Project and Forget: Solving Large-Scale Metric Constrained Problems

Rishi Sonthalia[1]
Work with: Anna C. Gilbert

ICML 2023

[1]Twitter: @RishiSonthalia

Graph Metrics

One important case of the problem is when S is the space of metrics defined on a certain graph.

Definition 1

$M \subset \mathbb{R}^{\binom{n}{2}}$ be the space such that for all $1 \leq i < j < k \leq n$ we have that

$$M_{ij} + M_{jk} \geq M_{ik}$$

This space is known an the metric polytope.

Definition 2

Give a graph G, the metric polytope defined by G is the projection of M onto the coordinates that correspond to the edges in G. We shall denote this space $\mathrm{MET}(G)$.

What does MET(G) look like

Broken Cycles

In general given a cycle $C = v_1 \ldots, v_k$ we say this cycle is *broken* if

$$w(v_1, v_k) - \sum_{i=1}^{k-1} w(v_i, v_{i+1}) > 0$$

We shall the edge with the large weight as the heavy edge and the other edges as the light edge. We say call the difference (i.e. the quantity on the left) the deficit of a cycle.

Remark

$w \in \text{MET}(G)$ if and only if $G = (V, E, w)$ has no broken cycles and $w(e) \geq 0$ for all $e \in E$

Graph Metric Optimization

Metric Constrained Problems

We can formulate the above problem as the following constrained optimization problem.

$$\begin{aligned} \text{minimize} \quad & f(D) \\ \text{subject to} \quad & D \in \text{MET}(G) \end{aligned}$$

where f is a strictly convex function and G is a given graph.

Specific Problem Formulations

- **Metric nearness**
  - Given D, $n \times n$ matrix of distances, find closest metric

$$\hat{M} = \arg\min \|D - M\|_p \quad \text{s.t.} \quad M \in \text{MET}_n.$$

- **(Weighted) Correlation clustering**
  - Given graph G and (dis)similarity measures on each edge $e = (i, j)$, $w^+(e)$ and $w^-(e)$, partition nodes into clusters by minimizing

$$\min \sum_{e \in E} w^+(e) x_e + w^-(e)(1 - x_e) \quad \text{s.t. } x \in \text{MET}(G).$$

Optimization Techniques: Existing Methods

- Constrained optimization problems with **many** constraints: $O(n^3)$ for simple triangle inequality constraints, possibly exponentially many for graph cycle constraints.

- Existing methods don't scale.
  - Run out of memory, or
  - take too long to converge.

Project and Forget

We propose a new iterative algorithm Project and Forget

Steps

1. Find a well chosen **subset** of violated constrained using an oracle to add constraints to the problem.
2. Iteratively project onto to our current chosen constraints.
3. Forget some constraints.

Theoretical results: Summary

Theorem 1 (Sonthalia and Gilbert)

If $f \in \mathcal{B}(S)$, H_i *are strongly zone consistent with respect to* f, *and* $\exists\, x^0 \in S$ *such that* $\nabla f(x^0) = 0$, *then*

1. *Then any sequence* x^n *produced by Project and Forget converges to the optimal solution of problem.*
2. *If* x^* *is the optimal solution,* f *is twice differentiable at* x^*, *and the Hessian* $H := Hf(x^*)$ *is positive definite, then there exists* $\rho \in (0, 1)$ *such that*

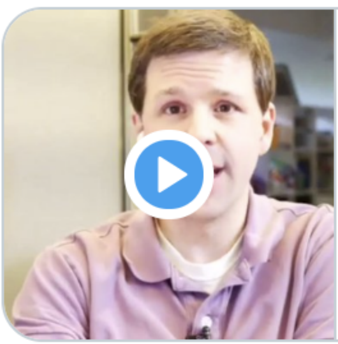
$$\lim_{\nu \to \infty} \frac{\|x^* - x^{\nu+1}\|_H}{\|x^* - x^\nu\|_H} \leq \rho \tag{0.1}$$

where $\|y\|_H^2 = y^T H y$.

David Gleich
@dgleich

Replying to @dgleich

Or just skip to the video with @RishiSonthalia @n_veldt on doing a deep dive into Rishi Sonthalia and Anna Gilbert paper with them.



