Imaging techniques for Raman Spectroscopy

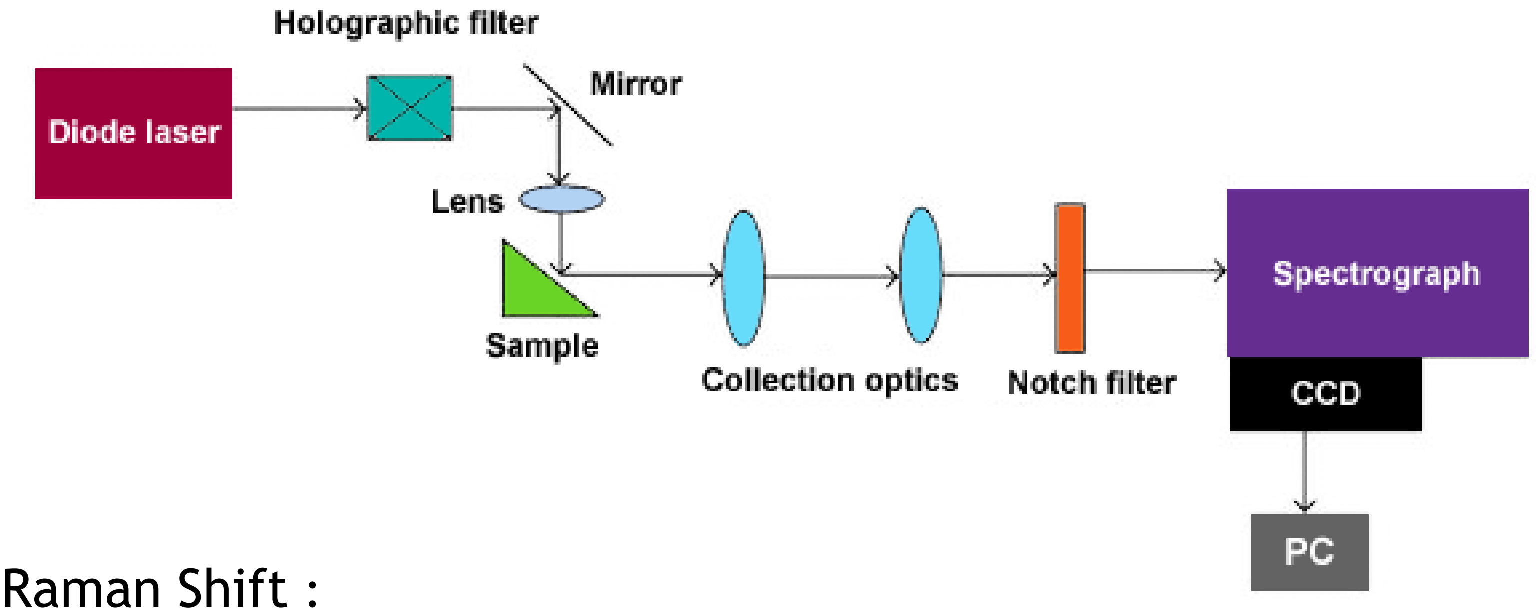
Arijit Pramanik

150020094

Dept. of Computer Science and Engineering

Raman spectral imaging

- Used for structural fingerprint for the identification of molecules

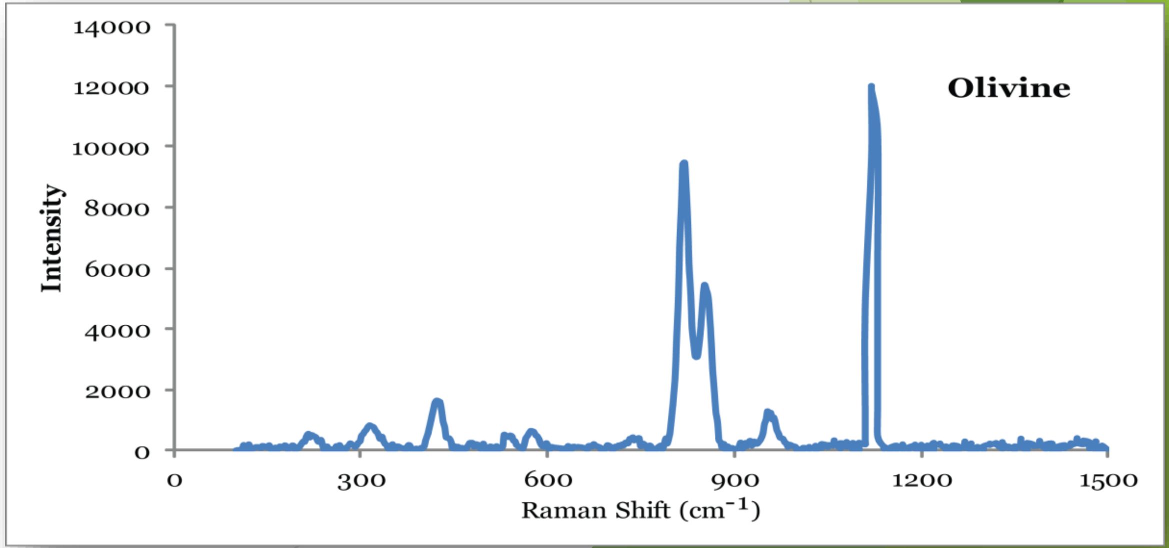


Raman Shift :

$$\Delta w = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)$$

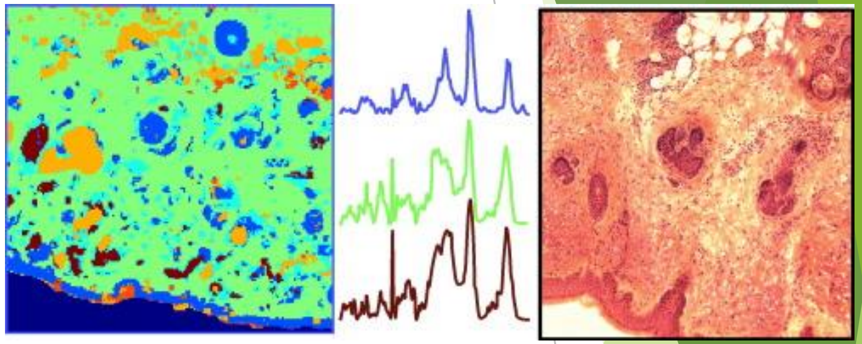
Applications of Raman Spectroscopy

1. Pharmaceuticals and Cosmetics – compound distribution in tablets, contaminant detection
2. Geology and Mineralogy – mineral and phase distribution in rock sections
3. Semiconductors – Purity, alloy composition and superlattice structures
4. Life Sciences – Characterization of bio-molecules and medical diagnosis
5. Carbon Materials – Purity and electrical properties of carbon nanotubes



Let's see its impact on medical diagnostics

The presence or absence of certain spectral bands when Raman imaging is done on biological biopsy samples plays a cardinal role in cancer detection.



However, paraffin is commonly used to preserve samples from decay, but has an intense Raman signature that prevents the study of the underlying tissue.

In vitro Raman analysis can be done on frozen or dewaxed paraffin embedded biopsies.

Any attempt to perform a chemical dewaxing on biopsy tissue can catastrophically alter its signature spectrum!

A few drawbacks..

- High sample acquisition time

  1. Higher resolution images
  2. High desired spectral quality
  3. Some materials have extremely low Raman scattering properties

