Computational Optimal Transportation

Michael Rawson

November 29, 2021

Sinkhorn Algorithm

Regularized optimal plan $P = \min_{P'} \sum_{i,j} C_{i,j} P'_{i,j} - \epsilon H(P)$, where H is entropy.

Sinkhorn Algorithm

Regularized optimal plan $P = \min_{P'} \sum_{i,j} C_{i,j} P'_{i,j} - \epsilon H(P)$, where H is entropy.

Algorithm 2: Sinkhorn Algorithm

Input:

$s, t \in \mathbb{P}(\{1, ..., n\})$ are distributions

$C \in \mathbb{R}^{n \times n}$ is cost matrix

$\epsilon : 0 < \epsilon << 1$

$M \in \mathcal{N}$ is max iterations

Sinkhorn Algorithm

Regularized optimal plan $P = \min_{P'} \sum_{i,j} C_{i,j} P'_{i,j} - \epsilon H(P)$, where H is entropy.

Algorithm 3: Sinkhorn Algorithm

Input:

$s, t \in \mathbb{P}(\{1, ..., n\})$ are distributions
$C \in \mathbb{R}^{n \times n}$ is cost matrix
$\epsilon : 0 < \epsilon << 1$
$M \in \mathcal{N}$ is max iterations

Output:

$d \in \mathbb{R}$

Sinkhorn Algorithm

Regularized optimal plan $P = \min_{P'} \sum_{i,j} C_{i,j} P'_{i,j} - \epsilon H(P)$, where H is entropy.

Algorithm 4: Sinkhorn Algorithm

Input:

$s, t \in \mathbb{P}(\{1, ..., n\})$ are distributions

$C \in \mathbb{R}^{n \times n}$ is cost matrix

$\epsilon : 0 < \epsilon << 1$

$M \in \mathcal{N}$ is max iterations

Output:

$d \in \mathbb{R}$

Begin:

$K = exp(-C/epsilon)$; $u = v = ones(n, 1)$

$T = diag(u) \cdot K \cdot diag(v)$

for *i = 1,...,M* **do**

$u = s/(Kv)$

$v = t/(K^T u)$

$T = diag(u) \cdot K \cdot diag(v)$

end

$d = \|C \cdot T\|_{fro}$

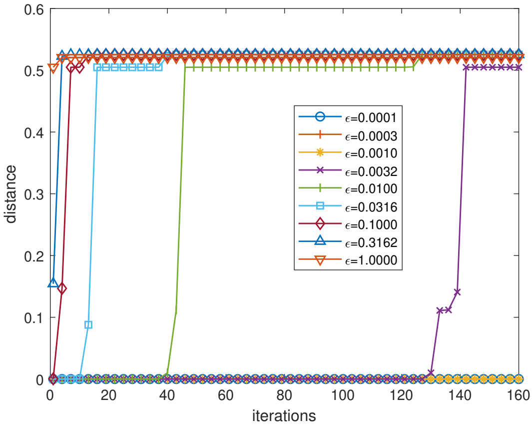
Sinkhorn Algorithm

Sinkhorn Algorithm

Sinkhorn Convergence over Iterations:

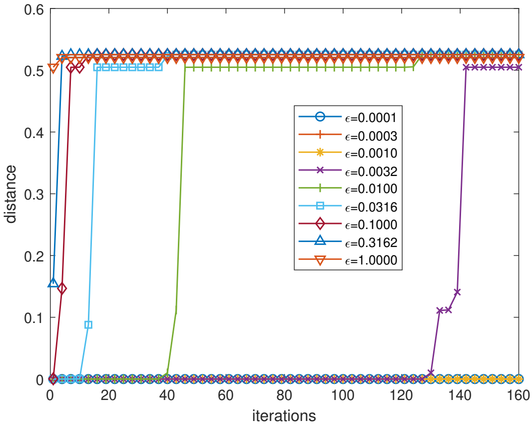
Sinkhorn Algorithm

Sinkhorn Convergence over Iterations:



Sinkhorn Algorithm

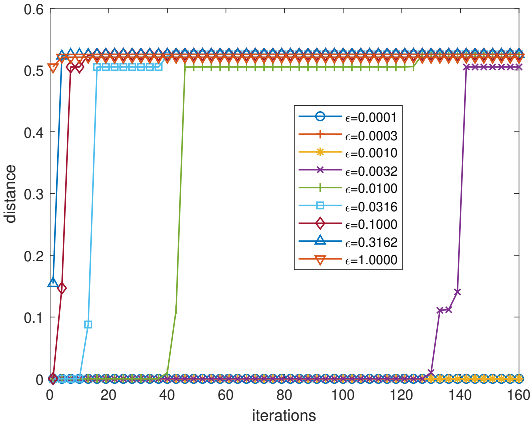
Sinkhorn Convergence over Iterations:



$$x, y \in \mathbb{P}(\{1, ..., 5\});\;\; x, y \sim \chi_1^2 \text{ normalized},\;\; C(i,j) = |i - j|$$

Sinkhorn Algorithm

Sinkhorn Convergence over Iterations:

