Centering noisy image with application to Cryo-EM

黃峻祿 for the paper A. Heimowitz, N. Sharon and A. Singer, centering noisy image with application to cryo-EM, SIAM J. Imaging Science, (2021) 14:689.

Group meeting at ISSAS, Taipei

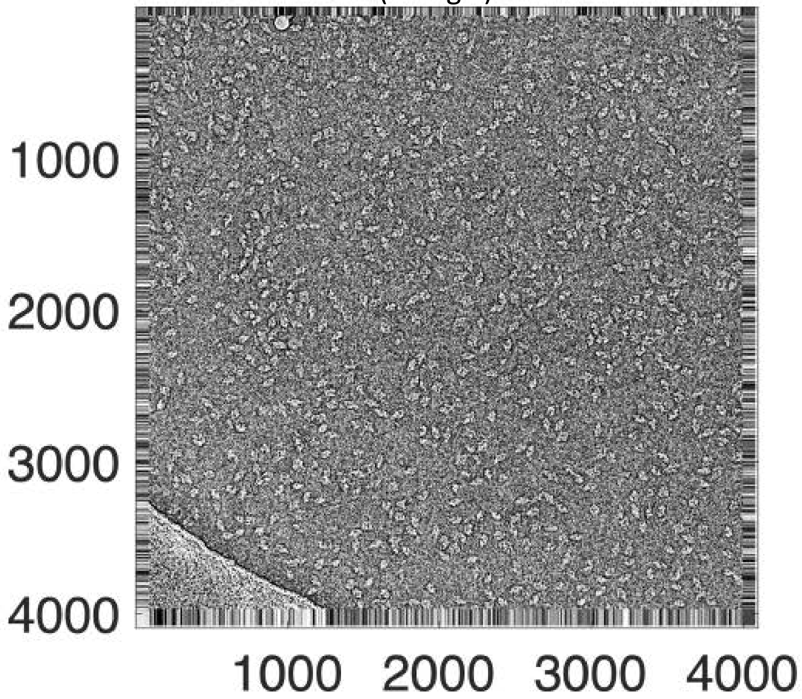
Content

- Preliminary
- Centering noise image
- Paper's suggested method
- Application to the cryo-EM data

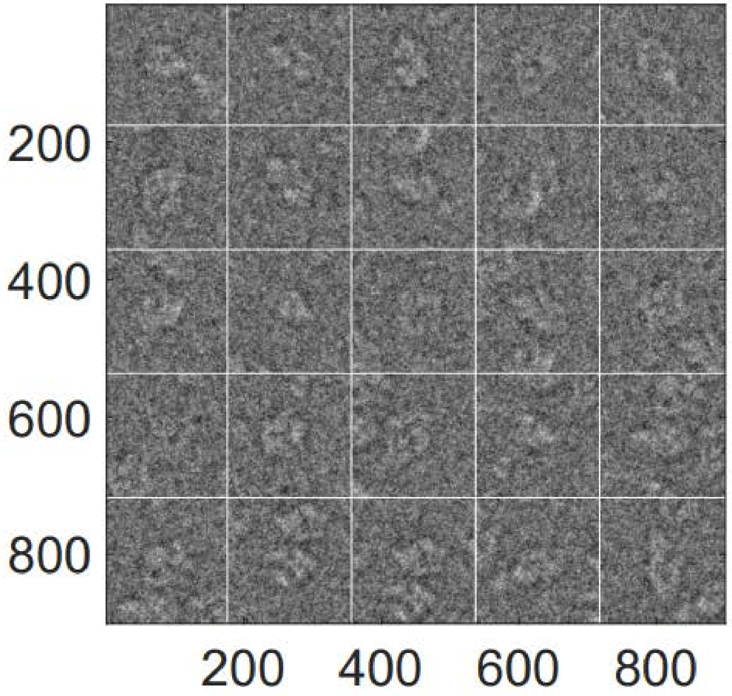
Preliminary: particle picking

In the single particle analysis, we have a stage for picking up particles from a micrograph.

Experimental data from EMPIAR-10017 (betagal)



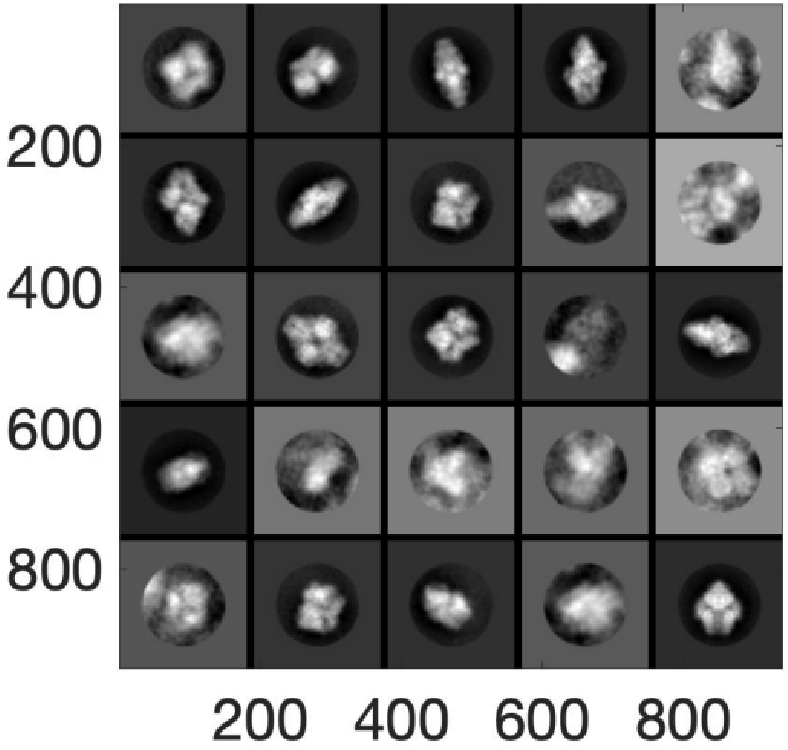
Single particle images extracted from the left micrograph



The single particle images are selected by using the APPLE picker. The picker provides roughly centered projection images.

Preliminary: 2D classification

After picking the single particle images, we then classify and average the particle images with similar viewing direction.



Some of the class averages have high quality while others are blurred. The blurred averages are discarded. **Centering the particle images before the 2D classification reduce the number of blurred classes**.

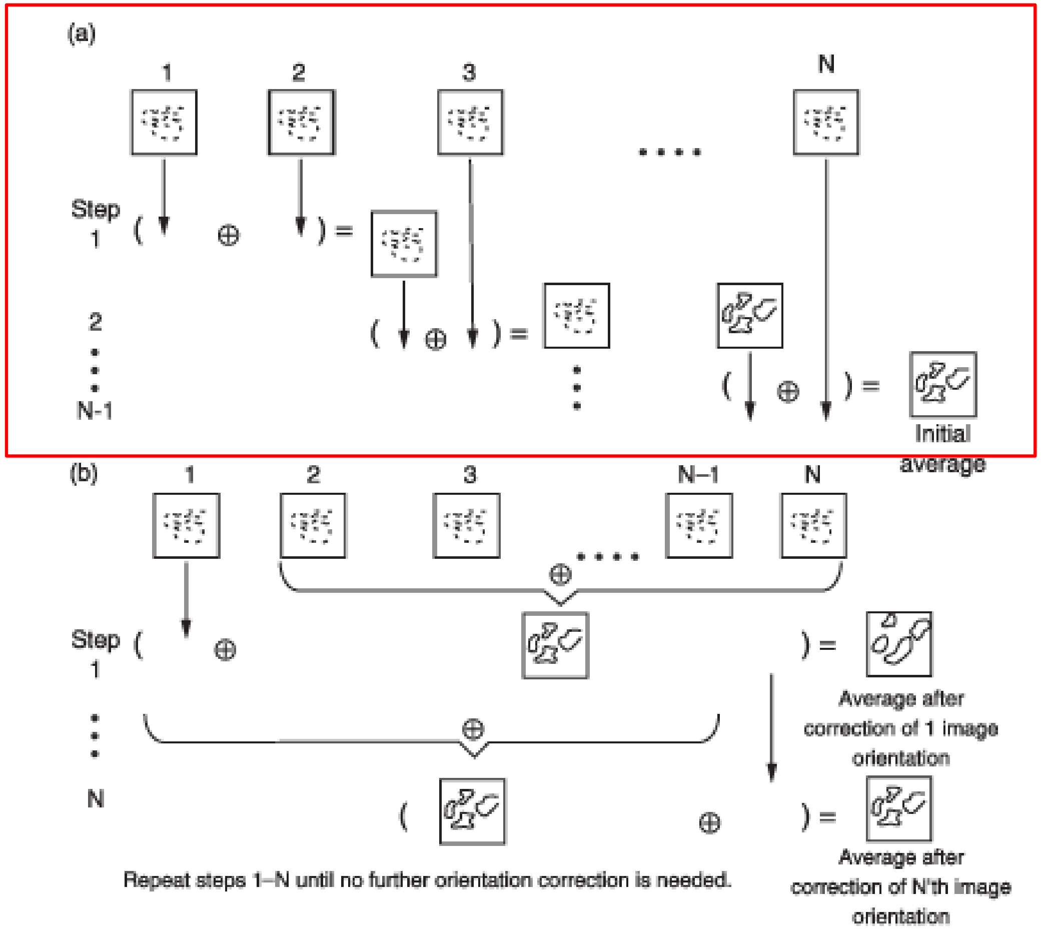
Preliminary: reference-free alignment (RFA)

