The Simple Wasserstein Barycenter Problem

Henry Wolkowicz

Dept. Comb. and Opt., University of Waterloo, Canada

Workshop on Recent Advances in Optimization
October 11 - 12, 2023, Fields Institute, Stewart Library

work with:
A. Alfakih, Univ. of Windsor;
L. Jung and **W.M. Moursi**, Univ. of Waterloo.

"Nothing takes place in the world whose meaning is not that of some maximum or minimum."
Leonhard Euler

Problem

Problem (simplified Wasserstein barycenter problem)

- *given k sets consisting of n points each; select exactly one point from each set to*
- *minimize sum of distances to barycenter of k chosen pts.*

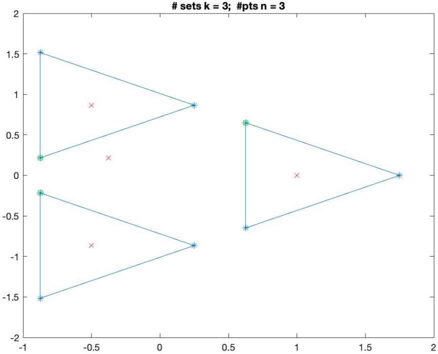


Figure: k=3=n: wheel of wheels; k odd; duality gaps; multiple opts

Outline

simplified Wasserstein barycenter problem

- problem is NP-hard.
- exploit Euclidean Distance Matrix structure;
  apply facial reduction to a doubly nonnegative relaxation;
- EMPHASIZE: obtain natural splitting for applying symmetric alternating direction method of multipliers, sADMM
- Empirics on random problems are surprisingly successful; find provable exact solution from upper/lower bounds
- examples with special symmetric structure result in duality gaps.

Open Question

• Surprisingly, we generally solve these NP-hard random problems to optimality, i.e., we find the exact optimal barycenter and optimal choice of points in each set and this yields a zero duality gap.
e.g., time for random problems $k = n = 25$ $O(10sec)$.
In contrast, MATLAB-CVX with Gurobi: $2, 348, 18000$ secs for $n = k = 5, 7, 8$, resp.

• We can find problems with positive duality gaps by generating problems with multiple solutions using special structure, e.g. symmetry.

• QUESTION: What is the key to characterizing problems with positive duality gaps? Is this related to rigidity of graph or uniqueness of optimal solutions?

Notation I

- $S \in \mathcal{S}^n$ space of $n \times n$ symmetric matrices with trace $S = \langle S, T \rangle$ inner prodcut;
  $\operatorname{diag}(S) \in \mathbb{R}^n$ is diagonal of S;
  adjoint is $\operatorname{diag}^*(v) = \operatorname{Diag}(v) \in \mathcal{S}^n$.
- $\mathbb{S}^n_+ \subset \mathcal{S}^n$, $X \succeq 0$, positive semidefinite cone;
  $\mathbb{S}^n_{++}$, $X \succ 0$, positive definite matrices
- $\mathcal{N}^n$ $n \times n$ nonnegative matrices;
- $\text{DNN} = \mathbb{S}^n_+ \cap \mathcal{N}^n$, doubly nonnegative cone
- e vector of ones

Notation II

- given points $p_i \in \mathbb{R}^d$, let $P^T = \begin{bmatrix} p_1 & p_2 & \dots & p_t \end{bmatrix} \in \mathbb{R}^{dt}$;
  Wlog points span $\mathbb{R}^d$ and are centered:

  $$P^T e = 0, \; e \text{ vector of ones.}$$

  $(P^T \mapsto P^T - ve^T, v := \frac{1}{n} P^T e)$
- corresponding: Euclidean distance matrix, EDM,
  $D_{ij} = \|p_i - p_j\|^2$ and
  Gram matrix, $G = PP^T$; and by Schoenberg [3]
  the Lindenstrauss operator, $\mathcal{K}(G)$ relates D, G:

  $$D = \mathcal{K}(G) = \operatorname{diag}(G)e^T + e\operatorname{diag}(G)^T - 2G.$$

Notation III

$D = \mathcal{K}(G) = \mathrm{diag}\,(G)e^T + e\mathrm{diag}\,(G)^T - 2G$

- Moreover, $\mathcal{K}$ (Lindenstrauss operator): one-one and onto between centered subspace, $\mathcal{S}_C^n$ and hollow subspace, $\mathcal{S}_H^n$

$$\mathcal{K} : \quad \mathcal{S}_C^n \leftrightarrow \mathcal{S}_H^n$$

$$\mathcal{S}_C^n = \{X \in \mathcal{S}^n \ : \ Xe = 0\}, \quad \mathcal{S}_H^n = \{X \in \mathcal{S}^n \ : \ \mathrm{diag}\,X = 0\}.$$

(Note centering $P^T e = 0 \implies G \in \mathcal{S}_C^n$.)

