

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Bayesian Radiation Image Reconstruction and Uncertainty Quantification

Permalink

<https://escholarship.org/uc/item/9hg9f1m9>

Author

Lee, Jaewon

Publication Date

2024

Peer reviewed|Thesis/dissertation

Bayesian Radiation Image Reconstruction and Uncertainty Quantification

By

Jaewon Lee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering- Nuclear Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kai Vetter, Chair

Professor Youngho Seo

Professor Lee A. Bernstein

Associate Professor Gabriel Orebi Gann

Fall 2024

Bayesian Radiation Image Reconstruction and Uncertainty Quantification

Copyright 2024
by
Jaewon Lee

Abstract

Bayesian Radiation Image Reconstruction and Uncertainty Quantification

by

Jaewon Lee

Doctor of Philosophy in Engineering- Nuclear Engineering

University of California, Berkeley

Professor Kai Vetter, Chair

This dissertation reports the development of radiation image reconstruction and uncertainty quantification algorithms incorporating Bayesian principles. The proposed algorithms effectively combat the ill-posed nature of radiation image reconstruction problems, while providing deeper insights into the physical quantities of interest through their uncertainty quantification capabilities.

The first algorithm, dubbed Kullback-Leibler (KL) divergence maximum *a posteriori* expectation maximization (KLD MAP-EM), incorporates a sparsity-inducing prior distribution based on KL divergence. As a result, the algorithm yields parsimonious solution images effectively minimizing image artifacts caused by measurement noise, while accurately localizing radiation source hotspots. The algorithm is computationally efficient and requires minimal modification of the conventional maximum likelihood-expectation maximization (ML-EM) algorithm, making it readily applicable to a variety of radiation imaging applications where sparsity in solution images is required.

The second algorithm, named the Gaussian process prior (GPP) image reconstruction algorithm, imposes spatial correlations between image voxels through an underlying Gaussian process (GP) variable with a choice of link function. The algorithm is highly robust to overfitting, and the GP covariance kernel can be modified to incorporate structural priors, thereby enhancing the quality of reconstructed images. Furthermore, a complete Bayesian framework for uncertainty quantification and hyperparameter selection methods was developed for the GPP algorithm, leveraging the Laplace approximation and preconditioned Crank-Nikolson Markov Chain Monte Carlo (pCN MCMC). As a result, the GPP algorithm not only produces high quality reconstructed images but also provides accompanying uncertainty images, offering a more comprehensive understanding of the imaging subject. Furthermore, the method provides uncertainties for derived quantities such as radiation dose and total reconstructed activity, enhancing decision-making reliability in high-stakes scenarios.

Lastly, the full-spectral GPP algorithm was developed, in which the following three different inverse problems are solved simultaneously: radiation spectra decomposition, deconvolution, and image reconstruction. The result is a novel algorithm decomposing a time-series of radiation energy spectra collected from a single, free-moving detector into their constituent contributions from radiological components, while reconstructing source distribution images corresponding to the identified components. The full-spectral GPP algorithm demonstrates exceptional performance, reconstructing source images and components even from low signal-to-noise ratio (SNR) measurements with fluctuating background contributions. Moreover, it enables deeper understanding of radiological environment by identifying salient radiological source and background features present in the imaging space.

The proposed algorithms have been rigorously tested using both simulated and real-world measurement data, that are challenging for conventional image reconstruction algorithms. Although primarily demonstrated for free-moving single detector imaging modality, these methods are broadly applicable across various imaging modalities, offering numerous avenues for future research.

To Mr. Jang, my former teacher, who turned my life around when I was 11.

Contents

Contents	ii
List of Figures	iv
1 Motivation and Introduction	1
1.1 Motivation for Radiation Image Reconstruction	1
1.2 Dissertation Overview	2
2 Radiation Imaging Modalities and Reconstruction Algorithms	3
2.1 Radiation Imaging Modalities	3
2.2 Radiation Imaging Data Formation and Reconstruction methods	10
3 Kullback-Leibler Divergence Regularization for Free-moving Imaging	28
3.1 Problem formulation	29
3.2 Kullback-Leibler divergence regularization	31
3.3 Results	34
3.4 Discussion	42
3.5 Conclusions	44
4 Radiation Image Reconstruction and Uncertainty Quantification Using a Gaussian Process Prior	45
4.1 Introduction	46
4.2 Methods	47
4.3 Results	59
4.4 Discussion	73
4.5 Supplementary Information	75
5 Simultaneous Spectral Decomposition and Full Spectral Imaging	80
5.1 Methods	81
5.2 Results	97
5.3 Discussion and Future Work	113
5.4 Conclusions	116

6	Closing Remarks	117
6.1	Summary	117
6.2	Future Research Avenues	118
	Bibliography	120

List of Figures

2.1	Schematic illustration of Compton imaging.	4
2.2	Relative probabilities of photoelectric absorption, Compton scattering and pair production as a function of photon energy and effective atomic number of matter. Data was obtained from NIST XCOM photon cross sections database [8].	4
2.3	Schematic illustration of coded aperture imaging. Depending on the source position, a distinct mask shadow is cast on the detector.	5
2.4	A schematic illustration of SDF applied to Compton imaging and single detector imaging. The locations and orientations of an imager are tracked, and Compton events recorded at different times are combined in a global map, localizing the source. In free-moving single detector imaging, the variation in count rate is used to deduce the source distribution and its strength.	7
2.5	(a) Polaris-LAMP, which combines a commercial gamma-ray imager with contextual sensors, enabling SDF. (b) A free-moving measurement of three 30-A cylinders for DUF_6 . Two cylinders near the trunk were empty, while the third one contains 1/2-2/3 capacity of DUF_6 , emitting 185.7 keV and 1001 keV gamma rays from ^{235}U and ^{238}U , respectively. Note that the measurement path and count rate are shown in the map. (c) Free-moving coded aperture imaging with 185.7 keV gamma-ray source. (d) Compton imaging mapping 1001 keV gamma-ray source. Image from [28] with permission.	8
2.6	(A): Shepp-Logan phantom (B): Noiseless sinogram of the Shepp-Logan phantom (C): Simple backprojection of the noiseless sinogram. Note that it introduces blurr, which is characterized by the PSF matrix $\mathbf{P}^\top \mathbf{P}$. (D): Filtered backprojection of the noiseless sinogram. In the absence of noise, the FBP produces perfect reconstruction.	16
2.7	(A): Shepp-Logan phantom (B): A measured sinogram corrupted with Gaussian noise. (C): Simple backprojection reconstruction result (D): Filtered backprojection reconstruction result (E)(F): Reconstruction results after 20 and 100 iterations of SART. With increasing number of iterations, the reconstructed image exhibits pronounced high spatial frequency features, including noise.	21

3.1	(Top) An example of single-detector imaging scenario with a 10 μCi point source placed at $(x, y, z)=(-0.5, 0.5, 0)$ m. (Bottom) The measured count data. The fluctuation in the measurement data primarily comes from the variation in source to detector distance. Also, the Poisson measurement noise is observed in the high-frequency noise feature in the measurement.	29
3.2	Reconstruction results using the ML-EM after (left) 20 and (right) 200 iterations. In both cases, the reconstruction fails to localize the point source. After 20 iterations of the ML-EM, the measurement data is clearly underfitted, whereas after 200 iterations clear signs of overfitting are observed.	31
3.3	(a) KLD MAP-EM reconstruction of the point source measurement scenario in figure 3.1. Unlike the ML-EM result shown in figure 3.2, the KLD MAP-EM successfully localizes the point source location. (b) Forward projection $\bar{y}$ from the reconstructed image. Note that the forward projection does not fit to the noise feature.	35
3.4	Measurement path and measurement/forward projection of (a) the ground truth (b) ML-EM reconstruction and (c) KLD MAP-EM reconstruction. Note that the ML-EM completely fails to map the distribution, exhibiting reconstruction biased towards the measurement path. Also, the forward projection of the ML-EM reconstruction indicates severe overfitting. The KLD MAP-EM reconstruction finds a good distribution estimate, along with the more regularized forward projection.	36
3.5	The measurement path and point cloud shown in (a) a top-down view, (b) a side view, and (c) a 3D perspective, illustrating the ^{137}Cs point source location within the house. (d) A time-series plot of counts within the energy region of interest (632.0-692.0 keV).	37
3.6	Images reconstructed using the (a) ML-EM and (b) KLD MAP-EM. 200 iterations were used for both reconstructions.	38
3.7	Setup for the 1 mCi ^{239}Pu -surrogate point source measurement. The source was placed on a 5-gallon bucket in an open field. A plastic scintillator integrated drone was flown to measure both the neutrons and gamma rays from the source.	40
3.8	Top: Image reconstruction results using the ML-EM (left) and the KLD MAP-EM algorithm (right). Bottom: Measured time-series radiation count data and the forward projection from the reconstructed images. Note that the SNR in the measurement is very low. Also, it is apparent that the measurement is corrupted with non-Poisson noise.	41
3.9	Left: Image reconstruction results the KLD MAP-EM when two point sources are present in the imaging space. Right: Image reconstruction results the KLD MAP-EM when three point sources are present in the imaging space. As we add more point sources, the measurements cannot be accounted for by a single hotspot, leading to erroneous reconstruction results.	43

- 4.1 Random samples from the prior distribution $p(\mathbf{x})$ on a 20×20 m 2-D imaging space with different sets of hyperparameters, σ and λ . Each sample was realized by first drawing a sample from $\boldsymbol{\xi}_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then transforming it by $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{L}\boldsymbol{\xi}_w)$. The properties and characteristics of the prior distribution can be better understood by examining random samples from the prior. Note that the σ parameter defines the characteristic correlation length, while the reciprocal of λ parameter determines the mean pixel values. 50
- 4.2 (a) A simulated measurement scenario with a 1-mCi anisotropic Gaussian distributed source. The human-walked path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters $\sigma=4.1$ m and $\lambda=1.4 \times 10^{-4}$ Bq $^{-1}$ found from the empirical Bayes method were used for reconstruction. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y=0$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals. 60
- 4.3 Contour plots of the negative log-marginal likelihood as a function of the hyperparameters σ and λ . The hyperparameters selected from Algorithm 2 are marked with red Xs. (a) The anisotropic Gaussian source measurement scenario. (b) The ring and Gaussian source measurement scenario. (c) The uniform square source measurement scenario with a structural prior. When using the structural prior, 2 different sets of hyperparameters were used for in-square and out-of-square. Therefore, only the in-square hyperparameters were used to create the contour plot, while the out-of-square hyperparameters were fixed at the values found from Algorithm 2. 61
- 4.4 To better understand the impact of the hyperparameters, the anisotropic Gaussian source images reconstructed from the GPP MAP with different choices of the hyperparameters σ and λ are displayed. With larger σ and λ , the prior distribution is spatially more correlated and the pixel values are pushed towards zeros. As a result, the reconstructed image with large hyperparameters (bottom right) tend to spread out. 62

4.5	(a) A simulated measurement scenario with a 2-mCi ring-shaped source and a 1-mCi anisotropic Gaussian distributed source. The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters $\sigma=2.4$ m and $\lambda=6.0\times 10^{-5}$ Bq $^{-1}$ found from the empirical Bayes method were used for reconstruction. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y=0$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.	63
4.6	(a) A simulated measurement scenario with a 1-mCi uniform located at $(x, y) = (2.5\text{m}, 2.5\text{m})$. The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the GPP MAP without incorporating the structural information. The hyperparameters found from Algorithm 2, $\sigma=2.03$ m and $\lambda=8.10$ Bq $^{-1}$ were used. (c) The reconstructed from the proposed GPP MAP with the structural prior. $\sigma = 1.0 \times 10^4$ m and $\lambda = 1.1 \times 10^{-5}$ Bq $^{-1}$ for the in-square hyperparameters, and $\sigma = 9.8 \times 10^3$ m and $\lambda = 7.7 \times 10^4$ Bq $^{-1}$ for the out-of-square hyperparameters. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y=2.5$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.	65
4.7	(a) Measurement setup comprising a 181 μCi ^{137}Cs point source and a CZT-based Compton camera, H420-LAMP. (b) Point cloud map of the laboratory where the measurement was conducted. (c) A 4π depth map showing the distance between the Compton camera and the surrounding objects.	67
4.8	Directional radiation flux maps reconstructed from (a) the ML-EM and (c) the GPP MAP with a structural prior. The flux maps are weighted using the depth map from Fig. 4.7, yielding the activity maps from (b) the ML-EM and (d) the GPP MAP. Note that the ML-EM flux map spreads across both the box and the background, whereas the GPP MAP with a structural prior effectively constrains the reconstructed flux within the box. Consequently, the ML-EM fails to accurately estimate the total activity within the box, while the GPP achieves a significantly more accurate estimate of 179 μCi	69
4.9	Histogram of the estimated total activity samples.	70
4.10	(a) Measurement scenario with the ground truth distribution. (b) GPP MAP reconstruction result. (c) Estimated dose rate map and example paths. (d) Dose rate uncertainty map.	71

4.11	Summary of total dose received along Paths 1 2 and 3 and their associated uncertainties. Samples from an approximate posterior distribution were used to derive the uncertainties.	72
5.1	(a)(b): Ground truth source activity distributions for (a) ^{134}Cs , ^{137}Cs and (b) ^{131}I . (c)(d)(e): Spatially varying count rate contribution from KUT progenies. (f) Ground truth source emission energy spectra. Note that the emission spectra include downscattered and backscattered photons from the ground, which manifest as a low-energy continuum. (g): Ground truth detected background spectra from KUT progenies. These spectra represent what would be observed by the detector. Therefore, they are influenced by detector-specific spectral features, such as the NaI(Tl) detector resolution and pronounced Compton continua.	99
5.2	2D time-energy count histogram from the Monte Carlo simulated measurement with a NaI(Tl) detector. The time-dimension bin width is 0.25 s, while in the energy-dimension 128 bins were spaced proportional to $\sqrt{E}$. The independent contributions from the five source and background components yielded complex and dynamic measurement data.	100
5.3	(a)(b): Ground truth emission rate distributions for (a) ^{134}Cs , ^{137}Cs and (b) ^{131}I . (c)(d): MLE image reconstruction results. Note that the reconstruction is corrupted with noise, which is a typical sign of overfitting. (e)(f): Full spectral GPP image reconstruction results. Component 1 and 2 appear to correspond to the Cs and I isotope distributions, respectively. (g)(h): 90% credible interval maps from the Laplace approximation.	101
5.4	(a)(b): Reconstructed source spectra from the MLE and full spectral GPP reconstruction. (c)(d): Reconstructed background detection spectra. (e)(f): Time-series of gross counts measurements and forward projections from the reconstructed source and background components. In (b)(c)(d), 90% credible intervals estimated from the Laplace approximation are indicated with shaded regions.	102
5.5	(Bottom left): A point cloud map of SRNL H-area obtained from a measurement with a LiDAR on MiniPRISM imaging system. (Top, Bottom right): A map of SRNL H-area.	103
5.6	MiniPRISM imaging system is equipped with 58 CZT detectors and contextual sensors, including LiDAR, visual camera and IMU. The imager can operate as either a coded aperture imager or a Compton camera. However, in this work, the measurements from the 58 CZT detectors were aggregated, yielding measurement data akin to that from a single detector.	104

5.7	(Left): The measurement path and gross counts measured from MiniPRISM. Notably, the gross counts level is elevated near the fence on the left, and decreases near the forest on the right side. (Right): The 2D time-energy histogram from the MiniPRISM measurement. For visibility, bins with 0 count were not colored. The background variability is observed from the horizontal streaking features. Additionally, a trace of ^{137}Cs photopeak counts at 662 keV is visible.	105
5.8	(a): The ^{137}Cs distribution reconstructed from the proposed approach. (b): The 90% credible interval map from the Laplace approximation(c) The reconstructed ^{137}Cs spectral component. Interestingly, the reconstructed spectrum clearly shows the Compton edge and Compton continuum at <477 keV as well as the 662 keV photopeak feature.	106
5.9	(Top): The reconstructed background spectral components. The background component 1 appears to correspond to downscattered photons, while the component 2 contains various photopeak features, such as 238 keV, 609 keV, and 1460 keV photopeaks from the KUT progenies. (Bottom) The forward projections from the reconstructed source and background components. In both figures, the shaded regions represent the 90% credible intervals. Note that the source forward projection is very low compared to that of background components, indicating the very low SNR in the data.	107
5.10	Negative log-marginal likeilhood (NLML) value as a function of hyperparameter optimization iteration numbers. The 1 source 1 background model does not possess enough explanatory power, leading to a large negative log-marginal likeilhood value after hyperparameter optimization. On the other hand, the 1 source 3 background model does not yield lower NLML than the 1 source 2 background model after hyperparameter optimization, suggesting 1 source 3 background model is superfluous for the measurement data.	108
5.11	(a) A simulated HEU point source measurement scenario. (b)(c) Simulated ground truth background and 1-cm lead shielded HEU source spectra. (d) Gross count forward projection from the source and the background.	109
5.12	(left) Measured time-energy 2D histogram data. (right) For visibility, the histogram colorbar scale was modified, reavealing the dynamically changing background spectral features.	110
5.13	Results from the gross count PSL algorithm. Note that the PSL appraoch cannot effectively deal with the dynamically chaning background countrate.	111
5.14	Left: Full spectral PSL MLE results, without any priors imposed. Right: Full spectral GPP-PSL results.(a)(b): Source location and strength estimation. (c)(d) Reconstructed source spectrum. (e)(f) Reconstructed background spectrum. (g)(h) Forward projection gross counts from the reconstructions. Note that by incorporating the GP prior, the source spectrum estimation is significantly improved. In addition, the forward projections in (g) and (h) indiciate that the GP prior effectively combat the overfitting of the MLE method.	112

Acknowledgments

Leaving Korea to begin graduate school, I was excited yet slightly anxious about the new life awaiting me in Berkeley as a PhD student. I remember feeling particularly nervous about meeting my advisor, Kai, as I knew almost nothing about him prior to coming here. However, I now realize that I couldn't have asked for a better advisor for my PhD journey. Kai, under your guidance, I have grown academically, professionally, and personally over the past five years. I truly appreciate your unwavering support and encouragement—none of the work in this dissertation would have been possible without you.

I was also fortunate enough to meet incredible mentors throughout my PhD: Ren, Tenzing, Don, and Mark. I sincerely thank you for taking the time to meet with me every week and for guiding me whenever I felt lost. I learned so much from each of you, and I truly enjoyed all of the discussions we had throughout my time here.

While I received incredible academic support here, I also made an effort to give back by mentoring other students. I thank Nolan, Mehul, Jimmy, Jisu, and Oscar for trusting me as their mentor and following through with my guidance. I only wish I had more time to support each of you fully.

I am immensely thankful for the friendships I forged in Berkeley. In particular, the members of the 1030 Cragmont house—Yves, Lorenzo, Andrew, Carla, Max, Chai, and Yulim—deserve special appreciation for all the laughter and drinks we shared. I never thought I would find a missing brother here in Berkeley, but then I met Matthew. Can't wait to see you graduate brother. I also thank Kalie for being an amazing lab mate who shared the PhD journey with me. Thank you, Tatiana, not only for all the fun we had but also for providing emotional support when I needed it most. I greatly enjoyed going to the gym with you, Robin, though we probably should have kept a more consistent workout routine. Lastly, the members of KAKOA—Jisu and Seungeun—helped me reconnect with my Korean roots. Without you two, I might have lost my Koreanness entirely.

Finally, I wish to express my gratitude to my mom, my dad, and my brother. Without their support I would not have come anywhere close to where I am today.

Chapter 1

Motivation and Introduction

1.1 Motivation for Radiation Image Reconstruction

Over the past century, radiation imaging has established itself as a central component in various applications, including medical diagnostic imaging [1], astrophysics [2], non-destructive testing (NDT) [3], and nuclear safety and security [4]. For example, medical diagnostic imaging techniques such as computed tomography (CT), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) have become vital tools for understanding the anatomical and functional features of human diseases. In NDT, radiography techniques using X-rays or neutrons are routinely performed to ensure the structural integrity of manufactured components. Furthermore, the mapping of dispersed radioactivity over large areas can effectively aid in emergency response and contamination remediation efforts, whose importance was highlighted again after the Fukushima Daiichi nuclear power plant accident in 2011 [5]. Lastly, there have been growing interests in incorporating radiation imaging into nuclear safeguard tasks, as it provides better spatial awareness of radioactive materials while significantly reducing the dose to inspectors[6].

In most applications, the goal of radiation imaging is to produce images that reveal spatial distribution of radioactivity or properties of intervening materials. Such spatial information is often not straightforward to obtain; instead, it requires various forms of physical modulation of radiations to encode the desired spatial information. For example, in medical CT, X-rays are projected at multiple angles and attenuated inside the patient's body, encoding anatomical features based on the varying densities of tissues they traverse. In SPECT imaging, radiation emitted from radiotracers is physically blocked by collimators, allowing only those traveling in specific directions to penetrate, thereby encoding the distribution of the radiotracers. In both examples, the modulated radiation is then detected by various radiation detectors, converting it into readable signals.

However, the measured signals are often not directly interpretable by humans, necessitating an additional post-processing step: image reconstruction. This process converts raw signals from detectors into meaningful visual images. Image reconstruction is typically

mathematically formulated as an inverse problem. It begins with a mathematical model, called the forward operation (or forward model), which represents physical and stochastic processes involved in signal formation. The original image that generated the data is then determined by inverting the forward operation.

Unfortunately, image reconstruction is a challenging task for three main reasons.

1. Radiation generation, modulation, and detection are inherently stochastic processes. Consequently, the measured signals inevitably contain noise, which degrades the quality of the reconstructed images.
2. The signal formation process cannot be fully understood or accurately modeled, leading to inaccuracies in the forward model.
3. Lastly, certain imaging scenarios do not yield sufficient information in the measured signals to produce meaningful images.

These limitations together result in undesirable reconstructed images corrupted by statistical and physical perturbations. Therefore, significant efforts have been dedicated to the development of advanced image reconstruction techniques to address these challenges.

This work aims to contribute to these efforts by presenting the development of novel Bayesian radiation image reconstruction algorithms. Leveraging Bayesian framework, these algorithms demonstrate robust performance even in information-limited radiation imaging scenarios. As a result, they enable radiation imaging to be effectively applicable across a variety of challenging and versatile context. Throughout this dissertation, the development, applications and general implications of these algorithms will be discussed in depth.

1.2 Dissertation Overview

The primary objective of this work is to present the development of Bayesian image reconstruction algorithms. Readers are assumed to have a foundational understanding of common radiation imaging modalities and the basics of image reconstruction algorithms. However, to underscore the applicability of the proposed methods and their improvements over conventional approaches, Chapter 2 reviews several radiation imaging modalities and the fundamentals of radiation image reconstruction.

Chapter 3 details the development of the Kullback-Leibler maximum *a posteriori* (KLD MAP-EM) algorithm. Chapter 4 discusses the Gaussian process prior (GPP) image reconstruction and uncertainty quantification algorithm in detail. In Chapter 5, the full-spectral GPP algorithm—an extension of the GPP algorithm leveraging spectral information in measurement data—is presented in detail.

Finally, Chapter 6 summarizes the proposed algorithms, highlights their broad implications, and outlines motivations for future research.

Chapter 2

Radiation Imaging Modalities and Reconstruction Algorithms

2.1 Radiation Imaging Modalities

In this section, several widely used radiation imaging modalities for astrophysics, nuclear medicine, nuclear security and safety applications are briefly introduced. In particular, the principles of Compton imaging, Coded aperture imaging, and recently developed free-moving 3D radiation imaging techniques are discussed.

Compton Imaging

In Compton imaging, the kinematics of Compton scattering is used to deduce and image the origin of photons [7]. Consider a photon with a known initial energy E_0 undergoing Compton scattering at position $\mathbf{r}_1$ depositing some energy, followed by a photoelectric absorption at $\mathbf{r}_2$. Denoting the energy deposited in the interactions E_1 and E_2 , we have $E_0 = E_1 + E_2$. Note that the interaction positions and deposited energies can be measured from a position sensitive radiation detection system. Then, the Compton kinematics formula allows us to deduce the angle θ at which the photon scattered as

$$\theta = \cos^{-1} \left(1 - m_e c^2 \left(\frac{1}{E_0 - E_1} - \frac{1}{E_2} \right) \right), \quad (2.1)$$

where m_e is the rest mass of an electron and c is the speed of light. The pair of Compton scattering and photoelectric absorption is referred to as a *Compton event*. Hence, for a given Compton event, the origin of the photon is constrained to lie on a conic surface, whose opening angle is θ and the rotation axis is around $\mathbf{r}_1 - \mathbf{r}_2$. See Fig. 2.1 for illustration.

As the first interaction in a Compton event must be Compton scattering, the sensitivity of a Compton imaging device depends on both the incident gamma-ray energy and the detector material. Figure 2.2 presents a ternary map illustrating the relative probabilities of different

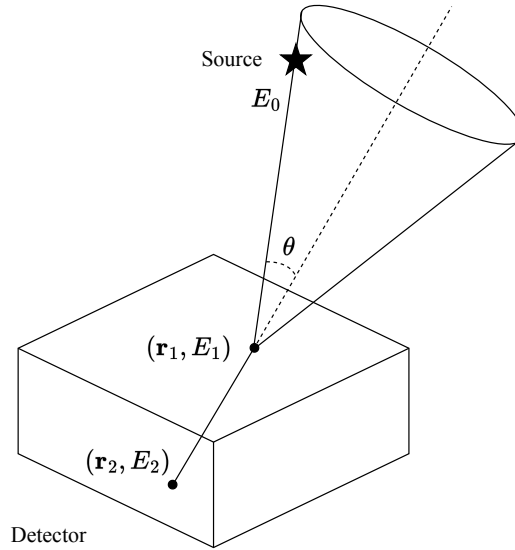


Figure 2.1: Schematic illustration of Compton imaging.

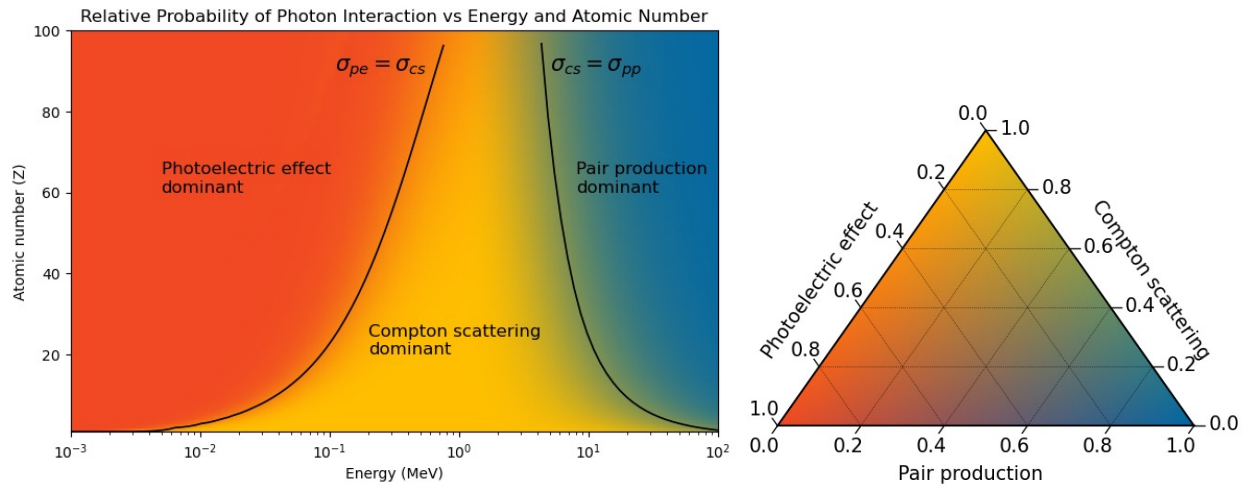


Figure 2.2: Relative probabilities of photoelectric absorption, Compton scattering and pair production as a function of photon energy and effective atomic number of matter. Data was obtained from NIST XCOM photon cross sections database [8].

photon interaction modes as a function of the incoming photon energy and the material atomic number. Note that the incoming photon energy is application specific, varying widely from few hundreds keV to 10 MeV. For example, in nuclear molecular imaging, gamma rays emitted from radiopharmaceuticals are typically less than 1 MeV [9]. On the other hand, Gamma-ray Compton telescopes are commonly designed for photon energies up to 10 MeV [10]. Hence, the detector material must be carefully chosen to maximize the sensitivity of the Compton imaging system for a given application.

Since Compton imaging does not rely on mechanical collimators, it is particularly well-suited when a wide field of view (FOV) or high sensitivity is required. Due to these advantages, Compton imaging has been applied in astrophysics [10][11][12], nuclear medicine [13], and nuclear threat detection [14].

Coded Aperture Imaging

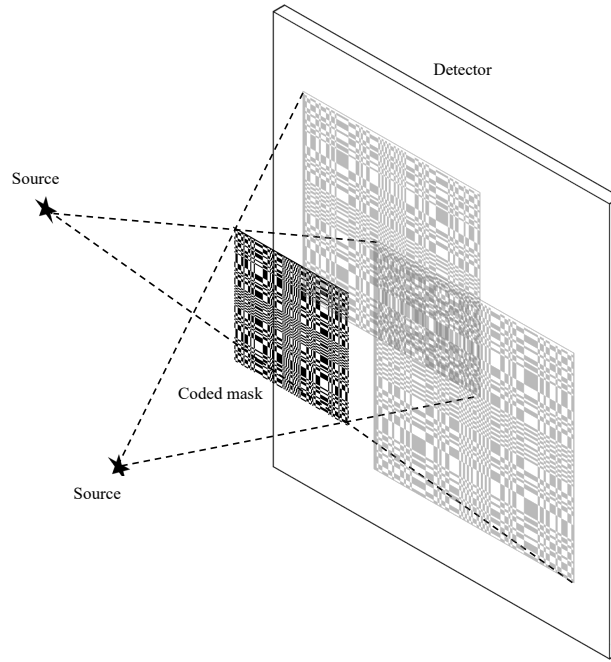


Figure 2.3: Schematic illustration of coded aperture imaging. Depending on the source position, a distinct mask shadow is cast on the detector.

In coded aperture imaging, a photon-opaque mask with a known pattern, called a coded aperture mask, is placed in front of a position sensitive detector. Therefore, radiations emitted from radioisotopes are blocked by the mask, casting a unique patterned shadow on the detector, depending on the location of the radiation source. Subsequently, this shadow pattern is then used to reconstruct image of radioactive source distribution [15].

The pinhole camera is a special case of coded aperture imaging, where a single aperture is used to modulate incident photons. In pinhole camera imaging, the radiation count distribution on the detector directly reveals the radioactivity distribution in the imaging space, requiring minimal image reconstruction effort. However, because most emitted photons are blocked by the mask, it suffers from a low signal-to-noise ratio (SNR), requiring long exposure time for image formation. One can simply enlarge the pinhole to allow more photons to pass through, thereby increasing the SNR. However, the increased SNR comes at the cost of degraded resolution, as the pinhole camera resolution is inversely proportional to the diameter of pinhole [16].

The key idea behind coded aperture imaging is to place multiple small apertures across the mask, increasing the open fraction of the mask up to a range of 25-50% [17]. Therefore, the SNR is significantly improved compared to that of a pinhole camera, while minimizing resolution degradation by maintaining a small aperture size. Figure 2.3 illustrates the principle of coded aperture imaging. Note that unlike the pinhole camera, data obtained from a coded aperture imaging requires a non-trivial image reconstruction process to “decode” the image from the measurement, due to overlapping “pinhole” projections [18].

Coded aperture masks are typically made of high-density metals, such as tungsten, to effectively block incoming photons via photoelectric absorption. From Fig. 2.2, it can be noted that for tungsten, photoelectric absorption is the primary mode of photon interaction below 300 keV, making coded aperture effective for imaging low-energy gamma rays and X-ray emitters. For this reason, coded aperture imaging is sometimes complemented with Compton imaging, enabling radiation imaging across wide range of photon energies [19][20]. Coded aperture imaging has found its applications in X-ray telescopes [21], nuclear medicine [22], non-destructive testing [23] and nuclear power plant decommissioning[24].

Scene Data Fusion and Single Detector Imaging

The Compton and coded aperture imaging introduced are stationary radiation imaging modalities, meaning that the entire measurements are taken with the imaging system held fixed in place. If these imaging systems are to be moved, it must follow well-planned paths around a constrained FOV. For applications such as medical imaging, the use of such stationary imaging systems is practical, as a fixed region-of-interest (ROI) is typically assumed. However, when the potential source location is not known *a priori* within a large imaging space, such static imaging systems pose the following two main challenges. First, because the radiation intensity follows the inverse-square law, imaging systems are only sensitive to radiation emitters in their vicinity, necessitating multiple time-consuming measurements at different locations. Secondly, imaging space is only locally defined for a stationary measurement; therefore, “stitching together” different imaging spaces from multiple stationary measurements becomes a non-trivial task.

Scene Data Fusion (SDF), a recently developed concept that combines advances in computer vision and radiation imaging technologies, overcomes these challenges and provides a novel capability to map and localize radiological sources in 3D on free-moving platforms

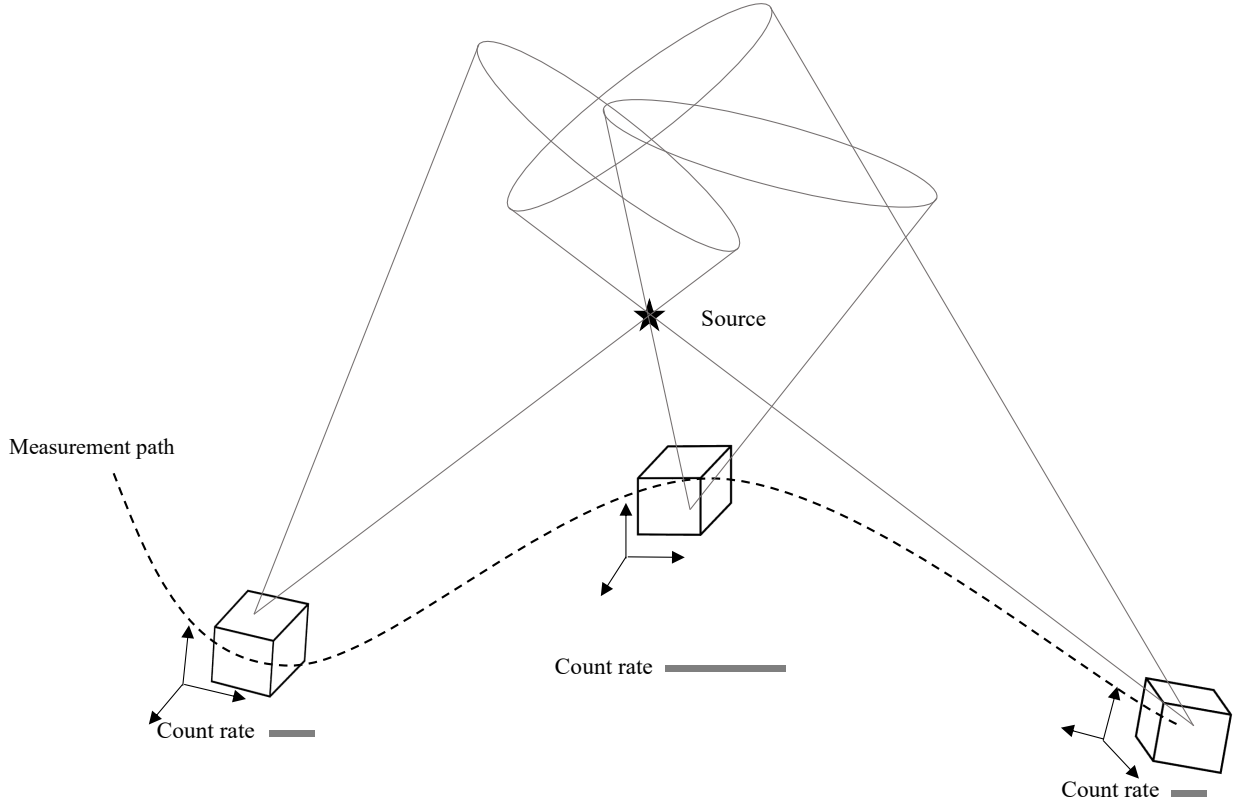


Figure 2.4: A schematic illustration of SDF applied to Compton imaging and single detector imaging. The locations and orientations of an imager are tracked, and Compton events recorded at different times are combined in a global map, localizing the source. In free-moving single detector imaging, the variation in count rate is used to deduce the source distribution and its strength.

[25][4][26][27]. In SDF, a conventional stationary imaging system (e.g., a Compton or coded aperture imager) is equipped with a position tracking system such as Simultaneous Localization and Mapping (SLAM). This allows the system to track both locations and orientations of the imager, while it is being freely moved around in the imaging space. As a result, data collected at each timestep is associated with a specific detector position and orientation. The time-dependant measurement data is then reconstructed in a global imaging space, enabling 3D, free-moving radiation imaging. Figure 2.4 illustrates the principle of SDF applied to Compton imaging.

An example of SDF with a real-world measurement is illustrated in Fig. 2.5. The measurement was conducted with Polaris-LAMP, a free-moving gamma-ray imager developed at Lawrence Berkeley National Laboratory (LBNL) [28]. Polaris-LAMP combines Polaris-H, a commercial off-the-shelf stationary imager with Compton and coded aperture imaging

capability, and a suit of contextual sensors enabling SLAM, such as a LiDAR, visual camera and inertial measurement unit (IMU). As illustrated in the figure, Polaris-LAMP effectively localizes and maps the depleted uranium hexafluoride (DUF_6) radiation source in container cylinders using measurement data collected while being freely moved around in the imaging space.

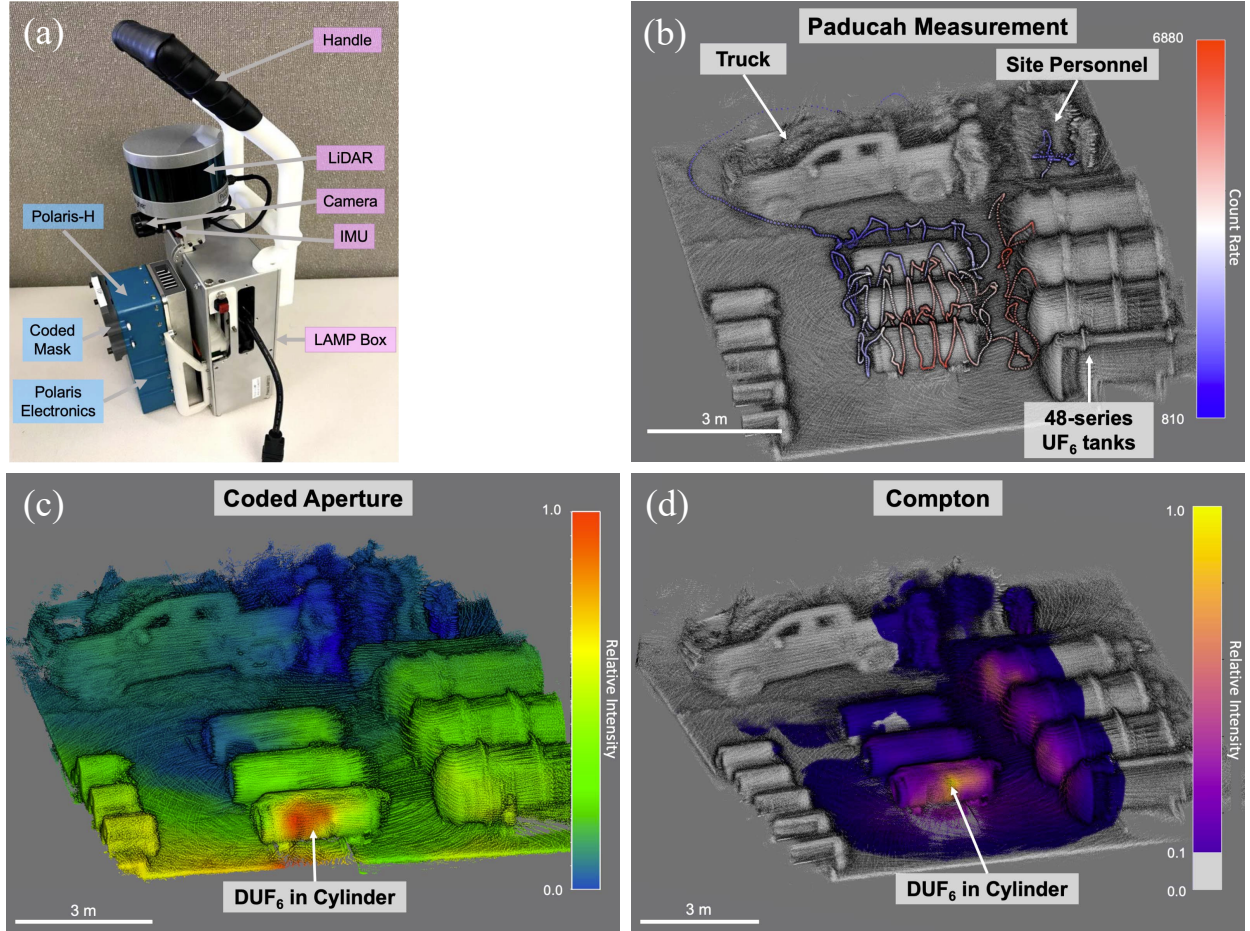


Figure 2.5: (a) Polaris-LAMP, which combines a commercial gamma-ray imager with contextual sensors, enabling SDF. (b) A free-moving measurement of three 30-A cylinders for DUF_6 . Two cylinders near the trunk were empty, while the third one contains 1/2-2/3 capacity of DUF_6 , emitting 185.7 keV and 1001 keV gamma rays from ^{235}U and ^{238}U , respectively. Note that the measurement path and count rate are shown in the map. (c) Free-moving coded aperture imaging with 185.7 keV gamma-ray source. (d) Compton imaging mapping 1001 keV gamma-ray source. Image from [28] with permission.

Due to its free-moving nature, SDF is particularly effective when the potential source location is unconstrained in a large environment. Earlier works have demonstrated its utilities

in a variety of potential applications, including nuclear threat detection, emergency response contamination remediation, and nuclear safeguard [25][4][28][29].

Moreover, SDF enables the use of a series of count rate data measured from a single detector to map radiological source distributions[29][30][31]. In single detector mapping, variations in count rate is recorded as the detector is moved around in an imaging space. According to the inverse square-law, the radiation count rate increases as the detector approaches the source and decreases as it moves farther away. Therefore, the fluctuation in count rate data encodes information about the location and distribution of the source. See Fig. 2.4 for illustration of single detector imaging.

The only requirement for single detector imaging is a non-position sensitive radiation detector equipped with a position tracking system. Consequently, compared to conventional Compton and coded aperture imagers, it can be constructed with simple electronics and detectors, making it a robust and cost-effective alternative. Moreover, single detector imaging offers very high sensitivity with 4π FOV, owing to the absence of collimation. Due to these advantages, it has been studied for potential applications in environmental radiation mapping [32], lost source search [33], and accident remediation [28][34].

Despite the promise of single detector imaging, it is often challenged by the lack of reliable image reconstruction methods [30]. Note that when using a single detector, the directionality of radiation is not encoded due to the lack of collimation. As a result, conventional image reconstruction algorithms often fail to produce desirable images. Throughout this dissertation, limitations of conventional image reconstruction methods in the context of single detector imaging will be demonstrated in detail.

2.2 Radiation Imaging Data Formation and Reconstruction methods

In this section, some of the foundational principles of the measurement data formation and reconstruction methods are briefly explained.

Radiation Image Formation

Suppose we represent a true 2D image as a function, $f(x, y)$ with its input arguments being a point in Cartesian coordinate. The true image $f(x, y)$ is then observed with an imaging system, yielding an observed image $\hat{f}(x, y)$. A simple image formation model assumes linear and shift-invariant (LSI) relationship between $f(x, y)$ and $\hat{f}(x, y)$. For an LSI imaging system, formation of the observed image $\hat{f}(x, y)$ can be expressed in terms of convolution

$$\hat{f}(x, y) = a(x, y) * f(x, y) + n(x, y), \quad (2.2)$$

where $a(x, y)$ is called the point spread function (PSF) or impulse response function, and $n(x, y)$ denotes a random noise process. Note that the PSF $a(x, y)$ is the observed image when $f(x, y) = \delta(x, y)$, where $\delta(x, y)$ is a 2D direct delta function. Typically, $a(x, y)$ is a blurred image of the delta function, and the convolution $a(x, y) * f(x, y)$ introduces the same blurring effect across the entire image domain. Then, the blurred image is subject to random noise $n(x, y)$, yielding a noisy measured image $\hat{f}(x, y)$.

The LSI model provides useful insights in image formation as the convolution model can be easily analyzed in the Fourier domain. Through Fourier transformation of (2.2), the PSF function $a(x, y)$ is interpreted as a frequency filter suppressing high frequency components of the true image $f(x, y)$. However, in many imaging modalities, the imaging system is not strictly LSI; although most imaging systems operate in a relatively linear response region, imaging systems are rarely exactly shift-invariant. In other words, the impulse response (i.e., PSF) may vary depending on the location of the impulse in the imaging domain. Moreover, in most radiation imaging modalities, the measurement is not a continuous image function $\hat{f}(x, y)$. Instead, it is often digitized signals read from a radiation detection system.

For these reasons, it is more realistic to consider the following measurement formation model for radiation imaging:

$$\mathbf{y}[m] = \iint a_m(x, y) \mathbf{f}(x, y) dx dy + \mathbf{n}[m], \quad (2.3)$$

where $\mathbf{y}[m]$ is the m -th element of a discretized observed measurement vector $\mathbf{y} \in \mathbb{R}^M$, $a_m(x, y)$ is the measurement contribution of $f(x, y)$ to $\mathbf{y}[m]$, and $\mathbf{n}[m]$ is the m -th element of a noise vector $\mathbf{n} \in \mathbb{R}^M$. Note that the measurement vector $\mathbf{y}$ does not necessarily represent an observed image. For example, it could be a vector of measured radiation counts from an imaging system.

In radiation image reconstruction, the goal is to estimate the unknown true image $f(x, y)$, given the measurement vector $\mathbf{y}$. Hence, image reconstruction is a prime example of inverse problems whose forward model is given in equation 2.3.

For the purpose of image reconstruction, the image to be estimated is also discretized in a vector form. Therefore, the model in 2.3 can be simply written with a matrix-vector multiplication as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2.4)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is commonly referred to as a system matrix, and $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$ is the unknown image to be estimated.

Note that with the noise vector $\mathbf{n}$, we are assuming an additive noise model, such as Gaussian noise. The additive noise model is valid in some radiation imaging modalities such as CT. However, in most radiation imaging modalities, Poisson model more aptly describes the measurement noise characteristics due to the quantized nature of radiation emission and detection. With Poisson noise in the measurement, the model in equation 2.4 becomes¹

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}) \quad \text{where} \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}. \quad (2.5)$$

In cases where the system matrix $\mathbf{A}$ is invertible and no noise is present, direct inversion of $\mathbf{A}^{-1}\mathbf{y}$ guarantees a good reconstruction of $\mathbf{x}$. However, such an ideal scenario is never encountered in reality, and direct inversion yields an undesirable reconstructed image with amplified noise. Moreover, the system matrix $\mathbf{A}$ is rarely invertible in practice, due to its non-square matrix shape and rank deficiency.

Due to the aforementioned practical difficulties, the image reconstruction problem is considered to be a classic example of an ill-posed inverse problem. Hadamard first listed the properties of a well-posed problem as follows [35]:

1. The problem has a solution.
2. The solution is unique.
3. The solution's behavior changes continuously with the initial conditions.

Problems that do not satisfy any of these properties are considered ill-posed. In the case of image reconstruction, the first and third conditions are usually satisfied. However, there may be many—if not infinite—solutions that could produce the measurement $\mathbf{y}$, and hence the solution is rarely unique. As a result, most image reconstruction techniques incorporate prior knowledge to specify a unique solution.

¹For Poisson measurement noise model, elements of the measurement vector $\mathbf{y} \in \mathbb{Z}_{\geq 0}^M$ are non-negative integers.

Measurement Data Types (Binned and List-mode Data)

Due to the quantized nature of radiation, raw output from radiation detectors is recorded on an event by event basis, and represented as a collection of J vectors $\{\mathbf{a}_j\}_{j=1}^J$, where each $\mathbf{a}_j \in \mathbb{R}^K$ is referred to as an attribute vector. The dimensions of attribute vectors K is determined by the number of relevant attributes for a given imaging modality. In radiation imaging, some common attributes include measured radiation energies, interaction positions and measurement timestamps. These attribute vectors can be regarded as a collection of Cartesian coordinates in an K -dimensional hyperspace.

The attribute vectors are randomly and independently generated according to a Poisson point process. A Poisson point process is a stochastic process where the number of occurrence in a bounded region in an underlying space is Poisson distributed. In other words, suppose we are given an intensity function $\lambda(\mathbf{a}) : \mathbb{R}^K \rightarrow \mathbb{R}_{\geq 0}$ and any bounded region $\mathcal{B}$ in $\mathbb{R}^K$. Then, under the Poisson point process, the number of events in the bounded region $N(\mathcal{B})$ is Poisson distributed as follows:

$$N(\mathcal{B}) \sim \text{Poisson}(\lambda_{\mathcal{B}}), \quad \text{where} \quad \lambda_{\mathcal{B}} = \int_{\mathcal{B}} \lambda(\mathbf{a}) d\mathbf{a} \quad (2.6)$$

The collection of vectors $\{\mathbf{a}_i\}_{i=1}^J$ produced from an underlying Poisson process is referred to as list-mode data [36]. Another way of representing the same data is to use a frequency distribution in $\mathbb{R}^K$. Suppose we have M number of disjoint bounded regions $\{\mathcal{B}_m\}_{m=1}^M$ that are called *bins*. Then, by counting the number of attribute vectors within each bin, we have a length- M vector $\mathbf{b} \in \mathbb{Z}_{\geq 0}^M$, whose entries are frequencies of attribute vector occurrence within the bins. The frequency vector $\mathbf{b} \in \mathbb{R}^M$, together with the bins $\{\mathcal{B}_m\}_{m=1}^M$, are commonly referred to as binned data.

When binning the list-mode data, some loss of information is inevitable as attribute vectors close to each other are lump together in bins. The amount of information loss depends on the size of the bins used, $\{\mathcal{B}_m\}_{m=1}^M$; the smaller the bin size, the more information is retained. Therefore, in practice, the size of the bins (or equivalently, the number of bins M) should be chosen such that the amount of information loss is tolerable with negligible degradation in image reconstruction process. For more detailed discussion of information loss in binned data, see [37]

Whether to use binned or list-mode data in reconstruction tasks mostly depends on imaging modalities. For example, in Compton imaging, relevant attributes from the raw data include the positions in Cartesian coordinate and energies deposited in the first (Compton scattering) and second (photoelectric absorption) interactions, resulting in attribute vectors with 8 elements, $\mathbf{a} = (x_1, y_1, z_1, E_1, x_2, y_2, z_2, E_2)$. Even if we discretize each attribute dimension very coarsely with 10^2 bins, the resulting binned data will be of size 10^{16} . Naively storing such binned data requires about 10 petabytes of memory. One can compress the large binned data as Compton events are not common and most of the bins will contain no events. However, resolution of the Compton imaging is highly sensitive to precision of the attributes, and therefore coarsely binned data severely impacts the image quality. For these reasons, Compton imaging is almost always performed with list-mode data [38].

