

Angular Embedding: A Robust Quadratic Criterion

Stella X. Yu, *Member, IEEE*

Abstract—Given the size and confidence of pairwise local orderings, angular embedding (AE) finds a global ordering with a near-global optimal eigensolution. As a quadratic criterion in the complex domain, AE is remarkably robust to outliers, unlike its real domain counterpart LS, the least squares embedding. Our comparative study of LS and AE reveals that AE's robustness is due not to the particular choice of the criterion, but to the choice of representation in the complex domain. When the embedding is encoded in the angular space, we not only have a non-convex error function that delivers robustness, but also have a Hermitian graph Laplacian that completely determines the optimum and delivers efficiency. The high quality of embedding by AE in the presence of outliers can hardly be matched by LS, its corresponding L_1 norm formulation, or their bounded versions. These results suggest that the key to overcoming outliers lies not with additionally imposing constraints on the embedding solution, but with adaptively penalizing inconsistency between measurements themselves. AE thus significantly advances statistical ranking methods by removing the impact of outliers directly without explicit inconsistency characterization, and advances spectral clustering methods by covering the entire size-confidence measurement space and providing an ordered cluster organization.

Index Terms—Least squares methods, spectral methods, graph algorithms, constrained optimization, linear programming, statistical computing, clustering, modeling and recovery of physical attributes.



1 INTRODUCTION

DETERMINING a global ranking of elements from their pairwise local comparisons is a fundamental problem in decision science [1]–[3], social choice theory [4], financial economics [5], statistics [6]–[10], machine learning [11]–[15], and computer vision [16]–[18].

These pairwise comparisons are increasingly given in terms of cardinal scores (*how much is one better than the other?*) instead of ordinal orderings (*which one of the two is better?*) [10]. Whether the scores are obtained from preference judgement by humans or from attribute extraction by a computational routine, as relative measures over partial observations of all the elements, these local orderings are often inconsistent between subsets of the elements to reach a unanimous global ordering.

The basic problem in unmasking the true global ranking inherent in the inconsistent local ranking data is to prevent rank reversal and preserve the rank magnitude as much as possible [2], [3], [12], [15], [17].

Formally, all the pairwise local orderings between n elements can be captured in a pair of $n \times n$ matrices (O, C) , with O for the size of the ordering and C for the confidence in the ordering. We seek an $n \times 1$ global ordering vector X such that if with consistency a global

ranking algorithm would produce $X_a > X_b$ for two elements a and b , then with slight perturbation or even large outliers in the local orderings (O, C) , a good global ranking algorithm would not produce a rank reversal: $X_a < X_b$, and preserve the magnitude of local orderings as much as possible: $X_a - X_b \rightsquigarrow O_{a,b}$, whenever $C_{a,b} \neq 0$. The problem of solving the global ordering $X_{n \times 1}$ from pairwise local orderings $(O_{n \times n}, C_{n \times n})$ is a data embedding problem, where n elements with inconsistent pairwise local measurements are now positioned into a metric space and can be compared with each other on a global scale.

This paper compares the newly proposed angular embedding (AE) with the conventional least squares embedding (LS). AE was first used for modeling subjective brightness from the objective intensity of an image [17], where pairwise local orderings are computed from intensity differences at multiple scales and integrated into a global brightness ordering. AE has also been applied to the figure-ground segmentation of natural scene images [18]. LS has been widely used in surface reconstruction [19], shape from shading [20], high dynamic range compression [21], image matting and fusion [16], and various model fitting scenarios [22], [23].

LS and AE differ on the representation used for encoding the global ordering and the objective function used for measuring the quality of rank preservation (Fig. 1).

1. Representation: linear versus angular space? While LS ranks the elements on a line, AE ranks them on a circle. While the size and confidence of a local ordering are separated as a difference and an importance weight on the difference in LS, they are integrated as the phase and magnitude of a single complex number in AE.

• The author is with the Computer Science Department, Boston College, 140 Commonwealth Ave, Chestnut Hill, MA 02467. Email: stella.yu@bc.edu.

Manuscript received 4 July 2009; revised 17 Feb. 2010; accepted 18 Apr. 2011; published online 13 May 2011.

Recommended for acceptance by F. Bach.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-07-0421.

Digital Object Identifier no. 10.1109/TPAMI.2011.107.

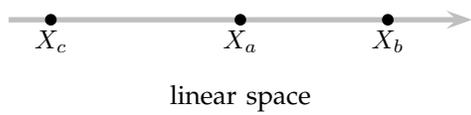
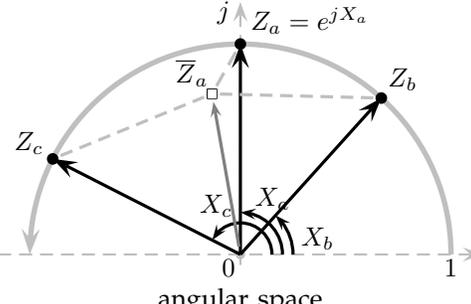
	a: embedding representation	b: embedding criterion
LS	 <p style="text-align: center;">linear space</p>	$\varepsilon_{LS}(X) = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (X_a - X_b - O_{a,b})^2$ <p style="text-align: center;">minimize variance</p>
AE	 <p style="text-align: center;">angular space</p>	$\varepsilon_{AE}(Z) = \sum_{a=1}^n D_{a,a} (Z_a - \bar{Z}_a)^2$ <p> $D_{a,a}$ = the total confidence associated with a. \bar{Z}_a = the average estimate of Z_a given $Z, (O, C)$. </p> <p style="text-align: center;">minimize distance</p>

Fig. 1. LS and AE differ in the representation used for encoding the global ordering X (LS' linear space vs. AE's angular space) and the objective function $\varepsilon(X)$ used for measuring the quality of rank preservation (minimize the variance of neighbour estimates of the embedding for LS vs. minimize the distance between the embedding and its neighbourhood average for AE).

2. Criterion: minimize variance versus minimize distance? While LS minimizes the variance of neighbour estimates of the embedding, AE minimizes the distance between the embedding and its neighbourhood average.

We investigate whether the representation or the criterion is more crucial for the distinctions between LS and AE, whether and how AE outperforms LS.

Consider ordering pixels by their surface heights, i.e., reconstructing a surface from noisy measurements of its gradients or more generally pairwise height differences. Without noise, the surface gradient field has zero curl and is integrable: The integral along any closed path of pixels is 0, and the height difference computed between two pixels does not depend on the choice of the path.

There are two basic approaches to this 2D integration problem: One focuses on obtaining an integrable gradient field by incorporating the integrability constraints into the estimation of the gradient field itself [24]–[27], and the other applies integrability during the estimate of the surface when a non-integrable gradient field is already provided [19], [25], [27]–[30]. These formulations often minimize an LS cost function between the reconstruction and a smooth surface space, assuming no measurement outliers. Outliers are often tackled explicitly and iteratively to reduce their impact on LS [16].

Fig. 2 compares LS and AE on reconstructing a surface given the same pairwise height differences. With Gaussian noise, the LS and AE results both closely match the ground-truth surface. With outliers, the LS surface loses the ground-truth surface to an overwhelming amount of spiky noise, whereas the AE surface has no noticeable change in the quality of reconstruction, with only a slight increase in the overall approximation error.

We will first formulate standard LS and AE criteria, and then develop an LS variant in terms of AE's distance

criterion and an AE variant in terms of LS' variance criterion. Examining these 4 criteria, ε_{LS} , ε_{AE} , ε_{LSD} , ε_{AEV} , allows us to not only establish that it is not the criterion but the representation that sets AE and LS apart, but also tease out AE's non-convex error function and robust error weight function, providing a compelling explanation to AE's remarkably robust embedding performance.

Quadratic formulations are notorious for their sensitivity to outliers, yet by seeking an embedding in the complex domain, AE deviates from LS and achieves robustness without identifying outliers [16] or resorting to regularization as in Bayesian least squares [31], [32]. Our comparative study on LS and AE suggests that, the key to overcoming outliers lies not with additionally imposing constraints on the solution, which is often not warranted in applications, but with adaptively penalizing inconsistency between measurements themselves.

2 EMBEDDING FROM PAIRWISE ORDERINGS

We formulate our embedding problem in graph theory and then proceed from the standard LS and AE criteria to their cross-combination variants. We use the same notations (e.g. L for Laplacian) for both LS and AE to draw similarity, sometimes adding subscripts for clarity.

2.1 Problem Formulation in Spectral Graph Theory

Given n elements and their pairwise local ordering measurements in the d -dimensional space, with size $O_{n \times n \times d}$ and confidence $C_{n \times n \times d}$, we would like to establish their global ordering in the same space, with coordinates $X_{n \times d}$. We focus on $d = 1$, as it will soon become evident that, for all our embedding criteria, solving the d -dimensional embedding problem is equivalent to solving d independent 1-dimensional embedding problems.

For elements a and b , their local ordering size $O_{a,b}$ can be arbitrary, whereas the confidence level $C_{a,b}$ in this size is always between 0 and 1. A positive (negative) $O_{a,b}$ can be interpreted as a measurement on how much better (worse) a is than b . While there often holds $O_{a,b} = -O_{b,a}$ and $C_{a,b} = C_{b,a}$, neither O nor C is required to have such consistency. Naturally, $O_{a,a} = 0$ and $C_{a,a} = 1$.

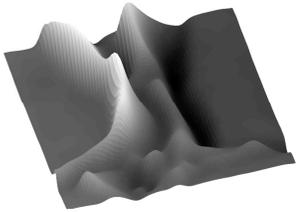
The goal of embedding is to find a global ordering X , such that its relative differences between elements match those pairwise local ordering measurements:

$$E_{a,b} = (X_a - X_b) - O_{a,b} \rightsquigarrow 0, \text{ if } C_{a,b} > 0 \quad (1)$$

When no global ordering X can fulfill all these measurements, a criterion assessing the overall fulfillment errors is needed to help choose the optimal X .

In graph theory, the pairwise local ordering measurements are fully captured in a weighted graph of n nodes, where each node denotes an element, and the edge going from node a to node b is associated with two weights: $O_{a,b}$ and $C_{a,b}$, describing how much a is better than b and how much confidence we have in this information.

a: ground truth



b: Gaussian noise

c: + 10% outliers

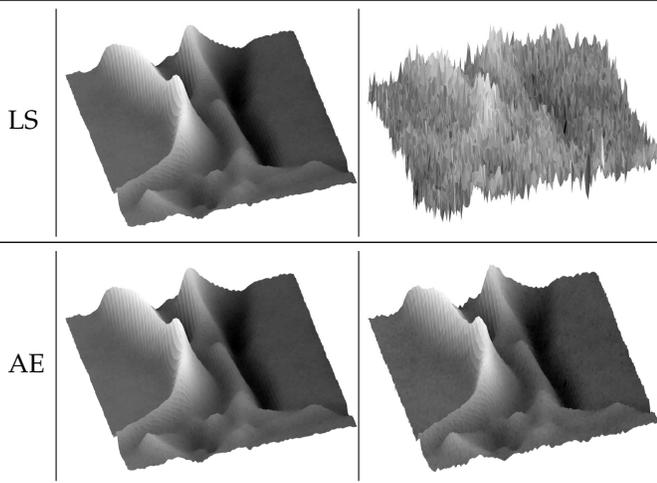


Fig. 2. AE with a quadratic criterion in the complex domain is remarkably robust to outliers, unlike its real domain counterpart LS. a) A ground-truth surface with a height range of 1 and a maximal gradient magnitude of 0.16. b) Gaussian noise with $\sigma = 0.05$ is added to the pairwise height differences between pixels within a city-block distance of 2. Integration of these differences by LS and AE shows comparable reconstructions, both with a standard error of 0.0085. c) Uniformly distributed outliers of values ± 3 are added to 10% of these noisy height differences. The standard error becomes 0.1666 for LS and 0.0184 for AE. The surface reconstructed by LS is barely recognizable, while that by AE remains a close approximation to the ground truth surface.