Simplified Wasserstein Barycenters and EDMs

Main Problem; Wasserstein Barycenter

- given sets $S_1, ..., S_k$, each with n points in $\mathbb{R}^d$
- Find optimal barycenter point y after choosing exactly one point from each set:

$$\begin{aligned} p_W^* &:= \min_{\substack{y \in \mathbb{R}^d \\ p_i \in S_i}} \sum_{i \in [k]} \|p_i - y\|^2 \\ &= \min_{p_i \in S_i} \min_{y \in \mathbb{R}^d} \sum_{i \in [k]} \|p_i - y\|^2 \end{aligned}$$

$([k] = \{1, 2, \ldots, n\}$

Connection to EDM

Lemma (minimal property of standard barycenter)

Suppose that we are given k points $q_i \in \mathbb{R}^d$, $i = 1, \ldots k$. Let $y = \frac{1}{k}\sum_{i=1}^{k} q_i$ denote the barycenter. Then sum of squared distances are minimized:
$\sum_{i=1}^{k} \|q_i - y\|^2 < \sum_{i=1}^{k} \|q_i - (y + v)\|^2$, $\forall 0 \neq v \in \mathbb{R}^d$.

Proof.

Wlog, assume points are centered at origin, i.e., translate $q_i \to q_i - y$. Since $ky = \sum_i q_i = 0$, for any $0 \neq v \in \mathbb{R}^d$,

$$\sum_{i=1}^{k} \|q_i\|^2 < \sum_{i=1}^{k} \|q_i\|^2 + k\|v\|^2 = \sum_{i=1}^{k} \|q_i - v\|^2.$$

□

Equivalent Problem

Proposition

Consider the main problem that consists in finding the optimal barycenter y of the optimal points p_i, $y = \frac{1}{k}\sum_{i\in[k]} p_i$. This is equivalent to finding exactly one point in each set that minimizes the sum of squared distances:

$$(WIQP) \qquad 2kp_W^* = p^* := \min_{p_1\in S_1,\ldots,p_k\in S_k} \sum_{i,j\in[k]} \left\|p_i - p_j\right\|^2.$$

Equivalent Sum of Squared Differences

$$(WIQP) \qquad 2kp_W^* = p^* := \min_{p_1\in S_1,\dots,p_k\in S_k} \sum_{i,j\in[k]} \left\|p_i - p_j\right\|^2.$$

Proof.

Let $p_i, i \in [k]$ be optimal; let y be barycenter. Wlog, translate $p_j \leftarrow p_j - y, \forall j$, $y = 0$. Therefore, redefine P with points $p_i, i \in [k]$ in rows of P, and centered, i.e., $P^Te = 0$, $PP^Te = Ge = 0$. Now

$$\begin{aligned}
\sum_{i,j\in[k]} \left\|p_i - p_j\right\|^2 &= e^TDe \\
&= e^T\left(\operatorname{diag}(G)e^T + e\operatorname{diag}(G)^T - 2G\right)e \\
&= 2k \operatorname{trace} G \quad (Ge = 0) \\
&= 2k\sum_{i\in[k]} \|p_i\|^2 \\
&= 2kp_W^*, \quad \text{from previous lemma.}
\end{aligned}$$

□

Reformulation Using EDM

x binary; $A = I \otimes e$

$$x := [v_1^T, \ldots, v_k^T]^T \in \mathbb{R}^{nk}, \quad A := \text{blkdiag}[e^T, \ldots, e^T] \in \mathbb{R}^{k \times nk}.$$

$Ax = e$ pick exactly one point from each set

BCQP binary-constrained quadratic problem; $D \in \mathrm{EDM}$

$$(BCQP) \qquad \begin{array}{rll} p^* = & \min & x^T D x = \operatorname{trace} D x x^T = \langle D, x x^T \rangle \\ & \text{s.t.} & Ax = e \\ & & x \in \{0,1\}^{kn} \end{array}$$

Note EDMs are nsd on $e^{\perp}$; constraints imply $x^T e$ is constant

therefore projected Hessian is nsd, i.e., minimizing a concave function, NP-HARD problem. AND constraints are totally unimodular.

