QUADRATICALLY REGULARIZED OPTIMAL TRANSPORT ON GRAPHS
FAST ITERATIVE SOLUTION OF THE OPTIMAL TRANSPORT PROBLEM ON GRAPHS

Society for Industrial and Applied Mathematics(SIAM)
2018 & 2021

Montacer Essid and Justin Solomon
Enrico Facca and Michele Benzi
Reporter: Fengjiao Gong

March 10, 2022

Outline

Optimal transport problem (OTP)
Monge–Kantorovich problem

Consider a measurable space $\mathcal{X}$

- f^+ and f^-: nonnegative measures with equal mass
- cost function $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ where $c(x, y)$: cost paid for transporting one unit of mass from x to y
- $\mathcal{M}_+(\mathcal{X} \times \mathcal{X})$: space of nonnegative measures on space $\mathcal{X} \times \mathcal{X}$
- γ^*: the optimal plan:

$$\inf_{\gamma \in \mathcal{M}_+(\mathcal{X}\times\mathcal{X})} \left\{ \int_{\mathcal{X}\times\mathcal{X}} c(x,y) \mathrm{d}\gamma(x,y) : \begin{array}{c} \text{for all } A, B \text{ measurable subsets of } \mathcal{X} \\ \gamma(A, \mathcal{X}) = f^+(A), \quad \gamma(\mathcal{X}, B) = f^-(B) \end{array} \right\} \tag{1}$$

Optimal transport problem (OTP)

Space $\mathcal{X}$ is a connected, weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

- n vertices and m edges, $m = \mathcal{O}(n)$ if sparse
- weights $\boldsymbol{w} \in \mathbb{R}^m$: strictly positive
- cost function: *the shortest weighted path metric*

$$c\left(v_i, v_j\right) = \mathrm{dist}_{\mathcal{G},\boldsymbol{w}}\left(v_i, v_j\right) = \inf_{\mathcal{P}\left(v_i,v_j\right)} \left\{ \sum_{e \in \mathcal{P}\left(v_i,v_j\right)} w_e : \mathcal{P}\left(v_i, v_j\right) = \left(\begin{array}{c} \text{edge-path} \\ \text{from } v_i \text{ to } v_j \end{array} \right) \right\} \tag{2}$$

- f^+ and f^-: vector $\boldsymbol{b}^+, \boldsymbol{b}^- \in \mathbb{R}^n_+ = [0, +\infty[^n$

$$\sum_{i=1}^{n} b_i^+ = \sum_{i=1}^{n} b_i^- \tag{3}$$

Optimal transport problem (OTP)
Linear-programming problem

Linear-programming problem:

$$\inf_{\gamma \in \mathbb{R}_{+}^{n,n}} \left\{ \operatorname{tr}\left(\boldsymbol{C}^{T}\gamma\right) : \begin{array}{rl} \boldsymbol{\gamma 1} & = \boldsymbol{b}^{+} \\ \boldsymbol{1}\boldsymbol{\gamma}^{T} & = \boldsymbol{b}^{-} \end{array} \right\}, \tag{4}$$

where C is the cost matrix.

Minimal-cost Flow Problem(LP)

Equivalent - ℓ^1-minimal cost flow problem:

$$\min_{\boldsymbol{q}\in\mathbb{R}^m} \left\{ \sum_{e=1}^{m} w_e \left|q_e\right| \; s.t. \; \boldsymbol{E}^T\boldsymbol{q} = \boldsymbol{b}^+ - \boldsymbol{b}^- \right\} \tag{5}$$

where $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ is the signed incidence matrix

$$E_{e,i} = \begin{cases} 1 & \text{if } i = \text{ “head” of edge } e, \\ -1 & \text{if } i = \text{ “tail” of edge } e \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Minimal-cost Flow Problem(LP)

Denote

- forcing term: $\boldsymbol{b} = \boldsymbol{b}^+ - \boldsymbol{b}^-$
- diag matrix $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$
- matrix $\boldsymbol{G} = \boldsymbol{W}^{-1}\boldsymbol{E}$

