Northeastern University
College of Computer and Information Science

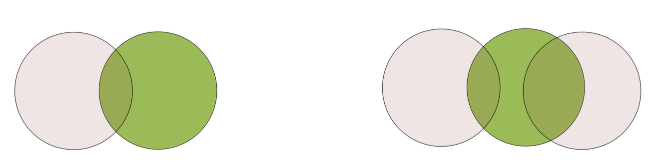
Decomposable Submodular Function Minimization
Discrete and Continuous

Alina Ene   Huy L. Nguyen   Laszlo Vegh

BOSTON UNIVERSITY   LSE The London School of Economics and Political Science

Submodular Set Functions

For all A ⊆ B and $x \notin B$



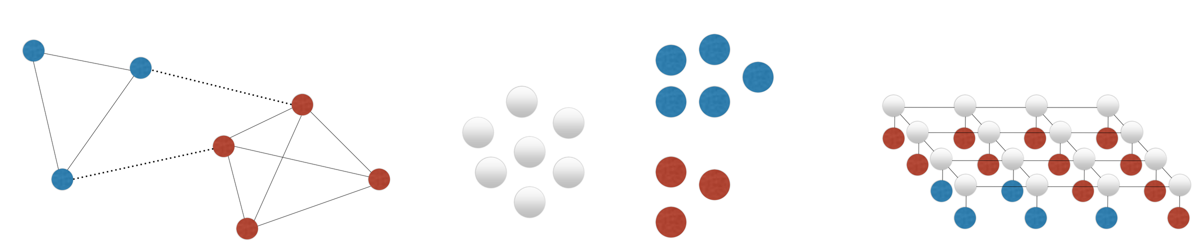
$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$$

For all A ⊆ B

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

Submodular Function Minimization

$$\min_{S \subseteq V} f(S)$$



Minimum Cut   Clustering   MAP Inference

Many algorithms: combinatorial, ellipsoid, cutting plane,…

$O(n^5 T + n^6)$	[Orlin '09]	$n = \lvert V \rvert$
$O((n^4 T + n^5)\log M)$	[Iwata '03]	$M = \max_{S \subseteq V} f(S)$

Decomposable Functions

$$\min_{S \subseteq V} \sum_{i=1}^{r} f_i(S)$$

Simple f_i: fast subroutine for minimizing $f_i(S) + w(S)$ for any modular function w

$O\left(\frac{r^2}{l}\log\left(\frac{M}{\varepsilon}\right)Q\right)$	Reflection [Nishihara, Jegelka, Jordan '14]
$O\left(\frac{r}{l}\log\left(\frac{M}{\varepsilon}\right)Q\right)$	Coordinate descent [Ene, Nguyen '15]
$O\left(\frac{r}{\sqrt{l}}\log\left(\frac{M}{\varepsilon}\right)Q\right)$	Accelerated coordinate descent [Ene, Nguyen '15]

[NJJ'14] shows restricted strong convexity param $l = \Omega\left(\frac{1}{n^2 r}\right)$

Our result: $\boldsymbol{l = \Omega\left(\frac{1}{n^2}\right)}$

Q: time for one "projection" of f_i

Random Coordinate Descent Algorithm

Base polytope

$$B(f) = \{y \mid \langle y, \mathbf{1}_S \rangle \leq f(S) \;\; \forall S \subseteq V, \; \langle y, \mathbf{1}_V \rangle = f(V)\}$$

Exact formulation

$$\min_{x \in [0,1]^n} \hat{f}(x) = \min_{x \in \{0,1\}^n} f(x)$$

Convex but non-smooth objective

Lovász extension

$$\hat{f}(x) = \max_{y \in B(f)} \langle y, x \rangle$$

Regularization for free

$$\min_{x \in \mathbb{R}^n} \hat{f}(x) + \frac{1}{2}\|x\|^2$$

Primal formulation

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{r} \left(\hat{f}_i(x) + \frac{1}{2r}\|x\|^2\right)$$