One of the centering methods is the RFA. Iterative alignment method [2] is introduced.

We have a set of particle image $P = \{p_i; i = 1, \ldots, N\}$

In (a) algorithm, it is to obtain the global average:

- Pick randomly two images $p_{i=2}$ and $p_{k=1}$ from *P*
- Align p_i and p_k in terms of minimizing $\|p_i$-$p_k\|$
- Construct the initial average $a_{m=1} = p_i \oplus p_k$
- Align $p_{i=3}$ and $a_{m=1}$
- Update the average $a_{m=2} = (p_{i=3} \oplus (m-1)a_{m=1})/i$
- Increment *m* by one until *m*=N to obtain final average $A = a_{m=N}$



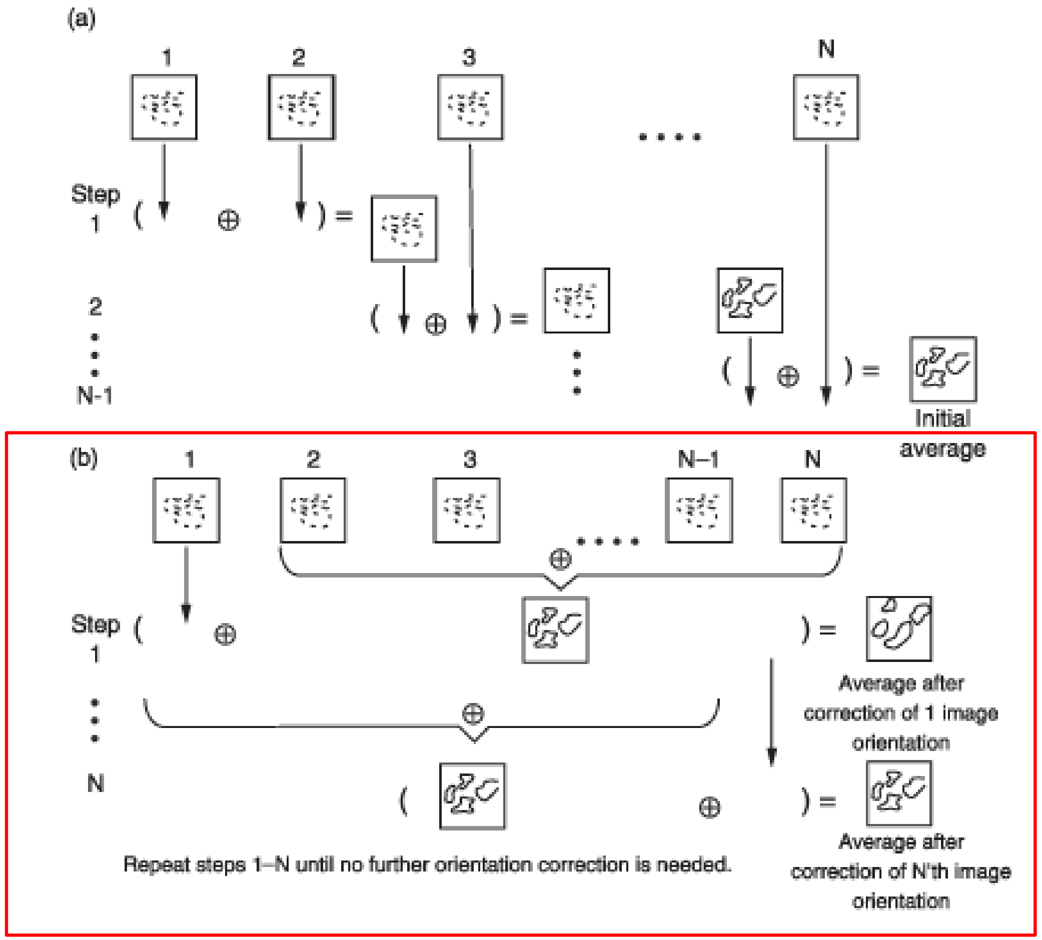
Preliminary: reference-free alignment (RFA)

One of the centering methods is the RFA. Iterative alignment method [2] is introduced.

We have a set of particle image $P = \{p_i; i = 1, \dots, N\}$

In (b) algorithm, it is to refine the global average:

1. Set counter *m*=1
2. Subtract A from p_k to obtain a modified average: $A' = \frac{(NA - p_k)}{N-1}$
3. Align A' with p_k
4. Update the $A = [p_k \oplus (N-1)A']/N$
5. Increment *m* by one, if *m*<N, go to 2
6. If 3 shows the image's position is changed significantly, go back to 1



Motivation

To identify the center of mass (CM) in noisy data becomes challenging. This paper proposes a different approach which is the geometric median (GM) serving as a robust estimator of the CM.

Center of mass and geometric median

We define center of mass, CM:

$$\mu = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{i-1}^{n} w_i p_i$$

where $\{p_i\}_{i=1}^{n}$ is a set of points and their associate weights $\{w_i\}_{i=1}^{n}$.

Alternative center of mass is defined by using Fréchet variance* $\psi(x) = \sum_{i=1}^{n} w_i d^2(p_i, x)$ for any point *x*,

$$\mu = \arg\min_x \psi(x)$$

*: the sum of weighted squared distances from *x* to p_i

Fréchet mean, μ, can minimize Fréchet variance. The Fréchet mean is influenced by outliers and extreme values. Fréchet median is suggested to use.