Equivalent form

$$\min_{\boldsymbol{q}\in\mathbb{R}^m} \left\{ \|\boldsymbol{W}\boldsymbol{q}\|_1 \text{ s.t. } \boldsymbol{E}^T\boldsymbol{q} = \boldsymbol{b} \right\} \tag{7}$$

Minimal-cost Flow Problem(LP)
Dual Problem

Dual Theory

- Slater condition
- Lagrangian function
- Strong Duality
- Complementary slackness
  Karush-Kuhn-Tucker (KKT) conditions

Dual Problem

Slater condition

- For a normal problem

$$\begin{cases} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{cases}$$

  it satisfies Slater's condition if it is strictly feasible, that is

$$\exists\, x_0 \in \mathcal{D} : \quad f_i(x_0) < 0 \quad i = 1, \ldots, m, \quad h_i(x_0) = 0, i = 1, \ldots, p.$$

- For a convex problem

$$\begin{cases} f_0, f_1, \ldots, f_m & \to & \textit{convex} \\ h_1, \ldots, h_p & \to & \textit{affine} \end{cases}$$

  it satisfies a weak form of Slater's condition if function f_i is affine.
- Denote

$$p^\star = \min_x f_0(x)$$

Lagrangian function

Define Lagrangian $\mathcal{L} : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$

$$\mathcal{L}(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Define function $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$

$$g(\lambda, \nu) := \min_x \mathcal{L}(x, \lambda, \nu)$$

Dual problem: Suppose $\lambda_i > 0, i = 1, \ldots, m$, then $\lim_{\|\lambda\| \to +\infty} g = -\infty$. Since g is concave, so it would have a maximum denoted as

$$d^* = \max_{\lambda \geq 0, \nu} g(\lambda, \nu), \; \lambda \geq 0$$

Strong Duality

- **Theorem:** If the primal problem is convex, and satisfies the weak Slater's condition, then strong duality holds, that is $p^\star = d^\star$.
- **Connection:** Existence of an optimal solution to any one of these two problems guarantees an optimal solution to the other, with that their extreme values are equal[1].
- **Duality Gap**

$$p^\star - d^\star$$

which is non-negative.

Complementary slackness conditions

Assume that strong duality holds, both primal and dual problems are attained by x^* and (λ^*, ν^*) respectively. Then

$$f_0(x^*) = g(\lambda^*, \nu^*) \leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \leq f_0(x^*)$$

Since every term in that sum is non-positive, each term is zero:

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m \quad (KKT)$$

If the problem is convex, and satisfies Slater's condition, then a primal point is optimal if and only if there exist (λ, ν) such that the *KKT* conditions are satisfied.

Minimal-cost Flow Problem(LP)
Dual Problem

Primal problem

$$\min_{\boldsymbol{q} \in \mathbb{R}^m} \left\{ \|\boldsymbol{W}\boldsymbol{q}\|_1 \text{ s.t. } \boldsymbol{E}^T \boldsymbol{q} = \boldsymbol{b} \right\} \tag{8}$$

Dual problem[2]

$$\max_{\boldsymbol{u} \in \mathbb{R}^n} \left\{ \boldsymbol{b}^T \boldsymbol{u} \text{ s.t. } \|\boldsymbol{G}\boldsymbol{u}\|_\infty \leq 1 \right\} \tag{9}$$

Minimal-cost Flow Problem(LP)
Dual Problem

Dual problem

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ \boldsymbol{b}^T \boldsymbol{u} \text{ s.t. } \|\boldsymbol{G}\boldsymbol{u}\|_\infty \leq 1 \right\}$$