Conversely, in coded aperture imaging, a large number of attribute vectors are recorded due to its high sensitivity, and each attribute vector contains a 2-D Cartesian coordinate (x, y) of the interaction position on a detector. The large number of events, combined with the low-dimensional attribute space facilitates the use of binned data for compact representation.

Analytical Image Reconstruction Methods

Here, we review several classical analytical image reconstruction methods and examine their properties. Although the focus of this thesis is on iterative (statistical) reconstruction methods, these analytical methods provide insights into several important general characteristics of image reconstruction.

Analytical image reconstruction techniques were first developed and extensively studied in the context of medical CT image reconstruction [39]. Here, we introduce several backprojection-based CT image reconstruction methods, and discuss their connection to reconstruction methods in other imaging modalities and iterative reconstruction methods. In 2D CT imaging, a measurement generates a collection of 1D projections of a ground truth image $f(x, y)$ taken at multiple angles. This imaging procedure produces a 2D function $s(l, \theta)$ called a *sinogram*, where l and θ denote the 1D detector position and projection angle, respectively. Mathematically, the sinogram $s(l, \theta)$ is obtained via the Radon transform of the original image $f(x, y)$. In practice, the sinogram is discretized and stored as binned data, $\mathbf{s} \in \mathbb{R}^M$. Hence, with an operator $\mathcal{P}$ representing the Radon transform, we have a data generation model,

$$\mathbf{s} = \mathcal{P}f(x, y). \quad (2.7)$$

For image reconstruction, the target image is also discretized and represented in a vector form, $\mathbf{f} \in \mathbb{R}_{\geq 0}^N$. In addition, we define a discrete Radon transform with a matrix $\mathbf{P} \in \mathbb{R}^{M \times N}$. With the discretized original image and Radon transform operator, we have

$$\mathbf{s} = \mathbf{P}\mathbf{f}. \quad (2.8)$$

Each column of the projection matrix $\mathbf{P}[:, n]$ corresponds to a discrete sinogram generated from an impulse valued image $\mathbf{e}_n$, where $\mathbf{e}_n$ is a standard basis vector. In other words, $\mathbf{P}[:, n] = \mathbf{P}\mathbf{e}_n$. Also, it should be emphasized that (2.7) is *the actual model reflecting the underlying physical reality*, whereas (2.8) is *a constructed model for the purpose of image reconstruction*.

Since the sinogram $\mathbf{s}$ consists of projections of the unknown image $f(x, y)$ at different angles, a natural approach to image reconstruction is to reverse the projection process by smearing each projection back into the imaging space, a procedure known as “backprojection”. Backprojection of the sinogram is achieved by applying transpose of the projection matrix to the sinogram, $\mathbf{P}^\top \mathbf{s}$. Simply speaking, the backprojection matrix $\mathbf{P}^\top$ can be thought of as equally redistributing each sinogram element $\mathbf{s}[l]$ to its possible origins in the imaging vector $\mathbf{f}$. The simple backprojection (SBP) is performed by applying the backprojection operator to the measured sinogram, as follows:

$$\mathbf{f}_{\text{SBP}} = \mathbf{P}^\top \mathbf{s} = \mathbf{P}^\top \mathcal{P}f(x, y) = \mathbf{P}^\top \mathbf{P}\mathbf{f} \quad (2.9)$$

Note that the SBP reconstruction is perfect if and only if $\mathbf{P}^\top \mathbf{P} = \mathbf{I}$. In other words, $\mathbf{P}$ is an orthonormal matrix whose transpose is its inverse. An example of such imaging system

is a perfect pinhole camera; however, such condition is never met in practice and the SBP inevitably introduces blur in the reconstructed image $\mathbf{f}_{\text{SBP}}$.

The matrix $\mathbf{P}^\top \mathbf{P}$ in (2.9) possesses an important physical meaning, as the n -th column of $\mathbf{P}^\top \mathbf{P}$ is a blurred response from SBP when the input image is $\mathbf{e}_n$. For this reason, $\mathbf{P}^\top \mathbf{P}$ is often referred to as a PSF matrix. When the response is shift-invariant, the $\mathbf{P}^\top \mathbf{P}$ is a Toeplitz matrix and interpreted as a discrete convolution matrix. Therefore, the PSF matrix applies a low-pass frequency filter to the original image $\mathbf{f}$, leading to a blurred image $\mathbf{f}_{\text{SBP}}$.

Understanding the effect of SBP, a natural choice to reverse the blurring is to apply the inverse of the PSF matrix $(\mathbf{P}^\top \mathbf{P})^{-1}$ as follows:

$$\mathbf{f}_{\text{FBP}} = (\mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top \mathbf{s} = (\mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top \mathbf{P} \mathbf{f} = \mathbf{f}. \quad (2.10)$$

Applying $(\mathbf{P}^\top \mathbf{P})^{-1}$ is equivalent to applying an inverse frequency filter to undo the effect of blur added by SBP. Hence, the equation (2.10) first backproject the sinogram $\mathbf{s}$ onto the imaging space via SBP, then the blur introduced in SBP is removed with the inverse filter $(\mathbf{P}^\top \mathbf{P})^{-1}$. For this reason, (2.10) is known as the *backproject-then-filter* algorithm in CT image reconstruction. In fact, applying the inverse filter $(\mathbf{P}^\top \mathbf{P})^{-1}$ to the SBP image is equivalent to applying a filter to the sinogram $\mathbf{s}$, such that the backprojection of the filtered sinogram retrieves the original image $\mathbf{f}$. This approach is the traditional filtered backprojection (FBP) algorithm and therefore, the backproject-then-filter algorithm in (2.10) is equivalent to FBP. For more detailed discussions on the equivalence between the backproject-then-filter algorithm and FBP, see [40][41]. Figure 2.6 presents the results from the SBP and the FBP reconstruction methods when applied to the noiseless sinogram data from the Shepp-Logan phantom [42].

The FBP algorithm is traditionally understood with the filtering of the projection data $\mathbf{s}$ and its backprojection onto imaging space, without introducing the projection matrix $\mathbf{P}^\top$ [43]. Here we used a different approach to derive the FBP, and this linear-algebraic derivation of the FBP provides interesting insights into the properties of FBP.

Although the FBP in equation (2.10) yields perfect reconstruction with the noiseless measurement $\mathbf{s}$, the reconstructed images degrade quickly with addition of measurement noise. With the zero-mean Gaussian noise vector $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$ added to $\mathbf{s}$, the performance of the FBP can be analyzed with the help of singular vector decomposition (SVD).

Using SVD, we write $\mathbf{P} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are the orthonormal singular vector matrices and $\mathbf{\Sigma}$ is the diagonal matrix with singular values. The FBP with noise then is

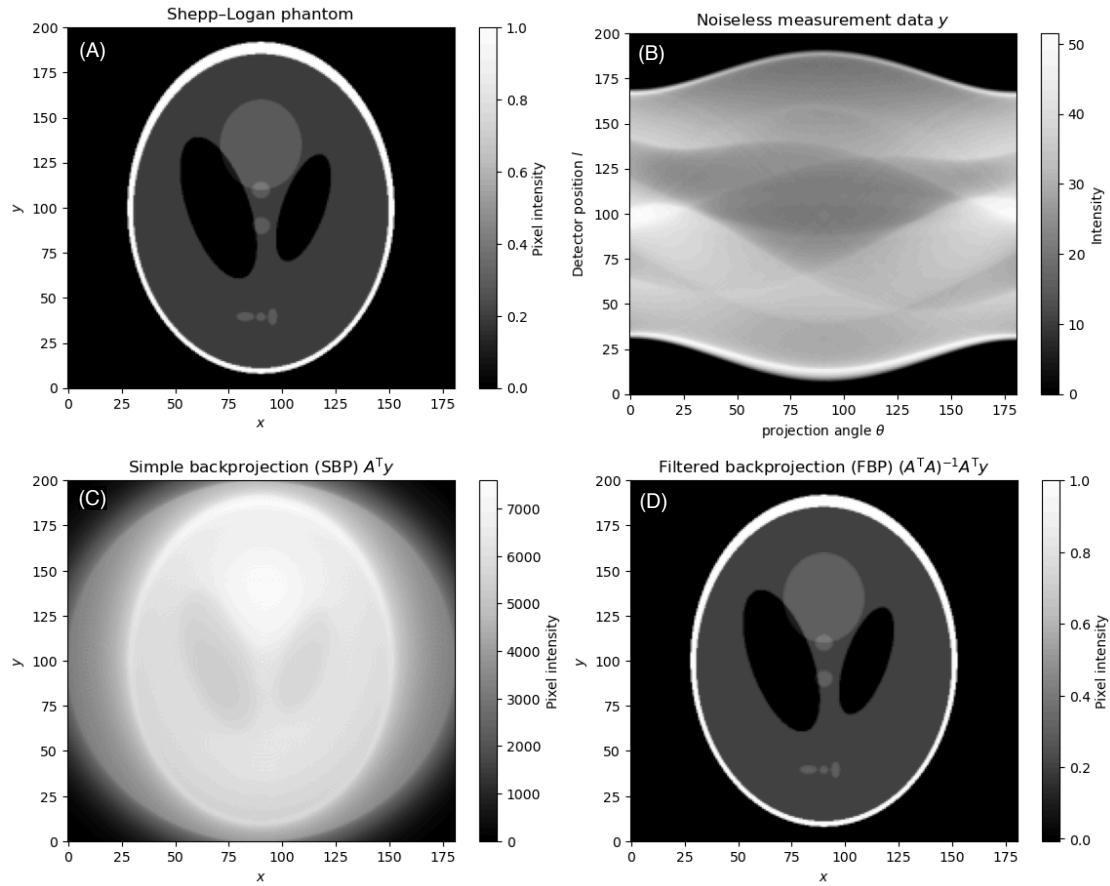


Figure 2.6: (A): Shepp-Logan phantom (B): Noiseless sinogram of the Shepp-Logan phantom (C): Simple backprojection of the noiseless sinogram. Note that it introduces blurr, which is characterized by the PSF matrix $\mathbf{P}^T \mathbf{P}$. (D): Filtered backprojection of the noiseless sinogram. In the absence of noise, the FBP produces perfect reconstruction.

written as,

$$\mathbf{f}_{\text{FBP}} = (\mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top (\mathbf{s} + \mathbf{n}) \quad (2.11)$$

$$= (\mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^\top)^{-1} \mathbf{V} \mathbf{\Sigma} \mathbf{U}^\top (\mathbf{s} + \mathbf{n}) \quad (2.12)$$

$$= (\mathbf{V} \mathbf{\Sigma}^{-2} \mathbf{V}^\top) \mathbf{V} \mathbf{\Sigma} \mathbf{U}^\top (\mathbf{s} + \mathbf{n}) \quad (2.13)$$

$$= (\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top) (\mathbf{s} + \mathbf{n}) \quad (2.14)$$

$$= (\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top) (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top \mathbf{f} + \mathbf{n}) \quad (\because \mathbf{s} = \mathbf{P} \mathbf{f} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top \mathbf{f}) \quad (2.15)$$

$$= \mathbf{f} + (\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top) \mathbf{n} \quad (2.16)$$

$$= \underbrace{\mathbf{f}}_{\text{original image}} + \underbrace{(\mathbf{V} \mathbf{\Sigma}^{-1} \mathbf{U}^\top) \mathbf{n}}_{\text{noise in FBP}} \quad (2.17)$$

$\mathbf{U}^\top$ and $\mathbf{V}$ are orthonormal matrices, and therefore applying it to the noise vector only rotates it without any changes in the magnitude $\|\mathbf{n}\|_2$. However, the inverse singular values in $\mathbf{\Sigma}^{-1}$ become very large when the projection matrix $\mathbf{P}$ is poorly conditioned (i.e., its smallest singular value is close to 0). As a result, the noise in FBP “blows up” and significantly degrades the reconstructed image $\mathbf{f}_{\text{FBP}}$. For this reason, the inverse filter $(\mathbf{P}^\top \mathbf{P})^{-1}$ is typically augmented to prevent exploding noise in the reconstructed image.

Another important insight can be found by inspecting the FBP equation (2.10). That is, the FBP in (2.10) is the solution to the following ordinary least-square (OLS) optimization problem:

$$\underset{\mathbf{f}}{\text{argmin}} \|\mathbf{P} \mathbf{f} - \mathbf{s}\|_2^2. \quad (2.18)$$

This optimization problem has been extensively studied, and is also known to yield maximum likelihood estimation (MLE) solution when the measurement is corrupted with independently and identically distribute (i.i.d.) Gaussian noise. In CT imaging, i.i.d. Gaussian noise is sometimes assumed due to high photon flux at the detector. However, in many other imaging modalities with photon starve measurements, Poisson noise more aptly describes the stochasticity of measurements, and therefore, the FBP is not an optimal choice for reconstruction. Moreover, the optimization problem (2.18) does not assume non-negativity in $\mathbf{f}$, leaving the risk of obtaining negative values in the reconstruction $\mathbf{f}_{\text{FBP}}$.

As described before, the FBP was first developed in the context of CT image reconstruction. However, its connection to the OLS (2.18) has prompted the use of the term “FBP” when the OLS is used for image reconstruction [44][45]. However, in many of these cases the PSF matrix $\mathbf{P}^\top \mathbf{P}$ is not a Toeplitz matrix (i.e., the PSF is not shift-invariant), and therefore its inverse $(\mathbf{P}^\top \mathbf{P})^{-1}$ cannot be interpreted as a conventional frequency filter.

Despite its limitations, the FBP is still widely adapted in clinical tomography image reconstruction due to its simple implementation, fast reconstruction speed, and a reasonably good reconstruction quality when the SNR is sufficiently high.

Iterative (Statistical) Image Reconstruction Methods

The limitations of the FBP are greatly alleviated with the iterative image reconstruction techniques. In iterative image reconstruction, the task of image reconstruction is formulated as an optimization problem, where the objective function is derived from the statistical characteristics of the imaging modalities. For this reason, these methods are often referred to as statistical reconstruction methods. Once the objective function is defined, an optimization algorithm is employed to determine the parameters (i.e., the reconstructed image) that solve the optimization problem. These optimization algorithms typically operate in an iterative manner, hence the name iterative image reconstruction methods.

We first mathematically describe formulation of binned data $\mathbf{y} \in \mathbb{R}^M$ with an unknown image $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$ and a known system matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{M \times N}$ as,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad \text{where } \mathbf{n} \in \mathbb{R}^M \quad \text{and} \quad \mathbf{n} \sim \mathcal{N}(0, \sigma^2) \quad (2.19)$$

In other words, we assume i.i.d. zero-mean Gaussian noise with variance σ^2 in the measurement $\mathbf{y}$. The system matrix $\mathbf{A}$ is constructed from the underlying physics of imaging systems, such that

$$\mathbf{A}[i, j] = T \cdot P(i \leftarrow j) := T \cdot P(\text{a radiation emission from voxel } j \text{ collected in bin } i), \quad (2.20)$$

where T denotes the measurement time². Note that the role of T is to keep the units of voxel values “radiation emission per unit time”.

From the probability density function (pdf) of the Gaussian distribution, the likelihood function $p(\mathbf{y}|\mathbf{x})$ is written as

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^M}} \exp\left(-\frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma^2}\right), \quad (2.21)$$

where $\|\cdot\|_2$ denotes l_2 norm. Note that the measurement $\mathbf{y}$ is known, so that the likelihood $p(\mathbf{y}|\mathbf{x})$ is a function of $\mathbf{x}$. In the MLE, the goal is to obtain $\mathbf{x}_{\text{MLE}}$ by solving the following optimization problem:

$$\mathbf{x}_{\text{MLE}} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{y}|\mathbf{x}) \quad (2.22)$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}} -\log p(\mathbf{y}|\mathbf{x}) \quad (2.23)$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}} \frac{M}{2} \log 2\pi\sigma^2 + \frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma^2} \quad (2.24)$$

$$= \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad (2.25)$$

²This particular probabilistic description of the system matrix is only true when the image vector $\mathbf{x}$ represents radiation emission rate (e.g., emission tomography). In the case of transmission tomography such as CT, it contains weights that model line-integration (i.e., Radon transform)

Equation (2.23) is derived from the monotonicity of the logarithm function, which ensures that the solution to the optimization problem remains unchanged. The transition from equation (2.24) to (2.25) relies on the fact that adding constants or multiplying non-negative constants to the optimization objective does not alter the optimization solution.

The optimization problem (2.25) is commonly referred to as an OLS problem. Upon inspection, it becomes evident that the optimization problem (2.25) has the same mathematical form as (2.10), which is solved using the FBP. Therefore, the solution to (2.25) is given as³,

$$\mathbf{x}_{\text{MLE}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}. \quad (2.26)$$

This observation implies that the underlying assumption of the FBP algorithm is that the measurement noise follows an i.i.d. Gaussian distribution. A significant caveat of the least-square problem (2.10) is that the image vector $\mathbf{x}$ is unconstrained. Consequently, in cases where the system matrix $\mathbf{A}$ is poorly conditioned and the measurement had low signal-to-noise ratio (SNR), there is a risk of obtaining negative elements in the MLE solution $\mathbf{x}_{\text{MLE}}$, which is not physically justified.

A non-negative MLE solution is found by imposing non-negativity constraint to the optimization problem (2.25) as follows:

$$\mathbf{x}_{\text{MLE}} = \underset{\mathbf{x}, \mathbf{x} \succcurlyeq 0}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2, \quad (2.27)$$

where we used $\mathbf{x} \succcurlyeq 0$ to denote element-wise non-negativity. The optimization problem (2.27) is commonly referred to as non-negative least-square problem.

With the addition of the non-negativity constraint, the (2.27) has no analytical closed-form solution. Instead, iterative approaches such as the projected gradient descent algorithm is employed to find the solution. Here, we briefly introduce the image space reconstruction algorithm (ISRA) and its derivation via gradient descent [46].

The gradient of the optimization objective function in (2.25) is,

$$\nabla_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 = 2\mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{y}) = 2(\mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{A}^\top \mathbf{y}) \quad (2.28)$$

One may view the gradient as twice of the difference between the SBP with the estimated image $\mathbf{A}^\top \mathbf{A}\mathbf{x}$ and the SBP from the measurement $\mathbf{A}^\top \mathbf{y}$. Obviously, when the estimate $\mathbf{x}$ is close to the solution of (2.27), the magnitude of the gradient is small. The gradient descent algorithm iterative updates $\mathbf{x}^{(t)}$, the estimate of $\mathbf{x}$ at iteration t , by taking a “step” towards the negative gradient direction as follows:

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \eta \nabla_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \big|_{\mathbf{x}=\mathbf{x}^{(t)}} \\ &= \mathbf{x}^{(t)} - 2\eta (\mathbf{A}^\top \mathbf{A}\mathbf{x}^{(t)} - \mathbf{A}^\top \mathbf{y}), \end{aligned} \quad (2.29)$$

³The MLE solution (2.26) is for when the system matrix $\mathbf{A}$ has more rows than columns. When $\mathbf{A}$ has more columns than rows, the OLS has infinite solutions and $\mathbf{x}_{\text{MLE}} = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{y}$ gives a solution with the smallest $\|\mathbf{x}_{\text{MLE}}\|_2$.

where $\boldsymbol{\eta}$ is the stepsize hyperparameter determining the amount of “leap” towards the negative gradient direction. Instead of keeping a constant $\boldsymbol{\eta}$ for the entire gradient descent algorithm, we choose an element-wise, adaptive stepsize by setting $\boldsymbol{\eta} = \frac{\mathbf{x}^{(t)}}{2\mathbf{A}^\top \mathbf{A}\mathbf{x}^{(t)}}$. Applying $\boldsymbol{\eta}$, the update rule (2.29) becomes

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \frac{2\mathbf{x}^{(t)}}{2\mathbf{A}^\top \mathbf{A}\mathbf{x}^{(t)}} \odot (\mathbf{A}^\top \mathbf{A}\mathbf{x}^{(t)} - \mathbf{A}^\top \mathbf{y}) \quad (2.30)$$

$$= \mathbf{x}^{(t)} \odot \frac{\mathbf{A}^\top \mathbf{y}}{\mathbf{A}^\top \mathbf{A}\mathbf{x}^{(t)}}, \quad (2.31)$$

where $\odot$ denotes an element-wise multiplication operator. Note that with the choice of adaptive hyperparameter $\boldsymbol{\eta}$, the negative terms in (2.29) is removed and the gradient update rule becomes multiplicative. Therefore, as long as the initial estimate $\mathbf{x}^{(0)}$ does not contain any negative elements, the non-negativity constraint is automatically satisfied. With a sufficient number of iterations, the ISRA algorithm converges to the solution of (2.27).

The ISRA algorithm introduced solves the non-negative least-squares problem, which assumes i.i.d. Gaussian noise in the measurements. However, in most radiation imaging modalities, measurement noise is better modeled by a Poisson distribution due to the discrete nature of radiation measurements. Under the Poisson noise model, the data measurement is described as,

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}) \quad \text{where} \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}. \quad (2.32)$$

The Poisson distribution is parameterized by its mean $\bar{\mathbf{y}}$, which is the forward projection of the image $\mathbf{x}$. From the definition of Poisson distribution probability density function (pdf), the Poisson likelihood is given as

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^M \frac{e^{-[\mathbf{A}\mathbf{x}]_i} [\mathbf{A}\mathbf{x}]_i^{\mathbf{y}_i}}{\mathbf{y}_i!}. \quad (2.33)$$

With the Poisson likelihood function, the MLE problem is written as follows:

$$\mathbf{x}_{\text{MLE}} = \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmax}} p(\mathbf{y}|\mathbf{x}) \quad (2.34)$$

$$= \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmin}} -\log p(\mathbf{y}|\mathbf{x}) \quad (2.35)$$

$$= \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmin}} \sum_{i=1}^M [\mathbf{A}\mathbf{x}]_i - \mathbf{y}_i \log [\mathbf{A}\mathbf{x}]_i + \log \mathbf{y}_i! \quad (2.36)$$

$$= \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmin}} \sum_{i=1}^M [\mathbf{A}\mathbf{x}]_i - \mathbf{y}_i \log [\mathbf{A}\mathbf{x}]_i \quad (2.37)$$

The quantity in (2.37) is sometimes referred to as Poisson loss. For notational brevity, it is often written in a vectorized form, $\mathbf{1}^\top (\mathbf{A}\mathbf{x} - \mathbf{y} \odot \log \mathbf{A}\mathbf{x})$, with $\mathbf{1}$ denoting a vector of ones.

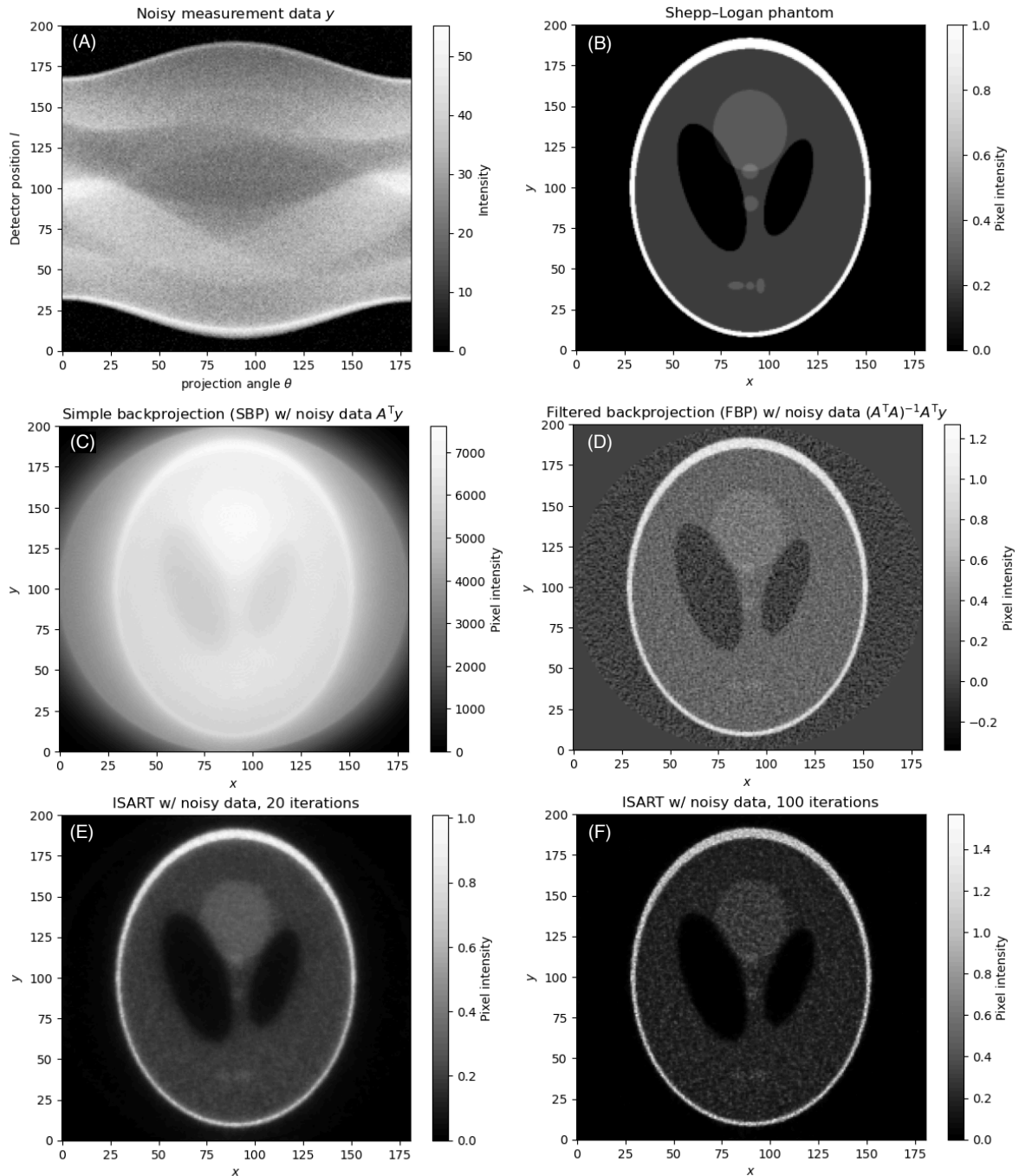


Figure 2.7: (A): Shepp-Logan phantom (B): A measured sinogram corrupted with Gaussian noise. (C): Simple backprojection reconstruction result (D): Filtered backprojection reconstruction result (E)(F): Reconstruction results after 20 and 100 iterations of SART. With increasing number of iterations, the reconstructed image exhibits pronounced high spatial frequency features, including noise.

The MLE problem with Poisson noise is most commonly solved using the Maximum Likelihood-Expectation Maximization (ML-EM) algorithm. The ML-EM algorithm was first derived by L.A. Shepp and Y.Vardi using the Expectation-Maximization (EM) algorithm [47]. The EM algorithm introduces the concept of complete data, representing the contribution of each pixel to each measurement element [48]. It alternates between updating the complete data and the image vector, ultimately converging to the solution, $\mathbf{x}_{\text{MLE}}$.

Here, instead of the EM based derivation, we introduce a gradient descent based derivation analogous to the ISRA algorithm. Recently, the gradient descent based optimization algorithms have been extensively studied for machine learning applications. Therefore, understanding the conventional ML-EM algorithm through the lens of gradient descent provides useful insights that can be readily adopted from the deep neural network research.

We first proceed by finding the gradient of the Poisson loss in (2.37) as follows:

$$\nabla_{\mathbf{x}} \mathbf{1}^\top (\mathbf{A}\mathbf{x} - \mathbf{y} \odot \log \mathbf{A}\mathbf{x}) = \mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}}. \quad (2.38)$$

With the gradient found, the gradient descent update rule is written as

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \boldsymbol{\eta} \nabla_{\mathbf{x}} \mathbf{1}^\top (\mathbf{A}\mathbf{x} - \mathbf{y} \odot \log \mathbf{A}\mathbf{x}) \big|_{\mathbf{x}=\mathbf{x}^{(t)}} \\ &= \mathbf{x}^{(t)} - \boldsymbol{\eta} (\mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}). \end{aligned} \quad (2.39)$$

Similarly to the derivation of the ISRT, we choose an adaptive hyperparameter $\boldsymbol{\eta}$ to remove the negative terms in the update rule (2.39). By setting $\boldsymbol{\eta} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1}}$, we have

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1}} \odot (\mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}) \\ &= \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1}} \odot \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}, \end{aligned} \quad (2.40)$$

Similarly to the ISRA, the ML-EM update rule in (2.40) is multiplicative. Therefore, with non-negative initialization of $\mathbf{x}$, the non-negativity constraint is satisfied automatically. Additionally, it is known that the ML-EM algorithm asymptotically converges to the solution to the MLE problem in (2.37).

The ML-EM algorithm introduced assumes binned measurement data. In the case of list-mode data, Barret and Parra *et al.*, [36][49] have derived the list-mode ML-EM algorithm from the first principles, using the fundamental stochasticity of list-mode data and its likelihood function. However, perhaps an easier, albeit mathematically less rigorous, way to “derive” the list-mode ML-EM is to consider list-mode data as binned data with infinitesimally small bins.

In the limit of bin widths approaching zero, the measurement vector $\mathbf{y} \in \{0, 1\}^\infty$ becomes infinitely large with its elements being either 1 or 0. Consequently, the number of rows in the system matrix $\mathbf{A} \in \mathbb{R}^{\infty \times N}$ becomes infinite. With the dimensionality of $\mathbf{y}$ and $\mathbf{A}$ in mind, recall that the binned data ML-EM update rule is given as

$$\mathbf{x}^{(t+1)} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1}} \odot \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}. \quad (2.41)$$

Upon inspection, it can be noticed that the computation of $\mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}$ is unaffected if we remove all zeroed entries in $\mathbf{y}$ and the corresponding rows in $\mathbf{A}$. Let $\mathbf{1} \in \{1\}^M$ and $\mathbf{A}' \in \mathbb{R}_{\geq 0}^{M \times N}$ denote such modified data vector and system matrix. Then the ML-EM update rule becomes

$$\mathbf{x}^{(t+1)} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1}} \odot \mathbf{A}'^\top \frac{\mathbf{1}}{\mathbf{A}' \mathbf{x}^{(t)}}. \quad (2.42)$$

And rewriting the update rule element-wise,

$$\begin{aligned} \mathbf{x}^{(t+1)}[j] &= \frac{\mathbf{x}^{(t)}[j]}{\sum_{l=1}^{\infty} \mathbf{A}[l, j]} \sum_{i=1}^M \frac{\mathbf{A}'[i, j]}{\sum_{k=1}^N \mathbf{A}'[i, k] \mathbf{x}^{(t)}[k]} \\ &= \frac{\mathbf{x}^{(t)}[j]}{T \sum_{l=1}^{\infty} P(l \leftarrow j)} \sum_{i=1}^M \frac{P(i \leftarrow j)}{\sum_{k=1}^N P(i \leftarrow k) \mathbf{x}^{(t)}[k]} \end{aligned} \quad (2.43)$$

Next, we note that the probability $P(i \leftarrow j)$ (defined in (2.20)) can be written as follows:

$$\begin{aligned} P(i \leftarrow j) &= P(i|j)P(j), \quad \text{where} \\ P(j) &:= \text{Probability that a radiation emission from voxel } j \text{ results in an event} \\ P(i|j) &:= \text{Given that a radiation emission from voxel } j \text{ results in an event,} \\ &\quad \text{the probability that it is the } i\text{-th event} \end{aligned} \quad (2.44)$$

The probability $P(j)$ is often referred to as the *sensitivity* of voxel j , and it can be computed as,

$$P(j) = \sum_{i=1}^{\infty} \mathbf{A}[i, j] = \sum_{i=1}^{\infty} P(i \leftarrow j). \quad (2.45)$$

With the probabilities defined in (2.44) and (2.45), the ML-EM update rule in (2.43) becomes,

$$\mathbf{x}^{(t+1)}[j] = \frac{\mathbf{x}^{(t)}[j]}{T \cdot P(j)} \sum_{i=1}^M \frac{P(i|j)P(j)}{\sum_{k=1}^N P(i|k)P(k) \mathbf{x}^{(t)}[k]} \quad (2.46)$$

Note that the computation of $P(i|j)$ is difficult for many imaging systems. However, $P(j|i)$, which represents probability of the i -th event resulting from a radiation emitted from the j -th voxel, is much easier to compute. Hence, using Bayes' rule, we write $P(i|j) = P(j|i)P(i)/P(j)$ and the update rule becomes,

$$\begin{aligned} \mathbf{x}^{(t+1)}[j] &= \frac{\mathbf{x}^{(t)}[j]}{T \cdot P(j)} \sum_{i=1}^M \frac{P(j|i)P(i)/P(j) \cdot P(j)}{\sum_{k=1}^N P(k|i)P(i)/P(k) \cdot P(k) \mathbf{x}^{(t)}[k]} \\ &= \frac{\mathbf{x}^{(t)}[j]}{T \cdot P(j)} \sum_{i=1}^M \frac{P(j|i)}{\sum_{k=1}^N P(k|i) \mathbf{x}^{(t)}[k]} \end{aligned} \quad (2.47)$$

A few remarks follow from (2.47). First, in the derivation of (2.47), the sensitivity of the j -th voxel $P(j)$ was derived from the sum of the system matrix $\sum_{i=1}^{\infty} \mathbf{A}[i, j] = \sum_{i=1}^{\infty} P(i \leftarrow j)$. Obviously, this summation is difficult to perform as we must consider all possible infinite number of events. Therefore, in practice, it is estimated from a Monte Carlo particle simulation or a real-world measurement.

Second, it should be noted that a collection of $P(j|i)$ across all $j = 1, 2, \dots, N$ voxels creates a probability map of possible radiation origin for the i -th event. For example, in Compton imaging, the map is a conic surface given by the i -th compton event. In the case of PET imaging, it is a line-of-response connecting two coincidence interaction positions.

Lastly, for notational convenience, we define the sensitivity vector $\mathbf{s}[j] = P(j)$ and the list-mode system matrix $\mathbf{A}_{\text{list}}[i, j] = P(j|i)$ for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$. Then, the element-wise list-mode ML-EM update rule (2.47) is re-written as

$$\mathbf{x}^{(t+1)} = \frac{\mathbf{x}^{(t)}}{T\mathbf{s}} \odot \mathbf{A}_{\text{list}}^{\top} \frac{\mathbf{1}}{\mathbf{A}_{\text{list}}\mathbf{x}^{(t)}}. \quad (2.48)$$

Notably, the list-mode ML-EM (2.48) is largely similar to the binned data ML-EM in (2.40). For this reason, it is often said that the binned data ML-EM is applicable to the list-mode data. However, care should be exercised when using the list-mode ML-EM, since the system matrix and the sensitivity vector are fundamentally different in those two algorithms.

From a classical convergence-rate analysis, it has been shown that the ISRA and the ML-EM algorithms exhibit a very slow sub-linear convergence rate [50]. Furthermore, these algorithms initially adapt to low spatial frequency components in the image during earlier iterations, followed by high spatial frequency components in later iterations [51]. Although high spatial frequency components contain important fine details in an image (e.g., edges, small point-like features, etc.), they also introduce undesirable image noise. For this reason, when using the ISRA or the ML-EM algorithm in practice, the algorithm is typically not run until full convergence. Instead, the process is terminated after a preset number of iterations, a technique known as *early stopping*. However, finding a reasonable stopping rule may be a difficult task, as it depends on the imaging system and measurements used for reconstruction [52].

Regularization and Bayesian image reconstruction

As we have seen in the previous section, reconstructed images are inevitably subject to noise, unless noiseless measurement data is used. Early stopping of the ISRA or the ML-EM algorithm is often helpful in reducing such unwanted noise. However, it is often difficult to decide when to stop the iterative process, as premature termination may lead to lack of desirable image details. A more direct method to reduce noise and enhance image quality involves an application of Bayesian statistics. To understand Bayesian statistics, it is helpful to first understand frequentist statistics, which the MLE is based on.

Suppose we have noisy measurement data $\mathbf{y}$, which is generated from an image $\mathbf{x}$. When the noise in $\mathbf{y}$ is Poisson distributed, $\mathbf{y}$ is modelled as $\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x})$. In frequentist framework, the image $\mathbf{x}$ is considered an unknown but fixed parameter. The randomness arises solely from the measurement process that generates the observed data $\mathbf{y}$. Therefore, the likelihood $p(\mathbf{y}|\mathbf{x})$ describes how probable it is to observe the given data for a particular fixed $\mathbf{x}$. Maximum likelihood estimation (MLE) seeks to find the $\mathbf{x}$ that makes the observed $\mathbf{y}$ most likely.

In contrast, the Bayesian framework treats the parameter $\mathbf{x}$ itself as a random variable, equipped with a prior distribution $p(\mathbf{x})$ that represents our knowledge or assumptions about $\mathbf{x}$ before seeing the data. Once the data $\mathbf{y}$ is observed, it is considered fixed, and the uncertainty shifts to the parameter. The posterior distribution $p(\mathbf{x}|\mathbf{y})$ combines the prior and the likelihood, capturing our updated understanding of $\mathbf{x}$ given the observed $\mathbf{y}$. Estimating $\mathbf{x}$ by selecting the value that maximizes $p(\mathbf{x}|\mathbf{y})$ is known as maximum *a posteriori* (MAP) estimation, the Bayesian analog to the frequentist MLE.

In order to obtain the posterior $p(\mathbf{x}|\mathbf{y})$, we use Bayes' rule as follows:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \quad (2.49)$$

where $p(\mathbf{x})$ and $p(\mathbf{y})$ are called prior distribution and marginal likelihood (or evidence), respectively. Note that $p(\mathbf{x})$ reflects our prior knowledge of $\mathbf{x}$ before making any measurements. Therefore, the posterior $p(\mathbf{x}|\mathbf{y})$ represents our updated uncertainty about $\mathbf{x}$ after the measurement $\mathbf{y}$. With the Bayes' rule (2.49), the MAP image reconstruction problem is formulated as follows:

$$\mathbf{x}_{\text{MAP}} = \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y}) \quad (2.50)$$

$$= \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmax}} \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \quad (2.51)$$

$$= \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) + \log p(\mathbf{y}) \quad (2.52)$$

$$= \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) \quad (2.53)$$

Note that $\log p(\mathbf{y})$ is a constant and removing it does not affect the optimization. The difference between the MLE and MAP is the negative log-prior $-\log p(\mathbf{x})$. If we choose

$p(\mathbf{x}) = \exp(-\lambda U(\mathbf{x}))$, (2.53) becomes

$$\mathbf{x}_{\text{MAP}} = \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} -\log p(\mathbf{y}|\mathbf{x}) + \lambda U(\mathbf{x}). \quad (2.54)$$

The optimization objective in (2.54) explicitly shows the tradeoff between the data fidelity (i.e., the likelihood function) $-\log p(\mathbf{y}|\mathbf{x})$ and the prior knowledge $\lambda U(\mathbf{x})$. Due to this tradeoff, $U(\mathbf{x})$ is sometimes referred to as a penalty function. The tradeoff is adjusted with a hyperparameter $\lambda \in \mathbb{R}_{\geq 0}$.

The ML-EM algorithm (2.40), which solves the MLE problem with Poisson loss can be modified to solve the MAP problem in (2.54). With the Poisson likelihood function, (2.54) is written as,

$$\mathbf{x}_{\text{MAP}} = \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} \mathbf{1}^\top (\mathbf{A}\mathbf{x} - \mathbf{y} \odot \log \mathbf{A}\mathbf{x}) + \lambda U(\mathbf{x}). \quad (2.55)$$

Similarly to the ML-EM derivation, we employ the gradient descent algorithm with an adaptive stepsize parameter to solve (2.55). The gradient descent rule is given as,

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{x}^{(t)} - \boldsymbol{\eta} \nabla_{\mathbf{x}} (\mathbf{1}^\top (\mathbf{A}\mathbf{x} - \mathbf{y} \odot \log \mathbf{A}\mathbf{x}) + \lambda U(\mathbf{x}))|_{\mathbf{x}=\mathbf{x}^{(t)}} \\ &= \mathbf{x}^{(t)} - \boldsymbol{\eta} (\mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}} + \lambda \nabla_{\mathbf{x}} U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(t)}}). \end{aligned} \quad (2.56)$$

Similarly to the derivation of the ML-EM algorithm, we choose an adaptive hyperparameter $\boldsymbol{\eta} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1} + \lambda \nabla_{\mathbf{x}} U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(t)}}}$, yielding the multiplicative update rule

$$\mathbf{x}^{(t+1)} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1} + \lambda \nabla_{\mathbf{x}} U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(t)}}} \odot \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A}\mathbf{x}^{(t)}}. \quad (2.57)$$

The algorithm (2.57) is often referred to as the MAP-EM algorithm. Note that the MAP-EM update rule can be equivalently derived from the EM algorithm [53][54]. The EM based derivation involves estimation of $\nabla_{\mathbf{x}} U(\mathbf{x})$ with respect to the current image estimate $\mathbf{x}^{(t+1)}$, which is generally intractable. Therefore, often the gradient with respect to the previous estimate $\mathbf{x}^{(t)}$ is used as an approximation, eventually leading to the same update rule given in (2.57). For this reason, the MAP-EM algorithm is sometimes referred to as the *one-step-late* algorithm [54].

When the vector $\mathbf{A}^\top \mathbf{1} + \lambda \nabla_{\mathbf{x}} U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(t)}}$ in (2.57) contains non-negative entries only, the MAP-EM algorithm monotonically decreases the objective function (2.55) and eventually converges to a solution [55]. However, it has been shown that with negative entries in $\mathbf{A}^\top \mathbf{1} + \lambda \nabla_{\mathbf{x}} U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(t)}}$, the convergence is not guaranteed and more concerningly, the non-negativity constraint may not be satisfied. As a result, several modifications to alleviate the problems have been suggested [56][57].

Various MAP techniques have been studied extensively in the context of image reconstruction, ranging from Tikhonov regularization [58], total variation (TV) [59] regularization to plug-and-play priors [60] that utilize pre-trained deep neural networks. It has been shown

that the MAP approaches can significantly improve the quality of reconstructed images compared to the MLE, while alleviating the ill-posedness of the image reconstruction problem [61]. Moreover, unlike frequentist approaches such as MLE, the Bayesian paradigm offers a straightforward interpretation of uncertainties, including both aleatoric uncertainties, arising from the inherent randomness in data, and epistemic uncertainties, resulting from a lack of data [62].

Chapter 3

Kullback-Leibler Divergence Regularization for Free-moving Imaging

As described in Chapter 2, localization and mapping of radiological sources in 3D with free-moving detector platforms is a promising technology, with a variety of potential applications including nuclear proliferation detection and emergency response. Moreover, the use of Simultaneous Localization and Mapping (SLAM), in combination with portable radiation detectors, has opened up the possibility of using a series of time-integrated count data measured from a single detector to reconstruct radiological source locations and distributions in three dimensions.

For free-moving single detector imaging, image reconstruction problems can be solved with the standard Maximum Likelihood Expectation Maximization (ML-EM). However, due to the limited angular modulation of a single detector and effectively unconstrained 3-D imaging space, image reconstruction problems are highly ill-posed and underdetermined. As a result, the ML-EM algorithm is not robust and suffers from severe overfitting.

To overcome the limitations of the ML-EM algorithm, studies have been done on the cases where sparsity in the reconstructed image can be assumed, i.e. point sources [30][33]. The point source assumption is valid in many radiological source search scenarios; however, in situations where one wishes to map distributed source activities, a more general formalism is required.

In this work, we propose a Bayesian Maximum *a Posteriori* Expectation Maximization (MAP-EM) image reconstruction algorithm with a Kullback-Leibler (KL) divergence prior. In the following, we briefly lay out the mathematical derivation of the proposed algorithm from the perspective of Bayesian statistics. Then, the proposed algorithm is benchmarked against the standard ML-EM algorithm on a simulated source mapping scenario to prove its effectiveness.

3.1 Problem formulation

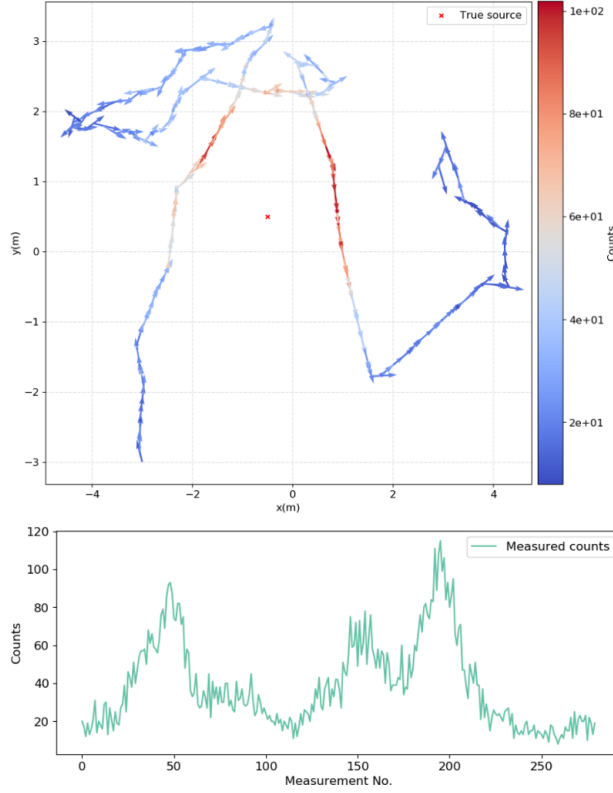


Figure 3.1: (Top) An example of single-detector imaging scenario with a $10 \mu\text{Ci}$ point source placed at $(x, y, z) = (-0.5, 0.5, 0)$ m. (Bottom) The measured count data. The fluctuation in the measurement data primarily comes from the variation in source to detector distance. Also, the Poisson measurement noise is observed in the high-frequency noise feature in the measurement.

In a single detector imaging scenario, the objective is to estimate an image vector $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$ and a constant background rate $b \in \mathbb{R}_{\geq 0}$ from time-series of count measurement data $\mathbf{y} \in \mathbb{Z}_{\geq 0}^M$. The measurement data and the image are related through a system matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{M \times N}$. Therefore, the data generation model is formulated as

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}) \quad \text{where} \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + b\mathbf{t}, \quad (3.1)$$

where $\mathbf{t}$ is a vector of integration time of each measurement. The problem formulation in (3.1) can be simplified by augmenting the system matrix and image vector as,

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix}. \quad (3.2)$$

For notational simplicity, $\mathbf{A}$ and $\mathbf{x}$ will denote the augmented system matrix and image vector, henceforth. Vector $\bar{\mathbf{y}}$ is forward projection of image $\mathbf{x}$, which forms the mean of the Poisson random variable $\mathbf{y}$. Elements of the system matrix $\mathbf{A}[i, j]$ represent the measurement time multiplied by the probability of detection at position i , when a single particle is emitted from voxel j . This probability is calculated as

$$\mathbf{A}[i, j] = \frac{\text{Area}_{\text{eff}}(i, j)}{4\pi \|\mathbf{r}(i, j)\|_2^2} \Delta t, \quad (3.3)$$

where $\|\mathbf{r}(i, j)\|_2$ and Δt denotes the distance between the i -th measurement position and the j -th voxel and the measurement time, and $\text{Area}_{\text{eff}}(i, j)$ is the effective area. The effective area is defined as the ratio of detection rate (detections per unit time) to the intensity of a parallel radiation beam (particles per unit area per unit time). It is important to note that the effective area depends on the orientation of the detector at the i -th measurement relative to the j -th voxel. Also, the efficiency formula (3.3) is not valid when the detector is too close to the voxel j (i.e., $\|\mathbf{r}(i, j)\|_2 \approx 0$). In such cases, a small correction factor ϵ may be added to the denominator of (3.3) [33].

A simulated example of the single detector imaging scenario is shown in Fig. 3.1. A point source with an activity of 10 μCi was placed at $(x, y, z) = (-0.5, 0.5, 0)$ m, and a spherical detector with a 20 cm radius was moved around the source at 1 m above the ground level. The entire measurement was taken for about 30 seconds, and the detector moving path is indicated with arrows, where each arrow marks a measurement position with an integration time of 0.1 second. The corresponding measurement data is also presented in Fig. 3.1. As observed, the measured counts vary across the measurement, depending on the source-to-detector distance. Note that the measured count data is corrupted by Poisson noise, which manifests as high frequency fluctuations. Figure 3.2 shows the result of the ML-EM reconstruction after 20 and 200 iterations. As observed, the reconstruction fails to localize the point source in both cases. Moreover, it illustrates the pathological problem of the image reconstruction with a single detector; the reconstructed activity is biased towards the moving path of the detector. This behavior is observed regardless of the ML-EM iterations. Moreover, with a large number of ML-EM iterations, the forward projection fits to the high frequency Poisson noise in the measurement, which is a typical sign of overfitting.

With a single detector, the directionality of radiation cannot be retrieved. Therefore, it is not difficult to intuitively understand why the ML-EM reconstruction fails to localize the source. More insights in the failure of the ML-EM can be found by examining the system matrix $\mathbf{A}$. Let j and $\mathbf{a}_j$ be an index of a voxel that lies right beneath the measurement path, and the column vector of $\mathbf{A}$ corresponding the the voxel. As the detection probability follows the inverse-square law, elements of $\mathbf{a}_j$ rapidly peaks as the detector passes by the voxel. As a result, the vector forms a spike-like trend, akin to a standard basis vector.