In this directed graph, node a has 2 connections with node b : an outgoing edge $(O_{a,b}, C_{a,b})$ and an incoming edge $(O_{b,a}, C_{b,a})$. We consider b a *neighbour* of a 's if $C_{a,b} > 0$ or $C_{b,a} > 0$. The *neighbourhood* of a is thus $N(a) = \{b : C_{a,b} > 0 \text{ or } C_{b,a} > 0\}$. Confidence C can be regarded as the strength of connection between nodes, i.e. *weight* in the traditional sense, with 0 for a non-existing edge and 1 for a maximal edge, while ordering O can be regarded as the relative displacement between nodes.

The problem of embedding is to assign X_a to node a , $\forall a$, so that all the nodes become ordered by their X numbers. To find the global ordering X that fulfills (O, C) , the local views of (O, C) from all the neighbourhoods need to be reconciled and propagated to the entire graph.

It is helpful to express a criterion as a sum of n errors, one per node. From node a 's point of view, the degree of weights or the total confidence from all its edges is:

$$\text{degree or total confidence: } D_{a,a} = \sum_{b=1}^n (C_{a,b} + C_{b,a}) \quad (2)$$

Its neighbour b has 2 estimates of a 's position:

	neighbour estimate	relative confidence
outgoing	$\vec{X}_{a,b} = X_b + O_{a,b}$	$\vec{C}_{a,b} = \frac{C_{a,b}}{D_{a,a}}$

(3)

incoming	$\overleftarrow{X}_{a,b} = X_b - O_{b,a}$	$\overleftarrow{C}_{a,b} = \frac{C_{b,a}}{D_{a,a}}$
----------	---	---

(4)

The average of the neighbour estimate of a 's position is:

$$\text{average: } \bar{X}_a = \sum_{b=1}^n (\vec{C}_{a,b} \vec{X}_{a,b} + \overleftarrow{C}_{a,b} \overleftarrow{X}_{a,b}) \quad (5)$$

These definitions lend themselves to matrix notations.

	outgoing	incoming
total confidence	$D = \text{Diag}(C1_n + C'1_n)$	

(6)

relative confidence	$\vec{C} = D^{-1}C$	$\overleftarrow{C} = D^{-1}C'$
---------------------	---------------------	--------------------------------

(7)

neighbour estimate	$\vec{X} = 1_n X' + O$	$\overleftarrow{X} = 1_n X' - O'$
--------------------	------------------------	-----------------------------------

(8)

neighbourhood average	$\bar{X} = (\vec{C} \bullet \vec{X} + \overleftarrow{C} \bullet \overleftarrow{X})1_n$	
-----------------------	--	--

(9)

where $'$ denotes matrix conjugate transpose, $\text{Diag}(\cdot)$ the diagonal matrix formed from its vector argument, \bullet the Hadamard (element-wise) product of two matrices, and 1_n the $n \times 1$ vector of 1s.

The 4 embedding criteria we will consider can all be written as a sum of n node-centric errors, each pertinent to a single node in the graph and capturing the local view of the embedding X with respect to the neighbour estimate $(\vec{X}, \overleftarrow{X})$ or the neighbourhood average \bar{X} .

2.2 LS Standard: Minimize Variance

The standard LS criterion seeks to minimize the embedding errors weighted by their confidence levels:

$$\varepsilon_{LS} = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} E_{a,b}^2 = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (X_a - X_b - O_{a,b})^2 \quad (10)$$

Writing it as n node-centric errors, we see that the usual weighted LS is symmetrical with respect to \vec{X} and \overleftarrow{X} :

$$2\varepsilon_{LS} = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} E_{a,b}^2 + \sum_{a=1}^n \sum_{b=1}^n C_{b,a} E_{b,a}^2 \quad (11)$$

$$= \sum_{a=1}^n D_{a,a} \sum_{b=1}^n \overrightarrow{C}_{a,b} (X_a - \overrightarrow{X}_{a,b})^2 + \overleftarrow{C}_{a,b} (X_a - \overleftarrow{X}_{a,b})^2 \quad (12)$$

$$= D (\overrightarrow{C} \bullet (X1'_n - \overrightarrow{X}) \bullet (X1'_n - \overrightarrow{X}) + \overleftarrow{C} \bullet (X1'_n - \overleftarrow{X}) \bullet (X1'_n - \overleftarrow{X})) 1_n. \quad (13)$$

If we assume that \overrightarrow{X}_a and \overleftarrow{X}_a each has a mean of X_a , ε_{LS} aims to minimize the variance of the neighbour estimate (Fig. 3).

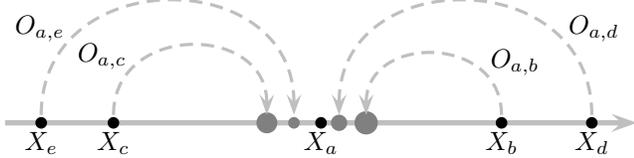


Fig. 3. LS seeks to minimize the variance in the neighbour estimate of the embedding. For element a , every neighbour b casts its estimate of X_a based on its own position X_b and local ordering $O_{a,b}$ with confidence $C_{a,b}$. ε_{LS} measures the total variance of these estimates around X_a from all as .

We write ε_{LS} as a quadratic form of vector X :

$$\varepsilon_{LS}(X) = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (X_a^2 + X_b^2 - 2X_a X_b) \quad (14)$$

$$- \sum_{a=1}^n \sum_{b=1}^n 2(X_a - X_b) C_{a,b} O_{a,b} + \sum_{a=1}^n \sum_{b=1}^n C_{a,b} O_{a,b}^2 \quad (14)$$

$$= X'(D - C - C')X - 2X'(C \bullet O - C' \bullet O)1_n \quad (15)$$

$$+ 1'_n(C \bullet O \bullet O)1_n \quad (\text{constant, to be dropped}). \quad (16)$$

Dropping the last constant term, we rewrite $\varepsilon_{LS}(X)$ in terms of graph Laplacian L and measurement M :

$$\text{Laplacian: } L = D - (C + C'), \quad (17)$$

$$\text{measurement: } M = C \bullet O - (C \bullet O)', \quad (18)$$

$$\text{minimize: } \varepsilon_{LS}(X) = X' L X - 2X' M 1_n. \quad (19)$$

The optimal LS embedding X_{LS}^* is thus $X_{LS}^* = L^{-1} M 1_n$. Since $L 1_n = 0$, L is rank deficient and L^{-1} is interpreted as the pseudo-inverse of L . Given that L depends only on C , the propagation aspect of global integration is regulated by C only, while the local ordering reconciliation aspect is primarily achieved by $M 1_n$, the confidence-weighted sum of local orderings in each neighbourhood.

In spectral graph theory [33], the normalized Laplacian \tilde{L} is often more revealing of the graph structure. Given any degree D and Laplacian L , we define

$$\text{normalized Laplacian: } \tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}. \quad (20)$$

We can rewrite LS as follows:

$$\text{minimize: } \varepsilon_{LS}^*(\tilde{X}) = \tilde{X}' \tilde{L} \tilde{X} - 2\tilde{X}' \tilde{M} 1_n, \quad (21)$$

$$\text{scaled embedding: } \tilde{X} = D^{\frac{1}{2}} X, \quad (22)$$

$$\text{normalized Laplacian: } \tilde{L} = I - D^{-\frac{1}{2}} (C + C') D^{-\frac{1}{2}}, \quad (23)$$

$$\text{normalized measurement: } \tilde{M} = D^{-\frac{1}{2}} M, \quad (24)$$

$$\text{optimal solution: } \tilde{X}_{LS}^* = \tilde{L}^{-1} \tilde{M} 1_n. \quad (25)$$

If (O, C) are d -dimensional measurements, we have

$$\varepsilon_{LS}(X) = \sum_{k=1}^d \sum_{a=1}^n \sum_{b=1}^n C_{a,b,k} (X_{a,k} - X_{b,k} + O_{a,b,k})^2. \quad (26)$$

Since each of the d dimensions (column k) of $X_{n \times d}$ can be optimized independently with respect to its own 1-dimensional pairwise measurements, the d -dimensional embedding X_{LS}^* is simply d 1-dimensional embeddings.

2.3 AE Standard: Minimize Distance

While LS realizes the global ordering X in the positions of points on a line, AE realizes it in the angles of points on a unit circle (Fig. 4). These points can be conveniently represented by single numbers in the complex domain:

$$Z_a = e^{jX_a}, \quad j = \sqrt{-1}, \quad a = 1, \dots, n. \quad (27)$$

In this new representation, the neighbour estimate has a multiplicative instead of an additive adjustment:

$$\text{outgoing: } \overrightarrow{Z}_{a,b} = e^{j\overrightarrow{X}_{a,b}} = e^{j(X_b + O_{a,b})} = Z_b e^{jO_{a,b}}, \quad (28)$$

$$\text{incoming: } \overleftarrow{Z}_{a,b} = e^{j\overleftarrow{X}_{a,b}} = e^{j(X_b - O_{b,a})} = Z_b e^{-jO_{b,a}}. \quad (29)$$

The neighbourhood average in the complex plane is

$$\text{average: } \overline{Z}_a = \sum_{b=1}^n (\overrightarrow{C}_{a,b} \overrightarrow{Z}_{a,b} + \overleftarrow{C}_{a,b} \overleftarrow{Z}_{a,b}). \quad (30)$$

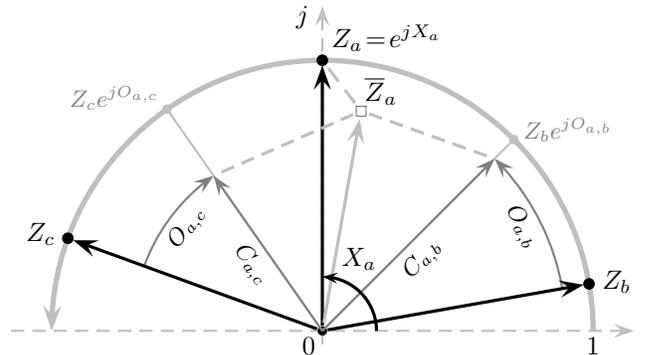


Fig. 4. AE places elements on the unit circle such that their angular displacements fulfill local ordering measurements. It seeks to minimize the total (squared) distance between position Z_a and its neighbourhood average \overline{Z}_a from all as .

Similarly to locally linear embedding (LLE) in the real domain [34], AE minimizes the L_2 distance in the complex domain between the embedding Z_a and its neighbourhood average \bar{Z}_a from all the a 's, weighted by its total confidence $D_{a,a}$:

$$\varepsilon_{AE}(Z) = \sum_{a=1}^n D_{a,a} |Z_a - \bar{Z}_a|^2 \quad (31)$$

$$= \sum_{a=1}^n D_{a,a} \left| Z_a - \sum_{b=1}^n Z_b (\vec{C}_{a,b} e^{jO_{a,b}} + \overleftarrow{C}_{a,b} e^{-jO_{b,a}}) \right|^2. \quad (32)$$

Yet, unlike LLE or the original proposal of AE [17], where only the outgoing neighbour estimate $\vec{Z}_{a,b}$ is used to define the neighbourhood average, here we also use the incoming estimate $\overleftarrow{Z}_{a,b}$ to define the average \bar{Z}_a . \bar{Z}_a is identical to that in the original AE only when $(O', C') = (-O, C)$. This modification renders AE symmetrical with respect to \vec{X} and \overleftarrow{X} , just like the standard LS.