The geometric median, GM, is defined by using the Fréchet median

$$\mu_1 = \arg\min_x \sum_{i=1}^{n} w_i d \quad (p_i, x)$$

Proof of Fréchet mean without weight involved

$$\phi(x) = \frac{1}{n}\sum_{i=1}^{n} d^2(p_i, x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_j - p_{i,j}\right)^2$$

Proof from [1]

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} {x_j}^2 - 2\sum_{j=1}^{m} x_j p_{i,j} + \sum_{j=1}^{m} {p_{i,j}}^2\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} {x_j}^2 - \frac{2}{n}\sum_{j=1}^{m}\left(\sum_{i=1}^{n} p_{i,j}\right)x_j + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} {p_{i,j}}^2\right)$$

$$= \sum_{j=1}^{m}\left[x_j - \left(\frac{1}{n}\sum_{i=1}^{n} p_{i,j}\right)\right]^2 - \sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} p_{i,j}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} {p_{i,j}}^2\right)$$

$$= \sum_{j=1}^{m}\left[x_j - \left(\frac{1}{n}\sum_{i=1}^{n} p_{i,j}\right)\right]^2 + \sum_{j=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n} {p_{i,j}}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} p_{i,j}\right)^2\right]$$

$$\mu = \arg\min \phi(x) = \frac{1}{n}\sum_{i=1}^{n} p_{i,j}$$

Application of center of mass to image (0)

- Scipy::nidimage::center_of_mass() is reviewed

We have an array [0,0,0,0], and we want to know its center of mass.
[0,1,1,0],
[0,1,1,0],
[0,1,1,0]

The calculation in nidimage is shown below:

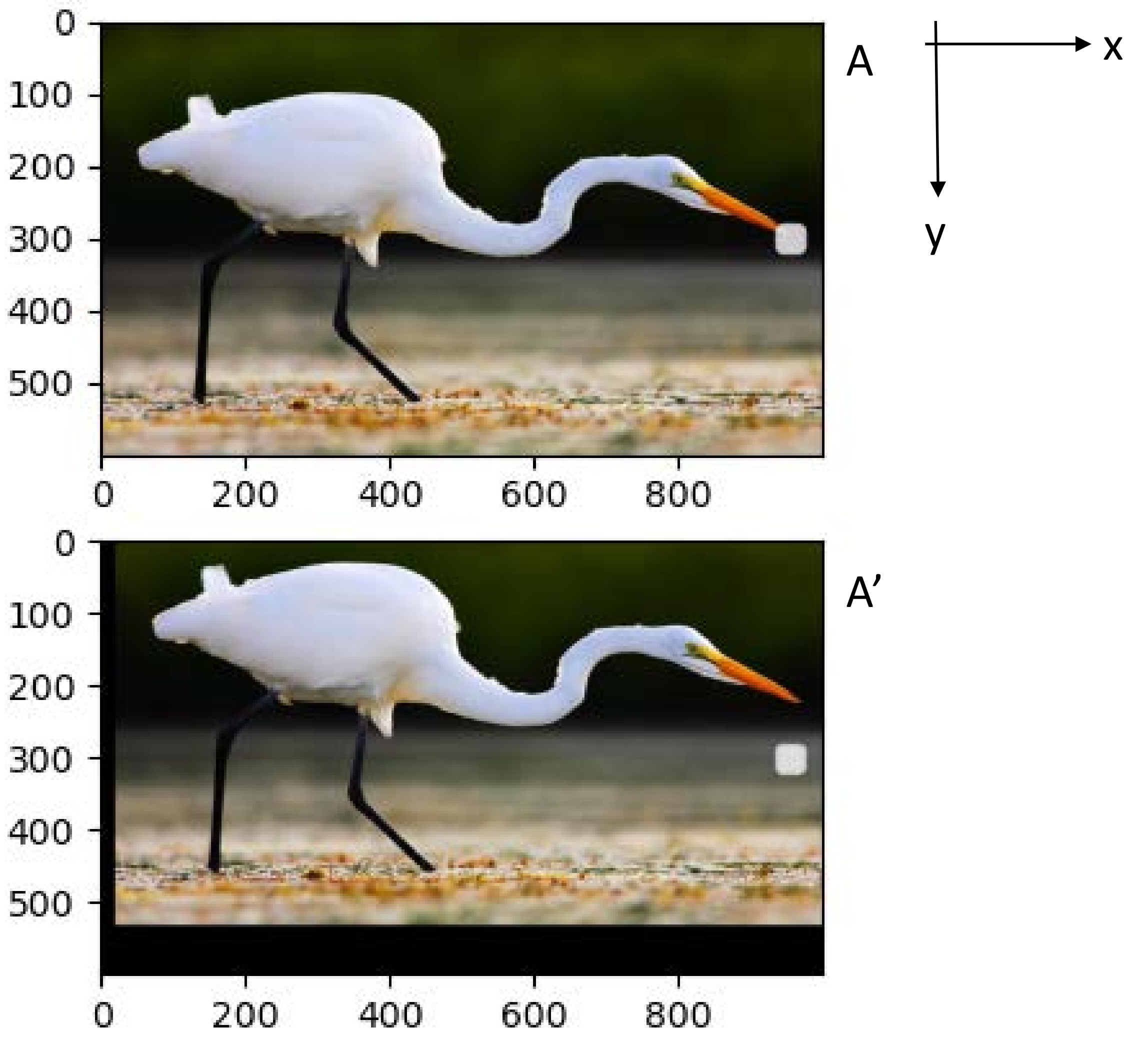
- Get dimension of the array, which is 2
- Get sum over column in the array, which is [0,3,3,0] as a normalizer
- Get x index, which is [0,1,2,3] and y index which is $[0,1,2,3]^T$
- Get product of the array and the x-y index.
- Divide the product by the normalizer

Application of center of mass to image (1)

A simple example for showing center of mass in an image.

Scipy::nidimage::center_of_mass() is used.

The size of image A is 600x1000 and its center of mass is computed as (*y*=366.66, *x*=479.98). We can shift A by an amount, (-66.66, 20.02), to become A'.



Application of center of mass to image (2)

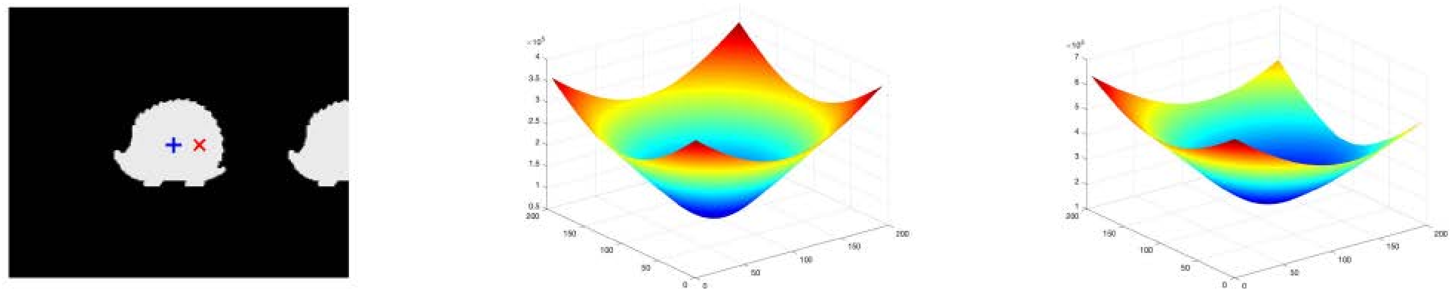
- Center of mass, **CM**, of an image can be written

$$\mu = \arg\min_{x\in\mathcal{P}} \sum_{p_i\in\mathcal{P}} I(p_i)d^2(p_i,x)$$

where $\mathcal{P} = \{p_i\}_{i=1}^{n}$ the set of points (image grid in image) and $I(p_i)$ the weights or pixel values. *I* is the image that we want to center.

- **GM** is reformulated as:

$$\mu = \arg\min_{x\in\mathcal{P}} \sum_{p_i\in\mathcal{P}} I(p_i)d\quad(p_i,x)$$



(a) GM of the hedgehog (blue +)
GM of the whole image (red ×)

(b) Landscape for blue +

(c) Landscape for red ×

Application of center of mass to image (2)

- Center of mass, **CM**, of an image can be written

$$\mu = \arg\min_{x \in \mathcal{P}} \sum_{p_i \in \mathcal{P}} I(p_i) d^2(p_i, x)$$

where $\mathcal{P} = \{p_i\}_{i=1}^{n}$ the set of points (image grid in image) and $I(p_i)$ the weights or pixel values.
I is the image that we want to center.