- Raman signal is weak

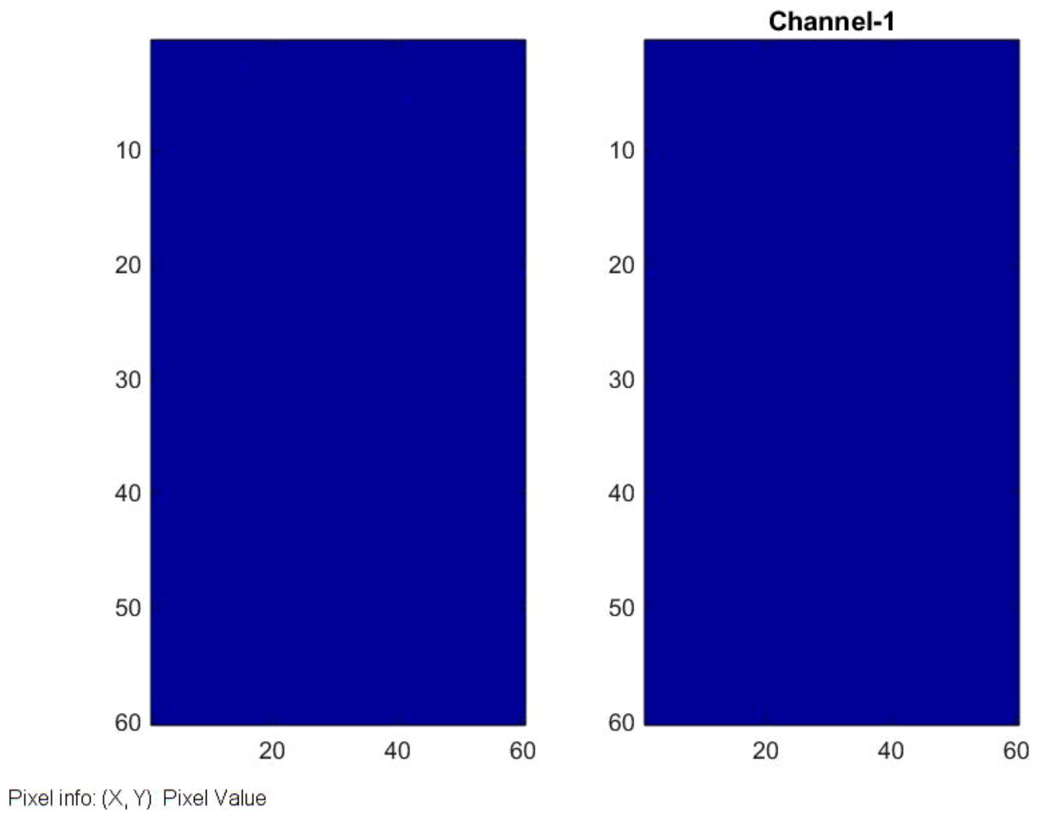
  Only around 1 in every 1 million photons is Raman scattered
  (*Ref* : *http://spie.org/newsroom/combine-and-conquer*)

  Cost-effective CCD detectors are inefficient : *have inherent dark noise, optic responses*.
  Shot noise from fluorescence may further reduce SNR

    1. More scans of the sample
    2. Longer exposure time (might destroy sample)
    3. Higher sample concentration

Our innovation…Compressed Sensing

Given Raman shifts for only a fraction of the entire sample, we reconstruct the Raman shifts for the missing pixels using information from the spectra of the randomly chosen pixels



Advantages

Sampling rate : A sampling rate of 50% will require only half of the original acquisition time

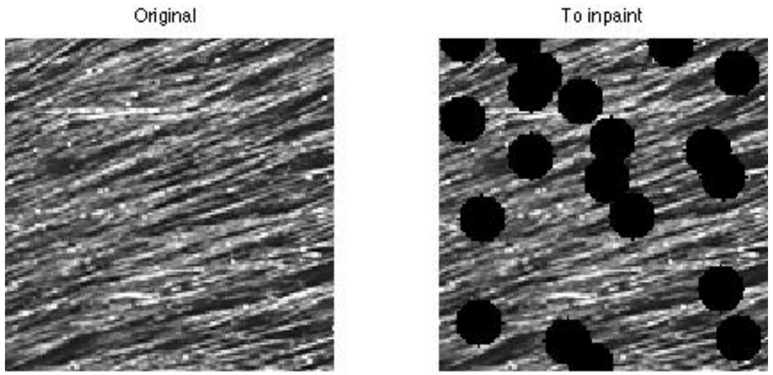
- Can be used for obtaining higher spectral quality for the recorded fraction of pixels
- Complete reconstruction done in software. No expensive hardware like coded apertures required
  (*Ref* : *Colored Coded Aperture Design by Concentration of Measure in Compressive Spectral Imaging -Arguello H. and Gonzalo R.*)
- Random sampling : No complicated sampling heuristic
- Lesser scans help to preserve sample integrity over time
- Denoising of the input image

Proof-of-concept : Inpainting

Inpainting is the process of reconstructing lost or deteriorated parts of images and videos
It is generally used for :

