Optimal Transport for Super Resolution Applied to Astronomy Imaging

Michael G. Rawson[1,2], Jakob Hultgren[2]

[1]Pacific Northwest National Laboratory, Seattle, WA, USA
[2]University of Maryland at College Park, College Park, MD, USA

July 15, 2022

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Contents

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Super Resolution

Super resolution seeks to improve image resolution without further data collection.

Super Resolution

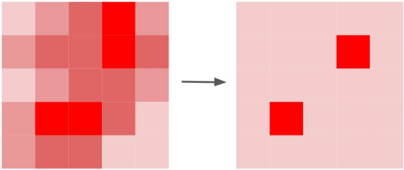
Super resolution seeks to improve image resolution without further data collection.

This is possible, given constraints that give a well-posed inverse problem, for example smoothness or **sparsity**.

Super Resolution

Super resolution seeks to improve image resolution without further data collection.

This is possible, given constraints that give a well-posed inverse problem, for example smoothness or **sparsity**.



Optimal Transport

Definition

Let $\mu \in \mathbb{R}^n_{\geq 0}, \nu \in \mathbb{R}^m_{\geq 0}$ be probability vectors and $C \in \mathbb{R}^n \times \mathbb{R}^m$.

Optimal Transport

Definition

Let $\mu \in \mathbb{R}^n_{\geq 0}, \nu \in \mathbb{R}^m_{\geq 0}$ be probability vectors and $C \in \mathbb{R}^n \times \mathbb{R}^m$. Then the optimal transport plan from μ to ν is

$$\arg \min_{P \in \Pi(\mu,\nu)} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} P_{ij} \tag{1}$$

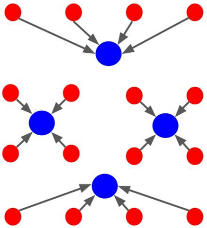
Optimal Transport

Definition

Let $\mu \in \mathbb{R}^n_{\geq 0}, \nu \in \mathbb{R}^m_{\geq 0}$ be probability vectors and $C \in \mathbb{R}^n \times \mathbb{R}^m$. Then the optimal transport plan from μ to ν is

$$\arg \min_{P \in \Pi(\mu,\nu)} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} P_{ij} \tag{1}$$

where $\Pi(\mu, \nu) = \{P \in \mathbb{R}^{n \times m} : \sum_{i=1}^{n} P_{ij} = \nu_j \ \forall j, \sum_{j=1}^{n} P_{ij} = \mu_i \ \forall i\}$.



Entropy

Entropy

Definition

For a probability mass function $p : \mathcal{J} \to \mathbb{R}$ (non-negative and sums to 1), we will use $H(p)$ to denote its entropy,

$$H(p) = -\sum_{\iota \in \mathcal{J}} p(\iota) \ln(p(\iota)). \tag{2}$$

Entropy

Definition

For a probability mass function $p : \mathcal{J} \to \mathbb{R}$ (non-negative and sums to 1), we will use $H(p)$ to denote its entropy,

$$H(p) = -\sum_{\iota \in \mathcal{J}} p(\iota) \ln(p(\iota)). \tag{2}$$

Key point: Sparse arrays have low entropy.

Entropic regularisation

Approximate optimal transport distances can be computed by solving the entropic regularization of optimal transport

$$\arg\min_{P\in\Pi(\mu,\nu)} \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}P_{ij} - \epsilon H(P).$$

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Entropic regularisation

Approximate optimal transport distances can be computed by solving the entropic regularization of optimal transport

$$\arg \min_{P \in \Pi(\mu,\nu)} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} P_{ij} - \epsilon H(P).$$

This is equivalent to the matrix scaling problem, more precisely finding positive multipliers $f \in \mathbb{R}^n, g \in \mathbb{R}^m$ such that the matrix

$$(f_i e^{-C_{ij}/\epsilon} g_j) \in \Pi(\mu, \nu).$$

Sinkhorn

The matrix scaling problem can be solved with the Sinkhorn algorithm, alternating between solving

$$\sum_{j=1}^{m} f_i e^{-C_{ij}/\epsilon} g_j = \mu_i$$

for f

Sinkhorn

The matrix scaling problem can be solved with the Sinkhorn algorithm, alternating between solving

$$\sum_{j=1}^{m} f_i e^{-C_{ij}/\epsilon} g_j = \mu_i$$

for f and then solving

$$\sum_{i=1}^{n} f_i e^{-C_{ij}/\epsilon} g_j = \nu_j$$

for g.