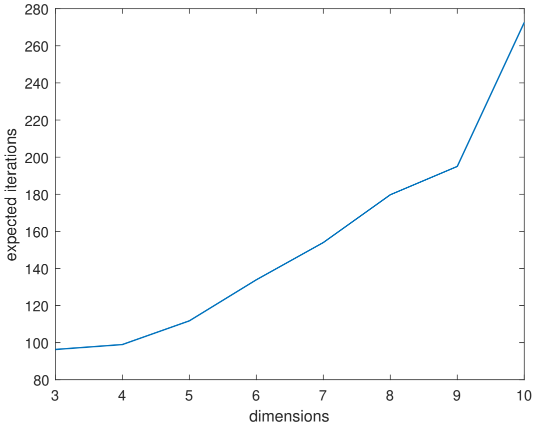


$$x, y \in \mathbb{P}(\{1, ..., 5\});\ \ x, y \sim \chi_1^2 \text{ normalized},\ \ C(i,j) = |i - j|$$
$$\|x - y\|_1 = 0.8270, \|x - y\|_2 = 0.4964, \|x - y\|_\infty = 0.4135$$

Sinkhorn Algorithm

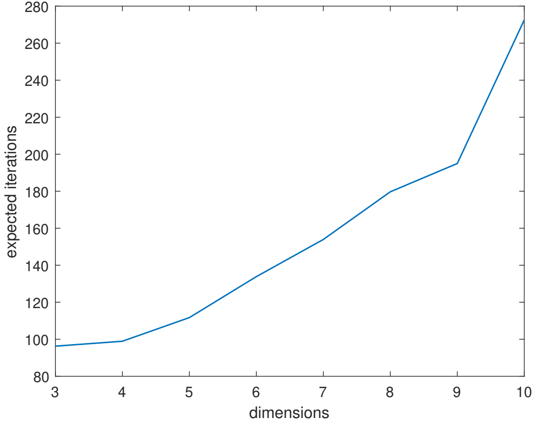
Sinkhorn Convergence over Dimensions:

Sinkhorn Convergence over Dimensions:



Sinkhorn Algorithm

Sinkhorn Convergence over Dimensions:



$$x, y \in \mathbb{P}(\{1, \ldots, n\});\ \ x, y \sim \chi_1^2 \text{ normalized},\ \epsilon = 0.01,\ C(i, j) = |i - j|$$

Sinkhorn Application

Sinkhorn Application

Removing Noise from Signals

Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance

Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

Sinkhorn Application

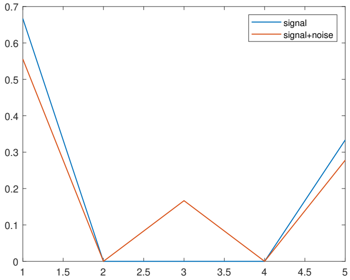
Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$
$H(v) = -\sum v_i \log(v_i)$

Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$
$H(v) = -\sum v_i \log(v_i)$
$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$
$H(v) = -\sum v_i \log(v_i)$
$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$



Removing Noise from Signals

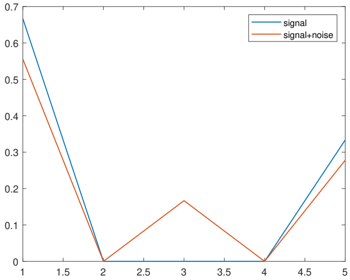
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

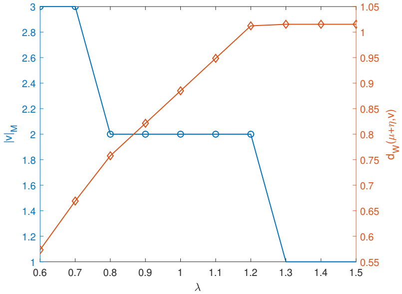
Noise is high entropy a.s. since independent with positive variance

Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

$H(v) = -\sum v_i \log(v_i)$

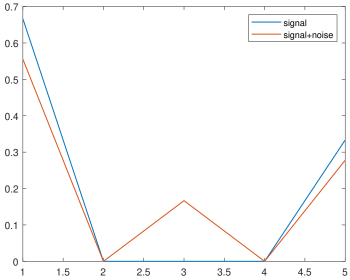
$\mu \in \mathbb{P}(\{1, \ldots, 5\}), \mu = (0.65, 0, \ldots, 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

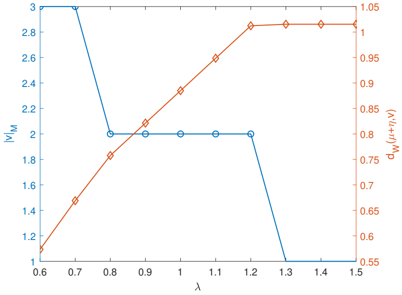




Sinkhorn Application

Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s. since independent with positive variance
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$
$H(v) = -\sum v_i \log(v_i)$
$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$





$$\mu + \eta \in \mathbb{P}(\{1, ..., 5\}), \epsilon = 0.01, \; C(i, j) = |i - j|$$

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*,*

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$, *the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2, *both must agree in sparsity (*$\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$*)*

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ $(\|\mu\|_0 < n)$ *with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$, *the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2, *both must agree in sparsity* $(\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0)$

Remark: Noise is high entropy a.s.