Rewrite

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ \boldsymbol{b}^T \boldsymbol{u} \; s.t. \; |(\boldsymbol{G}\boldsymbol{u})_e| \leq 1, \forall e = 1, \ldots, \right\} \tag{10}$$

$$\equiv \quad \max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ \boldsymbol{b}^T \boldsymbol{u} \; s.t. \; \tfrac{w_e}{2} \left(|(\boldsymbol{G}\boldsymbol{u})_e|^2 - 1 \right) \leq 0, \forall e = 1, \ldots, m \right\} \tag{11}$$

Minimal-cost Flow Problem(LP)
Dual Problem

Lagrange functional, here $\boldsymbol{\mu} \in \mathbb{R}_+^m$ is nonnegative

$$\inf_{\boldsymbol{\mu}\in\mathbb{R}_+^m} \sup_{\boldsymbol{u}\in\mathbb{R}^n} \Phi(\boldsymbol{u}, \boldsymbol{\mu}) := \boldsymbol{b}^T\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{L}[\mu]\boldsymbol{u} + \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{w} \tag{12}$$

Denote

$$\mathcal{E}(\boldsymbol{\mu}) = \sup_{\boldsymbol{u}\in\mathbb{R}^n} \left(\boldsymbol{b}^T\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{L}[\mu]\boldsymbol{u}\right) = \frac{1}{2}\boldsymbol{b}^T\boldsymbol{\mathcal{U}}[\mu] = \frac{1}{2}\boldsymbol{b}^T\boldsymbol{L}^\dagger[\mu]\boldsymbol{b} \tag{13}$$

Problem only in terms of $\boldsymbol{\mu}$

$$\min_{\boldsymbol{\mu} \in \mathbb{R}^m_+} \left\{ \mathcal{L}(\boldsymbol{\mu}) = \mathcal{E}(\boldsymbol{\mu}) + \frac{1}{2}\boldsymbol{w}^T\boldsymbol{\mu} \right\} \tag{14}$$

Here it could be seen as a minimization problem for an energy functional $\mathcal{L}(\boldsymbol{\mu})$, while μ can be interpreted as a conductivity associated to the edges of the graph[3].

Proposition 1. Problems (7), (9), and (14) are equivalent. This means that the following equalities hold:

$$\max_{\boldsymbol{u}\in\mathbb{R}^n}\left\{\boldsymbol{b}^T\boldsymbol{u} : \|\boldsymbol{G}\boldsymbol{u}\|_\infty \leq 1\right\} = \min_{\boldsymbol{q}\in\mathbb{R}^m}\left\{\|\boldsymbol{W}\boldsymbol{q}\|_1 : \boldsymbol{E}\boldsymbol{q} = \boldsymbol{b}\right\} = \min_{\boldsymbol{\mu}\in\mathbb{R}^m\geq 0}\mathcal{L}(\boldsymbol{\mu})$$

Moreover, given a solution $\boldsymbol{\mu}^*$ that minimizes the functional $\mathcal{L}$ we can recover a maximizer $\boldsymbol{u}^*$ of problem (2.2) by solving

$$\boldsymbol{L}\left[\mu^*\right]\boldsymbol{u}^* = \boldsymbol{b} \tag{15a}$$

$$\left|(\boldsymbol{G}\boldsymbol{u}^*)_e\right| \leq 1 \quad \forall e = 1,\ldots,m \tag{15b}$$

$$\left|(\boldsymbol{G}\boldsymbol{u}^*)_e\right| = 1 \quad \mu_e^* > 0 \tag{15c}$$

(the KKT equations for problem (9)). A minimizer $\boldsymbol{q}^*$ of problem (7) can be represented in the form

$$\boldsymbol{q}^* = \operatorname{diag}\left(\boldsymbol{\mu}^*\right)\boldsymbol{G}\boldsymbol{u}^* \tag{16}$$

Moreover, the following equality holds:

$$\boldsymbol{\mu}^* = |\boldsymbol{q}^*| \tag{17}$$

Minimal-cost Flow Problem(LP)
Solution

- Define $\Psi : \mathbb{R}^m \to \mathbb{R}^m_+$

$$\boldsymbol{\mu} = \Psi(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^2/4 \tag{18}$$

