REDUCING ACQUISITION TIME AND RADIATION DAMAGE: DATA-DRIVEN SUBSAMPLING FOR SPECTROMICROSCOPY

Maike Meier, Lorenzo Lazzarino, Boris Shustin, Hussam Al-Daas, and Paul Quinn

Mathematical Institute - University of Oxford

Computational Mathematics Theme - STFC UKRI

Internal Seminar, 5th June 2025

University of Oxford
Mathematical Institute

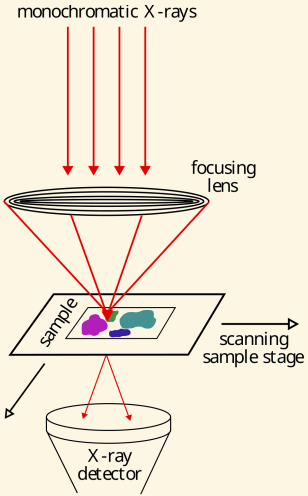
Oxford Mathematics

DATA-DRIVEN SUBSAMPLING FOR SPECTROMICROSCOPY

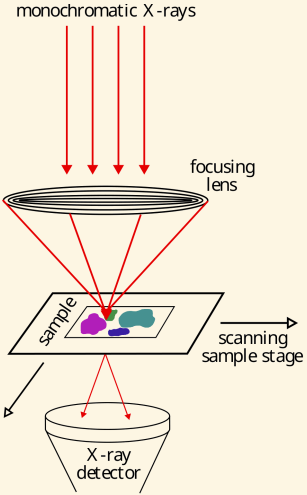
PROBLEM DESCRIPTION

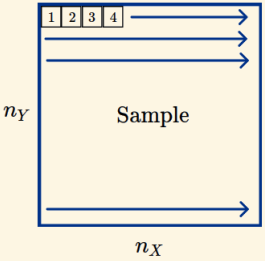
1

SPECTROMICROSCOPY INTRO

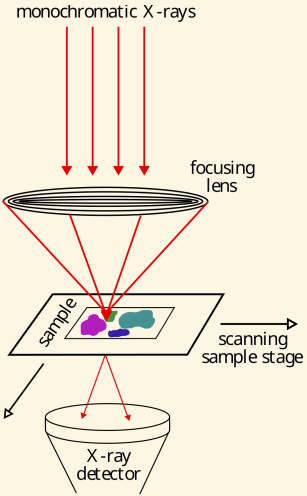


SPECTROMICROSCOPY INTRO

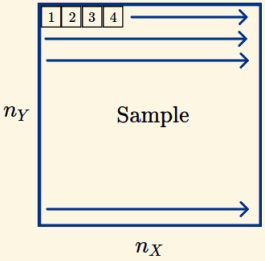




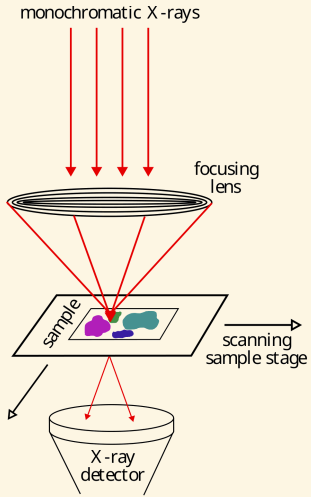
SPECTROMICROSCOPY INTRO



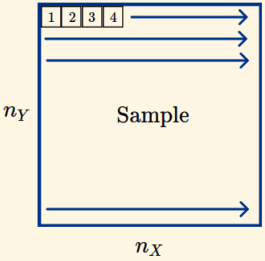
Overall Goal: Determine identity and spacial distributioin of unknown materials contained in the non-homogeneous sample

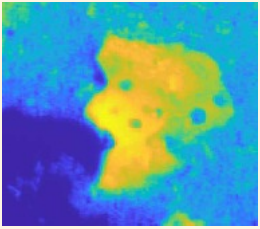


SPECTROMICROSCOPY INTRO

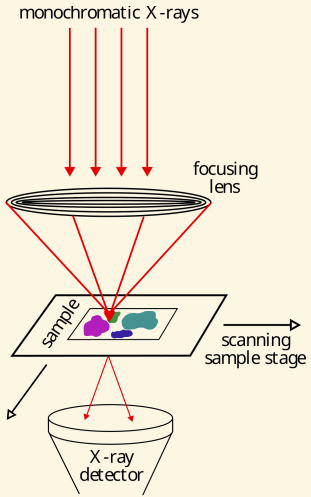


Overall Goal: Determine identity and spacial distributioin of unknown materials contained in the non-homogeneous sample





