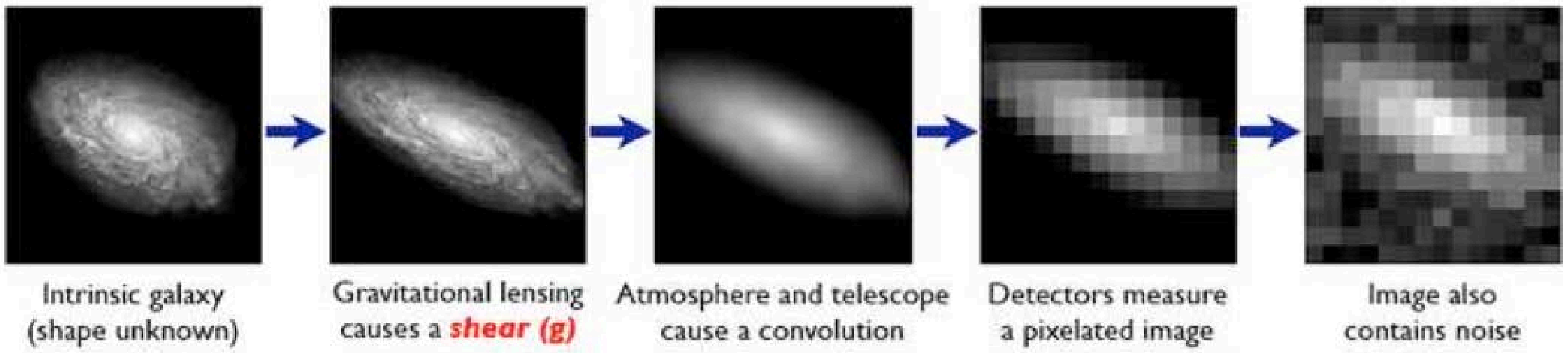
Optimal Transport and PSF Modeling

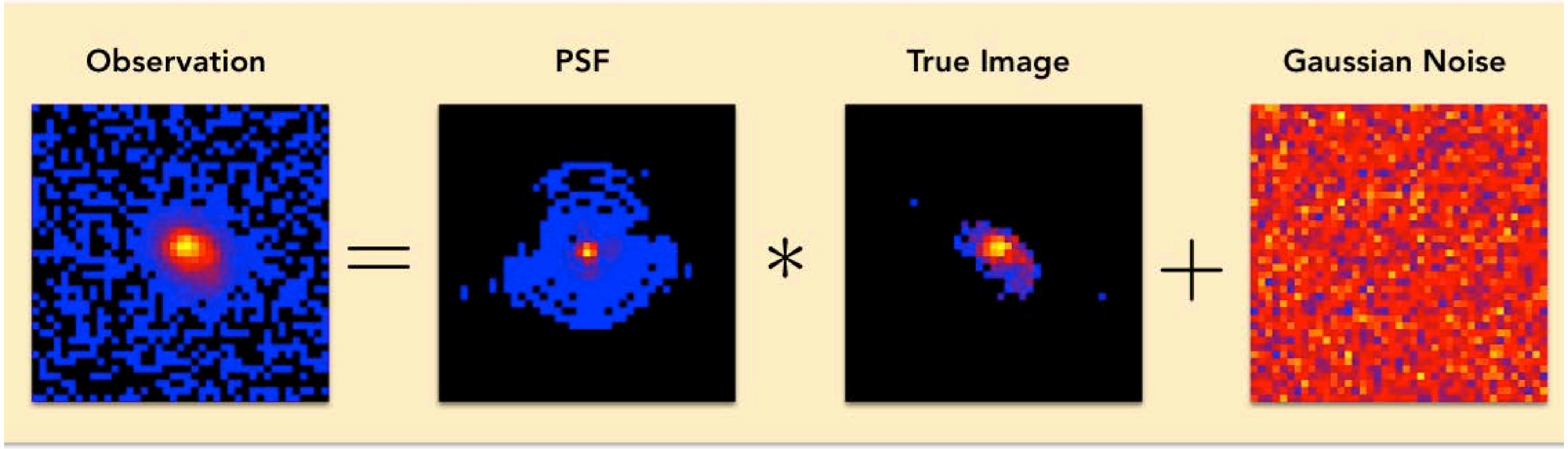
Rebeca Araripe, Morgan A. Schmitz, Fred Ngolè, Jean-Luc Starck

French-Chinese Days on Weak Lensing

05/10/2018

Motivation


Intrinsic galaxy
(shape unknown)
Gravitational lensing
causes a shear (g)
Atmosphere and telescope
cause a convolution
Detectors measure
a pixelated image
Image also
contains noise


Observation
PSF
True Image
Gaussian Noise

Inverse problem

Dependency on position

$$\min_{X} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_2^2$$

- $\boldsymbol{X} \in \mathbb{R}_+^{N \times K}$: true PSFs at different positions of the FOV.
- $\boldsymbol{M} = [\boldsymbol{M}_1, \cdots, \boldsymbol{M}_K]$, $\boldsymbol{M}_k \in \mathbb{R}_+^{N' \times N}$: decimation and shifting.
- $\boldsymbol{Y} \in \mathbb{R}_+^{N' \times K}$: observed PSFs.

Inverse problem

Dependency on position

$$\min_{X} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_2^2$$

- $\boldsymbol{X} \in \mathbb{R}_+^{N \times K}$: true PSFs at different positions of the FOV.
- $\boldsymbol{M} = [\boldsymbol{M}_1, \cdots, \boldsymbol{M}_K],\ \boldsymbol{M}_k \in \mathbb{R}_+^{N' \times N}$: decimation and shifting.
- $\boldsymbol{Y} \in \mathbb{R}_+^{N' \times K}$: observed PSFs.

Fig: PSF variation for different wavelengths (400 to 900 nm)

Inverse problem

Dependency on position and wavelength

Spectral energy density

$$\min_{(\boldsymbol{X}_v)_{1\leq v\leq m}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{M}\sum_v \beta_v \boldsymbol{X}_v\|_2^2$$

- $\boldsymbol{X}_v \in \mathbb{R}_+^{N\times K}$: true PSFs at different positions of the FOV.
- $\boldsymbol{M} = [\boldsymbol{M}_1, \cdots, \boldsymbol{M}_K]$, $\boldsymbol{M}_k \in \mathbb{R}_+^{N'\times N}$: decimation and shifting.
- $\boldsymbol{Y} \in \mathbb{R}_+^{N'\times K}$: observed PSFs.

Fig: PSF variation for different wavelengths (400 to 900 nm)