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2*, both must agree in sparsity (*$\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$*)*

Remark: Noise is high entropy a.s. Minimize entropy of $\mu + \eta$.

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu+\eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2*, both must agree in sparsity (*$\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$*)*

Remark: Noise is high entropy a.s. Minimize entropy of $\mu + \eta$. If noise is small enough, the noise can be removed, recovering sparse structure.

Proof:

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2*, both must agree in sparsity (*$\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$*)*

Remark: Noise is high entropy a.s. Minimize entropy of $\mu + \eta$. If noise is small enough, the noise can be removed, recovering sparse structure.

Proof: Apply triangle inequality to show feasible set does not have sparsity that μ does not have.

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ ($\|\mu\|_0 < n$) with added noise $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$, the solution of $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ will identify the structure of μ s.t. $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ if η is such that

1. *for $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$*
2. *for solutions v_1, v_2, both must agree in sparsity ($\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$)*

Remark: Noise is high entropy a.s. Minimize entropy of $\mu + \eta$. If noise is small enough, the noise can be removed, recovering sparse structure.

Proof: Apply triangle inequality to show feasible set does not have sparsity that μ does not have. We have μ is in the feasible set.

Optimal Transportation Reconstruction Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will identify the structure of* μ *s.t.* $\|\tilde{\mu} - \tilde{v}\|_0 = 0$ *if* η *is such that*

1. *for* $v' \in Ball(\mu, 2d_W(\mu, \mu + \eta)), \|v'\|_0 \geq \|\mu\|_0$
2. *for solutions* v_1, v_2*, both must agree in sparsity (*$\|\tilde{v}_1 - \tilde{v}_2\|_0 = 0$*)*

Remark: Noise is high entropy a.s. Minimize entropy of $\mu + \eta$. If noise is small enough, the noise can be removed, recovering sparse structure.

Proof: Apply triangle inequality to show feasible set does not have sparsity that μ does not have. We have μ is in the feasible set. Minimizing H maximizes sparsity. So v agrees with the sparsity of μ.

Optimal Transportation Reconstruction Failure

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s.

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s.
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals

Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

Noise is high entropy a.s.

Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals

Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

Noise is high entropy a.s.

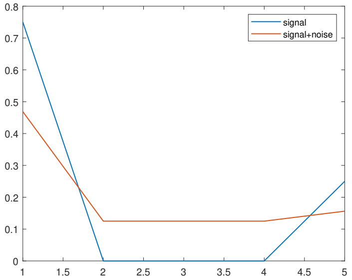
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

$\mu + \eta \in \mathbb{P}(\{1, ..., 5\}), \epsilon = 0.01, \ C(i,j) = |i - j|$

Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis
Noise is high entropy a.s.
Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$
$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$
$\mu + \eta \in \mathbb{P}(\{1, ..., 5\}), \epsilon = 0.01,\ C(i, j) = |i - j|$



Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals

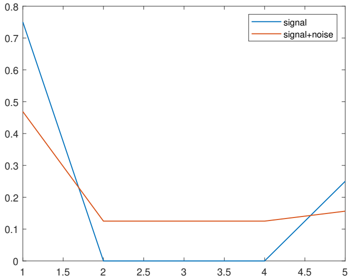
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

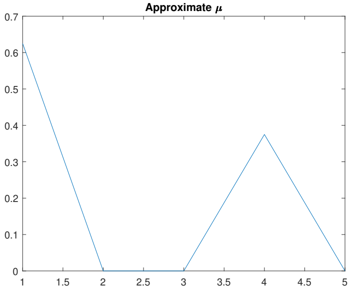
Noise is high entropy a.s.

Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

$\mu \in \mathbb{P}(\{1, \dots, 5\}), \mu = (0.65, 0, \dots, 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

$\mu + \eta \in \mathbb{P}(\{1, \dots, 5\}), \epsilon = 0.01, \; C(i,j) = |i - j|$





Optimal Transportation Reconstruction Failure

Failing Removing Noise from Signals

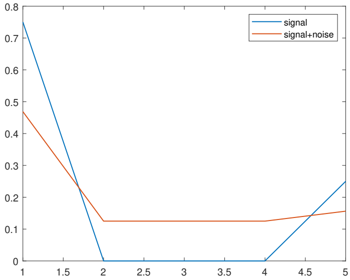
Signals have low entropy: $\mu \in \mathbb{R}^n$ is sparse in a localized basis

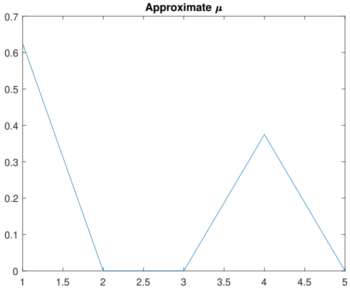
Noise is high entropy a.s.

Minimize entropy: $v = \arg\min_{v'} d_W(\mu + \eta, v') + \lambda H(v')$

$\mu \in \mathbb{P}(\{1, ..., 5\}), \mu = (0.65, 0, ..., 0, 0.35), \|\mu\|_M = rank(H_0(\mu^{-1}([0.13, 1]))) = 2$

$\mu + \eta \in \mathbb{P}(\{1, ..., 5\}), \epsilon = 0.01, \; C(i, j) = |i - j|$





Approximate signal (left: lambda=[0.3, 0.7]) not equal to true signal for any lambda.

