Adaptive Undersampling in Spectromicroscopy

Oliver Townsend

October 27, 2021

Introduction

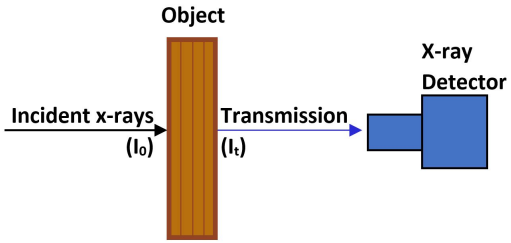
- Spectromicroscopy is a powerful tool used in many disciplines to understand the chemical speciation of a specimen at the microscopic scale.
- Typically the intensity of precise x-ray beams are measured over a 100×100 grid and at 150 energy levels.
- Experiments take several hours, creating a bottleneck.

Proposed Solution

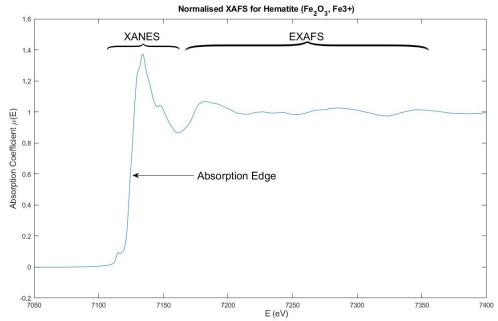
Undersample the measurements taken of a specimen and use numerical techniques to recover the missing entries.

X-ray Absorption Spectroscopy (XAS)

- Used to determine **material composition**.
- Measures the intensity of transmitted x-rays: $I^t(E)$. From this we infer the absorption coefficient: $\mu(E)$.
- Absorption is based on the atoms' core electron excitation energies.
- $I^t(E) = I^0 \exp(-\mu(E)t)$.



(a) Transmission Setup



(b) X-ray Absorption Fine Structure, $mu(E)$

Figure 1: Spectroscopy Illustrations

Scanning Transmission X-ray Microscopy (STXM)

- Spectromicroscopy combines spectroscopy with microscopic techniques - this particular experiment is called STXM.
- STXM repeats the spectroscopy experiment at many points in a grid pattern (raster) at the microscopic level.
- We store the optical density in tensor $D \in \mathbb{R}^{n_E \times n_1 \times n_2}$, so that

$$D_{i,j,k} = -\ln\left[\frac{I^t_{j,k}(E_i)}{I^0}\right] = \mu(E_i)t_{j,k}.$$

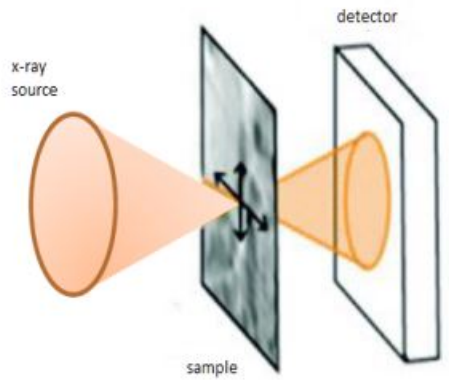


Figure 2: STXM Illustration

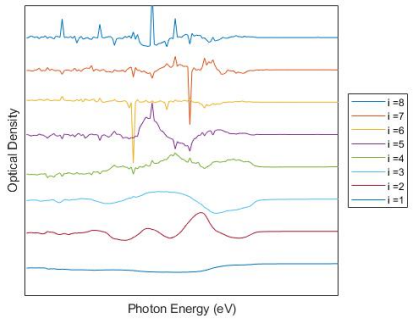
PCA

- Flatten the data: $A \in \mathbb{R}^{n_E \times N}, N = n_1 n_2$.
- We wish to filter out noise.
- Use PCA to characterise the **most significant variations** of the data:

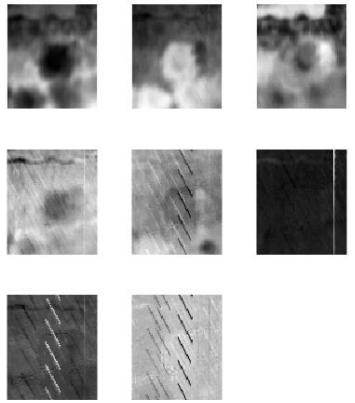
$$A = C \cdot R,$$

with $C \in \mathbb{R}^{n_E \times n_E}$, $R \in \mathbb{R}^{n_E \times N}$.