Optimal Transport

The Monge-Kantorovich problem

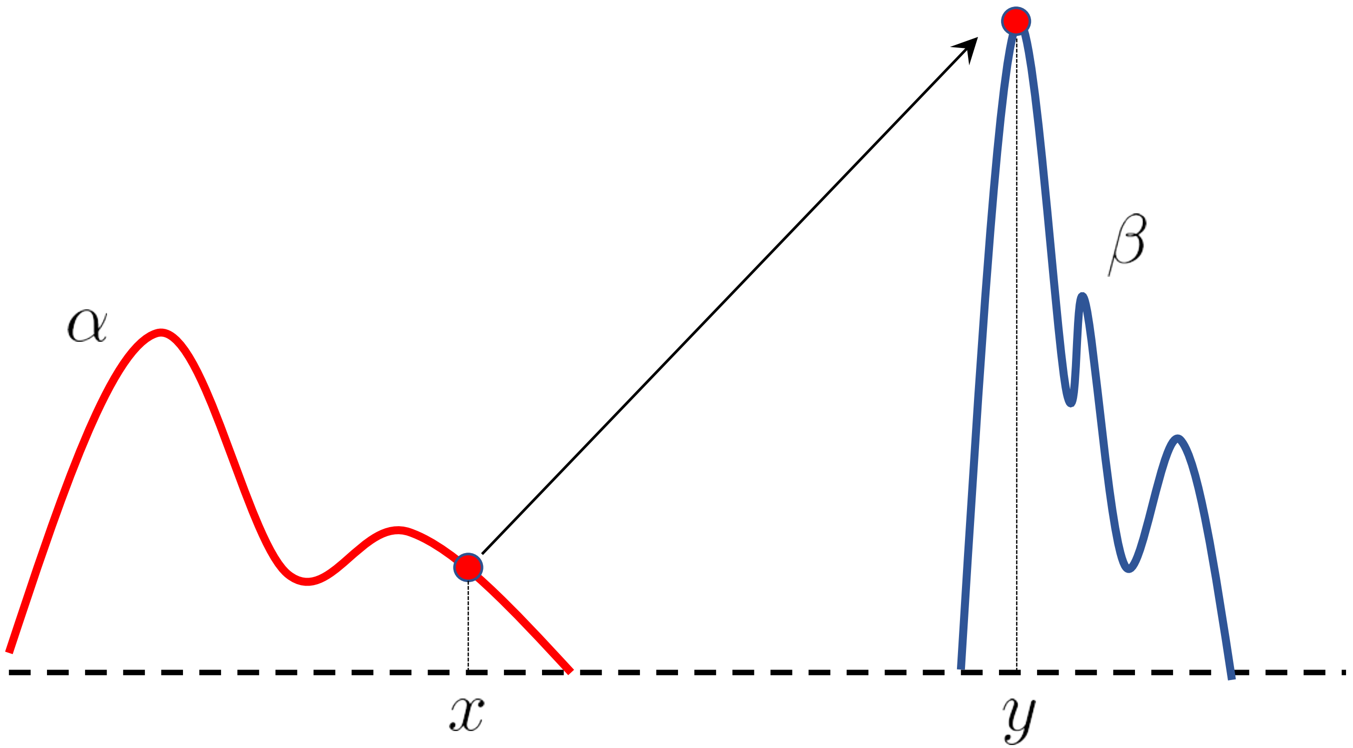


Fig: Graphical representation of mass transportation problem: find the optimal way of transporting a heap of sand α into the shape of a heap of sand β, knowing the cost of moving grains of sand to and from any position.

Optimal Transport

Discrete formalism

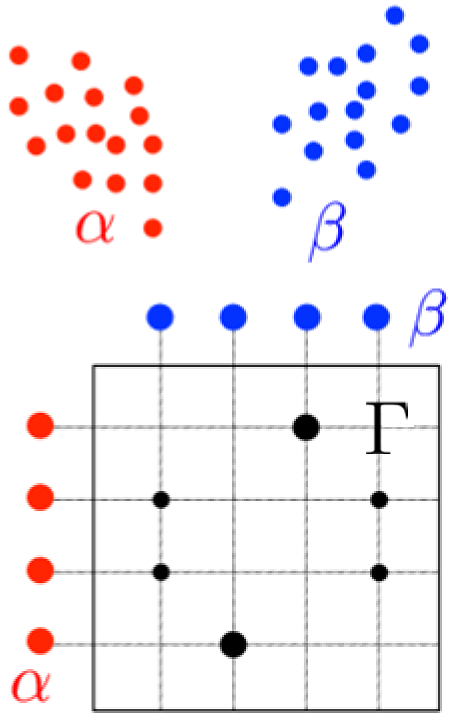


Fig: coupling matrix between two discrete distributions.[1]

$$\boldsymbol{\alpha} = \sum_{i=1}^{N} \alpha_i \delta_{\boldsymbol{p}_i} \qquad \boldsymbol{\beta} = \sum_{i=1}^{N} \beta_i \delta_{\boldsymbol{p}_i}$$

- $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \sum_N \qquad \sum_N := \left\{ \boldsymbol{d} \in \mathbb{R}_+^N ; \sum_i d_i = 1 \right\}$
- $[\boldsymbol{p}_1, \cdots, \boldsymbol{p}_N] \in \mathbb{R}^{d \times N}$

Transport plan (coupling matrix)

$$\boldsymbol{\Gamma} = (\gamma_{i \to j})_{1 \leq i,j \leq N}$$

Monge-Kantorovich problem

$$\min_{\boldsymbol{\Gamma} \in \boldsymbol{\Pi}} < \boldsymbol{C}, \boldsymbol{\Gamma} >$$

Cost matrix

- $\boldsymbol{\Gamma} \in \boldsymbol{\Pi}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad , \qquad \boldsymbol{\Pi}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left\{ \boldsymbol{\Gamma} \in \mathbb{R}_+^{N \times N} \;\middle|\; \boldsymbol{\Gamma} \mathbb{1}_N = \boldsymbol{\alpha}, \quad \boldsymbol{\Gamma}^{\mathrm{T}} \mathbb{1}_N = \boldsymbol{\beta} \right\}$

[1] Cuturi & Peyré, Computational Optimal Transport

Optimal Transport

Discrete formalism

- Given that $C_{i,j} = [d(i,j)]^d$ where $d(i,j)$ is a distance:

Wasserstein distance

$$W_2(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \left(\min_{\boldsymbol{\Gamma} \in \boldsymbol{\Pi}(\boldsymbol{\alpha}, \boldsymbol{\beta})} < \boldsymbol{C}, \boldsymbol{\Gamma} > \right)^{1/2}$$

Optimal Transport

Displacement interpolation

- Given that $C_{i,j} = [d(i,j)]^d$ where $d(i,j)$ is a distance:

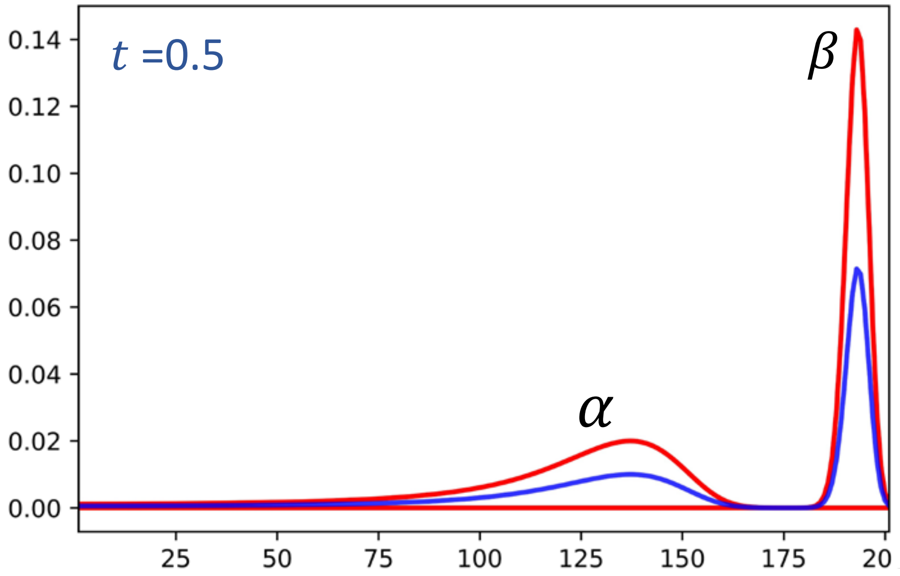
Wasserstein distance

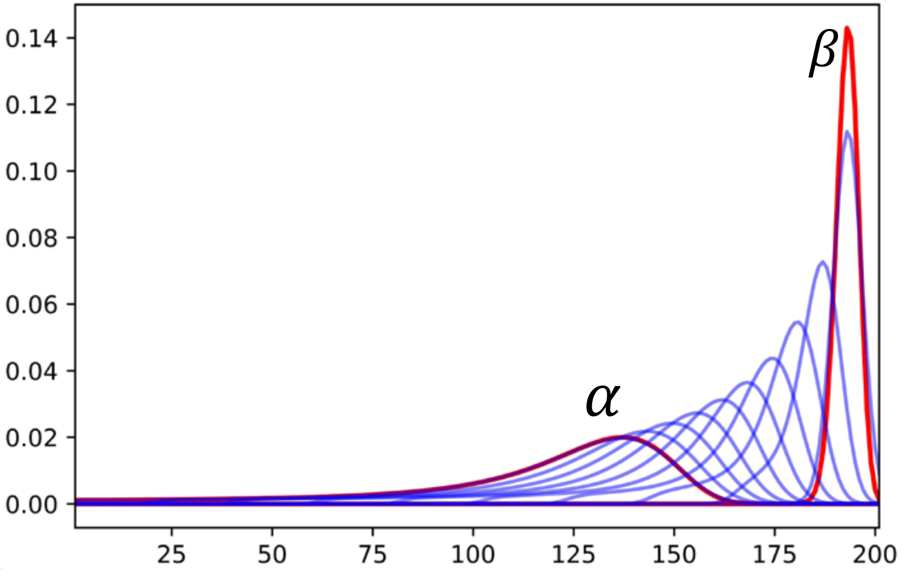
$$W_2(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \left(\min_{\boldsymbol{\Gamma} \in \boldsymbol{\Pi}(\boldsymbol{\alpha}, \boldsymbol{\beta})} < \boldsymbol{C}, \boldsymbol{\Gamma} > \right)^{1/2}$$

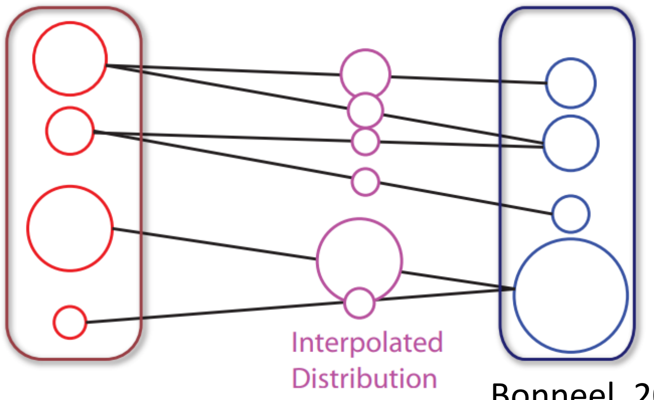
Wasserstein barycenter

$$\boldsymbol{P}((\boldsymbol{d}_1, \cdots, \boldsymbol{d}_S), \boldsymbol{\lambda}) = \underset{\boldsymbol{u}}{\operatorname{argmin}} \sum_{i=1}^{S} \lambda_i W_2^2(\boldsymbol{u}, \boldsymbol{d}_i)$$

- $\boldsymbol{d}_i \in \sum_N$, $\boldsymbol{\lambda} \in \sum_S$







Bonneel, 2011

Optimal Transport

Displacement interpolation

ATTRACTIVE TOOL,
BUT COSTLY TO COMPUTE: $\mathcal{O}(N^3)$

- Given that $C_{i,j} = [d(i,j)]^d$ where $d(i,j)$ is a distance:

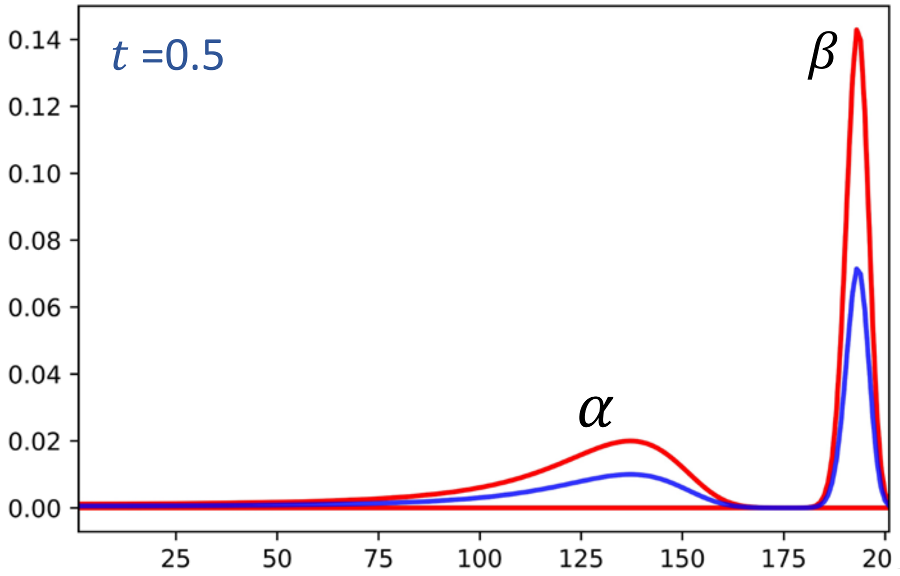
Wasserstein distance

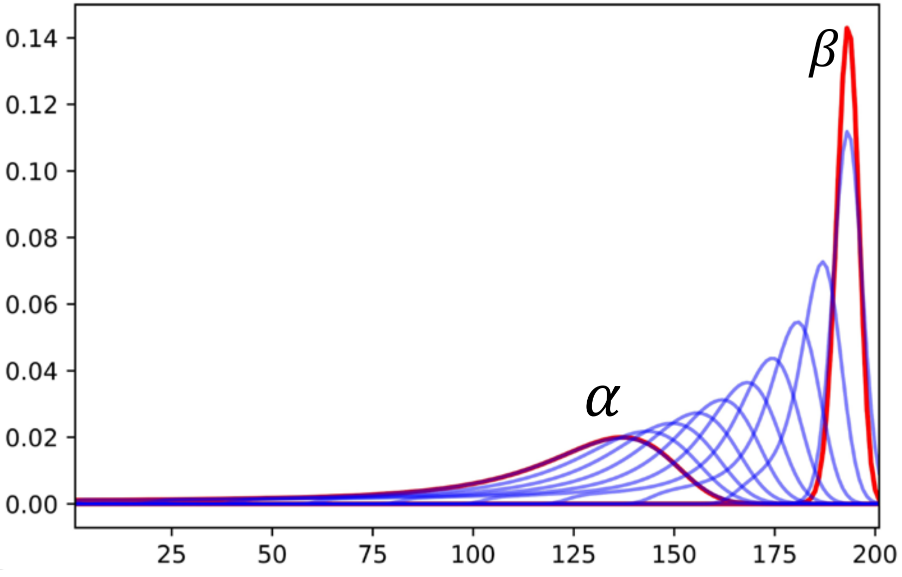
$$W_2(\boldsymbol{\alpha},\boldsymbol{\beta}) := \left(\min_{\boldsymbol{\Gamma}\in\boldsymbol{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta})} < \boldsymbol{C}, \boldsymbol{\Gamma} > \right)^{1/2}$$

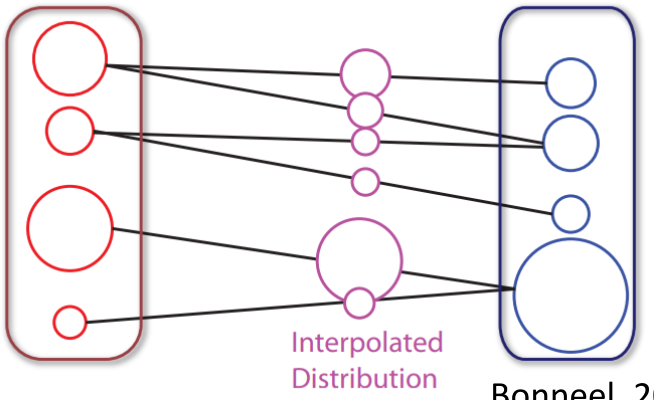
Wasserstein barycenter

$$\boldsymbol{P}((\boldsymbol{d}_1,\cdots,\boldsymbol{d}_S),\boldsymbol{\lambda}) = \operatorname*{argmin}_{\boldsymbol{u}} \sum_{i=1}^{S} \lambda_i W_2^2(\boldsymbol{u},\boldsymbol{d}_i)$$

