Fast Decomposable Submodular Function Minimization using Constrained Total Variation

K. S. Sesh Kumar — DSI, Imperial College London, UK

Francis Bach — INRIA/ENS/PSL, Paris, France

Thomas Pock — Graz Institute of Technology, Graz, Austria

1 . Decomposable SFM

- **Ground Set :** V of n elements
- **Submodular functions :** $\forall i \in [r], F_i : 2^V \to \mathbb{R}, F_i(\varnothing) = 0,$

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

- **Goal :** $(\text{SFM}\mathbf{D}) \ \min_{A \subset V} \sum_{i=1}^{r} F_i(A)$
- **Oracles :** $(\text{SFM}\mathbf{D}_\text{i}) \ \min_{A \subset V} F_i(A)$
- **Contribution :** Propose a related new continuous optimization problem and algorithms to optimize it using SFM$\mathbf{D}_\text{i}$ oracles. The solution of the new continuous optimization problem solves SFM$\mathbf{D}$

2 . Continuous Optimization

- Given all submodular function $i \in [r]$, $F_i : 2^V \to \mathbb{R}$, the corresponding Lovász extension $f_i : \mathbb{R}^n \to \mathbb{R}$ is given by

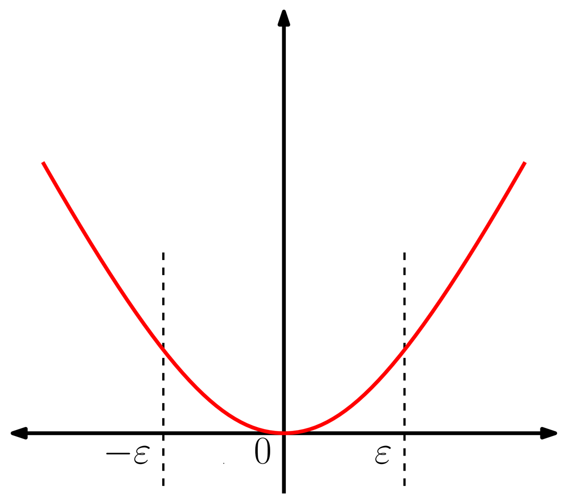
$$f_i(w) = \sum_{k=1}^{n-1} (w_{j_k} - w_{j_{k+1}}) F_i(\{j_1, \ldots, j_k\}) + w_{j_p} F_i(\{j_1, \ldots, j_p\}),$$

where $w \in \mathbb{R}^n$ such that $w_{j_1} \geqslant \cdots \geqslant w_{j_p}$.

- **Continuous Optimization**

$$(\text{SFM}\mathbf{C}) \ \min_{w \in \mathbb{R}^n} \sum_{i=1}^{r} f_i(w) + \sum_{j=1}^{n} \psi(w_j),$$

$$(\text{SFM}\mathbf{C}_\text{i}) \ \min_{w \in \mathbb{R}^n} f_i(w) + \sum_{j=1}^{n} \psi(w_j),$$



where $\psi : \mathbb{R} \to \mathbb{R}$ is a convex function with Fenchel-conjugate everywhere.

- **Related work.** Total variation [1, 2, 3] 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:** SFM$\mathbf{D}_\text{i}$ $\to$ SFM$\mathbf{C}_\text{i}$ $\to$ SFM$\mathbf{C}$ $\to$ SFM$\mathbf{D}$

3 . SFM$\mathbf{D}_\text{i}$ $\to$ SFM$\mathbf{C}_\text{i}$

1: **Input :** SFM$\mathbf{D}_i$ for $F_i : 2^V \to \mathbb{R}$ and $\varepsilon \in \mathbb{R}_+$.
2: **Output :** (w^*, s^*) primal-dual optimal pair of SFM$\mathbf{C}_i$.
3: $A_+ = \text{argmin}_{A \subset V} F_i(A) + \varepsilon|A|$ with a dual certificate $s_+ \in B(F_i)$.
4: $A_- = \text{argmin}_{A \subset V} F_i(A) - \varepsilon|A|$ 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 $g_i : \mathbb{R}^{|U|} \to \mathbb{R}$.
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 [2] to obtain (w_U^*, s_U^*).
8: $(w^*(U), s^*(U)) = (w_U^*, s_U^*)$