Semidefinite, SDP, Relaxation of BCQP

A lifting to matrix space

use vector $\begin{pmatrix} 1 \\ x \end{pmatrix}$ and lift to rank-1 matrix $Y_x := \begin{pmatrix} 1 \\ x \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix}^T$,
and then relax the nonconvex rank-1 constraint. During the relaxation stage, we impose the constraints that we have on x, such as the $\{0,1\}$: $x_i^2 - x_i = 0$, constraints on x represented as:

$$\text{arrow}(Y_x) = e_0 \quad (0 - \text{th unit vector});$$

$$\text{arrow} : \mathbb{S}_+^{n+1} \to \mathbb{R}_+^{n+1} : \begin{bmatrix} s_0 & s^T \\ s & \bar{S} \end{bmatrix} \mapsto \begin{pmatrix} s_0 \\ \operatorname{diag}(\bar{S}) - s \end{pmatrix}.$$

The binary constraint on vector x is equivalent to the arrow constraint on lifted matrix Y_x if rank-one holds.

SDP reformulation via facial reduction

$$\hat{D} := \begin{bmatrix} 0 & 0 \\ 0 & D \end{bmatrix} \in \mathbb{S}^{kn+1}; \; K := \begin{bmatrix} -e^T \\ A^T \end{bmatrix} \begin{bmatrix} -e^T \\ A^T \end{bmatrix}^T \in \mathbb{S}_+^{kn+1}$$

Reformulate objective/constraint

objective function of BCQP: $\langle D, xx^T \rangle = \langle \hat{D}, Y_x \rangle$
For linear equality constraint, $K, Y_x \succeq 0$,

$$\begin{aligned}
Ax = e \iff & \begin{bmatrix} 1 \\ x \end{bmatrix}^T \begin{bmatrix} -e^T \\ A^T \end{bmatrix} = 0 \\
\iff & Y_x K := \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T \begin{bmatrix} -e^T \\ A^T \end{bmatrix} \begin{bmatrix} -e^T \\ A^T \end{bmatrix}^T = 0 \\
\iff & \langle Y_x, K \rangle = 0 \\
\iff & KY_x = 0, \text{ i.e: } \mathrm{Range}(Y_x) \subseteq \mathrm{Null}(K).
\end{aligned}$$

The last step follows since both $K, Y_x \succeq 0$.

Facial Reduction

$KY = 0$ both $K, Y \succeq 0$

If we choose V so that $\mathrm{Range}(V) = \mathrm{Null}(K)$, then we can *facially reduce* the problem using the substitution

$$\left\{ Y \leftarrow VRV^T \right\} \trianglelefteq \mathbb{S}_+^{kn+1}, \quad R \in \mathbb{S}_+^{nk+1-k}.$$

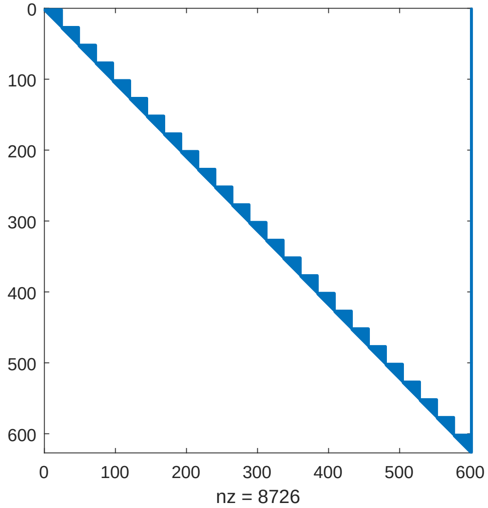
This makes the constraint $KY = 0$ redundant.
Therefore, the SDP reformulation is

$$\text{(SDP)} \qquad \begin{array}{rll} p^* = & \min_{Y \in \mathbb{S}^{nk+1}} & \langle \hat{D}, Y \rangle \\ & & \mathrm{arrow}(Y) = e_0 \\ & & \mathrm{rank}\,(Y) = 1 \\ & & {\color{red} KY = 0} \text{ discard after FR} \\ & & Y \succeq 0 \end{array}$$

If ${\color{red} Y \leftarrow VRV^T}$, then we can discard $KY = 0$ constraint.