- Define

$$\tilde{\mathcal{L}}(\boldsymbol{\sigma}) = \mathcal{L}(\Psi(\boldsymbol{\sigma})) \tag{19}$$

- Nicer convergence behavior

$$\min \tilde{\mathcal{L}} = \mathcal{L} \circ \Psi \Leftrightarrow \min \mathcal{L}$$

- Minimize $\tilde{\mathcal{L}}(\boldsymbol{\sigma})$ via **gradient descent** approach

$$\boldsymbol{\sigma}^* = \lim_{t \to +\infty} \boldsymbol{\sigma}(t) \tag{20}$$

$$\partial_t \boldsymbol{\sigma}(t) = -\nabla \tilde{\mathcal{L}}(\boldsymbol{\sigma}(t)), \quad \boldsymbol{\sigma}(0) = \boldsymbol{\sigma}_0 \tag{21}$$

Discretization approches(backward Euler time stepping)

Minimal-cost Flow Problem(LP)

Equivalent - ℓ^1-minimal cost flow problem:

$$\min_{\boldsymbol{q}\in\mathbb{R}^m} \left\{ \sum_{e=1}^{m} w_e \left| q_e \right| \; s.t. \; \boldsymbol{E}^T \boldsymbol{q} = \boldsymbol{b}^+ - \boldsymbol{b}^- \right\} \tag{22}$$

where $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ is the signed incidence matrix

$$E_{e,i} = \begin{cases} 1 & \text{if } i = \text{ "head" of edge } e, \\ -1 & \text{if } i = \text{ "tail" of edge } e \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Quadratically Regularized LP(QP)

$$\mathcal{W}_1(\rho_0, \rho_1) := \begin{cases} \min_{J \in \mathbb{R}^{|E|}} \sum_{e \in E} c_e J_e \\ \text{s.t. } J \geq 0, \\ \qquad D^\top J = \rho_1 - \rho_0 \end{cases} \tag{LP}$$

- J one directed flow per edge : $\rho_0 \to \rho_1$[4]

Beckmann's minimal flow problem

Problem 4.5. Consider the minimization problem

$$\text{(BP)} \quad \min \left\{ \int |\mathbf{w}(x)| \, \mathrm{d}x \ : \ \mathbf{w} : \Omega \to \mathbb{R}^d, \ \nabla \cdot \mathbf{w} = \mu - \nu \right\}, \tag{4.4}$$

where the divergence condition is to be read in the weak sense, with no-flux boundary conditions, i.e., $-\int \nabla\phi \cdot \mathrm{d}\mathbf{w} = \int \phi \, \mathrm{d}(\mu - \nu)$ for any $\phi \in C^1(\overline{\Omega})$.

Quadratically Regularized LP(QP)

convex but not strictly convex

$\Downarrow$

output is **unpredictable/non-unique**: J.

$\Downarrow$

adding stronger convexity: **Quadratically** regularizer

$$\mathcal{W}_{1,\alpha}\left(\rho_0, \rho_1\right) := \begin{cases} \min\limits_{J \in \mathbb{R}^{|E|}} \sum\limits_{e \in E} c_e J_e + \dfrac{\alpha}{2} \sum\limits_{e} \boldsymbol{J_e^2} \\ \text{s.t. } J \geq 0, \\ \qquad D^\top J = \rho_1 - \rho_0. \end{cases} \tag{QP}$$

strong duality by the affine Slater condition[5]

Quadratically Regularized LP(QP)
Dual Problem

PROPOSITION 1 (duality). Quadratically regularized transport on graphs (LP) can be computed as follows:

$$\mathcal{W}_{1,\alpha}(\rho_0, \rho_1) = \frac{1}{\alpha} \sup_{\boldsymbol{p} \in \mathbb{R}^{|V|}} \left[\alpha \boldsymbol{f}^\top \boldsymbol{p} - \frac{1}{2} \left| (\boldsymbol{D}\boldsymbol{p} - \boldsymbol{c})_+ \right|_2^2 \right] \tag{DQP}$$

where $\boldsymbol{f} := \rho_1 - \rho_0$ and $\boldsymbol{c} \in \mathbb{R}^{|E|} \times 1$ is the vector of costs per edge. Furthermore, the primal variable J_e on edge $e \in E$ is zero whenever $(\boldsymbol{D}\boldsymbol{p} - \boldsymbol{c})_e \leq 0$.