- $\boldsymbol{d}_i \in \sum_N$, $\boldsymbol{\lambda} \in \sum_S$







Bonneel, 2011

Optimal Transport

Wasserstein distance approximations

Entropic regularization [2]

$$W_2^2(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \inf_{\boldsymbol{\Gamma} \in \boldsymbol{\Pi}(\boldsymbol{\alpha}, \boldsymbol{\beta})} < \boldsymbol{C}, \boldsymbol{\Gamma} > -\gamma H(\boldsymbol{\Gamma})$$

- $\gamma > 0$
- $H(\boldsymbol{\Gamma}) := -\sum_{i,j} \Gamma_{i,j} \log(\Gamma_{i,j} - 1)$

Advantages:

1. Problem γ–strongly convex.
2. Gain in stability. The greater γ, more smooth the coupling matrix.
3. Solution is a diagonal scaling of $e^{-\boldsymbol{C}/\gamma}$ $\rightarrow$ allows the use of Bregman's optimization tools, wich are easily paralellizable.

[2] Cuturi 2013

Optimal Transport

Wasserstein distance approximations

Entropic regularization: a Kulback-Leibler projection

$$W_2^2(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{\Gamma}\in\boldsymbol{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta})} \mathrm{KL}(\boldsymbol{\Gamma}|\boldsymbol{K})$$

- $\boldsymbol{K} = \exp(-\boldsymbol{C}/\gamma)$
- $\boldsymbol{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \underbrace{\{\boldsymbol{\Gamma}\in\mathbb{R}_+^{N\times N},\quad \boldsymbol{\Gamma}\mathbb{1}_N = \boldsymbol{\alpha}\}}_{\mathcal{C}^1} \cap \underbrace{\{\boldsymbol{\Gamma}\in\mathbb{R}_+^{N\times N},\quad \boldsymbol{\Gamma}^\mathrm{T}\mathbb{1}_N = \boldsymbol{\beta}\}}_{\mathcal{C}^2}$

Sinkhorn iterations

$\boldsymbol{b}^0 := \mathbb{1}_N$

1. $\boldsymbol{\Gamma}^l := \Delta(\boldsymbol{b}^l)\boldsymbol{K}^\mathrm{T}\Delta(\boldsymbol{a}^l)$
2. $\boldsymbol{a}^l := \dfrac{\boldsymbol{\beta}}{\boldsymbol{K}\boldsymbol{b}^{l-1}}$
3. $\boldsymbol{b}^l := \dfrac{\boldsymbol{\alpha}}{\boldsymbol{K}^\mathrm{T}\boldsymbol{a}^l}$

$\xrightarrow{\text{Linear convergence}}$ $\boldsymbol{\Gamma}^\star$

Optimal Transport

Wasserstein distance approximations

Entropic regularization: a Kulback-Leibler projection

$$W_2^2(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{\Gamma}\in\boldsymbol{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta})} \mathrm{KL}(\boldsymbol{\Gamma}|\boldsymbol{K})$$

- $\boldsymbol{K} = \exp(-\boldsymbol{C}/\gamma)$
- $\boldsymbol{\Pi}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \underbrace{\{\boldsymbol{\Gamma}\in\mathbb{R}_+^{N\times N},\quad \boldsymbol{\Gamma}\mathbb{1}_N = \boldsymbol{\alpha}\}}_{\mathcal{C}^1} \cap \underbrace{\{\boldsymbol{\Gamma}\in\mathbb{R}_+^{N\times N},\quad \boldsymbol{\Gamma}^{\mathrm{T}}\mathbb{1}_N = \boldsymbol{\beta}\}}_{\mathcal{C}^2}$

Sinkhorn iterations

$\boldsymbol{b}^0 := \mathbb{1}_N$

1. $\boldsymbol{\Gamma}^l := \Delta(\boldsymbol{b}^l)\boldsymbol{K}^{\mathrm{T}}\Delta(\boldsymbol{a}^l)$
2. $\boldsymbol{a}^l := \dfrac{\boldsymbol{\beta}}{\boldsymbol{K}\boldsymbol{b}^{l-1}}$
3. $\boldsymbol{b}^l := \dfrac{\boldsymbol{\alpha}}{\boldsymbol{K}^{\mathrm{T}}\boldsymbol{a}^l}$

$\Longrightarrow$ Linear convergence $\quad \boldsymbol{\Gamma}^\star$

Notes:

- If cost the squared Euclidean distance, $\boldsymbol{K}$ dot product can be replaced by convolution with Gaussian kernel with std equals to γ.
- Small values of γ can lead to instabilities. Alternatively , we can perform sinkhorn iterations in log domain.

Optimal Transport

Wasserstein barycenter approximation

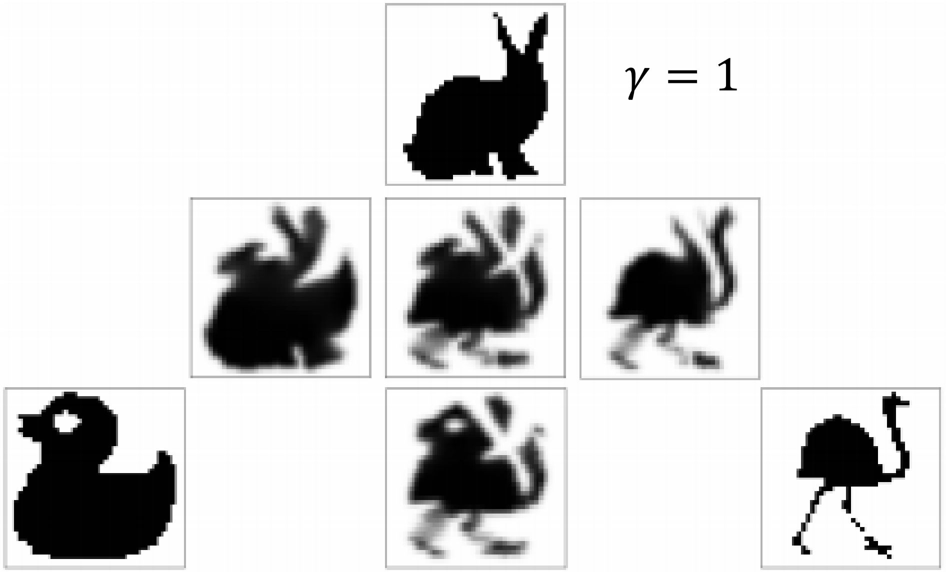
Extending to the barycenter computation..