We relax the n unit-length constraints in Eqn. 27 to a single constraint $Z' D Z = 1'_n D 1_n$, and rewrite $\varepsilon_{AE}(Z)$ in terms of the real degree matrix D , complex measurement M , and complex graph Laplacian L :

$$\text{measurement: } M = C \bullet e^{jO} + (C \bullet e^{jO})', \quad (33)$$

$$\text{Laplacian: } L = D - M, \quad (34)$$

$$\text{average: } \bar{Z} = D^{-1} M Z, \quad (35)$$

$$\begin{aligned} \text{minimize: } \varepsilon_{AE}(Z) &= (Z - \bar{Z})' D (Z - \bar{Z}) \\ &= Z' (L' D^{-1} L) Z, \end{aligned} \quad (36)$$

$$\text{subject to: } Z' D Z = 1'_n D 1_n. \quad (37)$$

Note that e^A denotes entry-wise exponentiation of A , i.e., $(e^A)_{a,b} = e^{A_{a,b}}$. While the Laplacian for LS depends only on confidence C , the Laplacian for AE depends also on ordering O and, in fact, on the entire measurement M .

As with the standard LS criterion, we can rewrite the AE criterion using the normalized Laplacian \tilde{L} :

$$\text{minimize: } \varepsilon_{AE}(\tilde{Z}) = \tilde{Z}' \tilde{L}^2 \tilde{Z} \quad (38)$$

$$\text{subject to: } \tilde{Z}' \tilde{Z} = 1'_n D 1_n \quad (39)$$

$$\text{scaled embedding: } \tilde{Z} = D^{\frac{1}{2}} Z \quad (40)$$

$$\text{normalized Laplacian: } \tilde{L} = I - D^{-\frac{1}{2}} M D^{-\frac{1}{2}} \quad (41)$$

$$\text{normalized measurement: } \tilde{M} = D^{-\frac{1}{2}} M \quad (42)$$

As a result of standard Rayleigh quotient optimization over \tilde{L}^2 , the optimum \tilde{Z}^* is the eigenvector of \tilde{L}^2 with the smallest eigenvalue. Since $\tilde{L}V = \lambda V \Rightarrow \tilde{L}^2 V = \lambda^2 V$, it is also the eigenvector of \tilde{L} with the eigenvalue of the smallest magnitude. AE optimum X_{AE}^* is thus encoded in the angles of V_1 (Eqn. 27):

$$\text{optimum: } X_{AE}^* = \angle Z^* = \angle D^{\frac{1}{2}} Z^* = \angle \tilde{Z}^* = \angle V_1, \quad (43)$$

$$\text{where } \tilde{Z}^* = V_1, \quad \varepsilon_{AE}^* = \lambda_1^2 (1'_n D 1_n), \quad (44)$$

$$\tilde{L} V_k = \lambda_k V_k, \quad |\lambda_1| \leq \dots \leq |\lambda_n|. \quad (45)$$

2.4 LS Variant: Minimize Distance

We consider an LS variant (Fig. 5) which adopts LS' representation of X in the linear space and AE's criterion of distance minimization, now between X_a and its neighbourhood average \bar{X}_a (Eqn. 5):

$$\varepsilon_{LSD}(X) = \sum_{a=1}^n D_{a,a} (X_a - \bar{X}_a)^2 \quad (46)$$

$$= \sum_{a=1}^n D_{a,a} \left(X_a - \sum_{b=1}^n (\vec{C}_{a,b} \vec{X}_{a,b} + \overleftarrow{C}_{a,b} \overleftarrow{X}_{a,b}) \right)^2. \quad (47)$$

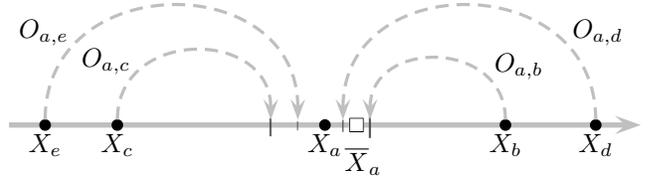


Fig. 5. The LS variant seeks to minimize the distance between X_a and its neighbourhood average \bar{X}_a (\square) for all a s, employing the same criterion as AE but in the linear embedding space for X .

We write ε_{LSD} using the Laplacian L and the measurement M for the standard LS criterion:

$$\text{Laplacian: } L = D - (C + C'), \quad (48)$$

$$\text{measurement: } M = C \bullet O - (C \bullet O)', \quad (49)$$

$$\begin{aligned} \text{average: } \bar{X} &= D^{-1} (C + C') X \\ &\quad + D^{-1} (C \bullet O - C' \bullet O') 1_n \\ &= X - D^{-1} (LX - M 1_n), \end{aligned} \quad (50)$$

$$\begin{aligned} \text{minimize: } \varepsilon_{LSD}(X) &= (X - \bar{X})' D (X - \bar{X}) \\ &= (LX - M 1_n)' D^{-1} (LX - M 1_n). \end{aligned} \quad (51)$$

Further simplifying with LS' normalized Laplacian $\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, normalized measurement $\tilde{M} = D^{-\frac{1}{2}} M$, and scaled embedding $\tilde{X} = D^{\frac{1}{2}} X$, we have

$$\text{minimize: } \varepsilon_{LSD}(\tilde{X}) = (\tilde{L} \tilde{X} - \tilde{M} 1_n)' (\tilde{L} \tilde{X} - \tilde{M} 1_n), \quad (52)$$

$$\text{optimum: } \tilde{X}_{LSD}^* = \tilde{L}^{-1} \tilde{M} 1_n. \quad (53)$$

Despite the difference in the criterion, the optimum for the LS variant is identical to that for the standard LS.

2.5 AE Variant: Minimize Variance

Conversely, we consider an AE variant (Fig. 6) which adopts AE's representation of X in the angular space of embedding Z and LS' criterion of variance minimization, now of the neighbour estimate of the embedding Z :

$$\varepsilon_{AEV}(Z) = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} |Z_a - Z_b e^{jO_{a,b}}|^2. \quad (54)$$

Since $|Z_b - Z_a e^{jO_{b,a}}| = |Z_b e^{-jO_{b,a}} - Z_a| = |Z_a - \overleftarrow{Z}_{a,b}|$,

$$\begin{aligned} 2\varepsilon_{AEV} &= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} |Z_a - Z_b e^{jO_{a,b}}|^2 + C_{b,a} |Z_b - Z_a e^{jO_{b,a}}|^2 \\ &= \sum_{a=1}^n D_{a,a} \sum_{b=1}^n \vec{C}_{a,b} |Z_a - \vec{Z}_{a,b}|^2 + \overleftarrow{C}_{a,b} |Z_a - \overleftarrow{Z}_{a,b}|^2. \end{aligned} \quad (55)$$

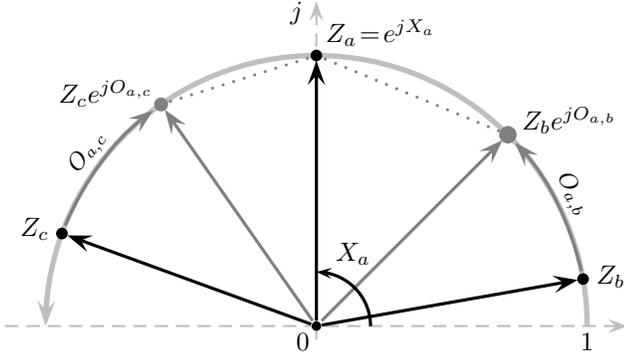


Fig. 6. The AE variant seeks to minimize the variance in the neighbour estimate $Z_b e^{jO_{a,b}}$ of Z_a for all a s, using the same criterion as LS but in the angular embedding space for X .

Just like the standard LS criterion, the AE variant is symmetrical with respect to $\vec{Z}_{a,b}$ and $\vec{Z}_{a,b}$, the neighbour estimate of Z_a in both outgoing and incoming directions.

However, unlike the standard LS criterion, the AE variant is a quadratic of Z without any linear terms:

$$\begin{aligned} \varepsilon_{AEV}(Z) &= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} |Z_a - Z_b e^{jO_{a,b}}|^2 \\ &= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (|Z_a|^2 + |Z_b|^2 - Z_a' Z_b e^{jO_{a,b}} - Z_a Z_b' e^{-jO_{a,b}}) \\ &= Z'(D - (C \bullet e^{jO} + C' \bullet e^{-jO}))Z. \end{aligned} \quad (56)$$

We write the AE variant as a complete quadratic form of L , subject to the same single relaxed norm constraint:

$$\text{measurement: } M = C \bullet e^{jO} + (C \bullet e^{jO})', \quad (57)$$

$$\text{Laplacian: } L = D - M, \quad (58)$$

$$\text{minimize: } \varepsilon_{AEV} = Z' LZ, \quad (59)$$

$$\text{subject to: } Z' D Z = 1_n' D 1_n. \quad (60)$$

Further simplifying with AE's normalized Laplacian $\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, normalized measurement $\tilde{M} = D^{-\frac{1}{2}} M$, and scaled embedding $\tilde{Z} = D^{\frac{1}{2}} Z$, we have

$$\text{minimize: } \varepsilon_{AEV}(\tilde{Z}) = \tilde{Z}' \tilde{L} \tilde{Z}, \quad (61)$$

$$\text{subject to: } \tilde{Z}' \tilde{Z} = 1_n' D 1_n. \quad (62)$$

We obtain a near-global optimum X_{AEV}^* in the angles of the eigenvector of \tilde{L} with the smallest eigenvalue:

$$\text{optimum: } X_{AEV}^* = \angle Z^* = \angle D^{\frac{1}{2}} Z^* = \angle \tilde{Z}^* = \angle V_1, \quad (63)$$

$$\text{where } \tilde{Z}^* = V_1, \quad \varepsilon_{AEV}^* = \lambda_1(1_n' D 1_n), \quad (64)$$

$$\tilde{L} V_k = \lambda_k V_k, \quad \lambda_1 \leq \dots \leq \lambda_n. \quad (65)$$

3 LS-AE DISTINCTIONS AND CONNECTIONS

Fig. 7 summarizes the four embedding criteria we have investigated, along with new results to be proven in this section. There are a few commonalities:

1. They all use the same degree matrix D , which has the total level of confidence for each element.

2. They all can be written compactly as a quadratic form of scaled embedding, normalized Laplacian, and normalized measurement matrices.
3. For both LS and AE, minimizing variance or minimizing distance results in the same optimum.

There are also several differences:

1. Whereas the Laplacian and the measurement are respectively symmetrical and skew-symmetrical for LS, they are always complex Hermitian for AE.
2. While the measurement M is defined by $C \bullet O$ for LS, it is $C \bullet e^{jO}$ for AE. For LS, the measurement is the same whether it is a large value with a small confidence or a small value with a large confidence.
3. While the Laplacian for LS is exclusively determined by confidence C , it also involves size O for AE.
4. While the LS optimum requires a linear solver for fitting $\tilde{M} 1_n$ in the column space of \tilde{L} , the AE optimum needs the smallest eigenvector of \tilde{L} only.
5. Minimizing variance or distance only changes the kernel for the quadratic objective function. Compared to the distance criterion, the variance criterion introduces an extra \tilde{L}^{-1} for LS and \tilde{L} for AE.