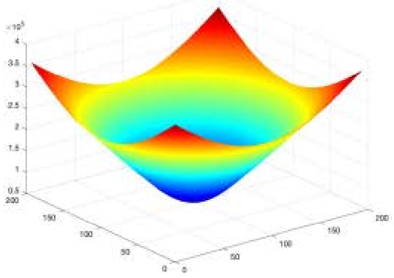
- **GM** is reformulated as:

Hedgehog is:

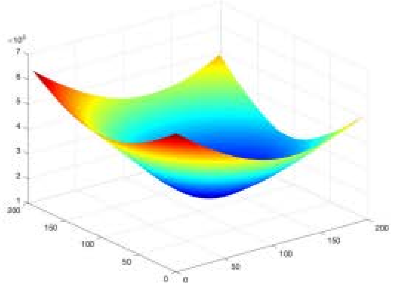
Figure from [Natioanl Geographic in April 2014]

GM of the he

$$\mu = \arg\min_{x \in \mathcal{P}} \sum_{p_i \in \mathcal{P}} I(p_i) d \quad (p_i, x)$$



(b)



(c)

Application of center of mass to image (2)

- Center of mass, **CM**, of an image can be written

$$\mu = \arg\min_{x \in \mathcal{P}} \sum_{p_i \in \mathcal{P}} I(p_i) d^2(p_i, x)$$

where $\mathcal{P} = \{p_i\}_{i=1}^{n}$ the set of points (image grid in image) and $I(p_i)$ the weights or pixel values. *I* is the image that we want to center.

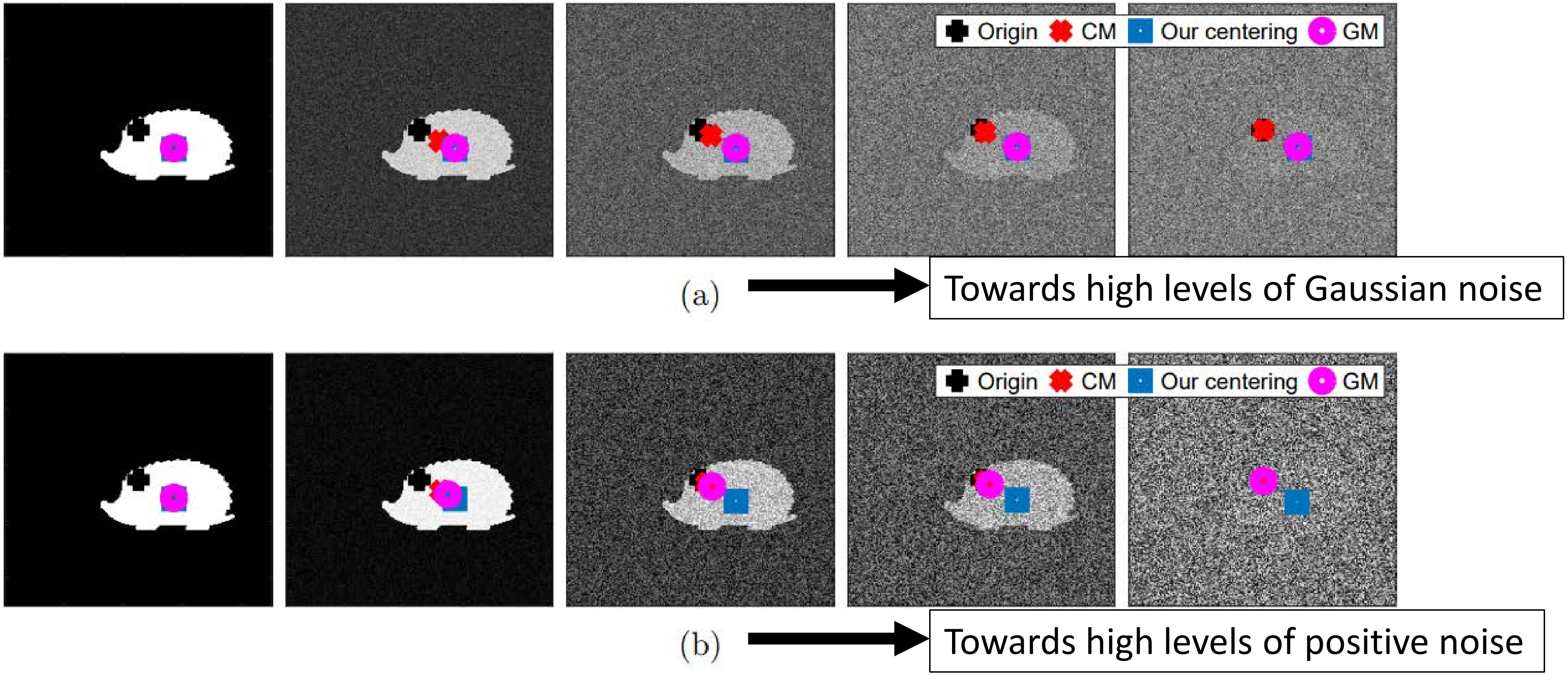
- **GM** is reformulated as:

$$\mu = \arg\min_{x \in \mathcal{P}} \sum_{p_i \in \mathcal{P}} I(p_i) d \quad (p_i, x)$$

GM of the hedgehog has contribution from other hedgehog. If we apply GM to the cryo-EM images that has high levels of noise, Fréchet mean is mostly determined by the noise. That GM then reflects the noise center instead of the object one. This paper proposes a surrogate function to the CM which can indicate the center of object in high noisy image.

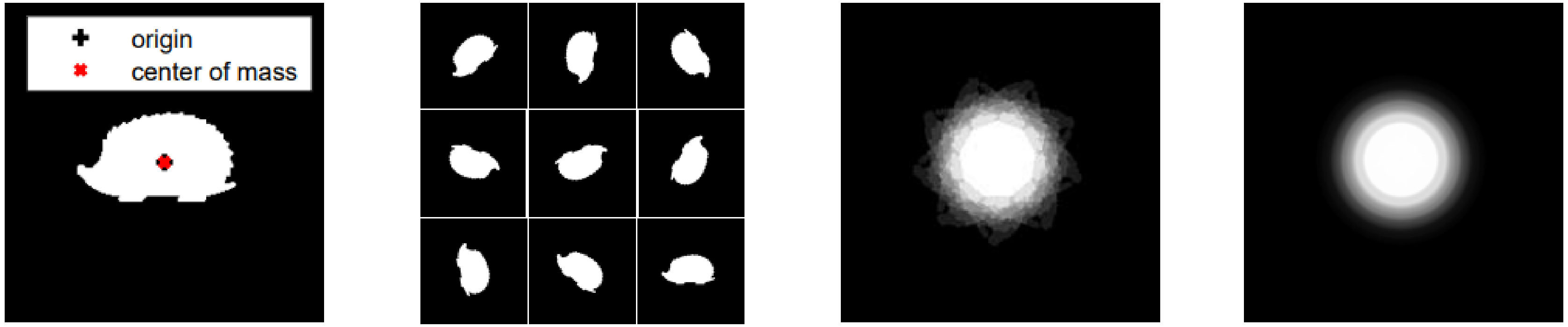
GM of the whole image (red ×)

Effect of noise on the centers

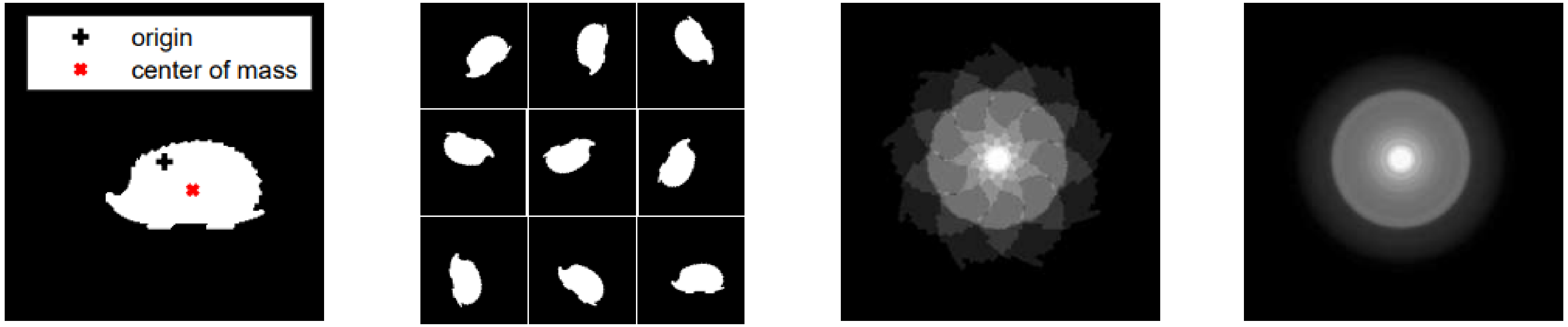


With increasing levels of the noise in (a) and (b), paper's centering can always point the object's center. GM fails at centering in (b) and CM fails in (a) and (b).

