Labs
Optimization for Machine Learning
Spring 2020

EPFL
School of Computer and Communication Sciences
Martin Jaggi & Nicolas Flammarion
github.com/epfml/OptML_course

Problem Set 8, due May 1, 2020 (Frank-Wolfe)

Convergence of Frank-Wolfe

Exercise 1:

Assuming $h_0 \leq 2C$, and the sequence $h_0, h_1, \ldots$ satisfies

$$h_{t+1} \leq (1-\gamma)h_t + \gamma^2 C \qquad t = 0, 1, \ldots$$

for $\gamma = \frac{2}{t+2}$, prove that

$$h_t \leq \frac{4C}{t+2} \qquad t = 0, 1, \ldots$$

Applications of Frank-Wolfe

Exercise 2:

Derive the LMO formulation for matrix completion, that is

$$\min_{Y \in X \subseteq \mathbb{R}^{n \times m}} \sum_{(i,j) \in \Omega} (Z_{ij} - Y_{ij})^2$$

when $\Omega \subseteq [n] \times [m]$ is the set of observed entries from a given matrix Z.

Where our optimization domain X is the unit ball of the trace norm (or nuclear norm), which is defined the convex hull of the rank-1 matrices

$$X := conv(\mathcal{A}) \ \text{ with } \ \mathcal{A} := \left\{ \mathbf{u}\mathbf{v}^\top \;\middle|\; \substack{\mathbf{u} \in \mathbb{R}^n,\ \|\mathbf{u}\|_2 = 1 \\ \mathbf{v} \in \mathbb{R}^m,\ \|\mathbf{v}\|_2 = 1} \right\} .$$

1. Derive the LMO for this set X for a gradient at iterate $Y \in \mathbb{R}^{n \times m}$.
2. Dervie the *projection* step onto X. How does the computational operations (or costs) needed to compute the LMO and the projection step compare?

Practical Implementation

Follow the Python notebook provided here:

github.com/epfml/OptML_course/tree/master/labs/ex08/