- Columns of C are the eigenvectors of the covariance matrix $Z = A \cdot A^T$, so orthogonal.
- Define $R = C^{-1}A = C^T A$.



(a) Columns of C



(b) (Reshaped) Rows of R

Choice of Low Rank

To remove noise and interference, we compute **low rank** $A' = C'R'$, where C', R' are the first L columns, rows of C, R respectively.

- Set L at the elbow point of the SVs of A - point of max curvature.
- To find this, use the KNEEDLE algorithm.
- KNEEDLE finds the **maximum orthogonal distance** to the line through the 1^{st} and 20^{th} data point.

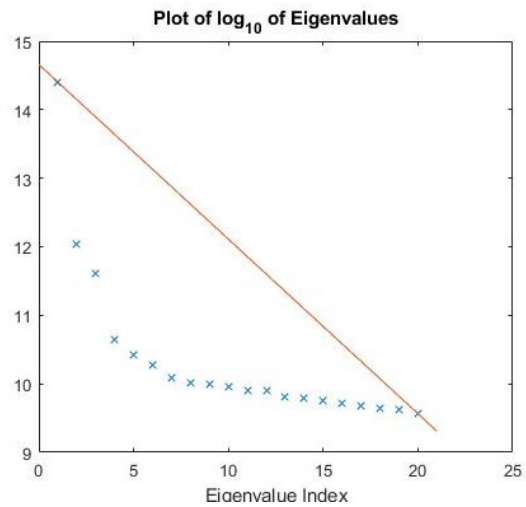


Figure 4: Illustration of Kneedle on DS1

Cluster Analysis

- We identify pixels with similar spectra through cluster analysis.
- Use kmeans to cluster the columns of R' in L-dimensional space.
- Taking the mean of the columns of A corresponding to pixels in a given cluster approximates the absorption spectra of that material.

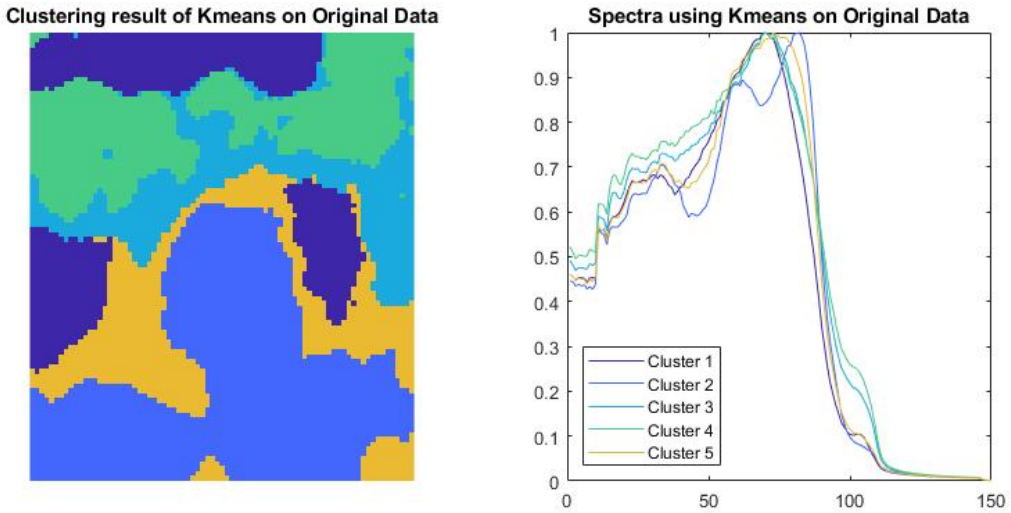


Figure 5: Cluster results of DS2 with $k = 5$.