Dual formulation

$$\min_{y_i \in B(f_i)} g(y) := \frac{1}{2}\left\|\sum_{i=1}^{r} y_i\right\|^2$$

Smooth dual objective, minimize via random coordinate descent [Nesterov '12; Fercoq, Richtárik '13]

Random Coordinate Descent Algorithm

Start with $y_0 = (y_0^{(1)}, \ldots, y_0^{(r)})$, where $y_0^{(i)} \in B(f_i)$

In each iteration k ($k \geq 0$)

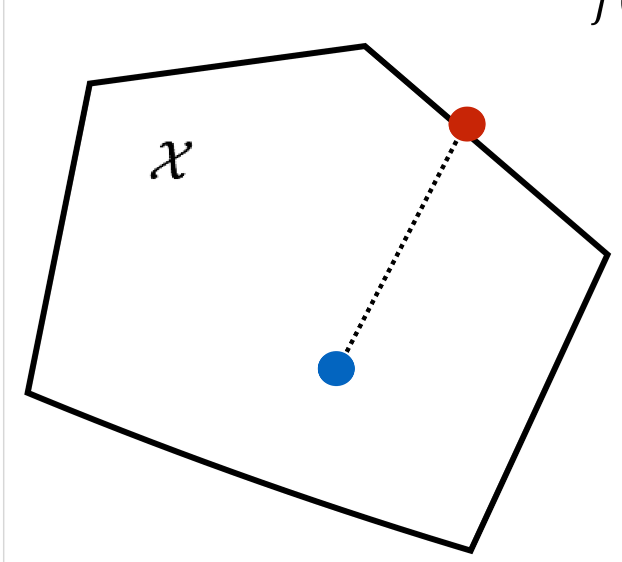
Pick an index $i_k \in \{1, 2, \ldots, r\}$ uniformly at random

⟪*Update block* i_k⟫

$$y_{k+1}^{(i_k)} \leftarrow \operatorname*{argmin}_{y \in B(f_{i_k})} \left(\left\langle \nabla_{i_k} g(y_k), y - y_k^{(i_k)} \right\rangle + \|y - y_k^{(i_k)}\|^2 \right)$$

Convergence Analysis

$$\min \{ f(x) : x \in \mathcal{X} \}$$



l-strong Convexity

$$f(■) - f(■) \geq \langle \nabla f(■), ■ - ■ \rangle + l \cdot d(■, ■)^2/2$$

for any ■, ■ in X

What we really need

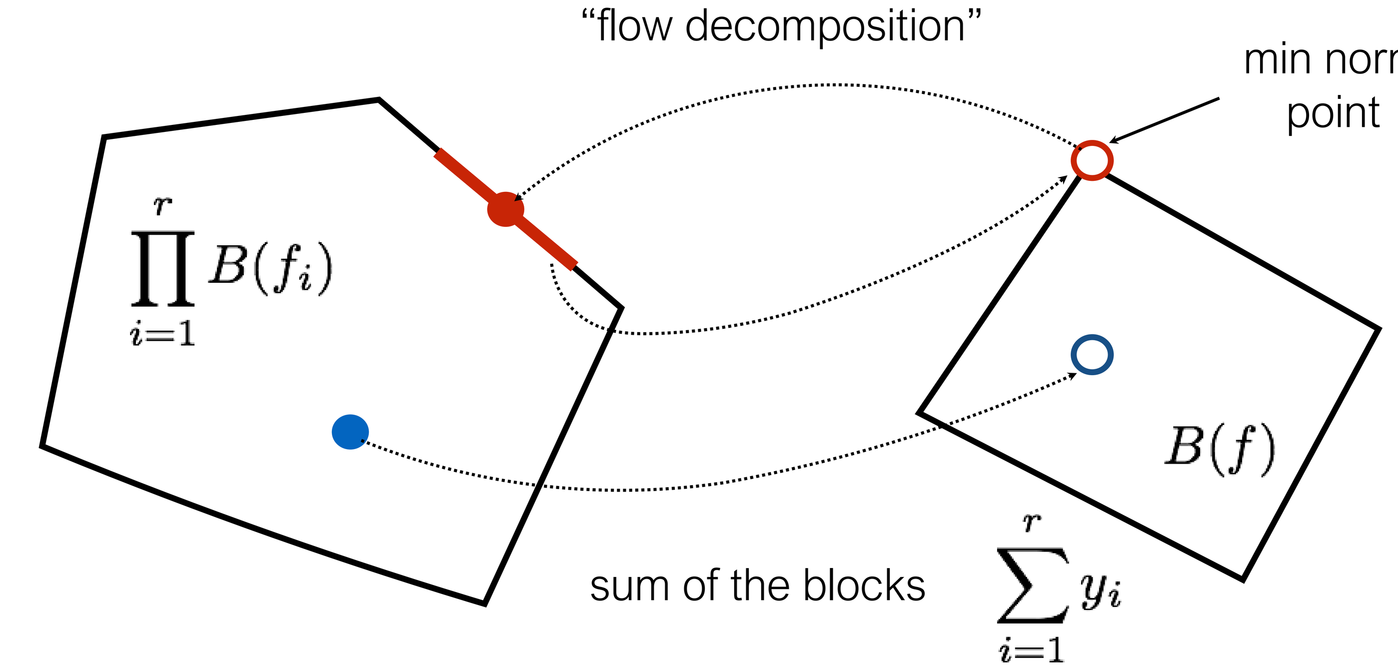
$$f(■) - f(■) \geq l \cdot d(■, ■)^2/2$$