Notations

Denote

- v_+: positive part of vector $v \in \mathbb{R}^n$

$$(v_+)_i := \max\{v_i, 0\}$$

- $|v|$: Euclidean norm of v

Quadratically Regularized LP(QP)
Dual Problem

Proof:

$$\begin{aligned}\mathcal{L}(J,p) &= \min_{J\in\mathbb{R}_+^{|E|}} \max_{p\in\mathbb{R}^{|V|}} \left[c^\top J + \frac{\alpha}{2}J^\top J + \left(f - \left(D^\top J\right)\right)^\top p\right] \\ &= \max_{p\in\mathbb{R}^{|V|}} \left[f^\top p + \min_{J\in\mathbb{R}_+^{|E|}} \left(J^\top (c - Dp) + \frac{\alpha}{2}J^\top J\right)\right]\end{aligned} \tag{24}$$

Here switched max and min by strong duality, then inner minimum can be explicitly computed since it is a quadratic function w.r.t J

$$J = \frac{(Dp - c)_+}{\alpha} \tag{25}$$

Substituting into (24):

$$\mathcal{W}_{1,\alpha}\left(\rho_0, \rho_1\right) = \sup_{p\in\mathbb{R}^{|V|}} \left[f^\top p - \frac{1}{2\alpha}\left|(Dp - c)_+\right|_2^2\right] \tag{26}$$

Quadratically Regularized LP(QP)
Active Set

DEFINITION 2 **(active set).** The active set of edges in E associated with a dual variable p is the set $S(p) := \{e \in E : (Dp - c)_e > 0\}$. In a minor abuse of notation, we will denote the active set of edges in E associated with a primal variable J as the set $S(J) := \{e \in E : J_e > 0\}$.

- Denote J^α: solution of the primal problem (QP)
- Denote p^α: solution of the dual problem (DQP)
- Denote $S(\alpha) := S(J^\alpha) = S(p^\alpha)$ the active set of edges with regularizing coefficient α

Quadratically Regularized LP(QP)
Proposition

Define $M(p)$ a diagonal matrix:

$$M(p)_{ee} := \begin{cases} 0 & \text{if } e \notin S(p) \\ 1 & \text{if } e \in S(p) \end{cases} \tag{27}$$

rewrite (DQP)

$$\sup_{p \in \mathbb{R}^{|V|}} \underbrace{\left[\alpha f^\top p - \frac{1}{2}(Dp - c)^\top M(p)(Dp - c) \right]}_{g(p)} \tag{28}$$

Differentiating

$$\nabla g(p) = \alpha f - D^\top M(p)(Dp - c) \tag{29}$$

$$\mathrm{Hess}[g](p) = -D^\top M(p) D \tag{30}$$

Quadratically Regularized LP(QP)
Proposition

In spectral graph theory [6], the unweighted Laplacian matrix is $L := D^\top D$.

Proposition 3**.** The Hessian of the dual problem (DQP) is the unweighted Laplacian $L(p)$ of the active subgraph $G(p) := (V, S(p))$.