Simple Structure of V

same diagonal upper-triangular blocks



(powerful) Gangster constraint

PROP: fixes zeros/shoots holes in certain entries of matrix

Let x be feasible for BCQP. Then

$$[A^TA - I] \circ xx^T = 0 \quad \text{(Hadamard product)},$$

and $A^TA - I \geq 0, xx^T \geq 0$. Define the gangster indices

$$\mathcal{J} := \left\{ ij : \left(A^TA - I\right)_{ij} > 0 \right\}.$$

The gangster constraint on feasible Y is $Y_{00} = 1$ and

$$\mathcal{J}(Y) = Y_{\mathcal{J}} = 0 \in \mathbb{R}^{|\mathcal{J}|}.$$

Proof

$A = I_k \otimes e$ (Kronecker)

$\operatorname{Diag}(\operatorname{diag}(A^T A)) = I_{kn}$

$$
\begin{aligned}
Ax = e \iff & \quad A^T A x = A^T e = \operatorname{diag}(A^T A) \\
\iff & \quad A^T A x - I x = A^T e - I x \\
& \qquad = \operatorname{diag}(A^T A) - \operatorname{Diag}[\operatorname{diag}(A^T A)] x \\
\iff & \quad (A^T A - I) x = \operatorname{diag}(A^T A) \circ (e - x) = e - x \\
\iff & \quad (A^T A - I) x x^T = (e - x) x^T = e x^T - x x^T \\
\iff & \quad \operatorname{trace}[(A^T A - I) x x^T] = \operatorname{trace}[e x^T - x x^T] \\
& \qquad = \sum_{i=1}^{nk} x_i - x_i^2 = 0 \\
\iff & \quad (A^T A - I) \circ x x^T = 0.
\end{aligned}
$$

The final conclusion now follows from the nonnegativities in the Hadamard product. ■

Update SDP relaxation

complete gangster index: $\hat{\mathcal{J}} := \{(0,0)\} \cup \mathcal{J}$

gangster indices J are nonzeros of (hollow) block diagonal matrix $A^T A - I$, i.e., the set of off-diagonal indices of the n-by-n diagonal blocks of the bottom right of Y_x.

SDP relaxation becomes

$$
\begin{array}{rll}
p^* = & \min_{Y \in \mathbb{S}^{nk+1}} & \langle \hat{D}, Y \rangle \\
& & \operatorname{arrow}(Y) = e_0 \\
& & \mathcal{G}_{\hat{J}}(Y) = e_0 \\
& & KY = 0 \\
& & Y \succeq 0
\end{array}
$$

where by abuse of notation, $\mathcal{G}_{\hat{J}}(Y) \cong Y_{\hat{J}}$

Doubly nonnegative (DNN) relaxation with FR

natural splitting use FR: $Y = VRV^T, V^TV = I$

variables $R \in \mathbb{S}_+^{nk+1-k}$, $Y \in \mathbb{S}_+^{nk+1}$;
use facial vector V with orthonormal columns;
constraint $KY = 0$ is discarded.

$x_i \in \{0, 1\}$

Lifting for feasible Y implies $0 \leq Y \leq 1$.

Trace constraint

Lemma (redundant trace constraint)

$$KY = 0, \operatorname{arrow}(Y) = e_0, V^T V = I$$

$$\implies k + 1 = \operatorname{trace}(Y) = \operatorname{trace} VRV^T = \operatorname{trace}(R)$$

Proof.

$$K := \begin{bmatrix} -e^T \\ A^T \end{bmatrix} \begin{bmatrix} -e^T \\ A^T \end{bmatrix}^T \implies \operatorname{Null}(K) = \operatorname{Null}\left(\begin{bmatrix} -e^T \\ A^T \end{bmatrix}^T \right) \implies$$

$$0 = KY \iff 0 = \begin{bmatrix} -1 & e^T & \dots & 0^T \\ \dots & \dots & \dots & \dots \\ -1 & 0^T & \dots & e^T \end{bmatrix} \begin{bmatrix} Y_{0,0} & \dots & Y_{0,nk} \\ \dots & \dots & \dots \\ Y_{nk,0} & \dots & Y_{nk,nk} \end{bmatrix},$$

implies $\operatorname{trace}(Y) = Y_{0,0} + \sum_{j=1}^{k} \sum_{i=1}^{n} Y_{jn+i,0} = 1 + k.$ □