Optimal Transportation Failure Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Optimal Transportation Failure Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*,*

Optimal Transportation Failure Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will not agree with* μ*, that is* $\|\tilde{\mu} - \tilde{v}\|_0 > 0$*, if*

Optimal Transportation Failure Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will not agree with* μ*, that is* $\|\tilde{\mu} - \tilde{v}\|_0 > 0$*, if* $d_W(\mu, \mu + \eta) > d_W(v', \mu + \eta)$ *with* $\|v'\|_0 \leq \|\mu\|_0$ *and* $\|\tilde{v}' - \tilde{\mu}\|_0 > 0$*.*

Optimal Transportation Failure Theorem

Let $\tilde{w} = sign(w)\ \forall w \in \mathbb{R}^n$.

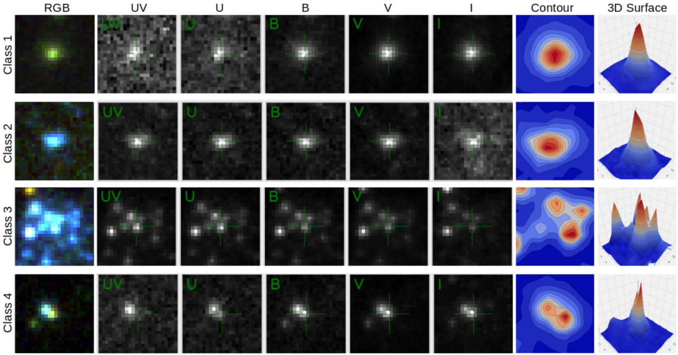
Theorem

Given a sparse signal $\mu \in \mathbb{P}(\{1, ..., n\})$ *(*$\|\mu\|_0 < n$*) with added noise* $\eta : \mu + \eta \in \mathbb{P}(\{1, ..., n\})$*, the solution of* $v = \arg\min_{v' : d_W(\mu+\eta, v') \leq d_W(\mu, \mu+\eta)} H(v')$ *will not agree with* μ*, that is* $\|\tilde{\mu} - \tilde{v}\|_0 > 0$*, if* $d_W(\mu, \mu+\eta) > d_W(v', \mu+\eta)$ *with* $\|v'\|_0 \leq \|\mu\|_0$ *and* $\|\tilde{v}' - \tilde{\mu}\|_0 > 0$*.*

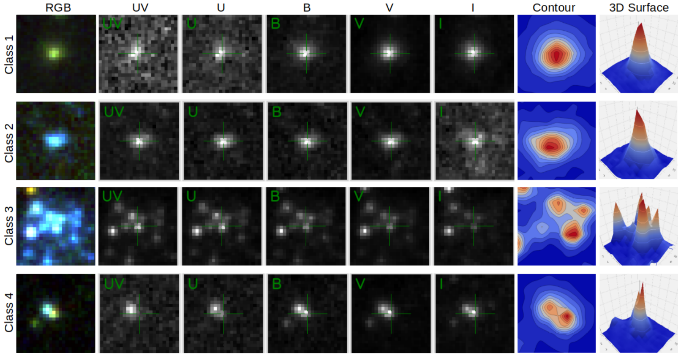
Remark: When the signal-to-noise ratio (SNR) is too low, the reconstruction is underdetermined.

Application: Star Cluster Identification

Application: Star Cluster Identification

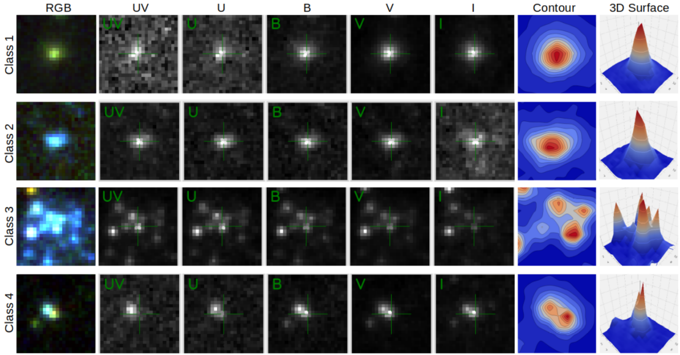


Application: Star Cluster Identification



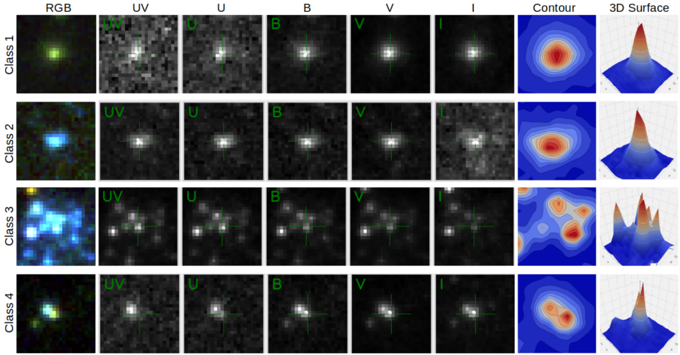
From [1], LEGUS classification scheme. Examples of candidates from the LEGUS images of NGC 1566 classified as Class 1 (symmetric star cluster; top),