Project and Forget Metric Optimization - Explained
Rishi Sonthalia and Anna Gilbert proposed a faster way to solve metric constrained optimization problems in a recen...
youtube.com

9:44 AM · Nov 11, 2020 · Twitter Web App

Project and Forget: Results - Metric Nearness

Here we have $f(x) = 0.5\|x\|^2$ and $S = \text{MET}(K_n)$.

Metric Nearness: Time

Algorithm	Number of Nodes									
	100	200	300	400	500	600	700	800	900	1000
PF	13.5	32.7	85.1	170	271	458	720	983	1356	1649
Cyclic Bregman	1.77	10.5	47.1	141	322	558	910	1472	2251	3167
Mosek	11.7	542	Out of Memory							
SCS	1632	19466	Timed Out							
OSQP	64.5	3383	Timed Out							
ProxSDP	353	684	Timed Out							
Ipopt	2792	Timed Out								
ECOS	597	Timed Out								
CPLEX	Out of Memory									
SLSQP	Timed Out									
COSMO	Timed Out									

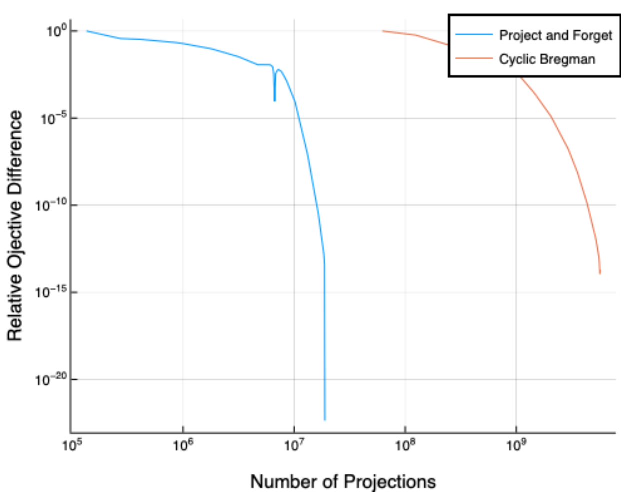
Weighted Correlation Clustering

Graph	n	# Constraints	Time	# Active Constraints	Iters.
Slashdot	82140	5.54×10^{14}	46.7 hours	384227	145
Epinions	131,828	2.29×10^{15}	121.2 hours	579926	193

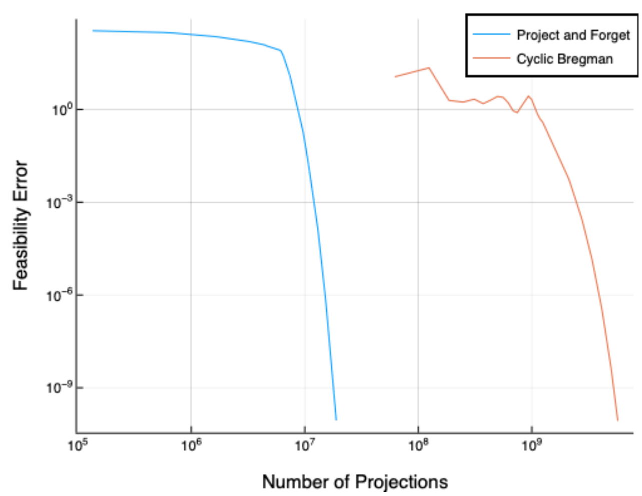
Table 1: Time taken and quality of solution returned by Project and Forget when solving the weighted correlation clustering problem for sparse graphs. The table also displays the number of constraints the traditional LP formulation would have.

Metric Nearness

Forget step is the crucial step!



