MULTI-SCALE BAYESIAN RECONSTRUCTION OF COMPRESSIVE X-RAY IMAGE

Jiaji Huang, Xin Yuan, and Robert Calderbank

Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708, U.S.A

ABSTRACT

A novel multi-scale dictionary based Bayesian reconstruction algorithm is proposed for compressive X-ray imaging, which encodes the material's spectrum by Poisson measurements. Inspired by recently developed compressive X-ray imaging systems [1], this work aims to recover the material's spectrum from the compressive coded image by leveraging a reference spectrum library. Instead of directly using the huge and redundant library as a dictionary, which is cumbersome in computation and difficult for selecting those active dictionary atoms, a *multi-scale* tree structured dictionary is refined from the spectrum library, and following this a Bayesian reconstruction algorithm is developed. Experimental results on real data demonstrate superior performance in comparison with traditional methods.

Index Terms— Compressive sensing, multi-scale, dictionary, poisson, X-ray imaging

1. INTRODUCTION

Dictionary based representations have demonstrated effectiveness in signal and image recovery problems [2]. Recently, multi-scale dictionary has received much attention due to its better depiction of the data geometry [3] compared with its single-scale counterpart [2]. Multi-scale representation achieves stronger signal restoration ability [4], and has been effectively applied to many machine learning tasks, including classification [5], novelty detection [6] and topic modelling [7].

A note on the notations before introducing our problem: matrices and vectors are denoted as bold upper and lower case letters, e.g., $\mathbf{D}$ is a matrix and $\boldsymbol{\theta}$ is a vector. Scalars are denoted as plain letters, e.g., ℓ, Z. In this paper, we introduce multi-scale dictionary based representation into solving the Poisson compressive sensing (CS) inverse problem. In a Poisson CS inverse problem, measurements $\mathbf{y} \in \mathbb{Z}_+^M$ are from a Poisson distribution,

$$\mathbf{y} \sim \text{Pois}(\mathbf{H}\mathbf{f} + \boldsymbol{\mu}), \tag{1}$$

where $\mathbf{H} \in \mathbb{R}_+^{M\times N}$ characterizes the system. $\mathbf{f} \in \mathbb{R}_+^N$ is the underlying signal (vectorized image) and $\boldsymbol{\mu}$ is the "dark current". The task is to estimate $\mathbf{f}$ given $\mathbf{y}$, $\mathbf{H}$ and $\boldsymbol{\mu}$. While the maximum likelihood estimator [8] is effective, recently, significant improvement has been achieved by exploring the image's nature, *e.g.*, smoothness and sparsity [9, 10, 11].

The Poisson CS algorithm proposed in this paper is inspired by an X-ray machine—Coded Aperture Coherent Scatter Spectral Imaging (CACSSI) [1, 12]. This machine allows simultaneous measure of location and molecular signature of a material under test. Thus scan time is significantly reduced in comparison with traditional X-ray imaging systems. The imaging mechanism in CACSSI is governed by the Bragg's law, $q = \frac{1}{2d} = \frac{E}{hc}\sin\frac{\delta}{2}$, where q is the momentum transfer, d is the effective lattice spacing of the material, E is the energy of the X-ray, h is the Planck's constant, c is the speed of light in vacuum, and δ is the angle between the incident and scatted X-ray. Different materials occupy different q values. When an X-ray is projected on a material, it scatters at one or several specific angles with specific energy. The angle and energy information is recorded by spatially distributed energy sensitive pixels. The resultant X-ray image at the detector is modeled as Poisson distributed due to its limited energy.

To reconstruct $\mathbf{f}$ and estimate the spectrum, the point process based recovery algorithms, *e.g.*, Maximum Likelihood Estimator (MLE) [8], Maximum Posterior with TV penalty (MAP-TV) [13], have been shown effective [1, 14]. However, they can not well accommodate the uncertainty in background noise and can only give a point estimation. Herein, we adopt a Bayesian framework which simultaneously estimates the background noise $\boldsymbol{\mu}$ and the signal $\mathbf{f}$. And particularly, significant improvement in reconstruction is achieved by leveraging a multi-scale dictionary.