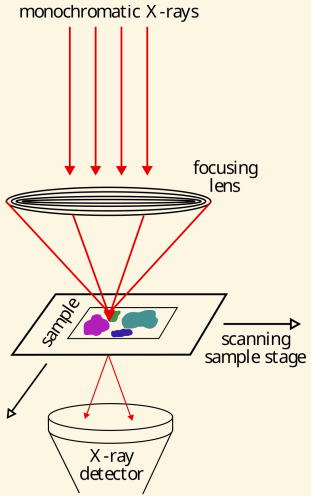
SPECTROMICROSCOPY INTRO *Sub-taks*



Overall Goal: Determine identity and spacial distribution of unknown materials contained in the non-homogeneous sample

1. (Sub)sampling strategy
2. Analysis

SPECTROMICROSCOPY INTRO *Sub-taks*



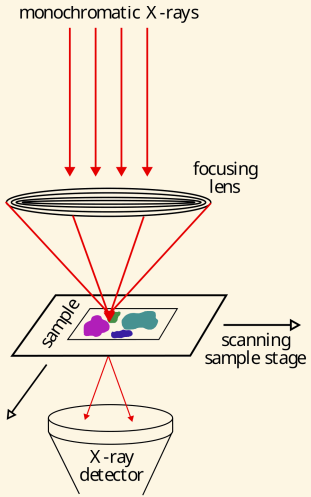
Overall Goal: Determine identity and spacial distribution of unknown materials contained in the non-homogeneous sample

1. (Sub)sampling strategy 2. Analysis

↓

- Measuring the whole dataset takes hours!
- Matrix Completion
- Machinery can't measure a single pixel without measuring its full spatial row/column

SPECTROMICROSCOPY INTRO *Sub-taks*



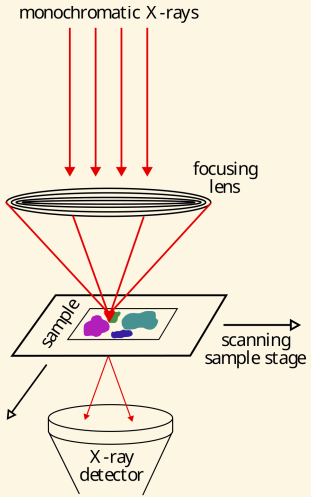
Overall Goal: Determine identity and spacial distribution of unknown materials contained in the non-homogeneous sample

1. (Sub)sampling strategy 2. Analysis

↓

- Compute SVD of completed matrix → eigenspectra (dominant left singular vectors), eigenimage (dominant right singular vectors)
- Cluster analysis on eigenimages to cluster pixels that have similar eigenspectra
- Average absorption spectra for a cluster found by mean of measured data of the cluster

SPECTROMICROSCOPY INTRO *Sub-taks*



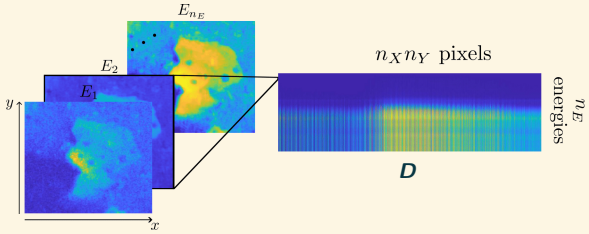
Overall Goal: Determine identity and spacial distribution of unknown materials contained in the non-homogeneous sample

1. (Sub)sampling strategy

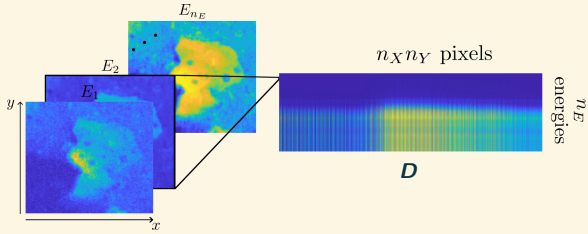
2. Analysis

- Compute SVD of completed matrix $\rightarrow$ eigenspectra (dominant left singular vectors), eigenimage (dominant right singular vectors)
- Cluster analysis on eigenimages to cluster pixels that have similar eigenspectra
- Average absorption spectra for a cluster found by mean of measured data of the cluster