Notations

Nonregularized problem

- Denote J_0: a solution of (LP) with active set $S\left(J_0\right)$ (nonunique)
- Define the union of all active sets:

$$S_0 = \bigcup_{J_0 \text{ solution of (LP)}} S\left(J_0\right) \tag{31}$$

Quadratically regularized problem

- Denote J^α: the solution of (QP) with active set $S(\alpha)$
- Denote $|J^\alpha|$: Euclidean norm of vector J^α

Quadratically Regularized LP(QP)
Proposition

Rewrite (QP)

$$V_\alpha(J) = \boldsymbol{c}^\top J + \frac{\alpha}{2}|J|^2 \tag{32}$$

PROPOSITION 4. Let $0 < \alpha < \alpha'$, and J^α be the solution of (QP) . If $J^\alpha \neq J^{\alpha'}$, then $\boldsymbol{c}^\top J^\alpha < \boldsymbol{c}^\top J^{\alpha'}$ and $|J^\alpha|^2 > |J^{\alpha'}|^2$

Here α tunes the influence of the two terms

Quadratically Regularized LP(QP)
Proposition

PROPOSITION 5 (sparsity). There exists a constant $\tilde{\alpha} > 0$ depending on the graph G and data f such that for all $\alpha \in (0, \tilde{\alpha})$, the solution J^α of (QP) is also a solution of (LP).

Behavior of J_α for small $\alpha > 0$

Quadratically Regularized Dual Problem
Optimization Algorithm

PROPOSITION 1 **(duality).** Quadratically regularized transport on graphs (LP) can be computed as follows:

$$\mathcal{W}_{1,\alpha}(\rho_0, \rho_1) = \frac{1}{\alpha} \sup_{p \in \mathbb{R}^{|V|}} \left[\alpha f^\top p - \frac{1}{2} |(Dp - c)_+|_2^2 \right] \tag{DQP}$$

Iterative Algorithm:

$$\begin{aligned} s_k &\leftarrow \text{ SEARCH-DIRECTION } (p_k) \\ t_k &\leftarrow \text{LINE-SEARCH} (p_k, s_k) \\ p_{k+1} &\leftarrow p_k + t_k s_k \end{aligned}$$

Given current p_k and search direction s_k, we want to increase the objective

Optimization Algorithm
Search direction

The objective is quadratic but not strictly convex; in particular, there are directions along which it may be flat. Here adopts an alternating strategy:

$$s_k \leftarrow \begin{cases} \alpha f - D^\top M_k v_k & \text{if } k \text{ is odd} \\ L_k^+ \left(\alpha f - D^\top M_k v_k\right) & \text{if } k \text{ is even} \end{cases} \quad \begin{array}{|l} \text{gradient direction} \\ \text{pseudo-Newton direction} \end{array} \tag{33}$$

"pseudo-Newton" to refer to the fact that Hessian L_k is not invertible.

Quadratically Regularized Dual Problem
Optimization Algorithm

Every iteration of our algorithm is far more *expensive algorithmically*: the Laplacian L_k changes anytime the active set changes, preventing the use of a fixed factorization to apply L_k^+.

The algorithm could take upwards of $O\left(|V|^3\right)$ time to apply L_k^+, whereas individual iterations of gradient descent take $O(|V|)$ time.

Experiments verify that the choice of search directions is useful in practice for *reducing the number of iterations*

Quadratically Regularized Dual Problem
Experiments

Notations

- HessUpdate: our Code
- GradDescent: classical gradient descent (“ascent” in this case)
- PrecondGrad: prefactored full Laplacian each Newton step
- Simplex: a classical simplex algorithm on *the unregularized problem*
- Fast L_1: using a finite-volume discretization of mass flow, by applying a first order primal-dual method[7] for a quadratically regularized L_1 problem[8]

Graph Data[9]

- 5 different sizes: 50, 100, 500, 1,000 and 5,000 nodes
- 10 different graphs for each size