We first prove the above claims regarding the distance and variance criteria, then offer a diffusion interpretation to both LS and AE optima, and finally establish connections between LS' and AE's error functions.

3.1 Equivalent Variance and Distance Criteria

The LS variance and distance criteria are equivalent, since they have identical optima: $X_{LS}^* = \tilde{L}^{-1} \tilde{M} 1_n$. Their connections can be further clarified when both are written as complete quadratic forms of \tilde{X} , \tilde{L} , and \tilde{M} .

The LS variant is already a complete quadratic form. We accordingly complete the square for the standard LS by adding a constant $1_n' \tilde{M}' \tilde{L}^{-1} \tilde{M} 1_n$ (different from the earlier dropped constant $1_n' (C \bullet O \bullet O) 1_n$ in Eqn. 16):

$$\begin{aligned} \varepsilon_{LS} &= \tilde{X}' \tilde{L} \tilde{X} - 2 \tilde{X}' \tilde{M} 1_n \\ &\Leftrightarrow (\tilde{X}' \tilde{L}) \tilde{L}^{-1} (\tilde{L} \tilde{X}) - 2 (\tilde{L} \tilde{X})' \tilde{L}^{-1} (\tilde{M} 1_n) \\ &\quad + 1_n' \tilde{M}' \tilde{L}^{-1} \tilde{M} 1_n \quad (\text{constant added}) \quad (66) \\ &= (\tilde{L} \tilde{X} - \tilde{M} 1_n)' \tilde{L}^{-1} (\tilde{L} \tilde{X} - \tilde{M} 1_n). \quad (67) \end{aligned}$$

That is, the two LS criteria only differ in the quadratic kernel: identity matrix I for the LS variant, and diffusion matrix \tilde{L}^{-1} for the LS standard, both acting on $\tilde{L} \tilde{X} - \tilde{M} 1_n$, the error between the embedding differences and the neighbourhood averages of local ordering measurements.

The AE variance and distance criteria can be written as quadratic forms of \tilde{Z} , with kernels \tilde{L}^2 and \tilde{L} respectively. The optimum is the smallest eigenvector of \tilde{L} , in terms of the absolute size ($\min_{k=1}^n |\lambda_k|$) for the former and the algebraic size ($\min_{k=1}^n \lambda_k$) for the latter. They are identical when all the eigenvalues are nonnegative. We prove that in fact \tilde{L} 's eigenvalues are bounded between 0 and 2.

problem:	(local ordering O , confidence C) \Rightarrow global ordering X	
approach:	Least Squares Embedding (LS)	Angular Embedding (AE)
representation:	linear space	angular space
embedding:	$Y = X$	$Y = Z = e^{jX}$
measurement:	$M = (C \bullet O) - (C \bullet O)'$	$M = (C \bullet e^{jO}) + (C \bullet e^{jO})'$
Laplacian:	$L = D - (C + C)'$	$L = D - M$
common:		
degree:	$D = \text{Diag}((C + C')1_n)$	
scaled embedding:	$\tilde{Y} = D^{\frac{1}{2}}Y$	
normalized measurement:	$\tilde{M} = D^{-\frac{1}{2}}M$	
normalized Laplacian:	$\tilde{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$	
pairwise transition:	$P = I - D^{-1}L$	
criterion:		
minimize variance:	$\varepsilon_{LS} = (\tilde{L}\tilde{Y} - \tilde{M}1_n)' \tilde{L}^{-1} (\tilde{L}\tilde{Y} - \tilde{M}1_n)$	$\varepsilon_{AEV} = \tilde{Y}' \tilde{L}^2 \tilde{Y}$
minimize distance:	$\varepsilon_{LSD} = (\tilde{L}\tilde{Y} - \tilde{M}1_n)' (\tilde{L}\tilde{Y} - \tilde{M}1_n)$	$\varepsilon_{AE} = \tilde{Y}' \tilde{L} \tilde{Y}$
optimum:		
scaled embedding:	$\tilde{Y}^* = \tilde{L}^{-1} \tilde{M} 1_n$	$\tilde{Y}^* = V_1, \quad \tilde{L}V_k = \lambda_k V_k, \quad \lambda_1 \leq \dots \leq \lambda_n$
global ordering:	$X_{LS}^* = D^{-\frac{1}{2}} \tilde{Y}^*$	$X_{AE}^* = \angle Y^* = \angle D^{-\frac{1}{2}} \tilde{Y}^* = \angle \tilde{Y}^*$
diffusion:		
initial solution:	$Y^{(0)} = X^{(0)} = D^{-1}M1_n$	$Y^{(0)} = e^{jX^{(0)}} \text{ or any}$
recursion:	$Y^{(k)} = PY^{(k-1)} + Y^{(0)}$	$Y^{(k)} \propto PY^{(k-1)} + Y^{(k-1)}$
convergence:	$X_{LS}^* = (I - P)^{-1}Y^{(0)}$	$X_{AE}^* = \angle (I + P)^\infty Y^{(0)}$
sensitivity:		
error function:	$\rho_{LS}(\delta) = \delta^2 \times \frac{1}{\pi^2}$	$\rho_{AE}(\delta) = \sin^2 \frac{\delta}{2}$
derivative:	$\rho'_{LS}(\delta) = 2\delta \times \frac{1}{\pi^2}$	$\rho'_{AE}(\delta) = \sin \delta \times \frac{1}{2}$
weight function:	$\frac{\rho'_{LS}(\delta)}{\delta} = 2 \times \frac{1}{\pi^2}$	$\frac{\rho'_{AE}(\delta)}{\delta} = \frac{\sin \delta}{\delta} \times \frac{1}{2}$

Fig. 7. Summary of LS and AE criteria. The input of embedding is arbitrary pairwise local ordering measurements in value $O_{n \times n}$ and confidence $C_{n \times n}$ (nonnegative). The output of embedding is global ordering $X_{n \times 1}$. LS and AE have distinctive representations of the embedding Y , the measurement matrix M , and the Laplacian matrix L . They share the same degree matrix D , as well as the definitions of scaled embedding \tilde{Y} , normalized measurement \tilde{M} , normalized Laplacian \tilde{L} , and pairwise transition P from their own Y , M , and L . For LS or AE, the variance and distance optimality criteria lead to identical optima for quadratic forms with different quadratic kernels. The LS and AE optima can be interpreted, respectively, as passive and active diffusion from the same initial solution. Their distinction in the robustness to outliers is revealed by LS' constant and AE's sinc error weight functions.

Theorem 1. For any nonnegative matrix C and arbitrary matrix O , let $M = C \bullet e^{jO} + (C \bullet e^{jO})'$, $D = \text{Diag}((C + C')1_n)$, $\tilde{L} = I - D^{-\frac{1}{2}}MD^{-\frac{1}{2}}$. For any eigenvalue λ of L , there must be $0 \leq \lambda(\tilde{L}) \leq 2$, and \tilde{L} is semi-positive definite.

Proof: Since C is nonnegative, D must be semi-positive definite, i.e. $Z'DZ \geq 0, \forall Z$. Let $Z_a = r_a e^{jX_a}$, $Z_b = r_b e^{jX_b}$, $r_a, r_b \geq 0$, $E_{a,b} = X_a - X_b - O_{a,b}$. We have

$$\begin{aligned} & Z'(D - M)Z \\ &= Z' \left(\text{Diag}((C + C')1_n) - (C \bullet e^{jO} + C' \bullet e^{-jO'}) \right) Z \\ &= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (r_a^2 + r_b^2 - r_a r_b (e^{-jE_{a,b}} + e^{jE_{a,b}})) \quad (68) \end{aligned}$$

$$= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (r_a^2 + r_b^2 - 2r_a r_b \cos(E_{a,b})) \quad (69)$$

$$\geq \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (r_a^2 + r_b^2 - 2r_a r_b), \quad \text{since } r_a, r_b \geq 0, \quad (70)$$

$$= \sum_{a=1}^n \sum_{b=1}^n C_{a,b} (r_a - r_b)^2 \geq 0, \quad \forall Z, \quad (71)$$

$$\Rightarrow \lambda(\tilde{L}) \geq \min_Z \frac{Z' \tilde{L} Z}{Z' Z} = \min_Z \frac{Z'(D - M)Z}{Z' D Z} \geq 0. \quad (72)$$

Since $-2r_a r_b \cos(E_{a,b}) \leq 2r_a r_b \leq r_a^2 + r_b^2$, we also have

$$Z'(D - M)Z \leq \sum_{a=1}^n \sum_{b=1}^n C_{a,b} 2(r_a^2 + r_b^2) = 2Z' D Z \quad (73)$$

$$\Rightarrow \lambda(\tilde{L}) \leq \max_Z \frac{Z' \tilde{L} Z}{Z' Z} = \max_Z \frac{Z'(D - M)Z}{Z' D Z} \leq 2. \quad (74)$$

Thus, $0 \leq \lambda(\tilde{L}) \leq 2$, and \tilde{L} is semi-positive definite. \square

Therefore, whether LS or AE is formulated in the LS' standard variance criterion or AE's standard distance criterion makes no difference in the optimal solution.

3.2 Passive vs. Active Diffusion

For any graph, we can write the symmetrically normalized Laplacian \tilde{L} in terms of the asymmetrically normalized Laplacian $D^{-1}L$, or transition matrix P :

$$\tilde{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = D^{\frac{1}{2}}(I - P)D^{-\frac{1}{2}} = I - D^{\frac{1}{2}}PD^{-\frac{1}{2}}, \quad (75)$$

$$\text{transition: } P = I - D^{-1}L. \quad (76)$$

For LS, since $L = D - (C + C')$, P is determined by C exclusively. It is nonnegative and row-normalized:

$$\text{transition for LS: } P = D^{-1}(C + C'), \quad (77)$$

$$\text{row normalized: } P1_n = 1_n. \quad (78)$$

P can be interpreted as a transition probability matrix. The larger the combined confidence of $C_{a,b} + C_{b,a}$, the more likely the transition from a to b . What PX does to X is simple averaging or smoothing, with X_a replaced by its neighbourhood average $(PX)_a$.

For AE, since $L = D - M$, P also depends on O :

$$\text{transition for AE: } P = D^{-1}(C \bullet e^{jO} + C' \bullet e^{-jO'}). \quad (79)$$

P is row-normalized in the magnitudes of individual complex components. It can thus still be interpreted as a probability transition matrix, with a diffusion process in the magnitude space just like LS and a new jump process in the phase space. What PZ does to $Z = e^{jX}$ is *jump averaging*: While Z_a is also replaced by its neighbourhood average $(PZ)_a$, Z_b is first adjusted in phase by $O_{a,b}$ or $O_{b,a}$, and then averaged with others in the complex plane according to $\vec{C}_{a,b}$ or $\vec{C}_{b,a}$.

If O comes from pairwise differences of some underlying global ordering X , we prove that the transition matrices for LS and AE have the same spectrum, and their eigenvectors are related by e^{jX} , the angular embedding of the global ordering X in the complex domain.

Theorem 2. Assume $O = X1'_n - 1_nX'$ and C is nonnegative. Let $D = \text{Diag}((C + C')1_n)$, $P_{LS} = D^{-1}(C + C')$, and $P_{AE} = D^{-1}((C \bullet e^{jO}) + (C' \bullet e^{jO}'))$.

