Optimal Transport for Super Resolution Applied to Astronomy Imaging

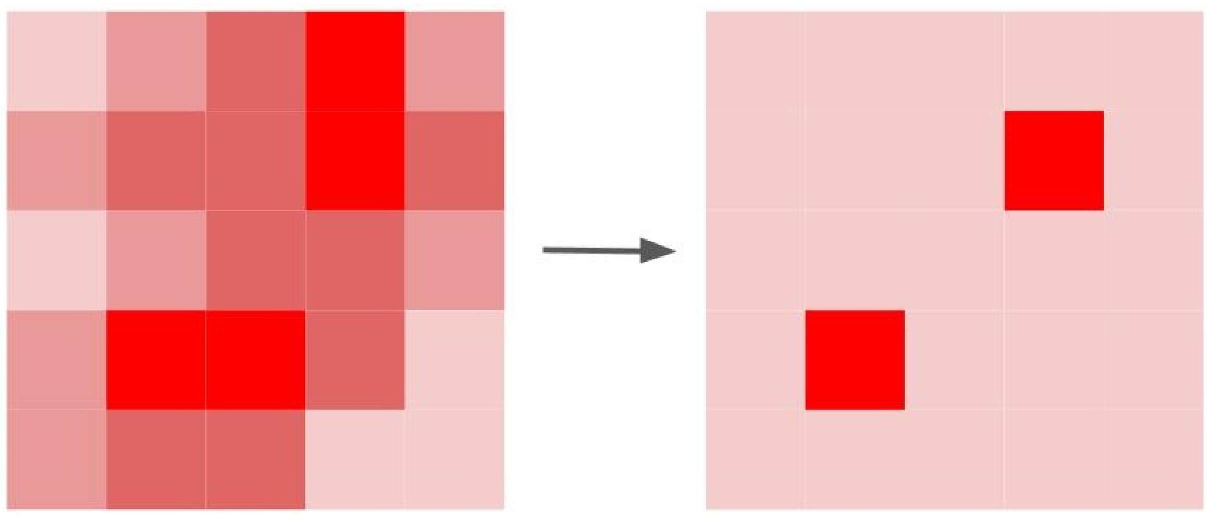
Michael G. Rawson[1,2], Jakob Hultgren[2]

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

Pacific Northwest NATIONAL LABORATORY

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

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

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

$$\mu_* = \arg \min_{\mu \in \mathbb{P}(X)} d_W^{\epsilon}(\mu, \nu) + \lambda H(\mu). \tag{2}$$

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.

Theoretic Results

Theorem ([2])

Let $\nu = \frac{1}{k}\sum_{i=1}^{k} \delta_{p_i}$ where $p_i \in \mathbb{R}^d$ and $\bar{\nu}$ 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.

Theorem ([2])

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.

Remark

The conditions in Theorem 2 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 2 is violated by some μ closer than δ to $\bar{\nu}$, then the solution of (2) does not identify the structure of ν.

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.

Theorem ([2])

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 $\|\mu_* - \mu'_*\| < \epsilon'$.

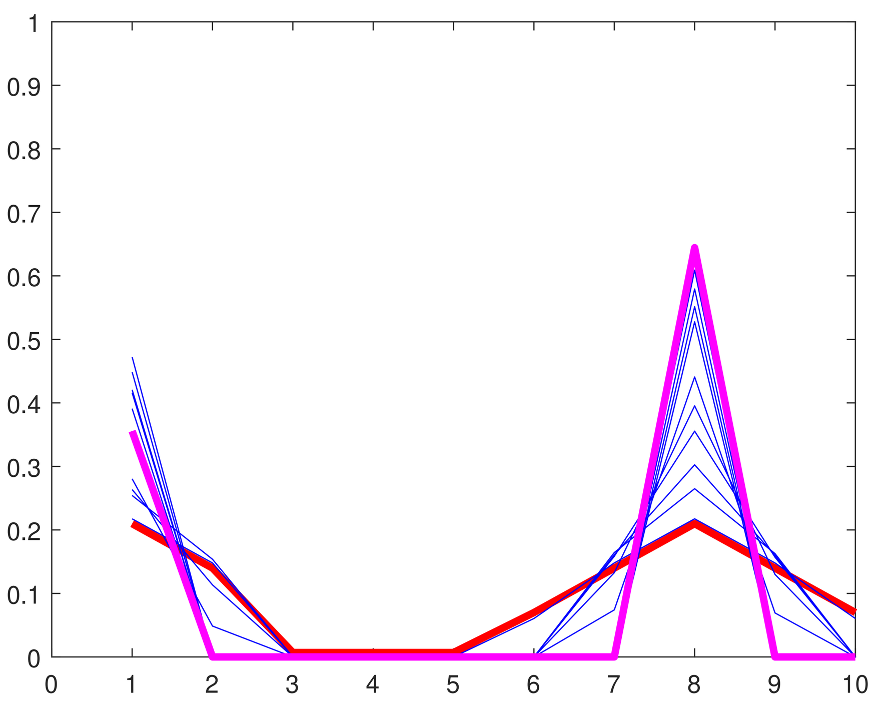
Method

Algorithm 1: O.T. Super Resolution Clustering [2]

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
Begin:
$K = 0$
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

Example

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



Plot of super resolution O.T. method. Red line is initial distribution. Blue lines are steps along gradient. Pink line is final, converged distribution. Sparsity level $\lambda = 10$. Solution $\nu = (0.35, 0, 0, 0, 0, 0, 0, 0.65, 0, 0)$.

Star Clustering Application

- The formation and evolution of star clusters[1]
- 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 [1]
- Neural networks are notoriously computationally expensive, sensitive to noise, and inflexible to appending or removing data variables

Confusion matrix of O.T. method on LEGUS data compared to StarcNet [1]. Column gives StarcNet classification and row gives O.T. classification [2]:

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

References

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

[2] M. Rawson and J. Hultgren. Optimal transport for super resolution applied to astronomy imaging. *Proceedings of EUSIPCO, 2022*, 2022.