SUBSAMPLING



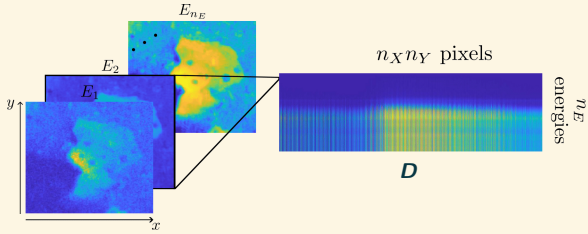
SUBSAMPLING



$$= \boxed{M} \quad \boxed{T}$$

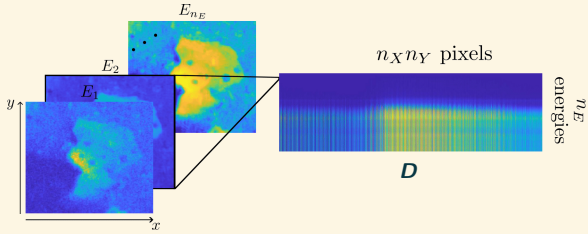
S

SUBSAMPLING



$$S = M \, T + E$$

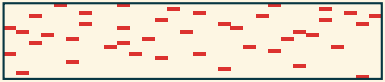
SUBSAMPLING



$$S = M \; T + E$$

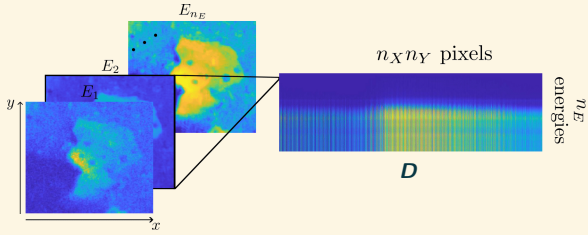
Now

- Uniform Raster sampling



- LoopedASD (Townsend et al, 2022)

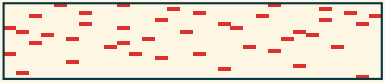
SUBSAMPLING



$$S = M \; T + E$$

Now

- Uniform Raster sampling



- LoopedASD (Townsend et al, 2022)

→ **Importance Sampling**

Leverage scores

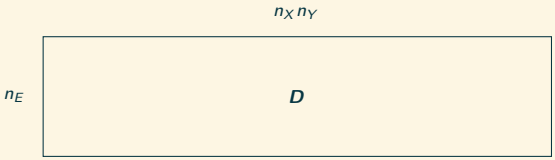
$$\ell_i(\boldsymbol{D})^2 := \|\boldsymbol{U}_{\text{dominant}}(i,:)\|_2^2$$

In principle, $\ell(D) \approx \ell(M)$

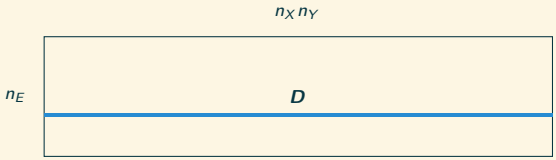
FIND IMPORTANCE DISTRIBUTIONS

2

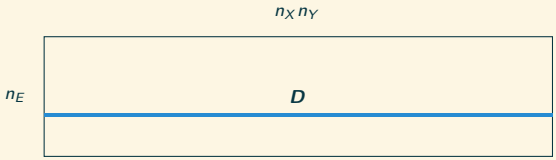
DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION



DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION

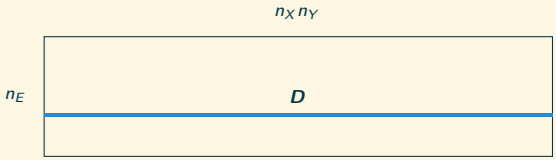


DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION



Problem: We don't have access to $\boldsymbol{D}$ (nor to $\boldsymbol{M}$)

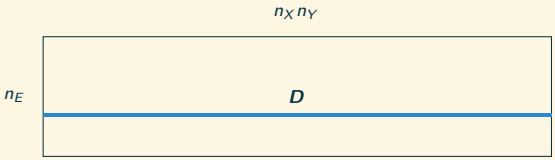
DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION



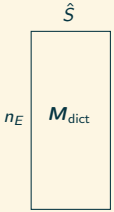
Problem: We don't have access to $\boldsymbol{D}$ (nor to $\boldsymbol{M}$)

