Part I, LNCS, vol. 5555, 634-652.

[12] A. Krause and C. Guestrin: Near-optimal non-

myopic value of information in graphical models, In *Proc. of the 21st Annual Conf. on Uncertainty in Artificial Intelligence* (UAI 2005), 324-331.

[13] L. Lovász : Submodular functions and convexity. In A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical Programming—The State of the Art*. Springer-Verlag, 1983, 235-257.

[14] K. Murota : *Discrete Convex Analysis*, SIAM, Philadelphia, 2003.

[15] K. Nagano, Y. Kawahara and S. Iwata : Minimum average cost clustering, To appear in *Advances in Neural Information Processing Systems* 23 (NIPS 2010).

[16] M. Narasimhan and J. Bilmes : Submodular-supermodular procedure with applications to discriminative structure learning, In *Proc. of the 21st Annual Conf. on Uncertainty in Artificial Intelligence* (UAI 2005), 404-410.

[17] M. Narasimhan and J. Bilmes: Local search for balanced submodular clusterings, In *Proc. of the 12th Int'l Joint Conf. on Artificial Intelligence* (IJCAI 2007), 981-986.

[18] M. Narasimhan, N. Jojic and J. Bilmes : Q-clustering, In *Advances in Neural Information Processing Systems* 18 (NIPS 2005), 979-986.

[19] G. L. Nemhauser, L. A. Wolsey and M. L. Fisher : An analysis of approximations for maximizing submodular set functions-I, *Mathematical Programming*, **14** (1978), 265-294.

[20] M. Queyranne : Minimizing symmetric submodular functions, *Mathematical Programming*, **82** (1998), 3-12.

[21] R. T. Rockafellar : *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[22] A. Schrijver : A combinatorial algorithm minimizing submodular functions in strongly polynomial time, *Journal of Combinatorial Theory* (B), **80** (2000), 346-355.

[23] M. Sviridenko : A note on maximizing a submodular set function subject to a knapsack constraint, *Operations Research Letters*, **32** (2004), 41-43.

[24] Z. Svitkina and L. Fleischer : Submodular approximation : sampling-based algorithms and lower bounds, *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science* (2008), 697-706.

[25] J. Vondrak : Optimal approximation for the Submodular Welfare Problem in the value oracle model, *Proc. of the 40th ACM Symp. on Theory of Computing* (STOC 2008), 67-74.

[26] L. Zhao, H. Nagamochi and T. Ibaraki : Approximating the minimum k-way cut in a graph via minimum 3-way cuts, *Journal of Combinatorial Optimization*, **5** (2001), 397-410.