Let $\phi(m)$ represent a function that inputs an index of a measurement number $m = 1, 2, \dots, M$ and outputs the index of a voxel closest to the m -th measurement position. Then, let $[\mathbf{a}_{\phi(1)}, \mathbf{a}_{\phi(1)}, \dots, \mathbf{a}_{\phi(M)}]$ and $\mathbf{A}_r$ denote a collection of $\phi(m)$ -th column vectors of $\mathbf{A}$ and

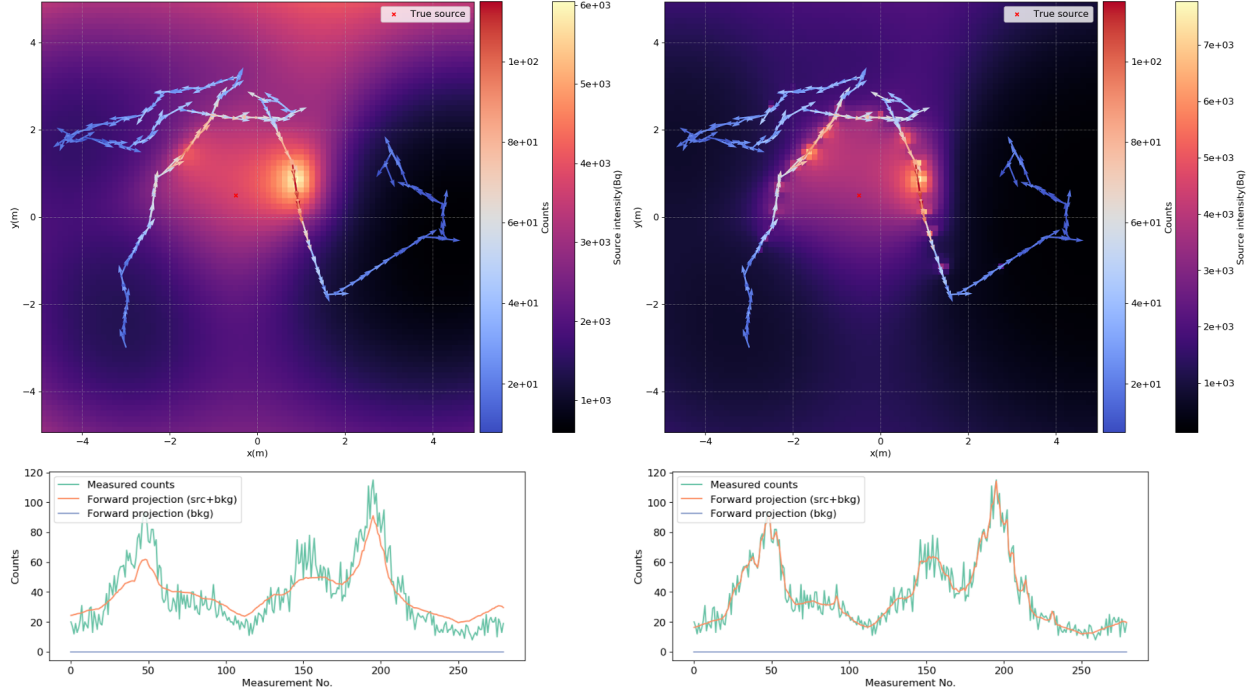


Figure 3.2: Reconstruction results using the ML-EM after (left) 20 and (right) 200 iterations. In both cases, the reconstruction fails to localize the point source. After 20 iterations of the ML-EM, the measurement data is clearly underfitted, whereas after 200 iterations clear signs of overfitting are observed.

the rest of the system matrix. Then, the system matrix can be re-arranged to be written as,

$$\mathbf{A} = [\mathbf{A}_{\phi(1)} \quad \mathbf{A}_{\phi(2)} \quad \dots \quad \mathbf{A}_{\phi(M)} \mid \mathbf{A}_r] \approx [k\mathbf{I} \mid \mathbf{A}_r], \quad (3.4)$$

where k is a constant. In other words, those voxels close to measurement positions collectively resemble an identity matrix, and therefore, it is sufficient to use only those voxels to fit the measurement $\mathbf{y}$, and even the noise in $\mathbf{y}$. As a result, the ML-EM algorithm prefers to use only those voxels to fit the data $\mathbf{y}$.

3.2 Kullback-Leibler divergence regularization

In order to improve the ML-EM reconstruction, here we propose a novel regularization method based on the Kullback-Leibler (KL) divergence, which is also known as relative entropy [63]. The KL divergence is a measure of dissimilarity between two probability

distributions. The definition of KL divergence between two pdfs $p(x)$ and $q(x)$ is given as:

$$D_{\text{KL}}(p \parallel q) = \mathbb{E}_{p(x)} \left[\log \frac{p(x)}{q(x)} \right] = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx. \quad (3.5)$$

The KL divergence is always non-negative ($D_{\text{KL}}(p \parallel q) \geq 0$) and zero if and only if $p(x) = q(x)$. Intuitively, KL divergence can be interpreted as a distance between two distributions, although it is not a true distance metric, as $D_{\text{KL}}(p \parallel q) \neq D_{\text{KL}}(q \parallel p)$. In the context of information theory, it represents an expected excess information from using $q(x)$ as a model instead of $p(x)$, when the actual distribution is $p(x)$.

In this work, we treat the normalized measurement vector $\hat{\mathbf{y}} = \frac{\mathbf{y}}{\mathbf{y}^\top \mathbf{1}}$ and normalized columns of the system matrix $\hat{\mathbf{a}}_j = \frac{\mathbf{a}_j}{\mathbf{a}_j^\top \mathbf{1}}$ as probability mass functions (pmf). The KL divergence between the pmfs is then calculated as follows:

$$D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j) = \sum_{i=1}^M \hat{\mathbf{y}}_i \log \frac{\hat{\mathbf{y}}_i}{\hat{\mathbf{a}}_{ij}}. \quad (3.6)$$

Note that the integration in (3.5) is replaced with a sum, as the distributions given are pmfs in (3.6). Also we define $\hat{\mathbf{y}}_i \log \frac{\hat{\mathbf{y}}_i}{\hat{\mathbf{a}}_{ij}} = 0$ when $\hat{\mathbf{y}}_i = 0$. With the KL divergence computed, we propose the following regularized optimization problem:

$$\mathbf{x}_{\text{KLD}} = \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmin}} \mathbf{Ax} - \mathbf{y} \odot \log \mathbf{Ax} + \beta \sum_{j=1}^N D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j) \mathbf{x}_j, \quad (3.7)$$

where $\beta \geq 0$ is a hyperparameter determining the regularization strength. Intuitively, the KL divergence $D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j)$ measures dissimilarity between the the measurement $\mathbf{y}$ and the forward projection from voxel j . Therefore, we penalize those voxels where the measurement data is unlikely to have come from.

A deeper understanding of the KL divergence regularization scheme can be gained by drawing a connection between the KL divergence and Poisson distributions. We re-write the

KL divergence in (3.7) as follows:

$$\sum_{j=1}^N D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j) \quad (3.8)$$

$$= \sum_{j=1}^N \left(\sum_{i=1}^M \hat{\mathbf{y}}_i \log \frac{\hat{\mathbf{y}}_i}{\hat{\mathbf{a}}_{ij}} \right) \quad (3.9)$$

$$= \sum_{j=1}^N \left(\sum_{i=1}^M \mathbf{y}_i \log \frac{\mathbf{y}_i}{C \hat{\mathbf{a}}_{ij}} \right) \quad (C = \mathbf{y}^\top \mathbf{1}) \quad (3.10)$$

$$= \sum_{j=1}^N \sum_{i=1}^M \mathbf{y}_i \log \mathbf{y}_i - \mathbf{y}_i \log C \hat{\mathbf{a}}_{ij} \quad (3.11)$$

$$\approx \sum_{j=1}^N \sum_{i=1}^M \log(\mathbf{y}_i!) + \mathbf{y}_i - \mathbf{y}_i \log C \hat{\mathbf{a}}_{ij} \quad (\because \text{Stirling's approximation}) \quad (3.12)$$

$$= \sum_{j=1}^N \sum_{i=1}^M \log(\mathbf{y}_i!) + C \hat{\mathbf{a}}_{ij} - \mathbf{y}_i \log C \hat{\mathbf{a}}_{ij} \quad (\because \sum_{i=1}^M \mathbf{y}_i = \sum_{i=1}^M C \hat{\mathbf{a}}_{ij}) \quad (3.13)$$

$$= -\log \left(\prod_{j=1}^N \prod_{i=1}^M \frac{e^{-C \hat{\mathbf{a}}_{ij}} (C \hat{\mathbf{a}}_{ij})^{\mathbf{y}_i}}{(\mathbf{y}_i!)} \right) \quad (3.14)$$

$$= -\log \left(\prod_{j=1}^N p(\mathbf{y}) \right), \quad \text{where } \mathbf{y} \sim \text{Poisson}(C \hat{\mathbf{a}}_j). \quad (3.15)$$

In other words, the KL divergence $D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j)$ encodes the negative log-likelihood of the entire measurement data $\mathbf{y}$ coming from the scaled forward projection from the j -th voxel, $C \hat{\mathbf{a}}_j$, with the scaling factor C ensuring the summation of the Poisson mean values matching the total measurement counts.

Another interesting fact about the KL divergence regularization is that it follows the formulation of adaptive LASSO regression [64]. In adaptive LASSO, the regularization term is given as

$$(\text{Adaptive LASSO regularizer}) = \beta \sum_{j=1}^N w_j |\mathbf{x}_j|, \quad (3.16)$$

For the proposed KL divergence regularizer, the weight term w_j is replaced with $D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j)$. Similarly to LASSO regression, adaptive LASSO regression induces sparse solution estimate of $\mathbf{x}$ [65]. This is a desirable property for imaging applications, as in many radiation imaging scenarios sparsely distributed radioactive source is assumed (e.g., point sources). Moreover, it has been shown that adaptive LASSO enjoys the oracle properties, meaning that it performs as well as if the true underlying model were given in advance [64].

With the KL divergence regularization term included, the MAP-EM update formula is modified as follows:

$$\mathbf{x}^{(t+1)} = \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1} + \nabla_{\mathbf{x}_j} \beta \sum_{j=1}^N D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_j) \mathbf{x}_j} \odot \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A} \mathbf{x}^{(t)}} \quad (3.17)$$

$$= \frac{\mathbf{x}^{(t)}}{\mathbf{A}^\top \mathbf{1} + \beta \mathbf{d}_{\text{KL}}} \odot \mathbf{A}^\top \frac{\mathbf{y}}{\mathbf{A} \mathbf{x}^{(t)}}, \quad (3.18)$$

where we defined $\mathbf{d}_{\text{KL}} = [D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_1), D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_2), \dots, D_{\text{KL}}(\hat{\mathbf{y}} \parallel \hat{\mathbf{a}}_N)]$ in equation (3.18). Note that $\mathbf{d}_{\text{KL}}$ does not involve $\mathbf{x}$, and therefore it needs to be computed only once before reconstruction, ensuring computational efficiency of the regularized MAP-EM algorithm. The KL divergence regularized MAP-EM algorithm (3.18) will be referred to as KLD MAP-EM henceforth.

3.3 Results

In this section, we present KLD MAP-EM reconstruction results on both simulated and real-world data, and compare them to the unregulated ML-EM reconstruction.

Point source reconstruction

First, figure 3.3 shows the result of the KLD MAP-EM with the measurement scenario shown in 3.3. The reconstruction was obtained after 200 iterations with $\beta = 5 \times 10^{-2}$. Note that with the KLD MAP-EM, the reconstruction is no longer biased towards the path and the point source is successfully localized. Moreover, the forward projection illustrated in figure 3.3 (b) reveals that the reconstruction is sufficiently regularized, and does not suffer from the overfitting problem.

Distributed source reconstruction

To show the general applicability of the KLD MAP-EM algorithm, a more challenging scenario with a distributed source was simulated and shown in figure 3.4. A 100 μCi isotropic Gaussian distributed source was simulated at $(x, y) = (5, 0)$ m, and a 10-cm diameter spherical detector was moved around for about 90 seconds. A constant background level was assumed with 50 cps, and integration time of 0.1 seconds was used.

Note that the SNR in the measurement is considerably worse than the point source measurement data, posing a challenge in image reconstruction. As a result, the ML-EM algorithm severely suffers from overfitting, as can be seen in figure 3.4 (b). Note that the typical bias of the ML-EM is apparent. Again, most source activities are concentrated around the path. Since the sensitivity of those voxels near the path is high, the total estimated source activity was 23.9 μCi , much lower than the ground truth of 100 μCi .

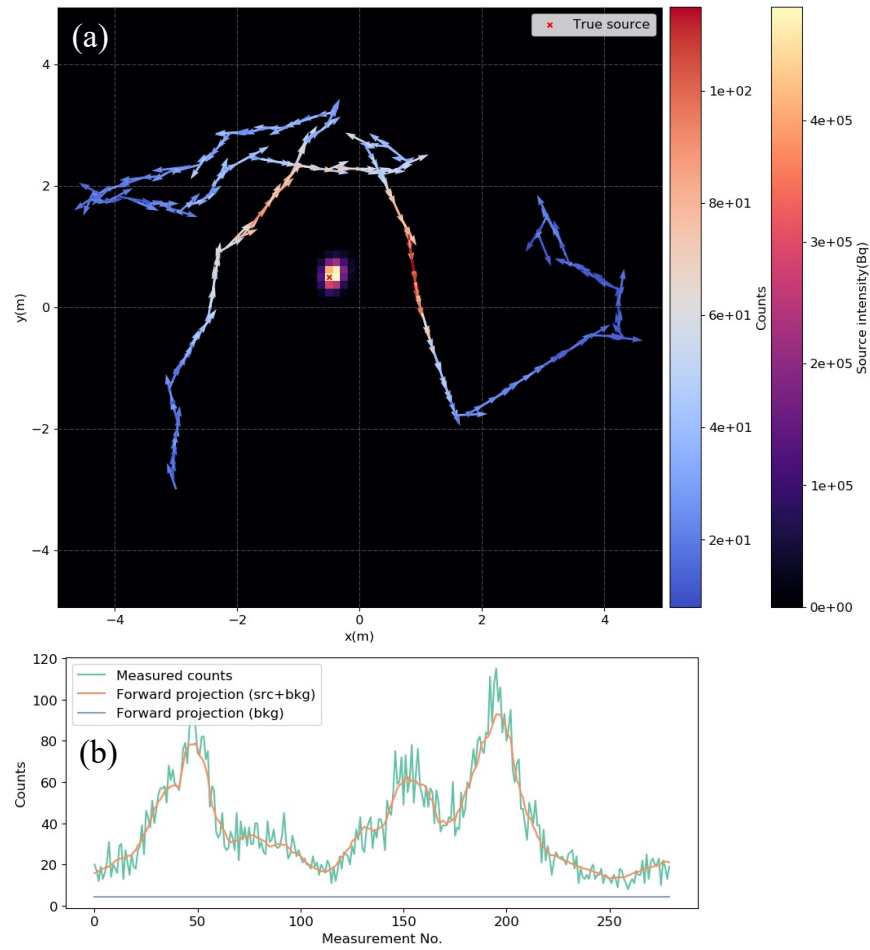


Figure 3.3: (a) KLD MAP-EM reconstruction of the point source measurement scenario in figure 3.1. Unlike the ML-EM result shown in figure 3.2, the KLD MAP-EM successfully localizes the point source location. (b) Forward projection $\bar{\mathbf{y}}$ from the reconstructed image. Note that the forward projection does not fit to the noise feature.

Figure 3.4 (c) shows the result of the KLD MAP-EM reconstruction after 200 iterations with $\beta = 5 \times 10^{-2}$. As can be seen, the KLD MAP-EM yields a good estimates of the source distribution and the background countrate, without overfitting the measurement data. The total reconstructed source activity was 92.2 μCi , which is in good agreement with the ground truth.

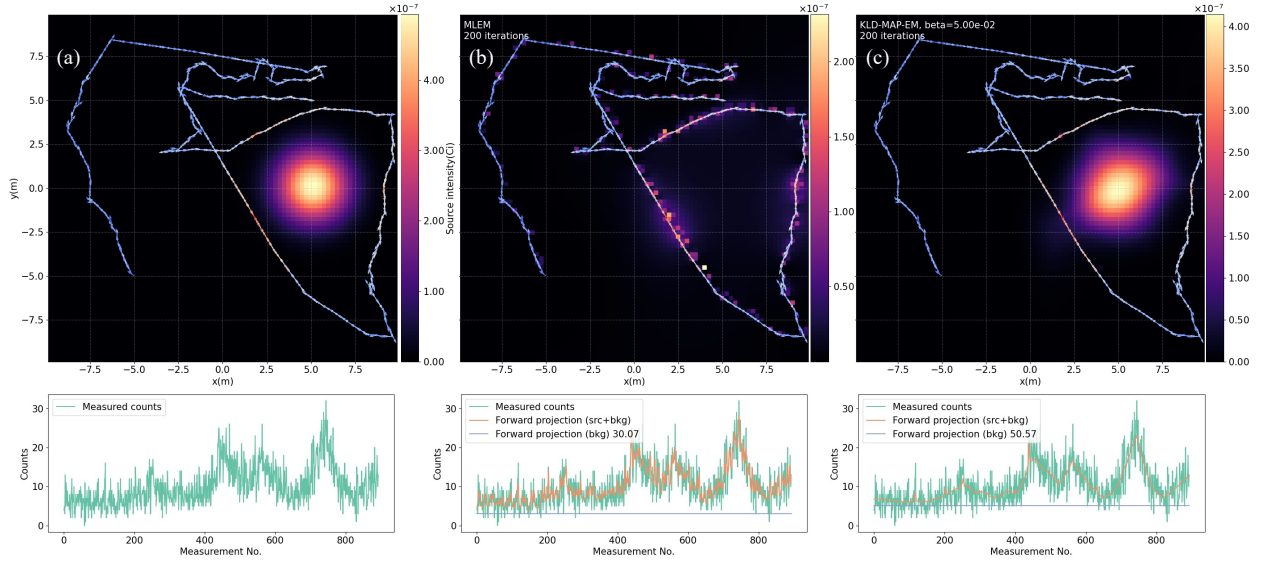


Figure 3.4: Measurement path and measurement/forward projection of (a) the ground truth (b) ML-EM reconstruction and (c) KLD MAP-EM reconstruction. Note that the ML-EM completely fails to map the distribution, exhibiting reconstruction biased towards the measurement path. Also, the forward projection of the ML-EM reconstruction indicates severe overfitting. The KLD MAP-EM reconstruction finds a good distribution estimate, along with the more regularized forward projection.

Real-world measurements

We present the KLD MAP-EM reconstruction results obtained from two real-world measurement scenarios. In addition to the reconstruction results, these scenarios highlight the challenges of image reconstruction from real-world data.

^{137}Cs Measurement with a CLLBC Scintillator Mounted Drone

In this real-world measurement scenario, a 1.8 mCi ^{137}Cs point source was placed within a house, and a CLLBC ($\text{Cs}_2\text{LiLa}(\text{Br},\text{Cl})_6:\text{Ce}$) scintillation detector based radiation imager (NG-LAMP [66]) was mounted on a drone and flown. Figure 3.5 illustrates the measurement

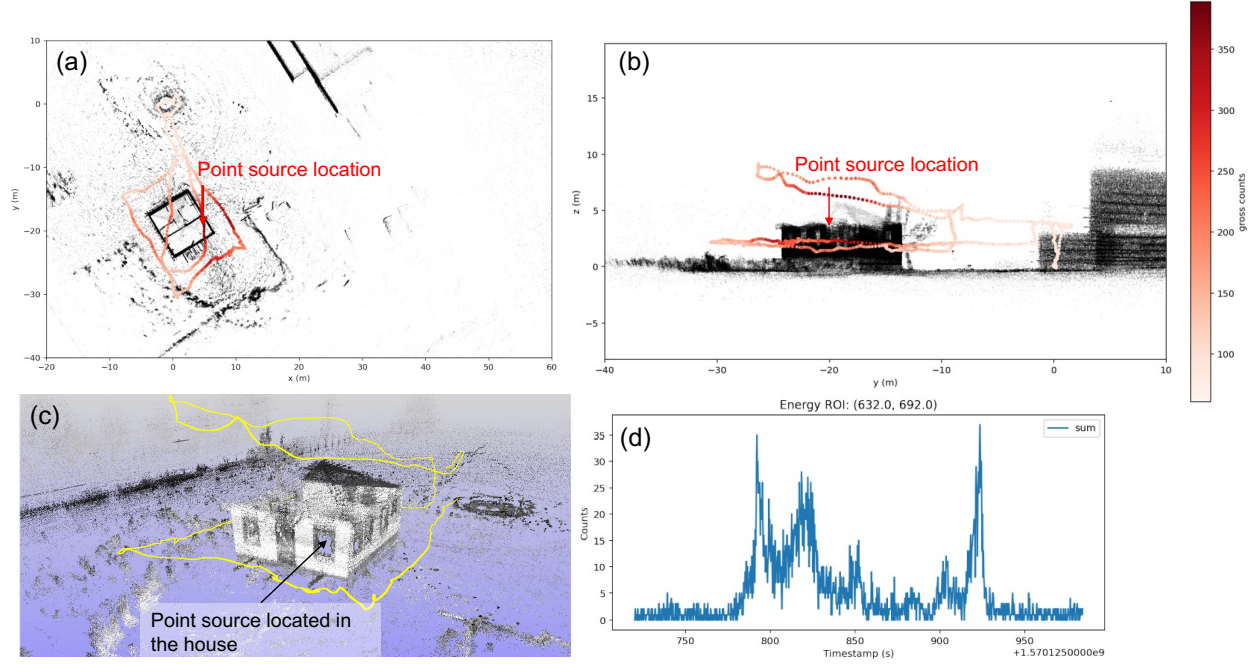


Figure 3.5: The measurement path and point cloud shown in (a) a top-down view, (b) a side view, and (c) a 3D perspective, illustrating the ^{137}Cs point source location within the house. (d) A time-series plot of counts within the energy region of interest (632.0-692.0 keV).

scenario with the LiDAR point cloud and the 4-minute drone measurement path. To improve the signal-to-background ratio (SBR) of the measurement, gamma-ray counts only within the ^{137}Cs photopeak energy window (662.0 ± 30 keV) was used for image reconstruction.

The imaging space was first constructed as a full 3D voxel space, then those voxels not containing point clouds were removed, leaving a sparse 3D voxel space. This approach is typically well-justified, as radioactive source is typically distributed on surfaces, and therefore removing the non-surface voxels considerably reduces computational cost of image reconstruction. However, the sparse voxel imaging space presents an additional challenge in image reconstruction; LiDAR point cloud was not created within the house and consequently, there were no image voxels near the ^{137}Cs source location.

Figure 3.6 shows the image reconstruction results from the ML-EM and the KLD MAP-EM algorithms. 200 iterations were performed for both algorithms, and $\beta = 5 \times 10^{-2}$ was used for the KLD MAP-EM. As observed in the simulated measurement scenarios, the ML-EM reconstruction fails to localize the source, with the reconstructed activities concentrated around the measurement paths. On the other hand, the KLD MAP-EM reconstruction exhibits significantly improved reconstruction, with the source activities localized in those voxels near the source. Note that since there was no voxels inside the house, those voxels on

the house wall and the roof were used to reconstruct the source.

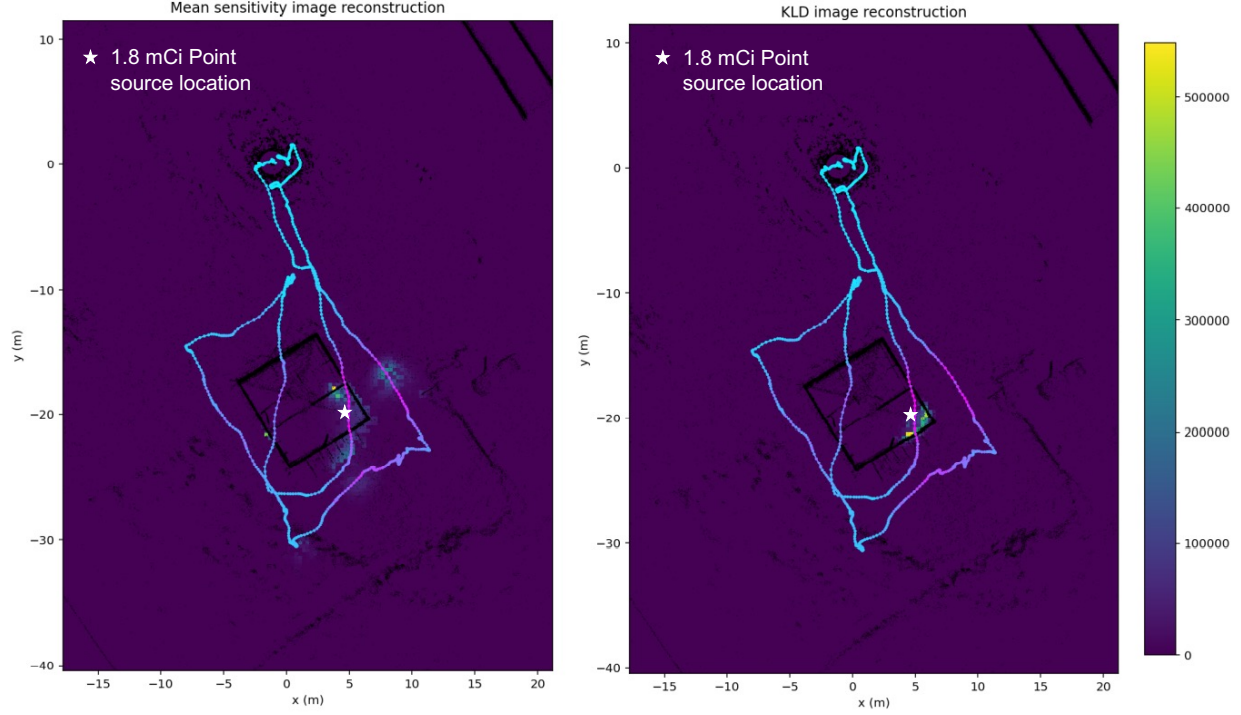


Figure 3.6: Images reconstructed using the (a) ML-EM and (b) KLD MAP-EM. 200 iterations were used for both reconstructions.

From the KLD MAP-EM reconstructed image, the estimated source activity was 0.82 mCi, which is significantly lower than the true source activity of 1.8 mCi. The discrepancy in source activity comes from two main reasons. First, the gamma-ray attenuation through the house wall was not accounted for and therefore, reconstructed source activity was underestimated. Secondly, those voxels on the surface of the house used for reconstruction are closer to the measurement path than the true source location. Note that the sensitivity is significantly higher in those voxels near the path. Consequently, the activity is further underestimated in reconstruction. These inaccurate modeling assumptions lead to an erroneous system matrix, which in turn results in degradation in reconstructed image quality.

^{239}Pu -surrogate Measurement with a Small Drone Composed of Structural Plastic Scintillators

In this measurement scenario, a 1 mCi ^{239}Pu -surrogate point source (a mixture of ^{133}Ba , ^{137}Cs , and ^{252}Cf sources to mimic a 1 mCi ^{239}Pu source) was placed in an open field, and a drone was flown for measurement. Interestingly, the detection medium was the drone

itself—some of its structural components were replaced with plastic scintillators, making it capable of detecting gamma rays and neutrons simultaneously [67]. Figure 3.7 illustrate the measurement setup. The source was placed 12.2 meters west of the launch position, and the drone was flown in a raster pattern, spanning around 23×23 meters with the constant altitude of 1.5 meters above the ground. The location of the drone was tracked using real-time kinematic positioning (RTK) based global positioning system (GPS).

Figure 3.8 shows the measurement count data and the image reconstruction results from both the ML-EM and the KLD MAP-EM algorithm. As observed in the measurement scenarios presented before, KLD MAP-EM shows significantly improved localization capability over the ML-EM. Notably, the radiation hotspot estimated from the algorithm is about 1 meter off to the south. This discrepancy can be better understood by examining the time-series measured counts plot. First of all, measurement time bin width was set at 1 second, which is too long to capture the variability in count rate near the source. Also, considering that the drone was moving at about 2.5 meters per second speed, 1 second measurement frequency leads to ~ 1 m error in source location estimation. More importantly, only < 10 cps was recorded when the drone was closest to the source, which is extremely low and comparable to the background countrate. Furthermore, the measurement count data contains non-Poisson noise, which is apparent with the count fluctuations at the measurement time within 10-20 s and 130-150 s ranges. These non-Poisson noise likely to have originated from electronics coupled to the plastic scintillators, and cannot be easily modeled into the system matrix.



Figure 3.7: Setup for the 1 mCi ^{239}Pu -surrogate point source measurement. The source was placed on a 5-gallon bucket in an open field. A plastic scintillator integrated drone was flown to measure both the neutrons and gamma rays from the source.

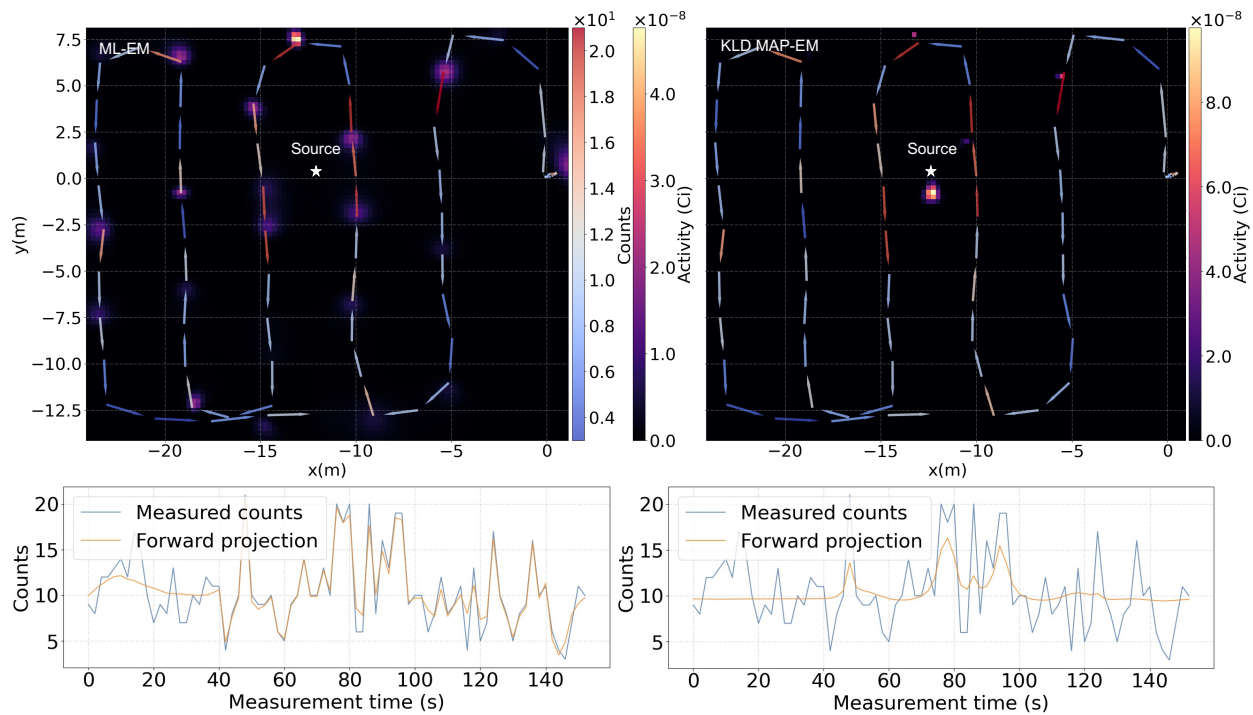


Figure 3.8: Top: Image reconstruction results using the ML-EM (left) and the KLD MAP-EM algorithm (right). Bottom: Measured time-series radiation count data and the forward projection from the reconstructed images. Note that the SNR in the measurement is very low. Also, it is apparent that the measurement is corrupted with non-Poisson noise.

3.4 Discussion

In this section, we discuss the strengths and weakness of the developed KLD MAP-EM algorithm, its applicability beyond single detector imaging scenarios, and potential future research directions.

Multiple radiation source reconstruction

As explained in section 3.2, the intuition behind the KL divergence based regularization is to prioritize those voxels that can “well-explain” the measurement data. This intuition is mostly valid when the source distribution is concentrated in a single radiation hot spot. However, when multiple scattered sources are in a single imaging space, the measurement data is a combination of contributions from these multiple sources, and therefore, it cannot be “explained” by a single hotspot, leading to potential degradation in reconstructed image quality.

To demonstrate this potential pitfall, the KLD MAP-EM algorithm was used on two simulated measurement scenarios and the results are shown in Fig. 3.9. In the first example scenario, two point sources were simulated, and while the reconstruction from the KLD MAP-EM shows some artifacts around source 2, overall the localization is satisfactory. Note that the artifact mostly originates from the fact that the measurement is a linear combination of the forward projections from both source 1 and 2.

The second measurement scenario adds an additional point source in the imaging domain, as seen in 3.9. Notably, with the addition of source 3, the measurement becomes more complex, and it cannot be easily explained by any of the point sources alone. As a result, the reconstructed point sources are more pulled towards the measurement path.

Extending to other imaging modalities and applications

Although the KL divergence regularization scheme introduced in this work was tested only on single detector imaging scenarios, its formulation is general enough to be broadly applicable to most radiation imaging modalities. Specifically, its utility is maximized in situations where a sparse source distribution with low SNR in measurement is expected. Examples include free-moving 3D Compton imaging or coded aperture, very low activity radioactive molecular imaging, and radiation monitoring system with stationary detectors. Moreover, the KLD MAP-EM can be trivially modified to accommodate measurement data with non-Poisson noise, such as Gaussian noise model.

Hyperparameter selection

One of the biggest challenges in using MAP-EM reconstruction methods, including the KLD MAP-EM, is selecting a good set of hyperparameters. For example, the KLD MAP-EM algorithm has two hyperparameters to be determined prior to image reconstruction: number

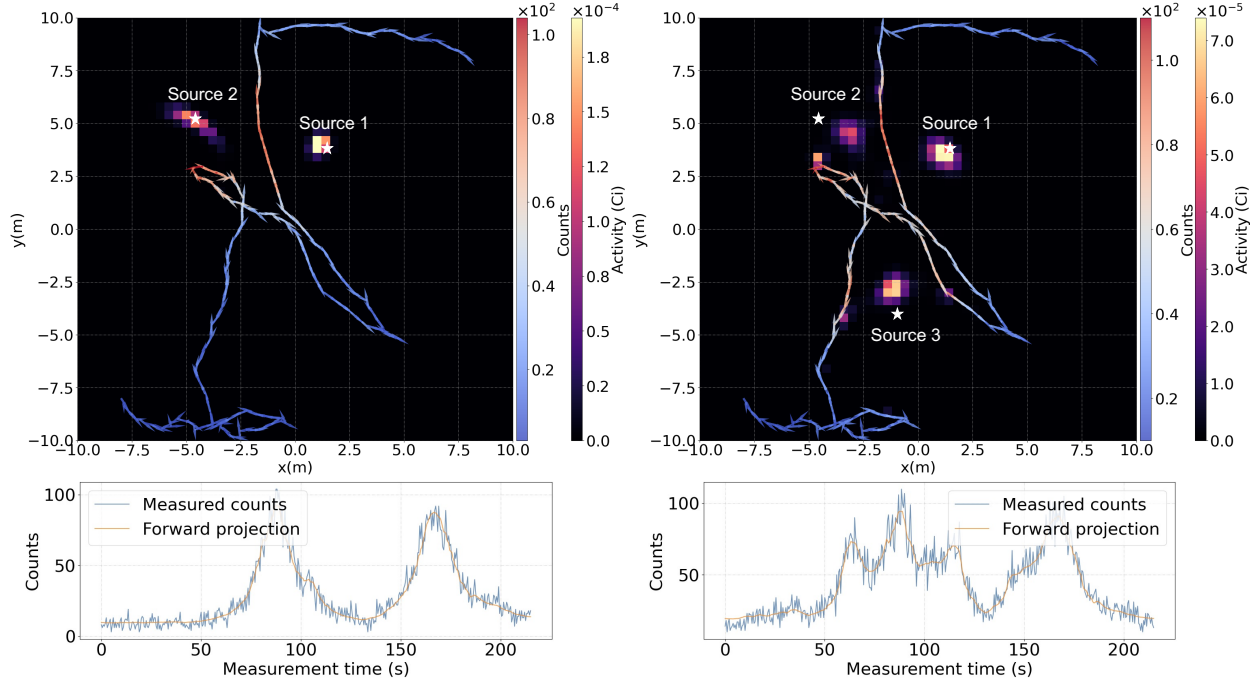


Figure 3.9: Left: Image reconstruction results the KLD MAP-EM when two point sources are present in the imaging space. Right: Image reconstruction results the KLD MAP-EM when three point sources are present in the imaging space. As we add more point sources, the measurements cannot be accounted for by a single hotspot, leading to erroneous reconstruction results.

of iterations and regularization strength β . Generally speaking, with increasing number of iterations and β parameter, reconstructed images gradually become more “refined” features (e.g., clear edges, etc). However, too many iterations with low β may cause overfitting, and too few iterations with high β might lead to underfitting. Hence, this tradeoff must be taken into account when the proposed algorithm is used. In this work, all image reconstructions were performed with 200 iterations and $\beta = 5 \times 10^{-2}$, indicating that some heuristics may be useful. However, the choice of hyperparameters may vary depending on specific applications.

A good practice for hyperparameter selection is to incorporate *a priori* application-specific domain knowledge. For example, in many imaging scenarios, point source configuration can often be safely assumed. In such cases, it is instructive to larger β value to promote more sparsity in reconstruction. Another effective strategy is to compare measurement data $\mathbf{y}$ and the forward projection from the reconstructed image $\bar{\mathbf{y}}$. Given that the measurement data follows Poisson distribution, a hypothesis testing can be performed to assess consistency between determine $\mathbf{y}$ and $\bar{\mathbf{y}}$. Then, iteratively adjusting the hyperparameters ensures an effective set of hyperparameters.

3.5 Conclusions

In this chapter, a KL divergence based image reconstruction technique, KLD MAP-EM was developed and rigorously tested across a variety of imaging scenarios. The KL divergence quantifies the likelihood that the measurement data originates from each individual image voxel. This metric was utilized to design a sparsity-inducing prior, which was integrated into regular MLE image reconstruction problems. Consequently, we proposed the KLD MAP-EM algorithm, which solves the regularized image reconstruction problem.

The performance of the KLD MAP-EM algorithm was demonstrated through simulations and real-world single detector imaging scenarios. The results clearly indicate that while the conventional ML-EM algorithm suffers from pathological overfitting, the proposed algorithm effectively localizes and maps the source distribution while inducing sparsity in the reconstructed images.

Future work could explore its extension to more complex radiation imaging modalities. Additionally, further investigations into optimization of hyperparameter will benefit the robustness of the algorithm. The promising results presented in this chapter suggest that the proposed method holds great potential for applications requiring sparsity in reconstructed images.

Chapter 4

Radiation Image Reconstruction and Uncertainty Quantification Using a Gaussian Process Prior

The KLD based regularization scheme introduced in chapter 3 produces regularized solution images by promoting sparsity. While the method is highly effective when sparsity in the solution image is expected, it may underperform in imaging scenarios where sources are widely distributed. In this chapter, we describe a complete framework for Bayesian image reconstruction and uncertainty quantification based on a Gaussian process prior (GPP) to overcome limitations of the maximum likelihood expectation maximization (ML-EM) and the KLD MAP-EM image reconstruction algorithms. The prior distribution is constructed with a zero-mean Gaussian process (GP) with a choice of a covariance function, and a link function is used to map the Gaussian process to an image. Unlike many other maximum *a posteriori* (MAP) approaches, our method offers highly interpretable hyperparameters that are selected automatically with the empirical Bayes method. Furthermore, the GP covariance function can be modified to incorporate *a priori* structural priors, enabling multi-modality imaging or contextual data fusion. Lastly, we illustrate that our approach lends itself to Bayesian uncertainty quantification techniques, such as the preconditioned Crank-Nicolson method (pCN) and the Laplace approximation.

The proposed framework is general and can be employed in most radiation image reconstruction problems, and we demonstrate it with simulated free-moving single detector radiation source imaging scenarios. We compare the reconstruction results from GPP and ML-EM, and show that the proposed method can significantly improve the image quality over ML-EM, all the while providing greater understanding of the source distribution via the uncertainty quantification capability. Furthermore, significant improvement of the image quality by incorporating a structural prior is illustrated.

4.1 Introduction

The Bayesian image reconstruction paradigm often faces several challenges. First, the hyperparameters (i.e., regularization parameters) reflecting *a priori* knowledge are often difficult to choose in a principled manner [68]. Secondly, even with carefully selected hyperparameters, the prior model itself may not be flexible enough to model the *a priori* knowledge that needs to be incorporated. Lastly, Bayesian uncertainty quantification of reconstructed images require very high dimensional integration, and analytical solutions are generally unavailable unless conjugate priors are employed [69].

To tackle the aforementioned challenges, we propose the use of a Gaussian process (GP)-based prior for image reconstruction and uncertainty quantification tasks. A GP is a stochastic process where the constituent random variables jointly follow a multivariate Gaussian distribution [70]. The GPs can be made very flexible with a choice of mean and covariance functions, and they have been extensively studied in the context of regression tasks [71]. In this work, we demonstrate that GPs can also serve as a powerful prior for Bayesian image reconstruction tasks. The proposed approach to use a GP prior will be referred to as Gaussian process prior (GPP) henceforth. The following are the contributions of our work:

- First, we demonstrate that the MAP estimate with the proposed GPP can significantly enhance the quality of the reconstructed radiation images compared to the conventional MLE-based approach. The computational cost in using the GP is greatly eased by exploiting the Kronecker product structure in the GP covariance matrix.
- We then show that uncertainties in the reconstructed images can be quantified using the preconditioned Crank-Nicolson (pCN) algorithm [72] and the Laplace approximation [73][74].
- We illustrate that the flexibility of the GPP allows for incorporation of structural prior, which can dramatically improve the reconstruction in some imaging scenarios.
- Lastly, we lay out a principled method for choosing the hyperparameters of the GPP, using the empirical Bayes method and Laplace approximation.

We note that a Bayesian approach using a GP prior for image reconstruction and uncertainty quantification was recently introduced by Zhou et al.[75], where a combination of a TV and a GP prior and the pCN-based uncertainty quantification method were proposed. However, due to the non-differentiable nature of the TV prior, it limits the use of gradient-based optimization and Hessian information. In contrast, in this work, we demonstrate that the Laplace approximation, which leverages the Hessian of the posterior, enables scalable, yet effective uncertainty quantification and hyperparameter selection.

The proposed framework is general and can be used for most radiation imaging modalities with Poisson likelihood and (with a small modification) Gaussian likelihood. In this work, we show the effectiveness of the proposed method with free-moving single detector imaging scenarios, which is based on the recently developed scene data fusion (SDF) concept

[4][25][26]. We chose this imaging modality for two reasons: firstly, due to the insufficient angular modulation from a single detector and vast imaging space, the imaging problem is highly ill-posed. As a result, the solutions found from the ML-EM algorithm are often undesirable [30], highlighting the benefits of the GPP image reconstruction and uncertainty quantification methods. Secondly, there has been growing interest in the free-moving imaging modalities due to its significant impact in nuclear safety and security applications [29]; however, the concept is often challenged by the absence of reliable image reconstruction methods. Hence the proposed approach can provide accurate reconstruction of radiation fields for a wide range of applications, thereby addressing the current needs of the research field.

4.2 Methods

Image Reconstruction with GPP

GPP MAP image reconstruction

The goal of radiation imaging is to recover an N -dimensional image vector $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$ from an M -dimensional non-negative measurement vector $\mathbf{y} \in \mathbb{Z}_{\geq 0}^M$. The two vectors are related by a system matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{M \times N}$ such that

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}), \quad \text{where} \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}. \quad (4.1)$$

Hence the system matrix contains the probabilities of radiation emitted from a voxel being detected in a measurement bin. With the Poisson measurement model, the likelihood function of an image vector $\mathbf{x}$ given the measurement vector $\mathbf{y}$ can be formulated as

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^M \frac{e^{-\bar{y}_i} \bar{y}_i^{y_i}}{y_i!}. \quad (4.2)$$

The well-established maximum likelihood (ML) image reconstruction approach seeks a solution $\hat{\mathbf{x}}_{\text{ML}}$, which minimizes the negative log-likelihood function $l(\mathbf{x}; \mathbf{y}) = -\log p(\mathbf{y}|\mathbf{x})$ up to a constant as follows:

$$\hat{\mathbf{x}}_{\text{ML}} = \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} l(\mathbf{x}; \mathbf{y}) \quad (4.3)$$

$$\stackrel{c}{=} \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} \sum_{i=1}^M \bar{y}_i - y_i \log \bar{y}_i \quad (4.4)$$

$$= \underset{\mathbf{x}, \mathbf{x} \geq 0}{\operatorname{argmin}} \sum_{i=1}^M [\mathbf{A}\mathbf{x}]_i - y_i \log [\mathbf{A}\mathbf{x}]_i, \quad (4.5)$$

where $\mathbf{x} \succeq 0$ imposes the non-negativity constraints and $\stackrel{c}{=}$ denotes equivalence up to a constant. It is well-known that the above optimization problem can be iteratively solved using the expectation maximization (EM) algorithm, leading to the popular maximum likelihood-expectation maximization (ML-EM) algorithm [47].

Although ML-EM has become a standard approach for image reconstruction tasks, for many radiation imaging problems the system matrix $\mathbf{A}$ is often not well-conditioned. Hence, the optimization problem (4.3) is ill-posed and the ML-EM algorithm is susceptible to over-fitting.

A popular approach to combat the ill-posedness is to incorporate a prior distribution $p(\mathbf{x})$ to make use of Bayesian formalism. Using Bayes' rule [76], the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is given by

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (4.6)$$

In Bayesian image reconstruction, often the image vector is recovered by locating the mode of the posterior distribution, leading to the Maximum *a Posteriori* (MAP) estimation as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}, \mathbf{x} \succeq 0}{\operatorname{argmin}} l(\mathbf{x}; \mathbf{y}) + l_p(\mathbf{x}), \quad (4.7)$$

where $l_p(\mathbf{x}) = -\log p(\mathbf{x})$ is the negative log prior distribution. An appropriate choice of a prior distribution is crucial for reliable image reconstruction, especially when the system matrix $\mathbf{A}$ is ill-conditioned.

In this work, for prior model construction, we assume that the image vector $\mathbf{x}$ is governed by a latent variable ξ , which follows a zero-mean Gaussian process (GP),

$$\xi \sim \mathcal{GP}(\mathbf{0}, k(\mathbf{r}, \mathbf{r}')), \quad (4.8)$$

where $k(\mathbf{r}, \mathbf{r}')$ is a covariance kernel, and $\mathbf{r}, \mathbf{r}' \in \mathbb{R}^3$ represents vectors containing the Cartesian coordinates of two points in the imaging space.

The choice of a covariance kernel $k(\mathbf{r}, \mathbf{r}')$ is crucial as it dictates the underlying spatial correlation of the image. In this work, we use the unit-variance squared exponential (SE) kernel to encode spatial correlation between pixels as follows:

$$k(\mathbf{r}, \mathbf{r}') = \exp \left(-\frac{\|\mathbf{r} - \mathbf{r}'\|^2}{2\sigma^2} \right), \quad (4.9)$$

where σ is the characteristic correlation length scale parameter of the prior distribution with the dimension of distance. It should be noted that various covariance kernels have been explored in the context of GPs, and the optimal covariance kernel may vary depending on applications [77].

For practical purposes, the imaging space is discretized with pixels and the GP in (4.8) is evaluated at the pixel centers. Then, the GP reduces to a zero-mean multivariate Gaussian distribution,

$$\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad (4.10)$$

$$\text{where } \Sigma_{ij} = k(\mathbf{r}_i, \mathbf{r}_j). \quad (4.11)$$

Hence,

$$p(\boldsymbol{\xi}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} \boldsymbol{\xi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi} \right). \quad (4.12)$$

Once the prior $p(\boldsymbol{\xi})$ is defined, a link function $f : \mathbb{R}_{\geq 0}^N \rightarrow \mathbb{R}^N$ is used to relate $\boldsymbol{\xi}$ and $\mathbf{x}$, keeping the positivity constraints imposed on $\mathbf{x}$. There are a variety of choices for the link function; for example, a simple logarithmic function, $f(x_i) = \log x_i = \xi_i$ could be used as a link function satisfying the non-negativity constraint on $\mathbf{x}$. However, for simpler statistical interpretation, we use the exponential-to-Gaussian link function [78], and its inverse can be used to recover $\mathbf{x}$ as follows:

$$\mathbf{x} = \mathbf{f}^{-1}(\boldsymbol{\xi}) = -\frac{1}{\lambda} \log \left(\frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\boldsymbol{\xi} \oslash (\sqrt{2} \operatorname{diag}(\boldsymbol{\Sigma})) \right) \right), \quad (4.13)$$

where λ is the scale parameter for the exponential distribution and erf is the error function. $\oslash$ denotes the element-wise division operator. The link function converts a zero-mean Gaussian distributed random variable ξ_i with the variance Σ_{ii} into a random variable following the exponential distribution with the scale parameter λ . As the reciprocal of λ defines the mean value of the exponential distributed random variable, it can be used to incorporate *a priori* information about the average pixel value, allowing for more statistically interpretable hyperparameter selection. Furthermore, due to the long-tailed nature of the exponential distribution, the link function can effectively model radiation source strength that varies over several orders of magnitudes. Samples from the prior distribution $p(\mathbf{x})$ with different hyperparameters σ and λ are shown in Fig. 4.1.

It should be emphasized that the appropriate choice of a link function may vary depending on the application, and the methods presented in this work are applicable with various link functions.

From (4.12), the negative log-prior for $\boldsymbol{\xi}$ can be written as follows:

$$\begin{aligned} l_p(\boldsymbol{\xi}) &= -\log p(\boldsymbol{\xi}) \\ &= \frac{1}{2} \log \sqrt{(2\pi)^n |\boldsymbol{\Sigma}|} + \frac{1}{2} \boldsymbol{\xi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi} \end{aligned} \quad (4.14)$$

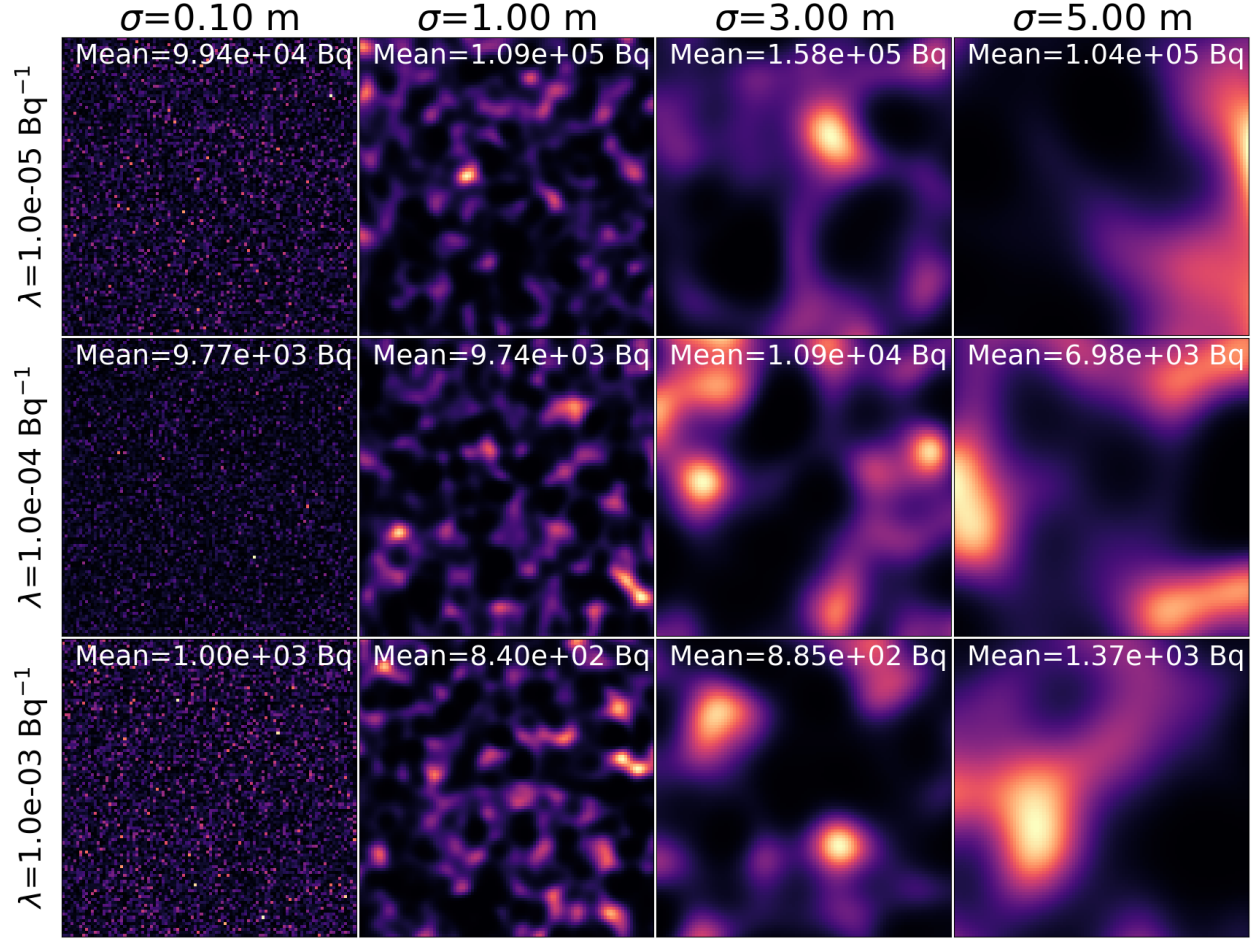


Figure 4.1: Random samples from the prior distribution $p(\mathbf{x})$ on a 20×20 m 2-D imaging space with different sets of hyperparameters, σ and λ . Each sample was realized by first drawing a sample from $\xi_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then transforming it by $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{L}\xi_w)$. The properties and characteristics of the prior distribution can be better understood by examining random samples from the prior. Note that the σ parameter defines the characteristic correlation length, while the reciprocal of λ parameter determines the mean pixel values.