Our solution: Data-driven approach

DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION

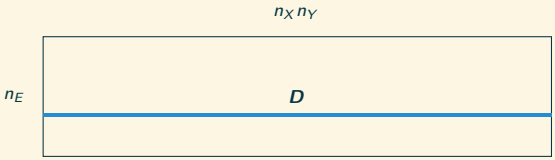


Problem: We don't have access to $\boldsymbol{D}$ (nor to $\boldsymbol{M}$)

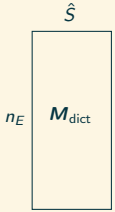


Our solution: Data-driven approach

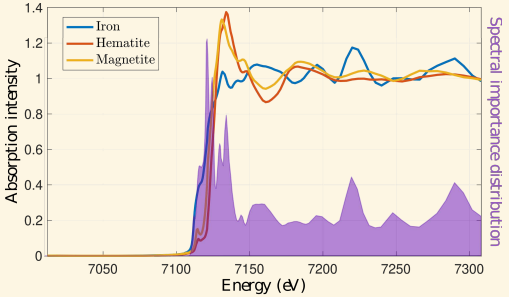
DETERMINING SPECTRAL IMPORTANCE DISTRIBUTION



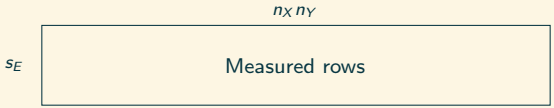
Problem: We don't have access to $\boldsymbol{D}$ (nor to $\boldsymbol{M}$)



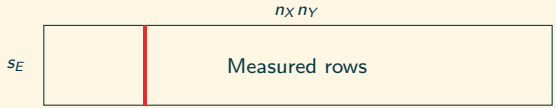
Our solution: Data-driven approach



DETERMINING SPACIAL IMPORTANCE DISTRIBUTION

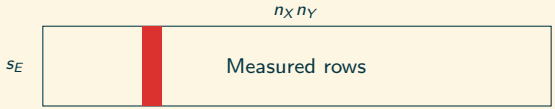


DETERMINING SPACIAL IMPORTANCE DISTRIBUTION



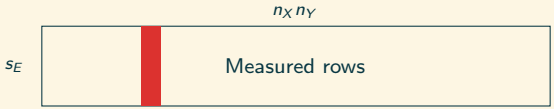
Problem: We can't measure single pixels

DETERMINING SPACIAL IMPORTANCE DISTRIBUTION



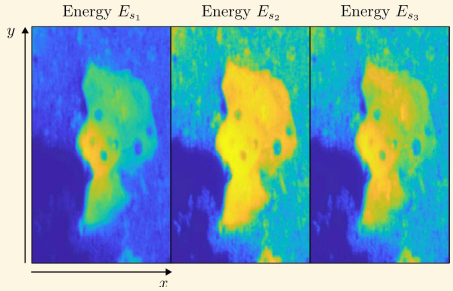
Problem: We can't measure single pixels $\rightarrow$ Goal: Determine importance distribution for block of columns

DETERMINING SPACIAL IMPORTANCE DISTRIBUTION

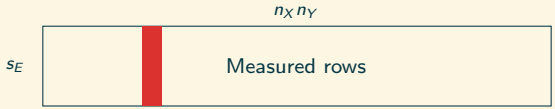


Problem: We can't measure single pixels $\rightarrow$ Goal: Determine importance distribution for block of columns

$\boldsymbol{F} =$

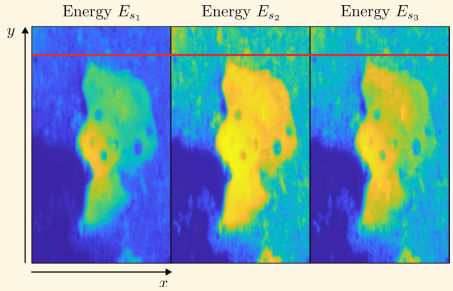


DETERMINING SPACIAL IMPORTANCE DISTRIBUTION



Problem: We can't measure single pixels $\rightarrow$ Goal: Determine importance distribution for block of columns

$\boldsymbol{F} =$



ADAPTIVE RANDOMIZED PIVOTING (ARP)

An algorithm that finds importance distribution by adaptively computing leverage-type scores

(Cortinovis, Kressner, 2024)

ADAPTIVE RANDOMIZED PIVOTING (ARP)

An algorithm that finds importance distribution by adaptively computing leverage-type scores