4 . SFM$\mathbf{C}_\text{i}$ $\to$ SFM$\mathbf{C}$

- SFM$\mathbf{C}$
  - 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

$$g_i(w) = \begin{cases} f_i(w) & \text{if } |w| \leqslant \varepsilon, \\ +\infty & \text{otherwise.} \end{cases}$$

  with the Fenchel-conjugate

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

- Optimization algorithms
  - BCD

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

  - Acceleration for $r = 2$

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

$$s_1^{\text{new}} = \underset{s_1^{\text{new}} \in \mathbb{R}^n}{\text{argmin}} \ g_1^*(s_1^{\text{new}}) + \frac{1}{2}\|s_1^{\text{new}} + s_2^{\text{new}}\|_2^2$$

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

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

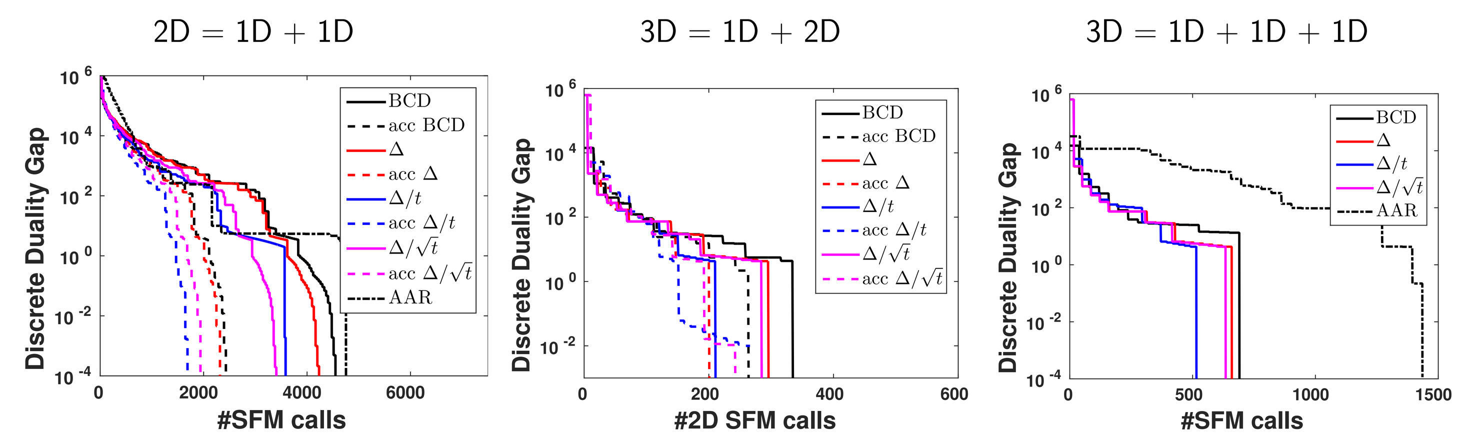
5 . SFM$\mathbf{C}$ $\to$ SFM$\mathbf{D}$