- Removing red-eye, datestamp from photographs
- Replacing corrupted pixels lost in transmission
- Removing logos in videos

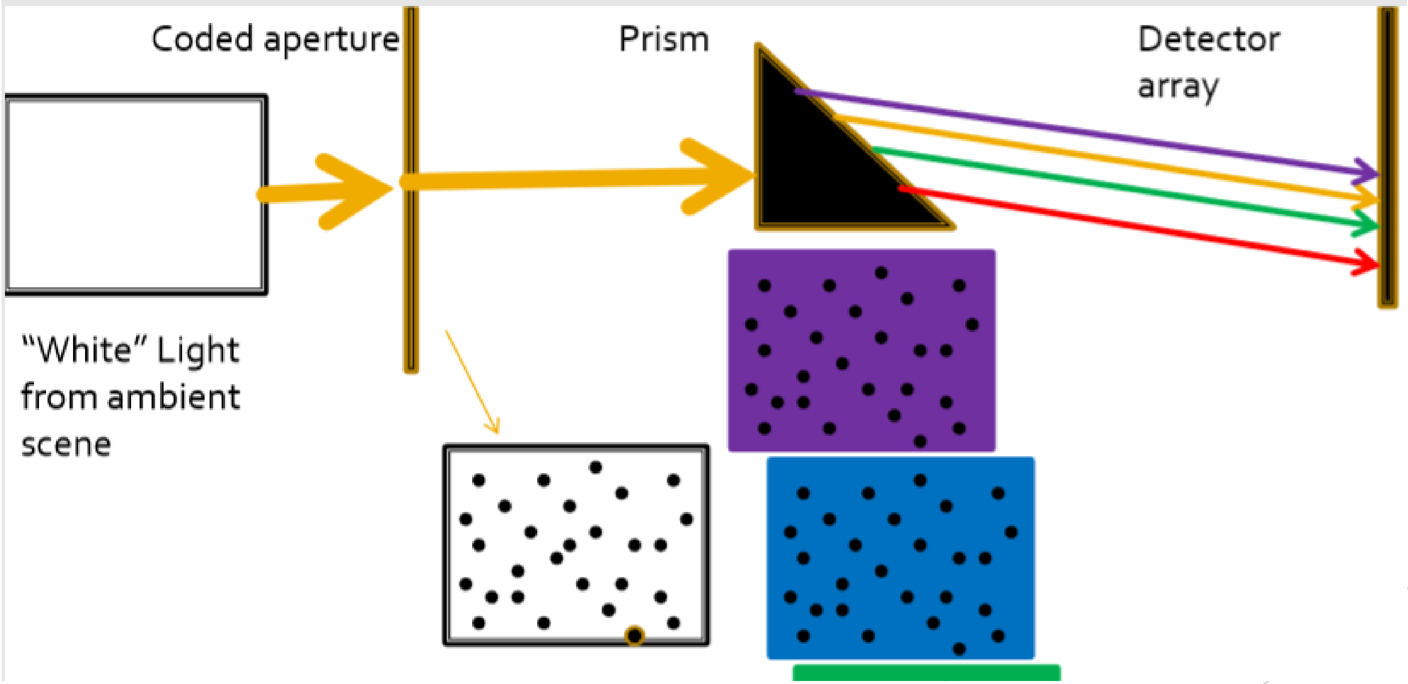
1. Structural inpainting
2. Textural inpainting



Revisting our AIP friend : CASSI

Mapping down the entire 3D spectral datacube to several 2D snapshots by means of dispersive prism induced wavelength dependent shifts and coupled with modulated binary masks

$$M_t(x,y)=\sum_{j=1}^{N_\lambda} S_{j,t}(x,y)=\sum_{j=1}^{N_\lambda} \hat{X}_{j,t}(x-l_j,y)=\sum_{j=1}^{N_\lambda} X_j(x-l_j,y)\bullet C_t(x-l_j,y)$$



Problem Statement

Given an image of m x n x L, where, for only a fraction, f of the m x n pixels, the Raman spectra is recorded, for all the given N channels,

Our goal is to reconstruct the entire image of m x n x L modeling the problem as a form of inpainting, given the undersampled image either by random subsampling or structural subsampling. We demonstrate two approaches for solving this problem taking advantage of the inherent sparsity of natural images :-