(Cortinovis, Kressner, 2024)

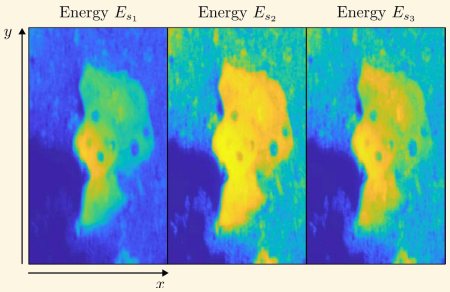
1. Compute $\bar{\boldsymbol{U}} := \boldsymbol{U}_{\text{dominant}}$ for $\boldsymbol{F}$ and leverage scores
2. Select an index i_1 accordingly
3. Update: $\bar{\boldsymbol{U}} \leftarrow \bar{\boldsymbol{U}}$ with $\bar{\boldsymbol{U}}(:, i_1)$ reduced to zero (via Householder reflectors)
4. Compute leverage scores of the updated $\bar{\boldsymbol{U}}$
5. Repeat until: all indices selected, $\bar{\boldsymbol{U}}$ fully reduced

ADAPTIVE RANDOMIZED PIVOTING (ARP)

An algorithm that finds importance distribution by adaptively computing leverage-type scores

(Cortinovis, Kressner, 2024)

1. Compute $\bar{\boldsymbol{U}} := \boldsymbol{U}_{\text{dominant}}$ for $\boldsymbol{F}$ and leverage scores
2. Select an index i_1 accordingly
3. Update: $\bar{\boldsymbol{U}} \leftarrow \bar{\boldsymbol{U}}$ with $\bar{\boldsymbol{U}}(:, i_1)$ reduced to zero (via Householder reflectors)
4. Compute leverage scores of the updated $\bar{\boldsymbol{U}}$
5. Repeat until: all indices selected, $\bar{\boldsymbol{U}}$ fully reduced

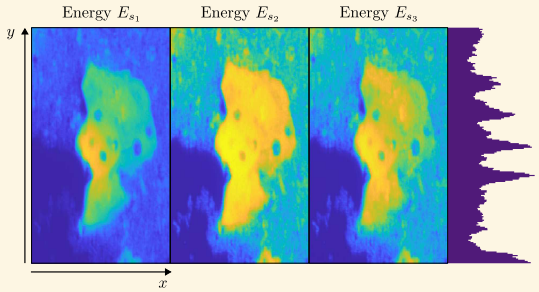


ADAPTIVE RANDOMIZED PIVOTING (ARP)

An algorithm that finds importance distribution by adaptively computing leverage-type scores

(Cortinovis, Kressner, 2024)

1. Compute $\bar{\boldsymbol{U}} := \boldsymbol{U}_{\text{dominant}}$ for $\boldsymbol{F}$ and leverage scores
2. Select an index i_1 accordingly
3. Update: $\bar{\boldsymbol{U}} \leftarrow \bar{\boldsymbol{U}}$ with $\bar{\boldsymbol{U}}(:, i_1)$ reduced to zero (via Householder reflectors)
4. Compute leverage scores of the updated $\bar{\boldsymbol{U}}$
5. Repeat until: all indices selected, $\bar{\boldsymbol{U}}$ fully reduced



GENERAL SCHEME

Algorithm 1

1. Compute leverage scores of $\boldsymbol{M}_{\text{dict}}$
2. Sample s_E energies accordingly
3. Change unfolding $\rightarrow \boldsymbol{F}$
4. Use ARP on $\boldsymbol{F}$

GENERAL SCHEME

Algorithm 1

1. Compute leverage scores of $\boldsymbol{M}_{\text{dict}}$
2. Sample s_E energies accordingly
3. Change unfolding $\rightarrow \boldsymbol{F}$
4. Use ARP on $\boldsymbol{F}$

Obtained: Spectral and spatial importance distributions

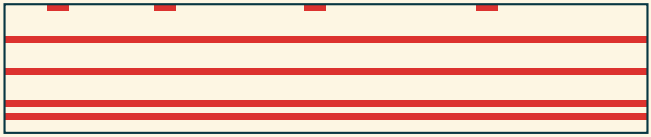
EXPERIMENT DESIGN

3

RASTER IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (RISS)

- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.

RASTER IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (RISS)

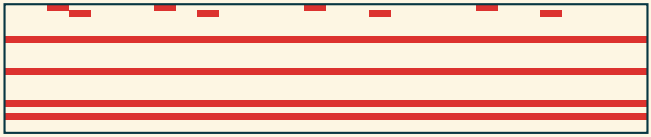


- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.