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

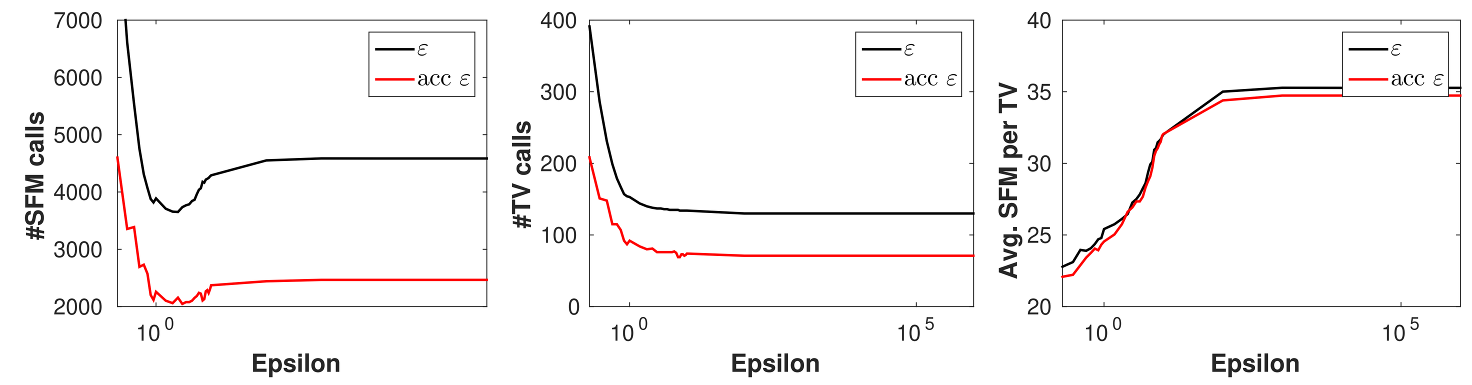
Proposition 1. *Given a feasible primal candidate w for SFM$\boldsymbol{C}$ with suboptimality η_C, one of the suplevel sets $\{w \geqslant \alpha\}$ of w is an η_D-optimal minimizer of SFM$\boldsymbol{D}$, with $\eta_\text{D} = \frac{\eta_\text{C}}{4\varepsilon} + \sqrt{\frac{\eta_\text{C} n}{2}}$.*

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

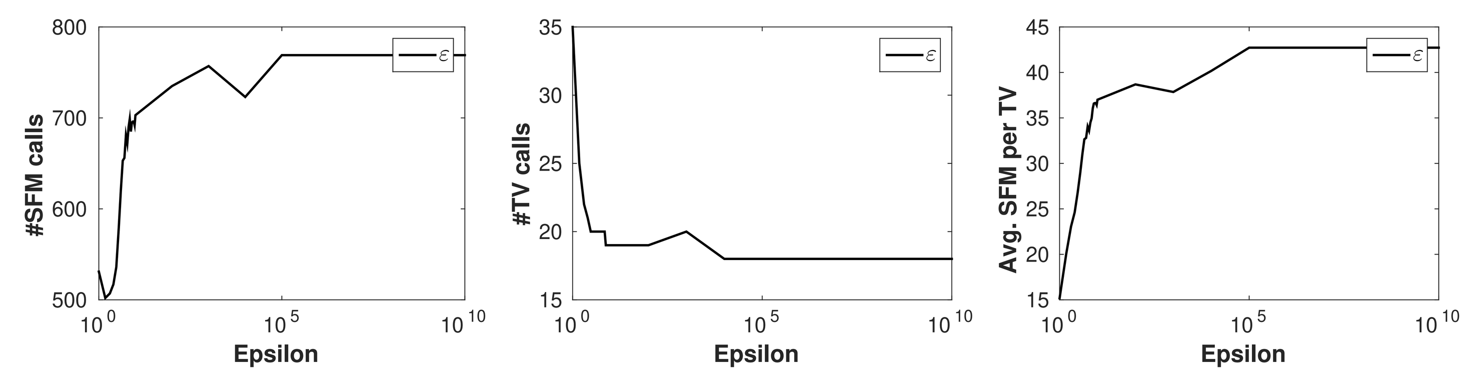
6 . Experiments - SFM of 2D and 3D cuts



Oracle calls for 2D = 1D + 1D for different ε values for both plain and accelerated BCD



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



- **Take home message.** Adding box constraints to the continuous optimization problem reduces individual SFM calls significantly.

7 . References

[1] S. Jegelka, F. Bach, and S. Sra, "Reflection methods for user-friendly submodular optimization," in *Advances in Neural Information Processing Systems*, 2013.

[2] K. S. Sesh Kumar and F. Bach, "Active-set methods for submodular minimization problems," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 4809–4839, 2017.

[3] A. Ene, H. Nguyen, and L. A. Végh, "Decomposable submodular function minimization: discrete and continuous," in *Advances in Neural Information Processing Systems*, 2017.