Towards the proposed method for centering (1)



(a) Centered (b) rotated copies (c) Sum of rotations in (b) (d) Average of all possible rotations

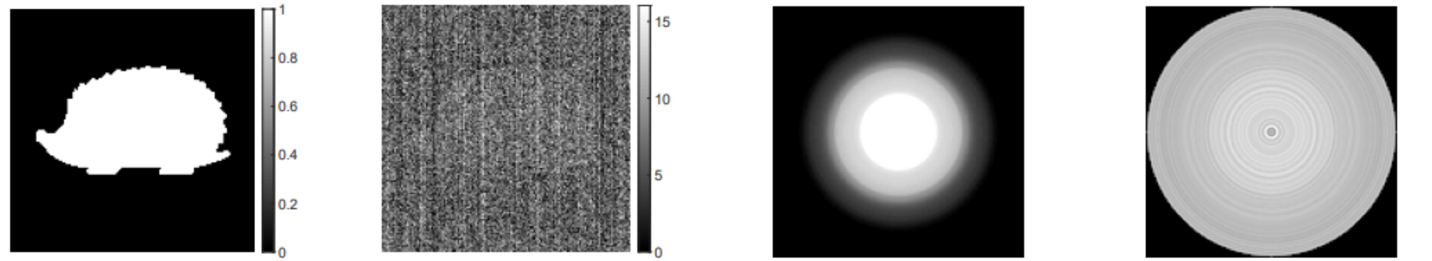


(e) Shifted (f) rotated copies (g) Sum of rotations in (f) (h) Average of all possible rotations

The difference in intensity spreads, (d) and (h), of rotational averaging of centered (top) and uncentered (bottom) hedgehogs.

Towards the proposed method for centering (2)

SNR(I) = $\|I_c\|^2 / \|\varepsilon\|^2$, I_c is the mean image and $\varepsilon = I - I_c$ the noise added to I_c

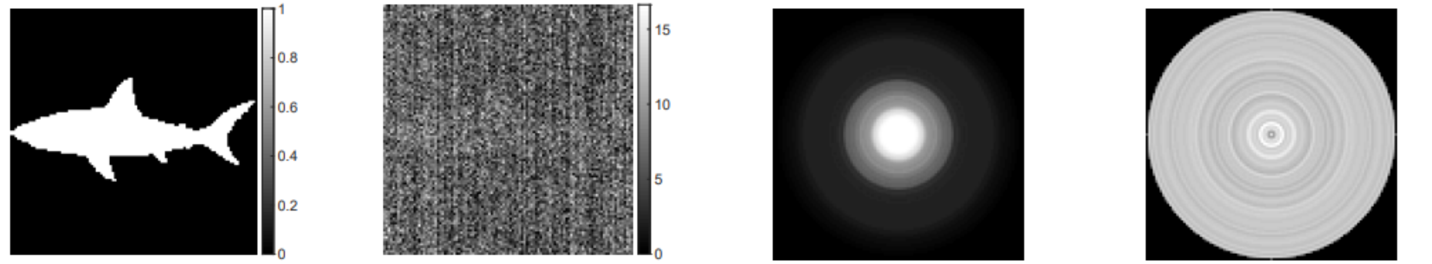


(a) Clean silhouette

(b) Noisy silhouette with SNR = 1/45

(c) Rotational averaging of (a)

(d) Rotational averaging of (b). In view of (c), the SNR = 1



(e) Clean silhouette

(f) Noisy silhouette with SNR = 1/45

(g) Rotational averaging of (e)

(h) Rotational averaging of (f). In view of (g), the SNR = 1/5

Rotational averaging has denoising effect.

Surrogate function

The paper presents a suggested surrogate function which serves as a robust estimate of the CM.

Radially rotational sum around the pixel *p* in image *I* of a radius, *R*, is written by:

$$u_p[R] = \sum_{l=0}^{R} \sum_{s \in B_l(p)} I(s)$$

*l*th ring of pixels around p, $A_l(p) = \{s \in \mathcal{P} | l-1 < d(p,s) < l\}$
Disk $B_l(p) = \bigcup_l A_l(p)$

We can quantify the distance between a given rotationally averaged projection and the ideal centered image and minimize the distance to find the sCM .

The minimization can be expressed as:

$\mu_s = \arg\min_{p \in P} \hat{L}_I(p)$, where the landscape of the sCM is $\hat{L}_I(p) = \left\| {}^{u_p}/_{\mathrm{E_{max}}} - \mathbf{1} \right\|_1$ (3.1)

$E_{max} = \max_{p \in P} u_p[R]$, a set of unit vector **1**=(1,…,1), and $\| \quad \|_1$ is the standard norm.

Rotational averaging

An image is expanded using 2D prolates spheroidal wave functions, PSWFs, to evaluate the rotational average. $I = \sum_{N=-\infty}^{\infty} \sum_{n=0}^{\infty} \alpha_{N,n} \Psi_{N,n}^{R}$

The wave function is:

$\Psi_{N,n}^{R}(\rho, \theta) = \begin{cases} \frac{1}{\sqrt{2\pi}} g_{N,n}^{R}(\rho) e^{iN\phi} & \rho < R, \\ 0 & \text{otherwise} \end{cases}$ where ρ and θ are radial and angular coordinates.

N and *n* are referred to angular and radial indexes.

The rotational average is

$\mathcal{A}_I(\rho) = \sum_{n=0}^{n_R} \alpha_n \Psi_{n,0}^{R}(n)$ where the expansion coefficients $\alpha_n = \int_{\rho \leq R} I(\rho) \Psi_{n,N=0}^{R}(\rho)\, d\rho$

Algorithm for finding center in image

Algorithm 3.1. Robust translational centering

Input: A (nonnegative) image I, and the particle radius R
Output: The chosen center μ_s
1: Define the set P of possible centers.
2: **for all** $p \in P$ **do**
3: $\quad I_p \leftarrow$ Crop I around the pixel p according to R
4: $\quad u_p \leftarrow$ Apply rotational averaging over I_p
5: $\quad \hat{L}_I(p) \leftarrow$ Calculate and store; see (3.1)
6: **end for**
7: $\mu_s \leftarrow \arg\min_{p \in P} \hat{L}_I(p)$
8: Return μ_s

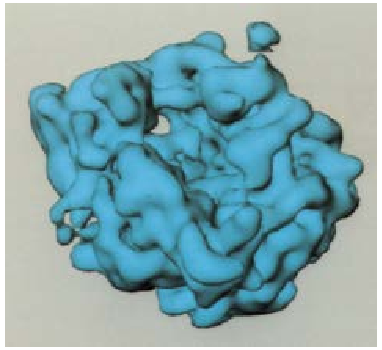
$$\hat{L}_I(p) = \left\| {u_p}/{\mathrm{E_{max}}} - \mathbf{1} \right\|_1 \quad (3.1)$$