The DNN Relaxation

polyhedral set constraints

$$\mathcal{Y} := \{Y \in \mathbb{S}^{nk+1} : \mathcal{G}_{\mathcal{J}}(Y) = Y_{\mathcal{J}} = e_0, \mathrm{arrow}(Y) = e_0, 0 \leq Y \leq 1\}$$

cone set constraints

$$\mathcal{R} := \{R \in \mathbb{S}_+^{nk+1-k} : \mathrm{trace}(R) = k+1\}.$$

DNN Relaxation

$$(\mathrm{DNN}) \qquad \begin{array}{rl} \min_{R,Y} & \langle \hat{D}, Y \rangle \\ \text{s.t.} & Y = VRV^T \\ & Y \in \mathcal{Y} \\ & R \in \mathcal{R} \end{array}$$

Optimality conditions; a Constraint Qualification holds

first-order optimality conditions for DNN

primal-dual pair (Y, R, Z) is optimal if, and only if,

$$\begin{array}{rcll} Y & = & VRV^T, \quad R \in \mathcal{R},\ Y \in \mathcal{Y} & \text{(primal feasibility)} \\ 0 & \in & -V^T Z V + \mathcal{N}_{\mathcal{R}}(R) & \text{(dual } R \text{ feasibility)} \\ 0 & \in & \hat{D} + Z + \mathcal{N}_{\mathcal{Y}}(Y) & \text{(dual } Y \text{ feasibility)} \end{array}$$

Optimality conditions using projections

primal-dual pair (R, Y, Z) is optimal for DNN

$$\iff$$

$$\begin{array}{rcl} R & = & \mathcal{P}_{\mathcal{R}}(R + V^T Z V) \\ Y & = & \mathcal{P}_{\mathcal{Y}}(Y - \hat{D} - Z) \\ Y & = & VRV^T \end{array}$$

(modified/symmetric) ADMM or PRSM algorithm

two "names"

symmetric alternating directions method of multipliers
Peaceman-Rachford splitting method (if on dual)

augmented Lagrangian for DNN; parameter $\beta > 0$

$$
\begin{array}{l}
\mathcal{L}_\beta(Y, R, Z) \\
\qquad := \quad \langle \hat{D}, Y \rangle + \langle Z, Y - VRV^T \rangle + \frac{\beta}{2}\| Y - VRV^T \|_F^2 \\
\qquad\qquad + \iota_{\mathcal{Y}} Y + \iota_{\mathcal{R}} R,
\end{array}
$$

where $\iota_S(\cdot)$ is indicator function for set S.

Updates

Update using augmented Lagrangian

we update the primal variables R, Y with intermediate (two) updates of dual multipliers

$$
\begin{aligned}
R_{k+1} &= \operatorname{argmin}_{R \in \mathbb{S}^{nk+1-k}} \mathcal{L}_\beta(R, Y_k, Z_k) \\
Z_{k+\frac{1}{2}} &= Z^k + \beta(Y_k - VR_{k+1}V^T) \\
Y_{k+1} &= \operatorname{argmin}_{Y \in \mathbb{S}^{nk+1}} \mathcal{L}_\beta(R_{k+1}, Y, Z_{k+\frac{1}{2}}) \\
Z_{k+1} &= Z_{k+\frac{1}{2}} + \beta(Y_{k+1} - VR_{k+1}V^T).
\end{aligned}
$$

Primal updates for R (explicit)

R update using spectral decomp. of M

$$
\begin{aligned}
R-\text{update} &= \operatorname{argmin}_{R\in\mathbb{S}^{nk+1-k}} \mathcal{L}_\beta(R, Y^k, Z^k) \\
&= \operatorname{argmin}_{R\in\mathcal{R}} \| Y^k - VRV^T + \tfrac{1}{\beta} Z_k \|_F^2 \\
&\qquad \text{by completing the square} \\
&= \operatorname{argmin}_{R\in\mathcal{R}} \| V^T Y_k V - R + \tfrac{1}{\beta} V^T Z_k V \|_F^2 \\
&\qquad \text{since } V^T V = I \\
&= \operatorname{argmin}_{R\in\mathcal{R}} \| R - V^T (Y_k + \tfrac{1}{\beta} Z_k) V \|_F^2 \\
&= \mathcal{P}_\mathcal{R}\left(V^T (Y^k + \tfrac{1}{\beta} Z^k) V \right) \\
&\qquad =: \mathcal{P}_\mathcal{R}(M); \; M = U \mathrm{Diag}\,(d) U^T \\
&= U \mathrm{Diag}\,[\mathcal{P}_{\Delta_{k+1}}(d)] U^T
\end{aligned}
$$