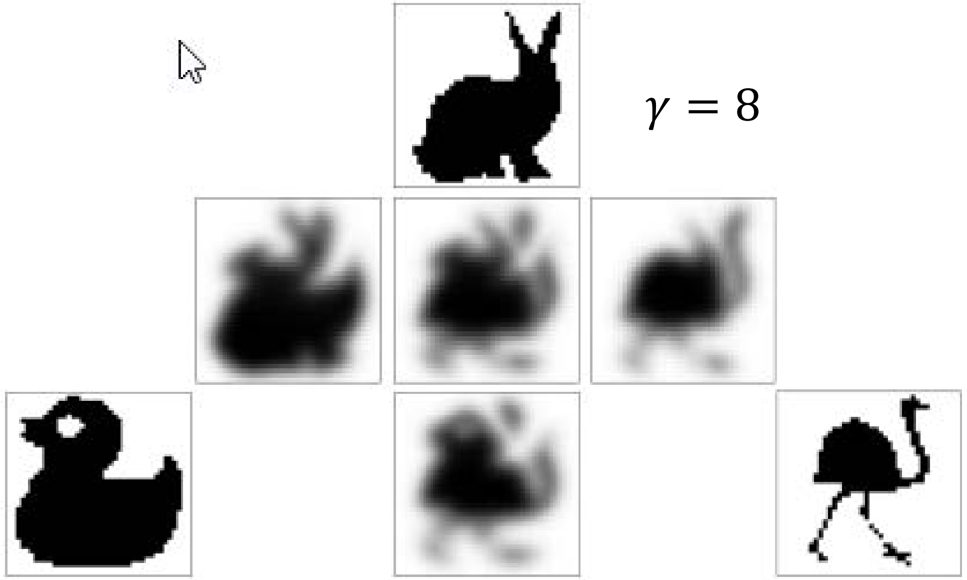
$$\boldsymbol{P}(\boldsymbol{D},\boldsymbol{\lambda}) = \operatorname*{argmin}_{\boldsymbol{u}} \sum_{s=1}^{S} \lambda_s W_\gamma(\boldsymbol{u},\boldsymbol{d}_s)$$ [4]

Generalized Sinkhorn algorithm

$\forall s, \quad \boldsymbol{b}_s^0 = \mathbb{1}_N$

1. $\forall s, \quad \boldsymbol{a}_s^l := \dfrac{\boldsymbol{d}_s}{\boldsymbol{K}\boldsymbol{b}_s^{l-1}}$ parallelizable
2. $P^l(\boldsymbol{D},\boldsymbol{\lambda}) := \Pi_{s=1}^{S}(\boldsymbol{K}^\top \boldsymbol{a}_s^l)^{\lambda_s}$
3. $\forall s, \quad \boldsymbol{b}_s^l = \dfrac{P^l(\boldsymbol{D},\boldsymbol{\lambda})}{\boldsymbol{K}^\top \boldsymbol{a}_s^l}$ parallelizable





Schmitz at al.,2018

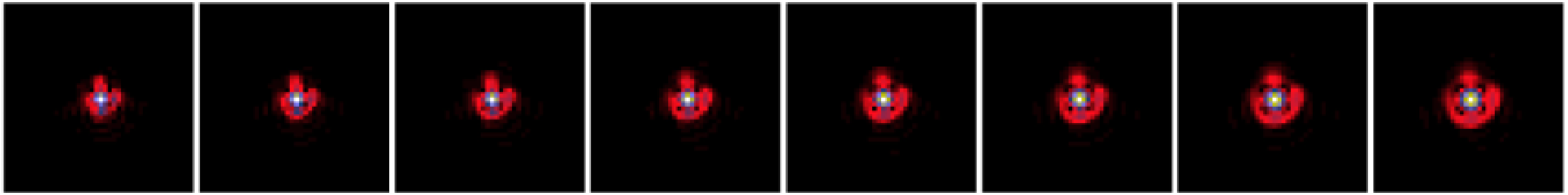
[4]Agueh & Carlier, 2011; Benamou et al., 2015; Chizat et al., 2016

λ-RCA

Transporting mass over wavelengths

Back to the PSF estimation problem:

$$\boldsymbol{Y} = \boldsymbol{M} \sum_v \beta_v \boldsymbol{X}_v + b$$



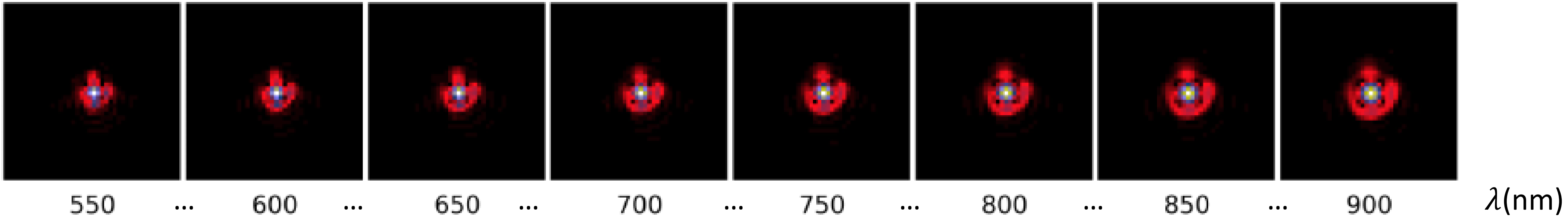
λ-RCA

Transporting mass over wavelengths

Back to the PSF estimation problem:

$$\boldsymbol{Y} = \boldsymbol{M} \sum_{v} \beta_v \boldsymbol{X}_v + b$$

How to break the degeneracy?



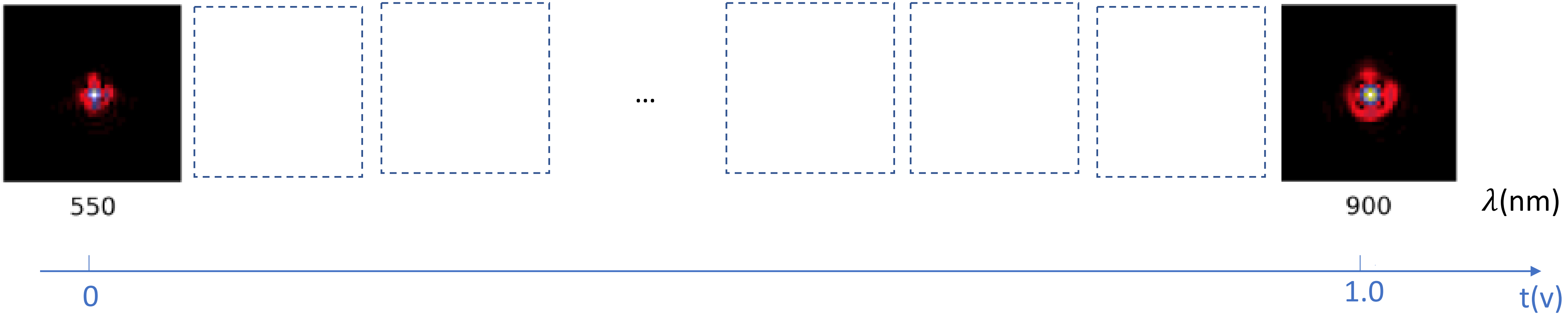
λ-RCA

Transporting mass over wavelengths

Back to the PSF estimation problem:

$$\boldsymbol{Y} = \boldsymbol{M} \sum_{v} \beta_v \boldsymbol{X}_v + b$$

How to break the degeneracy?