2. PROBLEM STATEMENT

In Eq. (1), the sensing matrix $\mathbf{H}$ of the X-ray machine can be estimated in advance by forward model. The underlying signal is considered as a two dimensional image [1], with one dimension $z = 1, \dots Z$ representing the spatial location (discretized grids), and the other dimension signifying the momentum transfer, $q = 1, \dots, Q$. Specifically, suppose that a material with spectrum $\mathbf{s} \in \mathbb{R}_+^Q$ occupies a subset $\Omega \subset \{1, 2, \dots, Z\}$ of these spatial grids. Then the input signal induced by this material is an image $\mathbf{F} \in \mathbb{R}_+^{Q\times Z}$, whose z-th column is $\mathbf{s}$ if $z \in \Omega$ and zero otherwise. And the $\mathbf{f}$ in equation (1) can be expressed as $\mathbf{f} = \text{vec}(\mathbf{F})$, stacking $\mathbf{F}$'s columns on one another.

By cascading all the reference spectra in the library, we can build a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_R] \in \mathbb{R}_+^{Q \times R}$, where R is the total number of spectra. And we express $\mathbf{F}$ as a linear combination of the atoms in $\mathbf{D}$, *i.e.*,

$$\mathbf{F} = \mathbf{D}\boldsymbol{\Theta}, \tag{2}$$

where $\boldsymbol{\Theta} \in \mathbb{R}_+^{R \times Z}$ linearly combines these dictionary atoms at each spatial grid. Denoting $\boldsymbol{\theta} = \text{vec}(\boldsymbol{\Theta}) \in \mathbb{R}^{RZ}$ and combining Eq (1) and (2), we can express the model as

$$\mathbf{y} \sim \text{Pois}(\mathbf{H}(\mathbf{I}_Z \otimes \mathbf{D})\boldsymbol{\theta} + \boldsymbol{\mu}). \tag{3}$$

The problem is to estimate the $\mathbf{F}$ given the $\mathbf{H}$ and imperfect measure of $\boldsymbol{\mu}$, which boils down to an estimate of $\boldsymbol{\theta}$. Obtaining $\mathbf{F}$, we can further estimate the spectrum by simply extracting the most intense column in $\mathbf{F}$.

As aforementioned, directly using this $\mathbf{D}$ is computationally expensive and difficult for selecting active atoms. Hence, we propose an approach to better leverage the dictionary and improve spectrum recovery performance. While our primary focus is on X-ray signal recovery, our work also sheds light on the widely existing signal recovery problems in photon limited imaging systems.

3. ALGORITHM

Rather than simply cascading all the reference spectra, we construct a multi-scale dictionary to refine the atoms and reduce the redundancy. Following this, we reconstruct the signal from the multi-scale dictionary. The flow-chart of the full algorithm is depicted in figure 1.

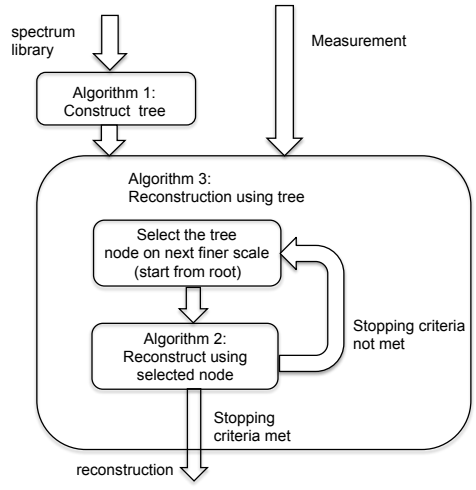


Fig. 1: Flow-chart of our approach

3.1. Construct multi-scale dictionary

We perform a nested partition of the over-complete dictionary to group similar spectra. Various methods can be employed for the partition, *e.g.*, iterative k-means [15], graph-cut [16], etc. Here we use the iterative k-means, where the dictionary atoms are partitioned into k subsets and each subset is further recursively partitioned. This results in a tree structure where each node $\mathcal{T}(\ell, n_\ell)$ is associated with a set of spectra, $\mathbf{P}(\ell, n_\ell)$. Here $\ell = 0, \ldots, L$ denotes the level of the node in the tree, with $\ell = 0$ implying the root node, and $n_\ell = 1, \ldots, N_\ell$ indices the node in the ℓ-th level. We then prune redundancy from the $\mathbf{P}(\ell, n_\ell)$ to get a refined dictionary $\mathbf{D}(\ell, n_\ell)$. As summarized in algorithm 1, Non-negative Matrix Factorization (NMF) [17] is applied to each $\mathbf{P}(\ell, n_\ell)$. The rank, $R(\ell, n_\ell)$, is chosen such that the product of two factor matrices is close enough to $\mathbf{P}(\ell, n_\ell)$. Notice that each node in the tree is a dictionary (a matrix). It is thus natural to ask the question, how to select the "best" node (dictionary) for representation. And we will present a strategy in the next section.