1. P_{LS} and P_{AE} have the same set of eigenvalues, which are all real and bounded between -1 and 1.
2. Their eigenvectors for the same eigenvalue λ are related as such: V for P_{LS} and $\text{Diag}(e^{jX})V$ for P_{AE} .
3. In particular, the largest eigenvalue-eigenvector pair is $(1, 1_n)$ for P_{LS} , and $(1, e^{jX})$ for P_{AE} .

Proof: For $O = X1'_n - 1_nX'$, we have: $O' = -O$, and

$$P_{AE} = D^{-1}((C \bullet e^{jO}) + (C' \bullet e^{jO'})) \quad (80)$$

$$= D^{-1}(C + C') \bullet e^{jO} \quad (81)$$

$$= P_{LS} \bullet e^{jO} = P_{LS} \bullet e^{jX1'_n} \bullet e^{-j1_nX'} \quad (82)$$

$$= \text{Diag}(e^{jX})P_{LS}\text{Diag}(e^{-jX}). \quad (83)$$

If $P_{LS}V = \lambda V$, then $P_{AE}(\text{Diag}(e^{jX})V) = \lambda(\text{Diag}(e^{jX})V)$. Likewise, P_{LS} shares the same spectrum as

$$D^{\frac{1}{2}}P_{LS}D^{-\frac{1}{2}} = D^{-\frac{1}{2}}(C + C')D^{-\frac{1}{2}}, \quad (84)$$

a symmetrical matrix with all real eigenvalues. Taking together with the well-known fact that the spectral radius of a transition matrix is 1, we have shown that both P_{LS} and P_{AE} have the same and all real eigenvalues bounded between -1 and 1. Their corresponding eigenvectors are related as $(V, \text{Diag}(e^{jX})V)$ for (P_{LS}, P_{AE}) respectively.

Since $P_{LS}1_n = 1_n$, 1_n is the eigenvector with the largest eigenvalue 1. P_{AE} thus also has the eigenvector of $\text{Diag}(e^{jX})1_n = e^{jX}$ with the largest eigenvalue 1. \square

P_{LS} depends just on C ; thus, its eigenvector can only encode confidence in a global ordering. P_{AE} depends on both O and C ; thus, its eigenvector can encode the size and confidence of an optimal global ordering separately in the angles and magnitudes of the same complex numbers. For (O, C) consistent with global ordering X , the above theorem shows that P_{AE} and P_{LS} 's corresponding eigenvectors reveal the same level of confidence in X .

This result generalizes the concept of transition matrix, useful for understanding the geometry of the data through diffusion [35]–[39], from the real domain to the complex domain. We show next that the LS (AE) optimum can be understood as the result of a passive (active) diffusion of local orderings.

The optimal LS embedding can be viewed as the result of an infinite diffusion of $X^{(0)}$, the confidence-weighted average measurement of each neighbourhood:

$$X^{(0)} = D^{-1}M1_n = D^{-\frac{1}{2}}\tilde{M}1_n, \quad (85)$$

$$X^{(k)} = X^{(0)} + PX^{(k-1)}, \quad k = 1, 2, \dots, \quad (86)$$

$$X_{LS}^* = D^{-\frac{1}{2}}\tilde{X}_{LS}^* = D^{-\frac{1}{2}}\tilde{L}^{-1}\tilde{M}1_n \quad (87)$$

$$= (I - P)^{-1}X^{(0)} = (I + P + P^2 + \dots)X^{(0)} \quad (88)$$

$$= \sum_{t=0}^{\infty} P^t X^{(0)} = \lim_{k \rightarrow \infty} X^{(k)} = X^{(\infty)}. \quad (89)$$

X_{LS}^* is thus an aggregation of $X^{(0)}$, a direct measurement on each element's overall superiority against neighbours, $PX^{(0)}$, the one-step propagation of neighbours' assessment of its superiority, $P^2X^{(0)}$, the two-step propagation of its neighbours' assessment of its superiority, and so on and so forth. How many steps the aggregation needs to converge depends on the transition matrix P , which is determined entirely by confidence C .

The optimal AE embedding is $\angle V_1$, where V_1 is the eigenvector of \tilde{L} with the smallest eigenvalue λ_1 . Since \tilde{L} 's eigenvalues are bounded between 0 and 2, V_1 must be the eigenvector of $2I - \tilde{L}$ with the largest eigenvalue $2 - \lambda_1$. V_1 can thus be solved via power iteration:

$$V_1 = \lim_{k \rightarrow \infty} V_1^{(k)} = V_1^{(\infty)}, \quad (90)$$

$$V_1^{(k)} \propto (2I - \tilde{L})V_1^{(k-1)}, \quad \text{s.t. } \|V_1^{(k)}\| = 1, \quad (91)$$

$$= (I + D^{\frac{1}{2}}PD^{-\frac{1}{2}})V_1^{(k-1)}. \quad (92)$$

This recursion can be simplified with $Z^{(k)} = D^{-\frac{1}{2}}V_1^{(k)}$. We obtain an interpretation comparable to that for LS:

$$Z^{(0)} = e^{jX^{(0)}}, \quad (93)$$

$$Z^{(k)} \propto (I + P)Z^{(k-1)} \propto (I + P)^k Z^{(0)}, \quad k > 0, \quad (94)$$

$$X_{AE}^* = \angle D^{-\frac{1}{2}}V_1 = \angle Z^{(\infty)}. \quad (95)$$

Unlike $X^{(0)}$ for LS, $Z^{(0)}$ can be any non-degenerate initial solution. The particular choice for $Z^{(0)}$ in the above satisfies $\angle Z^{(0)} = X^{(0)}$, which allows us to examine how the AE and LS optima emerge differently from $X^{(0)}$.

Therefore, both LS and AE optima can be regarded as some global ordering equilibrium reached upon diffusion of average local ordering $X^{(0)}$:

$$X_{LS}^* = (I - P)^{-1}X^{(0)}, \quad P = D^{-1}(C + C'), \quad (96)$$

$$X_{AE}^* = \angle (I + P)^\infty e^{jX^{(0)}}, \quad P = D^{-1}(C \bullet e^{jO} + C' \bullet e^{-jO'}). \quad (97)$$

or expressed in the following recursive iterations:

$$X_{LS}^* \leftarrow X^{(k)} = P_{LS}X^{(k-1)} + X^{(0)}, \quad k \rightarrow \infty, \quad (98)$$

$$Z_{AE}^* \leftarrow Z^{(k)} \propto P_{AE}Z^{(k-1)} + Z^{(k-1)}, \quad k \rightarrow \infty. \quad (99)$$

The LS optimum can be understood as an aggregation of progressively diminishing smoothed versions of $X^{(0)}$ via passive diffusion. In contrast, the AE optimum can be understood as an aggregation of progressively stabilized jump averages of any non-trivial initial global ordering via active diffusion, with a jump mechanism in the phase space exerting more inertia on the diffusion process, making the convergence steadier and smoother.

3.3 Sensitivity to Errors and Outliers

AE in its original distance criterion appears to be quite different from LS. Since AE can be equivalently cast in LS' variance criterion, we can reveal their connections. According to the variance interpretation of LS and AE,

$$\varepsilon_{LS} = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} E_{a,b}^2, \quad (100)$$

$$\frac{1}{4}\varepsilon_{AEV} = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} \sin^2 \frac{E_{a,b}}{2}, \quad (101)$$

$$\begin{aligned} \because |Z_a - Z_b e^{jO_{a,b}}|^2 &= |e^{jX_a} - e^{jX_b} e^{jO_{a,b}}|^2 = |e^{jE_{a,b}} - 1|^2 \\ &= (\cos E_{a,b} - 1)^2 + \sin^2 E_{a,b} = 4 \sin^2 \frac{E_{a,b}}{2}. \end{aligned} \quad (102)$$

We can thus write both as a weighted squared sum of errors (SSE), with their own error function $\rho(\delta)$:

$$\text{SSE:} \quad \varepsilon(X) = \sum_{a=1}^n \sum_{b=1}^n C_{a,b} \cdot \rho(E_{a,b}), \quad (103)$$

$$\text{error functions:} \quad \rho_{LS}(\delta) = \delta^2 \cdot \frac{1}{\pi^2}, \quad (104)$$

$$\rho_{AE}(\delta) = \sin^2 \frac{\delta}{2}. \quad (105)$$

Here we have scaled the LS and AE criteria in Eqns.100-101 so that their error functions have the same range of $[0, 1]$ over $\delta \in [-\pi, \pi]$, the central period of ρ_{AE} (Fig. 8a).

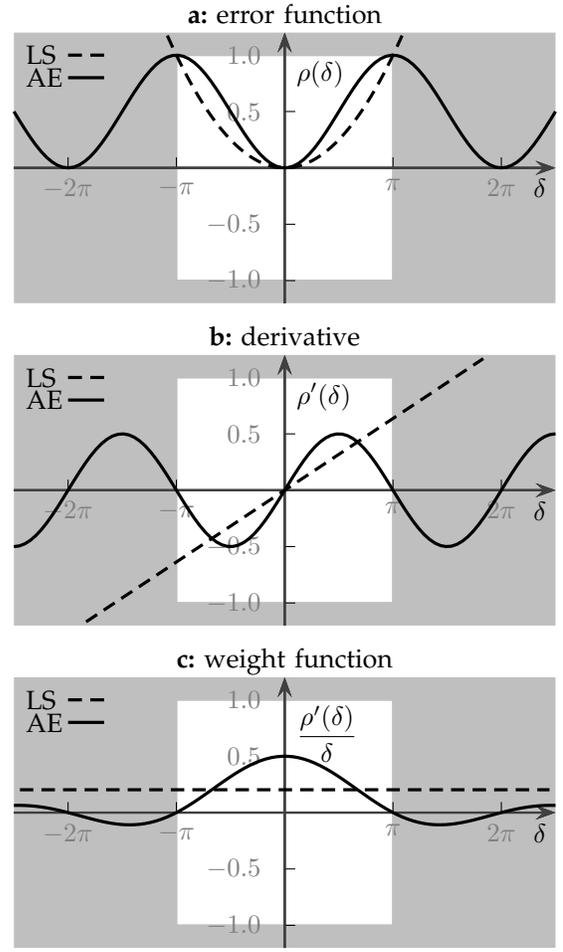


Fig. 8. AE is more sensitive to small errors yet more robust to large errors than LS. a) Since AE's error function is bounded and cyclic, we scale LS' error function to span the same range within the highlighted white box $(-\pi, \pi) \times (0, 1)$. b) While ρ'_{LS} increases with δ at a linear rate, ρ'_{AE} peaks at $\frac{\pi}{2}$. c) The weight function is constant (0.2) for LS, regardless of δ , and is a sinc function for AE that peaks (0.5) at 0 and vanishes at π , repeating the same pattern with a period of 2π and a decaying magnitude.

While LS and AE have different error functions, they become equivalent at very small δ , since

$$\rho_{AE}(\delta) = \sin^2 \frac{\delta}{2} = \left(\frac{\delta}{2} - \frac{\delta^3}{3!2^3} + \frac{\delta^5}{5!2^5} - \dots \right)^2 \quad (106)$$

$$\approx \frac{\delta^2}{4} \propto \frac{\delta^2}{\pi^2} = \rho_{LS}(\delta), \quad |\delta| \rightarrow 0, \quad (107)$$

i.e., LS and AE behave alike when the local ordering O slightly deviates from some true global ordering X .