Wasserstein Distance Gradient

The multipliers f and g play the role of Lagrange multipliers and can be used to approximate the variation of $W(\mu, \nu)$ with respect to the first and second argument respectively.

Wasserstein Distance Gradient

The multipliers f and g play the role of Lagrange multipliers and can be used to approximate the variation of $W(\mu, \nu)$ with respect to the first and second argument respectively.

This can be used to implement gradient descent methods.

Wasserstein Inverse Problem for Super Resolution

For a measurement ν and positive regularization parameter λ, we define the *sparse approximation* of ν as a minimizer

$$\mu_* = \arg\min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{3}$$

Wasserstein Inverse Problem for Super Resolution

For a measurement ν and positive regularization parameter λ, we define the *sparse approximation* of ν as a minimizer

$$\mu_* = \arg \min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{3}$$

- At least one minimizer exists by compactness of the finite dimensional probability simplex.

Wasserstein Inverse Problem for Super Resolution

For a measurement ν and positive regularization parameter λ, we define the *sparse approximation* of ν as a minimizer

$$\mu_* = \arg\min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{3}$$

- At least one minimizer exists by compactness of the finite dimensional probability simplex.
- The entropy term favors sparse solutions. There is a trade-off between sparsity of the solution and proximity to the measurement.

Wasserstein Inverse Problem for Super Resolution

For a measurement ν and positive regularization parameter λ, we define the *sparse approximation* of ν as a minimizer

$$\mu_* = \arg\min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{3}$$

- At least one minimizer exists by compactness of the finite dimensional probability simplex.
- The entropy term favors sparse solutions. There is a trade-off between sparsity of the solution and proximity to the measurement.
- This inverse problem is useful whenever there is a natural distance, or cost function, on the index set of ν.

Wasserstein Inverse Problem for Super Resolution

For a measurement ν and positive regularization parameter λ, we define the *sparse approximation* of ν as a minimizer

$$\mu_* = \arg\min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{3}$$

- At least one minimizer exists by compactness of the finite dimensional probability simplex.
- The entropy term favors sparse solutions. There is a trade-off between sparsity of the solution and proximity to the measurement.
- This inverse problem is useful whenever there is a natural distance, or cost function, on the index set of ν.
- We will let each entry in ν describe the intensity of a pixel in a 32×32 image and cost C_{ij} is chosen as the L^2-distance between the i^{th} pixel and j^{th} pixel.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Input: $\mu, \nu \in \mathbb{R}^n$: positive probability vectors, $C \in \mathbb{R}^{n \times n}$: cost matrix, ϵ : positive regularization parameter

Input: $\mu, \nu \in \mathbb{R}^n$: positive probability vectors, $C \in \mathbb{R}^{n\times n}$: cost matrix, ϵ : positive regularization parameter
Output: $d_W^\epsilon \in \mathbb{R}$: regularized distance between μ and ν, $F \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to μ at μ, $G \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to ν at ν

Input: $\mu, \nu \in \mathbb{R}^n$: positive probability vectors, $C \in \mathbb{R}^{n \times n}$: cost matrix, ϵ : positive regularization parameter
Output: $d_W^\epsilon \in \mathbb{R}$: regularized distance between μ and ν, $F \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to μ at μ, $G \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to ν at ν

Algorithm 3: The Sinkhorn Algorithm for Optimal Transport Distances

$f = (1, \ldots, 1) \in \mathbb{R}^n$
$g = (1, \ldots, 1) \in \mathbb{R}^n$
while *f and g have not converged* **do**
  for $1 \leq i \leq n$ **do**
    $f_i = \mu_i / \left(\sum_j \exp(-C_{ij}/\epsilon) g_j \right)$
  end

DEPARTMENT OF MATHEMATICS
Pacific Northwest NATIONAL LABORATORY

Input: $\mu, \nu \in \mathbb{R}^n$: positive probability vectors, $C \in \mathbb{R}^{n \times n}$: cost matrix, ϵ : positive regularization parameter
Output: $d_W^\epsilon \in \mathbb{R}$: regularized distance between μ and ν, $F \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to μ at μ, $G \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu, \nu)$ with resp. to ν at ν

