Fast Decomposable Submodular Function Minimization using Constrained Total Variation

K S Sesh Kumar [1]

Francis Bach [2] Thomas Pock [3]

[1]Data Science Institute
Imperial College London, UK
s.karri@imperial.ac.uk

[2]INRIA and Ecole normale superieure
PSL Research University, Paris France.
francis.bach@inria.fr

[3]Institute of Computer Graphics and Vision,
Graz University of Technology, Graz, Austria.
pock@icg.tugraz.at

Neural Information Processing Systems, 2019.

Submodular functions

- **Ground Set :** V of n elements
- **Definition**: $F : 2^V \to \mathbb{R}$ is **submodular** if and only if

$$\forall A, B \subseteq V, F(A) + F(B) \geq F(A \cup B) + F(A \cap B)$$

  - Equality for *modular* functions
  - Always assume $F(\varnothing) = 0$.

- **Equivalent definition** :

$$\forall k \in V, A \to F(A \cup \{k\}) - F(A) \text{ is non increasing}$$

  - Diminishing returns property

Decomposable SFM

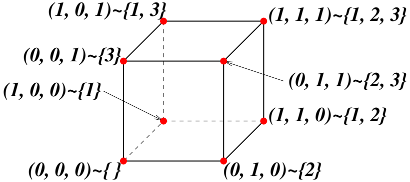
- **Ground Set :** V of n elements
- **Definition**:$\forall i \in [r]$, $F_i : 2^V \to \mathbb{R}$ is **submodular** if and only if

$$\forall A, B \subseteq V, F_i(A) + F_i(B) \geq F_i(A \cup B) + F_i(A \cap B)$$

  - Equality for *modular* functions
  - Always assume $F_i(\varnothing) = 0$.
- **Goal :** $(\mathrm{SFM}\mathbf{D})$ $$\min_{A \subset V} \sum_{i=1}^{r} F_i(A)$$
- **Oracles :** $(\mathrm{SFM}\mathbf{D}_\mathrm{i})$ $$\min_{A \subset V} F_i(A)$$
- **Contribution :**
  - Propose a new continuous optimization problem and algorithms to optimize it using $\mathrm{SFM}\mathbf{D}_\mathrm{i}$ oracles.
  - Derive the solution of $\mathrm{SFM}\mathbf{D}$ from the solution of the continuous optimization problem.

Subsets as pseudo boolean functions

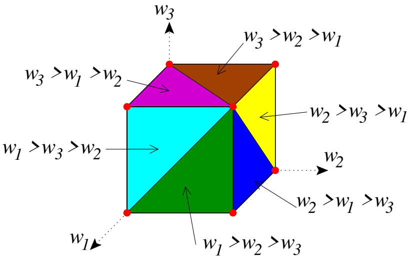
- Subsets may be identified with elements of $\{0,1\}^p$.



Lovász extension

- Given any set-function F and w such that $w_{j_1} \geqslant \cdots \geqslant w_{j_p}$,

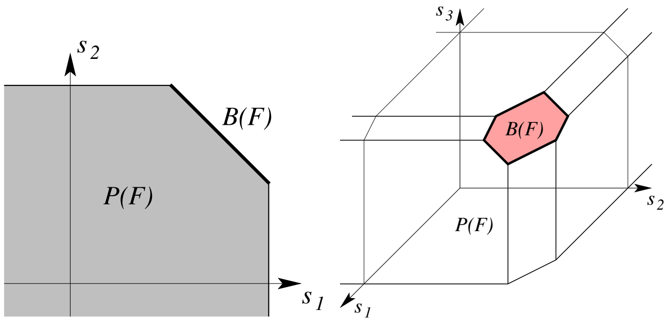
$$f(w) = \sum_{k=1}^{p-1} (w_{j_k} - w_{j_{k+1}}) F(\{j_1, \ldots, j_k\}) + w_{j_p} F(\{j_1, \ldots, j_p\})$$



- **Theorem**: F is submodular if and only if f is convex

Submodular Functions and base polyhedra

- Submodular polyhedron:
  $P(F) = \{s \in \mathbb{R}^p,\ \forall A \subset V,\ s(A) \leqslant F(A)\}$
- Base polyhedron: $B(F) = P(F) \cap \{s(V) = F(V)\}$



- **Representation of $f(w)$ as a support function**:

$$\boxed{f(w) = \max_{s \in B(F)} w^\top s}$$

Continuous Optimization

- Goal $(\mathrm{SFM}\mathbf{D})$ $\min_{A \subset V} \sum_{i=1}^{r} F_i(A)$,
- Oracles $(\mathrm{SFM}\mathbf{D}_\mathrm{i})$ $\min_{A \subset V} F_i(A)$,
- $(\mathrm{SFM}\mathbf{C})$ $\min_{w \in \mathbb{R}^n} \sum_{i=1}^{r} f_i(w) + \sum_{j=1}^{n} \psi(w_j)$,
- $(\mathrm{SFM}\mathbf{C}_\mathrm{i})$ $\min_{w \in \mathbb{R}^n} f_i(w) + \sum_{j=1}^{n} \psi(w_j)$,
- **Related work.** Total variation [Jegelka et al., 2013, Sesh Kumar and Bach, 2017, Ene et al., 2017] considers $\psi(v) = \frac{1}{2}v^2$.
- **Our work.** Let $\varepsilon \in \mathbb{R}_+$ and

$$\psi(w) = \begin{cases} \frac{1}{2}w^2 & \text{if } |w| \leqslant \varepsilon, \\ +\infty & \text{otherwise,} \end{cases}$$