Low Rank Matrix Completion

When presented with sampled data, assume there is an **underlying low rank** matrix. Find this by solving:

Definition (Low Rank Completion Problem (LRCP) for STXM)

Let Ω be the set of known entries of the data $A \in \mathbb{R}^{n_E \times N}$. Define $\mathcal{P}_\Omega : \mathbb{R}^{n_E \times N} \to \mathbb{R}^{n_E \times N}$ as the orthogonal projection onto the space supported by Omega:

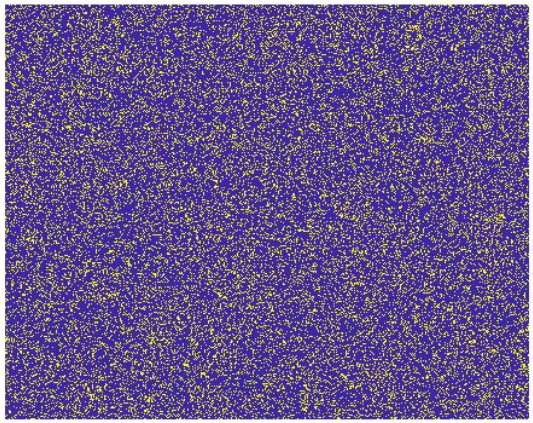
$$(\mathcal{P}_\Omega(A))_{ij} = A_{ij} \text{ if } (i,j) \in \Omega, \quad (\mathcal{P}_\Omega(A))_{ij} = 0 \text{ otherwise.}$$

We wish to solve:

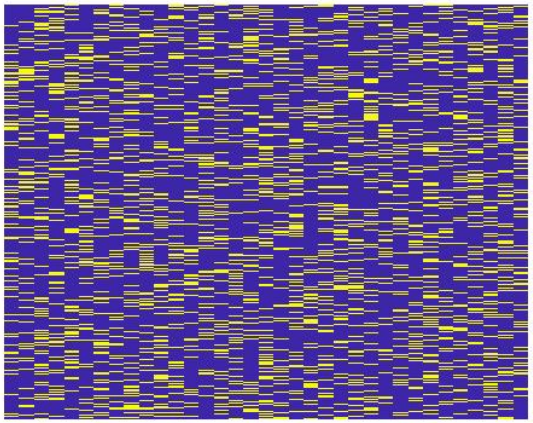
$$\min_{X \in \mathbb{R}^{n_E \times N}} \text{rank}(X), \quad \text{subject to } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(A).$$

Sampling Methods

- Define the **undersampling ratio** $p = |\Omega|/n_E n_1 n_2$.
- Bernoulli Method: Each entry sampled i.i.d. with probability p.
- Raster Sampling: block entries in their (physical) rows for greater experimental efficiency.
- Robust Raster Sampling: Slightly reduce randomness to ensure no zero-rows or zero-columns can occur.



(a) Bernoulli Sampling



(b) Raster Sampling

Alternating Steepest Descent (ASD)

ASD is the method of choice.

- Fix the rank r by setting the following decomposition:

$$A = XY, \quad A \in \mathbb{R}^{n_E \times N}, \; X \in \mathbb{R}^{n_E \times r}, \; Y \in \mathbb{R}^{r \times N}$$

  r must be set as an input.

- We now optimise the following function:

$$\min_{X,Y} f(X, Y) \qquad f(X, Y) = \frac{1}{2}||\mathcal{P}_\Omega(A) - \mathcal{P}_\Omega(XY)||_F^2.$$

- There are also scaled and looped variants: ScaledASD and LoopedASD.

ASD - Algorithm

Each iteration, one component of A = XY is fixed and $f(X, Y)$ is minimised using steepest descent with exact line searchwith respect to the other. This alternates between components.

Write $f(X, Y)$ as $f_X(Y)$ for fixed X, and as $f_Y(X)$ for fixed Y.