Algorithm 1 Construct Multi-scale Dictionary

Input: Reference spectrum library $\mathbf{D}$, NMF error tolerance ϵ
Output: Tree $\{\mathcal{T}(\ell, n_\ell)\}$ with refined dictionary $\{\mathbf{D}(\ell, n_\ell)\}$

```
Partition the atoms in D using iterative k-means. Tree node T(ℓ, n_ℓ) is associated with spectra P(ℓ, n_ℓ)
for all ℓ = 0, ..., L,  n_ℓ = 1, ..., N_ℓ do
    /*factorize each P(ℓ, n_ℓ) with appropriate rank*/
    rank R(ℓ, n_ℓ) ← 0
    factors B ← 0, W ← 0
    while ||P(ℓ, n_ℓ) − BW||_F^2 / ||P(ℓ, n_ℓ)||_F^2 > ε do
        R(ℓ, n_ℓ) ← R(ℓ, n_ℓ) + 1
        Perform rank-R(ℓ, n_ℓ) NMF of P(ℓ, n_ℓ) and assign the two factors to B, W.
    end while
    D(ℓ, n_ℓ) ∈ R_+^{Q × R(ℓ, n_ℓ)} ← column-normalized B
end for
```

3.2. Bayesian Reconstruction of $\mathbf{F}$ using $\{\mathbf{D}(\ell, n_\ell)\}$

Given the constructed multi-scale dictionary $\{\mathbf{D}(\ell, n_\ell)\}$ and measurements $\mathbf{y}$, we need to find the "best" node for reconstruction. Since a dictionary at a deeper level is supposed to be a finer representation of the to-be reconstructed spectrum, we seek the "best" node by selecting a path from root to leaf in $\{\mathcal{T}(\ell, n_\ell)\}$. Specifically, suppose we have arrived at node $\mathcal{T}(\ell - 1, n_{\ell-1})$, we then recover $\mathbf{F}$ using each dictionary associated with each child of $\mathcal{T}(\ell - 1, n_{\ell-1})$. We select the one with the smallest cost as the next node to land on. At this point, it is necessary to study how to reconstruct $\mathbf{F}$ given a particular dictiontary.

3.2.1. A Bayesian Model

We consider how to estimate $\boldsymbol{\Theta}(\ell, n_\ell) \in \mathbb{R}_+^{R(\ell, n_\ell) \times Z}$ (therefore the $\mathbf{F}$), given a dictionary $\mathbf{D}(\ell, n_\ell)$. Denote $\boldsymbol{\theta}(\ell, n_\ell) = \text{vec}\,[\boldsymbol{\Theta}(\ell, n_\ell)] \in \mathbb{R}^{ZR(\ell, n_\ell)}$ as the vectorized $\boldsymbol{\Theta}(\ell, n_\ell)$. Define

$$\mathbf{A} \triangleq \mathbf{H}(\mathbf{I}_Z \otimes \mathbf{D}(\ell, n_\ell)) \in \mathbb{R}^{M \times ZR(\ell, n_\ell)}. \tag{4}$$

Referring to the model (3), we have

$$\mathbf{y} \sim \text{Pois}(\mathbf{A}\boldsymbol{\theta}(\ell, n_\ell) + \boldsymbol{\mu}). \tag{5}$$