Algorithm 4: The Sinkhorn Algorithm for Optimal Transport Distances

$f = (1, \ldots, 1) \in \mathbb{R}^n$
$g = (1, \ldots, 1) \in \mathbb{R}^n$
while *f and g have not converged* **do**
  for $1 \le i \le n$ **do**
    $f_i = \mu_i / \left(\sum_j \exp(-C_{ij}/\epsilon) g_j \right)$
  end
  for $1 \le j \le n$ **do**
    $g_j = \nu_j / \left(\sum_i \exp(-C_{ij}/\epsilon) f_i \right)$
  end
end

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Input: $\mu, \nu \in \mathbb{R}^n$: positive probability vectors, $C \in \mathbb{R}^{n \times n}$: cost matrix, ϵ : positive regularization parameter

Output: $d_W^\epsilon \in \mathbb{R}$: regularized distance between μ and ν, $F \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu,\nu)$ with resp. to μ at μ, $G \in \mathbb{R}^n$: gradient of $d_W^\epsilon(\mu,\nu)$ with resp. to ν at ν

Algorithm 5: The Sinkhorn Algorithm for Optimal Transport Distances

$f = (1, \ldots, 1) \in \mathbb{R}^n$

$g = (1, \ldots, 1) \in \mathbb{R}^n$

while *f and g have not converged* **do**

  for $1 \leq i \leq n$ **do**

    $f_i = \mu_i / \left(\sum_j \exp(-C_{ij}/\epsilon) g_j \right)$

  end

  for $1 \leq j \leq n$ **do**

    $g_j = \nu_j / \left(\sum_i \exp(-C_{ij}/\epsilon) f_i \right)$

  end

end

$d_W^\epsilon = \sum_{i=1}^n \sum_{j=1}^n f_i g_j \exp(-C_{ij}/\epsilon) C_{ij}$

$F = -\epsilon \ln(f)$

$G = -\epsilon \ln(g)$

Input: $X \in \mathbb{R}^{N \times m \times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level, $0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2 \times m^2}$: cost matrix, $J_{\lambda,\epsilon}(x, v) := d_W^{\epsilon}(x, v) + \lambda H(v)$

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Input: $X \in \mathbb{R}^{N \times m \times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level,
$0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2 \times m^2}$: cost matrix,
$J_{\lambda,\epsilon}(x, v) := d_W^{\epsilon}(x, v) + \lambda H(v)$
Output: $K \in \mathbb{R}^N$: star cluster classification

Input: $X \in \mathbb{R}^{N\times m\times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level, $0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2\times m^2}$: cost matrix, $J_{\lambda,\epsilon}(x, v) := d^{\epsilon}_{W}(x, v) + \lambda H(v)$
Output: $K \in \mathbb{R}^{N}$: star cluster classification

Algorithm 8: O.T. Super Resolution Clustering [Rawson and Hultgren, 2022]

$K = 0$
for $i = 1, 2, ..., N$ **do**
  $v = X_i$
  while v *has not converged* **do**
    $w = \nabla d^{\epsilon}_{W}(X_i, \cdot)|_v + \lambda \nabla H|_v$
    $w = w - \langle w, \frac{1}{m}\mathbb{1}\rangle \cdot \frac{1}{m}\mathbb{1}$

Input: $X \in \mathbb{R}^{N\times m\times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level, $0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2\times m^2}$: cost matrix, $J_{\lambda,\epsilon}(x, v) := d_W^{\epsilon}(x, v) + \lambda H(v)$
Output: $K \in \mathbb{R}^N$: star cluster classification

Algorithm 9: O.T. Super Resolution Clustering [Rawson and Hultgren, 2022]

$K = 0$
for $i = 1, 2, ..., N$ **do**
  $v = X_i$
  while v *has not converged* **do**
    $w = \nabla d_W^{\epsilon}(X_i, \cdot)|_v + \lambda \nabla H|_v$
    $w = w - \langle w, \frac{1}{m}\mathbb{1}\rangle \cdot \frac{1}{m}\mathbb{1}$
    $\alpha = \sup\{\alpha \in \mathbb{R} : J_{\lambda,\epsilon}(v) > J_{\lambda,\epsilon}(v - \alpha w)\}$
    $\alpha = \min\{0.01, \alpha\}$