1. Using Non-Negative Matrix Factorization (NMF) or Non-Negative Sparse Coding (NNSC)

2. Using a Gaussian mixture model

Problem Statement continued…

Further, given an image of m x n x L , where, for only a fraction f or complete measurement of the m x n pixels, the Raman spectra is recorded, for all the given N channels, (typically for a paraffin preserved tissue sample of interest)

Modeling the problem as a form of source separation, given one of the component signal basis vectors, we try and reconstruct the unknown component. We demonstrate our previous approach utilising Non-Negative Matrix Factorization (NMF) or Non-Negative Sparse Coding (NNSC)

NMF & NNSC

NMF

Given matrix V, V = WH, subject to $v_i >= 0$, $w_i >=0$, $h_i >=0$

Usually, NMF finds applications in fields of astronomy, computer vision, audio signal processing, recommender systems and bioinformatics

More specifically, the approximation of $\mathbf{V}$ by $\mathbf{V} \simeq \mathbf{WH}$ is achieved by minimizing the error function

$$\min_{W,H} ||V - WH||_F, \text{ subject to } W \geq 0, H \geq 0.$$

NNSC

Introducing a sparsity prior for the NMF error minimization function,

$$\min_{W,H} ||V - WH||_F + \lambda \Sigma f(H_{ij}),\ W >= 0,\ H >= 0$$

Applying NNSC to our problem…

Blind Dictionary Learning :

- Extract image patches of size psz *x* psz.
- Let N be the total number of image patches under consideration
- Let L be the total number of channels involved

- Arrange image patches into columns of matrix X [(psz*psz*L) *x* N]
- Linear decomposition of X as X ~ AS
- A [(psz*psz*L) *x* l] represents the *dictionary matrix*, where each column depicts a *feature vector*
- S [l *x* N] represents the *coefficients matrix*, where each column represents *coefficients* of each basis vector in the input vector

Additional constraints…

NNSC : impose *sparsity constraint* on S

$$C(\mathbf{A},\mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|^2 + \lambda \sum_{ij} f(S_{ij}),$$

$$S_{ij} \geq 0 \ , \ A_{ij} \geq 0 \ , \ \|A_i\| = 1$$

All elements of A, S should be non-negative. Columns of A should be of unit norm

Typical choice for *f* : *f(s)* = |S|

λ balances tradeoff between *reconstruction error* and *sparsity*

What is Blind Dictionary Learning?

Static Dictionary Learning :

- Learn A on a set of representative images.
- Use above A to infer S for given test image

J(S) = $\sum_{i} |y_i - \dot{A}s_i|^2 + \lambda \sum_{ij} |s_{ij}|$

Blind Dictionary Learning :

Learn A, S simultaneously on the same image

J(A, S) = $\sum_{i} |y_i - \dot{A}s_i|^2 + \lambda \sum_{ij} |s_{ij}|$

Mask for each patch..

Given randomly sampled input image,

- Different patches have different number of missing pixels, i.e., a patch might have all pixel spectral values recorded, while another might have over half of them missing
- Accounted for by using a mask Φ_i for the i^{th} patch

Modified cost function for our problem hence evolves as :

$$J(A,S) = \sum_i |y_i - \phi_i A s_i|^2 + \lambda \sum_{ij} |s_{ij}|$$

y_i – actual image patch vector $[r_i \; x \; 1]$
Φ_i - mask for the image patch $[r_i \; x \; (psz*psz*L)]$
S_i – column of S $[l \; x \; 1]$
A – dictionary matrix $[(psz*psz*L) \; x \; l]$

Need for different $\phi_i' s$

$$\sum_i \left|y_i - \phi_i A_{s_i}\right|^2 + \lambda \sum_i \left|s_{ij}\right| = J(A, S)$$

Differentiating wrt A, we get ,

$$-2\phi_i^t\left(y_i - \phi_i A_{s_i}\right)s_i^t = 0$$

$$\sum_i \phi_i^T y_i s_i^t = \sum_i \phi_i^t \phi_i A_{s_i} s_i^t$$

$$A = \left(\sum_i \phi_i^t \phi_i s_i s_i^t\right)^{-1} \left(\sum_i \phi_i^t y_i s_i^t\right)$$

Learning A, S simultaneously

We update A, S via a stochastic gradient descent algorithm.