We are thus directed to infer $\boldsymbol{\theta}(\ell, n_\ell)$. We impose the following conjugate prior:

$$\begin{array}{lcl} \boldsymbol{\theta}(\ell, n_\ell) & \sim & \text{Gamma}(\alpha_\theta, \beta_\theta) \\ \boldsymbol{\mu} & \sim & \text{Gamma}(\beta_\mu \cdot \hat{\boldsymbol{\mu}}, \beta_\mu) \end{array} \tag{6}$$

where $\hat{\boldsymbol{\mu}}$ is a measure of the "dark current" by placing no material in the X-ray body. Note that $\hat{\boldsymbol{\mu}}$ is not the perfect knowledge of $\boldsymbol{\mu}$. This fact would be more conveniently accounted by a bayesian instead of an optimization based approach. Also, Gamma is the conjugate prior of Poisson thus analytical form of posterior can be derived. And finally, shrinkage effect (sparsity) on $\boldsymbol{\theta}(\ell, n_\ell)$ can be easily imposed by setting the shape parameter $\alpha_\theta \leq 1$.

3.2.2. Bayesian Inference

Now we derive the inference in detail by following the Poisson Factor Analysis (PFA) [18] steps. We first introduce latent variables

$$\begin{array}{l} y_{m,n} \sim \text{Pois}\left(\mathbf{A}_{m,n}\boldsymbol{\theta}_n(\ell, n_\ell)\right), \forall\ 1 \leq n \leq ZR(\ell, n_\ell) \\ y_{m,ZR(\ell,n_\ell)+1} \sim \text{Pois}(\boldsymbol{\mu}_m),\ \forall\ 1 \leq m \leq M, \end{array} \tag{7}$$

where m, n are the row and column indices of $\mathbf{A}$. Assuming independency among the $y_{m,n}$'s, the m-th measurement $\mathbf{y}_m$ can be generated as

$$\mathbf{y}_m = \sum_{n=1}^{ZR(\ell,n_\ell)+1} y_{m,n}, \tag{8}$$

due to the additivity of independent Poisson variables. Lemma 4.1. in [18] states that

$$\begin{array}{l} (y_{m,1}, \ldots, y_{m,ZR(\ell,n_\ell)+1}) | \mathbf{y}_m \\ \qquad \sim \text{Mult}(\mathbf{y}_m; \xi_{m,1}, \ldots, \xi_{m,ZR(\ell,n_\ell)+1}), \end{array}$$

where

$$\begin{array}{l} \xi_{m,n} = \dfrac{\mathbf{A}_{m,n}\boldsymbol{\theta}_n(\ell, n_\ell)}{\sum_{n=1}^{N} \mathbf{A}_{m,n}\boldsymbol{\theta}_n(\ell, n_\ell) + \boldsymbol{\mu}_m},\ \forall\ 1 \leq n \leq ZR(\ell, n_\ell) \\ \text{and} \\ \xi_{m,ZR(\ell,n_\ell)+1} = \dfrac{\boldsymbol{\mu}_m}{\sum_{n=1}^{N} \mathbf{A}_{m,n}\boldsymbol{\theta}_n(\ell, n_\ell) + \boldsymbol{\mu}_m} \end{array} \tag{9}$$

We are thus able to sample the latent $y_{m,n}$'s from a multinomial distribution given the $\mathbf{y}$ and the current estimate of $\boldsymbol{\theta}(\ell, n_\ell)$ and $\boldsymbol{\mu}$. With the $y_{m,n}$'s obtained, the posterior of $\boldsymbol{\theta}_n(\ell, n_\ell)$ can be derived as

$$\begin{array}{rl} p(\boldsymbol{\theta}_n(\ell, n_\ell)|-) \propto & \prod_{m=1}^{M} \text{Pois}(y_{m,n}; \mathbf{A}_{m,n}\boldsymbol{\theta}_n(\ell, n_\ell)) \\ & \times \text{Gamma}(\boldsymbol{\theta}_n(\ell, n_\ell); \alpha_\theta, \beta_\theta) \\ = & \text{Gamma}\left(\alpha_\theta + \sum_{m=1}^{M} y_{m,n}, \beta_\theta + \sum_{m=1}^{M} \mathbf{A}_{m,n}\right) \end{array} \tag{10}$$

Algorithm 2 Infer $\boldsymbol{\theta}(\ell, n_\ell)$

Input: Dictionary $\mathbf{D}(\ell, n_\ell)$ at tree node $\mathcal{T}(\ell, n_\ell)$, $\mathbf{y}$, $\mathbf{H}$, "dark current" measurements $\hat{\boldsymbol{\mu}}$, parameters α_θ, β_θ, β_μ
Output: Dictionary atom weights $\boldsymbol{\theta}(\ell, n_\ell)$

