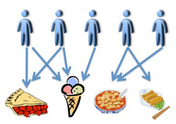
On the Convergence Rate of Decomposable Submodular Function Minimization

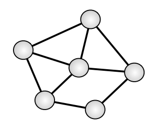
Robert Nishihara, Stefanie Jegelka, and Michael I. Jordan

UNIVERSITY OF CALIFORNIA Berkeley

Discrete Problems in Machine Learning


recommenda-tions

variable selection


graphical models

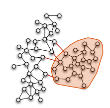

graph cuts


image segmentation

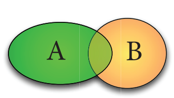

sensor placement

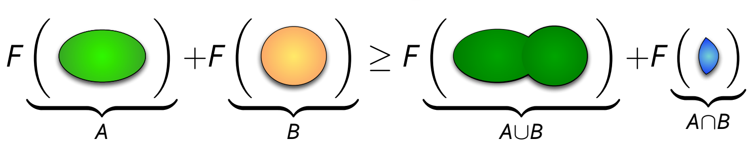
Submodularity

Let V be a finite set, and let $F\colon 2^V \to \mathbb{R}$ be a function on subsets of V. F is **submodular** if it satisfies

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

for all $A, B \subset V$.





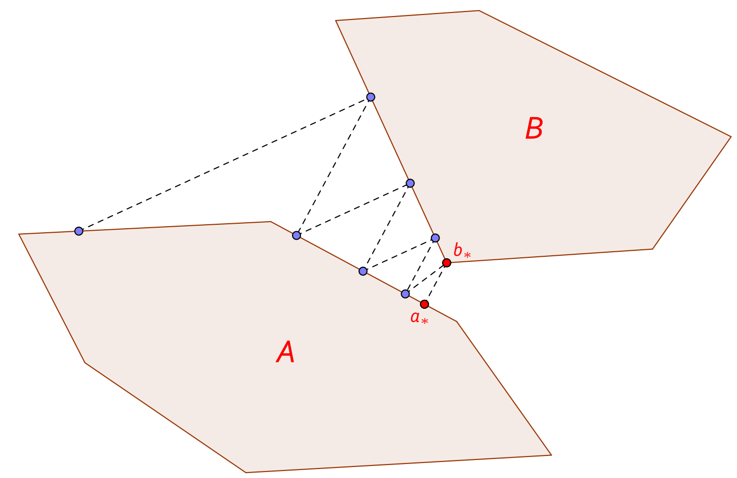
Structural Assumption

F is submodular and F decomposes as the sum of ``simple'' submodular functions $F = F_1 + \cdots + F_R$. Here, ``simple'' means that we have a fast subroutine for minimizing $F_r + s$, where $s(A) = \sum_{n \in A} s_n$ is modular.

The Minimization Problem

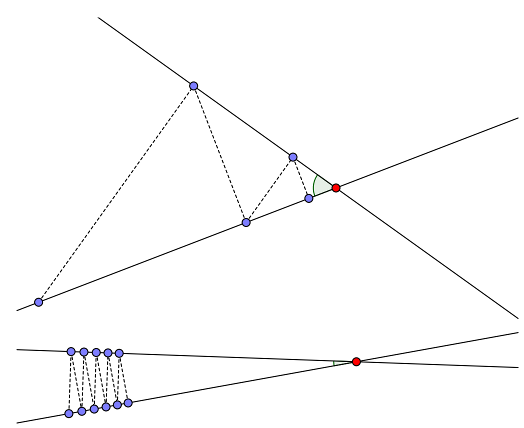
The most basic problem is $\min_{A \subset V} F(A)$. We consider the related problem $\min_{w \in \mathbb{R}^N} f(w) + \frac{1}{2}||w||^2$ and we approach it through its dual, which can be expressed as the best approximation problem $\min_{a \in \mathcal{A}, b \in \mathcal{B}} ||a - b||_2$, where $\mathcal{A} = \left\{(a_1, \ldots, a_R) \mid a_r \in \mathbb{R}^N, \sum_{r=1}^R a_r = 0\right\}$ and $\mathcal{B} = B(F_1) \times \cdots \times B(F_R)$. We solve this using alternating projections, which is well understood for subspaces but not for polyhedra.