For the non-measured energies

- Set the beam energy to E.
- Sample s_R spatial rows from the spatial importance distribution
- Measure the sampled spatial rows

RASTER IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (RISS)

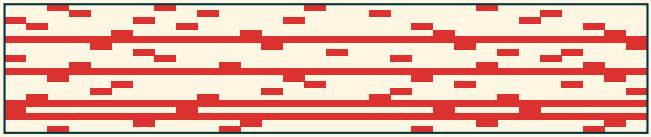


- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.

For the non-measured energies

- Set the beam energy to E.
- Sample s_R spatial rows from the spatial importance distribution
- Measure the sampled spatial rows

RASTER IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (RISS)



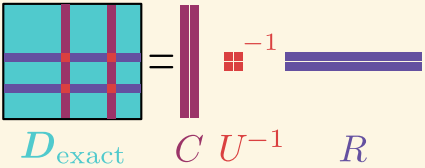
- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.

For the non-measured energies

- Set the beam energy to E.
- Sample s_R spatial rows from the spatial importance distribution
- Measure the sampled spatial rows

- Complete the measured dataset using loopedASD

CUR OVERVIEW



- "Motivation" of importance distributions (theoretical guarantees)
- Natural fit for experimental design
- Gives matrix completion (with interpolation of rows)

CUR IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (CURISS)

- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.
- Sample s_R spatial rows from the spatial importance distribution

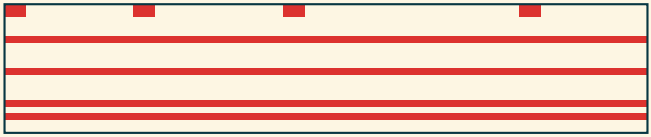
CUR IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (CURISS)

- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.
- Sample s_R spatial rows from the spatial importance distribution

For the non-measured energies

- Set the beam energy to E.
- Measure the sampled spatial rows

CUR IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (CURISS)

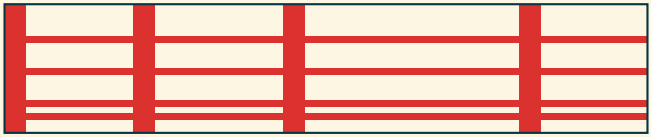


- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.
- Sample s_R spatial rows from the spatial importance distribution

For the non-measured energies

- Set the beam energy to E.
- Measure the sampled spatial rows

CUR IMPORTANCE SAMPLING FOR SPECTRO-MICROSCOPY (CURISS)



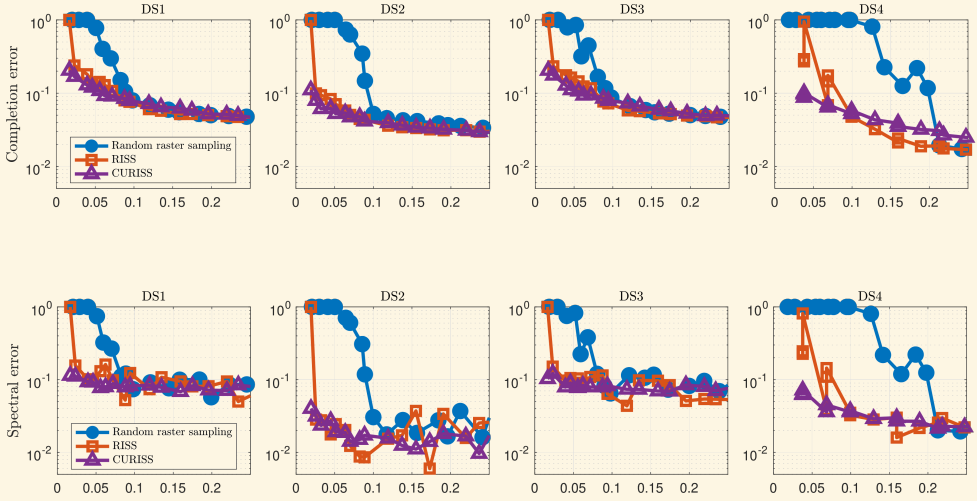
- Use Algorithm 1 to measure a small number of full energy scans at energies and determine a spatial importance distribution.
- Sample s_R spatial rows from the spatial importance distribution