Size	$\alpha = 10^{-5}$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	5	10
HessUpdate								
50	$4.4 \cdot 10^{-3}$	$6.48 \cdot 10^{-3}$	$4.93 \cdot 10^{-3}$	$5.73 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$	$2.33 \cdot 10^{-2}$	$3.42 \cdot 10^{-2}$
100	$1.52 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$1.39 \cdot 10^{-2}$	$6.52 \cdot 10^{-2}$	$7.51 \cdot 10^{-2}$	0.14	0.14
500	0.33	0.39	0.87	2.97	6.5	2.75	3.59	4.25
1000	1.2	6.32	6.78	8.6	11.14	13.61	13.77	13.55
5000	3.4	4.66	7.84	9	11.56	14.28	13.3	14.73
GradDescent								
50	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
100	0.89	0.89	0.9	0.9	0.9	0.89	0.9	0.89
500	21.19	21.17	21.23	21.14	21.16	21.16	21.17	21.14
1000	87.88	87.86	87.93	87.89	87.9	88.05	88.02	87.88
5000	586.73	583.2	582.73	540.24	539.91	540.64	539.58	540.17
PrecondGrad								
50	0.34	0.33	0.23	$6.92 \cdot 10^{-2}$	$2.18 \cdot 10^{-2}$	$1.09 \cdot 10^{-2}$	$6.41 \cdot 10^{-3}$	$4.74 \cdot 10^{-3}$
100	1.77	1.7	1.52	0.62	0.25	$8.57 \cdot 10^{-2}$	$4.95 \cdot 10^{-2}$	$3.67 \cdot 10^{-2}$
500	23.81	23.73	23.89	22.28	6.86	2.28	1.21	0.5
1000	94.67	94.65	94.87	94.53	45.2	29.56	8.5	4.04
5000	646.16	647.34	645.69	599.40	554.84	383.79	241.55	64.19
Simplex *on the unregularized problem*								
50	$1.07 \cdot 10^{-2}$							
100	$1.44 \cdot 10^{-2}$							
500	0.21							
1000	2.66							
5000	14.38							

Figure 1: Average runtimes in seconds for HessUpdate, GradDescent, and PrecondGrad as a function of the size of the graph and the parameter α, and average runtimes for Simplex as a function of the graph size

Size	$\alpha = 10^{-5}$	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	5	10
HessUpdate								
50	0	0	0	0	0	0	0	0
100	$1.2 \cdot 10^{-3}$	0	0	0	0	0	$4.8 \cdot 10^{-3}$	$4 \cdot 10^{-4}$
500	$1.4 \cdot 10^{-3}$	0	$5 \cdot 10^{-4}$	$1.87 \cdot 10^{-2}$	$2.19 \cdot 10^{-2}$	0	$3.84 \cdot 10^{-2}$	$6.64 \cdot 10^{-2}$
1000	$3.5 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$6.3 \cdot 10^{-3}$	$3.13 \cdot 10^{-2}$	$8.87 \cdot 10^{-2}$	$2.93 \cdot 10^{-2}$	0.27	0.46
5000	$5.9 \cdot 10^{-3}$	$5.5 \cdot 10^{-3}$	$1.23 \cdot 10^{-2}$	$7.55 \cdot 10^{-2}$	$8.9 \cdot 10^{-2}$	0.13	0.19	0.22
GradDescent								
50	0.88	0.26	$1.51 \cdot 10^{-2}$	$1.55 \cdot 10^{-2}$	∞	∞	∞	∞
100	0.86	0.27	$1.21 \cdot 10^{-2}$	$2.02 \cdot 10^{-2}$	∞	∞	∞	∞
500	0.85	0.3	$1.7 \cdot 10^{-2}$	∞	∞	∞	∞	∞
1000	0.87	0.32	$1.82 \cdot 10^{-2}$	∞	∞	∞	∞	∞
5000	0.25	$9.71 \cdot 10^{-2}$	$4.5 \cdot 10^{-3}$	∞	∞	∞	∞	∞
PrecondGrad								
50	0.22	$7.09 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$	0	0	0	0	0
100	0.26	$7.48 \cdot 10^{-2}$	$2.8 \cdot 10^{-3}$	0	0	0	0	0
500	0.37	0.1020	$5.7 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	0	0	0	0
1000	0.43	0.1112	$6.9 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	0	0	0	0
5000	0.52	0.3412	$2.01 \cdot 10^{-2}$	$1.2 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	0	0
Relative difference in L^1 cost between Simplex *and* HessUpdate								
50	0	0	0	0	$1.35 \cdot 10^{-2}$	$9.04 \cdot 10^{-2}$	$6.12 \cdot 10^{-2}$	$3.99 \cdot 10^{-2}$
100	0	0	0	0	$1.4 \cdot 10^{-2}$	$8.73 \cdot 10^{-2}$	$5.56 \cdot 10^{-2}$	$3.53 \cdot 10^{-2}$
500	0	0	0	0	$1.9 \cdot 10^{-2}$	$8.03 \cdot 10^{-2}$	$4.84 \cdot 10^{-2}$	$3.03 \cdot 10^{-2}$
1000	0	0	0	0	$2 \cdot 10^{-2}$	$7.41 \cdot 10^{-2}$	$4.31 \cdot 10^{-2}$	$2.67 \cdot 10^{-2}$
5000	0	0	0	0	$2.15 \cdot 10^{-2}$	$7.33 \cdot 10^{-2}$	$4.23 \cdot 10^{-2}$	$2.62 \cdot 10^{-2}$