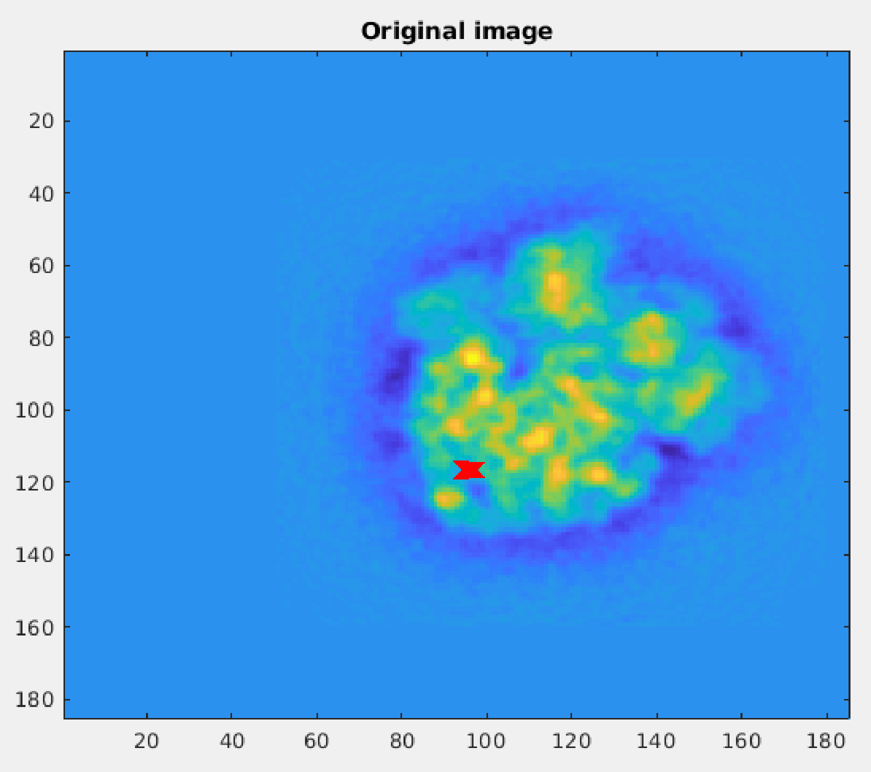
Example for centering image

70S ribosome data is used. This data is provided in its package on Github

Power by matlab R2017b

Visualization of 70S ribosome. Figure from [J. Mol. Biol. (1998) **280**, 103]





Index of center of mass, (95.4876, 116.0776)

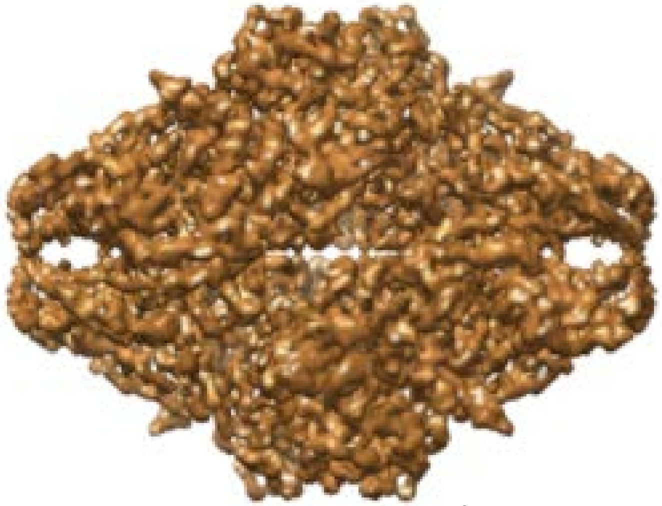
Image size: 91x91

Index of center of mass, (45.4676, 44.7353)

Application to cryo-EM data

Test data sample: beta-gal (EMPIAR-10017)

Resolution: 4.2 Å



Width: ~140 Å

Length: ~180 Å

The values of length and width are from [A. Bartesaghi, Structure of Beta-galactosidase at 3.2 Å resolution obtained by cryo-electron microscopy. Proc Natl Acad Sci USA 2014 111(32): 11709]

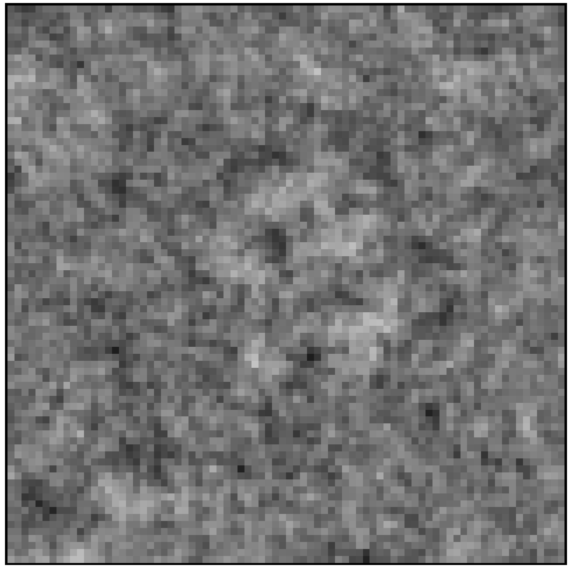
Microscopy condition
Acceleration voltage: 300 kV
C_s: 2 mm
Defocus: 1.359 – 4.692 µm

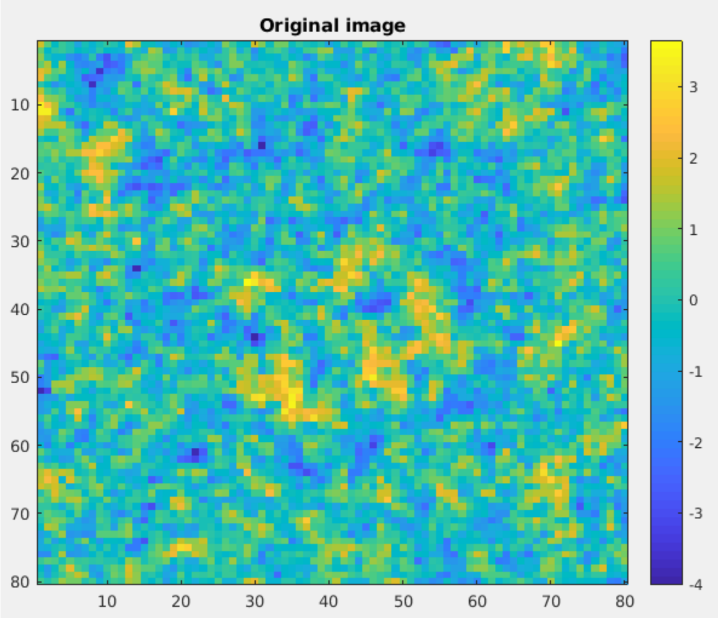
Application to the noisy data from cryo-EM

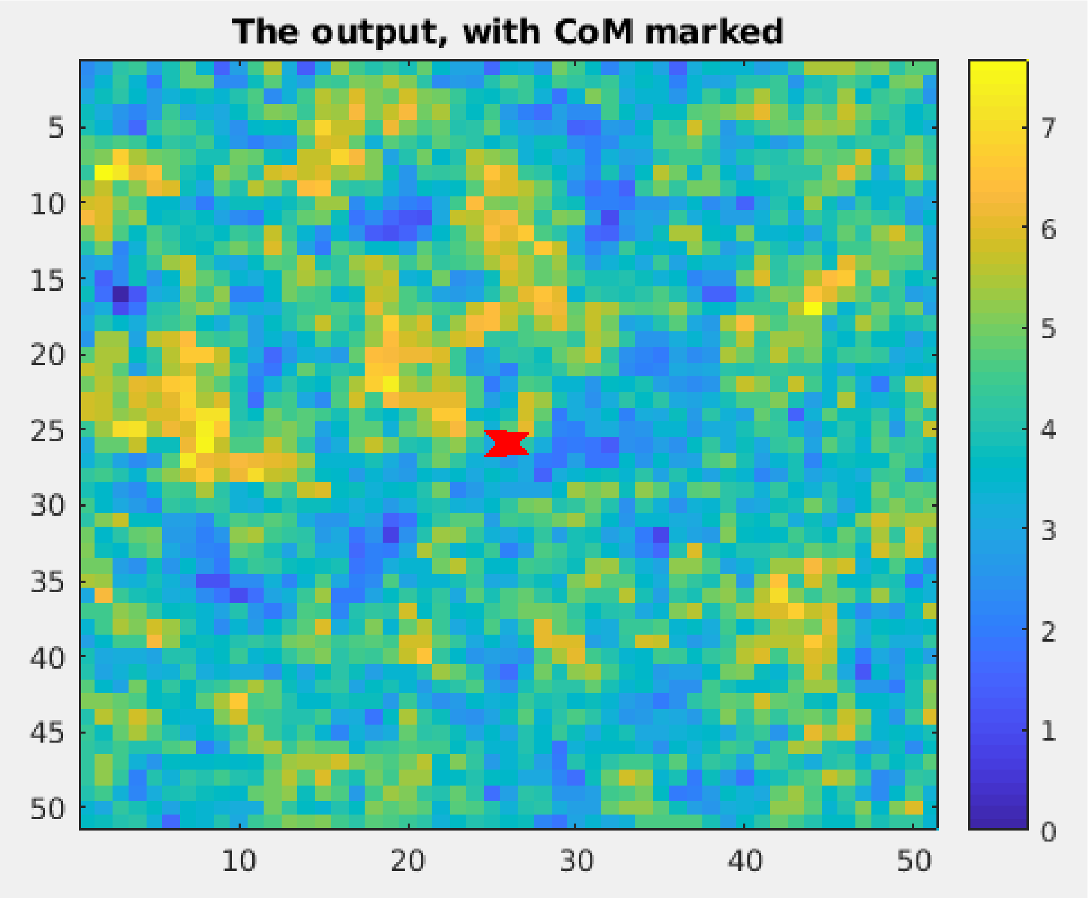
Apply this centering method to the noisy data.