Application: Star Cluster Identification



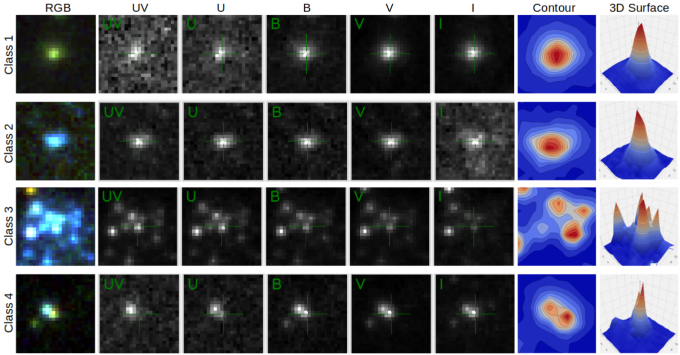
From [1], LEGUS classification scheme. Examples of candidates from the LEGUS images of NGC 1566 classified as Class 1 (symmetric star cluster; top), Class 2 (elongated star cluster; middle-top),

Application: Star Cluster Identification



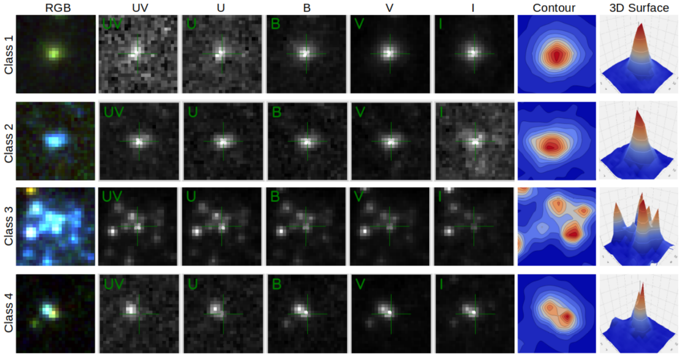
From [1], LEGUS classification scheme. Examples of candidates from the LEGUS images of NGC 1566 classified as Class 1 (symmetric star cluster; top), Class 2 (elongated star cluster; middle-top), Class 3 (compact, multi–peak association; middle-bottom), and

Application: Star Cluster Identification



From [1], LEGUS classification scheme. Examples of candidates from the LEGUS images of NGC 1566 classified as Class 1 (symmetric star cluster; top), Class 2 (elongated star cluster; middle-top), Class 3 (compact, multi–peak association; middle-bottom), and Class 4 (spurious object; bottom).

Application: Star Cluster Identification



From [1], LEGUS classification scheme. Examples of candidates from the LEGUS images of NGC 1566 classified as Class 1 (symmetric star cluster; top), Class 2 (elongated star cluster; middle-top), Class 3 (compact, multi–peak association; middle-bottom), and Class 4 (spurious object; bottom). The three–color image to the left is created using the NUV and U bands for the blue channel, the B band for green one, and the V and I bands for red one. The contour and 3D plots from the V–band are shown to the right of the figure.

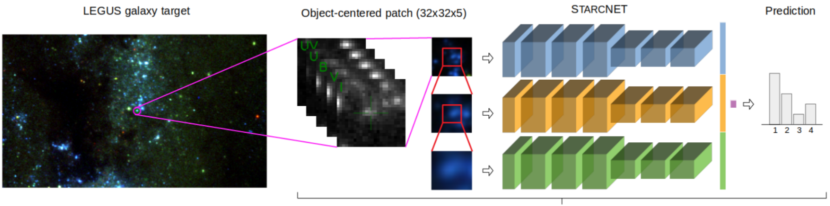
Star Cluster Identification

Star Cluster Identification

Convolutional Neural Network:

Star Cluster Identification

Convolutional Neural Network:

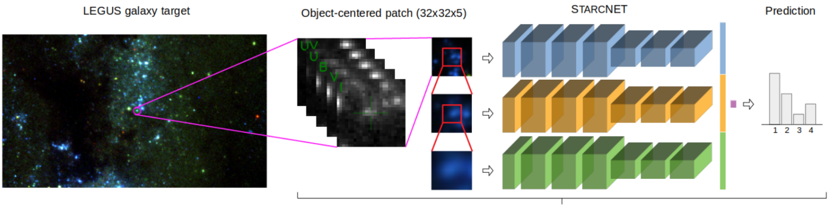


(a) Objects detected with SExtractor on HST observations

(b) Overview of the pipeline

Star Cluster Identification

Convolutional Neural Network:

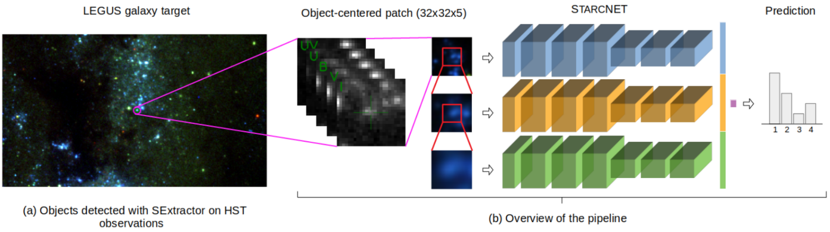


(a) Objects detected with SExtractor on HST observations

(b) Overview of the pipeline

From [1], the StarcNet pipeline.