for some optimal point ■

Restricted Strong Convexity Parameter

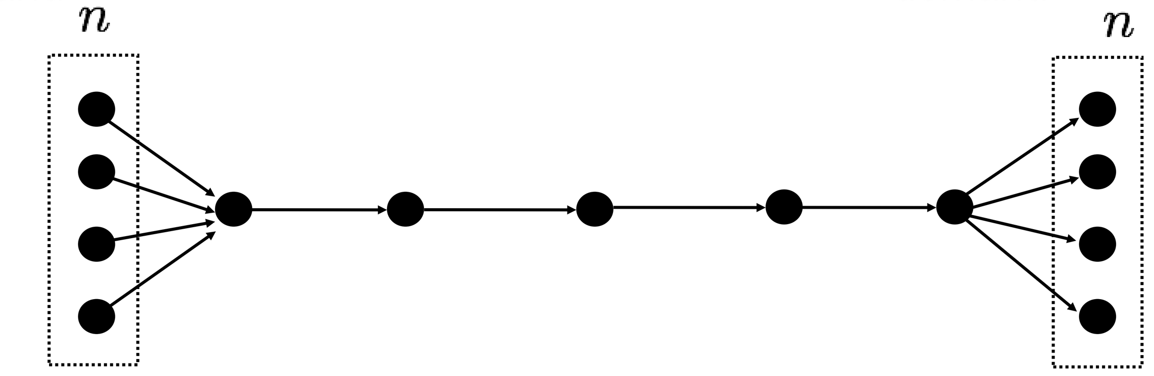
Our problem: $\min \left\{ g(y) = \left\|\sum_{i=1}^{r} y_i\right\|^2 : y \in \prod_{i=1}^{r} B(f_i) \right\}$

"flow decomposition"



g(●) - g(●) $\geq \frac{1}{2}$d(○, ○) ^{2} $\geq \frac{1}{n^2}\cdot$ d(●, ●) ^{2}

Tight example



$$\ell_2^2 \text{ of demand} = \frac{1}{n^2} \cdot n$$

$$\ell_2^2 \text{ of flow} = n$$

$$\ell_2^2 \text{ of demand} = \frac{1}{n^2} \cdot (\ell_2^2 \text{ of flow})$$

References

A. Ene, H. L. Nguyen. Random coordinate descent methods for minimizing decomposable submodular functions. ICML 2015.

R. Nishihara, S. Jegelka, M. I. Jordan. On the convergence rate of decomposable submodular function minimisation. NIPS 2014.

Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM Journal on Optimization, 2012.

O. Fercoq, P. Richtárik. Accelerated, parallel and proximal coordinate descent. arXiv, 2013.