

Fast Decomposable Submodular Function Minimization using Constrained Total Variation

K S Sesh Kumar ¹
Francis Bach ² Thomas Pock ³

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

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

³Institute 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

Submodular functions

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

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

Submodular functions

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $F : 2^V \rightarrow \mathbb{R}$ 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 $F(\emptyset) = 0$.

Submodular functions

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $F : 2^V \rightarrow \mathbb{R}$ 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 $F(\emptyset) = 0$.
 - ▶ **Equivalent definition** :
- $$\forall k \in V, A \rightarrow F(A \cup \{k\}) - F(A) \text{ is non increasing}$$

Submodular functions

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $F : 2^V \rightarrow \mathbb{R}$ 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 $F(\emptyset) = 0$.

- ▶ **Equivalent definition** :

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

- ▶ Diminishing returns property

Decomposable SFM

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $\forall i \in [r]$, $F_i : 2^V \rightarrow \mathbb{R}$ 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 $F_i(\emptyset) = 0$.
- ▶ **Goal** : (SFMD) $\min_{A \subseteq V} \sum_{i=1}^r F_i(A)$

Decomposable SFM

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $\forall i \in [r]$, $F_i : 2^V \rightarrow \mathbb{R}$ 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 $F_i(\emptyset) = 0$.
- ▶ **Goal** : (SFMD) $\min_{A \subseteq V} \sum_{i=1}^r F_i(A)$
- ▶ **Oracles** : (SFMD_i) $\min_{A \subseteq V} F_i(A)$

Decomposable SFM

- ▶ **Ground Set** : V of n elements
- ▶ **Definition**: $\forall i \in [r]$, $F_i : 2^V \rightarrow \mathbb{R}$ 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 $F_i(\emptyset) = 0$.

- ▶ **Goal** : (SFMD) $\min_{A \subseteq V} \sum_{i=1}^r F_i(A)$

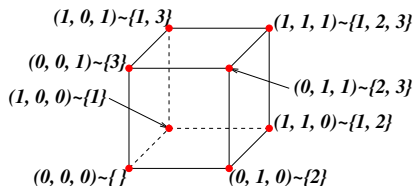
- ▶ **Oracles** : (SFMD_i) $\min_{A \subseteq V} F_i(A)$

- ▶ **Contribution** :

- ▶ Propose a new continuous optimization problem and algorithms to optimize it using SFMD_i oracles.
- ▶ Derive the solution of SFMD from the solution of the continuous optimization problem.

Subsets as pseudo boolean functions

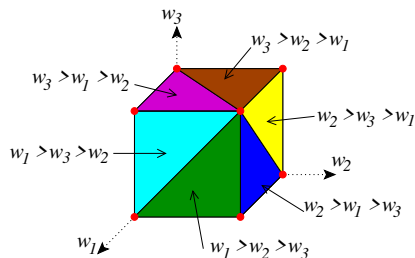
- Subsets may be identified with elements of $\{0, 1\}^P$.



Lovász extension

- Given **any** set-function F and w such that $w_{j_1} \geq \dots \geq w_{j_p}$,

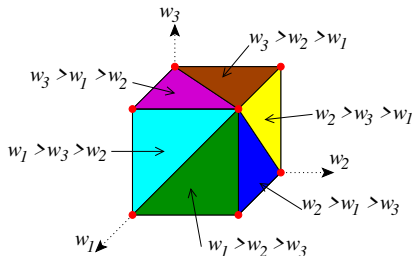
$$f(w) = \sum_{k=1}^{p-1} (w_{j_k} - w_{j_{k+1}}) F(\{j_1, \dots, j_k\}) + w_{j_p} F(\{j_1, \dots, j_p\})$$



Lovász extension

- ▶ Given **any** set-function F and w such that $w_{j_1} \geq \dots \geq w_{j_p}$,