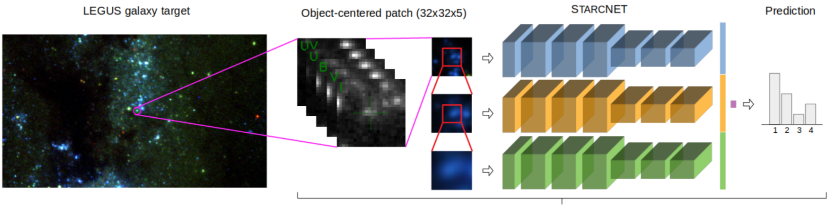
Convolutional Neural Network:



From [1], the StarcNet pipeline. Graphic sketch of the machine learning pipeline used in this work to classify cluster candidates in the LEGUS images.

Star Cluster Identification

Convolutional Neural Network:



(a) Objects detected with SExtractor on HST observations

(b) Overview of the pipeline

From [1], the StarcNet pipeline. Graphic sketch of the machine learning pipeline used in this work to classify cluster candidates in the LEGUS images. (Left): The Hubble Space Telescope images as processed by the LEGUS project through a custom pipeline to generate automatic catalogs of cluster candidates, which are part of the public LEGUS catalogs release (Calzetti et al. 2015; Adamo et al. 2017); we apply StarcNet to the LEGUS catalogs and images.

Star Cluster Identification

Star Cluster Identification

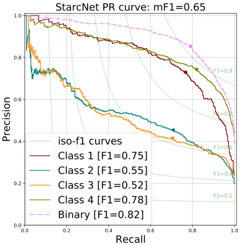
Convolutional Neural Network:

Star Cluster Identification

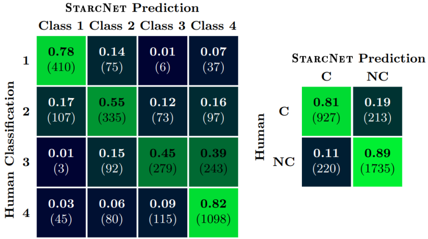
Convolutional Neural Network:

Human Classification \ StarcNet Prediction	Class 1	Class 2	Class 3	Class 4
1	0.78 (410)	0.14 (75)	0.01 (6)	0.07 (37)
2	0.17 (107)	0.55 (335)	0.12 (73)	0.16 (97)
3	0.01 (3)	0.15 (92)	0.45 (279)	0.39 (243)
4	0.03 (45)	0.06 (80)	0.09 (115)	0.82 (1098)

Human \ StarcNet Prediction	C	NC
C	0.81 (927)	0.19 (213)
NC	0.11 (220)	0.89 (1735)

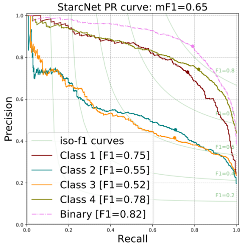


Convolutional Neural Network:



Human Classification \ StarcNet Prediction	Class 1	Class 2	Class 3	Class 4
1	0.78 (410)	0.14 (75)	0.01 (6)	0.07 (37)
2	0.17 (107)	0.55 (335)	0.12 (73)	0.16 (97)
3	0.01 (3)	0.15 (92)	0.45 (279)	0.39 (243)
4	0.03 (45)	0.06 (80)	0.09 (115)	0.82 (1098)

Human \ StarcNet Prediction	C	NC
C	0.81 (927)	0.19 (213)
NC	0.11 (220)	0.89 (1735)



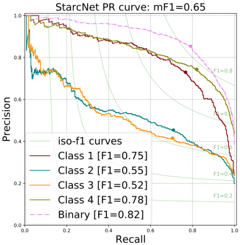
From [1], performance of StarcNet on the test set.

Star Cluster Identification

Convolutional Neural Network:

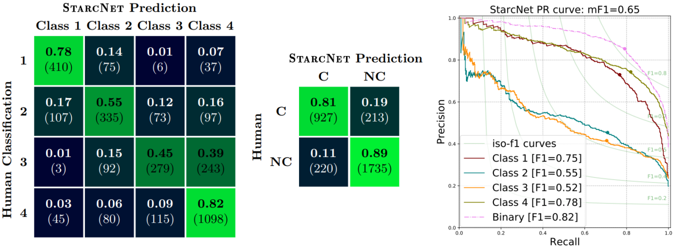
Human Classification \ StarcNet Prediction	Class 1	Class 2	Class 3	Class 4
1	0.78 (410)	0.14 (75)	0.01 (6)	0.07 (37)
2	0.17 (107)	0.55 (335)	0.12 (73)	0.16 (97)
3	0.01 (3)	0.15 (92)	0.45 (279)	0.39 (243)
4	0.03 (45)	0.06 (80)	0.09 (115)	0.82 (1098)

Human \ StarcNet Prediction	C	NC
C	0.81 (927)	0.19 (213)
NC	0.11 (220)	0.89 (1735)



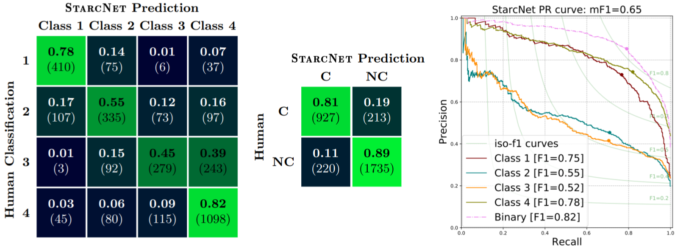
From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects).