Figure 2: Average relative error for HessUPDATE, GRADDESCENT, and PRECONDGRAD as a function of the size of the graph and the parameter α. ∞ means that the error was too big and the algorithm did not converge. Also represented is the average relative difference L^1 cost between HESSUPDATE and SIMPLEX

Grid size	$\alpha = 10^{-4}$	10^{-3}	10^{-2}	10^{-1}	1	5	10
HESSUPDATE							
10	0.9	0.11	$5.84 \cdot 10^{-2}$	$7.43 \cdot 10^{-2}$	$6.09 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$
20	4.14	6.42	0.9	1.26	0.7	0.59	0.69
30	9.4	22.71	11.89	6.08	8.4	6.02	6.09
40	19.46	40.56	60.49	37.89	36.41	40.19	39.56
FAST L^1							
10	0.15	0.3	0.32	$8.73 \cdot 10^{-2}$	0.54	1.33	12.2
20	0.59	1.23	1.72	1.53	2.9	51.09	31.87
30	1.57	4.87	5.27	10.28	14.63	173.4	143.23
40	4.69	20.1	23.47	41.28	49.31	550.65	569.87

Figure 3: Average runtime in seconds for HESsUPDATE and FAST L^1 for regular grids of $\mathbb{R}^2$ of different sizes and different parameters α . All solutions converged to within 0.5% of the ground truth value.

References I

- L. John Wiley Sons, “Linear programming and resource allocation modeling,” 2018.
- S. Boyd and L. Vandenberghe, “Convex optimization,” 2004.
- E. Facca, F. Cardin, and M. Putti, “Physarum dynamics and optimal transport for basis pursuit,” 12 2018.
- F. Santambrogio, “Optimal transport for applied mathematicians,” 2015.
- S. P. Boyd and L. Vandenberghe, “Convex optimization,” 2004.
- F. R. K. Chung, “Spectral graph theory,” 1996.
- A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” vol. 40, 2011.

References II

- W. G. Wuchen Li, Stanley Osher, "A fast algorithm for earth mover's distance based on optimal transport and l1 type regularization," pp. 16–66, Technical report, UCLA Computational and Applied Mathematics Report, 2016.
- F. Viger and M. Latapy, "Efficient and simple generation of random simple connected graphs with prescribed degree sequence," in *Computing and Combinatorics* (L. Wang, ed.), (Berlin, Heidelberg), pp. 440–449, Springer Berlin Heidelberg, 2005.

- Lagrange Duality - Prof. Daniel P. Palomar(Fall 2020-21 Slice)
  Hong Kong University of Science and Technology (HKUST)
- `https://inst.eecs.berkeley.edu/~ee227a/fa10/login/l_dual_strong`
- `https://inst.eecs.berkeley.edu/~ee127/sp21/livebook/l_sdual_opt`
- `https://en.wikipedia.org/wiki/Dual_norm`

Thanks!