For the non-measured energies

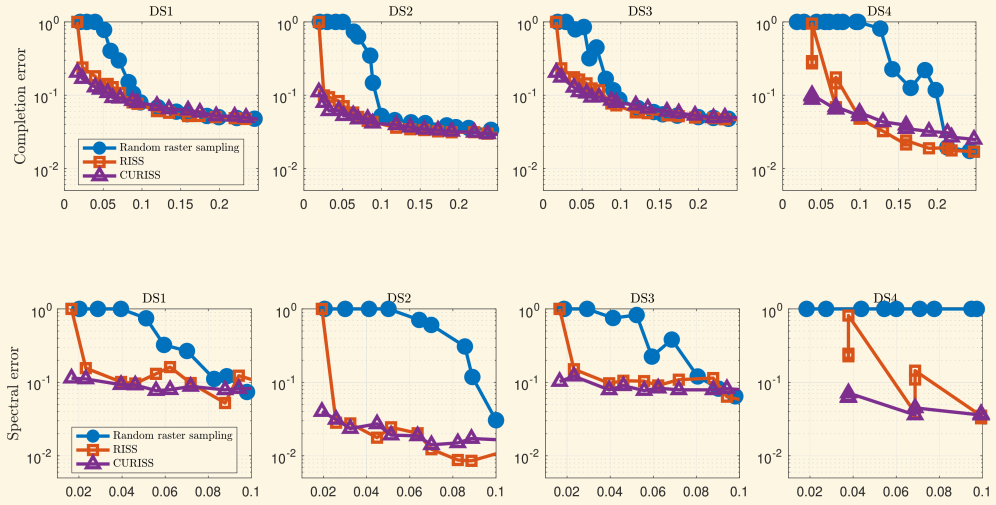
- Set the beam energy to E.
- Measure the sampled spatial rows

- Complete the measured dataset using CUR

RESULTS


DS1
DS2
DS3
DS4
Completion error
Spectral error
Random raster sampling
RISS
CURISS
10^0
10^{-1}
10^{-2}
0
0.05
0.1
0.15
0.2

RESULTS

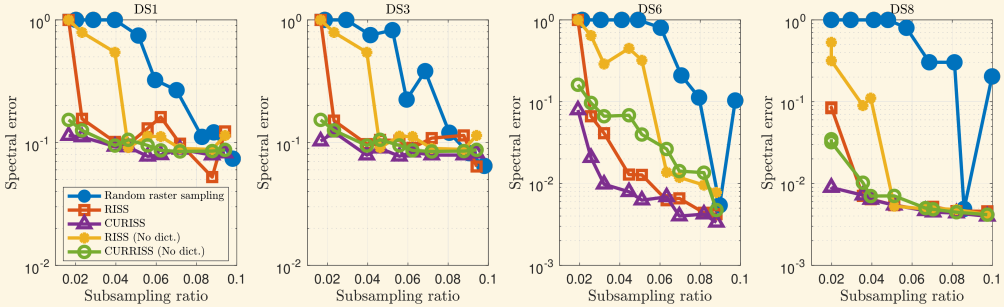

DS1
DS2
DS3
DS4
Completion error
Spectral error
Random raster sampling
RISS
CURISS

RESULTS

What if $\boldsymbol{M}_{\text{dict}}$ is not available? $\rightarrow$ Uniform sampling for energy scans

RESULTS

What if $\boldsymbol{M}_{\text{dict}}$ is not available? → Uniform sampling for energy scans



ADAPTIVE STRATEGY

4

ADAPTIVE CURISS (ACURISS)

Goal: Adaptively refine the subsampling, starting by CURISS with initial subsampling ratio p_0

ADAPTIVE CURISS (ACURISS)

Goal: Adaptively refine the subsampling, starting by CURISS with initial subsampling ratio p_0

Refinment

- How to update the importance distributions?

ADAPTIVE CURISS (ACURISS)

Goal: Adaptively refine the subsampling, starting by CURISS with initial subsampling ratio p_0

Refinment

- How to update the importance distributions?

Stopping Criteria

- How can we evaluate the accuracy of the refined CUR and understand when to stop?