where $\mathcal{P}_{\Delta_{k+1}}$ denotes the projection onto the simplex $\Delta_{k+1} := \{x \in \mathbb{R}^n_+ : \langle e, x \rangle = 1 + k\}$.

Primal update for Y (explicit)

Y with polyhedral constraints

$$
\begin{aligned}
\text{Y}-\text{update} &= \operatorname{argmin}_{Y \in \mathbb{S}^{nk+1}} \mathcal{L}_\beta(R_{k+1}, Y, Z_{k+\frac{1}{2}}) \\
&= \operatorname{argmin}_{Y \in \mathcal{Y}} \| Y - [V R_{k+1} V^T - \tfrac{1}{\beta}(\hat{D} + Z_{k+\frac{1}{2}})] \|_F^2 \\
&\qquad \text{by completing the square} \\
&= \mathcal{P}_{\mathcal{Y}} \left(V R_{k+1} V^T - \tfrac{1}{\beta}(\hat{D} + Z_{k+\frac{1}{2}}) \right) \\
&= \mathcal{P}_{arrowbox} \left(\mathcal{G}_{\hat{\mathcal{J}}} [V R_{k+1} V^T - \tfrac{1}{\beta}(\hat{D} + Z_{k+\frac{1}{2}})] \right)
\end{aligned}
$$

where $\mathcal{G}_{\hat{\mathcal{J}}}$ is the gangster constraint and $\mathcal{P}_{arrowbox}$ projects onto the polyhedral set $\{ Y \in \mathbb{S}^{nk+1} : Y_{ij} \in [0,1], \operatorname{arrow}(Y) = e_0 \}$.

Dual updates

Lagrange multipliers are essence of optimization

correct choice of Lagrange multiplier Z yields an unconstrained problem; important in obtaining strong lower bounds to prove optimality; (redundant) constraints on dual multipliers can be useful to speed up algorithm

Lemma (arrow projection)

Let $\mathcal{Z}_A := \Big\{ Z \in \mathbb{S}^{nk+1} : (Z + \hat{D})_{i,i} = 0,$

$$(Z + \hat{D})_{0,i} = 0, (Z + \hat{D})_{i,0} = 0, i = 1, ..., nk \Big\}.$$

Let (Y^*, R^*, Z^*) *be an optimal primal-dual pair for the* DNN*.*
Then, $Z^* \in \mathcal{Z}_A$*.*

Proof.

The proof of this fact uses the dual Y feasibility condition and a reformulation of the Y-feasible set. The details are in [2, Thm 2.14] and [1]. □

Modified dual variable update

project the dual variable onto $\mathcal{Z}_A$, i.e:

- $Z^{k+\frac{1}{2}} := Z^k + \beta\mathcal{P}_{\mathcal{Z}_A}(Y^k - VR^{k+1}V^T)$;
- $Z^{k+1} := Z^{k+\frac{1}{2}} + \beta\mathcal{P}_{\mathcal{Z}_A}(Y^{k+1} - VR^{k+1}V^T)$.

Algorithm (adaptive β)

rPRSM

- Initialization:
  $Y^0 = 0 \in S^{nk+1}, Z^0 = P_{Z_A}(0), \beta = \max(\lfloor \frac{nk+1}{k} \rfloor, 1)$
- WHILE: termination criteria are not met
  - $R^{k+1} = U\mathrm{Diag}\,[P_{\Delta_{k+1}}(d)]U^T$ where $U\mathrm{Diag}\,(d)U^T = eig(V^T(Y^k + \frac{1}{\beta}Z^k)V)$
  - $Z^{k+\frac{1}{2}} = Z^k + \beta P_{Z_A}(Y^k - VR^{k+1}V^T)$
  - $Y^{k+1} = P_{box}[G_{\hat{J}}(VR^{k+1}V^T - \frac{1}{\beta}(\hat{D} + Z^{k+\frac{1}{2}}))]$
  - $Z^{k+1} = Z^{k+\frac{1}{2}} + \beta P_{Z_A}(Y^{k+1} - VR^{k+1}V^T)$