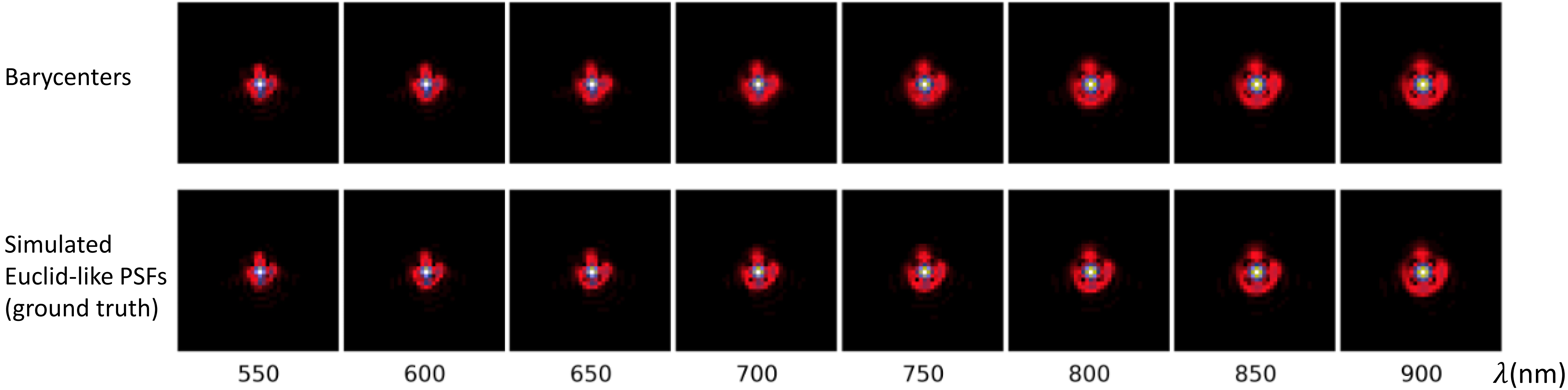
Hypothesis: monochromatic PSFs of a intermediary wavelength are Wasserstein barycenters of the extremes.

λ-RCA

Transporting mass over wavelengths

Hypothesis: monochromatic PSFs of a intermediary wavelength are Wasserstein barycenters of the extremes.

Experiment results:

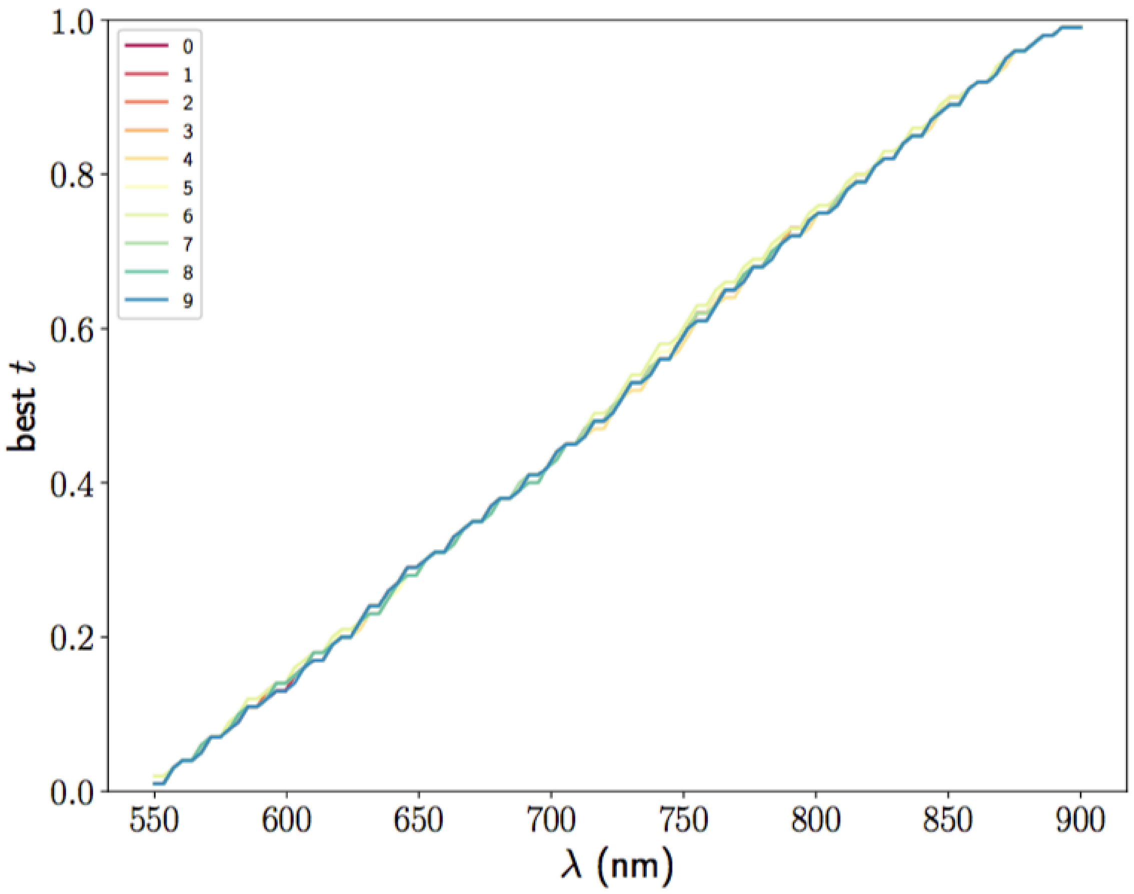


- linear projection of wavelengths into [0,1].
- γ=0.3

λ-RCA

Transporting mass over wavelengths

Is linear projection of λ's into [0,1] really the best option?



λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{X} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_2^2$$

λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{X} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_2^2 \quad \text{s.t.} \quad x_{i,j} > 0$$

- **Positivity constraint:** $X_{i,j} \geq 0$

λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{\boldsymbol{A},\boldsymbol{S}} \frac{1}{2}\sum_{k}^{K}\|\boldsymbol{Y}_k - \boldsymbol{M}\sum_{i}^{r} a_{i,k}\boldsymbol{s}_i\|_2^2 \quad \text{s.t.} \quad x_{i,j} > 0$$

- **Positivity constraint:** $X_{i,j} \geq 0$
- **Low-rank constraint:** $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$

  eigen-PSFs $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_r]$, $\boldsymbol{s}_i \in \mathbb{R}_+^{N\times r}$, $r << K$

  coefficients $\boldsymbol{A} \in \mathbb{R}^{r\times K}$

λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{\boldsymbol{A},\boldsymbol{S}} \frac{1}{2}\sum_{k}^{K} \|\boldsymbol{Y}_k - \boldsymbol{M}\sum_{i}^{r} a_{i,k}\boldsymbol{s}_i\|_2^2 \quad \text{s.t.} \quad x_{i,j} > 0$$

- **Positivity constraint:** $X_{i,j} \geq 0$
- **Low-rank constraint:** $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$ eigen-PSFs $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_r]$, $\boldsymbol{s}_i \in \mathbb{R}_+^{N\times r}$, $r << K$
  coefficients $\boldsymbol{A} \in \mathbb{R}^{r\times K}$
- **Smoothness constraint:** imposed on the atoms $\boldsymbol{s}_i$.

λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{\boldsymbol{A},\boldsymbol{S}} \frac{1}{2}\sum_{k}^{K} \|\boldsymbol{Y}_k - \boldsymbol{M}\sum_{i}^{r} a_{i,k}\boldsymbol{s}_i\|_2^2 \quad \text{s.t.} \quad x_{i,j} > 0$$

- **Positivity constraint:** $X_{i,j} \geq 0$
- **Low-rank constraint:** $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$
  eigen-PSFs $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_r]$, $\boldsymbol{s}_i \in \mathbb{R}_+^{N\times r}$, $r << K$
  coefficients $\boldsymbol{A} \in \mathbb{R}^{r\times K}$
- **Smoothness constraint:** imposed on the atoms $\boldsymbol{s}_i$.
- **Graph constraint:** imposed on the rows of $\boldsymbol{A}$. $\boldsymbol{A} = \boldsymbol{\alpha}\boldsymbol{V}^{\mathrm{T}}$

Coefficients learned from data.

Graph matrix.