Illustration for using the squared error (Euclidean distance) for minimizing cost between X and AS

A is updated via projected gradient descent, with an adaptive step size

$$A_{n+1} = A_n - \theta\left(\sum_i \phi_i^t(\phi_i A_n s_{ni} - X)s_{ni}^t\right)$$

S is updated via a multiplicative update rule. M_{ni} represents the i[th] column,

$$M_{ni} = \frac{(\phi_i A_{n+1})^t y_i}{\left((\phi_i A_{n+1})^t(\phi_i A_{n+1})s_{ni} + \lambda\right)}$$

$s_{(n+1)i} = s_{ni} {}^{\circ} M_{ni}$ where ° represents element-wise multiplication

Additional constraints…

We are employing Projected Gradient Descent, so

- Any negative entry in A is set to zero

- All columns in A are enforced to be of unit norm

  → To maximize sparsity, we can multiply an arbitrarily large value to elements of A, and scale down elements of S by the same factor, which minimizes the overall objective function. However, such a value of A is clearly not desirable

Super-resolution a.k.a structural sampling

- Here, the ϕ_i's are aren't randomly sampled, and are quite structural in nature
- For a p x p patch, there are only p^2 such matrices possible, much lesser than in the previous case of random sampling
- The equation shown on slide 14 will not have full rank, leading to a more difficult problem for reconstruction
- Hence, we start off with an initial guess. There are two proposed methods of which we leverage the latter

→ *Infer a dictionary on a part of the image which shall be used as an initial guess for the dictionary to be learnt for the entire image*

→ *Obtain an estimate of the image using bicubic interpolation. The dictionary inferred on the interpolated image is used as a seed for learning the dictionary on the given raw undersampled image*

Discussion

We see that the RMSE error value increases as the number of pixels sampled strictly decreases. We see particularly good preservation of spatial features as well with dictionary learning and see convincing results for smooth images.
For images with complex textures, the textures are not faithfully reconstructed.

For the super-resolution problem, we see a superior performance of the reconstructed image over the initial interpolated image used as the seed for dictionary inference guess.

Moving onto GMMs

Conventional Compressed Sensing to Statistical Compressed Sensing

- O(klog(n/k)) measurements VS O(k) measurements
- The average error of Gaussian SCS is tightly upper bounded by the best k-term linear approximation. The estimator error also has a closed-form formula

However, there are certain restrictions on GMMs so that a good reconstruction is obtained

- The covariance matrix needs to have good eigenvalue decay
- For higher dimensional signals, energy should be concentrated in the first few dimensions
- The Gaussian components should be 'orthogonal'

Decoding a signal from compressive measurements

$$\Delta(\Phi x) = argmax_x p(x|y) = \Sigma\Phi^T(\Phi\Sigma\Phi^T)^{-1}(\Phi x)$$

Here, $x \in \mathbb{R}^n$ is the signal
$\Phi \in \mathbb{R}^{m\,x\,n}$ is the binary mask (sensing matrix)
$y \in \mathbb{R}^m$ is the measured vector

Given a GMM, characterized by $\{\mu_i, \Sigma_i\}_{i=1}^N$,

where N is the number of components,
μ_i is the mean of ith component of size n
Σ_i is the covariance matrix of ith component of size n x n

In this case, we need $(\phi\Sigma\phi^t)$ to be of full rank, and hence in this case, the noise in the measurements adds to the rank of this above matrix.

Adapting it to our setting

Now, suppose we are given a set of measurement vectors $\{y_i\}_{\mathrm{i}=1}^{n}$, we seek to reconstruct the set of original signals $\{x_i\}_{\mathrm{i}=1}^{n}$ as follows :

Using the piecewise linear estimate leveraging the optimal decoder

$$\tilde{x}_j = \Sigma_j \Phi^T (\Phi \Sigma_j \Phi^T + \Sigma_\eta)^{-1} (y - \Phi \mu_j) + \mu_j$$

$$\tilde{j} = argmin_j \|y - \Phi \tilde{x}_j\|^2_{\Sigma_\eta} + \|\tilde{x}_j - \mu_j\|^2_{\Sigma_j} + log|\Sigma_j|$$