ACURISS *Refinment*

1. Initial CUR by CURISS with p_0
2. Alternate addition of a full energy scan (row of $\boldsymbol{D}$) and a spatial row scan (block of columns of $\boldsymbol{D}$)

ACURISS *Refinment*

1. Initial CUR by CURISS with p_0
2. Alternate addition of a full energy scan (row of $\boldsymbol{D}$) and a spatial row scan (block of columns of $\boldsymbol{D}$)

New energy scan

- Leverage scores of $\boldsymbol{M}_{\text{dict}}$ already computed
- Set to zero for already measured energies

ACURISS *Refinment*

1. Initial CUR by CURISS with p_0
2. Alternate addition of a full energy scan (row of $\boldsymbol{D}$) and a spatial row scan (block of columns of $\boldsymbol{D}$)

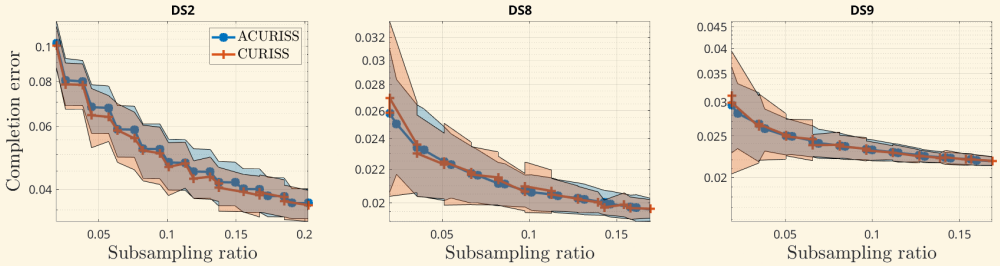
New energy scan

- Leverage scores of $\boldsymbol{M}_{\text{dict}}$ already computed
- Set to zero for already measured energies

New spatial row scan

- Create new $\boldsymbol{F}$
- Do ARP knowing already selected indices

ACURISS VS CURISS



ACURISS *Stopping Criteria*

Goal: Obtain an indicator of the accuracy BUT **Problem:** The full matrix $\boldsymbol{D}$ is not available!!

ACURISS *Stopping Criteria*

Goal: Obtain an indicator of the accuracy BUT **Problem:** The full matrix $\boldsymbol{D}$ is not available!!

Completion variation: Compare the completed matrix in the current refinement step with the one obtained in the previous refinement step

$$\|\boldsymbol{\Delta D}_i\| := \|\hat{\boldsymbol{D}}_i - \hat{\boldsymbol{D}}_{i-1}\|_F \leq \eta_{\boldsymbol{D}},$$

ACURISS *Stopping Criteria*

Goal: Obtain an indicator of the accuracy BUT **Problem:** The full matrix $\boldsymbol{D}$ is not available!!

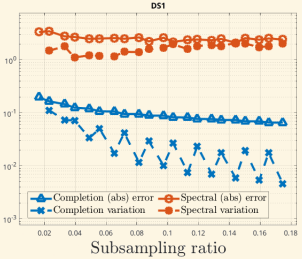
Completion variation: Compare the completed matrix in the current refinement step with the one obtained in the previous refinement step

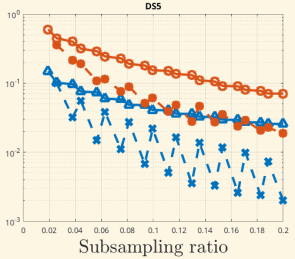
$$\|\boldsymbol{\Delta D}_i\| := \|\hat{\boldsymbol{D}}_i - \hat{\boldsymbol{D}}_{i-1}\|_F \leq \eta_{\boldsymbol{D}},$$

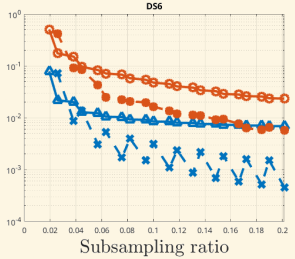
Spectral variation: At each refinement, compute spectral analysis, compare with previous step

$$\|\boldsymbol{\Delta M}_i\| := \|\boldsymbol{M}_{\text{cluster}}(\hat{\boldsymbol{D}}_i) - \boldsymbol{M}_{\text{cluster}}(\hat{\boldsymbol{D}}_{i-1})\|_F \leq \eta_{\boldsymbol{M}}$$

ACURISS *Stopping Criteria*







THANK YOU!

REDUCING ACQUISITION TIME AND RADIATION DAMAGE: DATA-DRIVEN SUBSAMPLING FOR SPECTROMICROSCOPY

Maike Meier, Lorenzo Lazzarino, Boris Shustin, Hussam Al-Daas, and Paul Quinn