ASD - Algorithm

$$\begin{cases} \text{Fix } Y_i, \text{ compute } \nabla f_{Y_i}(X_i) \\ X_{i+1} = X_i - \eta_{X+i} \nabla f_{Y_i}(X_i) \\ \text{Fix } X_{i+1}, \text{ compute } \nabla f_{X_{i+1}}(Y_i) \\ Y_{i+1} = Y_i - \eta_{Y_i} \nabla f_{X_{i+1}}(Y_i) \end{cases}$$

Experimental Results

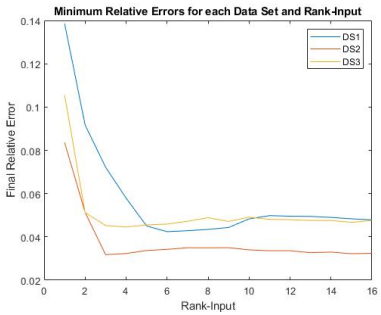
We created several numerical experiments to test all aspects ASD (and other algorithms) when used with spectromicroscopy.

In particular, we have examined ASD's completion results on:

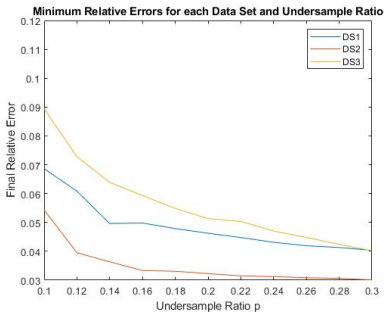
- Completing Spectromicroscopy data.
- The impact of completion on clustering.
- Parameter impact on clustering.

Real Data Results

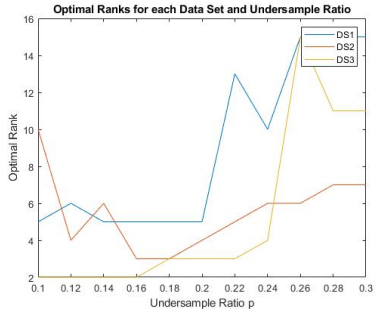
- LoopedASD achieves good completion errors for each data set within 1000 iterations (less than 30s).
- Minimum undersampling ratio of approximately $\hat{p} = 0.15$.
- For this p, optimal undersampling ratios of $r = 5$ for DS1, $r = 3$ for DS2 and DS3.



(a) Min. Completion Error vs Completion Rank



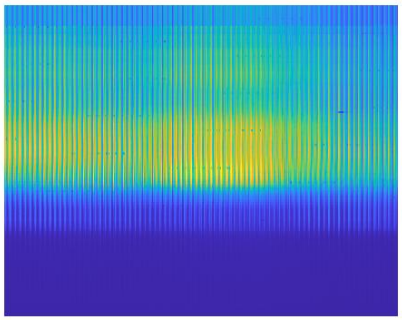
(b) Min. Completion Error vs Undersample Ratio



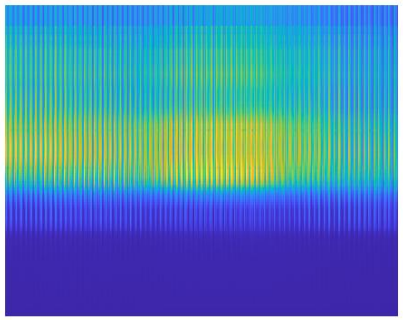
(c) Optimal Completion Rank per Undersample Ratio.

Figure 7: Average Results from Real Data

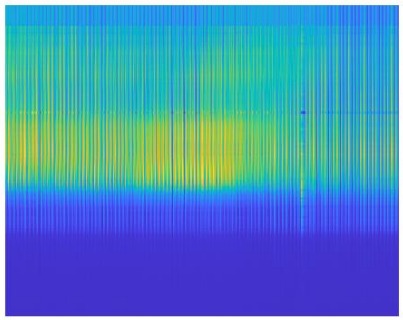
Scaled Images of Completions:



(a) Scaled Colour Image of Data Set 2



(b) LoopedASD completion of Data Set 2 ($p = 0.2,\ r = 5$)