Combining (4.13) and (4.14), the MAP solution in terms of ξ can be obtained by minimizing the unnormalized negative log posterior $\Psi(\xi) \triangleq l(\xi; \mathbf{y}) + l_p(\xi)$ as follows:

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \Psi(\xi) \quad (4.15)$$

$$= \underset{\xi}{\operatorname{argmin}} \sum_i^N [\mathbf{A}\mathbf{f}^{-1}(\xi) - \mathbf{y} \odot \log(\mathbf{A}\mathbf{f}^{-1}(\xi))]_i + \frac{1}{2}\xi^\top \Sigma^{-1}\xi. \quad (4.16)$$

The optimization problem (4.16) can be solved in a more efficient and numerically stable manner by introducing an uncorrelated multivariate unit Gaussian variable $\boldsymbol{\xi}_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Hence, we use a change of variable $\mathbf{L}\boldsymbol{\xi}_w = \boldsymbol{\xi}$, where $\mathbf{L}$ is the Cholesky factor of $\boldsymbol{\Sigma}$. The final optimization problem in terms of $\boldsymbol{\xi}_w$ is

$$\hat{\boldsymbol{\xi}}_w = \underset{\boldsymbol{\xi}_w}{\operatorname{argmin}} \Psi(\mathbf{L}\boldsymbol{\xi}_w) \quad (4.17)$$

$$= \underset{\boldsymbol{\xi}_w}{\operatorname{argmin}} \sum_i^N [\mathbf{A}f^{-1}(\mathbf{L}\boldsymbol{\xi}_w) - \mathbf{y} \odot \log(\mathbf{A}f^{-1}(\mathbf{L}\boldsymbol{\xi}_w))]_i + \frac{1}{2}\boldsymbol{\xi}_w^\top \boldsymbol{\xi}_w, \quad (4.18)$$

where $\odot$ is the element-wise product operation. Note that (4.17) is an unconstrained optimization problem, and it can be solved with a general purpose global optimization algorithms. We find that quasi-Newton algorithms, such as the Limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [79] achieves swift convergence. The gradient of (4.17), which is required for many optimization algorithms, can be analytically found as follows:

$$\begin{aligned} \nabla_{\boldsymbol{\xi}_w} \Psi(\mathbf{L}\boldsymbol{\xi}_w) = \mathbf{L}^\top & \left((\mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top (\mathbf{y} \oslash (\mathbf{A}\mathbf{x}))) \odot \left(\mathbf{1} \oslash \left(\sqrt{2\pi}\lambda \operatorname{diag}(\boldsymbol{\Sigma}) \right) \right. \right. \\ & \left. \left. \odot \exp(\lambda \mathbf{x} - \boldsymbol{\xi}^{\circ 2} \oslash (2 \operatorname{diag}(\boldsymbol{\Sigma}))) \right) \right) + \boldsymbol{\xi}_w, \end{aligned} \quad (4.19)$$

where $\circ$ is the element-wise power operations. Also, the $\operatorname{diag}(\cdot)$ operator extracts the main diagonal elements of a matrix into a vector. The derivation of the gradient expression (4.19) can be found in the supplementary information. Once the optimization problem is solved and $\hat{\boldsymbol{\xi}}_w$ is obtained, the image $\hat{\mathbf{x}}$ can be obtained by $\hat{\mathbf{x}} = \mathbf{f}^{-1}(\mathbf{L}\hat{\boldsymbol{\xi}}_w)$. The pseudocode for the GPP MAP estimation is shown in Algorithm 1.

Kronecker Product Representation of the Covariance Matrix

With a large number of voxels N , setting up and solving (4.17) is a burdensome task due mainly to the computational costs involved in the Cholesky factor $\mathbf{L} \in \mathbb{R}^{N \times N}$. For example, storing and computing the Cholesky factor requires $\mathcal{O}(N^2)$ memory and $\mathcal{O}(N^3)$ time, which can be prohibitively expensive for a large N .

Computation of the Cholesky decomposition is also the main computational bottleneck in the GP regression problems. To overcome the challenge, in [80] Saatci *et al.* showed that when the GP regression input variables lie on a multi-dimensional Cartesian grid, the Kronecker structure in the covariance matrix $\boldsymbol{\Sigma}$ can be exploited to significantly lessen the computational costs involving the Cholesky factor $\mathbf{L}$. As many imaging problems employ Cartesian grid imaging spaces, here the strategy can be readily extended to the GPP image reconstruction.

When a covariance kernel function is the product of kernels across each Cartesian input dimension, such that $k(\mathbf{r}_i, \mathbf{r}_j) = \prod_{d=1}^D k(r_i^d, r_j^d)$, the covariance matrix $\boldsymbol{\Sigma}$ constructed for a

Algorithm 1 GPP MAP estimation

Require: $\mathbf{A}$, $\mathbf{y}$, σ , λ , $\boldsymbol{\xi}_w^{\text{init}}$, ϵ (Convergence criterion)

$\boldsymbol{\Sigma} \leftarrow (4.11)$ ▷ Prior covariance matrix construction

$\mathbf{L} \leftarrow \text{Cholesky}(\boldsymbol{\Sigma})$ ▷ Cholesky decomposition

$k \leftarrow 0$

$\boldsymbol{\xi}_w^k \leftarrow \boldsymbol{\xi}_w^{\text{init}}$ ▷ Initialization

$\Psi^k \leftarrow \Psi(\mathbf{L}\boldsymbol{\xi}_w^k)$ ▷ Initialization

Convergence $\leftarrow$ False

while Convergence is False **do**

$\boldsymbol{\xi}_w^{k+1} \leftarrow \text{L-BFGS}(\boldsymbol{\xi}_w^k, (4.17), (4.19))$ ▷ L-BFGS iteration

$\Psi^{k+1} \leftarrow \Psi(\mathbf{L}\boldsymbol{\xi}_w^{k+1})$

if $|\Psi^k - \Psi^{k+1}| < \epsilon$ **then**

 Convergence $\leftarrow$ True

else

$k \leftarrow k + 1$

end if

end while

$\hat{\boldsymbol{\xi}}_w \leftarrow \boldsymbol{\xi}_w^{k+1}$

$\hat{\boldsymbol{\xi}} \leftarrow \mathbf{L}\hat{\boldsymbol{\xi}}_w$

$\hat{\mathbf{x}} \leftarrow (4.13)$ ▷ Reconstructed image

$\hat{\Psi} \leftarrow \Psi^{k+1}$ ▷ Negative log-posterior at the mode

multi-dimensional lattice imaging space can be expressed as a Kronecker product of much smaller covariance matrices for each input dimension $\boldsymbol{\Sigma}_d$ (i.e., x -, y - and z -dimensions), such that $\boldsymbol{\Sigma} = \bigotimes_{d=1}^D \boldsymbol{\Sigma}_d$. Furthermore, the Cholesky factor of a such covariance matrix is also a Kronecker product of Cholesky factor in each dimension, $\mathbf{L} = \bigotimes_{d=1}^D \mathbf{L}_d$. Hence, assuming the same number of voxels in each dimension, the memory and time complexity of the Cholesky factor computation are eased with $\mathcal{O}(DN^{2/D})$ and $\mathcal{O}(DN^{3/D})$.

Even with the obtained Kronecker product of the Cholesky factor, a naive matrix-vector multiplication $\mathbf{L}\boldsymbol{\xi}_w$ in (4.17) and (4.19) limits the scalability of the algorithm. Surprisingly, multiplication of a Kronecker product matrix on a vector can be cast as a tensor outer product and performed in $\mathcal{O}(DN^{\frac{1+D}{D}})$ time [80][81].

Equipped with the Kronecker product covariance matrix and its Cholesky factor, the overall time complexity of computing the cost function (4.17) and its gradient (4.19) is determined by the forward projection $\mathbf{A}f^{-1}(\mathbf{L}\boldsymbol{\xi}_w)$, which is computed in $\mathcal{O}(MN)$. Note that the ML-EM has the same complexity of $\mathcal{O}(MN)$ per iteration.

Incorporation of Structural Prior

With the growing interest in contextual data fusion and multi-modality radiation imaging [82][83][84], we discuss here the incorporation of *a priori* structural information using the GPP. Such *a priori* structural information can be obtained from other radiation imaging modalities (e.g., MRI, CT), commercial sensors (e.g., LiDAR, visual camera), or object detection algorithms.

We consider an imaging space $\mathcal{X}$ with N voxels $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ and further, L voxel clusters $C_l \subseteq \mathcal{X}$ for $l = 1, 2, \dots, L$, each belonging to different objects or structures in the imaging space $\mathcal{X}$. This structural information can be incorporated by modifying the covariance matrix of the Gaussian prior Σ (4.9). The modified covariance matrix is given by:

$$\Sigma_{ij} = \begin{cases} k(\mathbf{r}_i, \mathbf{r}_j), & \text{if } x_i, x_j \in C_l \quad \forall l = 1, 2, \dots, L \\ 0, & \text{otherwise.} \end{cases} \quad (4.20)$$

In other words, spatial correlation is defined only within the same object C_l . Note that different hyperparameters σ and λ can be applied to each voxel cluster C_l . The benefit of incorporating a structural prior is that it introduces correlations within salient features in the imaging space, leading to improved image quality and interpretability of the reconstruction results.

Note that, with the use of structural prior, the covariance matrix cannot be represented as a Kronecker product; however, by assigning zeroes for the covariances between different voxel clusters, the covariance matrix becomes a sparse block diagonal matrix, whose Cholesky decomposition can be computed and stored efficiently.

Bayesian Uncertainty Quantification

The use of a Gaussian process prior in this work furnishes the basis for further Bayesian inference such as Bayesian uncertainty quantification. In Bayesian statistics, the uncertainty in the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is often expressed in terms of a Bayesian credible interval. For a marginalized posterior variable x_j , the $(1-\alpha)100\%$ Bayesian credible interval is defined as the range $[x_{j-}, x_{j+}]$ within which x_j falls with a probability of $1-\alpha$. In other words, $p(x_j \in [x_{j-}, x_{j+}]|\mathbf{y}) = 1-\alpha$. The choice of such interval is not unique; therefore, we use an equal-tailed interval to define x_{j-} and x_{j+} as follows:

$$\int_{-\infty}^{x_{j-}} p(x_j|\mathbf{y}) dx_j = \frac{\alpha}{2} \quad (4.21)$$

$$\int_{x_{j+}}^{\infty} p(x_j|\mathbf{y}) dx_j = \frac{\alpha}{2}. \quad (4.22)$$

However, computation of the credible intervals is a daunting task, as it requires the integration of a very high dimensional distribution, $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) d\mathbf{x}$ (i.e., the marginal probability). Therefore, in this work, we investigate two different posterior approximation

methods for credible interval estimation.

Preconditioned Crank–Nicolson (pCN) Algorithm

In Bayesian statistics, Markov chain Monte Carlo (MCMC) algorithms have been widely used to generate a sequence of samples from an analytically intractable posterior distribution. However, for high dimensional posterior distributions (e.g., images), the MCMC sample convergence rate becomes very slow, requiring too many samples for reasonable posterior approximation [85].

In our approach, the use of GP prior in (4.17) lends itself to the use of the pCN algorithm [72], a dimension-robust MCMC algorithm specifically designed for posterior distributions with a Gaussian prior. The pCN convergence rate is independent of discretization of a parameter space, which is an attractive feature for high-dimensional inverse problems such as image reconstruction.

With the pCN algorithm, a chain of J samples $\{\boldsymbol{\xi}_n\}_{n=1}^J$ is obtained by following the proposal and acceptance/rejection steps. First, the pCN proposal in terms of $\boldsymbol{\xi}$ is given by,

$$\boldsymbol{\xi}'_{n+1} = \sqrt{1 - \beta^2} \boldsymbol{\xi}_n + \beta \boldsymbol{\Xi}_{n+1}, \quad (4.23)$$

where $\boldsymbol{\Xi}_{n+1}$ is a sample from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and the hyperparameter $0 < \beta < 1$ is a step size which is chosen manually. With the obtained $\boldsymbol{\xi}'_{n+1}$, the acceptance probability is computed as

$$\alpha(\boldsymbol{\xi}_n, \boldsymbol{\xi}'_{n+1}) = \min(1, \exp(l(\boldsymbol{\xi}_n; \mathbf{y}) - l(\boldsymbol{\xi}'_{n+1}; \mathbf{y}))). \quad (4.24)$$

Subsequently, the acceptance probability $\alpha(\boldsymbol{\xi}_n, \boldsymbol{\xi}'_{n+1})$ is compared to a random number Z_{n+1} drawn from a uniform distribution $Z_{n+1} \sim \text{Unif}([0, 1])$ and sets

$$\boldsymbol{\xi}_{n+1} = \begin{cases} \boldsymbol{\xi}'_{n+1}, & \text{if } Z_{n+1} < \alpha(\boldsymbol{\xi}_n, \boldsymbol{\xi}'_{n+1}) \\ \boldsymbol{\xi}_n, & \text{otherwise.} \end{cases} \quad (4.25)$$

In our experiments, the MCMC sample chains were started from the reconstructed image (i.e., the mode of the posterior). Once a sufficient number of samples are obtained, the pixelwise $(1 - \alpha)100\%$ confidence intervals can be found as follows. First, compute the pixelwise $(\alpha/2)100\%$ and $(1 - \alpha/2)100\%$ quantiles $\boldsymbol{\xi}_-$ and $\boldsymbol{\xi}_+$ from the obtained samples $\{\boldsymbol{\xi}_n\}_{n=1}^J$. Then, the quantiles in $\mathbf{x}$ are obtain as $\mathbf{x}_- = \mathbf{f}^{-1}(\boldsymbol{\xi}_-)$ and $\mathbf{x}_+ = \mathbf{f}^{-1}(\boldsymbol{\xi}_+)$. As the inverse link function $\mathbf{f}^{-1}$ is monotonically increasing, it ensures that the quantiles of $\mathbf{x}$ can be directly found by transforming the quantiles of $\boldsymbol{\xi}$. Lastly, the pixelwise $(1 - \alpha)100\%$ credible interval map $\mathbf{x}_{\text{CI}}$ can be found as $\mathbf{x}_{\text{CI}} = \mathbf{x}_+ - \mathbf{x}_-$.

Collecting S samples, the time complexity of the pCN method is $\mathcal{O}(MNS)$. Although each sample can be taken efficiently, ensuring convergence of the MCMC chain requires a large S , on the order of 10^5 to 10^6 samples. In the next section, we explore an alternative approach to posterior approximation, which typically runs much faster than the pCN method.

Laplace Approximation

Another posterior estimation method, the Laplace approximation, uses a multivariate Gaussian distribution with the mean located at the MAP estimate $\hat{\xi}$ found from (4.17) to directly approximate a posterior distribution.

Using second-order Taylor expansion of the unnormalized log posterior $-\Psi(\xi)$ at the mode $\hat{\xi}$, the posterior distribution can be approximated with $q(\xi|\mathbf{y})$ as follows:

$$p(\xi|\mathbf{y}) \approx q(\xi|\mathbf{y}), \quad \text{where} \quad (4.26)$$

$$q(\xi|\mathbf{y}) \sim \mathcal{N}\left(\hat{\xi}, \left(\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}\right)^{-1} = \left(\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} + \Sigma^{-1}\right)^{-1}\right), \quad (4.27)$$

A more detailed derivation as well as the expression for $\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}$ is presented in the supplementary information. Once the approximate posterior distribution $q(\xi|\mathbf{y})$ is obtained, the pixelwise $(\alpha/2)100\%$ and $(1 - \alpha/2)100\%$ quantiles ξ_- and ξ_+ are found directly from the Gaussian distribution quantile functions. Then, the $(1 - \alpha)100\%$ pixelwise credible interval map $\mathbf{x}_{\text{CI}}$ can be obtained as $\mathbf{x}_{\text{CI}} = \mathbf{f}^{-1}(\xi_+) - \mathbf{f}^{-1}(\xi_-)$. Note that the matrix $\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}}$ has the dimension of $N \times N$. With moderate number of voxles (i.e., $N \sim 10^4$), the inversion in (4.27) can be directly performed. However, with a large number of voxels, instantiating the covariance matrix and computing the inversion in (4.27) can be computationally very expensive.

In order for more efficient covariance matrix computation, the Hessian of the log-likelihood can be approximated with the Fisher information matrix at the mode $\mathbf{F}(\hat{\xi})$,

$$\nabla_{\xi}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \approx \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[\left(\nabla_{\xi} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right) \left(\nabla_{\xi} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right)^{\top} \right] \quad (4.28)$$

$$= (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}})^{\top} \mathbf{A}^{\top} \text{diag}((\mathbf{A} \mathbf{f}^{-1}(\xi))^{\circ -1}) \mathbf{A} (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}}) \quad (4.29)$$

$$= \mathbf{B} \mathbf{B}^{\top}, \quad (4.30)$$

where we define $\mathbf{B} \equiv (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)|_{\xi=\hat{\xi}})^{\top} \mathbf{A}^{\top} \text{diag}((\mathbf{A} \mathbf{f}^{-1}(\xi))^{\circ -\frac{1}{2}})$. The detailed derivation of (4.29) and the expression for $\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)$ is presented in the accompanying supplementary information. Fisher information has been studied extensively for neural network Hessian approximation, and it is well-justified as it is the expected Hessian of log-likelihood under $p(\mathbf{y}|\bar{\mathbf{y}})$. Equipped with the Hessian of the log-likelihood approximation, the expression for the covariance matrix in (4.27) becomes

$$\left(\nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}} \right)^{-1} \approx (\mathbf{B} \mathbf{B}^{\top} + \Sigma^{-1})^{-1} \quad (4.31)$$

$$= \mathbf{L} (\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L} + \mathbf{I})^{-1} \mathbf{L}^{\top}. \quad (4.32)$$

Next, we perform rank- K truncated eigendecomposition on $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L}$, such that $\mathbf{L}^{\top} \mathbf{B} \mathbf{B}^{\top} \mathbf{L} \approx \mathbf{Q} \mathbf{S} \mathbf{Q}^{\top}$ where $\mathbf{Q} \in \mathbb{R}^{N \times K}$ is a matrix containing eigenvectors and $\mathbf{S} \in \mathbb{R}^{K \times K}$ is a diagonal

matrix whose elements contain K largest eigenvalues of $\mathbf{L}^\top \mathbf{B} \mathbf{B}^\top \mathbf{L}$. Such decomposition can be efficiently obtained from randomized eigendecomposition algorithms. Randomized decomposition methods first proceed by forming a matrix $\mathbf{Y} \in \mathbb{R}^{N \times K}$ by multiplying a random matrix $\mathbf{P} \in \mathbb{R}^{N \times K}$ such that

$$\mathbf{Y} = (\mathbf{L}^\top \mathbf{B} (\mathbf{B}^\top \mathbf{L} \mathbf{P})). \quad (4.33)$$

The matrix $\mathbf{Y}$ captures most of the column space in $\mathbf{L}^\top \mathbf{B} \mathbf{B}^\top \mathbf{L}$ yet considerably smaller with $K \ll \min(M, N)$. Therefore, it can be further processed to efficiently estimate the eigendecomposition of $\mathbf{L}^\top \mathbf{B} \mathbf{B}^\top \mathbf{L}$. More details on the randomized eigendecomposition methods can be found in Halko *et al.* [86]. When the system matrix $\mathbf{A}$ is computed on-the-fly or too large to be loaded into core memory, the decomposition can be obtained efficiently with only single-pass of $\mathbf{A}$ using the matrix sketch algorithm [87]. Also, note that by computing the matrix-matrix multiplication in the order specified by the parentheses in (4.33), we never instantiate an $N \times N$ matrix, which is potentially too large to be stored in memory.

Once the eigendecomposition is obtained, the covariance matrix can be computed as follows:

$$\left(\nabla_{\xi}^2 \Psi(\xi) |_{\xi=\hat{\xi}} \right)^{-1} \approx \mathbf{L} (\mathbf{Q} \mathbf{S} \mathbf{Q}^\top + \mathbf{I})^{-1} \mathbf{L}^\top \quad (4.34)$$

$$= \mathbf{L} (\mathbf{I} - \mathbf{Q} (\mathbf{S}^{-1} + \mathbf{Q}^\top \mathbf{Q})^{-1} \mathbf{Q}^\top) \mathbf{L}^\top \quad (4.35)$$

$$= \mathbf{L} (\mathbf{I} - \mathbf{Q} (\mathbf{S} \oslash (\mathbf{S} + \mathbf{I})) \mathbf{Q}^\top) \mathbf{L}^\top \quad (4.36)$$

$$= \mathbf{L} \mathbf{L}^\top - \mathbf{L} \mathbf{Q} (\mathbf{S} \oslash (\mathbf{S} + \mathbf{I})) \mathbf{Q}^\top \mathbf{L}^\top \quad (4.37)$$

The second equality follows from the Woodbury matrix identity [88]. The resulting expression (4.37) for the covariance matrix can be partially computed to obtain necessary covariance elements for uncertainty quantification. For example, to create a pixel-wise uncertainty map, only the diagonal elements can be computed without forming an $N \times N$ matrix.

Due to the unimodality of the approximating Gaussian distribution, the Laplace approximation is well-justified when the posterior density is concentrated around the mode of the posterior, $\hat{\xi}$. The so-called ‘‘Bayesian central limit theorem’’ (i.e., Bernstein-von Mises theorem) states that with increasing number of conditionally independent measurements, posterior distributions tend toward Gaussian distributions under certain conditions [89]. Hence, with a sufficient number of measurements, the Laplace approximation provides fast, yet reasonably accurate approximation of the posterior distribution.

Hyperparameter Selection with Empirical Bayes

Although the hyperparameters of the GPP offers good physical and statistical interpretability, a principled method to infer the hyperparameters from the measurement is highly desirable. Here we show that the hyperparameters σ and λ can be systematically determined with the empirical Bayes method.

Including the hyperparameter vector $\boldsymbol{\theta} = [\sigma, \lambda]^\top$, the Bayes rule (4.6) can be re-written in terms of $\boldsymbol{\xi}$ as

$$p(\boldsymbol{\xi}|\mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\theta})p(\boldsymbol{\xi}|\boldsymbol{\theta})}{p(\mathbf{y}|\boldsymbol{\theta})}, \quad \text{where} \quad (4.38)$$

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\theta})p(\boldsymbol{\xi}|\boldsymbol{\theta}) d\boldsymbol{\xi}. \quad (4.39)$$

The empirical Bayes approach finds the hyperparameters $\hat{\boldsymbol{\theta}}$ maximizing the marginal likelihood as follows:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}, \boldsymbol{\theta} \succ 0}{\operatorname{argmin}} -\log p(\mathbf{y}|\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}, \boldsymbol{\theta} \succ 0}{\operatorname{argmin}} \Phi(\boldsymbol{\theta}), \quad (4.40)$$

where $\Phi(\boldsymbol{\theta}) \stackrel{c}{=} -\log p(\mathbf{y}|\boldsymbol{\theta})$ is the unnormalized log-marginal likelihood. However, as the computation of $\Phi(\boldsymbol{\theta})$ is intractable due to the high dimensional integration in (4.38), we use the Laplace approximation to estimate $\Phi(\boldsymbol{\theta})$ [90]. With the Laplace approximation, the unnormalized negative log-marginal likelihood $\Phi(\boldsymbol{\theta})$ at the posterior mode $\hat{\boldsymbol{\xi}}$ is given as

$$\Phi(\boldsymbol{\theta}) \approx \frac{1}{2} \log |\mathbf{I} + \nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \boldsymbol{\Sigma} + \Psi(\hat{\boldsymbol{\xi}}). \quad (4.41)$$

The derivation can be found in the supplementary information. Note that in (4.41), the $\Psi(\hat{\boldsymbol{\xi}})$ term is an output of Algorithm 1. The computation of the log-determinant term has the complexity of $\mathcal{O}(N^3)$, which prevents scalable computation of the log-marginal likelihood when N is large. Therefore, we exploit the low-rank structure of the matrix $\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \boldsymbol{\Sigma}$ to accelerate the computation of the log determinant.

As in (4.29), using the Fisher information matrix approximation to $\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$, the determinant can be approximated as,

$$\Phi(\boldsymbol{\theta}) \approx \frac{1}{2} \log |\mathbf{I} + \mathbf{B}\mathbf{B}^\top \boldsymbol{\Sigma}| + \Psi(\hat{\boldsymbol{\xi}}). \quad (4.42)$$

We also observe that using the precomputed Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top$, the matrix determinant lemma enables the following reformulation of the log-determinant term,

$$\frac{1}{2} \log |\mathbf{I} + \mathbf{B}\mathbf{B}^\top \boldsymbol{\Sigma}| = \frac{1}{2} \log |\mathbf{I} + \mathbf{B}^\top \mathbf{L}\mathbf{L}^\top \mathbf{B}| \quad (4.43)$$

$$= \frac{1}{2} \log |\mathbf{I} + \mathbf{L}^\top \mathbf{B}\mathbf{B}^\top \mathbf{L}|. \quad (4.44)$$

The Kronecker structure of the Cholesky factor $\mathbf{L}$ allows for very efficient matrix-matrix multiplication $\mathbf{B}^\top \mathbf{L}$ or $\mathbf{L}^\top \mathbf{B}$ in (4.43) and (4.44). Furthermore, depending on the dimensions of the system matrix $\mathbf{A}$, (4.43) (when $M \ll N$) or (4.44) (when $N \ll M$) can be used to compute the log determinant efficiently.

Algorithm 2 Hyperparameter selection with empirical Bayes

Require: $\mathbf{A}, \mathbf{y}, \boldsymbol{\xi}_w^{\text{init}}, \boldsymbol{\theta}^{\text{init}}, \epsilon_1, \epsilon_2$
 $k \leftarrow 0$
 $\hat{\boldsymbol{\xi}}_w^k \leftarrow \boldsymbol{\xi}_w^{\text{init}}$
 $\boldsymbol{\theta}^k \leftarrow \boldsymbol{\theta}^{\text{init}}$ ▷ Initialization
 Convergence $\leftarrow$ False
while Convergence is False **do**
 $\hat{\boldsymbol{\xi}}_w^{k+1}, \hat{\Psi}^k \leftarrow \text{Algorithm 1}(\mathbf{A}, \mathbf{y}, \boldsymbol{\theta}^k, \hat{\boldsymbol{\xi}}_w^k, \epsilon_1)$
 $\Phi(\boldsymbol{\theta}^k) \leftarrow (4.42)$ ▷ (4.43), (4.44), or (4.45) for the determinant
 $\log \boldsymbol{\theta}^{k+1} \leftarrow \text{L-BFGS}(\log \boldsymbol{\theta}^k, \Phi(\boldsymbol{\theta}^k))$
▷ L-BFGS iteration, logarithm for the positivity of the hyperparameters
if $|\Phi(\boldsymbol{\theta}^k) - \Phi(\boldsymbol{\theta}^{k+1})| < \epsilon_2$ **then**
 Convergence $\leftarrow$ True
else
 $k \leftarrow k + 1$
end if
end while
 $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}^{k+1}$ ▷ Selected hyperparameters

When both the M and N are too large, a low rank approximation of the system matrix $\mathbf{A} \approx \mathbf{U}_K \mathbf{V}_K^\top$ where $\mathbf{U}_K \in \mathbb{R}^{M \times K}$ and $\mathbf{V}_K \in \mathbb{R}^{N \times K}$ with $K \ll M, N$ can be used to accelerate the computation. Defining $\mathbf{C} = \mathbf{V}_K^\top \mathbf{J}_\xi \mathbf{f}^{-1}(\boldsymbol{\xi})|_{\xi=\boldsymbol{\xi}} \mathbf{L}$ and $\mathbf{D} = \mathbf{U}_K^\top \text{diag}(\mathbf{1} \oslash \bar{\mathbf{y}}^{\circ \frac{1}{2}})$, the log-determinant term can be approximated as,

$$\frac{1}{2} \log |\mathbf{I} + \mathbf{W}\boldsymbol{\Sigma}| \approx \frac{1}{2} \log |\mathbf{I} + \mathbf{C}\mathbf{C}^\top \mathbf{D}\mathbf{D}^\top|. \quad (4.45)$$

Hence using the rank- K approximation, the log-determinant computation has $\mathcal{O}(K^3)$ time complexity. The low-rank approximation of $\mathbf{A}$ needs to be computed only once using various techniques, and we found that randomized singular value decomposition (rSVD) provides a good approximation with the time complexity of $\mathcal{O}(MN \log K + (M + N)K^2)$ [86]. Such approximation of $\mathbf{A}$ may sacrifice some accuracy in image reconstruction; however, we found that the approximation has negligible impact for hyperparameter selection and once the hyperparameters are chosen, the full-rank system matrix can be used for more accurate image reconstruction. Additionally, the low rank approximation can also reduce the computation of forward projection $\mathbf{A}\mathbf{f}^{-1}(\boldsymbol{\xi})$, which helps to accelerate the overall optimization process.

Once an adequate log-marginal likelihood computation scheme is selected, the L-BFGS can be used to solve the optimization problem (4.40). The necessary gradient of $\Phi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ can be found with finite difference methods. The entire marginal likelihood maximization scheme is illustrated in Algorithm 2.

4.3 Results

In this section, we demonstrate the effectiveness of the proposed GPP MAP image reconstruction and uncertainty quantification techniques using simulated free-moving single detector imaging scenarios. Furthermore, we present two applications of the GPP algorithms: first, quantitative Compton imaging for safeguard applications and second, minimum radiation dose path planning.

GPP Image Reconstruction and Uncertainty Quantification on Simulated Free-moving Single Detector Measurements Scenarios

Each free-moving measurement scenario was simulated with a 10-cm diameter spherical detector (i.e., isotropic response) in a 20 m×20 m 2-D imaging space. The detector intrinsic efficiency was assumed to be 10% and the detector moving paths were simulated on the x-y plane. In all measurement scenarios, the z-coordinates of the pixel centers and detector paths were fixed at 0 m and 0.5 m, respectively. Along the detector paths, recorded counts were binned with the integration time of 0.1 s.

Anisotropic Gaussian Source

First, a measurement scenario with an anisotropic Gaussian source was simulated to demonstrate the effectiveness of the proposed algorithm. Gaussian source distributions are of particular interest because radioactive material dispersion often exhibits Gaussian-like distributional features [91].

Fig. 4.2 (a) illustrates the simulated 1-mCi anisotropic Gaussian distributed source positioned at the center of the imaging space and the 150-m path of the moving detector. Assuming the average human walking speed of 1.5 m/s, the total measurement time amounted to 98 s.

Fig. 4.2 (b) and Fig. 4.2 (c) show the image reconstruction results of the ML-EM after 200 iterations and the proposed GPP algorithm. As can be evidently seen, the ML-EM reconstructed image exhibits high frequency features clustered around the path, and we observed such patterns even after 20 iterations. On the other hand, the image reconstructed with the GPP MAP successfully localizes and maps the Gaussian distribution.

The hyperparameters of the GPP algorithm ($\sigma = 4.1$ m and $\lambda = 1.4 \times 10^{-4}$ Bq⁻¹) were chosen with Algorithm 2. Fig. 4.3 (a) shows the contour plot of the negative log-marginal likelihood as a function of σ and λ , together with the optimal hyperparameters selected from Algorithm 2. To investigate the impact of hyperparameter choices on image reconstruction, Fig. 4.4 shows the GPP MAP reconstructed images with various hyperparameters. As illustrated, the algorithm performs well over a large range of hyperparameter choices.

Fig. 4.2 (d) and (g) show the pixel-wise 90% credible interval maps from the pCN and Laplace approximation, respectively. For the pCN MCMC, 10^6 samples were collected and the pixel-wise credible intervals were computed. In both cases, there exists about 10%

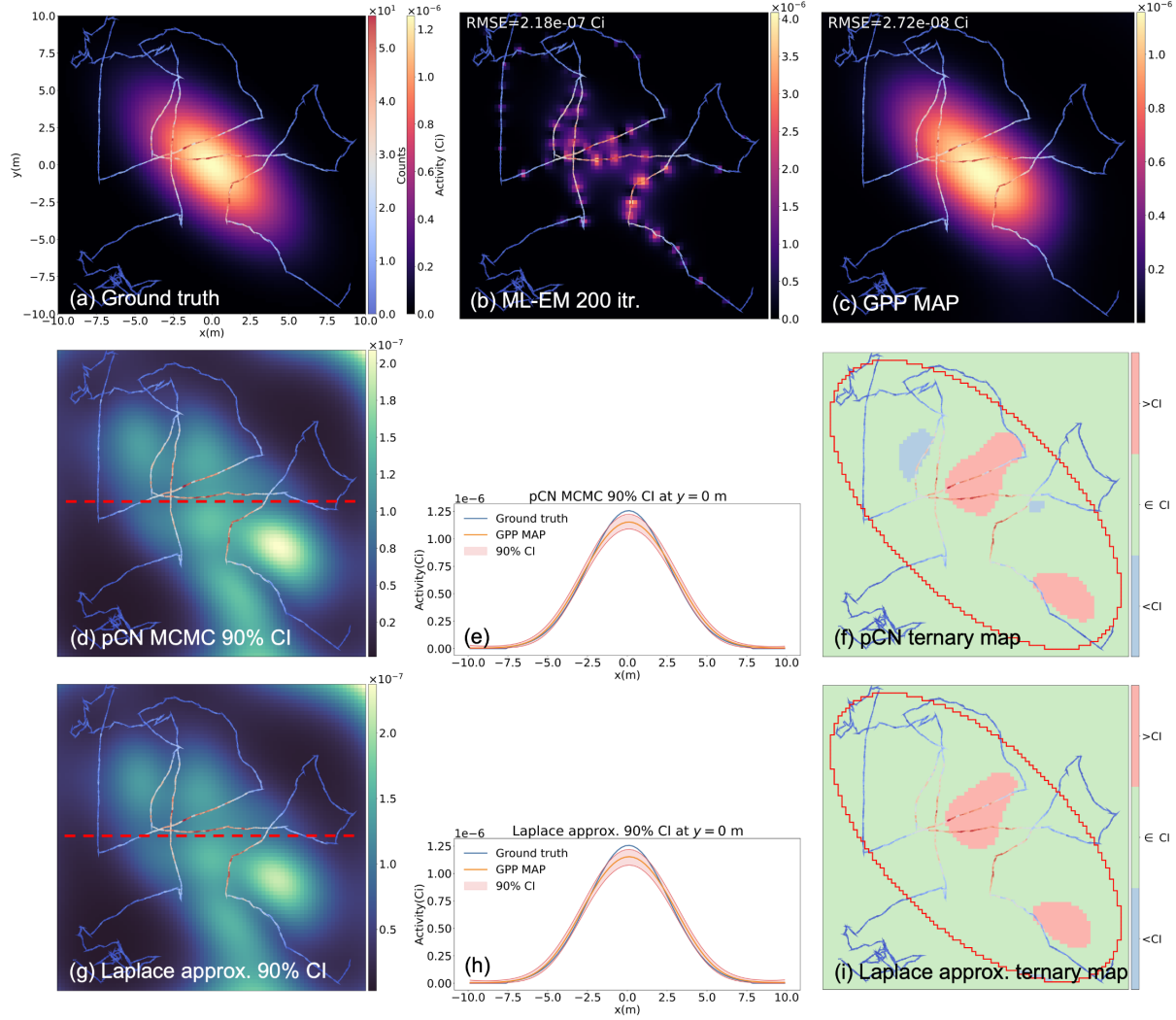


Figure 4.2: (a) A simulated measurement scenario with a 1-mCi anisotropic Gaussian distributed source. The human-walked path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters $\sigma=4.1$ m and $\lambda=1.4 \times 10^{-4}$ Bq $^{-1}$ found from the empirical Bayes method were used for reconstruction. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y=0$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

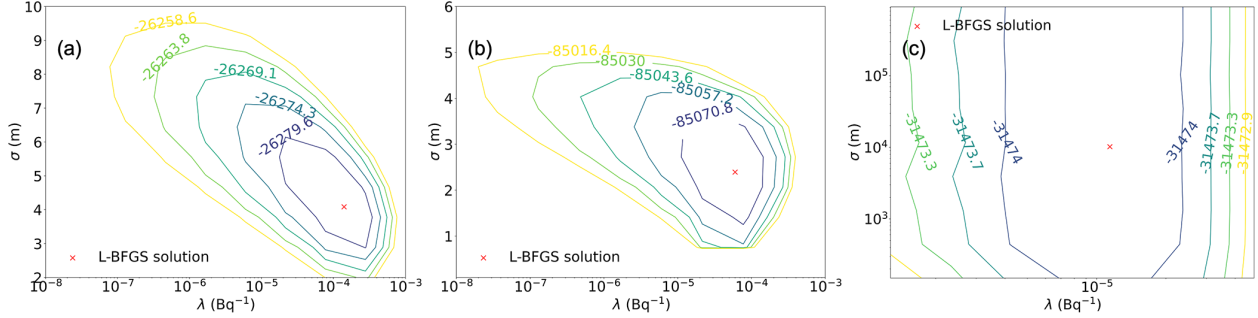


Figure 4.3: Contour plots of the negative log-marginal likelihood as a function of the hyperparameters σ and λ . The hyperparameters selected from Algorithm 2 are marked with red Xs. (a) The anisotropic Gaussian source measurement scenario. (b) The ring and Gaussian source measurement scenario. (c) The uniform square source measurement scenario with a structural prior. When using the structural prior, 2 different sets of hyperparameters were used for in-square and out-of-square. Therefore, only the in-square hyperparameters were used to create the contour plot, while the out-of-square hyperparameters were fixed at the values found from Algorithm 2.

of uncertainty around the source region, due to the ambiguity in the reconstructed images. Interestingly, the uncertainty is lower in those pixels close to the path, reflecting the fact that the source activity can be inferred with more confidence in the pixels with large sensitivity. Conversely, the unexplored region in the upper right corner exhibits high uncertainties, because the measurement is not sensitive to the pixels in the region. Note that with the Laplace approximation, the uncertainties in the upper right corner are slightly overestimated. We suspect that the marginal pixel posteriors in the upper right corner are exponentially distributed due to the lack of measurements, (i.e., $\text{prior} \sim \text{posterior}$). Capturing the skewness of the exponential distribution is challenging with the Laplace approximation because of the Gaussian posterior assumption. However, from Fig. 4.2 (e) and (h), which depict the slices of the credible interval maps at $y = 0$ m, it can be observed that both uncertainty quantification methods produce very similar results at the center of the imaging space where the source is distributed.

Fig. 4.2 (f) and (i) display the ternary maps, where 3 different colors are used depending on the ground truth pixel values relative to the estimated credible intervals. The pixels where the ground truth values are bounded by the credible intervals are colored green, while red and blue indicate the ground truth values exceeding and falling below the credible intervals, respectively. As illustrated, in most pixels the ground truth is well bounded by the credible intervals estimated from both the pCN and Laplace approximation.

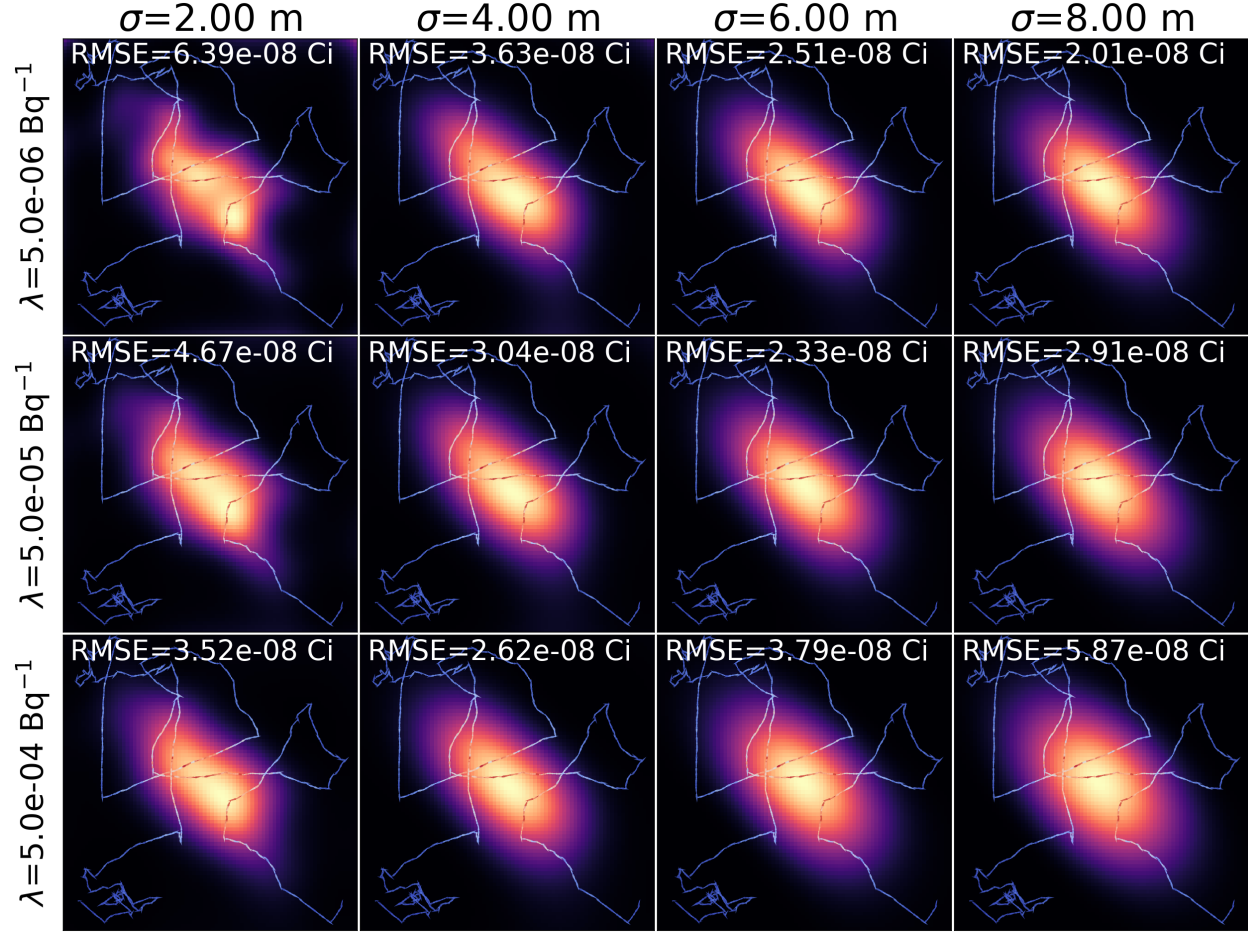


Figure 4.4: To better understand the impact of the hyperparameters, the anisotropic Gaussian source images reconstructed from the GPP MAP with different choices of the hyperparameters σ and λ are displayed. With larger σ and λ , the prior distribution is spatially more correlated and the pixel values are pushed towards zeros. As a result, the reconstructed image with large hyperparameters (bottom right) tend to spread out.

Ring and Gaussian Source

A more challenging measurement scenario involving a 2 mCi ring-shaped source and a 1 mCi anisotropic Gaussian source was simulated as depicted in Fig. 4.5 (a) along with the detector moving path. The total measurement duration was approximately 100 s. As can be seen in Fig. 4.5 (b) and (c), while the ML-EM produces source distribution concentrated around the measurement path, the GPP MAP estimate offers significantly more improved reconstruction results. The GPP hyperparameters were set to $\sigma = 2.4$ m and $\lambda = 6.0 \times 10^{-5}$ Bq⁻¹, which were selected with Algorithm 2. The contour plot of the negative marginal log-likelihood as

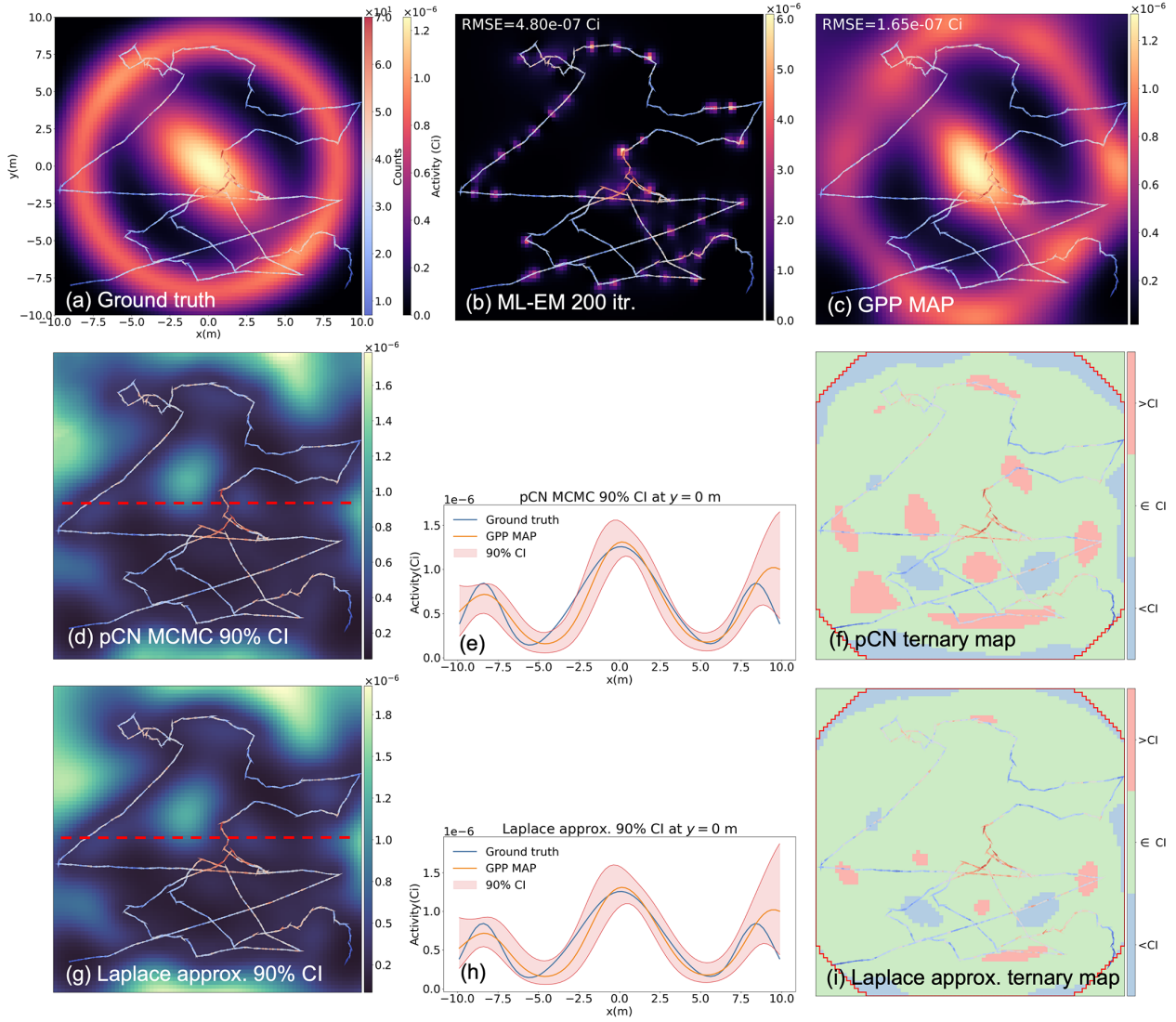


Figure 4.5: (a) A simulated measurement scenario with a 2-mCi ring-shaped source and a 1-mCi anisotropic Gaussian distributed source. The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the ML-EM algorithm after 200 iterations. (c) The reconstructed from the proposed GPP MAP. The hyperparameters $\sigma=2.4$ m and $\lambda=6.0\times 10^{-5}$ Bq $^{-1}$ found from the empirical Bayes method were used for reconstruction. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y=0$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

a function of the hyperparameters can be found in Fig. 4.3 (b).

Fig. 4.5 (d) and (g) present the results of the uncertainty quantification using the pCN and Laplace approximation, respectively. Similarly to the findings in section 4.3, the pixel uncertainties are predominantly influenced by the sensitivities of the pixels; uncertainties are lower around the detector path and elevated in those unvisited regions. Also, the Laplace approximation slightly overestimates the uncertainties, yet overall it remains largely comparable to the pCN as can be seen in Fig. 4.5 (e) and (h). In addition, the ternary maps Fig. 4.5 (f) and (i) demonstrate that the ground truths are mostly bounded by the credible intervals estimated from the both methods.

Uniform Square Source with a Structural Prior

We also demonstrate that incorporation of a structural prior using the GPP algorithm could significantly improve the image quality. Fig 4.6 (a) show another measurement scenario with a 1 mCi uniform square source locate at $(x, y) = (2.5 \text{ m}, 2.5 \text{ m})$. The total measurement duration was approximately 70 s. The reconstruction result using the GPP MAP is shown in Fig. 4.6 (b). Although the GPP MAP reconstruction successfully localized the source, the high frequency edges could not be recovered for two reasons: first, due to the lack of modulations in the single detector measurements, the forward projected measurements are blurred and noisy, leading to some ambiguity in the reconstructed images. Second, the GP prior with the SE kernel is not apt to model the high frequency edges of the ground truth distribution.