Compute $\mathbf{A}$ by equation (4)
Initialize $\boldsymbol{\theta}(\ell, n_\ell)$, $\boldsymbol{\mu}$ by equation (6)
for loop=1, . . . ,MAX_LOOPS **do**
  Compute $\xi_{m,n}$ by equation (9)
  /*Assign the mean of the posteriors to the variables*/
  $y_{m,n} \leftarrow \xi_{m,n} \cdot \mathbf{y}_m$
  $\boldsymbol{\theta}_n(\ell, n_\ell) \leftarrow \left(\alpha_\theta + \sum_{m=1}^{M} y_{m,n}\right) / \left(\beta_\theta + \sum_{m=1}^{M} \mathbf{A}_{m,n}\right)$
  $\boldsymbol{\mu}_m \leftarrow \left(\alpha_\mu + y_{m,N+1}\right) / \left(\beta_\mu + 1\right)$
end for

And similarly, the posterior of $\boldsymbol{\mu}_m$ is

$$p(\boldsymbol{\mu}_m|-) = \text{Gamma}\left(\alpha_\mu + y_{m,ZR(\ell,n_\ell)+1}, \beta_\mu + 1\right). \tag{11}$$

Full conjugation of the parameters allows us to perform the efficient mean-field variational Bayes [19] inference, where the mean values of the posterior distributions in (9)-(11) are used in the update equations. The inference is summarized in Algorithm 2. Note that the computational and memory cost in each loop in algorithm 2 depends on the dictionary size linearly. Thus in general, using the refined dictionary will be more efficient than using an over-complete one.

3.2.3. Path selection guided by reducing risk

We start from the root node, and select a path leading to the "best" node. Suppose we have arrived at node $\mathcal{T}(\ell, n_\ell)$ at

Algorithm 3 Reconstruct spectrum using tree $\{\mathcal{T}(\ell, n_\ell)\}$

Input: $\{\mathbf{D}(\ell, n_\ell)\}$, $\mathbf{y}$, $\mathbf{H}$, $\hat{\boldsymbol{\mu}}$, parameters α_θ, β_θ, β_μ
Output: Reconstructed $\widehat{\mathbf{F}}$

Reconstruct using root dictionary $\mathbf{D}(0, 1)$ by algorithm 2. Get risk $\mathcal{C}[0] = \mathcal{C}(0, 1)$ by equation (12)
$\ell \leftarrow 0$, $n_\ell^* \leftarrow 1$
while $\mathcal{T}(\ell, n_\ell^*)$ is not tree leaf **do**
  For each child of $\mathcal{T}(\ell, n_\ell^*)$, get its $\boldsymbol{\theta}(\ell+1, n_{\ell+1})$ using algorithm 2, compute its risk by equation (12)
  Take the child with the smallest risk, assign this risk to $\mathcal{C}[\ell+1]$, index to $n_{\ell+1}^*$
  if $\mathcal{C}[\ell+1] > \mathcal{C}[\ell]$ **then**
    break
  else
    $\ell \leftarrow \ell + 1$
  end if
end while
$\widehat{\mathbf{F}} \leftarrow \mathbf{D}(\ell, n_\ell^*)\boldsymbol{\Theta}(\ell, n_\ell^*)$

level ℓ, we then examine the cost of reconstructing the spectrum via algorithm 2 using each child of $\mathcal{T}(\ell, n_\ell)$. The $\ell+1$-th node to land on will be the one with the smallest risk. We stop delving deeper if the cost of using any of $\mathcal{T}(\ell, n_\ell)$'s children is larger than that of itself. A cost function is defined as the negative log-likelihood with certain penalty (*e.g.*,TV, ℓ_1) on the reconstructed spectrum $\hat{\mathbf{s}}$

$$\mathcal{C}(\ell, n_\ell) \triangleq -\log \text{Pois}(\mathbf{y}; \mathbf{A}\boldsymbol{\theta}(\ell, n_\ell)) + \lambda \text{pen}(\hat{\mathbf{s}}) \qquad (12)$$

The entire reconstruction framework is summarized in algorithm 3.