The Alternating Projection Algorithm



We prove that this algorithm converges linearly, and we give upper and lower bounds on the worst-case rate.

Intuition and Angles



Large angles lead to fast convergence and small angles lead to slow convergence. When $E = \mathcal{A} \cap \mathcal{B}$ is nonempty, define $\kappa_* = \sup_{x \in (\mathcal{A} \cup \mathcal{B}) \setminus E} \frac{d(x,E)}{\max\{d(x,\mathcal{A}), d(x,\mathcal{B})\}}$.

Proposition (Convergence Rate)

If alternating projections generates the sequence $\{a_k\} \in \mathcal{A}$ and $\{b_k\} \in \mathcal{B}$ then $||a_k - a_*|| \leq 2||a_0 - a_*||(1 - \frac{1}{\kappa_*^2})^k$ and $||b_k - b_*|| \leq 2||b_0 - b_*||(1 - \frac{1}{\kappa_*^2})^k$.

Define the angle $c_F(U, V)$ between subspaces U and V as

$$c_F(U,V) = \sup \left\{ u^\top v \;\middle|\; \begin{array}{l} u \in U \cap (U \cap V)^\perp \\ v \in V \cap (U \cap V)^\perp \\ ||u|| \leq 1, ||v|| \leq 1 \end{array} \right\}.$$

Generalize this to affine spaces.

Approach and Results

Proposition (κ_* and c_F)

If P and Q are polyhedra, then there exist faces P_x, Q_y such that

$$1 - \frac{1}{\kappa_*^2} \leq c_F(\text{aff}(P_x), \text{aff}(Q_y)).$$

Proposition (c_F and singular values)

If S and T are matrices with orthonormal rows and equal numbers of columns then $c_F(\text{null}(S), \text{null}(T))$ equals the largest singular value of $ST^\top$ that is less than one.

The relevant matrices for faces of $\mathcal{A}$ and $\mathcal{B}$ are

$$S = \frac{1}{\sqrt{R}} \Big(\underbrace{I_N \ \cdots \ I_N}_{\text{repeated } R \text{ times}} \Big) \quad \text{and } T = \begin{pmatrix} \frac{1_{A_{11}}^\top}{\sqrt{|A_{11}|}} & & \\ \vdots & & \\ \frac{1_{A_{1M_1}}^\top}{\sqrt{|A_{1M_1}|}} & & \\ & \ddots & \\ & & \frac{1_{A_{R1}}^\top}{\sqrt{|A_{R1}|}} \\ & & \vdots \\ & & \frac{1_{A_{RM_R}}^\top}{\sqrt{|A_{RM_R}|}} \end{pmatrix}$$

where each $A_{r1}, \ldots, A_{rM_r}$ is a partition of the set V. We can bound the largest non-one singular value of $ST^\top$ by the largest non-one eigenvalue of $(ST^\top)^\top(ST^\top)$, we do this by showing that $(ST^\top)^\top(ST^\top) = I - \frac{R-1}{R}\mathcal{L}$ where $\mathcal{L}$ is a graph Laplacian and applying Cheeger's inequality. This gives a bound on the angles c_F, which in turn gives a bound on κ_*, which in turn bounds the convergence rate.

Theorem (Upper Bound)

The alternating projection algorithm between the sets $\mathcal{A}$ and $\mathcal{B}$ converges linearly with rate $1 - \frac{1}{N^2R^2}$.

Theorem (Lower Bound)

The worst-case convergence rate of alternating projections between $\mathcal{A}$ and $\mathcal{B}$ is bounded below by $1 - \frac{\pi}{2N^2R}$.

Here $N = |V|$ and R is the number of functions in the sum $F = F_1 + \cdots + F_R$.

{rkn, stefje, jordan}@eecs.berkeley.edu