The above equation holds for estimating a signal x through its measurement vector y given the sensing matrix Φ with noise, i.e, $y = \Phi x + n$ using the GMM $\{\mu_i, \Sigma_i\}_{i=1}^{N}$

where $\|z\|_A$ denotes $z^T A^{-1} z$

Our problem formulation

So, for a given ensemble of signals $\{x_i\}_{i=1}^{n}$, we seek to minimize the following objective function :

$$J(\{x_i\}_{i=1}^{n}) = \sum_{i=1}^{n} \frac{\|y_i - \phi_i \tilde{x}_{\tilde{j}}\|^2}{2\sigma^2} + \|\tilde{x}_{\tilde{j}} - \mu_j\|_{\Sigma_j}^2 + \log|\Sigma_j|$$

This is accomplished using the MAP-EM algorithm where :

E-step : $(\tilde{x}, \tilde{j})$ = $argmax_{(x,j)} f(x|y, \mu_j, \Sigma_j)$

M-step : $\mu_j = \frac{1}{S_j} \sum_i w_{ij} \tilde{x}_i,$ $\Sigma_j = \frac{1}{S_j} \sum_i w_{ij} (\tilde{x}_i - \mu_j)(\tilde{x}_i - \mu_j)^T$

Discussion

At a higher sampling rate, the GMMs generally outperform the dictionary learning methods, but not quite so in the lower sampling regime as was guaranteed by SCS. The reason being that, the initial estimate of the mean and covariance are quite important as down the line, they yield the tuned means and variances.

It might be the case that the inferred Gaussian components do not show significant eigenvalue decay, and are not 'orthogonal' enough to each other. We know that, the model selection is accurate in the case of the these high-dimensional signals being such that the energy is concentrated in the first few principal dimensions. In only such a case, reconstruction with GMMs will work quite well with high-dimensional signals as in our case even with low sampling ratios.

Spectral Separation

This follows from the dictionary learning problem wit the only exception being that we decompose it into two dictionaries : A_ρ and A_s , corresponding to the paraffin and the tissue. So, we look forth to optimizing on the below mentioned objective function, with updates exactly the same as for dictionary learning with projected gradient descent with adaptive step size:

$$J(A_s, \{s_{pi}\}_{i=1}^{n}, \{s_{si}\}_{i=1}^{n}) = \|g_{si} - \Phi_i A_p s_{pi} - \Phi_i A_s s_{si}\|^2 + \lambda\|s_{pi}\|_1 + \lambda\|s_{si}\|_1,$$

$$\text{such that } A_s \succeq 0, \forall i(s_{pi} \succeq 0, s_{si} \succeq 0)$$

$$\forall j \in \{1, ..., K_s\}, \|A_{s_{\cdot,j}}\|_2^2 = 1$$

where A_ρ - paraffin dictionary of size n x k_1

A_s - tissue dictionary of size n x k_2

λ - tradeoff parameter for sparsity (can be kept different)

s_{s_i} - ith coefficient column for the tissue dictionary of size k_1 x 1

s_{pi} - ith coefficient column for the pure paraffin dictionary of size k_2 x 1

ϕ_i - sensing matrix for the ith patch of size m x n

$\geq$ denotes the element wise non-negativity constraint

Discussion

The spectral separation even with the iterative method doesn’t work quite as well as expected, owing to the sample spectra being overwhelmingly paraffin. Some of the skin estimate always leaks in into the paraffin component and hence is irrecoverable.

This also has the limitation of obtaining sample and pure paraffin spectra to be on the same scale since ad-hoc approaches for identifying peaks for scaling them is not generalizable across all such mixture problems.

Results..

- Silicon wafer image (71 X 71 X 201) with different sampling rates of 20%, 50% and 80%
- Silicon wafer image (101 X 101 X 201) with different sampling rates of 20%, 50% and 80%
- Silicon wafer image (41 X 41 X 201) with different sampling rates of 20%, 50% and 80%

- Paraffin

- Si + microcrystals of GaN

- Estimate of skin from paraffin + skin sampe

Future Work

- Implement deep neural networks for the problem of inpainting
- Perform undersampling in the spectral dimension for wavelengths as well

THANK YOU