Input: $X \in \mathbb{R}^{N \times m \times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level, $0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2 \times m^2}$: cost matrix, $J_{\lambda,\epsilon}(x, v) := d_W^{\epsilon}(x, v) + \lambda H(v)$
Output: $K \in \mathbb{R}^N$: star cluster classification

Algorithm 10: O.T. Super Resolution Clustering [Rawson and Hultgren, 2022]

$K = 0$
for $i = 1, 2, ..., N$ **do**
    $v = X_i$
    while v *has not converged* **do**
        $w = \nabla d_W^{\epsilon}(X_i, \cdot)|_v + \lambda \nabla H|_v$
        $w = w - \langle w, \frac{1}{m}\mathbb{1}\rangle \cdot \frac{1}{m}\mathbb{1}$
        $\alpha = \sup\{\alpha \in \mathbb{R} : J_{\lambda,\epsilon}(v) > J_{\lambda,\epsilon}(v - \alpha w)\}$
        $\alpha = \min\{0.01, \alpha\}$
        $v = v - \alpha w$; $v = diag(\mathbb{1}_{v>0})\, v$; $v = v/\|v\|_1$
    end

Input: $X \in \mathbb{R}^{N\times m\times m}$: N images size $m \times m$, $\lambda \in \mathbb{R}$: positive noise level, $0 < \epsilon < 1$: optimal transportation regularization, $C \in \mathbb{R}^{m^2\times m^2}$: cost matrix, $J_{\lambda,\epsilon}(x, v) := d_W^\epsilon(x, v) + \lambda H(v)$

Output: $K \in \mathbb{R}^N$: star cluster classification

Algorithm 11: O.T. Super Resolution Clustering [Rawson and Hultgren, 2022]

$K = 0$
for $i = 1, 2, ..., N$ **do**
  $v = X_i$
  while *v has not converged* **do**
    $w = \nabla d_W^\epsilon(X_i, \cdot)|_v + \lambda \nabla H|_v$
    $w = w - \langle w, \frac{1}{m}\mathbb{1}\rangle \cdot \frac{1}{m}\mathbb{1}$
    $\alpha = \sup\{\alpha \in \mathbb{R} : J_{\lambda,\epsilon}(v) > J_{\lambda,\epsilon}(v - \alpha w)\}$
    $\alpha = \min\{0.01, \alpha\}$
    $v = v - \alpha w;\ v = diag(\mathbb{1}_{v>0})\ v;\ v = v/\|v\|_1$
  end
  $V_i = v;\ \delta = \max V_i$
  if $rank(H_0(V_i^{-1}([0.75\delta,\ \delta]))) == 1$ **then**
    $K_i = 1$
  end
end

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Optimal Transport Bound on Gaussian Noise

Theorem ([Rawson and Hultgren, 2022])

Let $\nu = \frac{1}{k}\sum_{i=1}^{k} \delta_{p_i}$ where $p_i \in \mathbb{R}^d$ and $\tilde{\nu}$ be the noisy signal acquired by sampling, for each i, n points according to a normal distribution centered at p_i with independent components of variance σ^2.

Optimal Transport Bound on Gaussian Noise

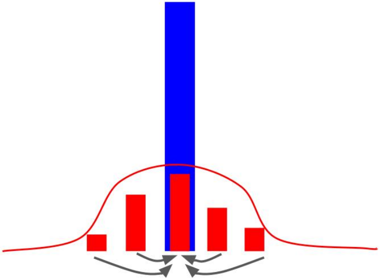
Theorem ([Rawson and Hultgren, 2022])

Let $\nu = \frac{1}{k}\sum_{i=1}^{k}\delta_{p_i}$ where $p_i \in \mathbb{R}^d$ and $\tilde{\nu}$ be the noisy signal acquired by sampling, for each i, n points according to a normal distribution centered at p_i with independent components of variance σ^2. Then the Wasserstein distance between ν and $\bar{\nu}$ is bounded by a random variable with expected value $d\sigma^2$ and variance $2d\sigma^4/N$, where $N = nk$ is the total number of points sampled.

Optimal Transport Bound on Gaussian Noise

Theorem ([Rawson and Hultgren, 2022])

Let $\nu = \frac{1}{k}\sum_{i=1}^{k}\delta_{p_i}$ where $p_i \in \mathbb{R}^d$ and $\tilde{\nu}$ be the noisy signal acquired by sampling, for each i, n points according to a normal distribution centered at p_i with independent components of variance σ^2. Then the Wasserstein distance between ν and $\bar{\nu}$ is bounded by a random variable with expected value $d\sigma^2$ and variance $2d\sigma^4/N$, where $N = nk$ is the total number of points sampled.