Convolutional Neural Network:



From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects). The rows show the distribution of the human–classified sources, while the columns are the predictions of StarcNet. Parenthesis in the confusion matrix refers to the unnormalized values.

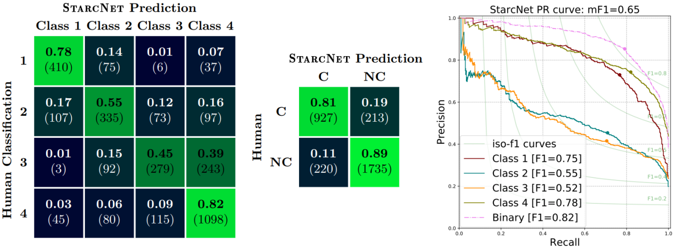
Convolutional Neural Network:



From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects). The rows show the distribution of the human–classified sources, while the columns are the predictions of StarcNet. Parenthesis in the confusion matrix refers to the unnormalized values. (Left) Overall accuracy evaluated for 4 class classification using raw bands as input.

Star Cluster Identification

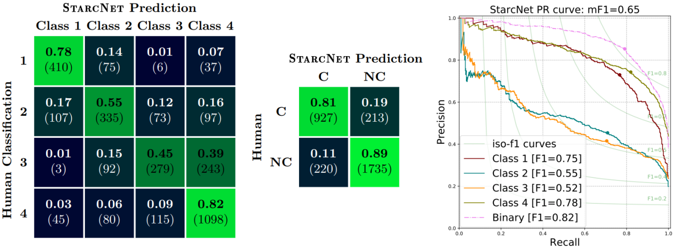
Convolutional Neural Network:



From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects). The rows show the distribution of the human–classified sources, while the columns are the predictions of StarcNet. Parenthesis in the confusion matrix refers to the unnormalized values. (Left) Overall accuracy evaluated for 4 class classification using raw bands as input. (Middle) Results calculated with 2 classes (cluster/non-cluster classification).

Star Cluster Identification

Convolutional Neural Network:

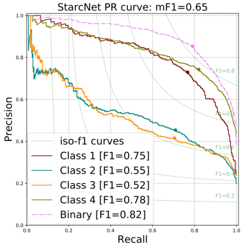


From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects). The rows show the distribution of the human–classified sources, while the columns are the predictions of StarcNet. Parenthesis in the confusion matrix refers to the unnormalized values. (Left) Overall accuracy evaluated for 4 class classification using raw bands as input. (Middle) Results calculated with 2 classes (cluster/non-cluster classification). (Right) Precision-recall curves for each of the four classes, as well as for binary classification.

Convolutional Neural Network:

Human Classification \ StarcNet Prediction	Class 1	Class 2	Class 3	Class 4
1	0.78 (410)	0.14 (75)	0.01 (6)	0.07 (37)
2	0.17 (107)	0.55 (335)	0.12 (73)	0.16 (97)
3	0.01 (3)	0.15 (92)	0.45 (279)	0.39 (243)
4	0.03 (45)	0.06 (80)	0.09 (115)	0.82 (1098)

Human \ StarcNet Prediction	C	NC
C	0.81 (927)	0.19 (213)
NC	0.11 (220)	0.89 (1735)



From [1], performance of StarcNet on the test set. Confusion matrix normalized over the classes in test set of the LEGUS dataset (20% of the total sources or about 3000 objects). The rows show the distribution of the human–classified sources, while the columns are the predictions of StarcNet. Parenthesis in the confusion matrix refers to the unnormalized values. (Left) Overall accuracy evaluated for 4 class classification using raw bands as input. (Middle) Results calculated with 2 classes (cluster/non-cluster classification). (Right) Precision-recall curves for each of the four classes, as well as for binary classification. The overall accuracy is 68.6% with 4 classes and 86.0% with binary classification.