Here we show that the limitations can be overcome by incorporating *a priori* structural information. The prior covariances between the pixels within and outside the source square can be zeroed to enforce independence between the two pixel clusters. The reconstruction result with the structural prior is presented in Fig. 4.6 (c). As can be observed, the reconstruction successfully recovers the uniform ground truth distribution with the edges. Note that two different sets of hyperparameters were used for the two pixel clusters, namely the in-square hyperparameters and the out-of-square hyperparameters. Using Algorithm 2, the hyperparameters were set at $\sigma = 1.0 \times 10^4 \text{ m}$ and $\lambda = 1.1 \times 10^{-5} \text{ Bq}^{-1}$ for the in-square hyperparameters, and $\sigma = 9.8 \times 10^3 \text{ m}$ and $\lambda = 7.7 \times 10^4 \text{ Bq}^{-1}$ for the out-of-square hyperparameters. The very large σ hyperparameters for both pixel clusters indicate that a piecewise constant function with the discontinuities at the pixel cluster boundaries well-explain the measurement data. Also the large difference in the λ parameters between the pixel clusters reflects the fact that the source is only present in the in-square pixel clusters. Fig. 4.3 (c) shows a contour plot illustrating the negative log-marginal likelihood as a function of in-square hyperparameters.

The 90% credible interval maps found from the pCN and Laplace approximation are displayed in Fig. 4.6 (d) and (g), respectively. Note that due to the strong correlations between the pixel clusters, the marginal credible interval maps are also step-function like. Also, similarly to the findings from the results presented in Sections 4.3 and 4.3, a slight overestimation of uncertainties was observed with the Laplace approximation. However,

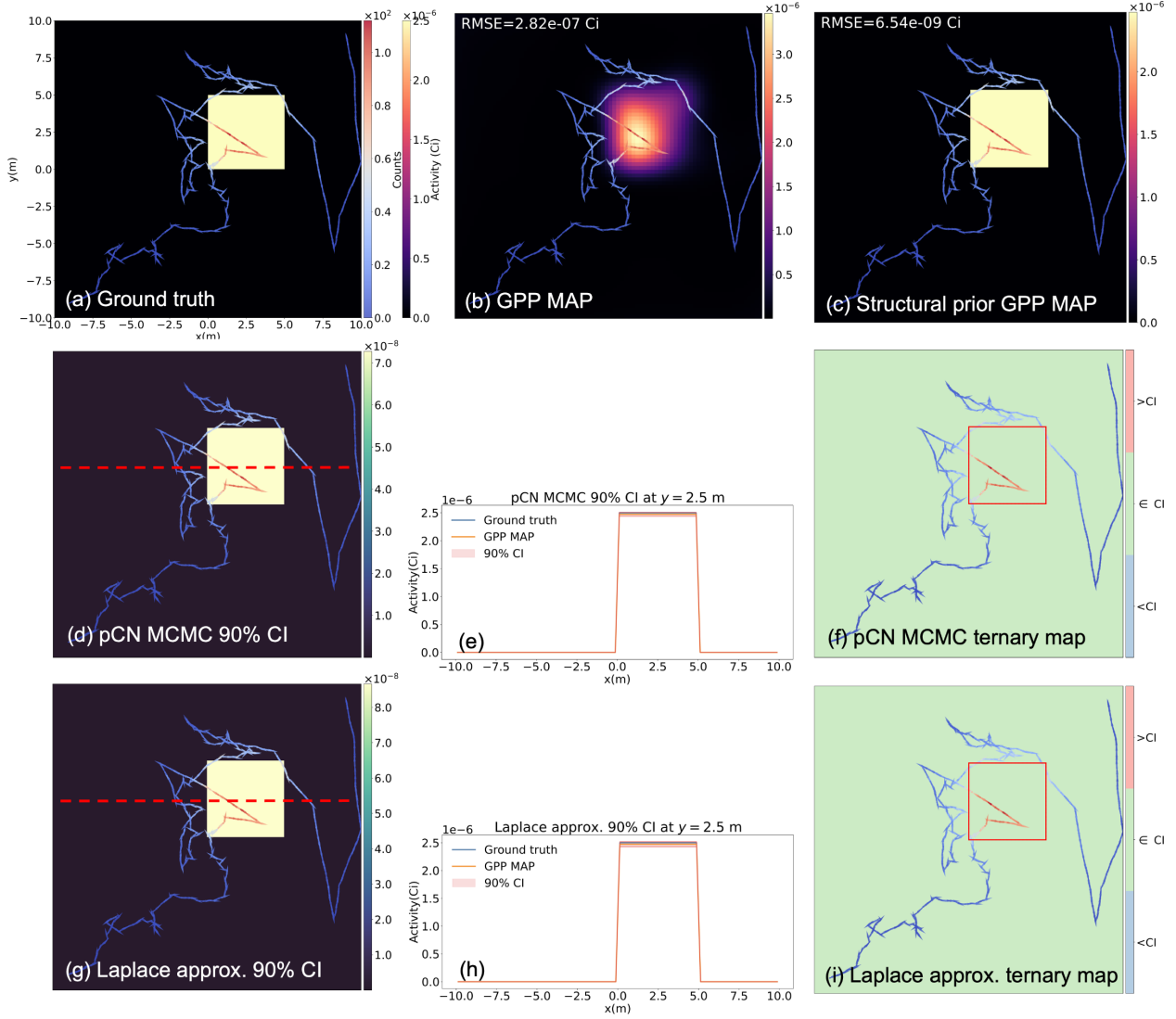


Figure 4.6: (a) A simulated measurement scenario with a 1-mCi uniform located at $(x, y) = (2.5\text{m}, 2.5\text{m})$. The detector moving path is marked with arrows and colored according to the measured counts. (b) The reconstructed image from the GPP MAP without incorporating the structural information. The hyperparameters found from Algorithm 2, $\sigma = 2.03$ m and $\lambda = 8.10$ Bq $^{-1}$ were used. (c) The reconstructed from the proposed GPP MAP with the structural prior. $\sigma = 1.0 \times 10^4$ m and $\lambda = 1.1 \times 10^{-5}$ Bq $^{-1}$ for the in-square hyperparameters, and $\sigma = 9.8 \times 10^3$ m and $\lambda = 7.7 \times 10^4$ Bq $^{-1}$ for the out-of-square hyperparameters. (d)(g) The pixelwise 90% credible interval maps from the pCN and the Laplace approximation, respectively. (e)(h) The credible intervals and the ground truth at $y = 2.5$ m (red dotted lines in (d) and (g)). (f)(i) Ternary maps in which the pixels are colored differently based on the ground truth pixel values relative to the estimated 90% credible intervals.

both methods produce comparable results as can be observed from Fig. 4.6 (e) and (h), where the slices of the uncertainty maps at $y = 2.5$ m are shown. Also, the ternary maps in Fig. 4.6 (f) and (i) indicate that the ground truth distribution is well-bounded by the estimated uncertainties.

Applications of the GPP Method

Quantitative Compton Imaging for Safeguard Applications

Here, we demonstrate that the GPP image reconstruction and uncertainty quantification methods are broadly applicable across various imaging modalities and data types. In particular, we apply the GPP method to quantitative Compton imaging in a 4π imaging space, highlighting its potential in nuclear safeguards applications.

As list-mode data is a common data type for Compton imaging, here we briefly outline how the GPP methods can be modified to accommodate list-mode data. Similarly to the previously introduced bin-mode data image reconstruction, the image vector $\hat{\mathbf{x}}$ is recovered by solving the MAP optimization problem,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} l(\mathbf{x}; \mathbf{y}) + l_p(\mathbf{x}). \quad (4.46)$$

where $l(\mathbf{x}; \mathbf{y})$ and $l_p(\mathbf{x})$ are the negative log-likelihood and the negative log-prior distribution. Given length- M list-mode data and a corresponding system matrix $\mathbf{A}$, the log-likelihood function is

$$l(\mathbf{x}; \mathbf{y}) \stackrel{c}{=} T\mathbf{s}^\top \mathbf{x} - \mathbf{1}^\top \log \mathbf{A}\mathbf{x} \quad (4.47)$$

where T is a detector dwell time and $\mathbf{s} \in \mathbb{R}^N$ is the sensitivity vector. Constructing the Gaussian process prior proceeds similarly to the bin-mode case. First, assuming that the image vector $\mathbf{x}$ is governed by an underlying vector $\boldsymbol{\xi}$, which follows a zero-mean multivariate Gaussian distribution with a positive definite covariance matrix $\boldsymbol{\Sigma}$:

$$p(\boldsymbol{\xi}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp(\boldsymbol{\xi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}). \quad (4.48)$$

Using the squared exponential covariance kernel, covariance matrix is defined as

$$\Sigma_{ij} = \exp\left(-\frac{\|d_{ij}\|_2^2}{2\sigma^2}\right), \quad (4.49)$$

where $d_{ij} = 2 \sin(\arccos(\hat{\Omega}_i \cdot \hat{\Omega}_j)/2)$ is a chordal length between pixel i and j . It should be noted that the imaging space is 4π spherical manifold, and using Euclidean distance for $d_{ij} = 2 \sin(\arccos(\hat{\Omega}_i \cdot \hat{\Omega}_j)/2)$ results in an invalid kernel function [92].

Once the Gaussian random variable $\boldsymbol{\xi}$ is defined, the Gaussian-to-exponential inverse link function in (4.13) link function $f : \mathbb{R}_{\geq 0}^N \longrightarrow \mathbb{R}^N$ can maps the Gaussian variable $\boldsymbol{\xi}$ to $\mathbf{x}$. With

the link function defined, the following optimization problem is solved for list-mode GPP image reconstruction,

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} Ts^{\top}f(\xi) - \mathbf{1}^{\top} \log \mathbf{A}f^{\top}(\xi) + \xi^{\top}\Sigma^{-1}\xi. \quad (4.50)$$

The subsequent steps are analogous to the bin-mode data case and therefore, it will not be described in detail here.

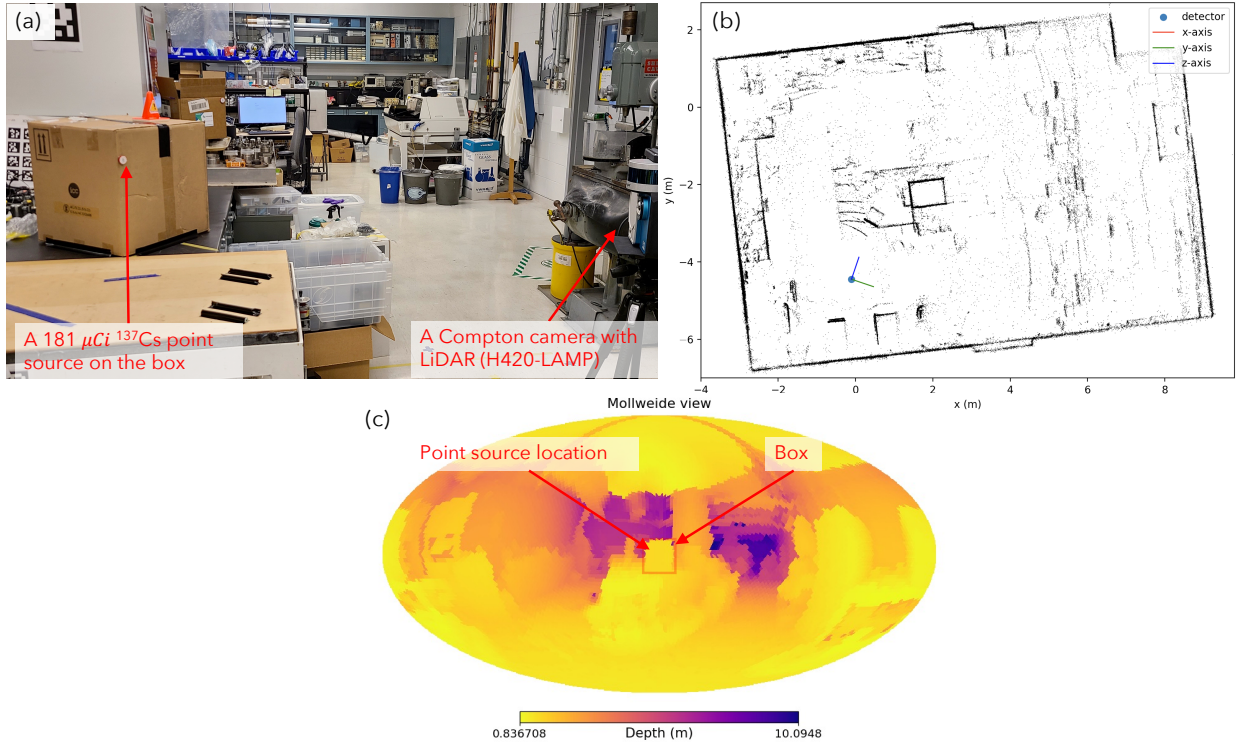


Figure 4.7: (a) Measurement setup comprising a 181 μCi ^{137}Cs point source and a CZT-based Compton camera, H420-LAMP. (b) Point cloud map of the laboratory where the measurement was conducted. (c) A 4π depth map showing the distance between the Compton camera and the surrounding objects.

Figure 4.7 (a) shows the 181 μCi ^{137}Cs point source measurement setup. The source was placed on the left edge of a cardboard box, and a H420-LAMP imager was positioned approximately 1.5 meters away and left to dwell for about 20 min. The H420-LAMP imager consists of a CZT-based static Compton camera (H3D H420) and a suite of contextual sensors including a LiDAR. Although the imager is capable of 3D free-moving imaging, it was employed here as a static Compton imager with a 4π imaging space. This setup emphasize quantification of source activity, rather than localization of it. Quantification of

radioactive source is relevant in applications such as nuclear safeguards, where the source position is typically well known.

For accurate radioactivity quantification, the source-to-detector distance must be measured to account for radiation intensity decrease with distance (i.e., inverse-square law). To achieve this, a point cloud map of the measurement room was generated prior to the measurement, as shown in Fig. 4.7 (b). This map was then converted into a 4π depth map, illustrating distances between the detector and nearby objects.

The quantitative image reconstruction was conducted in two-steps. First, image reconstruction was performed to yield a radiation flux map, which reveals directional radiation flux estimated at the detector position. The flux map was then converted into an quantitative activity map by multiplying the square of depth map. The depth weighting step is essential, as it accounts for decrease in radiation intensity with the inverse-square law. Figure 4.8 (a) and (c) show the radiation flux map reconstructed from the list-mode ML-EM and the GPP. For the GPP method, the depth map was used as a structural prior, and the pixels corresponding to the box were decorrelated from the rest of the pixels. As a result, while the ML-EM flux map localizes the point source direction with a typical Gaussian-like PSF, the GPP suppresses the pixel intensities outside of the box, concentrating intensities on the box only.

The incorporation of the structural prior plays a crucial role in converting the flux map into an activity map. Figure 4.8 (b) and (d) illustrate the activity maps from the list-mode ML-EM and the GPP with a structural prior. Notice that in the ML-EM activity map, the background is significantly amplified due to the large detector-to-background distance (~ 10 m). As a result, the sum of activity on the box pixels amount to $78.1 \mu\text{Ci}$, which is considerably lower than the ground truth activity, $181 \mu\text{Ci}$. On the other hand, the GPP with the structural prior effectively constrain the reconstructed activity within the box, yielding the estimated total activity on the box of $179 \mu\text{Ci}$.

Furthermore, the Laplace approximation with the GPP enables uncertainty quantification for estimated the total activity. Let $g(\mathbf{x})$ be a function that takes an activity map as input and outputs the total activity on the box. Samples from an approximate posterior distribution $\{\boldsymbol{\xi}_n\}_{n=1}^J$ are transformed as $\{g(\mathbf{f}^{-1}(\boldsymbol{\xi}_n))\}_{n=1}^J$, yielding samples from the posterior distribution, $p(g(\mathbf{x})|\mathbf{y})$, which allows computation of Bayesian uncertainty for the total activity.

Figure 4.10 presents the histogram of $\{g(\mathbf{f}^{-1}(\boldsymbol{\xi}_n))\}_{n=1}^J$ derived from the Laplace approximation, along with the mean and standard deviation of the samples. The mean of the total activity in the box was $179 \mu\text{Ci}$ with a standard deviation $7.40 \mu\text{Ci}$. This indicates that the GPP method can not only prove an accurate quantitative activity map, but also provides enhances robustness of the quantification with the uncertainty quantification. Moreover, the samples $\{g(\mathbf{f}^{-1}(\boldsymbol{\xi}_n))\}_{n=1}^J$ lends itself to more interesting statistical inferences, such as Bayesian hypothesis testing of radioactive isotope composition.

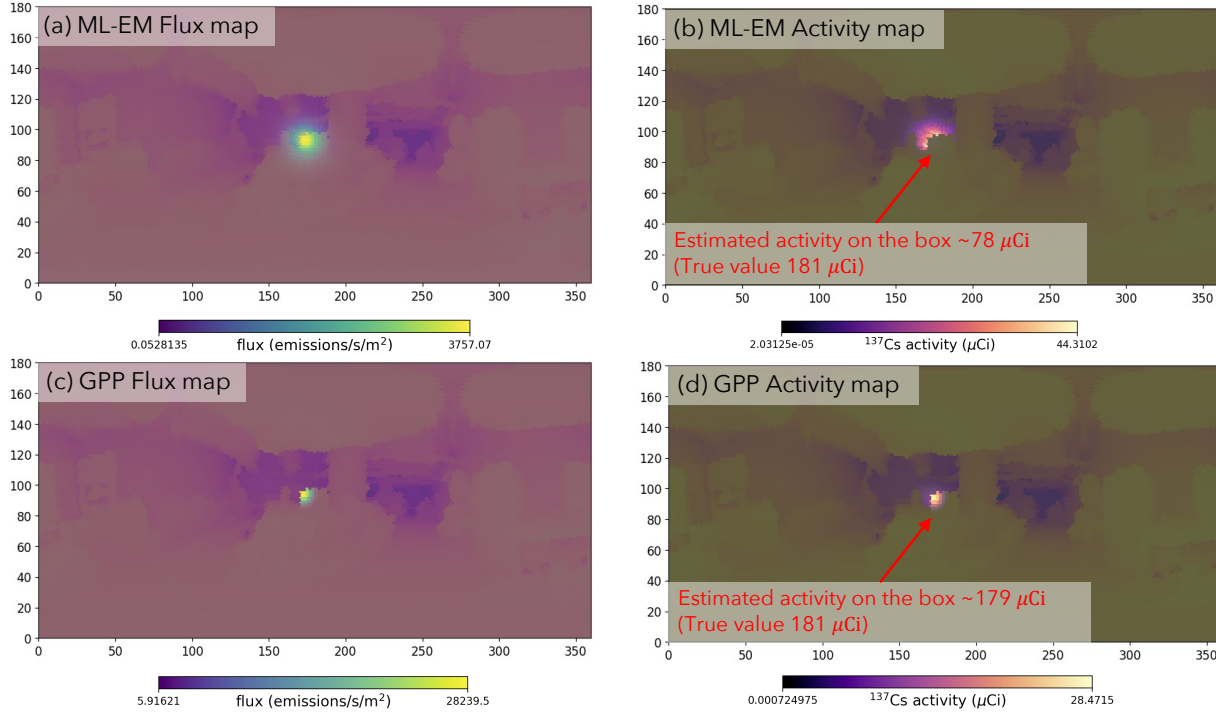


Figure 4.8: Directional radiation flux maps reconstructed from (a) the ML-EM and (c) the GPP MAP with a structural prior. The flux maps are weighted using the depth map from Fig. 4.7, yielding the activity maps from (b) the ML-EM and (d) the GPP MAP. Note that the ML-EM flux map spreads across both the box and the background, whereas the GPP MAP with a structural prior effectively constrains the reconstructed flux within the box. Consequently, the ML-EM fails to accurately estimate the total activity within the box, while the GPP achieves a significantly more accurate estimate of 179 μCi .

Minimum Dose Path Planning

Here, we introduce an application that highlights the importance of the uncertainty quantification capability developed in this work. Figure 4.10 (a) illustrates a simulated single detector measurement scenario involving a large distributed anisotropic Gaussian source within a $200\text{ m} \times 200\text{ m}$ imaging space. After a raster pattern measurement, the GPP MAP algorithm was used to successfully reconstruct the source distribution, as shown in Fig. 4.10 (b).

In certain radiation safety applications, it is often of interest to quantify spatial distribution of dose rate, instead of source activity. A dose rate map can be easily derived from the estimated activity map, by convolving it with a point dose kernel, which describes the dose rate from a point source as a function of distance. The resulting dose rate map is presented in Fig. 4.10 (c).

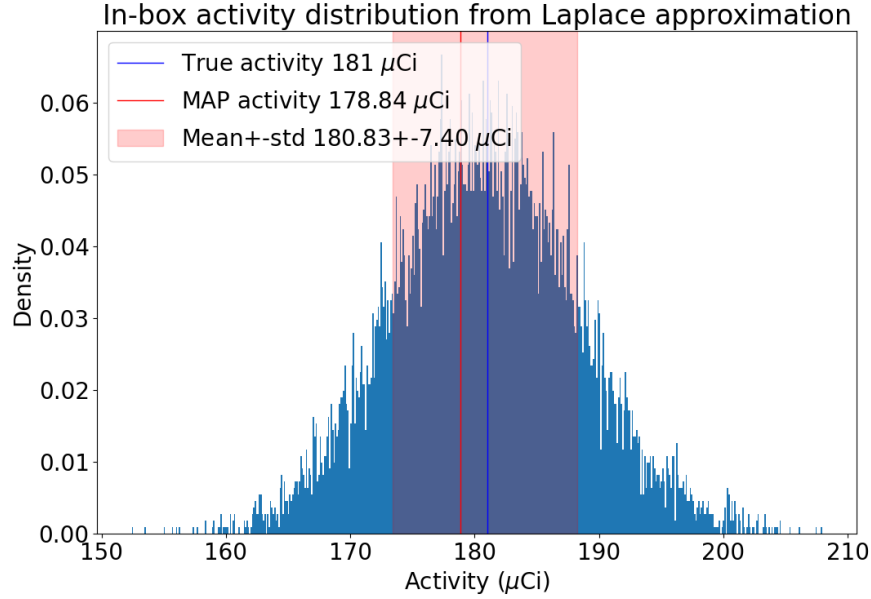


Figure 4.9: Histogram of the estimated total activity samples.

Suppose that, based on the reconstructed dose map, we wish to determine an optimal Path from point A to point B, that minimizes the dose received from the source. Three example paths are shown in Fig. 4.10 (c). Which Path should be chosen? Path 1, the shortest Path connecting points A and B, is clearly the worst choice, as it traverses through the estimated radiation hotspot. Excluding Path 1, Path 2 might appear to be the best option, as it stays farthest from the hotspot. Path 3, however, seems less ideal as the Path is somewhat closer to the hotspot than Path 2.

To rigorously answer this question, it is necessary to examine the uncertainties in the estimated dose map. Figure 4.10 (d) shows the dose uncertainty map derived from the Laplace approximation. Notably, the uncertainty is highest on the left edge of the imaging space due to the large distance from the measurement Path (i.e., low sensitivity). Consequently, the uncertainty in the total dose received from Path 2 is expected to be greater than Paths 1 and 3.

Figure 4.11 summarizes the total dose received along the three example paths and their associated uncertainties. As expected, the largest dose is received along Path 1. When comparing Path 2 and 3, the GPP MAP estimates that the total dose along Path 2 is slightly lower than Path 3. However, the uncertainty is much is significantly higher in Path 2, suggesting the possibility of undetected radioactive sources near the left edge of the imaging space.

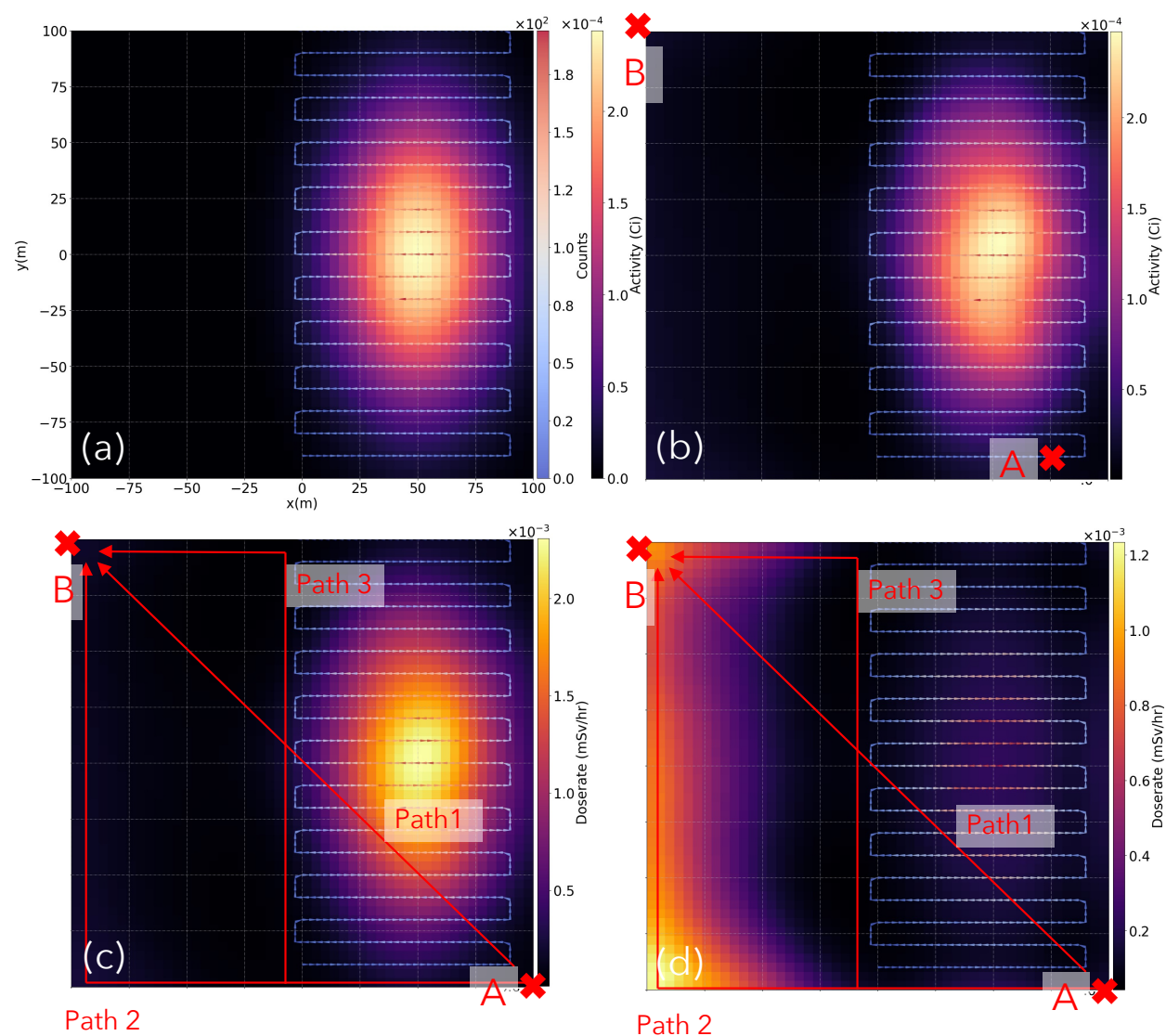


Figure 4.10: (a) Measurement scenario with the ground truth distribution. (b) GPP MAP reconstruction result. (c) Estimated dose rate map and example paths. (d) Dose rate uncertainty map.

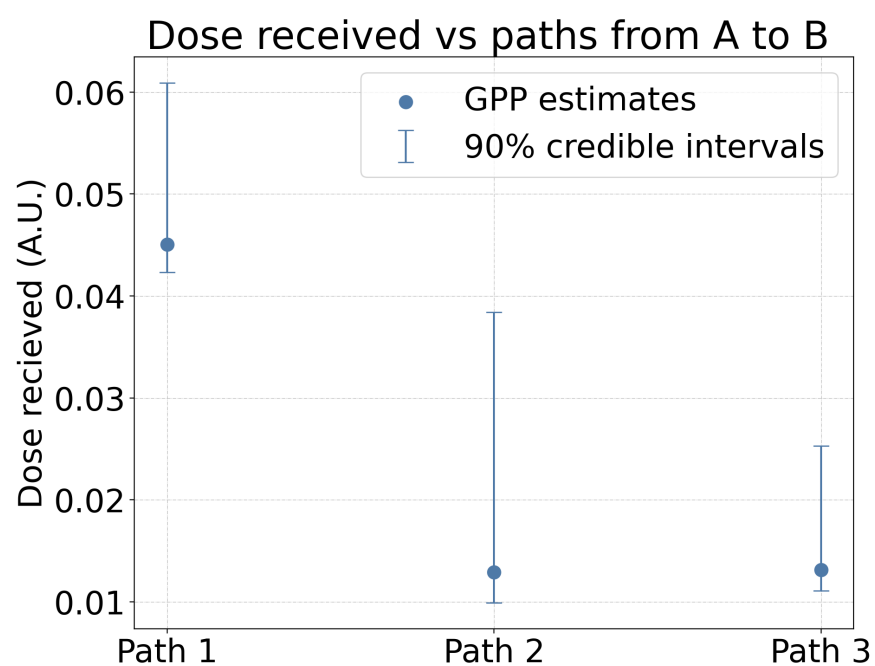


Figure 4.11: Summary of total dose received along Paths 1 2 and 3 and their associated uncertainties. Samples from an approximate posterior distribution were used to derive the uncertainties.

4.4 Discussion

In this section, we discuss the implications of the proposed methods and potential future research directions.

Implications of the Proposed Approach

First, the proposed GP prior can be made highly flexible with a choice of a covariance function and a link function. Therefore, the GP prior can effectively model desired properties of the reconstructed images, and is directly applicable to a wide range of highly ill-posed image reconstruction tasks. Such tasks arise from various radiation imaging modalities including Compton cameras [93], coded aperture imaging [94], limited angle computed tomography (CT) [95], low-dose positron emission tomography (PET) [96], among others.

Secondly, the uncertainty quantification capability of the proposed framework offers valuable insights into the reconstructed images, and it becomes particularly important when critical decisions are made based on the reconstructed images. Such problems commonly arise in many healthcare [97] and nuclear security applications [98]. Moreover, posterior distributions estimated from the pCN and the Laplace approximation can be employed for further inferences on quantities derived from the reconstructed images. For instance, suppose a quantity of interest u can be derived from an image $\mathbf{x}$ as $u = g(\mathbf{x})$. From a set of posterior samples obtained from the pCN, $\{\boldsymbol{\xi}_n\}_{n=1}^J$, the samples of the derived quantity $\{u_n\}_{n=1}^J$ can be obtained as $u_n = g(\mathbf{f}^{-1}(\boldsymbol{\xi}_n))$. In the context of the Laplace approximation, samples of $\boldsymbol{\xi}$ can be obtained from the approximating Gaussian distribution in (4.27). Some examples of such derived quantities are total dose registered in radiopharmaceutical therapy [99], in-beam ion ranges in ion cancer therapy, and *in-vivo* proton range verification in ion cancer therapy [100].

Future works

Incorporation of Background Contribution

The results presented in section 4.3 demonstrate that the proposed GPP algorithm can significantly improve the reconstruction quality where the ML-EM may be inadequate. However, the simulated measurement scenarios assume no present background contributions, which may be unrealistic in some radiation imaging applications. When the background contribution is known, it can be straightforwardly accounted for by changing the forward projection $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$ to $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{b}$, where the elements of the vector $\mathbf{b}$ are the known background contribution to the measurements. When the background rate is unknown, it has to be jointly estimated with the image $\mathbf{x}$, and it can be done by augmenting the system matrix $\mathbf{A}$ and the image vector $\mathbf{x}$ as $\tilde{\mathbf{A}} = [\mathbf{A} | \mathbf{a}_b]$ and $\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ b \end{bmatrix}$, where the vector $\mathbf{a}_b$ is assumed to be known (e.g., in the case of a constant background rate, a vector of integration time $\mathbf{a}_b =$

Δt). Subsequently, the prior covariance matrix Σ can be augmented as

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Sigma_b \end{bmatrix}, \quad (4.51)$$

where the Σ_b accounts for the variance of the background rate.

The use of the Gaussian-to-exponential link function may be inadequate as the background rate can be typically well specified within an order of magnitude. In such case, a different link function, such as the truncated Gaussian-to-Gaussian link function [78], can be used to effectively model the background prior distribution.

List-mode Data and Gaussian Likelihood

Although the work presented here assumes bin-mode data with the Poisson likelihood model, it should be emphasized that the proposed framework can be trivially modified to accommodate list-mode data [36] or Gaussian likelihood model. Essentially, the only differences in the formulation will be the likelihood function formulation in (4.2), and subsequently its gradient, Hessian expressions.

For example, when the measurements are given in terms of dose rates, Gaussian likelihood model can more aptly describe the measurement noise. In such cases, the likelihood function $l(\mathbf{x}; y)$ in (4.3) can be modified as follows:

$$l(\mathbf{x}; y) = \frac{1}{2}(\mathbf{y} - \mathbf{Ax})^\top \mathbf{W}(\mathbf{y} - \mathbf{Ax}), \quad (4.52)$$

where the matrix $\mathbf{W} \in \mathbb{R}^{M \times M}$ is the inverse of the measurement covariance matrix (i.e. a precision matrix). When independent and identical noise distribution is assumed, $\mathbf{W}$ is an identity matrix. With the Gaussian likelihood model, the GPP image reconstruction and uncertainty quantification presented in this paper can be similarly applied.

By accommodating various data types, the proposed framework is applicable to a wide range of radiation imaging applications.

Scalable Implementation

In section 4.2, we illustrated how the Kronecker product structure in the covariance matrix Σ can be exploited to avoid the computational challenges involving the Cholesky factor $\mathbf{L}$. The approach is particularly attractive as it provides exact representation of the covariance matrix; however, the method is not readily applicable when the imaging space is not in a dense Cartesian grid. For example, Compton cameras or active coded apertures imagers often employ a 4π spherical imaging space [101][102], and free-moving imaging scenarios often assume sparse voxel grid imaging spaces [103]. Thus, a different strategy is needed when more diverse imaging spaces are considered.

Several recent advances from the GP regression research community could potentially be employed to tackle the challenges. For example, in [104], Willson *et al.* introduced the

structured kernel interpolation (SKI) scheme, where the covariance matrix is interpolated from a Kronecker product represented sparse inducing points. The resulting approximate covariance matrix can be handled efficiently thanks to the sparsity of the interpolation coefficient matrix and the Kronecker product structure. Such formulation could be applied to imaging spaces on a non-structured grid.

Another interesting approach is Gaussian Markov random field (GMRF), where a sparse precision matrix Σ^{-1} is directly constructed and utilized. Such precision matrices admit rapid realization of GPs, removing the computational bottlenecks involving $\mathbf{L}\xi_w$. Moreover, Lindgren *et al.* showed that a GMRF can directly approximate GPs with a Matérn covariance kernel, and it can be efficiently applied to meshed imaging spaces [105].

Covariance Kernels

The SE kernel in (4.9) is a popular choice in many GP regression tasks, and in this work we demonstrated that it can be used to effectively model the prior covariance. In many radiation imaging modalities, the point spread functions (PSF) have smoothly falling edges and therefore, the use of the SE kernel could be well-justified. However, when it is required to recover high frequency features (e.g., edges or point sources), more expressive kernels may be desired. Rational quadratic kernel or Matérn covariance kernels could be readily used to provide more flexibility for better high frequency feature resolving capability. Moreover, recent works from the GP regression community suggest the use of spectral mixture (SM) kernel [106] or deep kernel learning framework [107], where a deep neural network is used to construct a highly expressive covariance kernel. It was demonstrated in [107] that such kernels can successfully recover piecewise constant functions with sharp falling edges, which is attractive for imaging applications.

4.5 Supplementary Information

Derivation of the GPP Posterior Gradient

The gradient of the negative log-posterior $\Psi(\mathbf{L}\xi_w)$ can be obtained as follows. Adopting the convention that the gradient of a scalar function is a column vector,

$$\nabla_{\xi_w} \Psi(\mathbf{L}\xi_w) \tag{4.53}$$

$$\begin{aligned} &= \nabla_{\xi_w} l(\mathbf{L}\xi_w; \mathbf{y}) + \nabla_{\xi_w} l_p(\mathbf{L}\xi_w) \\ &= \mathbf{J}_{\xi_w} \xi^\top \mathbf{J}_\xi \mathbf{f}^{-1}(\xi)^\top \nabla_{\mathbf{f}^{-1}(\xi)} l(\mathbf{L}\xi_w; \mathbf{y}) + \xi_w, \end{aligned} \tag{4.54}$$

where $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{L}\boldsymbol{\xi}_w)$ and $\boldsymbol{\xi} = \mathbf{L}\boldsymbol{\xi}_w$. The second equality follows from the chain rule. Each of the gradients and Jacobians in (4.54) is given as

$$\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\mathbf{L}\boldsymbol{\xi}_w; \mathbf{y}) = \mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top (\mathbf{y} \odot (\mathbf{A}\mathbf{x})), \quad (4.55)$$

$$\mathbf{J}_\xi \mathbf{f}^{-1}(\boldsymbol{\xi}) = \text{diag} \left(\mathbf{1} \odot \left(\sqrt{2\pi} \lambda \text{diag}(\boldsymbol{\Sigma}) \right) \odot \exp \left(\lambda \mathbf{x} - \boldsymbol{\xi}^{\odot 2} \odot (2 \text{diag}(\boldsymbol{\Sigma})) \right) \right), \quad (4.56)$$

$$\mathbf{J}_{\xi_w} \boldsymbol{\xi} = \mathbf{L}. \quad (4.57)$$

Note that depending on the choice of a link function, the Jacobian $\mathbf{J}_\xi \mathbf{f}^{-1}(\boldsymbol{\xi})$ is computed differently. The Jacobian given in (4.56) is for the exponential-to-Gaussian link function introduced in the main text. Combining (4.55), (4.56), and (4.57) together, the gradient is given as

$$\begin{aligned} \nabla_{\xi_w} \Psi(\mathbf{L}\boldsymbol{\xi}_w) = \\ \mathbf{L}^\top \left((\mathbf{A}^\top \mathbf{1} - \mathbf{A}^\top (\mathbf{y} \odot (\mathbf{A}\mathbf{x}))) \odot \left(\mathbf{1} \odot \left(\sqrt{2\pi} \lambda \text{diag}(\boldsymbol{\Sigma}) \right) \right. \right. \\ \left. \left. \odot \exp \left(\lambda \mathbf{x} - \boldsymbol{\xi}^{\odot 2} \odot (2 \text{diag}(\boldsymbol{\Sigma})) \right) \right) \right) + \boldsymbol{\xi}_w. \end{aligned} \quad (4.58)$$

Approximate Posterior and the Marginal Likelihood Using the Laplace Approximation

First we write the log-posterior distribution as $\log p(\boldsymbol{\xi}|\mathbf{y}) = \log \frac{1}{Z} - \Psi(\boldsymbol{\xi})$, where Z is the normalization constant. Then, the second order Taylor expansion of $-\Psi(\boldsymbol{\xi})$ around the mode of the posterior $\hat{\boldsymbol{\xi}} = \mathbf{L}\hat{\boldsymbol{\xi}}_w$ is given as,

$$-\Psi(\boldsymbol{\xi}) \approx -\Psi(\hat{\boldsymbol{\xi}}) - \frac{1}{2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^\top \nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \quad (4.59)$$

Note that the first order term is not included because the gradient at the mode $-\nabla_{\boldsymbol{\xi}} \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$ is zero. Using the Taylor expansion, the log-posterior distribution is written as,

$$\log p(\boldsymbol{\xi}|\mathbf{y}) \approx \log \frac{1}{Z} - \Psi(\hat{\boldsymbol{\xi}}) - \frac{1}{2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^\top \nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \quad (4.60)$$

Exponentiating both sides of (4.60) to recover the posterior $p(\boldsymbol{\xi}|\mathbf{y})$, we have

$$\begin{aligned} p(\boldsymbol{\xi}|\mathbf{y}) \approx \\ \frac{1}{Z} \exp(-\Psi(\hat{\boldsymbol{\xi}})) \exp \left(-\frac{1}{2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})^\top \nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \right) \end{aligned} \quad (4.61)$$

Since (4.60) is in the form of Gaussian distribution, it immediately follows that,

$$p(\boldsymbol{\xi}|\mathbf{y}) \approx q(\boldsymbol{\xi}|\mathbf{y}), \quad \text{where} \quad (4.62)$$

$$q(\boldsymbol{\xi}|\mathbf{y}) \sim \mathcal{N}\left(\hat{\boldsymbol{\xi}}, \left(\nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}\right)^{-1}\right), \quad (4.63)$$

The Hessian matrix $\nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})$ is given as,

$$\nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y}) + \nabla_{\boldsymbol{\xi}}^2 l_p(\boldsymbol{\xi}) \quad (4.64)$$

where, $\nabla_{\boldsymbol{\xi}}^2 l_p(\boldsymbol{\xi}) = \boldsymbol{\Sigma}^{-1}$. The Hessian of the negative log-likelihood function $\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})$ is given as,

$$\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y}) = \mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi}) \nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})}^2 l(\boldsymbol{\xi}; \mathbf{y}) \mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi}) + \nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y}) \nabla_{\boldsymbol{\xi}}^2 \mathbf{f}^{-1}(\boldsymbol{\xi}). \quad (4.65)$$

The expressions for $\mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi})$ and $\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y})$ are given in (4.56) and (4.55). $\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})}^2 l(\boldsymbol{\xi}; \mathbf{y})$ and $\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y}) \nabla_{\boldsymbol{\xi}}^2 \mathbf{f}^{-1}(\boldsymbol{\xi})$ are given as,

$$\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})}^2 l(\boldsymbol{\xi}; \mathbf{y}) = \mathbf{A}^T \text{diag}\left(\frac{\mathbf{y}}{\bar{\mathbf{y}}^{\odot 2}}\right) \mathbf{A} \quad \text{and} \quad (4.66)$$

$$\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y}) \nabla_{\boldsymbol{\xi}}^2 \mathbf{f}^{-1}(\boldsymbol{\xi}) = \text{diag}(\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y})) \mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi}) (\lambda \mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi}) - (\boldsymbol{\xi} \odot \text{diag}(\boldsymbol{\Sigma}))). \quad (4.67)$$

Note that $\nabla_{\boldsymbol{\xi}}^2 \mathbf{f}^{-1}(\boldsymbol{\xi})$ is a 3-D diagonal tensor, but the computation of $\nabla_{\mathbf{f}^{-1}(\boldsymbol{\xi})} l(\boldsymbol{\xi}; \mathbf{y}) \nabla_{\boldsymbol{\xi}}^2 \mathbf{f}^{-1}(\boldsymbol{\xi})$ is equivalent to (4.67). Combining (4.56), (4.55), (4.66) and (4.67), the Hessian of the negative log-likelihood function (4.65) is computed.

Fisher information for Hessian approximation

The Fisher information matrix provides a well-justified approximation to the Hessian of the negative log-likelihood $\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})$. The approximation can be expressed as

$$\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y}) \approx \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[\left(\nabla_{\boldsymbol{\xi}} l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right) \left(\nabla_{\boldsymbol{\xi}} l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right)^T \right], \quad (4.68)$$

where the right hand side is the definition of the Fisher information matrix. Using the chain rule, the gradient $\nabla_{\boldsymbol{\xi}} l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$ can be expressed as,

$$\nabla_{\boldsymbol{\xi}} l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} = (\mathbf{J}_{\boldsymbol{\xi}} \mathbf{f}^{-1}(\boldsymbol{\xi}))^T (\mathbf{J}_{\mathbf{f}^{-1}(\boldsymbol{\xi})} \log \bar{\mathbf{y}})^T \nabla_{\log \bar{\mathbf{y}}} l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}, \quad (4.69)$$

where $\bar{\mathbf{y}} = \mathbf{A} \mathbf{f}^{-1}(\boldsymbol{\xi})$ is forward projection. Plugging (4.69) into (4.68), we have

$$\begin{aligned} \nabla_{\xi}^2 l(\xi; \mathbf{y}) &\approx \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[(\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}})^{\top} \nabla_{\log \bar{\mathbf{y}}} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right. \\ &\quad \left. \nabla_{\log \bar{\mathbf{y}}} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}}^{\top} (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}}) (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)) \right] \end{aligned} \quad (4.70)$$

$$\begin{aligned} &= (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}})^{\top} \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[\nabla_{\log \bar{\mathbf{y}}} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right. \\ &\quad \left. \nabla_{\log \bar{\mathbf{y}}} l(\xi; \mathbf{y})|_{\xi=\hat{\xi}}^{\top} \right] (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}}) (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)) \end{aligned} \quad (4.71)$$

$$= (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}})^{\top} \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} \left[\nabla_{\log \bar{\mathbf{y}}}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} \right] (\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}}) (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)). \quad (4.72)$$

The equivalence between the expectation of the outer product and the Hessian follows from the property of the Fisher information matrix (see, [108]) The Jacobians and the Hessian in (4.72) are given as,

$$(\mathbf{J}_{\mathbf{f}^{-1}(\xi)} \log \bar{\mathbf{y}}) = \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \mathbf{A} \quad (4.73)$$

$$\nabla_{\log \bar{\mathbf{y}}}^2 l(\xi; \mathbf{y})|_{\xi=\hat{\xi}} = \text{diag}(\bar{\mathbf{y}}). \quad (4.74)$$

The Jacobian $(\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))$ is given in (4.56) for the exponential-to-Gaussian link function, and it may differ depending on the choice of a link function. Plugging the Jacobians and Hessian in (4.72), we have

$$\nabla_{\xi}^2 l(\xi; \mathbf{y}) \approx (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} \mathbf{A}^{\top} \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y}|\bar{\mathbf{y}})} [\text{diag}(\bar{\mathbf{y}})] \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \mathbf{A} (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)) \quad (4.75)$$

$$= (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} \mathbf{A}^{\top} \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \text{diag}(\bar{\mathbf{y}}) \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \mathbf{A} (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)) \quad (4.76)$$

$$= (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi))^{\top} \mathbf{A}^{\top} \text{diag}(\bar{\mathbf{y}}^{\circ-1}) \mathbf{A} (\mathbf{J}_{\xi} \mathbf{f}^{-1}(\xi)) \quad (4.77)$$

Marginal likelihood computation using the Laplace approximation

The marginal likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ can be expressed as,

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int \exp(\log p(\mathbf{y}|\xi) p(\xi|\boldsymbol{\theta})) d\xi \quad (4.78)$$

Similarly to (4.60), the $\log p(\mathbf{y}|\xi) p(\xi|\boldsymbol{\theta})$ can be approximated using the second-order Taylor expansion around the mode,

$$\begin{aligned} \log p(\mathbf{y}|\xi) p(\xi|\boldsymbol{\theta}) &\approx \\ \log p(\mathbf{y}|\hat{\xi}) p(\hat{\xi}|\boldsymbol{\theta}) &- \frac{1}{2} (\hat{\xi} - \xi)^{\top} \nabla_{\xi}^2 \Psi(\xi)|_{\xi=\hat{\xi}} (\hat{\xi} - \xi) \end{aligned} \quad (4.79)$$

Plugging (4.79) into (4.78), we have,

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|\hat{\boldsymbol{\xi}})p(\hat{\boldsymbol{\xi}}|\boldsymbol{\theta}) \int \exp\left(-\frac{1}{2}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\right)^\top \nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\right)\right) d\boldsymbol{\xi} \quad (4.80)$$

The integration in (4.80) can be analytically computed as $\left((2\pi)^N \left|\left(\nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}\right)^{-1}\right|\right)^{\frac{1}{2}}$ (i.e., Gaussian integration). Hence, the negative log-marginal likelihood up to constant, $\Phi(\boldsymbol{\theta}) \stackrel{c}{=} -\log p(\mathbf{y}|\boldsymbol{\theta})$ is given as,

$$\Phi(\boldsymbol{\theta}) = \Psi(\hat{\boldsymbol{\xi}}) + \frac{1}{2} \log \left| \left(\nabla_{\boldsymbol{\xi}}^2 \Psi(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right) \right| + \log |\boldsymbol{\Sigma}| \quad (4.81)$$

$$= \Psi(\hat{\boldsymbol{\xi}}) + \frac{1}{2} \log \left| \left(\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} + \boldsymbol{\Sigma}^{-1} \right) \right| + \log |\boldsymbol{\Sigma}| \quad (4.82)$$

$$= \Psi(\hat{\boldsymbol{\xi}}) + \frac{1}{2} \log \left| \mathbf{I} + \nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \boldsymbol{\Sigma} \right|. \quad (4.83)$$

Note that the expression for $\nabla_{\boldsymbol{\xi}}^2 l(\boldsymbol{\xi}; \mathbf{y})$ is given in (4.65).

Chapter 5

Simultaneous Spectral Decomposition and Full Spectral Imaging

In the previous chapter, we introduced a GP-based prior for radiation image reconstruction, uncertainty quantification method. When the method is used in the context of single detector imaging scenarios, variations in radiation detection counts over different measurement times is used as input for image reconstruction algorithms. However, a significant portion of the counts recorded by the single detector originates from background radiation, which is primarily from naturally occurring radioactive materials (NORM) and cosmic rays. These background radiations often obscure the desired source signals, resulting in a low signal-to-noise ratio (SNR) and consequently, a degradation in image quality. One straightforward method to enhance the SNR is to utilize radiation counts within the energy ranges of the source photopeaks only [30][33][31][109]. While this method can significantly improve the SNR, its applicability is limited for the following reasons:

- First, in many measurement scenarios, the radioisotopes to be mapped may not be known *a priori*, and thus the corresponding photopeak energies may also be unknown.
- Even if the photopeak(s) from the source can be identified, it is often unclear whether the photopeaks are from a mixture of radionuclides with the same spatial distributions, or if each photopeak is associated with a unique spatial distribution.
- In some measurement scenarios, the photopeaks may not be discernible from recorded energy spectra. For example, radiation spectra from heavily shielded or attenuated sources often lack distinct photopeak features.
- Lastly, background radiation fields may vary spatiotemporally, potentially interfering with the source signals of interests.

To address these challenges, this chapter introduces a novel approach, in which a series of time-integrated spectra are decomposed into salient source and background spectral components contributing to the total measurements. Simultaneously, image reconstruction is

performed alongside the decomposition task, producing distribution maps corresponding to the identified source components. The problem is inherently highly complex and as a result, the conventional maximum-likelihood estimation (MLE) method suffers from severe overfitting. To overcome this limitation, we formulate the problem within a Bayesian framework, imposing Gaussian process (GP) priors on spectral components and images [110]. Note that in Chapter 4, the GP prior was imposed only on the image to be estimated. In this chapter, we introduce a novel framework for estimating radiation images and their corresponding radiation energy spectra simultaneously, and impose GP priors on both the images and the spectra. This approach is robust against overfitting even with low SNR measurements and the reconstructed spectral components provide extensive insights into the radiological environment from a single-detector measurement. Furthermore, we present a computationally efficient Bayesian uncertainty quantification and model selection methods, which significantly enhances robustness of the approach.