  ENDWHILE

Lower bounds

Proving optimality; early stopping conditions

Lagrangian dual function to DNN model is

$$
\begin{aligned}
g(Z) &= \min_{R\in\mathcal{R},Y\in\mathcal{Y}} \langle \hat{D}, Y\rangle + \langle Z, Y - VRV^T\rangle \\
&= \min_{Y\in\mathcal{Y},R\in\mathcal{R}} \langle \hat{D} + Z, Y\rangle - \langle Z, VRV^T\rangle \\
&= \min_{Y\in\mathcal{Y}} \langle \hat{D} + Z, Y\rangle + \min_{R\in\mathcal{R}} (-\langle V^T Z V, R\rangle) \\
&= \min_{Y\in\mathcal{Y}} \langle \hat{D} + Z, Y\rangle - \max_{R\in\mathcal{R}} \langle V^T Z V, R\rangle \\
&= \min_{Y\in\mathcal{Y}} \langle \hat{D} + Z, Y\rangle - \max_{\|v\|^2=(k+1)} v^T V^T Z V v \\
&= \min_{Y\in\mathcal{Y}} \langle \hat{D} + Z, Y\rangle - (k+1)\lambda_{max}(V^T Z V).
\end{aligned}
$$

Upper bounds

rounding with 0-column

$Y(1 : end, 0$ and compute its nearest feasible solution to BCQP (an LSAP). It is equivalent to signal only the maximum weight index for each consecutive block of length n. The proof is in [1, section 3.2.2].

alternatively, use dominant eigenvector of Y

compute its nearest feasible solution to BCQP. It is again equivalent to signal only the maximum weight index for each consecutive block of length n.

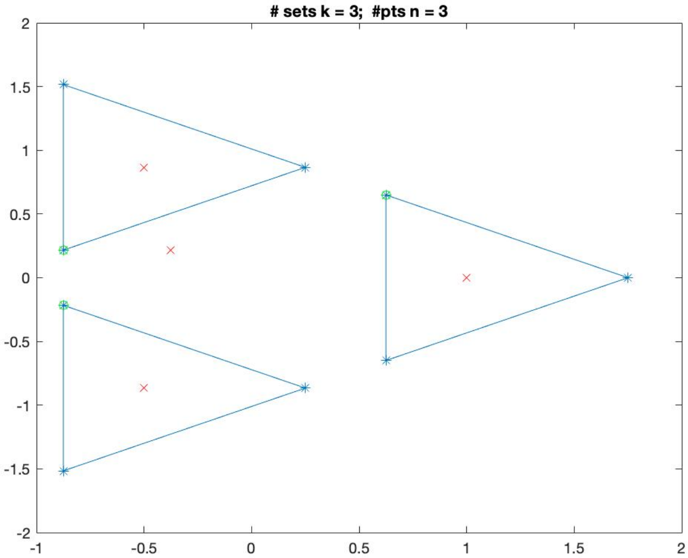
Random data

Specifications			Time (s)		Relative duality gap	
d	n	k	sADMM	Mosek	sADMM	Mosek
2	7	5	2.33e-01	3.66e-01	9.80e-08	2.41e-09
2	8	6	3.90e-01	6.94e-01	2.76e-10	5.91e-11
2	9	7	3.53e-01	1.30e+00	6.59e-07	1.55e-11
2	10	8	3.75e-01	3.92e+00	4.82e-08	4.96e-12
2	11	9	4.63e-01	1.30e+01	1.92e-09	2.21e-12
2	12	10	5.41e-01	3.09e+01	9.32e-10	8.41e-10
2	13	11	7.22e-01	7.31e+01	1.83e-08	2.94e-11