Procedure to obtain the particle image: alignment→ (ctffind and particle picking) →(particle extraction using relion)

Image size: 80x80, 3.54 Å/px







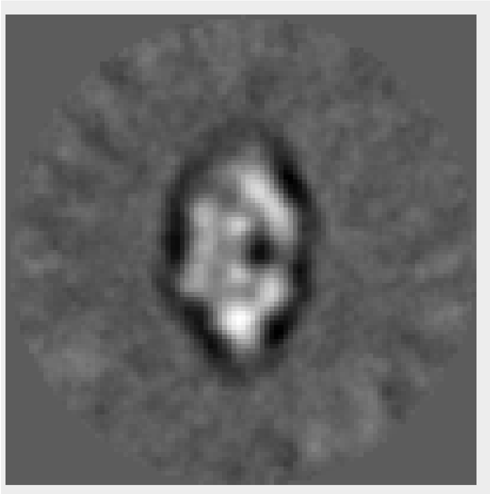
The length of beta-gal occupies about 51 pixels.

Application to classified data from cryo-EM (1)

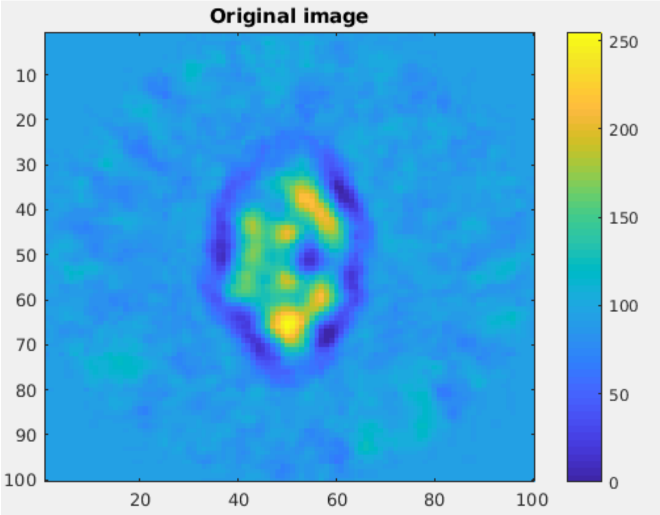
Apply this centering method to the classified data in 2D.

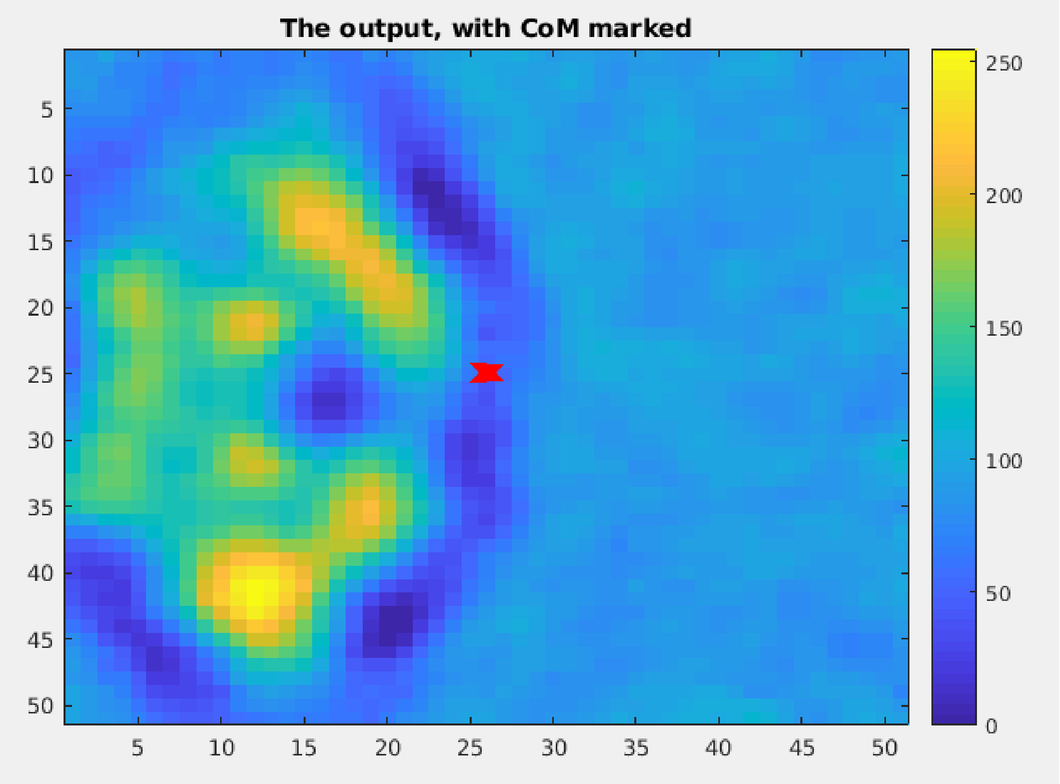
Procedure to obtain the particle image: alignment→ (ctffind and particle picking) → particle extraction → align with cl2d (using xmipp3)

Image size: 100x100, 3.54 Å/px



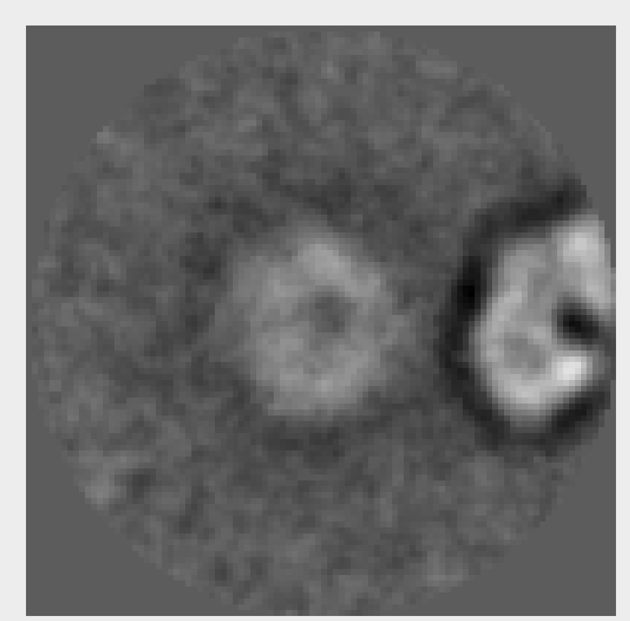
A clear particle image is used.