In the following, we provide a complete mathematical description of the reconstruction, uncertainty quantification and model selection algorithms. Additionally, the method is tested on both a challenging simulated measurement scenario and a real-world measurement to demonstrate the effectiveness of the proposed approach.

5.1 Methods

We consider a measurement scenario in which the objective is to estimate radioactivity distribution images across across all different radiation emission energies. These images can be represented in matrix form as $\mathbf{X} \in \mathbb{R}_{\geq 0}^{K \times L}$, where K denotes the number of image voxels and L represents the number of emission energy bins. The image matrix $\mathbf{X}$ is reconstructed from a 2-D time-energy histogram collected from a free-moving single detector, $\mathbf{Y} \in \mathbb{Z}_{\geq 0}^{I \times J}$. Note that I and J represent the number of time bins and recorded energy bins, respectively. Therefore, $\mathbf{Y}_{ij}$ denotes the detected radiation counts within the i -th time interval and j -th energy bin.

The two matrices are related through a system matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{IJ \times KL}$, such that

$$\mathbf{Y} \sim \text{Poisson}(\bar{\mathbf{Y}}), \quad \text{where} \quad \text{vec}(\bar{\mathbf{Y}}) = \mathbf{A} \text{vec}(\mathbf{X}). \quad (5.1)$$

Vectorization operation $\text{vec}(\cdot)$ converts a matrix into a vector by concatenating column vectors together. Note that the system matrix can be represented as a block matrix with $J \times L$ blocks, with each block is has dimension of $I \times K$. In other words, (i,k) -th element in the (j,l) -th block matrix encodes the probability that a radiation emitted from voxel k with an emission energy in bin l is detected in time bin i with the recorded energy in bin j .

Using the Poisson statistics of the measurement matrix $\mathbf{Y}$, a likelihood function $p(\mathbf{Y}|\mathbf{X})$ can be constructed, and the maximum likelihood estimation (MLE) of the image matrix $\mathbf{X}$ can be found using the conventional maximum likelihood-expectation maximization (MLEM) algorithm. However, the size of the system matrix $\mathbf{A}$ becomes prohibitively large to compute and store even with a moderate number of image voxels and time-energy histogram

bin discretization. Moreover, the MLE problem is susceptible to overfitting as the image matrix $\mathbf{X}$ contains too many unknown elements.

In order to make the problem computationally tractable and robust to overfitting, we make use of the inherent low-rank structures present in the system matrix $\mathbf{A}$ and the image matrix $\mathbf{X}$. First, we observe that the system matrix can be effectively decomposed into two smaller matrices, $\mathbf{A}_G \in \mathbb{R}_{\geq 0}^{I \times K}$ and $\mathbf{A}_R \in \mathbb{R}_{\geq 0}^{J \times L}$, which represent the geometric efficiencies and the detector response, respectively. In other words, $\mathbf{A}_{G[i,k]}$ denotes the solid angle of the detector at time i viewed from voxel k , and $\mathbf{A}_{R[j,l]}$ is the probability an incoming photon within energy bin j records a count in the energy bin l . Notably, this is equivalent to Kronecker product decomposition, $\mathbf{A} \approx \mathbf{A}_R^\top \otimes \mathbf{A}_G$, and it is justified from the fact that a single detector efficiency can be suitably approximated as the product of geometric efficiency and intrinsic efficiency [111].

Furthermore, we note that the image matrix $\mathbf{X}$ can be compactly represented with a low-rank decomposition, $\mathbf{X} = \mathbf{X}_I \mathbf{X}_C^\top$, where $\mathbf{X}_I \in \mathbb{R}_{\geq 0}^{K \times S}$, $\mathbf{X}_C \in \mathbb{R}_{\geq 0}^{L \times S}$ and $S \ll K, L$. Columns of $\mathbf{X}_I$ represent radioactivity distribution images corresponding to distinct radiation emission energy spectra, which are the columns of the matrix $\mathbf{X}_C$. In other words, the image matrix $\mathbf{X}$ is approximated with outer products of S distinct radiation images and corresponding spectral component vectors, such that $\mathbf{X} \approx \sum_{i=1}^S \mathbf{X}_I[:, i] \mathbf{X}_C[:, i]^\top$. The decomposition is reasonable, as a distinct spatial distribution of radioisotopes produces almost identical radiological emission spectrum across the entire measurement time. Incorporating the decomposition of $\mathbf{A}$ and $\mathbf{X}$, the problem can be rewritten as,

$$\mathbf{Y} \sim \text{Poisson}(\bar{\mathbf{Y}}), \quad \text{where} \quad \bar{\mathbf{Y}} = \mathbf{A}_G \mathbf{X}_I \mathbf{X}_C^\top \mathbf{A}_R. \quad (5.2)$$

The expression in the right-hand side of (5.2), $\text{vec}(\mathbf{A}_G \mathbf{X}_I \mathbf{X}_C^\top \mathbf{A}_R)$ comes from the property of Kronecker matrix-vector product, $(\mathbf{A} \otimes \mathbf{B}) \text{vec}(\mathbf{V}) = \text{vec}(\mathbf{B} \mathbf{V} \mathbf{A}^\top)$ [112].

Note that $\mathbf{X}_I \mathbf{X}_C^\top$ contains background radiation contributions. In many radiation mapping scenarios, spatial distribution of background radiation may not be of primary interest. Therefore, the background contributions can be modelled separately with another low-rank decomposition term, $\mathbf{X}_{BW} \mathbf{X}_{BC}^\top$. Analogous to the decomposition of $\mathbf{X} = \mathbf{X}_I \mathbf{X}_C^\top$, the columns of $\mathbf{X}_{BC} \in \mathbb{R}_{\geq 0}^{L \times B}$ represent the recorded background spectral components, and the columns of $\mathbf{X}_{BW} \in \mathbb{R}_{\geq 0}^{I \times B}$ correspond to the background count contributions (weights) from the associated background components.

$$\mathbf{Y} \sim \text{Poisson}(\bar{\mathbf{Y}}), \quad \text{where} \quad \bar{\mathbf{Y}} = \mathbf{A}_G \mathbf{X}_I \mathbf{X}_C^\top \mathbf{A}_R + \mathbf{X}_{BW} \mathbf{X}_{BC}^\top \quad (5.3)$$

Using the Poisson probability distribution function (PDF), the likelihood function can be constructed as

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^I \prod_{j=1}^J \exp(-\bar{Y}_{ij}) \bar{Y}_{ij}^{\mathbf{Y}_{ij}} / \mathbf{Y}_{ij}!, \quad (5.4)$$

where we define $\mathbf{X} := \{\mathbf{X}_I, \mathbf{X}_C, \mathbf{X}_{BW}, \mathbf{X}_{BC}\}$ to be a collection of all parameter matrices with a slight abuse of notation. With the likelihood function, the MLE problem can be formulated as

$$\begin{aligned} \hat{\mathbf{X}} (= \{\hat{\mathbf{X}}_I, \hat{\mathbf{X}}_C, \hat{\mathbf{X}}_{BW}, \hat{\mathbf{X}}_{BC}\}) &= \underset{\mathbf{X}, \mathbf{X} \succeq 0}{\operatorname{argmin}} -\log p(\mathbf{Y}|\mathbf{X}) \\ &= \underset{\mathbf{X}, \mathbf{X} \succeq 0}{\operatorname{argmin}} l(\mathbf{X}; \mathbf{Y}) \\ &= \underset{\mathbf{X}, \mathbf{X} \succeq 0}{\operatorname{argmin}} \sum_{i=1}^I \sum_{j=1}^J \bar{\mathbf{Y}}_{ij} - \mathbf{Y}_{ij} \log \bar{\mathbf{Y}}_{ij}, \end{aligned} \quad (5.5)$$

where $l(\mathbf{X}; \mathbf{Y})$ is the negative log-likelihood. The MLE problem can be solved using the block coordinate descent algorithm [113] citation, which has been previously used for non-negative matrix factorization (NMF) [114][115]. The iterative update rules are given as,

$$\begin{aligned} \mathbf{X}_I &\leftarrow \mathbf{X}_I \odot \frac{\mathbf{A}_G^\top (\mathbf{Y} \oslash \bar{\mathbf{Y}}) \mathbf{A}_R^\top \mathbf{X}_C^\top}{\mathbf{A}_G^\top \mathbf{1}_{I \times J} \mathbf{A}_R^\top \mathbf{X}_C^\top}, & \mathbf{X}_C &\leftarrow \mathbf{X}_C \odot \frac{\mathbf{X}_I^\top \mathbf{A}_G^\top (\mathbf{Y} \oslash \bar{\mathbf{Y}}) \mathbf{A}_R^\top}{\mathbf{X}_I^\top \mathbf{A}_G^\top \mathbf{1}_{I \times J} \mathbf{A}_R^\top}, \\ \mathbf{X}_{BW} &\leftarrow \mathbf{X}_{BW} \odot \frac{(\mathbf{Y} \oslash \bar{\mathbf{Y}}) \mathbf{X}_{BC}^\top}{\mathbf{1}_{I \times J} \mathbf{X}_{BC}^\top}, & \mathbf{X}_{BC} &\leftarrow \mathbf{X}_{BC} \odot \frac{\mathbf{X}_{BW}^\top (\mathbf{Y} \oslash \bar{\mathbf{Y}})}{\mathbf{X}_{BW}^\top \mathbf{1}_{I \times J}}, \end{aligned} \quad (5.6)$$

where $\mathbf{1}_{I \times J}$ denotes an $I \times J$ matrix of unities, and $\odot$ and $\oslash$ are element-wise multiplication and division, respectively.

The iterative update rule is guaranteed to converge to a minimum [114]. However, even with the decomposed form of $\mathbf{X}$, the problem is still highly ill-posed and therefore, the MLE approach suffers from overfitting.

Gaussian Process Prior Reconstruction

To combat the ill-posedness of the MLE problem, we make use of the Bayesian formalism by incorporating Gaussian processes (GP) priors with appropriate link functions, as done similarly in Chapter 4. Using Bayes rule, the posterior distribution $p(\mathbf{X}|\mathbf{Y})$ is given as,

$$p(\mathbf{X}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})}. \quad (5.7)$$

The maximum *a posteriori* (MAP) estimation $\hat{\mathbf{X}}$ is a point estimate at which the posterior

density is maximized as follows:

$$\hat{\mathbf{X}} = \operatorname{argmax}_{\mathbf{X}, \mathbf{X} \succeq 0} p(\mathbf{X}|\mathbf{Y}) \quad (5.8)$$

$$= \operatorname{argmin}_{\mathbf{X}, \mathbf{X} \succeq 0} -\log p(\mathbf{X}|\mathbf{Y}) \quad (5.9)$$

$$= \operatorname{argmin}_{\mathbf{X}, \mathbf{X} \succeq 0} -\log p(\mathbf{Y}|\mathbf{X}) - \log p(\mathbf{X}) \quad (5.10)$$

$$= \operatorname{argmin}_{\mathbf{X}, \mathbf{X} \succeq 0} l(\mathbf{X}; \mathbf{Y}) + l_p(\mathbf{X}), \quad (5.11)$$

where $l_p(\mathbf{X})$ is the negative log prior distributions. Next, we assume the parameters to be estimated, $\mathbf{X}_I, \mathbf{X}_C, \mathbf{X}_{BW}$, and $\mathbf{X}_{BC}$ are governed by underlying variables of same size, denoted $\mathbf{\Xi}_I, \mathbf{\Xi}_C, \mathbf{\Xi}_{BW}, \mathbf{\Xi}_{BC}$. The underlying variables are related to the original variables of interest via inverse link functions $\mathbf{f}^{-1}(\cdot)$, such that

$$\begin{aligned} \mathbf{f}_I^{-1}(\mathbf{\Xi}_I) &= \mathbf{X}_I & \mathbf{f}_C^{-1}(\mathbf{\Xi}_C) &= \mathbf{X}_C \\ \mathbf{f}_{BW}^{-1}(\mathbf{\Xi}_{BW}) &= \mathbf{X}_{BW} & \mathbf{f}_{BC}^{-1}(\mathbf{\Xi}_{BC}) &= \mathbf{X}_{BC}. \end{aligned} \quad (5.12)$$

Next, we impose zero-mean GP priors over these underlying variables, such that

$$\begin{aligned} \operatorname{vec}(\mathbf{\Xi}_I) &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_I), & \operatorname{vec}(\mathbf{\Xi}_C^\top) &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_C) \\ \operatorname{vec}(\mathbf{\Xi}_{BW}) &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{BW}), & \operatorname{vec}(\mathbf{\Xi}_{BC}^\top) &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{BC}). \end{aligned} \quad (5.13)$$

The covariance matrices $\mathbf{\Sigma}_I, \mathbf{\Sigma}_C, \mathbf{\Sigma}_{BW}$ and $\mathbf{\Sigma}_{BC}$ are populated using either the squared exponential (SE) kernel or the rational quadratic (RQ) kernel

$$\mathbf{\Sigma}_{ij}(\text{SE}) = \sigma^2 \exp\left(-\frac{\|\mathbf{r}_i - \mathbf{r}_j\|^2}{2l^2}\right) \quad (5.14)$$

$$\mathbf{\Sigma}_{ij}(\text{RQ}) = \sigma^2 \left(1 + \frac{\|\mathbf{r}_i - \mathbf{r}_j\|^2}{2\alpha l^2}\right)^{-\alpha}, \quad (5.15)$$

where $\|\mathbf{r}_i - \mathbf{r}_j\|$ is an Euclidean distance between two parameters indexed by i and j . The Euclidean distances are measured in terms of different physical quantities, depending on the variable matrix; for example, in the case of image matrix $\mathbf{\Xi}_I$, vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ denote cartesian coordinates of i -th and j -th image pixels. For the background counts contribution $\mathbf{X}_{BW}$, $\mathbf{r}_i$ and $\mathbf{r}_j$ represent timestamp of i -th and j -th measurements. For the component matrices $\mathbf{X}_C$ and $\mathbf{X}_{BC}$ the distance is difference in energy bin number, $|i - j|^1$. Additionally, we assume independence between columns of $\mathbf{\Xi}$, such that $\mathbf{\Sigma}_{ij} = 0$ if i and j are from different columns of $\mathbf{\Xi}$. This independence assumption leads to a block diagonal structure

¹More intuitive choice may be difference in bin center energy. However, radiation detectors typically exhibit $\sqrt{E}$ energy resolution dependence and therefore, we discretize the energy domain such that bin widths are proportional to $\sqrt{E}$, yielding constant energy resolution dependence in bin number units.

for Σ , where each block represents a covariance matrix for a column of Ξ . In conclusion, the covariance kernels encode spatial, temporal or spectral correlations between parameters, with the hyperparameters σ , l and α determining characteristic correlations within components.

Once the underlying variables are defined, inverse link functions $\mathbf{f}^{-1}(\cdot)$ are employed to relate the underlying variables to the original variables of interest. To satisfy the non-negativity constraints, the inverse link functions $\mathbf{f}^{-1}(\cdot)$ are required to be $\mathbf{f}^{-1}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$. Any functions satisfying the non-negativity constraints are valid choices for link function. In this work, we consider several link functions based on physical, and statistical characteristics of the variables.

The inverse link functions are required to map $\mathbf{f}^{-1} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, so that the underlying variables are unconstrained and the non-negativity constraint on $\mathbf{X}$ is automatically satisfied. Therefore, any functions with non-negative outputs can serve as inverse link functions. For instance, a simple exponential function, $\mathbf{f}^{-1}(\cdot) = \exp(\cdot)$ is a valid inverse link function. However, in this work, inverse link functions are chosen to reflect physical and statistical characteristics of the variables. For example, elements of the image matrix $\mathbf{X}_I$ and the background weight matrix $\mathbf{X}_{BW}$ are expected to vary widely across orders of magnitude. Therefore, we impose the inverse log link function for $\mathbf{X}_I$ and $\mathbf{X}_{BW}$ as follows:

$$\mathbf{X} = \mathbf{f}^{-1}(\Xi) = \frac{1}{\lambda} \exp(\Xi). \quad (5.16)$$

Notably, with the inverse log link function, transformation of the Gaussian variables Ξ yields $\mathbf{X}$ following log-normal distribution, whose properties are well-understood [116].

For the spectral component matrices, $\mathbf{X}_C$ and $\mathbf{X}_{BC}$, we use the softmax link function ,

$$\mathbf{X}[i, j] = \frac{\exp(\Xi[i, j]/T)}{\sum_{i=1}^I \exp(\Xi[i, j]/T)}, \quad (5.17)$$

where T is temperature hyperparameter, tuning the “peakedness” of the output vector [117]. The parameter plays an important role, as different radiation spectral components contain various combinations of photopeaks and smooth downscattered continua. The softmax link function normalizes the outputs such that the column-wise sum of $\mathbf{X}$ is unity. This property is particularly attractive for spectral component matrices, as each spectral component can be interpreted as a probability density function (pdf). Moreover, the normalization of spectral component matrices enhances interpretability of the image matrix $\mathbf{X}_I$ and the background weights matrix $\mathbf{X}_{BW}$, as their units become *radiation emission rate per unit time* and *background detection rate per unit time*, respectively.

With the GPs and link functions defined, the forward projection $\bar{\mathbf{Y}}$ in terms of Ξ is written as

$$\bar{\mathbf{Y}} = \mathbf{A}_G \mathbf{f}_I^{-1}(\Xi_I) \mathbf{f}_C^{-1}(\Xi_C) \mathbf{A}_R + \mathbf{f}_{BW}^{-1}(\Xi_{BW}) \mathbf{f}_{BC}^{-1}(\Xi_{BC}). \quad (5.18)$$

Because the prior distributions are defined on Ξ , $p(\mathbf{X})$ can be formulated with the change of variable theorem. However, it involves determinant of a Jacobian matrix, $|\mathbf{J}_{\Xi \mathbf{X}}|$, which

is costly to compute in general. As such, we formulate a MAP problem in terms of Ξ and solve the following optimization problem:

$$\hat{\Xi} = \underset{\Xi}{\operatorname{argmin}} l(\Xi; \mathbf{Y}) + l_p(\Xi). \quad (5.19)$$

Note that unlike (5.1), the above optimization problem is not subject to non-negativity constraints. The likelihood function $l(\Xi; \mathbf{Y})$ is unchanged from (5.5), and $l_p(\Xi) \triangleq \frac{1}{2} \Xi^\top \Sigma^{-1} \Xi$. Hence, the MAP estimation problem becomes

$$\begin{aligned} \hat{\Xi} &= \underset{\Xi}{\operatorname{argmin}} l(\Xi; \mathbf{Y}) + l_p(\Xi) \\ &= \underset{\Xi}{\operatorname{argmin}} \sum_{i=1}^I \sum_{j=1}^J \bar{\mathbf{Y}}_{ij} - \mathbf{Y}_{ij} \log \bar{\mathbf{Y}}_{ij} + \frac{1}{2} \Xi^\top \Sigma^{-1} \Xi, \\ &= \underset{\Xi}{\operatorname{argmin}} \Psi(\Xi) \end{aligned} \quad (5.20)$$

where we define $\Sigma = \{\Sigma_I, \Sigma_C, \Sigma_{BW}, \Sigma_{BC}\}$ for notational brevity. Therefore, any operations on Σ is performed variable-wise. Additionally, the negative log-posterior is parametrized with $\Psi(\Xi)$.

While the optimization problem (5.20) can be solved using a gradient-based optimization algorithm, better numerical stability and accelerated convergence is achieved through a change of variable

$$\mathbf{L} \Xi_w = \Xi, \quad (5.21)$$

where $\mathbf{L}$ is a Cholesky decomposition of Σ . Consequently, Ξ_w is an uncorrelated standard Gaussian variable. Rewriting (5.20) in terms of Ξ_w ,

$$\begin{aligned} \hat{\Xi}_w &= \underset{\Xi_w}{\operatorname{argmin}} \Psi(\mathbf{L} \Xi_w) \\ &= \underset{\Xi_w}{\operatorname{argmin}} \sum_{i=1}^I \sum_{j=1}^J \bar{\mathbf{Y}}_{ij} - \mathbf{Y}_{ij} \log \bar{\mathbf{Y}}_{ij} + \frac{1}{2} \Xi_w^\top \Xi_w \end{aligned} \quad (5.22)$$

It should be noted that the optimization problem (5.22) is not convex and therefore, the solution is not unique. For example, trivially permuting the sequence of source components and/or the background components yields another solution with the equivalent log posterior value. We found that a gradient descent algorithm such as ADAM or quasi-Newton algorithm such as Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) achieves swift convergence to a good minimum [79]. The gradient required for these optimization algorithms is obtained through automatic differentiation [118].

Once $\hat{\Xi}_w$ is found, the parameters of interest can be obtained by transforming $\hat{\mathbf{X}} = \mathbf{f}^{-1}(\mathbf{L} \hat{\Xi}_w)$. The pseudocode for the MAP estimation is given in Algorithm 3.

Algorithm 3 Full-spectral MAP estimation with a GP prior

Require: $\mathbf{A}_G, \mathbf{A}_R, \mathbf{Y}, \boldsymbol{\theta}$ (hyperparameters), $\boldsymbol{\Xi}_w^{\text{init}}, \epsilon$ (Convergence criterion)

$\boldsymbol{\Sigma} \leftarrow (4.11)$ ▷ Prior covariance matrix construction

$\mathbf{L} \leftarrow \text{Cholesky}(\boldsymbol{\Sigma})$ ▷ Cholesky decomposition

$k \leftarrow 0$

$\boldsymbol{\Xi}_w^k \leftarrow \boldsymbol{\Xi}_w^{\text{init}}$ ▷ Initialization

$\Psi^k \leftarrow \Psi(\mathbf{L}\boldsymbol{\Xi}_w^k)$ ▷ Initialization

Convergence $\leftarrow$ False

while Convergence is False **do**

$\boldsymbol{\Xi}_w^{k+1} \leftarrow \text{Optim. step}(\boldsymbol{\Xi}_w^k, (5.22))$ ▷ One iteration of ADAM or L-BFGS

$\Psi^{k+1} \leftarrow \Psi(\mathbf{L}\boldsymbol{\Xi}_w^{k+1})$

if $|\Psi^k - \Psi^{k+1}| < \epsilon$ **then**

 Convergence $\leftarrow$ True

else

$k \leftarrow k + 1$

end if

end while

$\hat{\boldsymbol{\Xi}}_w \leftarrow \boldsymbol{\Xi}_w^{k+1}$

$\hat{\boldsymbol{\Xi}} \leftarrow \mathbf{L}\hat{\boldsymbol{\Xi}}_w$

$\hat{\mathbf{X}} \leftarrow \mathbf{f}^{-1}(\hat{\boldsymbol{\Xi}})$ ▷ Apply inverse link functions from (5.12)

$\hat{\Psi} \leftarrow \Psi^{k+1}$ ▷ Negative log-posterior at the mode

Uncertainty Quantification

Due to the ill-posedness of the problem combined with the noisy nature of the measurement data $\mathbf{Y}$, uncertainty estimation in the MAP solution $\hat{\mathbf{X}}$ is highly desirable. In this work, we show that the Bayesian treatment in this work further lends itself to Bayesian uncertainty quantification of estimated parameters. In Bayesian statistics, the uncertainties of the parameter $\mathbf{X}$ are often quoted with Bayesian credible intervals. Given a marginal posterior distribution $p(\mathbf{X}[i]|\mathbf{Y})$, the $(1 - \alpha)100\%$ Bayesian credible interval is defined as $p(\mathbf{X} \in [\mathbf{X}[i]_-, \mathbf{X}[i]_+ | \mathbf{Y}) = 1 - \alpha$. In other words, the probability that $\mathbf{X}[i]$ falls into the interval is $(1 - \alpha)$. Note that such interval $[\mathbf{X}[i]_-, \mathbf{X}[i]_+]$ is not unique; therefore, we use equal-tail interval defined as

$$\frac{\alpha}{2} = \int_{-\infty}^{\mathbf{X}[i]_-} p(\mathbf{X}[i]|\mathbf{Y}) d\mathbf{X}[i] \quad \text{and} \quad (5.23)$$

$$\frac{\alpha}{2} = \int_{\mathbf{X}[i]_+}^{\infty} p(\mathbf{X}[i]|\mathbf{Y}) d\mathbf{X}[i]. \quad (5.24)$$

Analytical computation of the credible interval $[\mathbf{X}[i]_-, \mathbf{X}[i]_+]$ first requires posterior density function $p(\mathbf{X}[i]|\mathbf{Y})$, which is computed as,

$$p(\mathbf{X}[i]|\mathbf{Y}) = \int_{\mathbf{X}[j]} p(\mathbf{X}|\mathbf{Y}) d\mathbf{X}[j], \quad \text{for } j \neq i \quad (5.25)$$

$$= \int_{\mathbf{X}[j]} \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})} d\mathbf{X}[j], \quad \text{for } j \neq i \quad (5.26)$$

$$= \int_{\mathbf{X}[j]} \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{\int_{\mathbf{X}} p(\mathbf{Y}|\mathbf{X})p(\mathbf{X}) d\mathbf{X}} d\mathbf{X}[j], \quad \text{for } j \neq i \quad (5.27)$$

Therefore, computation of the credible interval involves high-dimensional integrations, which are generally intractable unless a conjugate prior is used. In this work, we use the Laplace approximation (LA) to find an approximate posterior distribution, enabling analytical computation of credible intervals [119].

Laplace Approximation (LA)

The LA aims to directly approximate the intractable posterior with a multivariate Gaussian distribution [120]. With the mode of the posterior (the MAP estimation) $\hat{\Xi}$ found from Algorithm 3, the Gaussian posterior approximation of $p(\Xi|\mathbf{Y})$ is given as,

$$p(\Xi|\mathbf{Y}) \approx q(\Xi|\mathbf{Y}), \quad \text{where} \quad (5.28)$$

$$q(\Xi|\mathbf{Y}) \sim \mathcal{N}(\hat{\Xi}, \Sigma_{\text{LA}} = (\nabla_{\Xi}^2 \Psi(\Xi)|_{\Xi=\hat{\Xi}})^{-1}). \quad (5.29)$$

The derivation of (5.29) is analogous to the mono-energetic GPP presented in chapter 4.5, and thus omitted here.

Once the approximating posterior $q(\Xi|\mathbf{Y})$ is obtained, the credible intervals can be found either directly or from samples of $q(\Xi|\mathbf{Y})$, depending on the link functions used. For example, the inverse log link function in (5.16) is a monotonically increasing function acting on input variables element-wise. As a result, the credible intervals in Ξ can be directly transformed into that in $\mathbf{X}$, such that $[\mathbf{X}[i]_-, \mathbf{X}[i]_+] = [\mathbf{f}^{-1}(\Xi[i]_-), \mathbf{f}^{-1}(\Xi[i]_+)]$. However, the softmax link function in (5.17) takes a multi-variate input, and it is not monotonically increasing. As a result, computing credible intervals in $\mathbf{X}$ directly from $q(\Xi|\mathbf{Y})$ is not straightforward with the softmax link function. Therefore, we use a Monte-Carlo estimation instead. From a set of samples drawn from $q(\Xi|\mathbf{Y})$, $\{\Xi_i\}_{i=1}^M$, the credible intervals are estimated as,

$$[\mathbf{X}_{i-}, \mathbf{X}_{i+}] \approx \text{quantile}\left(\{\mathbf{f}^{-1}(\Xi_i)\}_{i=1}^M, \left\{\frac{\alpha}{2}, 1 - \frac{\alpha}{2}\right\}\right), \quad (5.30)$$

where $\text{quantile}(\cdot)$ computes a quantile from a set of samples. In practice, $M \approx 10^4$ samples ensures a good estimation within $\sim 1\%$ error. Samples are obtained from a matrix square-root of the LA covariance matrix $(\Sigma_{\text{LA}}^{\frac{1}{2}})(\Sigma_{\text{LA}}^{\frac{1}{2}})^{\top} = \Sigma_{\text{LA}}$ as,

$$\Xi \sim \Sigma_{\text{LA}}^{\frac{1}{2}} \mathbf{z} + \hat{\Xi}, \quad \text{where} \quad (5.31)$$

$$\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}). \quad (5.32)$$

Computation of the LA covariance matrix $\Sigma_{\text{LA}} = (\nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}} + \Sigma^{-1})^{-1}$ requires Hessian of the likelihood function $\nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}}$ and large matrix inversions. For small scale problems, the Hessian matrix $\nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}}$ can be obtained from automatic differentiation, and the block diagonal structure of Σ^{-1} allows for efficient matrix inversion. However, the overall computation and memory cost scales with $\mathcal{O}(n^3)$ and $\mathcal{O}(n^2)$, respectively. Therefore, with a growing number of parameters, computation of the LA covariance matrix becomes prohibitively expensive. In section 5.1, we detail how the covariance matrix can be computed efficiently with a Fisher information matrix and randomized linear algebra.

Hyperparameter Optimization with Empirical Bayes

Various Gaussian process priors and link functions imposed on variables introduce a handful of hyperparameters that need to be determined prior to the MAP estimation of $\hat{\Xi}$. Although the highly interpretable nature of hyperparameters allows for selection of reasonable hyperparameters, a more systematic approach to automatically tuning hyperparameters is highly desirable. Here, we present automatic hyperparameter selection method using empirical Bayes [121].

With a vector $\boldsymbol{\theta}$ denoting all hyperparameters, the Bayes rule in (5.7) is rewritten as

$$p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}) = \frac{p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})}{p(\mathbf{Y}|\boldsymbol{\theta})} \quad (5.33)$$

$$= \frac{p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta})}{\int p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta})p(\mathbf{X}|\boldsymbol{\theta}) d\mathbf{X}}. \quad (5.34)$$

The quantity in the denominator $p(\mathbf{Y}|\boldsymbol{\theta})$ is often referred to as marginal likelihood, or model evidence. Intuitively, it encodes the average probability of a random sample from the prior $p(\mathbf{X}|\boldsymbol{\theta})$ gives rise to the measurement $\mathbf{Y}$. For example, when $\boldsymbol{\theta}$ imposes overly strong constraints on the prior distribution (e.g., large length scale parameter l), samples from the prior will be largely similar to each other, yet insufficiently complex enough to yield the measurement $\mathbf{Y}$. Conversely, too weak constraints on $\boldsymbol{\theta}$ will result in diverse samples of $\mathbf{X}$, yet most of them are unlikely to generate the measurement $\mathbf{Y}$ through forward projection. Therefore, with overly strong or weak hyperparameters yield low marginal likelihood value, $p(\mathbf{X}|\boldsymbol{\theta})$.

Consequently, the marginal likelihood $p(\mathbf{X}|\boldsymbol{\theta})$ is maximized only when the appropriate level of constraint is imposed through $\boldsymbol{\theta}$. Thus, the empirical Bayes method tunes the hyperparameters $\boldsymbol{\theta}$ by solving the following MLE problem,

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} -\log p(\mathbf{Y}|\boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \Phi(\boldsymbol{\theta}) \end{aligned} \quad (5.35)$$

where we defined $\Phi(\boldsymbol{\theta}) \stackrel{c}{=} -\log p(\mathbf{Y}|\boldsymbol{\theta})$ to be an unnormalized negative log-marginal likelihood.

Computation of marginal likelihood requires a very high dimensional integration over all parameters $\mathbf{X}$, which is intractable in general. Hence, we again use the Laplace approximation for a good approximation of $\Phi(\boldsymbol{\theta})$ in a computationally tractable manner.

First, the marginal likelihood $p(\mathbf{Y}|\boldsymbol{\theta})$ is given as,

$$p(\mathbf{Y}|\boldsymbol{\theta}) = \int \exp(\log p(\mathbf{Y}|\boldsymbol{\Xi})p(\boldsymbol{\Xi}|\boldsymbol{\theta})) d\boldsymbol{\Xi} \quad (5.36)$$

With Laplace approximation, $\log p(\mathbf{Y}|\boldsymbol{\Xi})p(\boldsymbol{\Xi}|\boldsymbol{\theta})$ term in the exponent is approximated using the second-order Taylor expansion around the mode of the posterior $\hat{\boldsymbol{\Xi}}$ as follows:

$$\begin{aligned} \log p(\mathbf{y}|\boldsymbol{\Xi})p(\boldsymbol{\Xi}|\boldsymbol{\theta}) &\approx \\ \log p(\mathbf{y}|\hat{\boldsymbol{\Xi}})p(\hat{\boldsymbol{\Xi}}|\boldsymbol{\theta}) - \frac{1}{2} \left(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \right)^\top \nabla_{\boldsymbol{\Xi}}^2 \Psi(\boldsymbol{\Xi})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}} \left(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \right). \end{aligned} \quad (5.37)$$

With the approximation, the marginal likelihood is

$$p(\mathbf{y}|\boldsymbol{\theta}) \approx \int \exp \left(-\frac{1}{2} \left(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \right)^\top \nabla_{\boldsymbol{\Xi}}^2 \Psi(\boldsymbol{\Xi})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}} \left(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \right) \right) d\boldsymbol{\Xi} \quad (5.38)$$

$$= \left((2\pi)^N \left| (\nabla_{\boldsymbol{\Xi}}^2 \Psi(\boldsymbol{\Xi})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}})^{-1} \right| \right)^{\frac{1}{2}} \quad (5.39)$$

Noting the similarity between the integrand in (5.38) and a multivariate Gaussian distribution $\mathcal{N}(\hat{\boldsymbol{\Xi}}, \nabla_{\boldsymbol{\Xi}}^2 \Psi(\boldsymbol{\Xi}))$, the integration can be analytically computed. Hence, the negative log-marginal likelihood $\Phi(\boldsymbol{\theta}) \triangleq -\log p(\mathbf{y}|\boldsymbol{\theta})$ is given as,

$$\Phi(\boldsymbol{\theta}) \approx \Phi_{\text{LA}}(\boldsymbol{\theta}, \hat{\boldsymbol{\Xi}}) \quad (5.40)$$

$$= \Psi(\hat{\boldsymbol{\Xi}}) + \frac{1}{2} \log \left| (\nabla_{\boldsymbol{\Xi}}^2 \Psi(\boldsymbol{\Xi})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}}) \right| + \frac{1}{2} \log |\boldsymbol{\Sigma}| \quad (5.41)$$

$$= \Psi(\hat{\boldsymbol{\Xi}}) + \frac{1}{2} \left(\log |\mathbf{L}^\top| + \log \left| (\nabla_{\boldsymbol{\Xi}}^2 l(\boldsymbol{\Xi}; \mathbf{y})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}} + \boldsymbol{\Sigma}^{-1}) \right| + \log |\mathbf{L}| \right) \quad (5.42)$$

$$= \Psi(\hat{\boldsymbol{\Xi}}) + \frac{1}{2} \log \left| \mathbf{L}^\top \nabla_{\boldsymbol{\Xi}}^2 l(\boldsymbol{\Xi}; \mathbf{y})|_{\boldsymbol{\Xi}=\hat{\boldsymbol{\Xi}}} \mathbf{L} + \mathbf{I} \right| \quad (5.43)$$

$$= \Psi(\hat{\boldsymbol{\Xi}}) + \frac{1}{2} \log \left| \nabla_{\boldsymbol{\Xi}_w}^2 l(\mathbf{L}\boldsymbol{\Xi}_w; \mathbf{y})|_{\boldsymbol{\Xi}_w=\hat{\boldsymbol{\Xi}}_w} + \mathbf{I} \right|. \quad (5.44)$$

Note that the Hessian $\nabla_{\boldsymbol{\Xi}}^2 l(\boldsymbol{\Xi}; \mathbf{y})$ is computed with automatic differentiation. The term $\frac{1}{2} \log |\boldsymbol{\Sigma}|$ comes from the GP negative log-prior distribution $-\log p(\boldsymbol{\Xi}|\boldsymbol{\theta})$. The term is irrelevant for the MAP estimation in (5.20) as it does not contain any parameters $\boldsymbol{\Xi}$. However, for hyperparameter optimization, it needs to be included as it is a function of hyperparameters $\boldsymbol{\theta}$.

An important observation here is that the MAP estimation $\hat{\boldsymbol{\Xi}}$ used for Laplace approximation is a function of hyperparameters $\boldsymbol{\theta}$. Therefore, the optimization problem (5.35) is

re-written as a bi-level optimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \Phi(\boldsymbol{\theta}) \approx \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \Phi_{\text{LA}}(\boldsymbol{\theta}, \hat{\Xi}), \quad \text{where} \quad (5.45)$$

$$\hat{\Xi} = \underset{\Xi}{\operatorname{argmin}} \Psi(\Xi|\boldsymbol{\theta}), \quad (5.46)$$

where $\Psi(\hat{\Xi}|\boldsymbol{\theta})$ is defined in (5.20), but its dependency on $\boldsymbol{\theta}$ is made more explicit. The optimization problem (5.45) can be solved with a gradient based optimization algorithm. The necessary gradient $\nabla_{\boldsymbol{\theta}}\Phi_{\text{LA}}(\boldsymbol{\theta})$ is decomposed into:

$$\boldsymbol{d}_{\boldsymbol{\theta}}\Phi_{\text{LA}}(\boldsymbol{\theta}|\hat{\Xi}) = \nabla_{\boldsymbol{\theta}}\Phi_{\text{LA}}(\boldsymbol{\theta}|\hat{\Xi}) + \mathbf{J}_{\boldsymbol{\theta}}\hat{\Xi} \cdot \nabla_{\hat{\Xi}}\Phi_{\text{LA}}(\boldsymbol{\theta}|\hat{\Xi}), \quad (5.47)$$

where $\boldsymbol{d}$ is used to denote total derivative. Computation of $\nabla_{\boldsymbol{\theta}}\Phi_{\text{LA}}(\boldsymbol{\theta}|\hat{\Xi})$ and $\nabla_{\hat{\Xi}}\Phi_{\text{LA}}(\boldsymbol{\theta}|\hat{\Xi})$ is straightforward with automatic differentiation. However, automatic differentiation of $\mathbf{J}_{\boldsymbol{\theta}}\hat{\Xi}$ requires caching gradients through the entire optimization iterations of Algorithm 3, which is memory intensive and may be impossible, when the optimization algorithm incorporates non-differentiable operations, such as line search. Hence, we invoke the Implicit Function Theorem (IFT), and rewrite the Jacobian as follows:

$$\mathbf{J}_{\boldsymbol{\theta}}\hat{\Xi} = -\nabla_{\Xi\Xi}^2\Psi(\Xi)^{-1} \cdot \nabla_{\Xi\boldsymbol{\theta}}^2\Psi(\Xi). \quad (5.48)$$

The IFT allows us to compute the Jacobian $\mathbf{J}_{\boldsymbol{\theta}}\hat{\Xi}$ by solving the systems of linear equations $\nabla_{\Xi\Xi}^2\Psi(\Xi) \cdot \mathbf{J}_{\boldsymbol{\theta}}\hat{\Xi} = -\nabla_{\Xi\boldsymbol{\theta}}^2\Psi(\Xi)$, without having to propagate the gradient through Algorithm 3.

Inversion of the Hessian matrix $\nabla_{\Xi\Xi}^2\Psi(\Xi)^{-1}$ is computationally expensive with $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$ memory and computational complexity. However, its computational cost can be greatly eased with an iterative approximate solvers, such as conjugate gradient (CG). Note thatt the use of CG in conjunction with the IFT has been studied in hyperparameter selection for large neural networks [122][123][124], and some Python packages provide readily available implementations for gradient computation using the IFT [125].

In addition to the optimization of hyperparameters $\boldsymbol{\theta}$, the low-rank structures in the problem formulation (5.3) naturally raises the question of how to select the number of source and background components. Here we show that this model selection problem can also be addressed with the empirical Bayes method. Let $\boldsymbol{\theta}_{S,B}$, $\forall S, B = 1, 2, 3 \dots$ represent a set of hyperparameters with S and B number of source and background components. The optimal number of source and background components combination is then determined by solving the following bi-level optimization problem:

$$\hat{S}, \hat{B} = \underset{S,B}{\operatorname{argmin}} \Phi(\hat{\boldsymbol{\theta}}_{S,B}), \quad \text{where} \quad (5.49)$$

$$\hat{\boldsymbol{\theta}}_{S,B} = \underset{\boldsymbol{\theta}_{S,B}}{\operatorname{argmin}} \Phi(\boldsymbol{\theta}_{S,B}). \quad (5.50)$$

The solution to the inner optimization problem (5.50) is obtained from (5.45) for a given choice of S and B . Note that the outer optimization problem is not differentiable with respect to S and B due to their discrete nature. Consequently, the outer optimization problem is solved by evaluating $\Phi(\hat{\theta}_{S,B})$ for all possible combinations of S and B . Typically, both S and B do not exceed 3 for radiation spectroscopy applications. Therefore, in this work the outer optimization problem (5.49) was solved with a manageable number of manual trials.

Computationally Efficient Implementation

As with many Bayesian inverse problems, the MAP estimation, uncertainty quantification, and model selection methods presented in this work may pose significant computational challenges with a growing number of parameters to be estimated [126]. In particular, operations involving large matrices such as the covariance matrix Σ or the Hessian matrix $\nabla_{\Xi\Xi}^2 \Psi(\Xi)$ can be prohibitively expensive, even with modern computing power [127]. In this section, we discuss practical techniques to alleviate these computational bottlenecks, enhancing the scalability of the algorithms.

Efficient computations involving the covariance matrix

As previously mentioned, the covariance matrix Σ is a block diagonal matrix, where its submatrices represent the component-wise covariance matrices. Consequently, Σ is never stored explicitly as an $N \times N$ matrix; instead, it is sufficient to store its component-wise submatrices for all computations involving Σ . For instance, its Cholesky factor $\mathbf{L}$ is another block diagonal matrix, whose submatrices are the Cholesky factors computed from the submatrices of Σ . Similarly, operations such as the matrix-vector multiplication and $\mathbf{L}\Xi_w$ in (5.21) is performed component-wise, thereby avoiding the $\mathcal{O}(N^2)$ computational cost of the matrix-vector multiplication. As previously mentioned, the covariance matrix Σ is a block diagonal matrix, where its submatrices represent the component-wise covariance matrices. Consequently, Σ is never stored explicitly as an $N \times N$ matrix; instead, it is sufficient to store its component-wise submatrices for all computations involving Σ . For instance, its Cholesky factor $\mathbf{L}$ is another block diagonal matrix, whose submatrices are the Cholesky factors computed from the submatrices of Σ . Similarly, operations such as the matrix-vector multiplication and $\mathbf{L}\Xi_w$ in (5.21) is performed component-wise, thereby avoiding the $\mathcal{O}(N^2)$ computational cost associated with the matrix-vector multiplication.

When performing operations on the component-wise covariance matrix, the computation bottleneck arises from the largest component vectors. In this work, the largest component vectors are the source image vectors in Ξ_I , whose dimensions ranges from 10^3 to 10^5 . As a result, the image vectors alone can significantly complicate computations involving the covariance matrix.

The computational cost associated with the source image vector covariance matrices is significantly reduced with the Kronecker product representation of the covariance matrices [80]. When GP inputs are defined on a rectilinear D -dimensional Cartesian grid (e.g., 2D

pixel image space), the covariance matrix can be expressed as a Kronecker product of covariance matrices along each Cartesian dimension. Furthermore, its Cholesky factor also has a Kronecker product structure, where its submatrices are the Cholesky factors for covariance matrix in each dimension. Consequently, computational burden of storing and constructing the Cholesky factor is greatly eased to $\mathcal{O}(DK^{2/D})$ memory and $\mathcal{O}(DK^{3/D})$ computational cost. Additionally, the Kronecker product structured Cholesky factor admits rapid matrix-vector multiplication with $\mathcal{O}(DK^{\frac{1+D}{D}})$ cost, represented as a tensor outer product.

With the source image covariance matrix represented as a Kronecker product, the overall computational cost of the Algorithm 3 is $\mathcal{O}(SIJKL + BIL)$, which is incurred from the forward projection (5.3).

Randomized linear algebra for scalable inverse Hessian and log-determinant computation

The main computational bottlenecks in the LA covariance (5.29) and the negative log-marginal likelihood (5.40) are the inversion of the Hessian matrix and the computation of log-determinant, as a naive approach incurs a computational cost of $\mathcal{O}(N^3)$ and a memory cost of $\mathcal{O}(N^2)$. As a result, with an increasing number of parameters N , the uncertainty quantification and hyperparameter optimization methods introduced in this work becomes prohibitively expensive.

In this work, we use randomized linear algebra (RLA) techniques to enable scalable computation of the Hessian matrix inversion and the log-determinant [128]. RLA methods summarizes the dominant subspace of a matrix by projecting it onto a set of random vectors. This projection results in a matrix much smaller than the original, while preserving its essential subspace structures. Consequently, algebraic operations performed on this smaller matrix yields a highly accurate approximation of the corresponding operations on the original matrix. In this work, we make use of the randomized eigenvalue decomposition (rEVD) and stochastic Lanczos quadrature for the computation of the inverse Hessian $(\nabla_{\Xi}^2 \Psi(\Xi)|_{\Xi=\hat{\Xi}})^{-1}$ and log-determinant $\log \left| \nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y})|_{\Xi_w=\hat{\Xi}_w} + \mathbf{I} \right|$, respectively.

First, the inverse Hessian matrix can be rewritten as

$$(\nabla_{\Xi}^2 \Psi(\Xi)|_{\Xi=\hat{\Xi}})^{-1} = (\nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}} + \Sigma^{-1})^{-1} \quad (5.51)$$

$$= (\nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}} + \mathbf{L}^{-\top} \mathbf{L}^{-1})^{-1} \quad (5.52)$$

$$= \mathbf{L} (\mathbf{L}^{\top} \nabla_{\Xi}^2 l(\Xi; \mathbf{y})|_{\Xi=\hat{\Xi}} \mathbf{L} + \mathbf{I})^{-1} \mathbf{L}^{\top} \quad (5.53)$$

$$= \mathbf{L} (\nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y})|_{\Xi_w=\hat{\Xi}_w} + \mathbf{I})^{-1} \mathbf{L}^{\top} \quad (5.54)$$

We then apply rEVD to the Hessian $\nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y})|_{\Xi_w=\hat{\Xi}_w}$ to obtain its approximate eigendecomposition $\nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y})|_{\Xi_w=\hat{\Xi}_w} \approx \mathbf{Q}\mathbf{S}\mathbf{Q}^{\top}$ where $\mathbf{Q} \in \mathbb{R}^{N \times R}$ and $\mathbf{S} \in \mathbb{R}^{R \times R}$ with $R \ll N$. The approximation is highly accurate, due to the rapid decay Hessian eigenvalues. The rEVD first proceeds by multiplying the input matrix by a random test matrix $\mathbf{\Omega} \in \mathbb{R}^{N \times K}$, a

procedure commonly known as *sketching*. Notably, sketching does not require explicit construction of the input matrix, which is memory-intensive. Instead, a matrix multiplication function, $\text{MatMul}(\cdot)$, suffices, taking input as the test matrix and returning the target matrix multiplication result. Thus, we have

$$\begin{aligned}\text{MatMul}(\mathbf{\Omega}) &= \nabla_{\mathbf{\Xi}_w}^2 l(\mathbf{L}\mathbf{\Xi}_w; \mathbf{y})|_{\mathbf{\Xi}_w=\hat{\mathbf{\Xi}}_w} \mathbf{\Omega} \\ &= \text{HMP}(\mathbf{\Omega}),\end{aligned}\tag{5.55}$$

where $\text{HMP}(\cdot)$ is the Hessian-matrix product operator. This operator represents a collection of Hessian-vector products of the Hessian $\nabla_{\mathbf{\Xi}_w}^2 l(\mathbf{L}\mathbf{\Xi}_w; \mathbf{Y})|_{\mathbf{\Xi}_w=\hat{\mathbf{\Xi}}_w}$ and the columns of $\mathbf{\Omega}$. Each Hessian-vector product is computed via reverse-mode automatic differentiation, which is readily available in modern computational frameworks such as PyTorch and TensorFlow [129][130]. Notably, computing each Hessian-vector product is only a constant-factor more expensive than evaluating $l(\mathbf{L}\mathbf{\Xi}_w; \mathbf{y})$ itself [122].

Equipped with the Hessian-matrix multiplication operator, the rest of rEVD follows Algorithm 4 in [128]. With the eigendecomposition we write

$$(\nabla_{\mathbf{\Xi}}^2 \Psi(\mathbf{\Xi})|_{\mathbf{\Xi}=\hat{\mathbf{\Xi}}})^{-1} = \mathbf{L} (\nabla_{\mathbf{\Xi}_w}^2 l(\mathbf{L}\mathbf{\Xi}_w; \mathbf{y})|_{\mathbf{\Xi}_w=\hat{\mathbf{\Xi}}_w} + \mathbf{I}_n)^{-1} \mathbf{L}^\top \tag{5.56}$$

$$\approx \mathbf{L} (\mathbf{Q}\mathbf{S}\mathbf{Q}^\top + \mathbf{I}_n)^{-1} \mathbf{L}^\top \tag{5.57}$$

$$= \mathbf{L} \underbrace{(\mathbf{I}_n - \mathbf{Q}(\mathbf{S}^{-1} + \mathbf{Q}^\top \mathbf{Q})^{-1} \mathbf{Q}^\top)}_{=\mathbf{B}\mathbf{B}^\top} \mathbf{L}^\top \quad (\because \text{Woodbury matrix identity}) \tag{5.58}$$

$$= \mathbf{L}\mathbf{B}\mathbf{B}^\top \mathbf{L}^\top, \tag{5.59}$$

where $\mathbf{B}$ is a matrix square root of $\mathbf{I}_n - \mathbf{Q}(\mathbf{S}^{-1} + \mathbf{Q}^\top \mathbf{Q})^{-1} \mathbf{Q}^\top$ and defined as

$$\mathbf{B} = \mathbf{I}_n + \mathbf{Q}((\mathbf{I}_m \oslash (\mathbf{S} + \mathbf{I}_m))^{\frac{1}{2}} - \mathbf{I})\mathbf{Q}^\top. \tag{5.60}$$

Therefore, the matrix square root of the LA covariance matrix $\mathbf{\Sigma}_{\text{LA}}^{\frac{1}{2}}$ is

$$\begin{aligned}\mathbf{\Sigma}_{\text{LA}}^{\frac{1}{2}} &= \mathbf{L}\mathbf{B} \\ &= \mathbf{L}(\mathbf{I}_n + \mathbf{Q}((\mathbf{I}_m \oslash (\mathbf{S} + \mathbf{I}_m))^{\frac{1}{2}} - \mathbf{I})\mathbf{Q}^\top)\end{aligned}\tag{5.61}$$

The matrix square root of the LA covariance matrix $\mathbf{\Sigma}_{\text{LA}}^{\frac{1}{2}}$ is used for sampling from the LA posterior distribution via its matrix-vector multiplication, as in (5.31). Note that $\mathbf{\Sigma}_{\text{LA}}^{\frac{1}{2}}$ does not need to be explicitly evaluated prior to posterior sampling. Instead, the low-rank structure of $\mathbf{Q}$, the Kronecker product structure of $\mathbf{L}$, and the diagonal structure of $(\mathbf{I}_m \oslash (\mathbf{S} + \mathbf{I}_m))^{\frac{1}{2}} - \mathbf{I}$ ensures efficient matrix-vector multiplications for sampling.