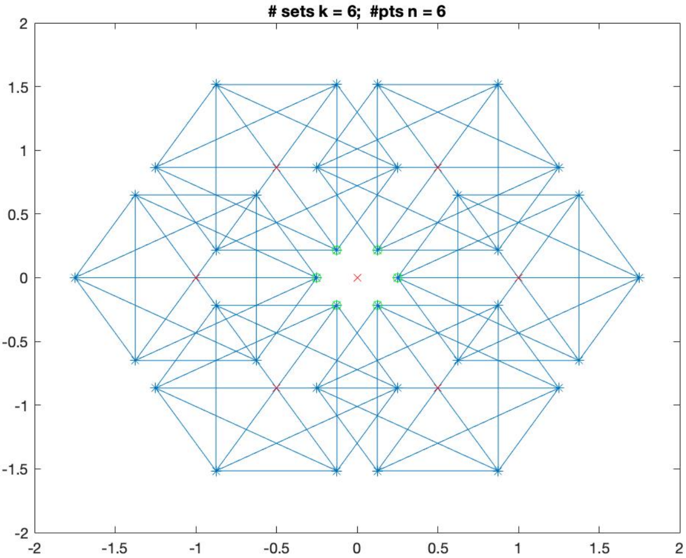
Scalability for large size

d	n	k	Time(s)	KKT residual	Relative duality gap
3	3	3	2.36e-02	2.20e-07	7.52e-15
4	4	4	1.38e-01	3.10e-08	9.95e-17
5	5	5	1.80e-01	7.02e-09	3.42e-16
6	6	6	3.06e-01	1.89e-08	9.09e-15
7	7	7	4.79e-01	1.19e-06	1.65e-14
8	8	8	3.16e-01	1.51e-06	5.83e-15
9	9	9	5.11e-01	1.43e-07	1.42e-14
10	10	10	5.46e-01	1.51e-07	1.46e-14
11	11	11	2.71e-01	7.38e-09	3.01e-14
12	12	12	1.01e+00	2.34e-08	2.02e-14
13	13	13	1.48e+00	4.76e-09	1.64e-14
14	14	14	2.98e+00	1.21e-06	2.75e-14
15	15	15	1.54e+00	9.83e-08	1.10e-14
16	16	16	1.27e+00	6.76e-08	1.70e-14
17	17	17	1.80e+00	1.36e-08	2.46e-14
18	18	18	2.44e+00	2.93e-06	3.17e-15
19	19	19	3.19e+00	9.19e-10	1.15e-14
20	20	20	5.53e+00	1.56e-09	4.15e-15
21	21	21	6.25e+00	1.53e-08	3.86e-14
22	22	22	1.38e+01	2.67e-06	1.32e-14
23	23	23	1.35e+01	4.16e-09	1.42e-14
24	24	24	1.64e+01	8.28e-07	3.56e-14
25	25	25	2.72e+01	1.73e-09	8.10e-16

wheel of wheels; k odd; duality gaps; multiple opts



k even unique opt



Conclusion

• the Simplified Wasserstein Barycenter problem, a NP-hard computational problem
• formulated as a binary constrained quadratic program
• applied doubly nonnegative relaxations and solved using facial reduction and symmetric alternating dirtection method of multipliers (sADMM) algorithm
• compute tight lower and upper bounds
• empirical results suggest: efficiency and accuracy and ability to exactly solve the NP-hard problem
• for input data with multiple optimal solutions, the algorithm has difficulty breaking ties and we get duality gaps
• • QUESTION: What is the key to characterizing problems with positive duality gaps? Is this related to rigidity of graph or uniqueness of optimal solutions?

- F. Burkowski, H. Im, and H. Wolkowicz.
  A Peaceman-Rachford splitting method for the protein side-chain positioning problem.
  Technical report, University of Waterloo, Waterloo, Ontario, 2022.
  arxiv.org/abs/2009.01450,21.
- N. Graham, H. Hu, H. Im, X. Li, and H. Wolkowicz.
  A restricted dual Peaceman-Rachford splitting method for a strengthened DNN relaxation for QAP.
  INFORMS J. Comput., 34(4):2125–2143, 2022.
- I.J. Schoenberg.
  Metric spaces and positive definite functions.
  Trans. Amer. Math. Soc., 44(3):522–536, 1938.

Thanks for your attention!

The Simple Wasserstein Barycenter Problem

Henry Wolkowicz

Dept. Comb. and Opt., University of Waterloo, Canada

Workshop on Recent Advances in Optimization
October 11 - 12, 2023, Fields Institute, Stewart Library

work with:
A. Alfakih, Univ. of Windsor;
L. Jung and **W.M. Moursi**, Univ. of Waterloo.

“Nothing takes place in the world whose meaning is not that of some maximum or minimum.”
Leonhard Euler