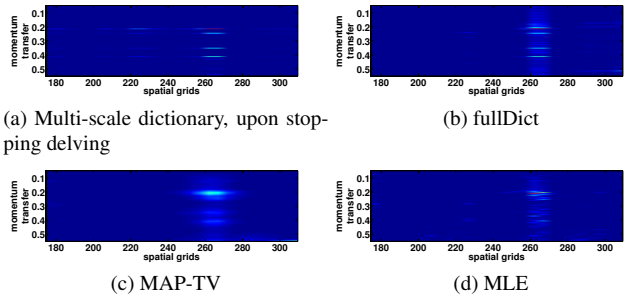
4. EXPERIMENTS

We test the performance of our algorithm on the real data collected by CACSSI. The comparators are MLE [8], MAP-TV [13], and reconstruction using the entire spectrum library as dictionary (fullDict), which is an over-complete one since $Q = 101$ and $R = 166$. Due to the limit of space, we only present two representative materials, aluminum and water.

Typically, the spectra of solids (e.g., aluminum) are spiky, whereas liquids' spectra are smooth and continuous. When partitioning these two categories of spectra, their difference might be difficult to capture by a commonly used distance metric, *e.g.*, Euclidean distance. Therefore, merely using k-means with Euclidean distance may not partition the spectrum library appropriately. So we first inspect the wavelet coefficients of all spectra. Those tend to occupy high frequency bands are spiky ones, otherwise are smooth ones. We separate smooth ones from spiky ones. This forms a first level bi-partition of the entire spectrum library. Then we further partition these two subsets recursively by iterative 2-means as in algorithm 1, where the NMF error tolerance ϵ is set to 0.05. Finally, we get a multi-scale dictionary, whose root node has 26 atoms. The two children of root represent liquid and solid. And they are with 2 and 23 atoms respectively. The smaller size of the liquids' dictionary is due to the fact that liquids' spectra well resemble each other.

The parameters for reconstruction are $\alpha_\theta = 0.01$, $\beta_\theta = 1$, $\beta_\mu = 1$ and MAX_LOOPS=100. In calculating the risk of equation (12), we choose TV penalty and $\lambda = 2 \times 10^4$. For the comparator methods, MAP-TV is with a TV penalty of 3000. FullDict is with $\beta_\mu = 1$, $\beta_\theta = 1$ and α_θ is set to a smaller value, 10^{-5}, to shrink more on $\boldsymbol{\theta}$. For each algorithm, the spectrum is estimated by taking the strongest column in $\widehat{\mathbf{F}}$. Exemplar $\widehat{\mathbf{F}}$ from aluminum measurements are shown in figure 2. The aluminum is placed at spatial grid 267. In figure 2, all methods are able to locate the material. But notice that by using multi-scale dictionary, very clean spikes at column 267 are produced, and there is no spurious response at the lower right conner compared with MAP-TV and MLE.

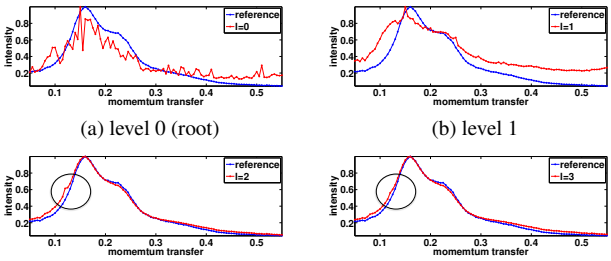
To illustrate the progress of traversing from coarse to fine, we demonstrate the reconstructed water spectrum at each level in figure 3. This traversal stops at level 3. We observe that as we delving deeper along a selected path, the reconstructed spectrum gradually approaches the reference till it is close enough and the path seeking stops.



(a) Multi-scale dictionary, upon stopping delving (b) fullDict

(c) MAP-TV (d) MLE

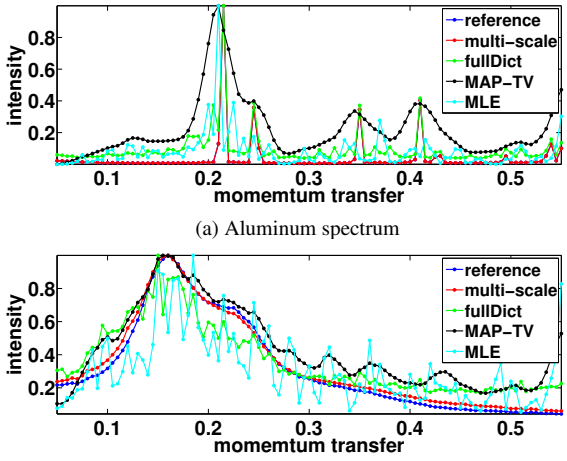
Fig. 2: Reconstructed $\widehat{\mathbf{F}}$ by four methods. The material is aluminum and placed at grid 267.