The sensitivity of LS and AE to the embedding error δ is revealed in the iterative re-weighted LS procedure for solving the SSE optimum X^* [40]:

$$0 = \frac{\partial \varepsilon}{\partial X_a} = \sum_{b=1}^n C_{a,b} \rho'(E_{a,b}) - C_{b,a} \rho'(E_{b,a}) = \sum_{b=1}^n \left(C_{a,b} \frac{\rho'(E_{a,b})}{E_{a,b}} \right) \cdot E_{a,b} - \left(C_{b,a} \frac{\rho'(E_{b,a})}{E_{b,a}} \right) \cdot E_{b,a}, \quad (108)$$

$$\sum_{b=1}^n W_{a,b}(X_a^* - X_b^* - O_{a,b}) - W_{b,a}(X_b^* - X_a^* - O_{b,a}) = 0, \quad (109)$$

$$X^* = (\text{Diag}((W + W')1_n) - W - W')^{-1} \cdot (W \bullet O - W' \bullet O')1_n, \quad (110)$$

$$\text{where } W_{a,b} = C_{a,b} \frac{\rho'(E_{a,b})}{E_{a,b}} = C_{a,b} \cdot w(E_{a,b}), \quad (111)$$

$$\text{weight function: } w(\delta) = \frac{\rho'(\delta)}{\delta}, \quad (112)$$

$$w_{LS}(\delta) = \frac{\rho'_{LS}(\delta)}{\delta} = 2 \cdot \frac{1}{\pi^2}, \quad (113)$$

$$w_{AE}(\delta) = \frac{\rho'_{AE}(\delta)}{\delta} = \frac{\sin \delta}{2\delta}. \quad (114)$$

Eqn. 110 prescribes a series of weighted LS embeddings for solving the optimum, the weights evolving with the embedding error. Each iteration solves a standard LS embedding problem for (O, C_{eq}) : While the local ordering O is always the original, the equivalent confidence C_{eq} is the original confidence C multiplied by the weight function w at the current embedding error E :

$$\text{iterative optimum: } X^* = L_{eq}^{-1} M_{eq} 1_n, \quad (115)$$

$$\text{equivalent confidence: } C_{eq} = W = C \bullet w(E), \quad (116)$$

$$\text{degree: } D_{eq} = \text{Diag}((C_{eq} + C'_{eq})1_n), \quad (117)$$

$$\text{measurement: } M_{eq} = C_{eq} \bullet O - (C_{eq} \bullet O)', \quad (118)$$

$$\text{Laplacian: } L_{eq} = D_{eq} - (C_{eq} + C'_{eq}). \quad (119)$$

For LS, $w_{LS}(E)$ is a constant; thus, $C_{eq} \propto C$, the optimum is solved in a single iteration. For AE, $w_{AE}(E)$ is a non-convex sinc function, with larger errors (e.g., outliers) given less weight (Fig. 8c); thus, more iterations are required for convergence, with local optimality guarantee only. This fact makes AE's eigensolution only more remarkable: The eigenvector is a near-global optimum solved in one step from the original (O, C) .

To summarize, what sets LS and AE apart is not their criteria but their representations. For both LS and AE, the variance and distance criteria lead to the same optima. AE represents the size and confidence of pairwise local orderings in single complex numbers, and encodes the size and confidence of an optimal global ordering in the eigenvector of its transition matrix. As a quadratic criterion in the complex domain, AE has a non-convex error function that is more sensitive to small errors yet at the same time more robust to large errors. The AE optimum can be viewed as the result of an iterative LS procedure where the confidence in local orderings gets adjusted according to the size of the embedding error.

4 EXPERIMENTAL RESULTS

We implement LS and AE in MATLAB (version: R2009b) on a PowerMac OS X with 2x3 Dual-Core CPU and 8GB memory. We compare LS and AE on accuracy (Fig. 9), efficiency (Fig. 10), and robustness (Fig. 11 and Fig. 12).

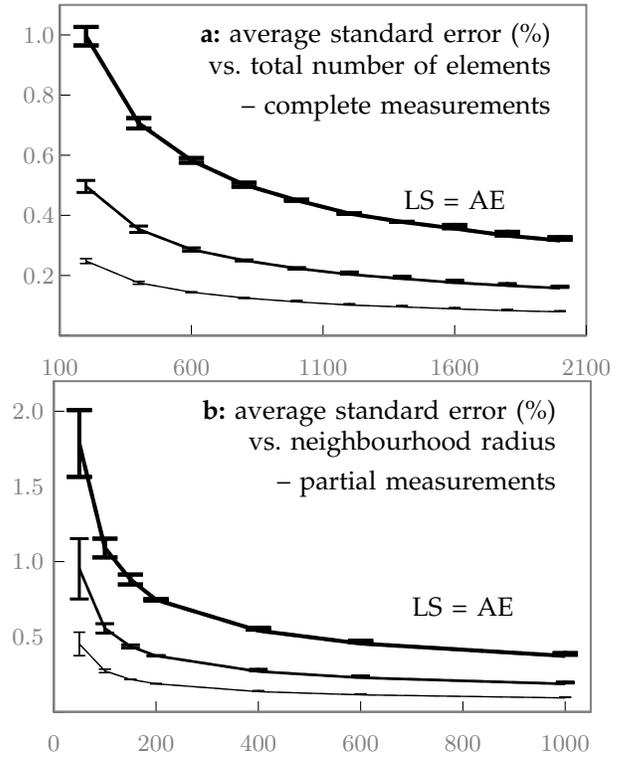


Fig. 9. LS and AE are equally accurate for Gaussian additive noise σ on complete or partial measurements. **a)** We generate ground-truth X as uniformly random numbers in $[-1, 1]$, obtain measurement O with confidence $C = 1$ by adding Gaussian noise to pairwise differences of X , and compute LS (gray) and AE (black) embeddings from (O, C) . We repeat this 20 times for each number of elements and noise level σ , and obtain the error bar plots for the overall difference between the embedding and X . Thicker lines for larger σ , $\sigma = 0.05, 0.1, 0.2$. LS and AE (curves overlap) always give the same optimum at each noise level, and the average standard error with respect to X increases with σ . **b)** For $n = 2000$ elements labeled from 1 to n , each has $2r+1$ neighbours (labeled from $k-r$ to $k+r$ for element k). We run the same experiment as in **a**, with $C(a, b) = 1$ if a, b are neighbours and 0 otherwise. LS and AE (curves overlap) converge quickly to that from complete measurements.

Given measurements (O, C) , we first compute the normalized Laplacian \tilde{L} for LS or AE. We then use MATLAB built-in function `lscov.m`, a Cholesky linear solver which is faster and more stable than pseudo-inverse, to obtain the optimum $\tilde{L}^{-1} \tilde{M} 1_n$ for LS, or use MATLAB function `eigs.m` to obtain the eigenvector V_1 of \tilde{L} with the smallest eigenvalue for AE. Since an ordering is subject to arbitrary translation, we center it at 0 by subtracting its mean value. With the arbitrary constants removed from all the embedding and ground-truth solutions, they can be directly compared.

We compare the accuracy of LS and AE upon Gaussian noise with respect to the number of elements and the number of measurements per element. Given all pairwise measurements, Fig. 9a shows that: **1)** The embedding becomes more accurate with more elements at any noise level; **2)** both LS and AE can achieve high accuracy in the presence of Gaussian noise, e.g., the standard error

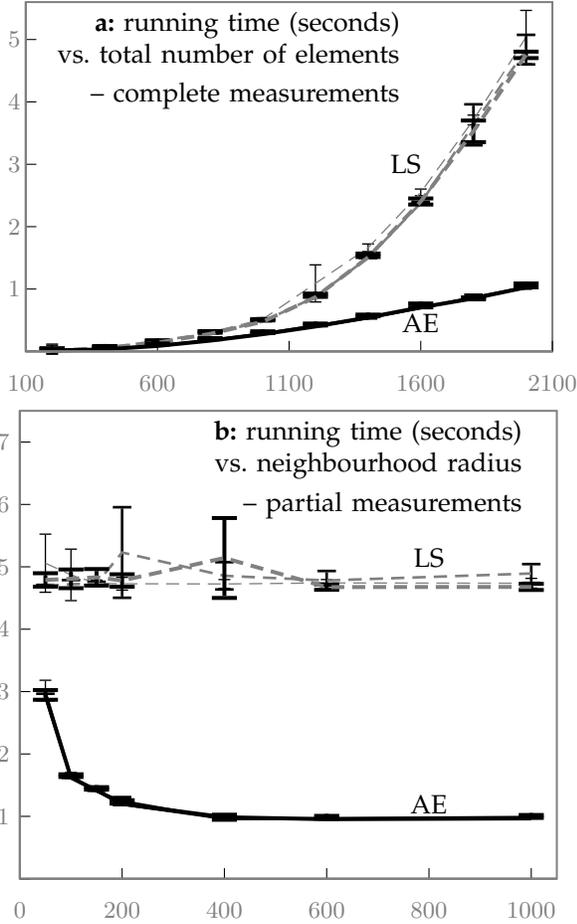


Fig. 10. Running times for Fig. 9. a) LS takes more time than AE with more elements, both varying little with σ (curves overlap) for complete measurements. b) LS takes the same amount of time while AE takes less time as the radius increases.

is less than 1% over signal range $[-1, 1]$ at noise level 0.2. Given measurements in limited neighbourhoods, Fig. 9b shows that: 1.) The accuracy is poor at a small radius, but becomes on par with that from complete measurements with only 20 ~ 40% of them, depending on the noise level; 2) the accuracy is similar for LS and AE at any neighbourhood radius and measurement noise level.

While LS and AE are equally accurate upon Gaussian noise, their running times are different. Fig. 10 shows that:

1. The level of measurement noise has little impact on the running time for either LS or LE;
2. While the time always increases with more elements, it does so more rapidly for LS than for AE;
3. The running time is insensitive to the radius for LS, but drops with an increasing radius for AE;
4. AE takes far less time than LS at a radius large enough to discover the same solution from complete measurements, i.e., AE is more effective at global integration from incomplete measurements.

It might appear counter-intuitive that AE’s running time actually decreases with an increasing radius. The



Fig. 11. Image reconstruction by AE, LS, bounded LS (LSB), L1, and bounded L1 (L1B) from pairwise intensity differences. a) Image X as the ground-truth ordering, with an intensity range of 1 over 180×160 pixels. b,c,d,e,f) Embedding results marked by their standard errors with respect to X and running times in seconds. Local ordering O is obtained with confidence 1 as intensity differences of X between pixels within a neighbourhood radius of 2, added with Gaussian noise of $\sigma = 0.05$. 10% of these measurements are further added with random noise of ± 3 (i.e. outliers). The AE optimum is far superior to others (despite the fact that the true range 1 is used in LSB and L1B). The time for AE (16s) is a little more than that for LS (6s) in this small radius case, but it is generally much less as the radius increases (Fig. 10, Fig. 12).

explanation is that while more matrix-vector operations are needed during each iteration of the eigensolver for a denser matrix \tilde{L} , the number of iterations is reduced as less cue propagation is required with a larger radius.

a: LS (0.059, 166s) — 20% outliers — **b:** AE (0.006, 18s)



c: LS (0.073, 166s) — 30% outliers — **d:** AE (0.010, 18s)



e: LS (0.083, 165s) — 40% outliers — **f:** AE (0.025, 19s)



Fig. 12. Image reconstruction by LS (a,c,e) and AE (b,d,f) over an increasing percentage of outliers. The same convention as Fig. 11. These results are obtained at $r = 8$. While robustness often increases with the neighbourhood radius, we observe a fairly stable break point for AE between 40% and 50%, the results showing more speckles with more outliers.