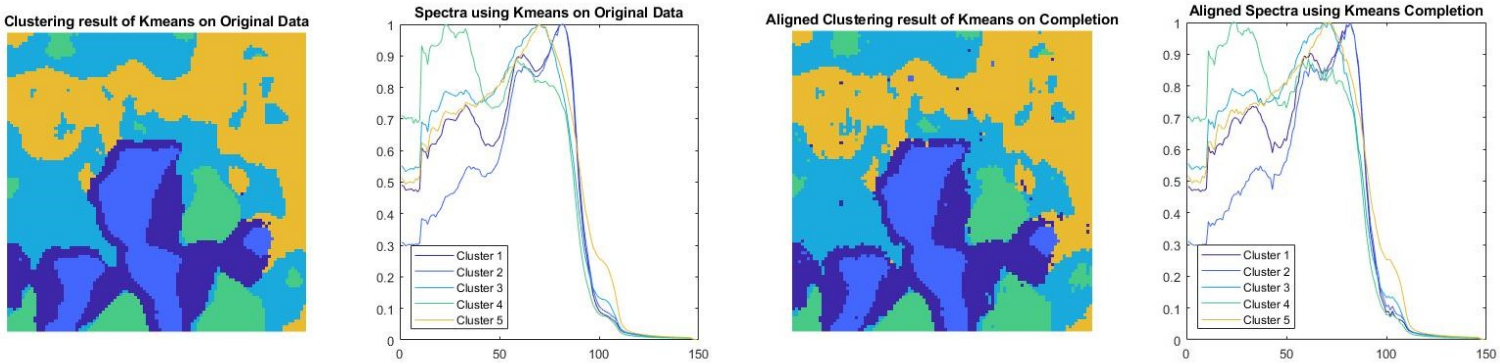


(c) Completion after sampling an artefact

Figure 8: Average Results from Real Data

Clustering

- We now apply PCA and Cluster Analysis to completions.
- Use Aligned Rand Index (ARI) to compare between clusters.
- Some parameter choices can be optimised (e.g. matching L with completion rank, cluster algorithm)
- Other variables depend on researcher aims (e.g. number of clusters).



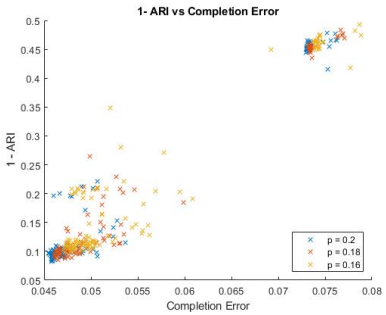
(a) Clustering Results of full data (Data Set 1), with mean absorption spectra

(b) Clustering Results of Completion, $p = 0.15$, $r = 5$, $L = 5$, # clusters $= 5$.

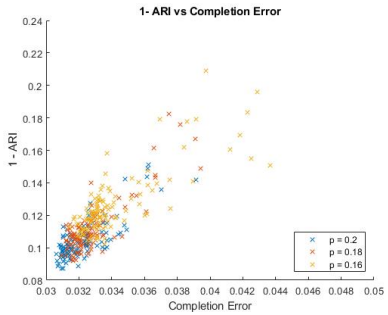
Figure 9: Comparing Clustering Results of completion against full data set.

Effect of Completion Parameters on Clustering

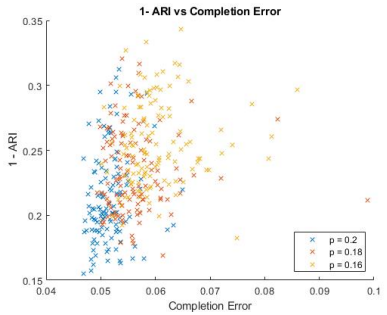
- We check that clusters similarity depends only on the completion error, not on any particular parameters (e.g. completion rank).
- Plot the completion error against $(1 - ARI)$, seeking a linear dependence.