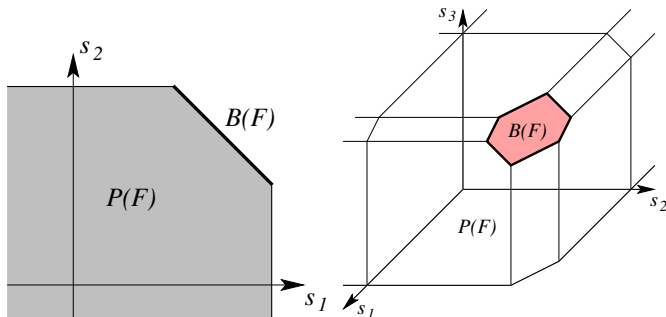
$$f(w) = \sum_{k=1}^{p-1} (w_{j_k} - w_{j_{k+1}}) F(\{j_1, \dots, j_k\}) + w_{j_p} F(\{j_1, \dots, j_p\})$$



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

Submodular Functions and base polyhedra

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



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

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

Continuous Optimization

- ▶ Goal (SFMD) $\min_{A \in V} \sum_{i=1}^r F_i(A),$
- ▶ Oracles (SFMD_i) $\min_{A \in V} F_i(A),$

Continuous Optimization

- ▶ Goal (SFMD) $\min_{A \subset V} \sum_{i=1}^r F_i(A),$
- ▶ Oracles (SFMD_i) $\min_{A \subset V} F_i(A),$
- ▶ (SFMC) $\min_{w \in \mathbb{R}^n} \sum_{i=1}^r f_i(w) + \sum_{j=1}^n \psi(w_j),$
- ▶ (SFMC_i) $\min_{w \in \mathbb{R}^n} f_i(w) + \sum_{j=1}^n \psi(w_j),$

Continuous Optimization

- ▶ Goal (SFMD) $\min_{A \subset V} \sum_{i=1}^r F_i(A),$
- ▶ Oracles (SFMD_i) $\min_{A \subset V} F_i(A),$
- ▶ (SFMC) $\min_{w \in \mathbb{R}^n} \sum_{i=1}^r f_i(w) + \sum_{j=1}^n \psi(w_j),$
- ▶ (SFMC_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 $\psi(v) = \frac{1}{2}v^2.$

Continuous Optimization

- ▶ Goal (SFMD) $\min_{A \subset V} \sum_{i=1}^r F_i(A),$
- ▶ Oracles (SFMD_i) $\min_{A \subset V} F_i(A),$
- ▶ (SFMC) $\min_{w \in \mathbb{R}^n} \sum_{i=1}^r f_i(w) + \sum_{j=1}^n \psi(w_j),$
- ▶ (SFMC_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

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

Continuous Optimization

- ▶ Goal (SFMD) $\min_{A \subset V} \sum_{i=1}^r F_i(A),$
- ▶ Oracles (SFMD_i) $\min_{A \subset V} F_i(A),$
- ▶ (SFMC) $\min_{w \in \mathbb{R}^n} \sum_{i=1}^r f_i(w) + \sum_{j=1}^n \psi(w_j),$
- ▶ (SFMC_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
$$\psi(w) = \begin{cases} \frac{1}{2}w^2 & \text{if } |w| \leq \varepsilon, \\ +\infty & \text{otherwise,} \end{cases}$$
- ▶ **Approach.** SFMD_i \rightarrow SFMC_i \rightarrow SFMC \rightarrow SFMD

$\text{SFMD}_i \rightarrow \text{SFMC}_i \rightarrow \text{SFMC} \rightarrow \text{SFMD}$

- 1: **Input** : SFMD_i for $F_i : 2^V \rightarrow \mathbb{R}$ and $\varepsilon \in \mathbb{R}_+$.
- 2: **Output** : (w^*, s^*) primal-dual optimal pair of SFMC_i for f_i .
- 3: $A_+ = \operatorname{argmin}_{A \subseteq V} F_i(A) + \varepsilon|A|$ with a dual certificate $s_+ \in B(F_i)$.
- 4: $A_- = \operatorname{argmin}_{A \subseteq V} F_i(A) - \varepsilon|A|$ 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: $(w^*(U), s^*(U)) = (w_U^*, s_U^*)$

$$\text{SFMD}_i \rightarrow \text{SFMC}_i \rightarrow \text{SFMC} \rightarrow \text{SFMD}$$