λ-RCA

Inverse problem: RCA background (cf. Morgan's talk)

$$\min_{\boldsymbol{A},\boldsymbol{S}} \frac{1}{2}\sum_{k}^{K}\|\boldsymbol{Y}_k - \boldsymbol{M}\sum_{i}^{r} a_{i,k}\boldsymbol{s}_i\|_2^2 \quad +\|\boldsymbol{\Phi}\boldsymbol{S}\|_1 \quad \text{s.t.} \quad x_{i,j} > 0$$

- **Positivity constraint:** $X_{i,j} \geq 0$
- **Low-rank constraint:** $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$ eigen-PSFs $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_r]$, $\boldsymbol{s}_i \in \mathbb{R}_+^{N\times r}$, $r << K$
  coefficients $\boldsymbol{A} \in \mathbb{R}^{r\times K}$
- **Smoothness constraint:** imposed on the atoms $\boldsymbol{s}_i$.
- **Graph constraint:** imposed on the rows of $\boldsymbol{A}$. $\boldsymbol{A} = \boldsymbol{\alpha}\boldsymbol{V}^{\mathrm{T}}$
- **Sparsity constraint:** PSF is sparse in starlet domain.

Coefficients learned from data.

Graph matrix.

λ-RCA

Inverse problem: introduction of barycenters

$$\min_{\boldsymbol{A},\boldsymbol{D}} \frac{1}{2}\sum_{k}^{K} \|\boldsymbol{Y}_k - \boldsymbol{M}\sum_{i}^{r} a_{i,k} \underbrace{\sum_{v} \beta_v \boldsymbol{P}_i(\boldsymbol{D},\boldsymbol{w}_v)}_{\boldsymbol{s}_i}\|_2^2 + \|\boldsymbol{\Phi}\boldsymbol{D}\|_1 \quad \text{s.t.} \quad \boldsymbol{D} \in \Sigma_N$$

Simplex constraint on dictionary: probability distribution.

- $\boldsymbol{D} = (\boldsymbol{d}_0^i, \boldsymbol{d}_1^i)_{1\le i\le S}\,, \quad \boldsymbol{d}_0^i \in \mathbb{R}_+^N$

Eigen-PSF extreme left

Eigen-PSF extreme right

- $\boldsymbol{w} = (1 - t_v, \quad t_v)_{1\le v\le \text{nb\_wvls}}$

λ-RCA

Alternated optimization

Algorithm 1 λ-RCA

Parameters estimation and initialization :
First guess super resolution $\rightarrow \boldsymbol{D}$
Spatial constraints parameters $\rightarrow \boldsymbol{\alpha}, \boldsymbol{V}$
Noise estimation $\rightarrow$ sigmas, threshold weights $\boldsymbol{W}$, shifts
Power method $\rightarrow$ gradient descent step

for $l = 0, ..., L$ **do**
    (I) dictionary update:

$$\boldsymbol{D}_{\text{new}} := \text{Condat-Vu}\left(\underset{\boldsymbol{D}}{\text{argmin}} \sum_k \frac{1}{2} \|\boldsymbol{y}_k - \sum_i a_{i,k} \sum_v \beta_v \boldsymbol{P}_i(\boldsymbol{D}, \boldsymbol{w}_v)\|_2^2 \quad + \|\boldsymbol{W}\boldsymbol{\Phi}\boldsymbol{D}\|_1 \quad + \delta_{\boldsymbol{D} \in \sum_N}\right)$$

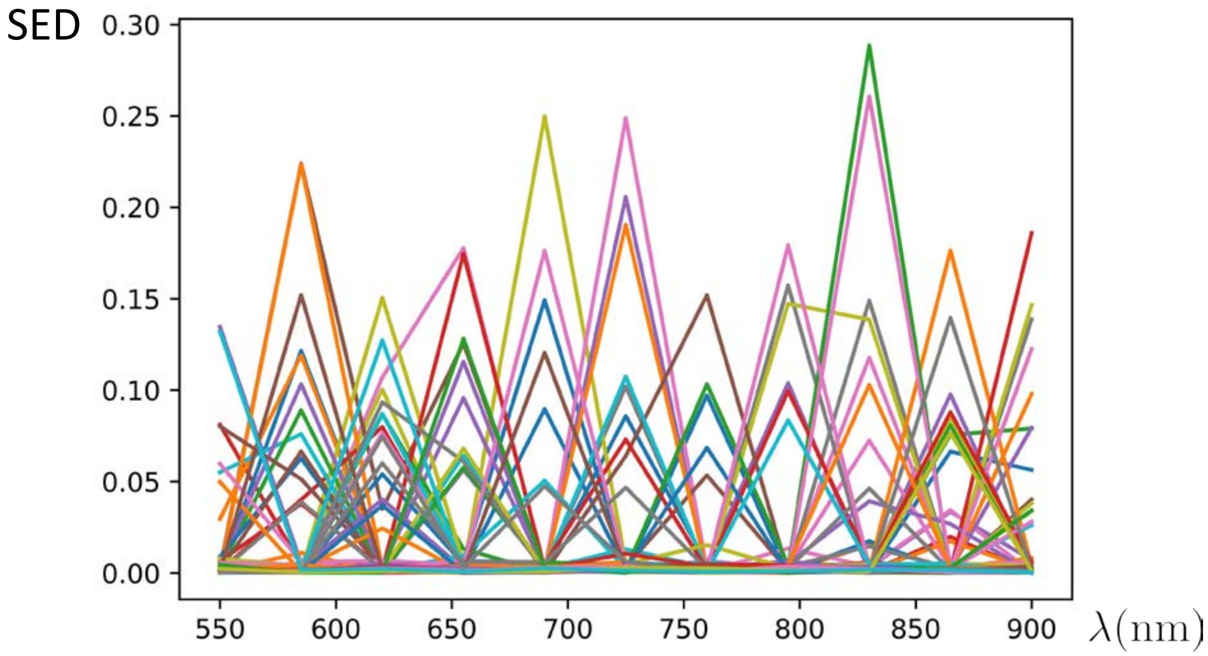
    (II) coefficients update:

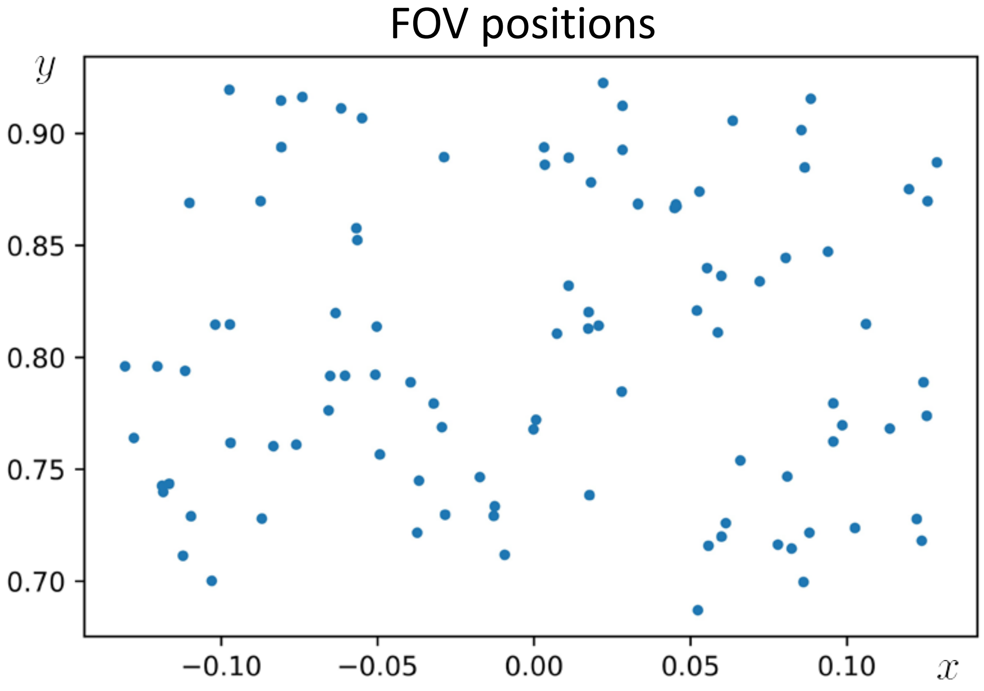
$$\boldsymbol{\alpha}_{\text{new}} := \text{Forward-Backward}\left(\underset{\boldsymbol{\alpha}}{\text{argmin}} \sum_k \frac{1}{2} \|\boldsymbol{y}_k - \sum_i a_{i,k} \sum_v \beta_v \boldsymbol{P}_i(\boldsymbol{D}, \boldsymbol{w}_v)\|_2^2 \quad + \|\boldsymbol{\alpha}\|_0\right)$$