The efficiency of LS and AE certainly evolves with new linear solvers and eigensolvers. Since the Laplacian for LS is singular, the linear solver must handle rank deficient matrices. Many solvers such as conjugate gradient methods (e.g. MATLAB functions `pcg.m`, `bicg.m`) are fast but fail to converge for the ill-conditioned LS Laplacian. We use MATLAB 7.6's built-in Cholesky linear solver and Lanczos eigensolver. For general matrices, the current best computational complexity is $O(n \log^c n)$ for the linear-solver [41] and $O(n^{1.5})$ for the eigen-solver

[42]. The former becomes $O(n \log^2 n \log \log n)$ when the graph is planar, and the latter becomes $O(n)$ when the graph is sparse, as in most computer vision applications.

Having studied the accuracy and efficiency of LS and AE with independent and identically distributed Gaussian noise, we turn to the robustness of LS and AE to singularly large noise that could potentially disrupt the entire embedding, i.e. measurement outliers.

We also investigate whether AE's robustness can be achieved by bounding the LS solution [43] or using L_1 instead of L_2 norm [23]. Specifically, if $X_a \in [X_{\min}, X_{\max}]$, $\forall a$, we first optimize regarding $\tilde{X} = D^{\frac{1}{2}} X$:

$$\min \quad \|\tilde{L}\tilde{X} - \tilde{M}1_n\|_k, \quad k = 1 \text{ or } 2 \quad (120)$$

$$\text{s.t.} \quad D^{\frac{1}{2}}(X_{\min}1_n) \leq \tilde{X} \leq D^{\frac{1}{2}}(X_{\max}1_n) \quad (121)$$

and then recover the optimum $X^* = D^{-\frac{1}{2}} \tilde{X}^*$ and remove its mean as the final solution. For $k=2$, without and with the constraint, we have the old LS and the new bounded LS (LSB) optima respectively. Likewise, for $k=1$, without and with the constraint, we have the L1 and bounded L1 (L1B) optima respectively. We use MATLAB function `lsqlin.m` for LSB and `linprog.m` for L1 and L1B.

Fig. 11 and Fig. 12 show that AE can recover the original image from noisy and outlier infected local intensity differences with a quality and computational efficiency that cannot be matched by LS, L1, or their bounded versions. AE can remain unaffected by as many as 40% outliers, with more speckles showing up as the percentage of outliers further increases. It eventually breaks down completely with 50% outliers.

For AE, the quality of embedding remains high before the percentage of outliers reaches the break point. This effect cannot be achieved by employing L_1 norm on the error or bounding the range of the embedding, where outliers with their large magnitudes force oscillations in the outcome to accommodate them. AE removes outliers not by imposing smoothness or bounds on the solution, but by penalizing inconsistency between the measurements themselves: small inconsistencies are sensitively tuned to, while large inconsistencies are simply ignored.

5 SUMMARY AND DISCUSSION

We conclude the paper with a summary of our results and a discussion on the connections of AE to statistical ranking and spectral clustering.

5.1 Summary

We compare LS and AE on integrating pairwise local ordering measurements to yield a global ordering. LS and AE are different on two accounts: seeking the embedding in the linear space or the angular space, and evaluating the embedding in terms of the variance or the distance minimization criterion.

Our theoretical analysis shows that:

1. Both distance and variance criteria result in the same LS or AE optimum.

2. It is thus not the criterion that sets LS and AE apart, but encoding the size and confidence of ordering in complex numbers and seeking the embedding in the angular space allows AE to be more effective than LS at integration.
3. The transition matrix for AE encodes the propagation of confidence (just like LS) in the magnitude space and the separation of ordering in the phase space, and its eigenvector completely determines the optimal embedding.
4. The representation in the complex domain leads to a non-convex error function that gives AE sensitivity to small inconsistencies and yet robustness to large inconsistencies in the embedding, and leads to an eigensolution that gives AE efficiency and near-global optimality in computation.

Our numerical experiments demonstrate that:

1. AE is faster and more effective than LS at global integration from partial measurements.
2. AE has a stronger capability to correct measurement outliers, mitigate and restrict embedding errors to infected pixels and their involved neighbours.
3. AE enjoys little loss in the embedding quality with an increasing percentage of measurement outliers until the level reaches a break point beyond 40%.
4. AE’s robustness to outliers cannot be matched by employing L_1 norm on the error or bounding the range of the embedding, suggesting that the key to overcoming outliers lies not with additionally imposing constraints on the solution, but with adaptively penalizing inconsistency between measurements themselves.

5.2 Connections to Statistical Ranking

AE significantly advances statistical ranking methods [2], [3], [8], [10] by removing the impact of outliers without expensive and explicit inconsistency characterization.

Pairwise rankings can be represented as edge flows on a graph, and pure inconsistency in these measurements is indicated by cyclic edge flows. The graph Helmholtzian (i.e., the vector Laplacian, as opposed to the scalar Laplacian) can thus unravel the ranking structure from edge flows [10]. According to the combinatorial Hodge theory, each edge flow can be decomposed into three orthogonal components, a curl-free *gradient flow* that represents the L_2 -optimal global ranking and two divergence-free flows that measure the validity of the global ranking (larger values for poorer rankings): a locally cyclic *curl flow* and a locally acyclic but globally cyclic *harmonic flow* that indicate whether the inconsistency in the measurements arises locally or globally.

However, the discrete Hodge decomposition [44] only provides some diagnostic information about outliers; it does not remove the impact of the outliers. It is an LS problem of higher computational complexity, since it deals with not just the number of nodes but the

number of triangles in the graph. Algorithms based on finite elements, multiscale, and smoothed particle hydrodynamics [45]–[48] have been proposed to reduce the complexity. While, in theory, the inconsistency is indicated by large curl or harmonic flow components, in practice it is hard to identify the few sparse outliers that could cause cyclic flows in a much larger scope. Furthermore, as our results have demonstrated, while an L_1 modification may enforce the sparseness in embedding errors [10], it cannot achieve robustness to outliers.

Another interesting connection between AE and statistical ranking is angular formulations. When AE is rewritten as a function of angular differences (as opposed to its complex representation), it exhibits a non-convex objective function that behaves like LS near 0 but is upper-bounded. This key property reveals why AE is more robust to outliers than LS, while still allowing for a global optimization via eigendecomposition as the constraint of embedding into the unit circle is relaxed. Curiously enough, while a χ^2 test for the goodness of fit between measured data and fitted data is usually performed on their differences directly, which requires estimating variances, an alternative is performed on the differences in the angles obtained by applying the inverse sine transformation to the data, which are normally distributed with variance independent of the underlying distribution [8]. A full investigation from a statistical perspective could shed more light into the advantage of formulating the embedding in the angular space.

5.3 Connections to Spectral Clustering

AE significantly advances spectral clustering (SC) methods [35], [49]–[52] by providing an ordered cluster organization and covering the entire size-confidence measurement space.

Popularized by normalized cuts for image segmentation [49], SC has been applied to motion segmentation and tracking [53], boundary detection [54], object recognition [55], image matching [56], contour grouping [57], spatial layout inference [58], image matting [59], etc. What makes them appealing is the convenience of pairwise grouping cues readily definable from applications, and the assurance of near-global optima efficiently computable via eigendecomposition [60].

However, knowing *what* goes with what in the same cluster is often not enough. It is desirable to know *how* different clusters go together (Fig. 13a). For example, image segmentation organizes pixels into disjoint regions, with little sense of the organization of the regions themselves, in terms of either visual saliency or depth ordering. SC with attraction and repulsion cues [61]–[63] does produce an ordered clustering, but it is problematic in theory and rather limited in practice.

Conventionally, SC views clustering as an emergent global property from local measurements of feature similarity. Consequently, there are three key limitations which have motivated the introduction of the dual cues of attraction and repulsion [61], [62].

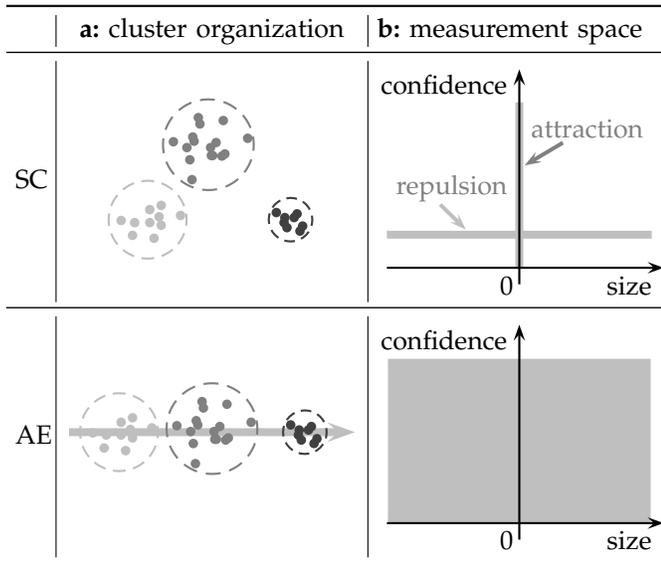


Fig. 13. AE improves SC with an ordered cluster organization and more complete measurement coverage. a) While SC finds a list of disjoint clusters, AE outputs an ordered partitioning. b) While AE covers the entire size-confidence measurement space, SC covers only two lines of different natures: Attraction measures the confidence in size 0, whereas repulsion measures the size of local orderings with a constant level of confidence.

1. The ambiguity of zero. The similarity relationships can be regarded as a form of attraction. Attraction is transitive, allowing us to supply local attraction fields and assume zero connections beyond. This step reduces the computational complexity of SC from $O(n^3)$ to $O(n^{\frac{3}{2}})$ [64], where n is the number of elements to be grouped. However, attraction of 0 is ambiguous: It could mean that the measurement is either zero or not available.

2. Fatal attraction. Attraction alone has no mechanism to check the growth of its transitivity. Suppose a and b are good friends of c , but a and b completely dislike each other. When grouping is carried out with attraction alone, a high affinity between (a, c) and (b, c) ensures that even zero affinity between (a, b) cannot interfere with clustering a, b, c into one group. This fatal drawback is caused by the innate transitivity of attraction, the very property that allows a global grouping to emerge from local attraction fields. However, with no explicit encoding of dislikes between elements, it is difficult to stop the unwanted propagation of attraction.

3. Disjoint use of similarity and dissimilarity. Feature dissimilarity contributes independently of feature similarity to the process of grouping. For examples, reds and greens pop out among surrounding blacks, due not so much to the similarity between them as to their dissimilarity from the common blacks; local occlusion cues that separate two regions in depth layers also bind different regions in the same depth layers together. However, such an active force for grouping has traditionally not been considered simultaneously with feature similarity.

With pairwise cues of attraction and repulsion natures, SC views clustering as the result of both grouping and

segregation processes, by feature similarity as well as feature dissimilarity. A positive (negative) cue suggests that two elements more likely to be together (separate), whereas 0 indicates a neutral tendency towards either outcome. To be grouped and not to be grouped can both be specified and considered simultaneously.

When repulsion is directional, an ordered clustering results: A zero difference between two elements in the outcome means they are in the same group, whereas a positive (negative) difference means they are in different groups with a positive (negative) advance [61].