- **Approach.** $\mathrm{SFM}\mathbf{D}_\mathrm{i} \rightarrow \mathrm{SFM}\mathbf{C}_\mathrm{i} \rightarrow \mathrm{SFM}\mathbf{C} \rightarrow \mathrm{SFM}\mathbf{D}$

$\mathrm{SFM}\mathbf{D}_{\mathrm{i}} \rightarrow \mathrm{SFM}\mathbf{C}_{\mathrm{i}} \rightarrow \mathrm{SFM}\mathbf{C} \rightarrow \mathrm{SFM}\mathbf{D}$

1: **Input :** $\mathrm{SFM}\mathbf{D}_i$ for $F_i : 2^V \rightarrow \mathbb{R}$ and $\varepsilon \in \mathbb{R}_+$.
2: **Output :** (w^*, s^*) primal-dual optimal pair of $\mathrm{SFM}\mathbf{C}_i$ for f_i.
3: $A_+ = \mathrm{argmin}_{A \subset V} F_i(A) + \varepsilon|A|$ with a dual certificate $s_+ \in B(F_i)$.
4: $A_- = \mathrm{argmin}_{A \subset V} F_i(A) - \varepsilon|A|$ with a dual certificate $s_- \in B(F_i)$.
5: $w^*(A_+) = -\varepsilon$, $s^*(A_+) = s_+$, $w^*(V \setminus A_-) = \varepsilon$, $s^*(V \setminus A_-) = s_-$
6: $U := A_- \setminus A_+$ and a discrete function $G_i : 2^U$ s.t. $G_i(B) = F_i(A_+ \cup B) - F_i(A_+)$ with Lovász extension $g_i : \mathbb{R}^{|U|} \rightarrow \mathbb{R}$.
7: Solve for optimal solutions of $\min_{w \in \mathbb{R}^{|U|}} g_i(w) + \frac{1}{2}w^2$ and its dual using divide-and-conquer algorithm [Sesh Kumar and Bach, 2017] to obtain (w_U^*, s_U^*).
8: $(w^*(U), s^*(U)) = (w_U^*, s_U^*)$

$\mathrm{SFM}\mathbf{D}_{\mathrm{i}} \rightarrow \mathrm{SFM}\mathbf{C}_{\mathrm{i}} \rightarrow \mathrm{SFM}\mathbf{C} \rightarrow \mathrm{SFM}\mathbf{D}$

- SFM**C**
  - Primal $\min_{w \in [-\varepsilon,\varepsilon]^n} \sum_{i=1}^r f_i(w) + \frac{1}{2}\|w\|_2^2$.
  - Dual $\max_{(s_1,\ldots,s_r)\in\mathbb{R}^{n\times r}} -\sum_{i=1}^r g_i^*(s_i) - \frac{1}{2}\|\sum_{i=1}^r s_i\|_2^2$, where

$$g_i^*(s_i) = \inf_{t_i \in B(F_i)} \sup_{w\in[-\varepsilon,\varepsilon]^n} w^\top(s_i - t_i) = \varepsilon \inf_{t_i\in B(F_i)} \|s_i - t_i\|_1.$$

- Optimization algorithms
  - BCD

$$\forall i \in [r],\ s_i^{\mathrm{new}} = \operatorname*{argmin}_{s_i^{\mathrm{new}}\in\mathbb{R}^n} \ g_i^*(s_i^{\mathrm{new}}) + \frac{1}{2}\|\sum_{j=1}^i s_j^{\mathrm{new}} + \sum_{j=i+1}^r s_j\|_2^2.$$

  - Acceleration for $r = 2$

$$\begin{aligned} s_2^{\mathrm{new}} &= \operatorname*{argmin}_{s_2^{\mathrm{new}}\in\mathbb{R}^n} g_2^*(s_2^{\mathrm{new}}) + \frac{1}{2}\|t_1 + s_2^{\mathrm{new}}\|^2 \\ s_1^{\mathrm{new}} &= \operatorname*{argmin}_{s_1^{\mathrm{new}}\in\mathbb{R}^n} g_1^*(s_1^{\mathrm{new}}) + \frac{1}{2}\|s_1^{\mathrm{new}} + s_2^{\mathrm{new}}\|_2^2 \\ t_1^{\mathrm{new}} &= s_1^{\mathrm{new}} + \beta(s_1^{\mathrm{new}} - s_1), \end{aligned}$$

with $\beta = (t-1)/(t+2)$ at iteration t

$\mathrm{SFMD_i} \to \mathrm{SFMC_i} \to \mathrm{SFMC} \to \mathrm{SFMD}$

$$\boxed{\min_{w\in\mathbb{R}^n} \sum_{i=1}^{r} f_i(w) + \sum_{j=1}^{n} \psi(w_j) \Rightarrow \min_{A\subset V} \sum_{i=1}^{r} F_i(A)}$$