Reconstructing the Support of a Sparse Signal

Theorem ([Rawson and Hultgren, 2022])

Assume ν is a sparse signal and $\bar{\nu}$ is a noisy signal such that $d_w(\nu, \bar{\nu}) < \delta$.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Reconstructing the Support of a Sparse Signal

Theorem ([Rawson and Hultgren, 2022])

Assume ν is a sparse signal and $\bar{\nu}$ is a noisy signal such that $d_w(\nu, \bar{\nu}) < \delta$. Then the solution of

$$\mu = \arg \min_{\mu : d_W(\bar{\nu}, \mu) \leq \delta} H(\mu)$$

will identify the structure of μ, i.e. have the same support as μ, if $||\nu||_0 \leq ||\mu||_0$ for all μ such that $d_W(\mu, \bar{\nu}) < 2\delta$, with equality only if μ and ν has the same support.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Reconstructing the Support of a Sparse Signal [Rawson and Hultgren, 2022]

Remark

The conditions in Theorem 4 can be summarized as a low enough noise level δ and enough sparsity of the true signal ν (making it a local minimizer of the L^0-norm).

Reconstructing the Support of a Sparse Signal [Rawson and Hultgren, 2022]

Remark

The conditions in Theorem 4 can be summarized as a low enough noise level δ and enough sparsity of the true signal ν (making it a local minimizer of the L^0-norm). It is interesting to note that these conditions are essentially necessary: if the inequality in Theorem 4 is violated by some μ closer than δ to $\bar{\nu}$, then the solution of (4) *does not identify the structure of ν.*

Reconstructing the Support of a Sparse Signal [Rawson and Hultgren, 2022]

Remark

The conditions in Theorem 4 can be summarized as a low enough noise level δ and enough sparsity of the true signal ν (making it a local minimizer of the L^0-norm). It is interesting to note that these conditions are essentially necessary: if the inequality in Theorem 4 is violated by some μ closer than δ to $\bar{\nu}$, then the solution of (4) does not identify the structure of ν.

Remark

Noise is high entropy, hence it is expected that the noise can be removed by minimizing the entropy.

Reconstructing the Support of a Sparse Signal [Rawson and Hultgren, 2022]

Remark

The conditions in Theorem 4 can be summarized as a low enough noise level δ and enough sparsity of the true signal ν (making it a local minimizer of the L^0-norm). It is interesting to note that these conditions are essentially necessary: if the inequality in Theorem 4 is violated by some μ closer than δ to $\bar{\nu}$, then the solution of (4) does not identify the structure of ν.

Remark

Noise is high entropy, hence it is expected that the noise can be removed by minimizing the entropy. However, if the signal-to-noise ratio is too low, this reconstruction is underdetermined.

Reconstructing the Support of a Sparse Signal

Theorem ([Rawson and Hultgren, 2022])

Fix a positive probability vector $\nu \in \mathbb{R}^d_{>0}$ *such that all elements of* ν *are distinct.*

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Reconstructing the Support of a Sparse Signal

Theorem ([Rawson and Hultgren, 2022])

Fix a positive probability vector $\nu \in \mathbb{R}^d_{>0}$ such that all elements of ν are distinct. Then the sparse recovery is continuous to perturbations around ν for small λ, i.e. for every $\epsilon' > 0$ there exists $\delta > 0$, such that if $d_W(\nu, \nu') < \delta$, $\mu_* = \arg\min_{\mu \in \mathbb{P}(X) : d_W(\mu,\nu) < \lambda} H(\mu)$, and $\mu'_* = \arg\min_{\mu \in \mathbb{P}(X) : d_W(\mu,\nu') < \lambda} H(\mu)$

Reconstructing the Support of a Sparse Signal

Theorem ([Rawson and Hultgren, 2022])