A close examination of the ordered clustering interpretation reveals that the dichotomy of attraction and repulsion is flawed. Attraction describes the *confidence* in a zero-difference outcome: Large attraction encourages a zero-difference outcome, and says nothing about any non-zero difference outcomes. On the other hand, repulsion describes the *size* of the difference itself: Large repulsion encourages the difference of this particular size in the outcome, and says nothing about the confidence in this or any other outcome (Fig. 13b).

Attraction being the confidence in size 0, repulsion being the size of a measurement itself, the two do not specify pairwise cues on the same terms. They work together only when their measurements complement each other. For example, region segmentation and depth ordering can be accomplished in one step, only when attraction acts to pull pixels together inside a region and repulsion acts to push pixels apart along region boundaries [61]. However, as observed in segmentation with matting cues [65], such a scheme works only if the right cues are at the right places. This requirement is obviously too much to ask of local cues in most applications. Consequently, ordered SC has not been so widely used as its orderless counterpart despite its larger demand and revived interest [18], [63], [65]–[67].

SC with attraction and repulsion is a remedy to an essentially orderless clustering, imposing an ordering out of grouping cues of two different natures, which cannot represent or distinguish low confidence in a large value from high confidence in a small value.

AE aims at a global ordering directly from its local ordering measurements, each equipped with two numbers: size and confidence (Fig. 13b). As SC with attraction and repulsion, AE also has a complex representation and an efficient eigensolution. However, AE overcomes the three limitations of SC and gives a principled interpretation to every cue, not by the nature of grouping it suggests, but by the outcome in the ordering it desires.

AE is superior to SC with attraction and repulsion, not only because it covers the entire measurement space and delivers an ordered clustering, but also because it is remarkably robust to measurement outliers.

Since AE overcomes the limitations of the widely used LS and SC, it is a valuable new tool for many applications other than brightness modeling [17] and figure-ground segmentation [18]. Quadratic criteria prone to outliers can nonetheless achieve robustness with com-

plex measurements. This idea is also applicable to problem formulations other than embedding.

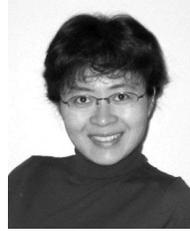
ACKNOWLEDGEMENTS

This research was funded by US National Science Foundation (NSF) CAREER IIS-0644204 and a Clare Boothe Luce Professorship. The author would like to thank Francis Bach and anonymous reviewers for thoughtful comments and excellent suggestions.

REFERENCES

- [1] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Mathematical Psychology*, vol. 15, pp. 234–81, 1977.
- [2] T. L. Saaty and L. G. Vargas, "Inconsistency and rank preservation," *Journal of Mathematical Psychology*, vol. 28, pp. 205–14, 1984.
- [3] T. L. Saaty and M. Sagir, "An essay on rank preservation and reversal," *Mathematical and Computer Modelling*, vol. 49, pp. 1230–43, 2009.
- [4] A. Sen, "The possibility of social choice," *Nobel Lecture on Economic Sciences*, 1998.
- [5] M. Ma, "A matrix approach to asset pricing in foreign exchange market," *Social Science Research Network*, 2006, <http://ssrn.com/abstract=921755>.
- [6] F. Mosteller, "Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations," *Psychometrika*, vol. 16, no. 1, pp. 3–9, 1951.
- [7] —, "Remarks on the method of paired comparisons: II. the effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed," *Psychometrika*, vol. 16, no. 2, pp. 203–6, 1951.
- [8] —, "Remarks on the method of paired comparisons: III. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed," *Psychometrika*, vol. 16, no. 2, pp. 207–18, 1951.
- [9] P. Diaconis, "A generalization of spectral analysis with application to ranked data," *The Annals of Statistics*, vol. 17, no. 3, pp. 949–79, 1989.
- [10] X. Jiang, L. heng Lim, Y. Yao, and Y. Ye, "Statistical ranking and combinatorial Hodge theory," *Mathematical Programming*, vol. 127, no. 1, pp. 203–44, 2011.
- [11] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things," *J. of Artificial Intelligence Research*, vol. 10, pp. 243–70, 1999.
- [12] R. Herbrich, T. Graepel, and K. Obermayer, "Large margin rank boundaries for ordinal regression," in *Advances in Large Margin Classifiers*. MIT Press, 2000, pp. 115–132.
- [13] A. Shashua and A. Levin, "Ranking with large margin principle: two approaches," in *Neural Information Processing Systems*, 2003.
- [14] R. M. Bell and Y. Koren, "Scalable collaborative filtering with jointly derived neighbourhood interpolation weights," in *Int'l Conf. on Machine Learning*, 2007.
- [15] C. Cortes, M. Mohri, and A. Rastogi, "Magnitude-preserving ranking algorithms," in *Int'l Conf. on Machine Learning*, 2007.
- [16] A. Agrawal, R. Raskar, S. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. Graph.*, vol. 24, pp. 828–35, 2005.
- [17] S. X. Yu, "Angular embedding: from jarring intensity differences to perceived luminance," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2009.
- [18] M. Maire, "Simultaneous segmentation and figure-ground organization using angular embedding," in *European Conf. Computer Vision*, 2010.
- [19] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 439–51, 1988.
- [20] B. K. P. Horn and M. J. Brooks, *Shape from Shading*. MIT Press, 1989.
- [21] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, pp. 249–56, 2002.
- [22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley and Sons, 2001.
- [23] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [24] D. Forsyth, "Shape from texture and integrability," in *Int'l Conf. Computer Vision*, 2001.
- [25] A. Yuille and D. Snow, "Shape and albedo from multiple images using integrability," in *IEEE Conf. Computer Vision & Pattern Recognition*, 1997.
- [26] N. Petrovic, I. Cohen, B. Frey, R. Koetter, and T. Huang, "Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2001.
- [27] P. Kovesi, "Shapelets correlated with surface normals produce surfaces," in *Int'l Conf. Computer Vision*, 2005.
- [28] T. Simchony, R. Chellappa, and M. Shao, "Direct analytical methods for solving Poisson equations in computer vision problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 435–46, 1990.
- [29] N. Petrovic, I. Cohen, B. Frey, R. Koetter, and T. Huang, "Enforcing integrability for surface reconstruction algorithms using belief propagation in graphical models," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2001.
- [30] A. S. Georghiadis, P. N. Bellhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 643–60, 2001.
- [31] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [32] I. Markovsky and S. V. Huffel, "Overview of total least squares methods," *Signal Processing*, vol. 87, pp. 2283–302, 2007.
- [33] F. R. K. Chung, *Spectral Graph Theory*. Am. Math. Soc., 1997.
- [34] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–6, 2000.
- [35] M. Meila and J. Shi, "Learning segmentation with random walk," in *Neural Information Processing Systems*, 2001.
- [36] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods," *Proc. Natl. Aca. Sci.*, vol. 102, no. 21, pp. 7432–37, 2005.
- [37] —, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps," *Proc. Natl. Aca. Sci.*, vol. 102, no. 21, pp. 7426–31, 2005.
- [38] B. Nadler and M. Galun, "Fundamental limitations of spectral clustering," in *Neural Information Processing Systems*, 2006.
- [39] A. Singer, Y. Shkolnisky, and B. Nadler, "Diffusion interpretation of nonlocal neighbourhood filters for signal denoising," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 118–39, 2009.
- [40] R. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. New York: Wiley, 1987.
- [41] D. A. Spielman and S. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *Annual ACM Symposium on Theory of Computing*, 2004.
- [42] J. Kuczynski and H. Wozniakowski, "Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start," *SIAM J. Matrix Anal. & Appl.*, vol. 13, no. 4, pp. 1094–122, 1992.
- [43] P. B. Stark and R. L. Parker, "Bounded-variable least squares: an algorithm and applications," *Comp. Stat.*, vol. 10, pp. 129–41, 1995.
- [44] K. Polthier and E. Preuss, "Variational approach to vector field decomposition," in *Eurographics Workshop on Scientific Visualization*, 2000.
- [45] —, "Identifying vector fields singularities using a discrete Hodge decomposition," in *Visualization and Mathematics III*, 2003.
- [46] Q. Guo, M. K. Mandal, and M. Y. Li, "Efficient Hodge-Helmholtz decomposition of motion fields," *Pattern Recognition Letters*, vol. 26, pp. 493–501, 2005.
- [47] Y. Tong, S. Lombeyda, A. N. Hirani, and M. Desbrun, "Discrete multiscale vector field decomposition," *ACM Trans. Graphics*, vol. 22, no. 3, pp. 445–52, 2003.
- [48] F. Petronetto, A. Paiva, M. Lage, G. Tavares, H. Lopes, and T. Lewiner, "Meshless Helmholtz-Hodge decomposition," *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 2, pp. 338–49, 2010.
- [49] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE Conf. Computer Vision & Pattern Recognition*, 1997.
- [50] Y. Weiss, "Segmentation using eigenvectors: a unifying view," in *Int'l Conf. Computer Vision*, 1999.
- [51] A. Y. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Neural Information Processing Systems*, 2002.

- [52] F. Bach and M. I. Jordan, "Learning spectral clustering," in *Neural Information Processing Systems*, 2004.
- [53] J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Int'l Conf. Computer Vision*, 1998.
- [54] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *Int'l J. Computer Vision*, vol. 43, pp. 7–27, 2001.
- [55] S. X. Yu and J. Shi, "Object-specific figure-ground segregation," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2003.
- [56] A. Toshev, J. Shi, and K. Daniilidis, "Image matching via saliency region correspondences," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2007.
- [57] Q. Zhu, G. Song, and J. Shi, "Untangling cycles for contour grouping," in *Int'l Conf. Computer Vision*, 2007.
- [58] S. X. Yu, H. Zhang, and J. Malik, "Inferring spatial layout from a single image via depth-ordered grouping," in *IEEE Workshop on Perceptual Organization in Computer Vision*, 2008.
- [59] A. Levin, A. Rav-Acha, and D. Lischinski, "Spectral matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 1699–712, 2008.
- [60] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *Int'l Conf. Computer Vision*, 2003.
- [61] —, "Segmentation with pairwise attraction and repulsion," in *Int'l Conf. Computer Vision*, 2001.
- [62] —, "Understanding popout through repulsion," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2001.
- [63] E. Bernardis and S. X. Yu, "Finding dots: Segmentation as popping out regions from boundaries," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2010.
- [64] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 888–905, 2000.
- [65] A. Stein, T. Stepleton, and M. Hebert, "Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2008.
- [66] S. Bagon, O. Brostkovski, M. Galun, and M. Irani, "Detecting and sketching the common," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2010.
- [67] S. Vitaladevuni and R. Basri, "Co-clustering of image segments using convex optimization applied to EM neuronal reconstruction," in *IEEE Conf. Computer Vision & Pattern Recognition*, 2010.



Stella X. Yu received the PhD degree in 2003 from the School of Computer Science at Carnegie Mellon University, where she studied robotics at the Robotics Institute and vision science at the Center for the Neural Basis of Cognition. She continued her computer vision research as a postdoctoral researcher in the Computer Science Department at the University of California Berkeley. Since she joined the faculty of Boston College on a Clare Booth Luce Professorship in 2005, Dr. Yu has been developing an interdisciplinary curriculum and research agenda on Art and Vision, for which she received the US National Science Foundation (NSF) CAREER award in 2007. Her research interests include spectral graph theory, perceptual organization, image segmentation, brightness modeling, scene classification, visual attention, and non-photorealistic rendering. She is a member of the IEEE and the IEEE Computer Society.