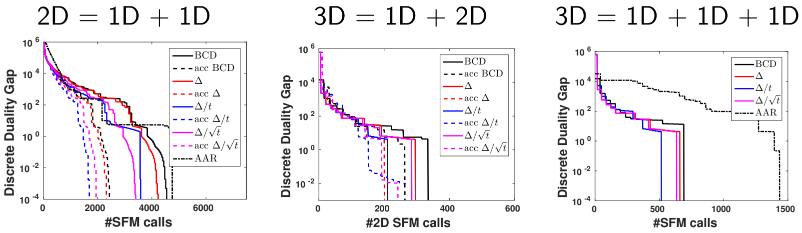
- If w^* is the optimal solution of $\mathrm{SFM}\mathbf{C}$, then $\{w^* \geq 0\}$ is the optimal solution of $\mathrm{SFM}\mathbf{D}$.

Proposition

Given a feasible primal candidate w for $\mathrm{SFM}\mathbf{C}$ with suboptimality η_{C}, one of the suplevel sets $\{w \geqslant \alpha\}$ of w is an η_{D}-optimal minimizer of $\mathrm{SFM}\mathbf{D}$, with $\eta_{\mathrm{D}} = \frac{\eta_{\mathrm{C}}}{4\varepsilon} + \sqrt{\frac{\eta_{\mathrm{C}} n}{2}}$.

- $\eta_{\mathrm{C}} = \frac{\Delta^2}{t^\alpha}$ where Δ is a notion of diameter of the base polytopes and $\alpha = 2$ for accelerated algorithms and $\alpha = 1$ for plain algorithms.
- $\eta_{\mathrm{D}} = \frac{\Delta\sqrt{n}}{t^{\alpha/2}} + \frac{\Delta^2}{\varepsilon t^\alpha}$.

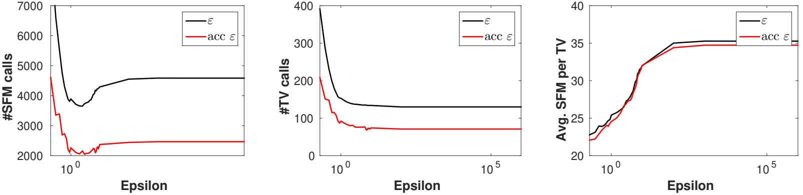
Experiments - Comparison to state-of-the-art



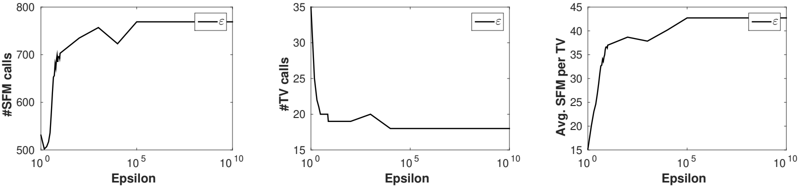
- 2D image of size $n = 2400 \times 2400 = 5.8 \times 10^6$.
- 3D volumetric surface of size $n = 102 \times 100 \times 79 = 8.1 \times 10^5$.

Experiments - Oracle calls for varying ε

Oracle calls for 2D = 1D + 1D for different ε



Oracle calls for 3D = 1D + 1D + 1D for different ε



- 2D image of size $n = 2400 \times 2400 = 5.8 \times 10^6$.
- 3D volumetric surface of size $n = 102 \times 100 \times 79 = 8.1 \times 10^5$.

Future work

- Further speed-ups may be achieved by extended the proposed algorithms to [Ene et al., 2017, Li and Milenkovic, 2018]
- Easily parallelizable and it would be interesting to compare to dedicated parallel algorithms for graph cuts [Shekhovtsov and Hlavác, 2011]
- Could be extended to more general submodular optimization problems [Bach, 2016]

References I

Francis Bach. Submodular functions: from discrete to continuous domains. Mathematical Programming, pages 1–41, 2016.

Alina Ene, Huy Nguyen, and László A Végh. Decomposable submodular function minimization: discrete and continuous. In Advances in Neural Information Processing Systems, 2017.

S. Jegelka, F. Bach, and S. Sra. Reflection methods for user-friendly submodular optimization. In Advances in Neural Information Processing Systems, 2013.

P. Li and O. Milenkovic. Revisiting decomposable submodular function minimization with incidence relations. In Advances in Neural Information Processing Systems, 2018.

K. S. Sesh Kumar and Francis Bach. Active-set methods for submodular minimization problems. Journal of Machine Learning Research, 18(1): 4809–4839, 2017.

References II

A. Shekhovtsov and V. Hlavác. A distributed mincut/maxflow algorithm combining path augmentation and push-relabel. In Energy Minimization Methods in Computer Vision and Pattern Recognition, 2011.

Thank you. Questions?