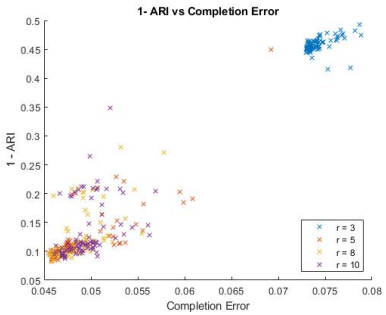
(a) DS1

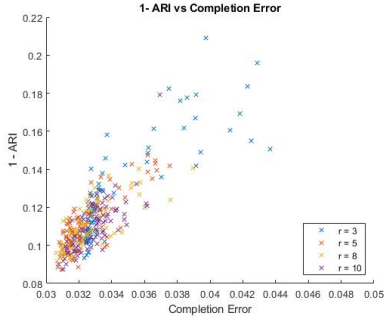


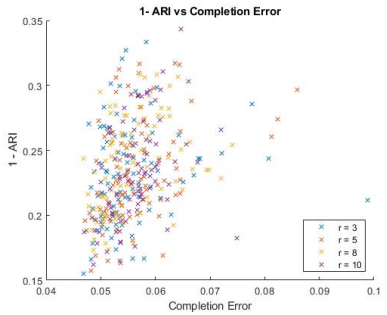
(b) DS2



(c) DS3







Where Next?

- Developing Rank finder to work with real data sets.
- Producing more rigorous guarantees for ASD completion.
- Tensor Completion.
- Non-negative Matrix Factorisation (NMF).
- Ptychography and Tomography.

Flattening the Data

Stack the optical densities of $N = n_1 \times n_2$ pixels to get $A \in \mathbb{R}^{n_E \times N}$.

If there is a **mix of S materials**, the coefficients $\mu(E)$ **sum linearly**.

Now write $\mu \in \mathbb{R}^{n_E \times S}$, $t \in \mathbb{R}^{S \times N}$: columns of μ are the materials' absorption spectra, rows of t are the corresponding thickness maps.

$$A_{i,j} = \sum_{s=1}^{S} \mu_s(E_i) t_j.$$

- $i = 1, ..., n_E$ runs over the energy levels of the x-ray flux,
- $j = 1, ..., N$ indicates the pixels.

Thickness Maps

The final step is to compute smooth thickness maps for each material. We have that,

$$\mu \cdot t \approx A' = C' \cdot R' = (C' \cdot T) \cdot (T^{\dagger} \cdot R),$$

for transformation T. $T^{\dagger}$ exists as long as $k \geq L$, and we associate

$$\mu = C' \cdot T \quad \text{and} \quad t = T^{\dagger} \cdot R$$

We use the mean spectra $\bar{A}$ as a stand in for μ, and set

$$T = (C')^{-1} \cdot \bar{A} = (C')^{T} \cdot \bar{A}$$

$$\begin{aligned} t &= T^{\dagger} \cdot R' \\ &= T^{\dagger} \cdot C^{T} \cdot A' \end{aligned}$$

Completion Algorithm Examples

Three general approaches considered for solving: convex optimisation, projected gradient descent and alternating optimisation. Some examples:

Singular Value Thresholding (SVT)

- Inspired by l_1 minimisation in signal processing, SVT implements convex optimisation through linearized Bregman iterations.
- Involves soft thresholding on the singular values of the iterates.

Singular Value Projection (SVP)

- Utilises a projected gradient method with efficient SVD.
- Truncates the iterates' number of nonzero singular values,
- Requires the true rank as an input.

Rank Finder Algorithm

- We wish to obtain the optimal rank from the sampled data $\mathcal{P}_\Omega(A)$.
- The optimal rank is clear for exact rank matrices - less so for smooth singular value decays.
- Run short trials at different completion ranks, using k-fold cross validation to penalise overfitting.
- True rank minimises the trial error.