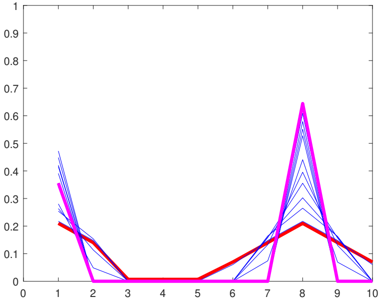
Fix a positive probability vector $\nu \in \mathbb{R}^d_{>0}$ such that all elements of ν are distinct. Then the sparse recovery is continuous to perturbations around ν for small λ, i.e. for every $\epsilon' > 0$ there exists $\delta > 0$, such that if $d_W(\nu, \nu') < \delta$, $\mu_* = \arg\min_{\mu \in \mathbb{P}(X) : d_W(\mu,\nu) < \lambda} H(\mu)$, and $\mu'_* = \arg\min_{\mu \in \mathbb{P}(X) : d_W(\mu,\nu') < \lambda} H(\mu)$ then $\|\mu_* - \mu'_*\| < \epsilon'$.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Let measurement $\hat{\nu} = (0.2, 0.15, 0, 0, 0, 0.1, 0.15, 0.2, 0.15, 0.1)$

Let measurement $\hat{\nu} = (0.2, 0.15, 0, 0, 0, 0.1, 0.15, 0.2, 0.15, 0.1)$

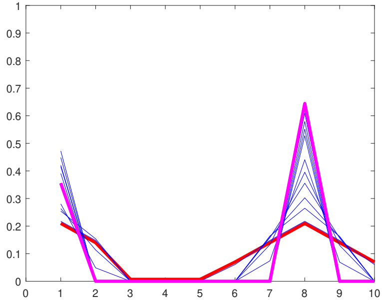


(a) Sparsity level $\lambda = 10$. Solution $\nu = (0.35, 0, 0, 0, 0, 0, 0, 0.65, 0, 0)$.

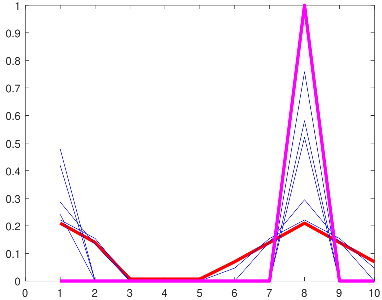
DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Let measurement $\hat{\nu} = (0.2, 0.15, 0, 0, 0, 0.1, 0.15, 0.2, 0.15, 0.1)$



(a) Sparsity level $\lambda = 10$. Solution $\nu = (0.35, 0, 0, 0, 0, 0, 0, 0.65, 0, 0)$.



(b) Sparsity level $\lambda = 100$. Solution $\nu = (0, 0, 0, 0, 0, 0, 0, 1, 0, 0)$.

Figure: Plot of super resolution O.T. method. Red line is initial distribution. Blue lines are steps along gradient. Pink line is final, converged distribution.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Star Clustering Application

- The formation and evolution of star clusters [Pérez et al., 2021]

Star Clustering Application

- The formation and evolution of star clusters [Pérez et al., 2021]
- Algorithmically detect star clusters in images of sky patches

Star Clustering Application

- The formation and evolution of star clusters [Pérez et al., 2021]
- Algorithmically detect star clusters in images of sky patches
- State of the art method trains a convolutional neural network (CNN) to classify each region in an image as containing a star cluster or not [Pérez et al., 2021]

Star Clustering Application

- The formation and evolution of star clusters [Pérez et al., 2021]
- Algorithmically detect star clusters in images of sky patches
- State of the art method trains a convolutional neural network (CNN) to classify each region in an image as containing a star cluster or not [Pérez et al., 2021]
- Neural networks are notoriously computationally expensive, sensitive to noise, and inflexible to appending or removing data variables

Star Clustering Application [Rawson and Hultgren, 2022]

Table: Confusion matrix of O.T. method on LEGUS data compared to StarcNet [Pérez et al., 2021]. Column gives StarcNet classification and row gives O.T. classification [Rawson and Hultgren, 2022].

	StarcNet Cluster	StarcNet Not Cluster
O.T. Cluster	25% (32)	13.3% (17)
O.T. Not Cluster	12.5% (16)	49.2% (63)

Star Clustering Application [Rawson and Hultgren, 2022]

Table: Confusion matrix of O.T. method on LEGUS data compared to StarcNet [Pérez et al., 2021]. Column gives StarcNet classification and row gives O.T. classification [Rawson and Hultgren, 2022].

	StarcNet Cluster	StarcNet Not Cluster
O.T. Cluster	25% (32)	13.3% (17)
O.T. Not Cluster	12.5% (16)	49.2% (63)

- The accuracy rate of clustering without sparsifying is 46% w.r.t. the CNN.