For efficient computation of the log-determinant $\log |\mathbf{L} \nabla_{\mathbf{\Xi}}^2 l(\mathbf{\Xi}; \mathbf{y})|_{\mathbf{\Xi}=\hat{\mathbf{\Xi}}} \mathbf{L}^\top + \mathbf{I}|$, we use the stochastic Lanczos quadrature (SLQ) [131][132][133][134]. Here we briefly introduce the

SLQ, and refer to for more detailed information [135]. First we note that for a symmetric positive definite matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ and its eigenvalues $s_1, s_2 \dots s_n$, computation of its log-determinant $\log |\mathbf{K}|$ can be re-written as,

$$\log |\mathbf{K}| = \log \left(\prod_{i=1}^N s_i \right) = \sum_{i=1}^N \log s_i = \text{Tr}(\log \mathbf{K}). \quad (5.62)$$

Therefore, the log-determinant is equivalent to the trace of $\log \mathbf{K}$.²

Exact estimation of the trace involves computing all eigenvalues of the matrix $\mathbf{K}$, which is expensive for large matrices. However, its Monte-Carlo approximation can be computed with the Hutchinson method, a stochastic trace estimator that only uses matrix-vector multiplication. In Hutchinson's method, $\text{Tr}(\mathbf{K})$ is estimated with N i.i.d. random vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_R$, with its mean $\mathbb{E}[\mathbf{v}] = \mathbf{0}$ and covariance $\mathbb{E}[\mathbf{v}\mathbf{v}^\top] = \mathbf{I}$ (typically Rademacher distribution with its entries being -1 or +1 with equal probability) as follows:

$$\text{Tr}(\mathbf{K}) = \frac{1}{R} \sum_{r=1}^R \mathbf{v}_r^\top \mathbf{K} \mathbf{v}_r. \quad (5.63)$$

Hence, the log-determinant is similarly computed as

$$\text{Tr}(\log \mathbf{K}) = \frac{1}{R} \sum_{r=1}^R \mathbf{v}_r^\top \log \mathbf{K} \mathbf{v}_r. \quad (5.64)$$

The scalar quantities $\mathbf{v}_r^\top \log \mathbf{K} \mathbf{v}_r$ is estimated using the Lanczos quadrature method. In the Lanczos quadratic method, the scalar quantities $\mathbf{v}_r^\top \log \mathbf{K} \mathbf{v}_r$ are first transformed into a Riemann-Stieltjes integral, which is approximated by the Gaussian quadrature rule. The nodes and weights required for the Gaussian quadrature rule are computed from the Lanczos algorithm. Lanczos algorithm generates an orthonormal basis $\mathbf{W}_{M+1}$ for the Krylov subspace $\text{Span}\{\mathbf{K}\mathbf{v}_r, \mathbf{K}^2\mathbf{v}_r, \dots, \mathbf{K}^M\mathbf{v}_r\}$. With the orthonormal basis, a tridiagonal matrix $\mathbf{T}_{M+1} = \mathbf{W}_{M+1}^\top \mathbf{K} \mathbf{W}_{M+1}$ is constructed. The nodes and weights of the Gaussian quadrature rule are estimated from the eigenvectors $q_1, q_2, \dots, q_{M+1}$ and eigenvalues $\lambda_1, \lambda_1, \lambda_2, \dots, \lambda_{M+1}$ of $\mathbf{T}_{M+1}$, such that the Riemann-Stieltjes integral is written as

$$\mathbf{v}_r^\top \log \mathbf{K} \mathbf{v}_r \approx \sum_{k=1}^{M+1} (\mathbf{e}_1^\top \mathbf{q}_k)^2 \log \lambda_k, \quad (5.65)$$

where $\mathbf{e}_1^\top \mathbf{q}_k$ is the first entry of the eigenvector $\mathbf{q}_k$. Using the Lanczos algorithm for matrix-vector multiplication, the Hutchinson estimation of the log-determinant (5.64) is given as,

$$\text{Tr}(\log \mathbf{K}) = \frac{1}{R} \sum_{r=1}^R \sum_{k=1}^{M+1} (\mathbf{e}_1^\top \mathbf{q}_k^n)^2 \log \lambda_k^n, \quad (5.66)$$

²In $\log \mathbf{K}$, note that the logarithm is a matrix function, and therefore it is applied to the eigenvalues of $\mathbf{K}$.

where $\mathbf{q}_k^n$ and λ_k^n denote the k -th eigenvector and eigenvalue of $\mathbf{v}_n$. Lanczos algorithm is numerically unstable, and several re-orthogonalization tricks have been developed for practical implementation. For more implementation information on stochastic Lanczos quadrature, see [135][136][137][138].

Applying the SLQ to compute $\log \left| \nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y}) \big|_{\Xi_w = \hat{\Xi}_w} + \mathbf{I} \right|$, we need to repeatedly multiply the random vectors $\mathbf{v}_i$ to $\nabla_{\Xi_w}^2 l(\mathbf{L}\Xi_w; \mathbf{y}) \big|_{\Xi_w = \hat{\Xi}_w} + \mathbf{I}$ to form the Krylov subspace. This matrix multiplication is computed without realizing the Hessian matrix making use of a matrix multiplication function $\text{MatMul}(\cdot)$ similarly to (5.55) as follows:

$$\text{MatMul}(\mathbf{V}) = \text{HMP}(\mathbf{V}) + \mathbf{V}, \quad (5.67)$$

where we defined $\mathbf{V}$ to be a matrix whose columns are the random vectors $\mathbf{v}_i$ for parallel computation. The computational cost of the log-determinant computation is $\mathcal{O}((N^2M + NM^2)R)$.

Full-spectral Point Source Localization

Although the work discussed thus far concerns imaging distributed radiological sources, here we briefly show that the idea of simultaneously estimating source images and their radiological spectra components can be applied to point source localization (PSL). In PSL, it is assumed that the spatial extent of radiological source is very small. Therefore, instead of reconstructing a source image, the PSL estimates an index of a single voxel where the source is most likely located. The PSL was proposed by Hellfeld *et al.* [30] without making use of spectral data. Therefore, the measurement data is a 1D vector representing a time-series of count rate data measured from a free-moving detector.

Recall from (2.5) that without spectral data, the single detector imaging problem formulation is written as,

$$\mathbf{y} \sim \text{Poisson}(\bar{\mathbf{y}}) \quad \text{where} \quad \bar{\mathbf{y}} = \mathbf{A}\mathbf{x}. \quad (5.68)$$

The PSL jointly estimates the optimal voxel index $\hat{j}$ and the source strength $\hat{x}$ from the following optimization problem:

$$\begin{aligned} \hat{j}, \hat{x} &= \underset{j, x \geq 0}{\text{argmin}} l_j(x; \mathbf{y}) \quad \text{where} \\ l_j(x; \mathbf{y}) &= \mathbf{1}^\top (\mathbf{A}[:, j]x - \mathbf{y} \odot \log \mathbf{A}[:, j]x) \end{aligned} \quad (5.69)$$

Note that $l_j(x; \mathbf{y})$ is the negative Poisson likelihood with a single j -th voxel. Hence, the PSL solves single-voxel MLE problems at all voxel locations, and outputs the voxel index j where the negative log-likelihood is the lowest and the corresponding MLE source strength³. Note that the MLE problem at each voxel can be solved with the ML-EM.

³Of course, one does not need to solve the MLE problems at all voxel locations to obtain the PSL solution. By employing an optimization algorithm, only a handful of MLE problems need to be solve and the brute-force method is not necessary.

Noting the similarity between the standard MLE image reconstruction problem and the PSL, one can easily formulate a full-spectral PSL approach from the full-spectral image reconstruction idea thus far discussed in this chapter.

Modifying the MLE problem formulation given in (5.5), the full-spectral PSL is given as,

$$\begin{aligned} \hat{j}, \hat{\mathbf{X}} (= \{\hat{x}, \hat{\mathbf{X}}_C, \hat{\mathbf{X}}_{BW}, \hat{\mathbf{X}}_{BC}\}) &= \underset{j, \mathbf{X}, \mathbf{X} \geq 0}{\operatorname{argmin}} -\log p(\mathbf{Y}|\mathbf{X}) \\ &= \underset{j, \mathbf{X}, \mathbf{X} \geq 0}{\operatorname{argmin}} l_j(\mathbf{X}; \mathbf{Y}), \end{aligned} \quad (5.70)$$

where

$$\begin{aligned} l_j(\mathbf{X}; \mathbf{Y}) &= \mathbf{1}^\top (\bar{\mathbf{Y}}_j - \mathbf{Y} \odot \log \bar{\mathbf{Y}}_j) \mathbf{1} \quad \text{and} \\ \bar{\mathbf{Y}}_j &= \mathbf{A}_G[:, j] x \mathbf{X}_C^\top \mathbf{A}_R + \mathbf{X}_{BW} \mathbf{X}_{BC}^\top. \end{aligned} \quad (5.71)$$

The MLE problem in each voxel is solved with the iterative block-coordinate descent algorithm in (5.6).

Although the full-spectral PSL approach is considerably less complicate than the MLE problem in (5.5), it is still severely ill-posed, due to the source spectrum and background estimation. To alleviate the ill-posedness of the problem, the GP priors can be imposed on the parameters to be estimated except the source strength x . The full-spectral PSL with the GP priors will be referred to as full-spectral GPP-PSL henceforth. The derivation of the full-spectral GPP-PSL algorithm will be omitted, as it can be straightforwardly derived from (5.22).

Uncertainty quantification in the full-spectral GPP-PSL approach is currently limited to single-voxel MAP estimations. Therefore, a different method is needed to quantify uncertainties across MAP estimates for different voxel indices, which will be addressed in future work.

5.2 Results

In this section, we demonstrate the effectiveness of the GP-based full-spectral imaging and uncertainty quantification method using a Monte Carlo-simulated single-detector imaging scenario and a real-world measurement data. The proposed method is compared with the MLE method, highlighting the advantages of the Bayesian approach.

Large area mapping of mixed ^{134}Cs , ^{137}Cs and ^{131}I

First, a complex measurement scenario with multiple distributed isotopes were simulated using MEGAlib—a Geant4 based Monte Carlo simulation package. Figure 5.2 illustrates the simulated distributions of three isotopes ^{134}Cs , ^{137}Cs and ^{131}I in a 120 m $\times$ 90 m imaging space, along with the measurement path of a hand-held 3" $\times$ 3" cylindrical NaI(Tl) scintillator

detector. The measurement was conducted over approximately 6 minutes, following a raster-pattern at a speed of 2 m/s, with the detector height fixed at 0.7 m above ground level.

The isotope distributions were determined with the Gaussian plume model as follows:

$$\text{Activity}(x, y) = \frac{\text{Activity}_{\max}}{\sqrt{2\pi}\sigma_{\text{gpm}}x} \exp\left(-\frac{y^2}{2(\sigma_{\text{gpm}}x)^2}\right). \quad (5.72)$$

For ^{134}Cs and ^{137}Cs isotopes, $\text{Activity}_{\max} = 12500 \text{ Bq/m}^2$ was applied, while for ^{131}I , it was set at 25000 Bq/m^2 . The Gaussian plume model offers a simple way to simulate dispersion of radioisotopes in a uniform wind field. To account for additional distribution shift from changes in wind direction, the Gaussian plume distribution was convolved with a 2D Gaussian filter with a standard deviation of 7 meters. Finally, the ^{134}Cs and ^{137}Cs plume distributions were rotated 67.5° while the ^{131}I distribution was rotated 22.5° about the x-axis, creating two distinct isotope distributions.

Additionally, spatially varying NORM background contribution from K, U and Th progenies were simulated as follows:

$$\text{K} : 150.0 \text{ cps} \quad (5.73)$$

$$\text{U} : 350.0 + 150 \sin(y/30) \text{ cps} \quad (5.74)$$

$$\text{Th} : 350.0 + 150 \sin(x/40) \text{ cps.} \quad (5.75)$$

Figure 5.2 illustrates the spatial distribution of these background count contributions. Thus, together with the cesium and iodine isotopes, these background radiation from the KUT progenies independently contribute to the total measurements, yielding a complex measurement scenario.

The system matrix components $\mathbf{A}_G$ and $\mathbf{A}_R$ were determined as follows. The elements of the geometric efficiency matrix $\mathbf{A}_G$ were computed based on the solid angle of the cylindrical detector at each measurement position

$$\mathbf{A}_G[i, j] = \frac{(\text{Det. height})(\text{Det. diameter}) \sin \theta + \frac{\pi}{4}(\text{Det. diameter})^2 \cos \theta}{4\pi \|\mathbf{r}\|_2^2}, \quad (5.76)$$

where $\mathbf{r}$ represents the vector connecting the j -th pixel to the detector location at the i -th measurement, and θ denotes an angle between $\mathbf{r}$ and the normal vector of the j -th pixel. The detector response matrix $\mathbf{A}_R$ was derived using a Monte Carlo simulation.

The Monte Carlo simulated data was reconstructed using the MLE approach (5.6) and the proposed method, and their image reconstruction results and spectral components reconstruction results are presented in Fig. 5.3 and Fig. 5.4, respectively. The MLE results were obtained after 50 iterations, while the GPP MAP reconstruction was performed following Algorithm 3 after hyperparameter optimization using empirical Bayes. As observed, the MLE approach fails in the reconstruction task, yielding overfitted noisy reconstruction results. On the other hand, the proposed method successfully reconstructs the underlying source distributions, their spectral components, and the three distinct background spectral components corresponding to KUT progenies.

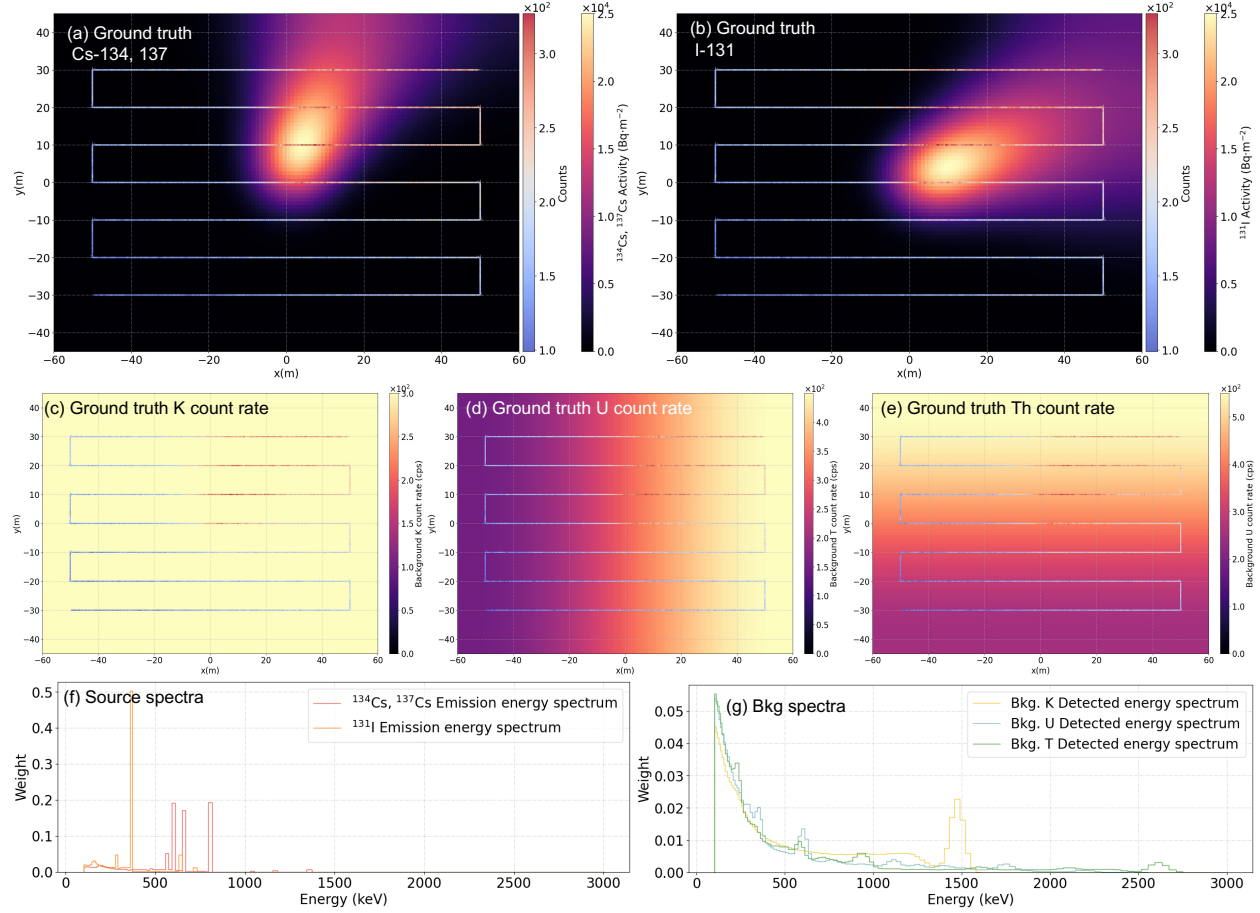


Figure 5.1: (a)(b): Ground truth source activity distributions for (a) ^{134}Cs , ^{137}Cs and (b) ^{131}I . (c)(d)(e): Spatially varying count rate contribution from KUT progenies. (f) Ground truth source emission energy spectra. Note that the emission spectra include downscattered and backscattered photons from the ground, which manifest as a low-energy continuum. (g): Ground truth detected background spectra from KUT progenies. These spectra represent what would be observed by the detector. Therefore, they are influenced by detector-specific spectral features, such as the NaI(Tl) detector resolution and pronounced Compton continua.

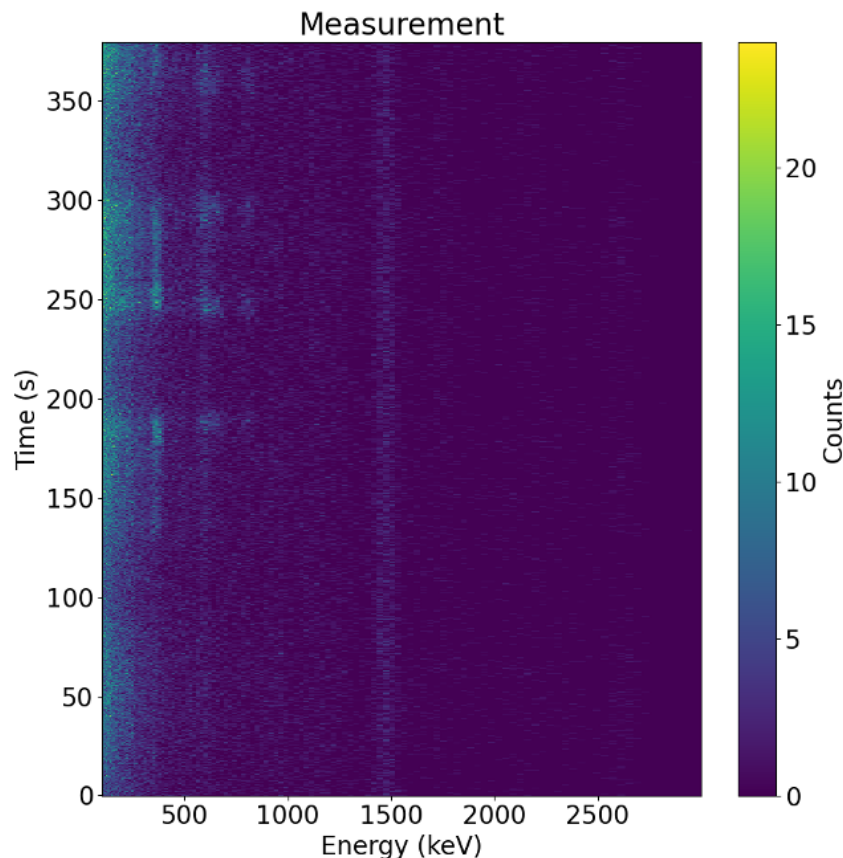


Figure 5.2: 2D time-energy count histogram from the Monte Carlo simulated measurement with a NaI(Tl) detector. The time-dimension bin width is 0.25 s, while in the energy-dimension 128 bins were spaced proportional to $\sqrt{E}$. The independent contributions from the five source and background components yields complex and dynamic measurement data.

Savannah River National Laboratory (SRNL) H-area contamination mapping

Next, we demonstrate the performance of the proposed approach on a challenging low-SNR real-world measurement. Figure 5.5 shows the map of Savannah River National Laboratory (SRNL) H-area, where an unknown amount of ^{137}Cs is distributed. The area was surveyed with MiniPRISM, a cadmium zinc telluride (CZT)-based hand-held radiation imager (see Fig. 5.6). Figure 5.7 presents its measurement path, as well as the measured 2D time-energy histogram. The survey lasted approximately 20 minutes and covered the area between the fence on the left and the forest on the right. Notably, the gross count rate exhibits a distinct spatial pattern: it is elevated near the fence, whereas it decreases towards the forest.

To determine whether the fluctuations in count rate are due to the background or the

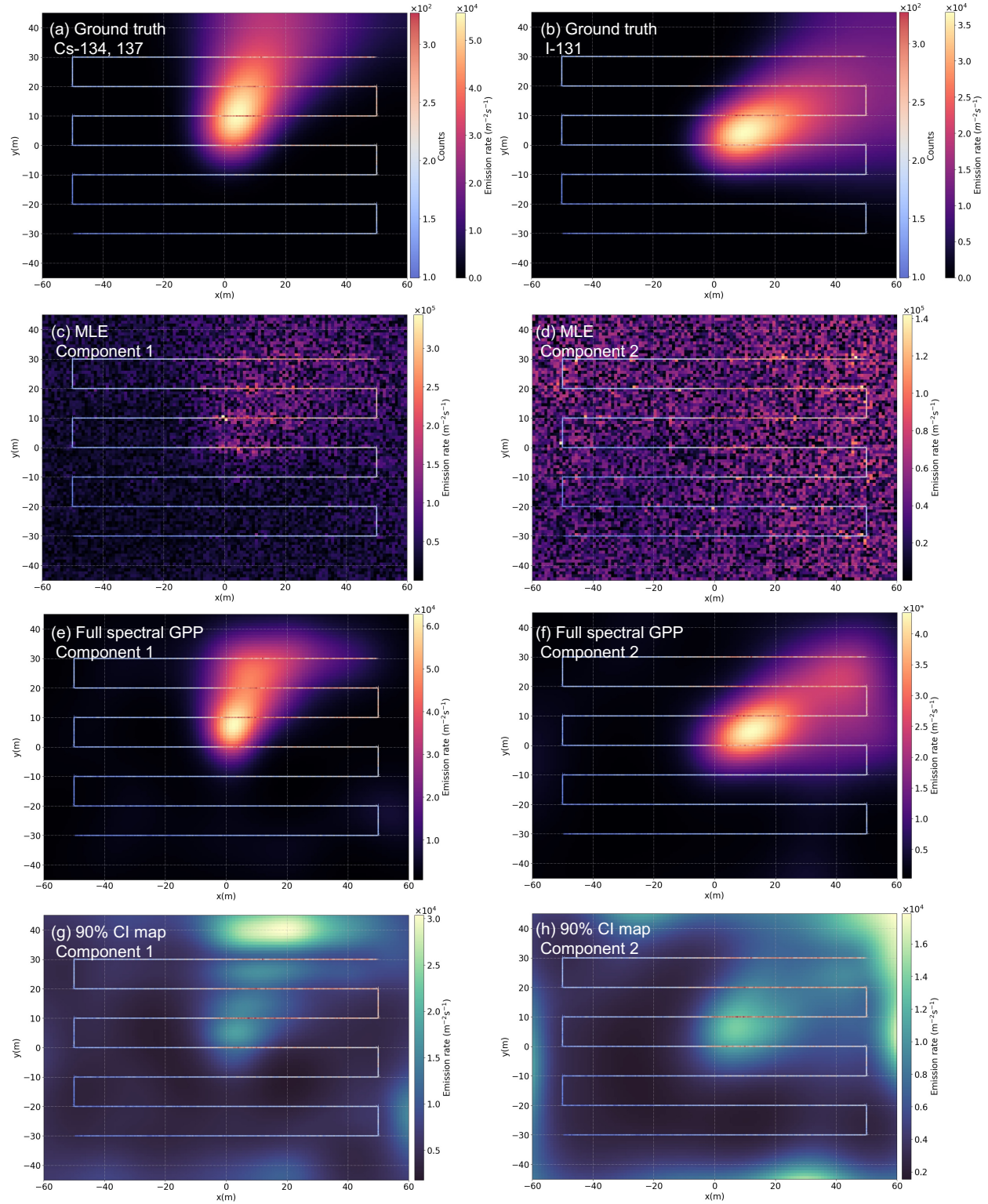


Figure 5.3: (a)(b): Ground truth emission rate distributions for (a) ^{134}Cs , ^{137}Cs and (b) ^{131}I . (c)(d): MLE image reconstruction results. Note that the reconstruction is corrupted with noise, which is a typical sign of overfitting. (e)(f): Full spectral GPP image reconstruction results. Component 1 and 2 appear to correspond to the Cs and I isotope distributions, respectively. (g)(h): 90% credible interval maps from the Laplace approximation.

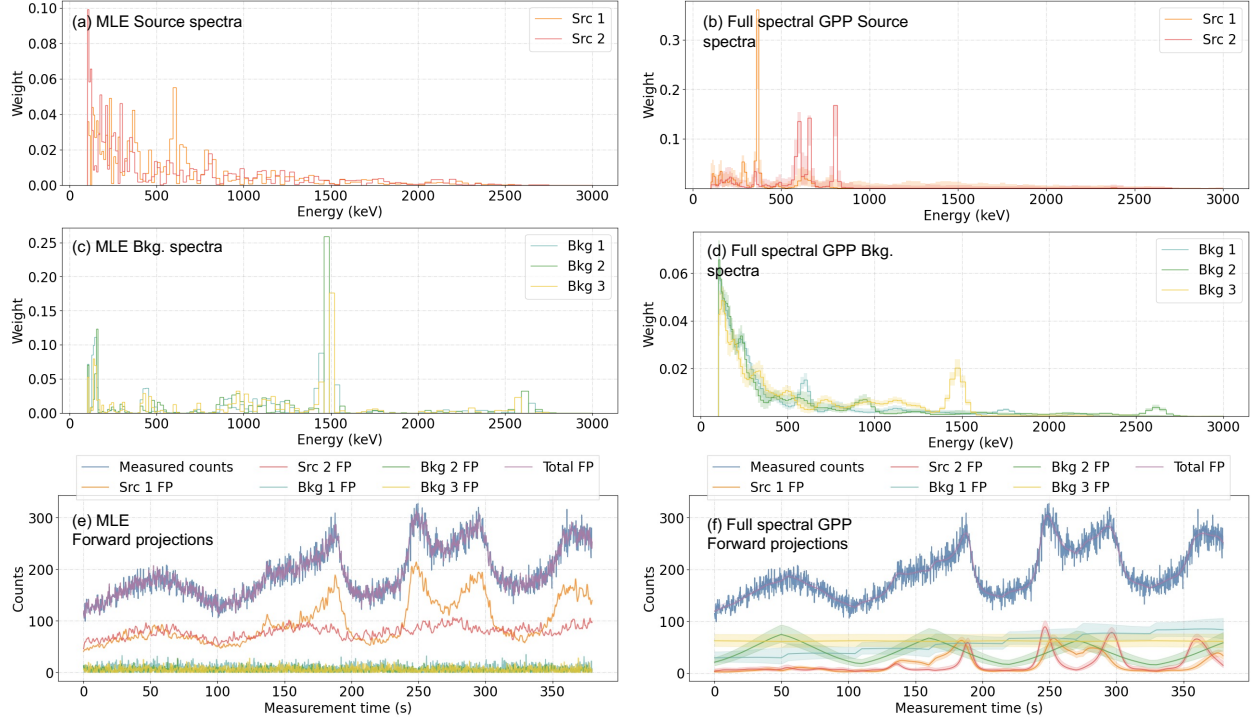


Figure 5.4: (a)(b): Reconstructed source spectra from the MLE and full spectral GPP reconstruction. (c)(d): Reconstructed background detection spectra. (e)(f): Time-series of gross counts measurements and forward projections from the reconstructed source and background components. In (b)(c)(d), 90% credible intervals estimated from the Laplace approximation are indicated with shaded regions.

^{137}Cs deposition, we analyzed the 2D energy-time histogram shown in Fig. 5.7. The histogram clearly reveals two observations: first, the low-energy downscattered continuum dynamically varies across the measurement period, and secondly, a trace amount of ^{137}Cs photopeak contribution at 662 keV was observed. These findings indicate that the gross count variation is attributed to the spatially varying background contribution. Moreover, the background contribution interferes with the ^{137}Cs measurement, lowering the SNR of the 662 keV photopeak counts measured, presenting a challenge in ^{137}Cs distribution mapping.

The reconstruction of the ^{137}Cs distribution was carried out as follows. First, the geometric efficiency matrix $\mathbf{A}_G$ was derived from an experimental measurement, where the average photon detection efficiency as a function of photon incident angle was determined. The detector response matrix $\mathbf{A}_R$ was assumed to be an identity matrix $\mathbf{A}_R = \mathbf{I}$. Consequently, reconstructed source spectral components do not represent emission spectra, but detected spectra from the MiniPRISM imager. With the detector response matrix being an identity matrix, the reconstruction results are no longer quantitative. Nevertheless, it still provides

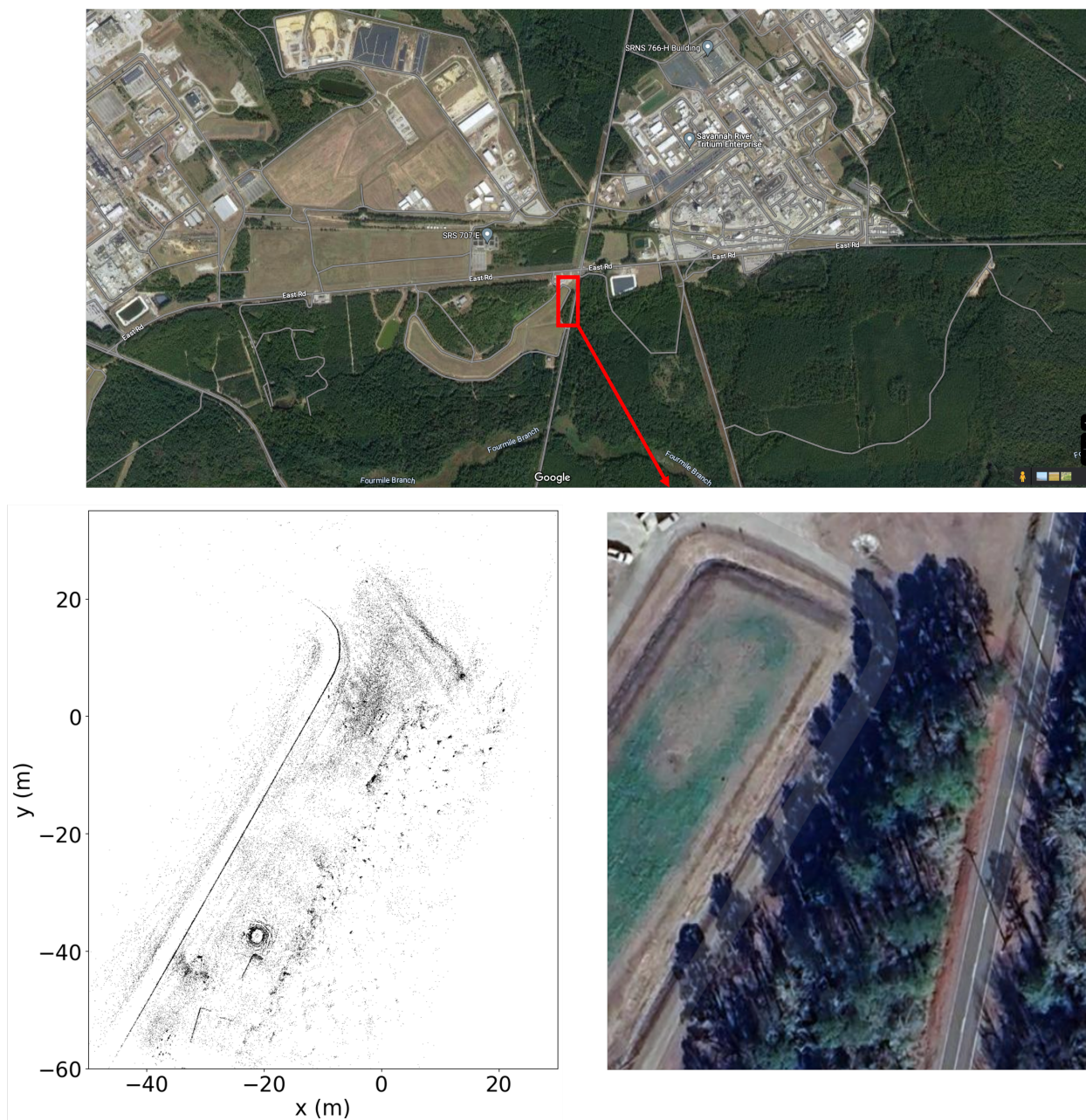


Figure 5.5: (Bottom left): A point cloud map of SRNL H-area obtained from a measurement with a LiDAR on MiniPRISM imaging system. (Top, Bottom right): A map of SRNL H-area.

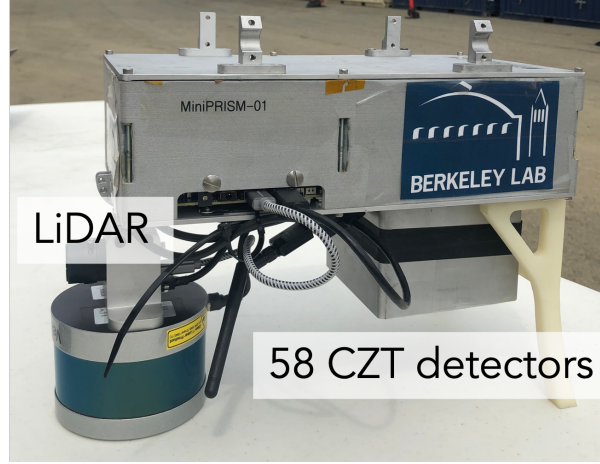


Figure 5.6: MiniPRISM imaging system is equipped with 58 CZT detectors and contextual sensors, including LiDAR, visual camera and IMU. The imager can operate as either a coded aperture imager or a Compton camera. However, in this work, the measurements from the 58 CZT detectors were aggregated, yielding measurement data akin to that from a single detector.

a relative source distribution, which remains valuable for various applications.

Once the system matrix factors $\mathbf{A}_G$ and $\mathbf{A}_R$ were determined, the optimal number of source and background components was derived using the empirical Bayes method. Three different candidate models, each with varying number of background components, were initialized, and hyperparameter optimization was carried out using the empirical Bayes method. In empirical Bayes, optimal hyperparameters are derived by minimizing the negative log-marginal likelihood (NLML) with an iterative optimization algorithm, and the results are shown in Fig 5.10. The results clearly indicates that the 1-background model lacks sufficient flexibility to fit the measurement data, leading to the highest NLML value. In contrast, the 2- and 3-background component models both converge to the same NLML value. Upon examining the hyperparameter optimized 3-background component model, we found that the background weights for one of the background components were zeroed, effectively reducing it to the 2-background component model. Therefore, we concluded that the 2-background component model provides sufficient flexibility without superfluous components.

The reconstruction results from the 2-background model are presented in Fig. 5.8. First, the strong 662 keV photopeak feature from the source spectral component suggests that the model successfully identified and separated the ^{137}Cs contribution from the measurement data. Furthermore, the reconstructed source spectrum includes the downscattered Compton continuum feature, which is expected from a ^{137}Cs measurement with a CZT detector. Interestingly, while only the 662 keV photopeak features are visible from the 2D time-energy histogram, the proposed method successfully extract salient spectral features of the ^{137}Cs

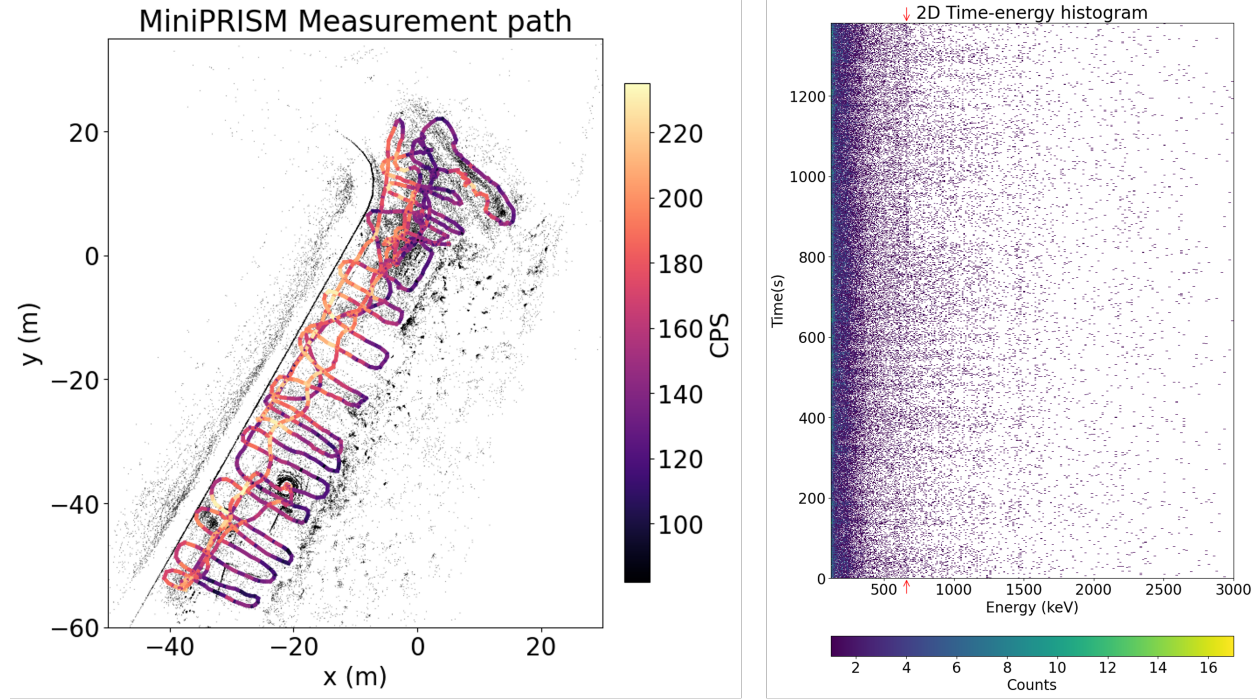


Figure 5.7: (Left): The measurement path and gross counts measured from MiniPRISM. Notably, the gross counts level is elevated near the fence on the left, and decreases near the forest on the right side. (Right): The 2D time-energy histogram from the MiniPRISM measurement. For visibility, bins with 0 count were not colored. The background variability is observed from the horizontal streaking features. Additionally, a trace of ^{137}Cs photopeak counts at 662 keV is visible.

energy spectrum from the measurement.

The corresponding source distribution successfully localizes and maps the ^{137}Cs hotspots around the measurement path. Moreover, the 90% credible interval map highlights elevated uncertainties around the identified hotspots and in unexplored regions of the imaging space. Note that the uncertainties in the hotspots reflect the statistical uncertainties inherent in the measurements, whereas the uncertainties in the unexplored regions arise from lack of sensitivity (i.e., epistemic uncertainties). Unfortunately, ground truth data is not available for this measurement; therefore, the reconstruction could not be validated in this work. Nevertheless, the uncertainty map serves as a valuable guide for planning future measurements in the area.

The reconstructed background spectral components and the forward projections from the source and background components are shown in Fig. 5.9. The background component 1 has large weights in the low-energy region, indicating that it accounts for detection of down-scattered photons. In contrast, the background component 2 contains various photopeaks

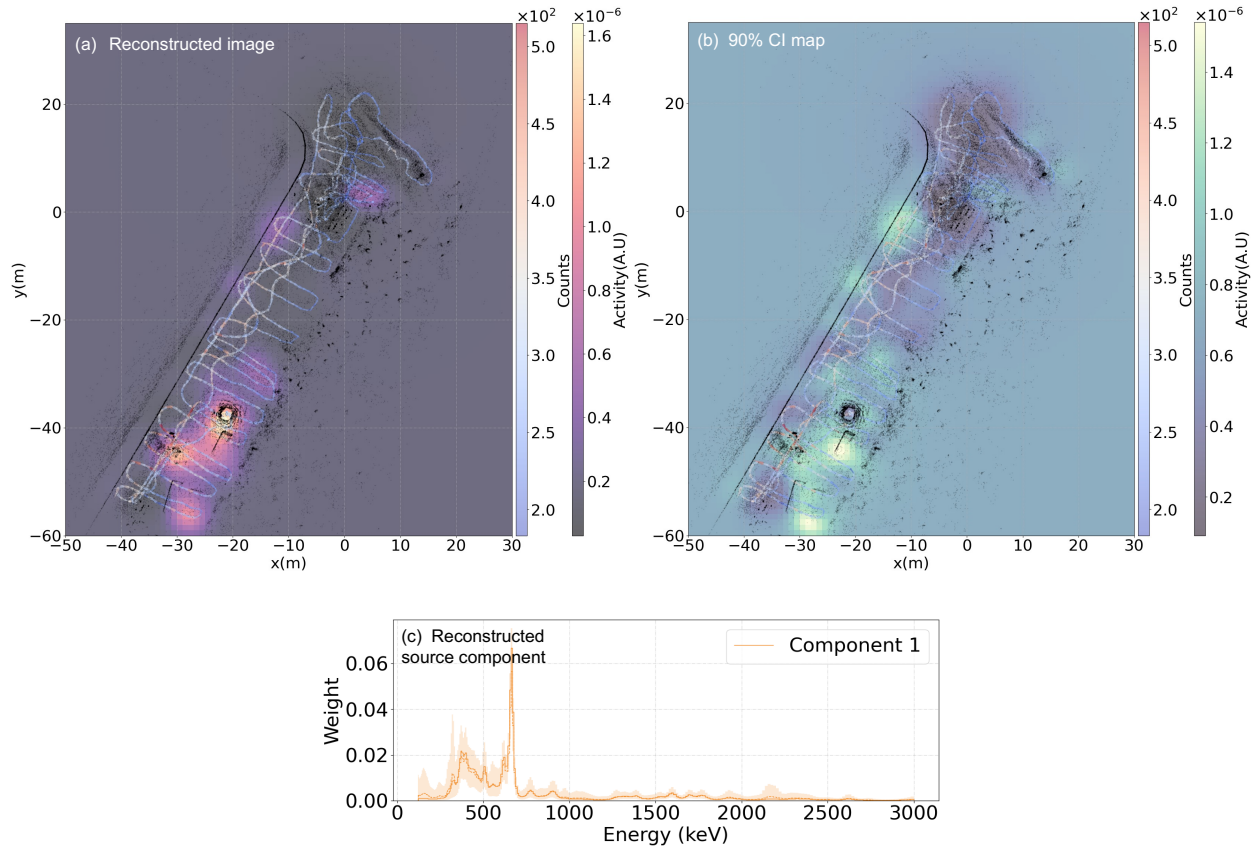


Figure 5.8: (a): The ^{137}Cs distribution reconstructed from the proposed approach. (b): The 90% credible interval map from the Laplace approximation(c) The reconstructed ^{137}Cs spectral component. Interestingly, the reconstructed spectrum clearly shows the Compton edge and Compton continuum at <477 keV as well as the 662 keV photopeak feature.

from the KUT NORM progenies, including the 238 keV, 609 keV, and 1460 keV photopeaks.

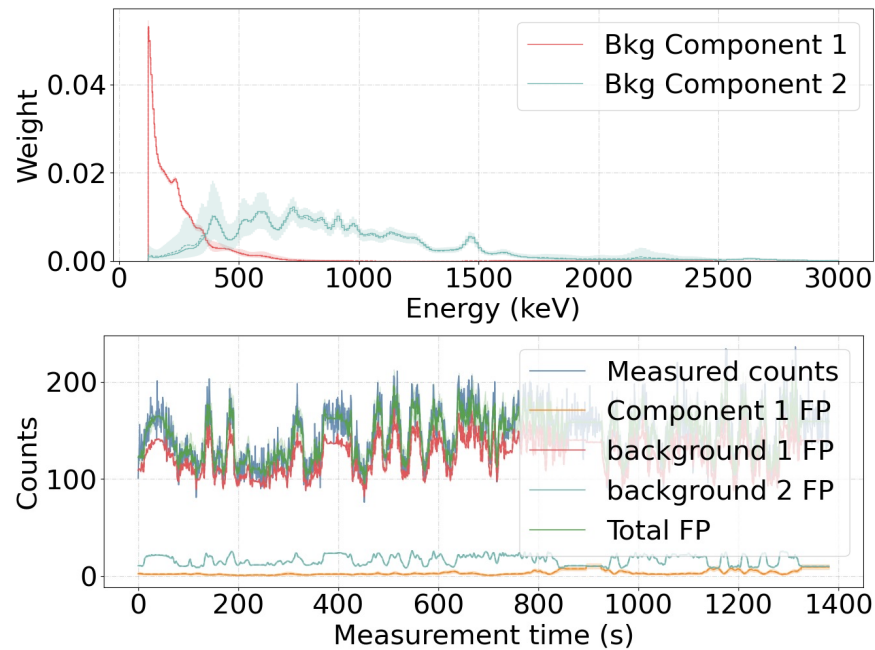


Figure 5.9: (Top): The reconstructed background spectral components. The background component 1 appears to correspond to downscattered photons, while the component 2 contains various photopeak features, such as 238 keV, 609 keV, and 1460 keV photopeaks from the KUT progenies. (Bottom) The forward projections from the reconstructed source and background components. In both figures, the shaded regions represent the 90% credible intervals. Note that the source forward projection is very low compared to that of background components, indicating the very low SNR in the data.

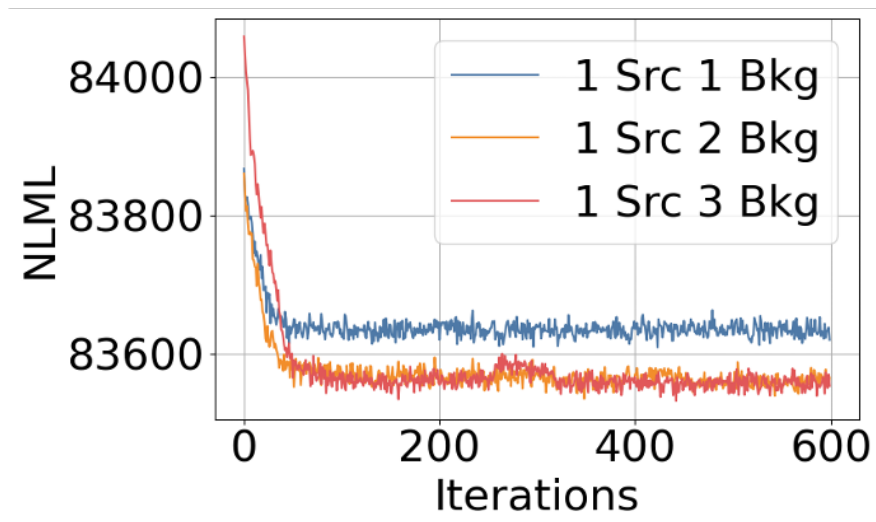


Figure 5.10: Negative log-marginal likelihood (NLML) value as a function of hyperparameter optimization iteration numbers. The 1 source 1 background model does not possess enough explanatory power, leading to a large negative log-marginal likelihood value after hyperparameter optimization. On the other hand, the 1 source 3 background model does not yield lower NLML than the 1 source 2 background model after hyperparameter optimization, suggesting 1 source 3 background model is superfluous for the measurement data.

Shielded highly enriched uranium point source localization

Here, we demonstrate the effectiveness of the Full spectral GPP-PSL approach with a simulated measurement scenario.

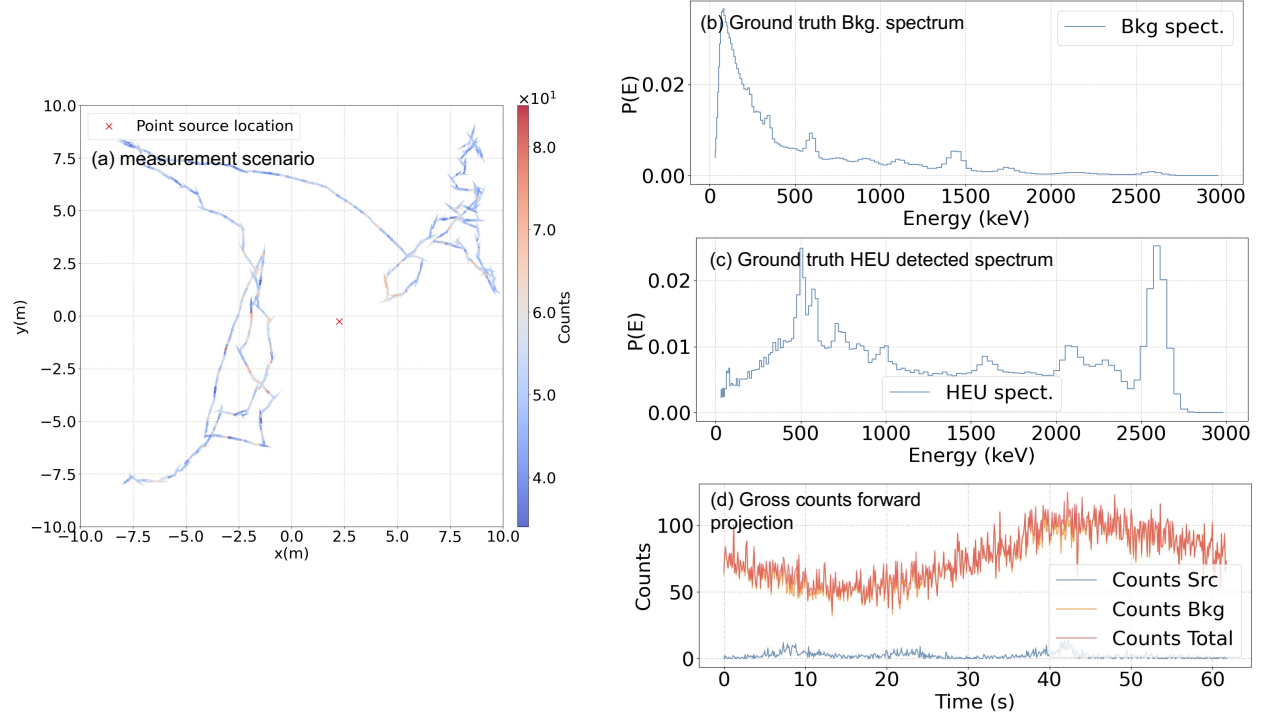


Figure 5.11: (a) A simulated HEU point source measurement scenario. (b)(c) Simulated ground truth background and 1-cm lead shielded HEU source spectra. (d) Gross count forward projection from the source and the background.