► SFMC

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

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

SFMD_i → SFMC_i → SFMC → SFMD

► SFMC

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

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

SFMD_i → SFMC_i → SFMC → SFMD

► SFMC

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

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

► Acceleration for $r = 2$

$$s_2^{\text{new}} = \underset{s_2^{\text{new}} \in \mathbb{R}^n}{\operatorname{argmin}} g_2^*(s_2^{\text{new}}) + \frac{1}{2} \|t_1 + s_2^{\text{new}}\|^2$$

$$s_1^{\text{new}} = \underset{s_1^{\text{new}} \in \mathbb{R}^n}{\operatorname{argmin}} g_1^*(s_1^{\text{new}}) + \frac{1}{2} \|s_1^{\text{new}} + s_2^{\text{new}}\|^2$$

$$t_1^{\text{new}} = s_1^{\text{new}} + \beta(s_1^{\text{new}} - s_1),$$

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

$$\text{SFMD}_i \rightarrow \text{SFM}\mathbf{C}_i \rightarrow \text{SFMC} \rightarrow \text{SFMD}$$

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

- If w^* is the optimal solution of SFMC, then $\{w^* \geq 0\}$ is the optimal solution of SFMD.

Proposition

Given a feasible primal candidate w for SFMC with suboptimality η_C , one of the suplevel sets $\{w \geq \alpha\}$ of w is an η_D -optimal minimizer of SFMD, with $\eta_D = \frac{\eta_C}{4\epsilon} + \sqrt{\frac{\eta_C n}{2}}$.

$$\text{SFMD}_i \rightarrow \text{SFMC}_i \rightarrow \text{SFMC} \rightarrow \text{SFMD}$$

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

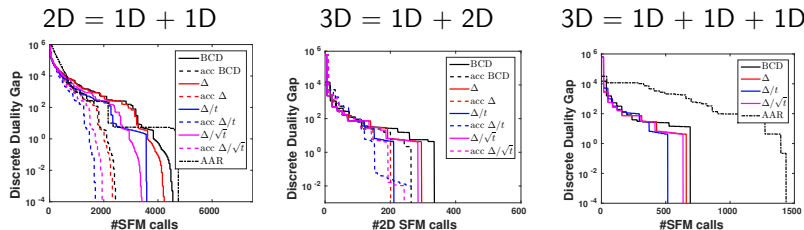
- If w^* is the optimal solution of SFMC, then $\{w^* \geq 0\}$ is the optimal solution of SFMD.

Proposition

Given a feasible primal candidate w for SFMC with suboptimality η_C , one of the suplevel sets $\{w \geq \alpha\}$ of w is an η_D -optimal minimizer of SFMD, with $\eta_D = \frac{\eta_C}{4\varepsilon} + \sqrt{\frac{\eta_C n}{2}}$.

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

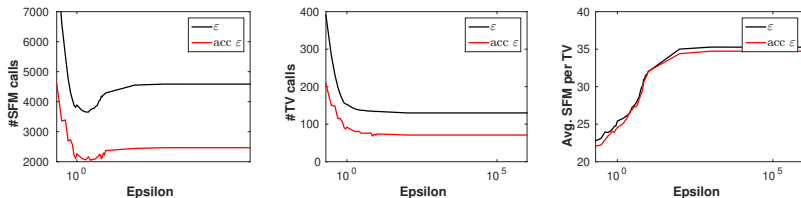
Experiments - Comparison to state-of-the-art



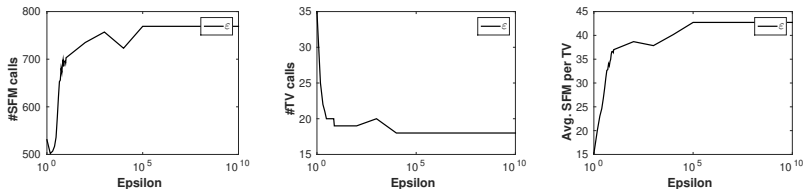
- ▶ 2D image of size $n = 2400 \times 2400 = 5.8 \times 10^6$.
- ▶ 3D volumetric surface of size $n = 102 \times 100 \times 79 = 8.1 \times 10^5$.

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 Hlaváč, 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. Hlaváč. 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?