(a) level 0 (root) (b) level 1

(c) level 2 (d) level 3

Fig. 3: Reconstructed water spectrum at each level



(a) Aluminum spectrum

(b) Water spectrum

Fig. 4: Reconstructed spectra: compare all methods

Figure 4 compares the finally (upon stopping delving) reconstructed spectrum with the comparators. Note that in figure 4a, the reconstructed spectrum by using multi-scale dictionary perfectly overlaps with reference. And for both materials, multi-scale dictionary yields estimates much closer to the references. This is particularly helpful for successive tasks like classification and threatening material identification.

5. REFERENCES

[1] J. A. Greenberg, K. Krishnamurthy, and D. J. Brady, "Snapshot molecular imaging using coded energy- sensitive detection," *Optics Express*, vol. 21, no. 21, pp. 25480–25491, October 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November 2006.

[3] W. K. Allard, G. Chen, and Mauro Maggioni, "Multi-scale geometric methods for data sets ii: Geometric multi-resolution analysis," *Applied and Computational Harmonic Analysis*, vol. 32, no. 3, pp. 435–462, 2012.

[4] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Modelling and Simulation*, vol. 7, no. 1, pp. 214–241, April 2008.

[5] R. Jenatton, J. Mairal, F. R. Bach, and G. R. Obozinski, "Proximal methods for sparse hierarchical dictionary learning," in *the 27th International Conference on Machine Learning*, 2010, pp. 487–494.

[6] Y. Xie, J. Huang, and R. Willett, "Changepoint detection for high-dimensional time series with missing data," *IEEE Journal of Selected Topics on Signal Processing*, vol. 7, no. 1, pp. 12–27, 2013.

[7] L. Li, X. Zhang, M. Zhou, and L. Carin, "Nested dictionary learning for hierarchical organization of imagery and text," in *UAI*, 2012, pp. 469–478.

[8] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA*, vol. 62, no. 1, pp. 55–59, 1972.

[9] K. Krishnamurthy, M. Raginsky, and R. Willett, "Multiscale photon-limited spectral image reconstruction," *SIAM journal on Imaging Science*, vol. 3, no. 3, pp. 619–645, Spetember 2010.

[10] Z. T. Harmany, F. M. Roummel, and R. Willett, "This is spiral-tap: sparse poisson intensity reconstruction algorithms—theory and practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[11] J. Salmon, C. A. Deledalle, R. Willett, and Z. T. Harmany, "Poisson noise reduction with non-local pca," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1109–1112.

[12] J. A. Greenberg, K. Krishnamurthy, and D. J. Brady, "Compressive single-pixel snapshot x-ray diffraction imaging," *Optics Letters*, vol. 39, no. 1, pp. 111–114, 2014.

[13] P. Getreuer, "Rudin-osher-fatemi total variation denoising using split bregman," *Image Processing On Line*, vol. 10, 2012.

[14] K. MacCabe, K. Krishnamurthy, A. Chawla, D. Marks, E. Samei, and D. Brady, "Pencil beam coded aperture x-ray scatter imaging," *Optics Express*, vol. 20, no. 15, pp. 16310–16320, 2012.

[15] K. Arai and A. R. Barakbah, "Hierarchical k-means: an algorithm for centroids initialization for k-means," *Reports of the Faculty of Science and Engineering*, vol. 36, no. 1, pp. 25–31, 2007.

[16] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[17] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[18] M. Zhou, L. Hannah, D. Dunson, and L. Carin, "Beta negative binomial process and poisson factor analysis," *International Conference on Artificial Intelligence and Statistic*, 2012.

[19] M. J. Beal, *Variational algorithms for approximate Bayesian inference*, Ph.D. thesis, University College London, 2003.