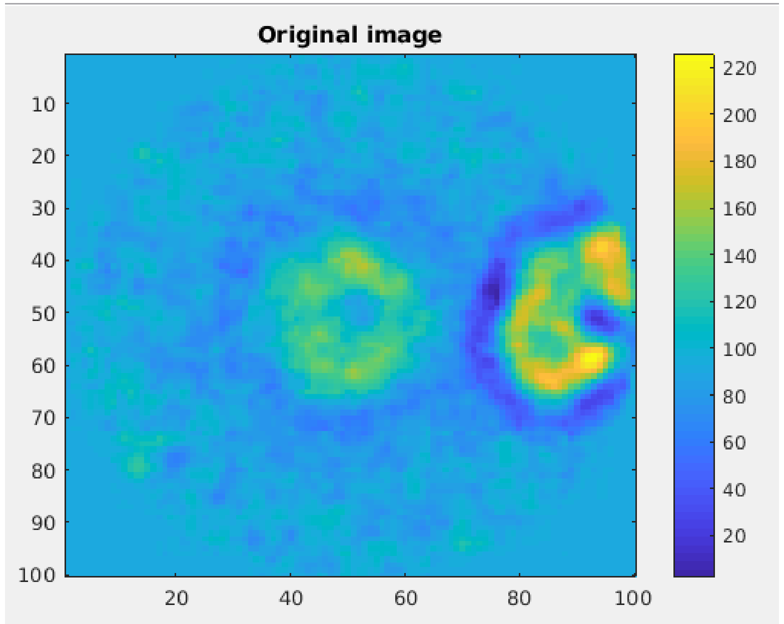


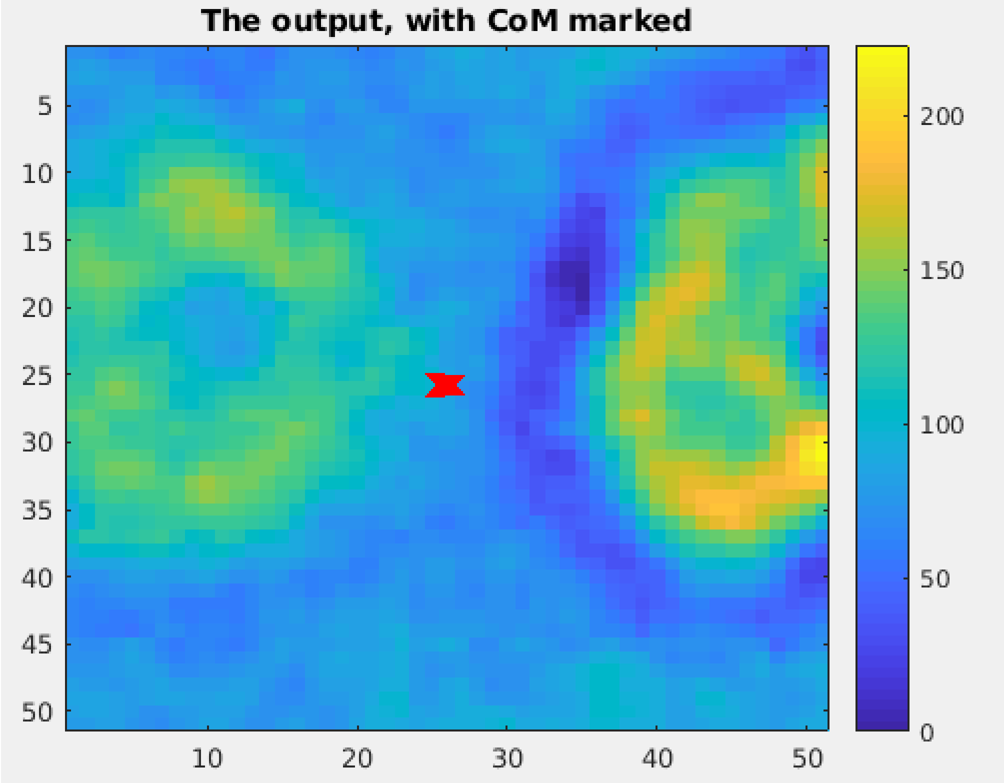


Application to classified data from cryo-EM (2)

Image size: 100x100, 3.54 Å/px







An image whish is easily misleading for identifying particles is used.

Comment on this centering method

- A novel approach for centering cryo-EM data by rotating image is used.
  - The experimental results shows centers are reasonable in the noisy data either in 70S ribosome or betagal.
  - The estimated centers in the classified data are misleading.

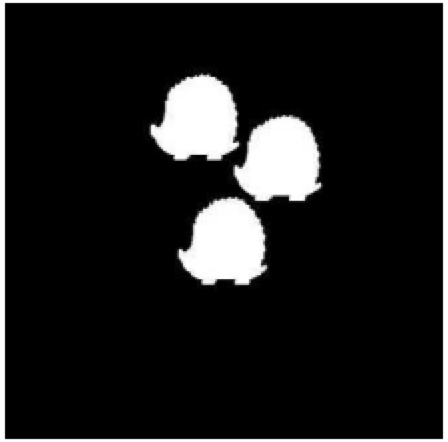
References (1)

[1] A. Lou et al., Differentiating through the Fréchet Mean, arXiv 2003.00335

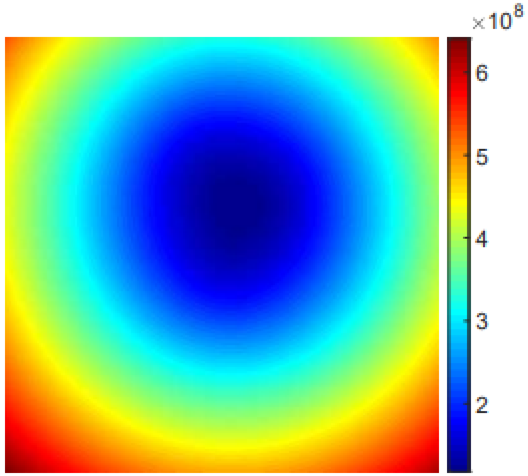
[2] Penczek et al., Three-dimensional reconstruction of single particles embedded in ice, Ultramicroscopy 40: 33 (1992)

Backup

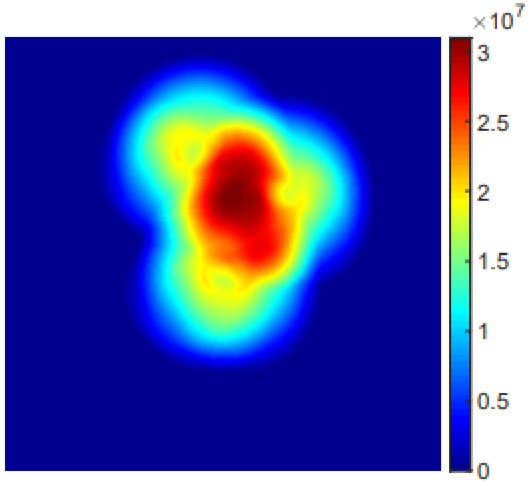
Comparison between the landscape functions



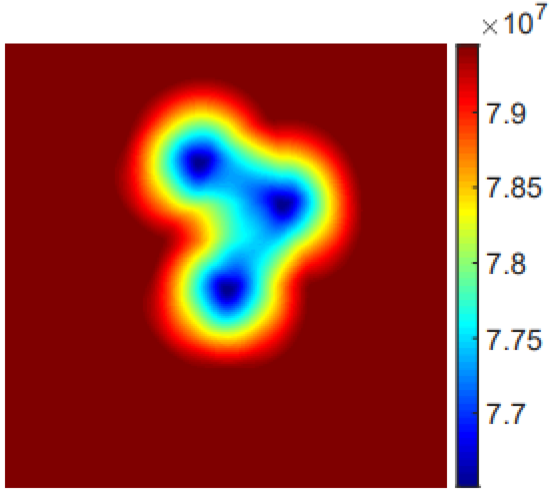
The hedgehog image used for calculating the centers shown on the right



Landscape of the GM



Landscape of the local GM



Landscape of the sCM