Optimal Transportation Star Cluster Identification

Our method:

Optimal Transportation Star Cluster Identification

Our method:

Algorithm 6: Optimal Transportation Star Cluster Identification

Input:

X : images $N \times m \times m$

δ : noise level

Our method:

Algorithm 7: Optimal Transportation Star Cluster Identification

Input:

X : images $N \times m \times m$

δ : noise level

Output:

C : star cluster classification $N \times 1$

Optimal Transportation Star Cluster Identification

Our method:

Algorithm 8: Optimal Transportation Star Cluster Identification

Input:

X : images $N \times m \times m$

δ : noise level

Output:

C : star cluster classification $N \times 1$

for $n = 1, 2, ..., N$ **do**

$V_n = \arg\min_{v' : d_W(X_n, v') < \delta} H(v')$

Optimal Transportation Star Cluster Identification

Our method:

Algorithm 9: Optimal Transportation Star Cluster Identification

Input:

X : images $N \times m \times m$

δ : noise level

Output:

C : star cluster classification $N \times 1$

for $n = 1, 2, ..., N$ **do**

  $V_n = \arg\min_{v' : d_W(X_n, v') < \delta} H(v')$

  if $rank(H_0(V_n^{-1}([0.9, 1]))) == 1$ **then**

Optimal Transportation Star Cluster Identification

Our method:

Algorithm 10: Optimal Transportation Star Cluster Identification

Input:
X : images $N \times m \times m$
δ : noise level
Output:
C : star cluster classification $N \times 1$
for $n = 1, 2, ..., N$ **do**
  $V_n = \arg\min_{v' : d_W(X_n, v') < \delta} H(v')$
  if $rank(H_0(V_n^{-1}([0.9, 1]))) == 1$ **then**
    $C_n = 1$
  else
    $C_n = 0$
  end
end

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v' : d_W(x, v') < \delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
    1. Converges first order (Taylors Remainder)

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
    1. Converges first order (Taylors Remainder)
    2. Computationally slow and exponentially slow in higher dimensions

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
    1. Converges first order (Taylors Remainder)
    2. Computationally slow and exponentially slow in higher dimensions
2. Random search:

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
    1. Converges first order (Taylors Remainder)
    2. Computationally slow and exponentially slow in higher dimensions
2. Random search:
    1. Converges with high probability

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
   1. Converges first order (Taylors Remainder)
   2. Computationally slow and exponentially slow in higher dimensions
2. Random search:
   1. Converges with high probability
   2. Exponentially slow in higher dimensions

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
    1. Converges first order (Taylors Remainder)
    2. Computationally slow and exponentially slow in higher dimensions
2. Random search:
    1. Converges with high probability
    2. Exponentially slow in higher dimensions
3. Gradient Descent:

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v' : d_W(x,v') < \delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
   1. Converges first order (Taylors Remainder)
   2. Computationally slow and exponentially slow in higher dimensions
2. Random search:
   1. Converges with high probability
   2. Exponentially slow in higher dimensions
3. Gradient Descent:
   1. Converges quickly to local minimum (usually)

Optimal Transportation Star Cluster Identification

Evaluating the expression $\arg\min_{v':d_W(x,v')<\delta} H(v')$
Note: Center the problem at x and add small vectors to stay in feasibility region.
Normalize vectors to stay in probability space.

1. Uniform grid search:
   1. Converges first order (Taylors Remainder)
   2. Computationally slow and exponentially slow in higher dimensions
2. Random search:
   1. Converges with high probability
   2. Exponentially slow in higher dimensions
3. Gradient Descent:
   1. Converges quickly to local minimum (usually)
   2. Misses global minimum and can even diverge

Results:
Preliminary results on class 2 cluster detection:

Optimal Transportation Star Cluster Identification

Results:
Preliminary results on class 2 cluster detection:
$precision = \frac{tp}{tp+fp} = 1$
$recall = \frac{tp}{tp+fn} = 2/5$
$F_1 = 2\frac{precision \cdot recall}{precision+recall} = \frac{4}{7} = 57\%$

Optimal Transportation Star Cluster Identification

Results:
Preliminary results on class 2 cluster detection:
$precision = \frac{tp}{tp+fp} = 1$
$recall = \frac{tp}{tp+fn} = 2/5$
$F_1 = 2\frac{precision \cdot recall}{precision+recall} = \frac{4}{7} = 57\%$

Future Work:

Optimal Transportation Star Cluster Identification

Results:
Preliminary results on class 2 cluster detection:

$$precision = \frac{tp}{tp+fp} = 1$$

$$recall = \frac{tp}{tp+fn} = 2/5$$

$$F_1 = 2\frac{precision \cdot recall}{precision + recall} = \frac{4}{7} = 57\%$$

Future Work:

1. Investigate transporting data across spectrum, not just spacially.

Optimal Transportation Star Cluster Identification

Results:
Preliminary results on class 2 cluster detection:
$precision = \frac{tp}{tp+fp} = 1$
$recall = \frac{tp}{tp+fn} = 2/5$
$F_1 = 2\frac{precision \cdot recall}{precision+recall} = \frac{4}{7} = 57\%$

Future Work:

1. Investigate transporting data across spectrum, not just spacially.
2. Estimate SNR parameter.

Optimal Transportation Star Cluster Identification

Results:
Preliminary results on class 2 cluster detection:
$precision = \frac{tp}{tp+fp} = 1$
$recall = \frac{tp}{tp+fn} = 2/5$
$F_1 = 2\frac{precision \cdot recall}{precision+recall} = \frac{4}{7} = 57\%$

Future Work:

1. Investigate transporting data across spectrum, not just spacially.
2. Estimate SNR parameter.
3. Speedup on GPU.

Bibliography

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.