Star Clustering Application [Rawson and Hultgren, 2022]

Table: Confusion matrix of O.T. method on LEGUS data compared to StarcNet [Pérez et al., 2021]. Column gives StarcNet classification and row gives O.T. classification [Rawson and Hultgren, 2022].

	StarcNet Cluster	StarcNet Not Cluster
O.T. Cluster	25% (32)	13.3% (17)
O.T. Not Cluster	12.5% (16)	49.2% (63)

- The accuracy rate of clustering without sparsifying is 46% w.r.t. the CNN.
- The O.T. method increases in accuracy to 74% with respect to the CNN.

UNIVERSITY OF MARYLAND 1856 DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Star Clustering Application [Rawson and Hultgren, 2022]

Table: Confusion matrix of O.T. method on LEGUS data compared to StarcNet [Pérez et al., 2021]. Column gives StarcNet classification and row gives O.T. classification [Rawson and Hultgren, 2022].

	StarcNet Cluster	StarcNet Not Cluster
O.T. Cluster	25% (32)	13.3% (17)
O.T. Not Cluster	12.5% (16)	49.2% (63)

- The accuracy rate of clustering without sparsifying is 46% w.r.t. the CNN.
- The O.T. method increases in accuracy to 74% with respect to the CNN.
- The CNN accuracy rate is 86% with respect to experts, but even experts agree with each other only around 70%-75% [Wei et al., 2020].

Star Clustering Application [Rawson and Hultgren, 2022]

Table: Confusion matrix of O.T. method on LEGUS data compared to StarcNet [Pérez et al., 2021]. Column gives StarcNet classification and row gives O.T. classification [Rawson and Hultgren, 2022].

	StarcNet Cluster	StarcNet Not Cluster
O.T. Cluster	25% (32)	13.3% (17)
O.T. Not Cluster	12.5% (16)	49.2% (63)

- The accuracy rate of clustering without sparsifying is 46% w.r.t. the CNN.
- The O.T. method increases in accuracy to 74% with respect to the CNN.
- The CNN accuracy rate is 86% with respect to experts, but even experts agree with each other only around 70%-75% [Wei et al., 2020].
- Our method provides a very high performance given that no neural network training, which often takes weeks of compute time, is required.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

Conclusion

- Optimal transportation is more
  - efficient,

Conclusion

- Optimal transportation is more
  - efficient,
  - robust,

Conclusion

- Optimal transportation is more
  - efficient,
  - robust,
  - flexible

  than CNNs.

Conclusion

- Optimal transportation is more
  - efficient,
  - robust,
  - flexible

  than CNNs.
- We proved that optimal transportation will reconstruct sparse sources and is robust to noise
- Relevant for correcting distortions and noise in imaging, ex: star cluster detection
- A predictive model for star clusters can produce a *policy* that informs where future surveys should look for star clusters [Rawson and Freeman, 2021, Freeman and Rawson, 2021].

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY

End

Thank You!

Questions?

michael.rawson@pnnl.gov, hultgren@umd.edu

References

Freeman, J. and Rawson, M. (2021).
Top-K ranking deep contextual bandits for information selection systems.
In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2209–2214. IEEE.

Pérez, G., Messa, M., Calzetti, D., Maji, S., Jung, D. E., Adamo, A., and Sirressi, M. (2021).
StarcNet: Machine learning for star cluster identification.
The Astrophysical Journal, 907(2):100.

Rawson, M. and Freeman, J. (2021).
Deep upper confidence bound algorithm for contextual bandit ranking of information selection.
Proceedings of Joint Statistical Meetings (JSM), Statistical Learning and Data Science Section, 2021.

Rawson, M. and Hultgren, J. (2022).
Optimal transport for super resolution applied to astronomy imaging.
Proceedings of EUSIPCO, 2022; arxiv 2202.05354.

Wei, W., Huerta, E., Whitmore, B. C., Lee, J. C., Hannon, S., Chandar, R., Dale, D. A., Larson, K. L., Thilker, D. A., Ubeda, L., and others (2020).
Deep transfer learning for star cluster classification: I. application to the PHANGS–HST survey.
Monthly Notices of the Royal Astronomical Society, 493(3):3178–3193.

DEPARTMENT OF MATHEMATICS

Pacific Northwest NATIONAL LABORATORY