    $\boldsymbol{A}_{\text{new}} = \boldsymbol{\alpha}_{\text{new}} \boldsymbol{V}^{\mathrm{T}}$
    (III) update parameters and threshold weights.
return $\boldsymbol{A}$, $\boldsymbol{D}$

λ-RCA

Preliminary results: Euclid-like test dataset





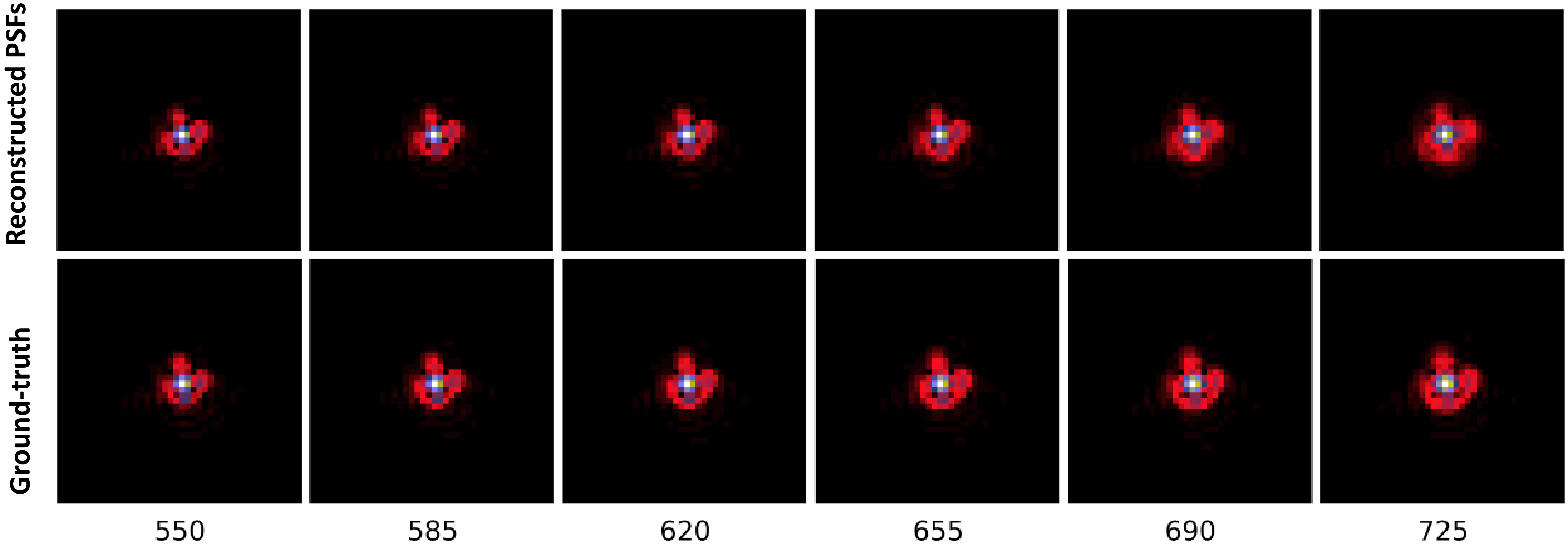
Data description:

- 100 stars in Euclid resolution 22x22p.
- Spatial positions in FOV.
- SEDs in 11 wavelengths. [550-900nm]

λ-RCA

Preliminary results

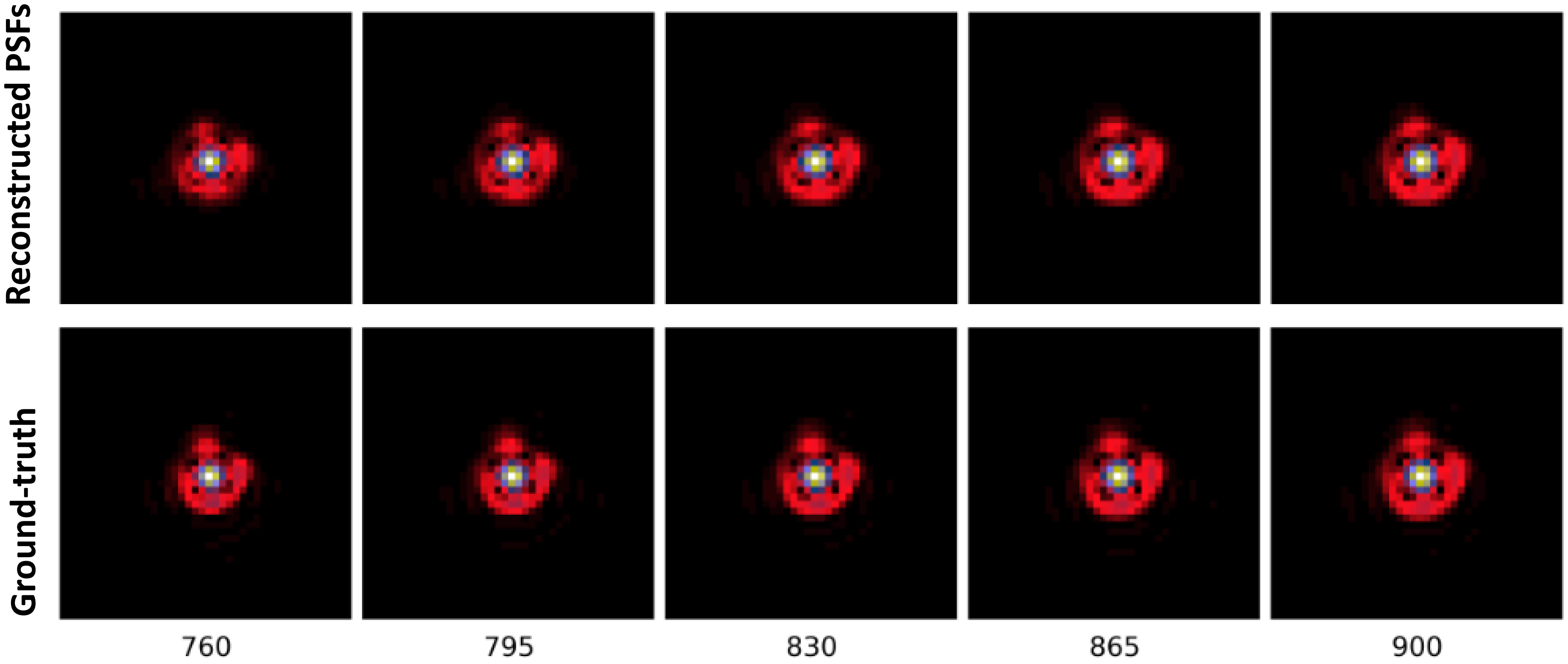
Dictionary first guess: ground-truth PSFs



λ-RCA

Preliminary results

Dictionary first guess: ground-truth PSFs



λ-RCA

Perspectives

- Test different dictionary first guess using observed information.
- Optimize λ-RCA algorithm to its full potential.
- Reconstruct PSFs of Euclid-like simulated galaxies on galsim and perform shape measurement.
- Evaluate sensibility with respect to SED uncertainty.
- Evaluate impact of band leakage in the barycenter estimation.