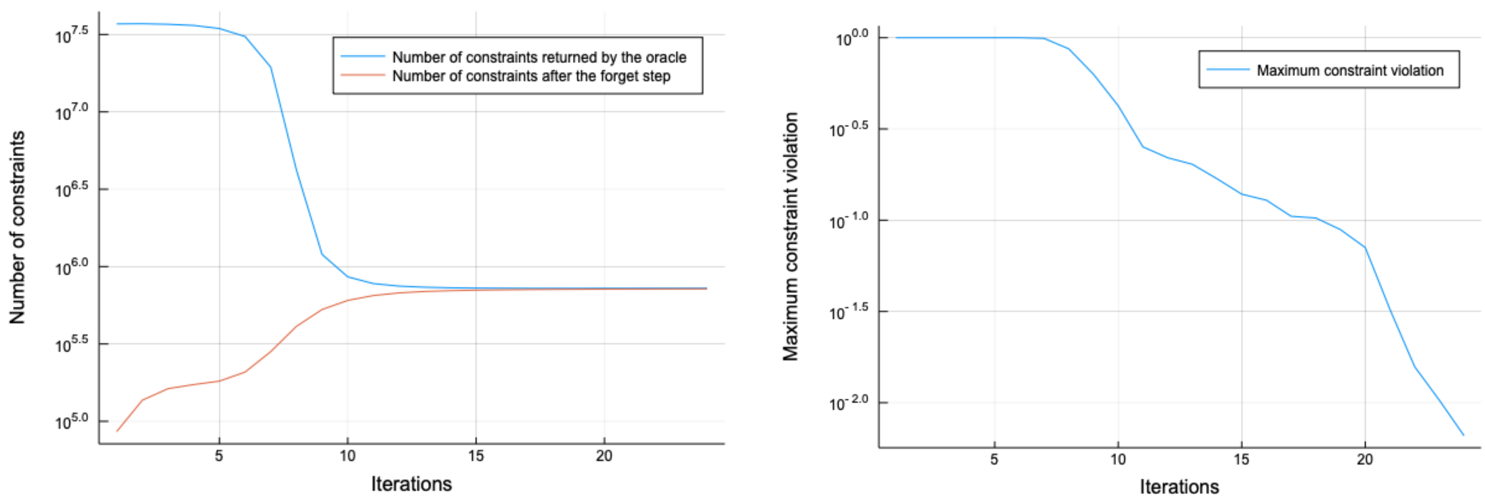
(a) Relative Objective Error



(b) Feasibility Error

Weighted Correlation Clustering

Forget step is still crucial!



(c) Number of constraints. (d) Max Violation.

Figure 1: Plots showing the number of constraints returned by the oracle, the number of constraints after the forget step, and the maximum violation of a metric constraint when solving correlation clustering on the Ca-HepTh graph

Non-Metric Constrained Problem - Optimal Transport

Primal

$$\begin{aligned} &OT(\boldsymbol{a}, \boldsymbol{b}) = \min \langle \boldsymbol{C}, \boldsymbol{P} \rangle \\ &\text{Subject to:} \quad \boldsymbol{a} = \boldsymbol{P}\mathbf{1}_m,\ \boldsymbol{b} = \boldsymbol{P}^T\mathbf{1}_n,\ \boldsymbol{P}_{ij} \geq 0 \end{aligned} \tag{0.2}$$

Quadratically Regularized

$$\begin{aligned} &\text{Minimize:} \quad \langle \boldsymbol{C}, \boldsymbol{P} \rangle + \gamma \|\boldsymbol{a} - \boldsymbol{P}\mathbf{1}_m\|^2 + \gamma \|\boldsymbol{b} - \boldsymbol{P}^T\mathbf{1}_n\|^2 \\ &\text{Subject to:} \quad \forall i \in [n], \forall j \in [m], \quad \boldsymbol{P}_{ij} \geq 0 \end{aligned} \tag{0.3}$$

Dual

$$\begin{aligned} &\text{Min:} \quad \frac{1}{\gamma}\|\boldsymbol{f}\|^2 + \frac{1}{\gamma}\|\boldsymbol{g}\|^2 - \langle \boldsymbol{f}, \boldsymbol{a} \rangle - \langle \boldsymbol{g}, \boldsymbol{b} \rangle \\ &\text{Subject to:} \quad \boldsymbol{f}_i + \boldsymbol{g}_j \leq \boldsymbol{C}_{ij} \end{aligned} \tag{0.4}$$

Results

Algorithm	501	1001	5001	10001	20001
PF	12	151	1972	5909	21665
LBFGSB	24	162	4080	Out of memory.	
Mosek dual	56	328	1927	Out of memory.	
Mosek primal	5	Out of memory.			
CPLEX primal	105	Out of memory.			
CPLEX dual	Out of memory.				
PGD	Did not converge.				

Table 2: Time taken in seconds to solve the quadratic regularized optimal transport problem. All experiments were run on a machine with 52 GB of RAM.

Extensions and applications

- The paper presents an extension of Project and Forget to general convex constraints, not simply metric or linear constraints.
- Applications include: sparse optimal transport (dual regularized), ℓ_2 SVMs, information theoretic metric learning.
- Can also be used to project onto the cone of submodular functions and for solving problems with graphical probability models. See Sonthalia, Seigal, Montufar 2023.