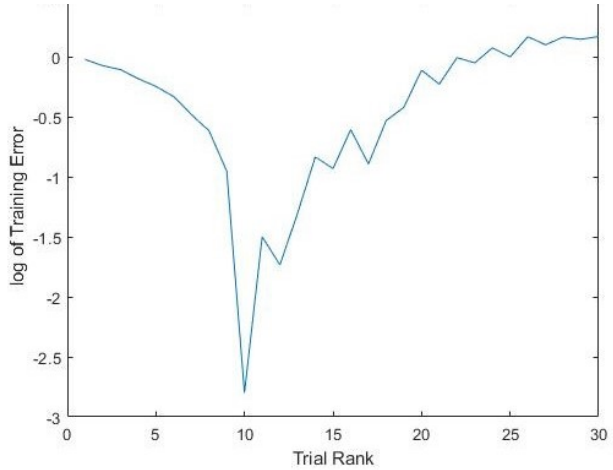
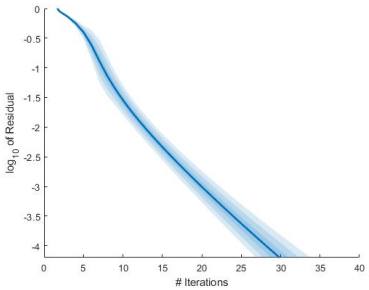


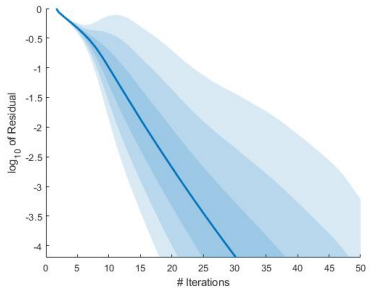
Figure 11: Trial Error for rank 10 matrix, square, sampled by point.

Exact Rank Results

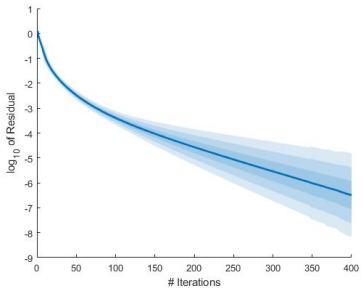
- Format 1: Square, Point Sample
- Format 2: Square, Raster Sample
- Format 3: Wide, Point Sample
- Format 4: Wide, Raster Sample



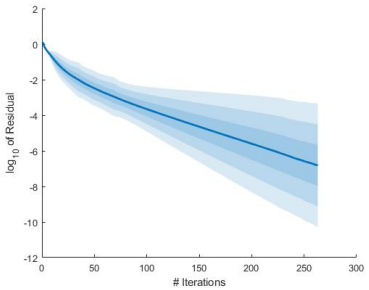
(a) Format 1, $p = 0.06$



(b) Format 2, $p = 0.11$

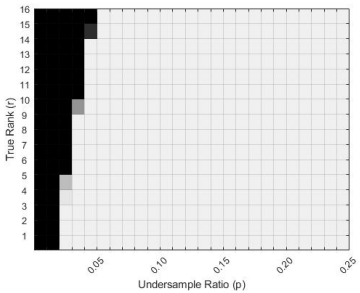


(c) Format 3, $p = 0.14$

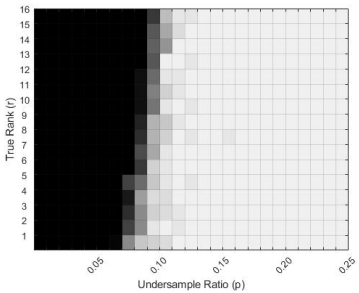


(d) Format 4, $p = 0.15$

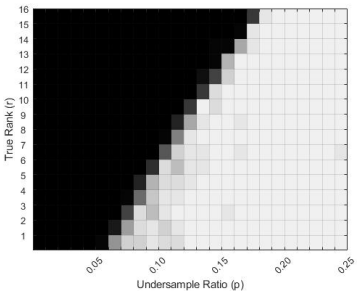
Figure 12: Above: Mean Residual Decay of Iterates (rank = 10). Below: Completion Rates varying parameters



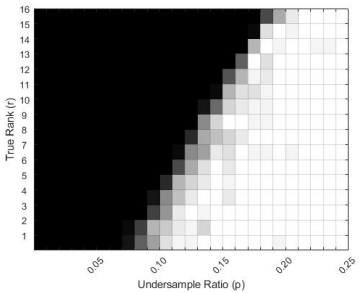
(a) Format 1



(b) Format 2



(c) Format 3



(d) Format 4