Figure 5.11 shows a simulated measurement scenario with a 1-cm lead shielded highly enriched uranium (HEU) source placed. The measurement was conducted with a 20-cm radius spherical detector for approximately a minute. The simulated background and the source energy spectrum can be found in Fig. 5.10. Note that the HEU source is heavily shielded; therefore, the source energy spectrum does not exhibit typical photopeak features expected from ^{235}U source, such as 185.7 keV gamma-ray photopeak. Instead, the source spectrum is largely spread out across energy domain, making it difficult to estimate the source isotope from the spectrum. Moreover, the gross counts forward projection from the source and background in Fig. 5.10 show that it is an extremely low SNR problem, and the background rate changes dynamically, further interfering with the counts collected from the HEU source.

Figure 5.12 shows the 2D time-energy histogram data from the simulated scenario. Notably, the source spectral features are almost invisible, indicating that an energy ROI cannot

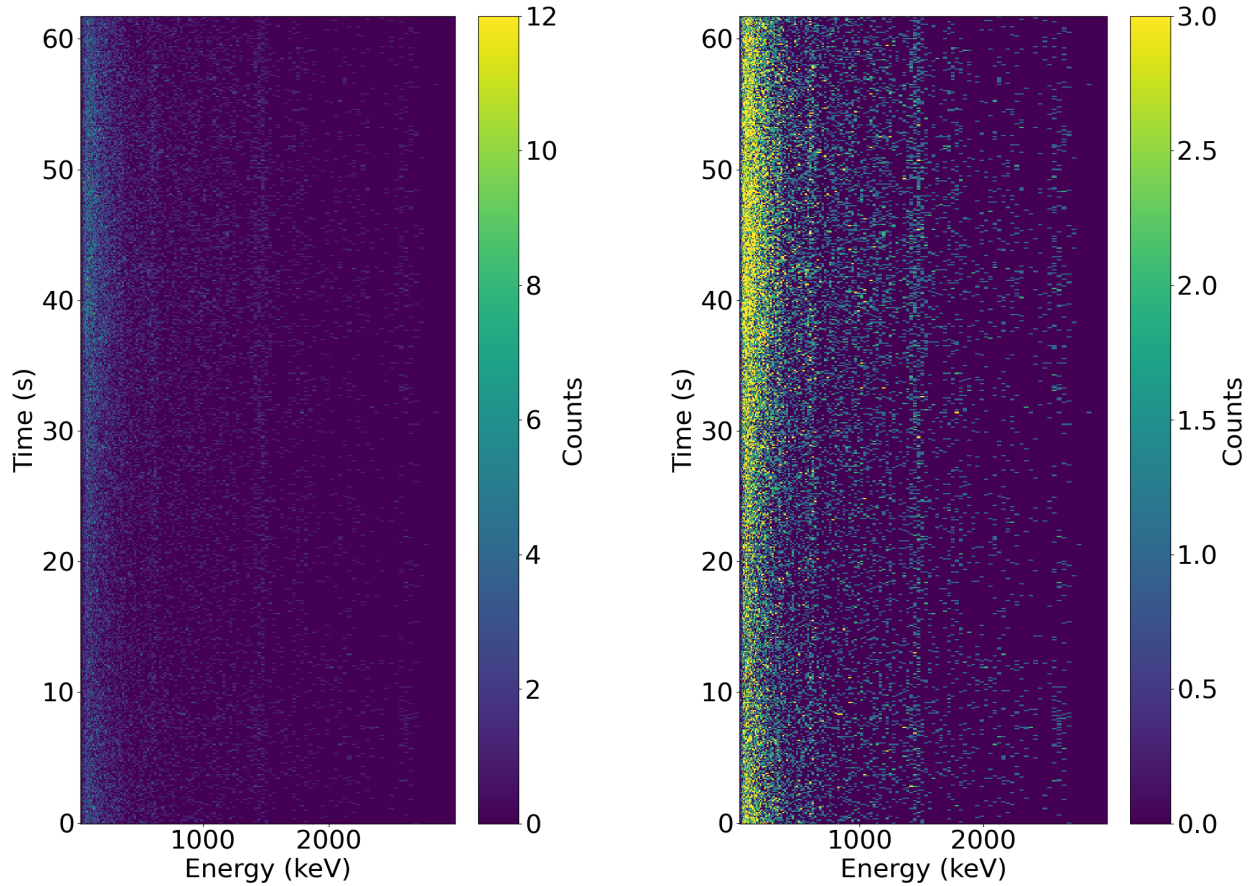


Figure 5.12: (left) Measured time-energy 2D histogram data. (right) For visibility, the histogram colorbar scale was modified, revealing the dynamically changing background spectral features.

be set around source photopeaks. On the other hand, clearly observable is the varying background spectral features.

With the time-serise of gross counts measured the PSL estimation results are shown in Fig. 5.13. First of all, the estimation is clearly erroneous, and the gross counts forward projection indicates that the PSL approach cannot effectively model the varying background countrate.

Fig. 5.14 illustrates the results of using the full-spectral PSL MLE without any priors and the full-spectral GPP-PSL estimation. Comparing the results to the vanilla PSL results in Fig. 5.13, it can be noitced that the source location and strength estimations are significantly improved. Furthremore, it is demonstrated that incorporation of the GP priors improves the source spectrum estimation significantly, while significantly alleviating the overfitting of the full spectral PSL MLE approach.

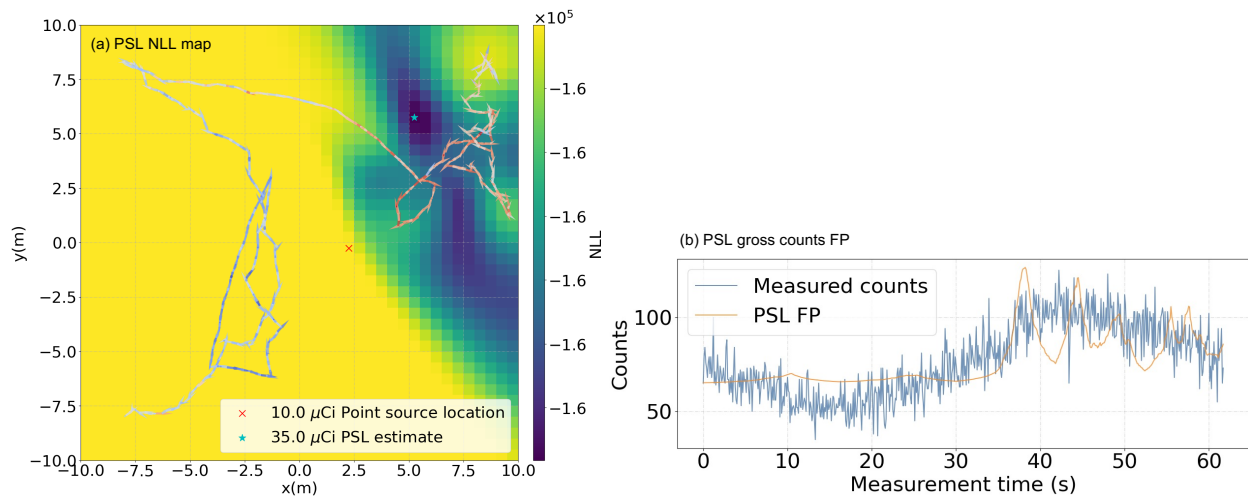


Figure 5.13: Results from the gross count PSL algorithm. Note that the PSL approach cannot effectively deal with the dynamically changing background count rate.

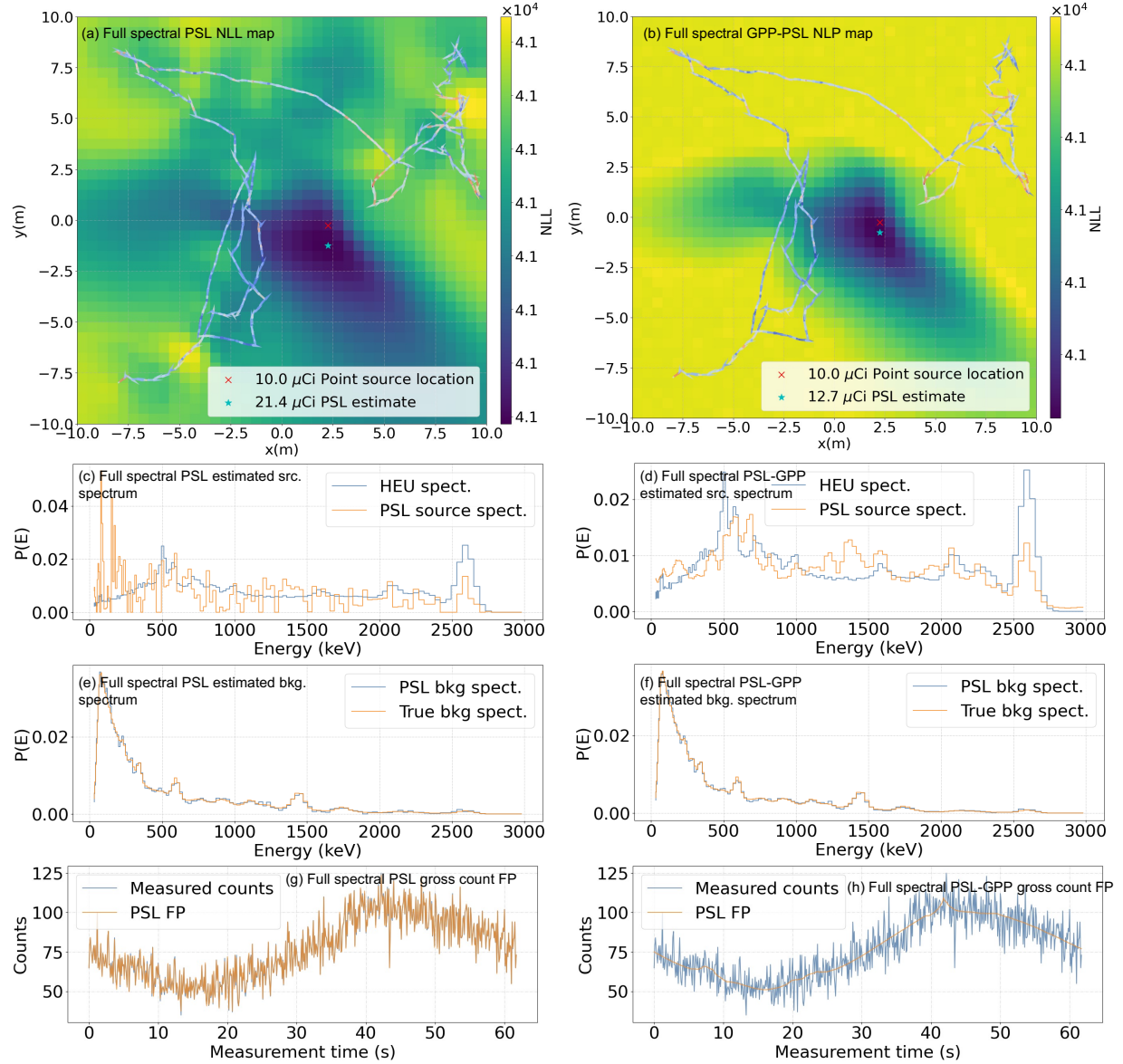


Figure 5.14: Left: Full spectral PSL MLE results, without any priors imposed. Right: Full spectral GPP-PSL results. (a)(b): Source location and strength estimation. (c)(d) Reconstructed source spectrum. (e)(f) Reconstructed background spectrum. (g)(h) Forward projection gross counts from the reconstructions. Note that by incorporating the GP prior, the source spectrum estimation is significantly improved. In addition, the forward projections in (g) and (h) indicate that the GP prior effectively combat the overfitting of the MLE method.

5.3 Discussion and Future Work

Validity of the Problem Formulation

One of the key contributions of this study is the novel approach to reformulating the very high-dimensional problem (5.1) into (5.3) using low-rank approximations. Specifically, the factorization of the system matrix, $\mathbf{A} \approx \mathbf{A}_G \otimes \mathbf{A}_R^\top$ significantly reduces the number of system matrix elements from $IJKL$ to $IJ + KL$, enabling computationally efficient forward projection. As previously noted, this approximation is well-justified because the probability of photon detection is often expressed as a product of the detector geometric efficiency (i.e., the solid angle of the detector) and its intrinsic efficiency.

The assumption holds particularly well for low-energy photons, as they primarily interact at the surface of the detector. Thus, its detection efficiency is directly proportional to the apparent surface area of the detector as seen from voxel positions. However, this approximation may be invalid with a significant portion of incident radiations passing through the detector without interacting. In such cases, the detection efficiency is no longer proportional to the apparent surface area, but instead to the volume of the detector. This might result from incident photons with very high energy, a low effective atomic number of the detector material, and/or a very small detector size. Consequently, when radioactive sources of interest emit both very high and low energy photons, this might lead to erroneous reconstructions.

In most practical applications where incident photon energies are less than 10 MeV and reasonably dense detectors are employed (e.g., NaI, CZT, etc.), this error can be safely ignored. However, When this error is anticipated, it can be mitigated by using a large detector with its dimensions significantly greater than the mean free path of the incident radiations.

Initialization of Parameters and Hyperparameters

It is well-known that matrix decomposition problems with non-negativity constraints are non-convex. Similarly, the MLE problem in (5.5) is non-convex and therefore, there is a high probability that the block coordinate descent algorithm (5.6) finds undesirable local minimum MLE solutions with different parameter initialization.

With the GP priors imposed on parameters, the MAP solution must minimize both the negative log-likelihood and negative log-prior, and consequently, the risk of finding an undesirable local minimum reduces considerably. With the GP prior, it is reasonable to initialize parameters with a sample from the GP prior distribution as follows:

$$\boldsymbol{\Xi}_w^{\text{init}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (5.77)$$

Note that with $\boldsymbol{\Xi}_w$ following standard normal distribution, the $\boldsymbol{\Xi}^{\text{init}} = \mathbf{L}\boldsymbol{\Xi}_w^{\text{init}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and therefore, $\boldsymbol{\Xi}^{\text{init}}$ follows the GP prior distribution. We found that with the initialization scheme, the MAP solution is consistent up to permutation of components.

Permutation between source components or background components are equivalent solutions. However, there is a chance that permutation between source and background components occurs. Typically, this problem can be resolved with a new initialization of Ξ_w and running Algorithm 3 again, but when the source SNR is low, it may be difficult to obtain a desirable solution.

In such cases, domain specific knowledge can be used in initialization to find a better solution. For example, in many measurement scenarios, the radiological sources of interest and their expected energy spectra are known *a priori*. Suppose we expect a MAP estimated source spectrum $\hat{\mathbf{X}}_C$ to be close to $\mathbf{X}'_C$. Then, initialization of $\hat{\Xi}_{w,C}^{\text{init}}$ can be done to satisfy the following:

$$\mathbf{f}_C^{-1}(\mathbf{L}_C \hat{\Xi}_{w,C}^{\text{init}}) = \mathbf{X}'_C. \quad (5.78)$$

Such initialization ensures that the MAP solution $\hat{\mathbf{X}}_C$ is close to the expected source spectra $\mathbf{X}'_C$. Also, similar initialization with domain knowledge can be applied to other parameters, $\mathbf{X}_I$, $\mathbf{X}_{BW}$, and $\mathbf{X}_{BC}$. In this work we did not assume any domain knowledge and used Gaussian initialization only. However, in our experiments we found that incorporation of domain knowledge significantly enhances the chance of successful reconstruction.

Good hyperparameter initialization is another essential ingredient for successful reconstruction. Although they are optimized using empirical Bayes, its objective function $\Phi(\boldsymbol{\theta})$ is also non-convex, and therefore, its optimization might result in an inadequate set of hyperparameters.

To ensure reasonable hyperparameter initialization, one can visualize a sample from the prior distribution after transforming it into $\mathbf{X} = \mathbf{f}^{-1}(\Xi)$ with the inverse link function $\mathbf{f}^{-1}$ defined with the hyperparameters. Then, by comparing the images, energy spectra, and forward projections, one can determine whether a reasonable choice of hyperparameters was made.

Additionally, hyperparameters are physically and statistically interpretable, and therefore a domain specific knowledge can easily be incorporated. For example, the lengthscale parameter σ can be set to reflect source distribution correlation, detector energy resolution, and background spatial correlation, depending on the parameter matrix. Furthermore, λ parameter in (5.16) can be interpreted as a constant mean for the GP prior as follows:

$$\mathbf{X} = \mathbf{f}^{-1}(\Xi) \quad (5.79)$$

$$= \frac{1}{\lambda} \exp(\Xi) \quad (5.80)$$

$$= \exp(-\log \lambda) \exp(\Xi) \quad (5.81)$$

$$= \exp(-\log \lambda + \Xi). \quad (5.82)$$

Therefore, Ξ follows a multivariate Gaussian distribution with mean $-\log \lambda$ and variance Σ and after exponentiated, it follows from the definition of log-normal distribution that the elements of $\exp(\Xi)$ has mean $\frac{1}{\lambda} \exp(l^2/2)$ and variance $\frac{1}{\lambda^2} (\exp(l^2) - 1) \exp l^2$. From this, one can set hyperparameters to match the anticipated mean and variance of the variables.

As a Standalone NMF Algorithm

The method presented here can be categorized as a special case of NMF. In particular, without the term for source contribution $\mathbf{A}_G \mathbf{X}_I \mathbf{X}_X \mathbf{A}_R$, (5.3) reduces to a typical NMF formulation. In NMF, a data matrix is factorized into two low-rank matrices with the condition that all three matrices are non-negative. It has been shown that the non-negativity constraints induce physically interpretable part-based decomposition [115]. For this reason, the NMF has been successfully applied in a variety of applications such as astronomy, data imputation, radiation spectral data analysis and bioinformatics, to name a few.

However, as demonstrated by the results in this work, NMF often fails to find meaningful solutions when data is insufficient and noisy. To address this limitation, the Bayesian treatment of NMF with a GP prior considered in this work can serve as a powerful alternative to standard NMF. The use of a GP prior in NMF has been explored by Schmidt *et al.* [78], who demonstrated significantly improved performance over the standard NMF with an example from chemical shift brain imaging. However, the work lacked a systematic approach to hyperparameter optimization, number of component selection and Bayesian uncertainty quantification, which are critical for robust Bayesian NMF algorithm.

In contrast, the empirical Bayes-based hyperparameter optimization and model selection method presented in this work provides a robust framework for NMF with a GP prior. Additionally, the uncertainty quantification capability offers valuable insights into decomposition tasks, enabling deeper understanding of results. Given its versatility and enhanced performance, this proposed approach makes an important contribution to the advancement of NMF methodologies with its utility in a broad range of applications where robust matrix factorization with uncertainty quantification is desired.

Future Work

The approach proposed in this work opens up a numerous avenues for future research. In the following, we highlight several promising directions. The Laplace approximation employed in this work plays a pivotal role in uncertainty quantification and hyperparameter optimization via empirical Bayes. While its use is justified by the Bernstein-von Mises theorem [139], it relies solely on information inferred at the posterior mode, limiting its accuracy when the posterior distribution significantly deviates from Gaussianity (e.g., when it is skewed, sub-Gaussian, or super-Gaussian). Variational inference and expectation propagation are attractive alternatives for posterior approximation [140][141]. In these methods, an intractable posterior distribution p is approximated by a parametric distribution q_{θ} (often Gaussian) through the minimization of a Kullback-Leibler (KL) divergence. Variational inference minimizes $D_{\text{KL}}(q_{\theta} \parallel p)$, while expectation maximization minimizes $D_{\text{KL}}(p \parallel q_{\theta})$. Unlike the Laplace approximation, these approaches consider the similarity of the two distributions across their entire domain, leading to more accurate posterior approximations when the true posterior is far from Gaussian.

Another popular approach for approximating intractable posteriors is Markov Chain Monte Carlo (MCMC), which generates a sequence of samples where each sample depends on the previous one (i.e., exhibiting the Markov property). MCMC methods are well-known for producing asymptotically exact samples from the target posterior distribution as the number of samples approaches infinity. They have been extensively employed in a broad range of Bayesian inference tasks. However, as the dimensionality increases, MCMC samples often become highly correlated, hindering efficiency in high-dimensional inference problems.

Recently, function-space MCMC algorithms have emerged as a promising solution to the sample degeneracy problem in high-dimensional inference tasks [72][142]. These methods are discretization agnostic, making them particularly appealing for imaging and spectroscopy applications. The preconditioned Crank–Nicolson (pCN) algorithm, for instance, is specifically designed for drawing samples from a posterior distribution when GP priors are employed [72]. Although these MCMC algorithms are often memory-intensive, once executed, they can serve as valuable benchmarks for evaluating the performance of other posterior approximation techniques.

5.4 Conclusions

In this work, we presented a novel Bayesian approach to simultaneously reconstruct radiological spectra and their corresponding spatial radioactivity distributions from single-detector measurement data. The problem is highly ill-posed by nature; however, we demonstrated that by employing low-rank approximations akin to NMF, a computationally efficient forward model can be obtained. The reconstruction problem is solved with incorporation of the GP priors, ensuring robust solution images and associated radiological spectra. Furthermore, the LA used in this work enables Bayesian uncertainty quantification and hyperparameter optimization via empirical Bayes, thereby enhancing the robustness and reliability of the reconstruction process. We also detailed methods for a computationally efficient implementation of the proposed method using the Kronecker product covariance matrix and randomized linear algebra. The effectiveness of the approach was demonstrated through highly challenging simulated and real-world measurement data.

The implications of the proposed method are significant. Radiation imaging with a single detector provides a fast, inexpensive, and robust means for large-area radiation mapping with very high sensitivity. Moreover, its simultaneous spectroscopic capability allows for a deeper understanding of the radiological environment in the mapped area. In addition, the proposed approach can be adapted as a Bayesian NMF algorithm, thereby opening up numerous potential applications.

Chapter 6

Closing Remarks

6.1 Summary

Image reconstruction is an integral part of any radiation imaging procedures, where raw measurement data are transformed into human-interpretable images. However, due to inevitable measurement noise and limited amount of information obtained from the imaging system, it is fundamentally an ill-posed inverse problem where a unique solution may not exist. Consequently, a considerable amount of effort has been devoted to the development of image reconstruction algorithms.

Bayesian image reconstruction is a popular approach to addressing the ill-posed nature of image reconstruction problems [143][144]. At the core of Bayesian methods lies the incorporation of *a priori* knowledge in the form of a prior distribution, which helps to specify a unique solution image. While Bayesian approaches have been extensively studied in the context of image reconstruction, their widespread adoption has been hindered by several key challenges. First, mathematically defining suitable prior distributions accurately reflecting prior assumptions is a complex task. Additionally, Bayesian methods face significant computational challenges due to the high-dimensional nature of image reconstruction problems.

Despite these practical challenges, the importance of Bayesian image reconstruction has become increasingly evident with the development of novel radiation imaging devices, where conventional image reconstruction methods struggle with low SNR in measurements. Furthermore, recent advancements in computing resources, combined with modern Bayesian statistical methodologies have made the adaptation of Bayesian approaches more feasible.

This dissertation aims to present three novel Bayesian image reconstruction algorithms mainly in the context of free-moving radiation imaging. Firstly, in Chapter 3, the development of a sparsity inducing image reconstruction algorithm—the KLD MAP-EM algorithm—was introduced. The KLD MAP-EM algorithm is particularly effective when sparsity in the solution image is assumed (e.g., point sources) and is implemented as a straightforward modification of the conventional ML-EM algorithm. The algorithm was evaluated in various free-moving single detector imaging scenarios including real-world measurements, demon-

strating its utility even in low SNR measurement conditions.

In Chapter 4, a complete Bayesian framework for radiation image reconstruction, uncertainty quantification, and model selection methods is developed. The method incorporates a Gaussian process prior (GPP) to impose spatial correlations between voxels, and the resulting intractable posterior is approximated with the Laplace approximation and the preconditioned Crank-Nicolson Markov chain Monte Carlo (pCN MCMC) algorithm. The posterior approximation offers enhanced understanding of reconstructed images through Bayesian uncertainty quantification. Moreover, a suitable prior distribution is automatically selected with the empirical Bayes method.

Lastly, in Chapter 5, the development of the full-spectral GPP method is introduced. The method aims to reconstruct radiation activity images across all energy ranges. For computational feasibility and robust reconstruction, a low-rank approximation technique, similar to non-negative matrix factorization (NMF), is employed to decompose the images into a small set of independent source images and corresponding radiological spectra. To ensure robust reconstruction, the GPP is applied to both the images and their corresponding radiological spectra, with the Laplace approximation employed for uncertainty quantification and model selection. This method demonstrates an unprecedented capability to map distributed radiological sources from a single detector while separating salient radiological source and background radiological features from highly noisy measurement data.

6.2 Future Research Avenues

The work presented in this dissertation opens up numerous directions for future research.

Firstly, the Bayesian image reconstruction methods were demonstrated primarily for single detector imaging applications. However, since the statistical image reconstruction formalism is identical across most radiation imaging systems, these methods can be readily adapted to various imaging modalities that require robust image reconstruction from low SNR data. Examples of such modalities include Compton imaging, coded aperture imaging, and low-dose positron emission tomography (PET), among others.

Moreover, while uncertainties in reconstructed images offer valuable information, it is also demonstrated in chapter 4 that the uncertainties can be propagated to quantities derived from the images for downstream tasks. This capability has the potential to significantly impact applications involving high-stakes decision making based on radiation images, such as nuclear medicine and radiation security.

Additionally, it is worth noting that the Bayesian approaches developed in this work may have applications in inverse problems beyond image reconstruction. For instance, sparsity inducing priors have been a significant focus of research in statistics and machine learning [145], where the KLD MAP-EM algorithm can be readily adapted. In addition, the GPP method in Chapter 4 provides an effective framework for inverse problems in which correlations between parameters can help to find robust solutions. Furthermore, the spectral

reconstruction method discussed in Chapter 5 can be interpreted as a special case of NMF algorithm, which has been applied across various scientific disciplines [146].

Lastly, it should be highlighted that Bayesian statistics is a rapidly evolving field of research. In particular, recent advances in machine learning have spurred significant interest within the Bayesian research community to incorporate the Bayesian paradigm into neural networks (NNs). Examples of such efforts include imposing prior distributions on NNs [147], using NNs for prior learning [107], and employing untrained or pre-trained NNs as Bayesian priors [148][149][150].

Modern NNs are parameterized by very high dimensional models. Consequently, a primary focus of these approaches have been developing scalable and accurate Bayesian inference methods. Since Bayesian image reconstruction problems often suffer from computational challenges arising from high dimensionality, these recent developments can be potentially adapted to yield more efficient, accurate and robust image reconstruction algorithms.

Bibliography

- [1] Shah Hussain et al. “Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review”. In: *BioMed Research International* 2022 (June 2022), p. 5164970. ISSN: 2314-6133. DOI: 10.1155/2022/5164970.
- [2] *History of Gamma-Ray Telescopes and Astronomy — Experimental Astronomy*. <https://link.springer.com/article/10.1007/s10686-009-9143-z>.
- [3] Bing Wang et al. “Non-Destructive Testing and Evaluation of Composite Materials/Structures: A State-of-the-Art Review”. In: *Advances in Mechanical Engineering* 12.4 (Apr. 2020), p. 1687814020913761. ISSN: 1687-8132. DOI: 10.1177/1687814020913761.
- [4] Kai Vetter et al. “Gamma-Ray Imaging for Nuclear Security and Safety: Towards 3-D Gamma-Ray Vision”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), pp. 159–168.
- [5] Kimiaki Saito et al. “Detailed Deposition Density Maps Constructed by Large-Scale Soil Sampling for Gamma-Ray Emitting Radioactive Nuclides from the Fukushima Dai-ichi Nuclear Power Plant Accident”. In: *Journal of Environmental Radioactivity* 139 (Jan. 2015), pp. 308–319. ISSN: 0265-931X. DOI: 10.1016/j.jenvrad.2014.02.014.
- [6] Hajir Al Hamrashdi, Stephen D. Monk, and David Cheneler. “Passive Gamma-Ray and Neutron Imaging Systems for National Security and Nuclear Non-Proliferation in Controlled and Uncontrolled Detection Areas: Review of Past and Current Status”. In: *Sensors (Basel, Switzerland)* 19.11 (June 2019), p. 2638. ISSN: 1424-8220. DOI: 10.3390/s19112638.
- [7] V. Schönfelder, A. Hirner, and K. Schneider. “A Telescope for Soft Gamma Ray Astronomy”. In: *Nuclear Instruments and Methods* 107.2 (Mar. 1973), pp. 385–394. ISSN: 0029-554X. DOI: 10.1016/0029-554X(73)90257-7.
- [8] “XCOM: Photon Cross Sections Database”. In: *NIST* (Sept. 2009).
- [9] Yun-sang Lee. “Radiopharmaceuticals for Molecular Imaging”. In: *AACE Clinical Case Reports* 7.1 (2021-01/2021-02), p. 1.

- [10] Carolyn Kierans, Tadayuki Takahashi, and Gottfried Kanbach. *Compton Telescopes for Gamma-ray Astrophysics*. Aug. 2022. DOI: 10.48550/arXiv.2208.07819. arXiv: 2208.07819.
- [11] John A. Tomsick et al. *The Compton Spectrometer and Imager Project for MeV Astronomy*. Sept. 2021. DOI: 10.48550/arXiv.2109.10403. arXiv: 2109.10403.
- [12] M. S. Bandstra et al. “DETECTION AND IMAGING OF THE CRAB NEBULA WITH THE NUCLEAR COMPTON TELESCOPE”. In: *The Astrophysical Journal* 738.1 (Aug. 2011), p. 8. ISSN: 0004-637X. DOI: 10.1088/0004-637X/738/1/8.
- [13] Hideaki Tashima and Taiga Yamaya. “Compton Imaging for Medical Applications”. In: *Radiological Physics and Technology* 15.3 (Sept. 2022), pp. 187–205. ISSN: 1865-0341. DOI: 10.1007/s12194-022-00666-2.
- [14] Laurel Sinclair. *A Compton Gamma Imager for Criminal and National Security Investigation*. Technical Report. Natural Resources Canada, Apr. 2014.
- [15] E. E. Fenimore and T. M. Cannon. “Coded Aperture Imaging with Uniformly Redundant Arrays”. In: *Applied Optics* 17.3 (Feb. 1978), pp. 337–347. ISSN: 2155-3165. DOI: 10.1364/AO.17.000337.
- [16] Cyrille Thomas et al. “X-Ray Pinhole Camera Resolution and Emittance Measurement”. In: *Physical Review Special Topics - Accelerators and Beams* 13.2 (Feb. 2010), p. 022805. DOI: 10.1103/PhysRevSTAB.13.022805.
- [17] Roberto Accorsi. “Design of a Near-Field Coded Aperture Cameras for High-Resolution Medical and Industrial Gamma-Ray Imaging”. Thesis. Massachusetts Institute of Technology, 2001.
- [18] Anico Kulow et al. “Comparison of Three Reconstruction Methods Based on Deconvolution, Iterative Algorithm and Neural Network for X-ray Fluorescence Imaging with Coded Aperture Optics”. In: *Journal of Analytical Atomic Spectrometry* 35.7 (2020), pp. 1423–1434. DOI: 10.1039/D0JA00146E.
- [19] Michelle Galloway et al. “A Combined Compton and Coded-Aperture Telescope for Medium-Energy Gamma-Ray Astrophysics”. In: *Astronomy & Astrophysics* 614 (June 2018), A93. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201731122.
- [20] Emily Frame et al. “Coded Aperture and Compton Imaging for the Development of ^{225}Ac -based Radiopharmaceuticals”. In: *Medical Physics* 50.10 (2023), pp. 6454–6468. ISSN: 2473-4209. DOI: 10.1002/mp.16717.
- [21] E. Caroli et al. “Coded Aperture Imaging in X- and Gamma-Ray Astronomy”. In: *Space Science Reviews* 45.3 (Sept. 1987), pp. 349–403. ISSN: 1572-9672. DOI: 10.1007/BF00171998.
- [22] S.R. Meikle et al. “A Prototype Coded Aperture Detector for Small Animal SPECT”. In: *IEEE Transactions on Nuclear Science* 49.5 (Oct. 2002), pp. 2167–2171. ISSN: 1558-1578. DOI: 10.1109/TNS.2002.803802.

- [23] P. R. T. Munro et al. “Medicine, Material Science and Security: The Versatility of the Coded-Aperture Approach”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2010 (Mar. 2014), p. 20130029. DOI: 10.1098/rsta.2013.0029.
- [24] Michał J. Cieślak, Kelum A. A. Gamage, and Robert Glover. “Coded-Aperture Imaging Systems: Past, Present and Future Development - A Review”. In: *Radiation Measurements* 92 (Sept. 2016), pp. 59–71. ISSN: 1350-4487. DOI: 10.1016/j.radmeas.2016.08.002.
- [25] Kai Vetter et al. “Advances in Nuclear Radiation Sensing: Enabling 3-D Gamma-Ray Vision”. In: *Sensors* 19.11 (Jan. 2019), p. 2541. ISSN: 1424-8220. DOI: 10.3390/s19112541.
- [26] Ross Barnowski et al. “Scene Data Fusion: Real-time Standoff Volumetric Gamma-Ray Imaging”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 800 (Nov. 2015), pp. 65–69. ISSN: 0168-9002. DOI: 10.1016/j.nima.2015.08.016.
- [27] Andrew Haefner et al. “Handheld Real-Time Volumetric 3-D Gamma-Ray Imaging”. In: *Nuclear Instruments and Methods in Physics Research. Section A, Accelerators, Spectrometers, Detectors and Associated Equipment* 857.C (June 2017). ISSN: 0168-9002. DOI: 10.1016/j.nima.2016.11.046.
- [28] J. Hecla et al. “Polaris-LAMP: Multi-Modal 3-D Image Reconstruction With a Commercial Gamma-Ray Imager”. In: *IEEE Transactions on Nuclear Science* 68.10 (Oct. 2021), pp. 2539–2549. ISSN: 1558-1578. DOI: 10.1109/TNS.2021.3110162.
- [29] Brian J. Quiter et al. “Ongoing Advancement of Free-Moving Radiation Imaging and Mapping”. In: *Hard X-Ray, Gamma-Ray, and Neutron Detector Physics XXIV*. Vol. 12241. SPIE, Oct. 2022, pp. 131–137. DOI: 10.1117/12.2635752.
- [30] Daniel Hellfeld et al. “Gamma-Ray Point-Source Localization and Sparse Image Reconstruction Using Poisson Likelihood”. In: *IEEE Transactions on Nuclear Science* 66.9 (Sept. 2019), pp. 2088–2099. ISSN: 0018-9499, 1558-1578. DOI: 10.1109/TNS.2019.2930294.
- [31] Jayson R. Vavrek et al. *Surrogate Distributed Radiological Sources III: Quantitative Distributed Source Reconstructions*. Dec. 2024. DOI: 10.48550/arXiv.2412.02926. arXiv: 2412.02926.
- [32] J. R. Vavrek et al. “Periodic Radiation Mapping of Contaminated Wetlands”. In: *2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD)*. Nov. 2023, pp. 1–1. DOI: 10.1109/NSSMICRTSD49126.2023.10338377.

- [33] Jayson R. Vavrek et al. “Reconstructing the Position and Intensity of Multiple Gamma-Ray Point Sources With a Sparse Parametric Algorithm”. In: *IEEE Transactions on Nuclear Science* 67.11 (Nov. 2020), pp. 2421–2430. ISSN: 1558-1578. DOI: 10.1109/TNS.2020.3024735.
- [34] Kai Vetter. “The Nuclear Legacy Today of Fukushima”. In: *Annual Review of Nuclear and Particle Science* 70. Volume 70, 2020 (Oct. 2020), pp. 257–292. ISSN: 0163-8998, 1545-4134. DOI: 10.1146/annurev-nuc1-101918-023715.
- [35] J. Hadamard. “Sur Les Problèmes Aux Dérivés Partielles et Leur Signification Physique”. In: *Princeton University Bulletin* 13 (1902), pp. 49–52.
- [36] Harrison H. Barrett, Timothy White, and Lucas C. Parra. “List-Mode Likelihood”. In: *Journal of the Optical Society of America. A, Optics, image science, and vision* 14.11 (Nov. 1997), pp. 2914–2923. ISSN: 1084-7529.
- [37] Eric Clarkson and Meredith Kupinski. “Quantifying the Loss of Information from Binning List-Mode Data”. In: *JOSA A* 37.3 (Mar. 2020), pp. 450–457. ISSN: 1520-8532. DOI: 10.1364/JOSAA.375317.
- [38] S.J. Wilderman et al. “List-Mode Maximum Likelihood Reconstruction of Compton Scatter Camera Images in Nuclear Medicine”. In: *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No.98CH36255)*. Vol. 3. Nov. 1998, 1716–1720 vol.3. DOI: 10.1109/NSSMIC.1998.773871.
- [39] Martin J. Willemink and Peter B. Noël. “The Evolution of Image Reconstruction for CT—from Filtered Back Projection to Artificial Intelligence”. In: *European Radiology* 29.5 (2019), pp. 2185–2195. ISSN: 0938-7994. DOI: 10.1007/s00330-018-5810-7.
- [40] Gengsheng Lawrence Zeng. *Medical Image Reconstruction*. Berlin, Heidelberg: Springer, 2010. ISBN: 978-3-642-05367-2 978-3-642-05368-9. DOI: 10.1007/978-3-642-05368-9.
- [41] Gengsheng L. Zeng. “A Filtered Backprojection Algorithm with Characteristics of the Iterative Landweber Algorithm”. In: *Medical Physics* 39.2 (Feb. 2012), pp. 603–607. ISSN: 0094-2405. DOI: 10.1118/1.3673956.
- [42] L. A. Shepp and B. F. Logan. “The Fourier Reconstruction of a Head Section”. In: *IEEE Transactions on Nuclear Science* 21.3 (June 1974), pp. 21–43. ISSN: 1558-1578. DOI: 10.1109/TNS.1974.6499235.
- [43] Gabor T. Herman. *Fundamentals of Computerized Tomography*. Ed. by Sameer Singh. Advances in Pattern Recognition. London: Springer, 2009. ISBN: 978-1-85233-617-2 978-1-84628-723-7. DOI: 10.1007/978-1-84628-723-7.
- [44] Jiyang Chu et al. “Unbiased Filtered Back-Projection in 4 π Compton Imaging With 3D Position Sensitive Detectors”. In: *IEEE Transactions on Nuclear Science* 63.6 (Dec. 2016), pp. 2750–2756. ISSN: 1558-1578. DOI: 10.1109/TNS.2016.2610980.

- [45] Daniel Shy et al. “Filtered Backprojection in Compton Imaging Using a Spherical Harmonic Wiener Filter With Pixelated CdZnTe”. In: *IEEE Transactions on Nuclear Science* 68.2 (Feb. 2021), pp. 211–219. ISSN: 1558-1578. DOI: 10.1109/TNS.2020.3045878.
- [46] M. E. Daube-Witherspoon and G. Muehllehner. “An Iterative Image Space Reconstruction Algorithm Suitable for Volume ECT”. In: *IEEE transactions on medical imaging* 5.2 (1986), pp. 61–66. ISSN: 0278-0062. DOI: 10.1109/TMI.1986.4307748.
- [47] L. A. Shepp and Y. Vardi. “Maximum Likelihood Reconstruction for Emission Tomography”. In: *IEEE Transactions on Medical Imaging* 1.2 (Oct. 1982), pp. 113–122. ISSN: 1558-254X. DOI: 10.1109/TMI.1982.4307558.
- [48] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data Via the EM Algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (Sept. 1977), pp. 1–22. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1977.tb01600.x.
- [49] L. Parra and H.H. Barrett. “List-Mode Likelihood: EM Algorithm and Image Quality Estimation Demonstrated on 2-D PET”. In: *IEEE Transactions on Medical Imaging* 17.2 (Apr. 1998), pp. 228–235. ISSN: 1558-254X. DOI: 10.1109/42.700734.
- [50] J.A. Fessler, N.H. Clinthorne, and W.L. Rogers. “On Complete-Data Spaces for PET Reconstruction Algorithms”. In: *IEEE Transactions on Nuclear Science* 40.4 (Aug. 1993), pp. 1055–1061. ISSN: 1558-1578. DOI: 10.1109/23.256712.
- [51] Eiichi Tanaka. “A Fast Reconstruction Algorithm for Stationary Positron Emission Tomography Based on a Modified EM Algorithm”. In: *IEEE Transactions on Medical Imaging* 6.2 (June 1987), pp. 98–105. ISSN: 1558-254X. DOI: 10.1109/TMI.1987.4307809.
- [52] Anastasios Gaitanis et al. “PET Image Reconstruction: A Stopping Rule for the MLEM Algorithm Based on Properties of the Updating Coefficients”. In: *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society* 34.2 (Mar. 2010), pp. 131–141. ISSN: 1879-0771. DOI: 10.1016/j.compmedimag.2009.07.006.
- [53] Peter J. Green. “On Use of the EM for Penalized Likelihood Estimation”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 52.3 (1990), pp. 443–452. ISSN: 0035-9246. JSTOR: 2345668.
- [54] P.J. Green. “Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm”. In: *IEEE Transactions on Medical Imaging* 9.1 (Mar. 1990), pp. 84–93. ISSN: 1558-254X. DOI: 10.1109/42.52985.
- [55] K. Lange. “Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing”. In: *IEEE Transactions on Medical Imaging* 9.4 (Dec. 1990), pp. 439–446. ISSN: 1558-254X. DOI: 10.1109/42.61759.

- [56] B. A. Mair and Jeff Zahnen. “A Generalization of Green’s One-Step-Late Algorithm for Penalized ML Reconstruction of PET Images”. In: *2006 IEEE Nuclear Science Symposium Conference Record*. Vol. 5. Oct. 2006, pp. 2775–2777. DOI: 10.1109/NSSMIC.2006.356454.
- [57] B. A. Mair, David R. Gilland, and Jing Sun. “Estimation of Images and Nonrigid Deformations in Gated Emission CT”. In: *IEEE transactions on medical imaging* 25.9 (Sept. 2006), pp. 1130–1144. ISSN: 0278-0062. DOI: 10.1109/tmi.2006.879323.
- [58] Mark S. Gockenbach. *Linear Inverse Problems and Tikhonov Regularization*. American Mathematical Soc., Dec. 2016. ISBN: 978-0-88385-141-8.
- [59] V.Y. Panin, G.L. Zeng, and G.T. Gullberg. “Total Variation Regulated EM Algorithm”. In: *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No.98CH36255)*. Vol. 3. Toronto, Ont., Canada: IEEE, 1998, pp. 1562–1566. ISBN: 978-0-7803-5021-2. DOI: 10.1109/NSSMIC.1998.773841.
- [60] Kai Zhang et al. *Plug-and-Play Image Restoration with Deep Denoiser Prior*. July 2021. DOI: 10.48550/arXiv.2008.13751. arXiv: 2008.13751 [cs, eess].
- [61] A. M. Stuart. “Inverse Problems: A Bayesian Perspective”. In: *Acta Numerica* 19 (May 2010), pp. 451–559. ISSN: 1474-0508, 0962-4929. DOI: 10.1017/S0962492910000061.
- [62] Yarin Gal et al. “Bayesian Uncertainty Quantification for Machine-Learned Models in Physics”. In: *Nature Reviews Physics* 4.9 (Sept. 2022), pp. 573–577. ISSN: 2522-5820. DOI: 10.1038/s42254-022-00498-4.
- [63] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177729694.
- [64] Hui Zou. “The Adaptive Lasso and Its Oracle Properties”. In: *Journal of the American Statistical Association* 101.476 (Dec. 2006), pp. 1418–1429. ISSN: 0162-1459. DOI: 10.1198/016214506000000735.
- [65] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer, 2009. ISBN: 978-0-387-84857-0 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7.
- [66] R. Pavlovsky et al. *3D Gamma-ray and Neutron Mapping in Real-Time with the Localization and Mapping Platform from Unmanned Aerial Systems and Man-Portable Configurations*. Aug. 2019. DOI: 10.48550/arXiv.1908.06114. arXiv: 1908.06114.
- [67] Michael Bondin. “Development of a Small Unmanned Aerial System Composed of Structural Pulse Shape Discriminating Plastic Scintillators”. PhD thesis. UC Berkeley, 2023.

- [68] Johnathan M. Bardsley and John Golde. “Regularization Parameter Selection Methods for Ill-Posed Poisson Maximum Likelihood Estimation”. In: *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data* 25.9 (Aug. 2009), p. 095005. ISSN: 0266-5611. DOI: 10.1088/0266-5611/25/9/095005.
- [69] Andrew Gelman et al. *Bayesian Data Analysis, Third Edition*. CRC Press, Nov. 2013. ISBN: 978-1-4398-4095-5.
- [70] David JC MacKay. “Introduction to Gaussian Processes”. In: *NATO ASI series F computer and systems sciences* 168 (1998), pp. 133–166.
- [71] *Gaussian Processes for Machine Learning: Contents*.
- [72] S. L. Cotter et al. “MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster”. In: *Statistical Science* 28.3 (Aug. 2013). ISSN: 0883-4237. DOI: 10.1214/13-STS421. arXiv: 1202.0709.
- [73] David J. C. MacKay. “A Practical Bayesian Framework for Backpropagation Networks”. In: *Neural Computation* 4.3 (May 1992), pp. 448–472. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.3.448.
- [74] Erik Daxberger et al. *Laplace Redux – Effortless Bayesian Deep Learning*. Mar. 2022. DOI: 10.48550/arXiv.2106.14806. arXiv: 2106.14806 [cs, stat].
- [75] Qingping Zhou et al. “Bayesian Inference and Uncertainty Quantification for Medical Image Reconstruction with Poisson Data”. In: *SIAM Journal on Imaging Sciences* 13.1 (Jan. 2020), pp. 29–52. ISSN: 1936-4954. DOI: 10.1137/19M1248352.
- [76] Thomas Bayes and null Price. “LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S”. In: *Philosophical Transactions of the Royal Society of London* 53 (Jan. 1997), pp. 370–418. DOI: 10.1098/rstl.1763.0053.
- [77] David Duvenaud. “Automatic Model Construction with Gaussian Processes”. PhD thesis. University of Cambridge, Nov. 2014.
- [78] Mikkil N. Schmidt and Hans Laurberg. “Nonnegative Matrix Factorization with Gaussian Process Priors”. In: *Computational Intelligence and Neuroscience* 2008 (2008), pp. 1–10. ISSN: 1687-5265, 1687-5273. DOI: 10.1155/2008/361705.
- [79] Dong C. Liu and Jorge Nocedal. “On the Limited Memory BFGS Method for Large Scale Optimization”. In: *Mathematical Programming* 45.1 (Aug. 1989), pp. 503–528. ISSN: 1436-4646. DOI: 10.1007/BF01589116.
- [80] Yunus Saatçi. “Scalable Inference for Structured Gaussian Process Models”. PhD thesis. Citeseer, 2012.
- [81] V. Pereyra and G. Scherer. “Efficient Computer Manipulation of Tensor Products with Applications to Multidimensional Approximation”. In: *Mathematics of Computation* 27.123 (1973), pp. 595–605. ISSN: 0025-5718. DOI: 10.2307/2005663.

- [82] Xiaolong Ouyang et al. “Incorporation of Correlated Structural Images in PET Image Reconstruction”. In: *IEEE Transactions on Medical Imaging* 13.4 (Dec. 1994), pp. 627–640. ISSN: 1558-254X. DOI: 10.1109/42.363105.
- [83] Christine Bathke et al. “Improved Image Reconstruction in Magnetic Particle Imaging Using Structural a Priori Information”. In: *International Journal on Magnetic Particle Imaging IJMPI* 3.1 (Mar. 2017). ISSN: 2365-9033.
- [84] Guobao Wang and Jinyi Qi. “PET Image Reconstruction Using Kernel Method”. In: *IEEE Transactions on Medical Imaging* 34.1 (Jan. 2015), pp. 61–71. ISSN: 1558-254X. DOI: 10.1109/TMI.2014.2343916.
- [85] Joshua S. Speagle. *A Conceptual Introduction to Markov Chain Monte Carlo Methods*. Mar. 2020. DOI: 10.48550/arXiv.1909.12313. arXiv: 1909.12313 [astro-ph, physics:physics, stat].
- [86] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. *Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions*. Dec. 2010. DOI: 10.48550/arXiv.0909.4061. arXiv: 0909.4061 [math].
- [87] Joel A. Tropp et al. “Practical Sketching Algorithms for Low-Rank Matrix Approximation”. In: *SIAM Journal on Matrix Analysis and Applications* 38.4 (Jan. 2017), pp. 1454–1485. ISSN: 0895-4798. DOI: 10.1137/17M1111590.
- [88] Max A. Woodbury. *Inverting Modified Matrices*. Princeton University, Princeton, NJ, 1950.
- [89] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998. ISBN: 978-0-521-78450-4. DOI: 10.1017/CB09780511802256.
- [90] *Pierre-Simon Laplace - Œuvres Complètes, Tome 6*.
- [91] O. G. Sutton. “The Problem of Diffusion in the Lower Atmosphere”. In: *Quarterly Journal of the Royal Meteorological Society* 73.317-318 (1947), pp. 257–281. ISSN: 1477-870X. DOI: 10.1002/qj.49707331704.
- [92] Jaehong Jeong, Mikyoung Jun, and Marc G. Genton. “Spherical Process Models for Global Spatial Statistics”. In: *Statistical science : a review journal of the Institute of Mathematical Statistics* 32.4 (Nov. 2017), pp. 501–513. ISSN: 0883-4237. DOI: 10.1214/17-STS620.
- [93] Raj Kumar Parajuli et al. “Development and Applications of Compton Camera—A Review”. In: *Sensors (Basel, Switzerland)* 22.19 (Sept. 2022), p. 7374. ISSN: 1424-8220. DOI: 10.3390/s22197374.
- [94] Michał J. Cieślak, Kelum A. A. Gamage, and Robert Glover. “Coded-Aperture Imaging Systems: Past, Present and Future Development – A Review”. In: *Radiation Measurements* 92 (Sept. 2016), pp. 59–71. ISSN: 1350-4487. DOI: 10.1016/j.radmeas.2016.08.002.

- [95] Kenneth M. Hanson and George W. Wecksung. “Bayesian Approach to Limited-Angle Reconstruction in Computed Tomography”. In: *JOSA* 73.11 (Nov. 1983), pp. 1501–1509. DOI: 10.1364/JOSA.73.001501.
- [96] Ciprian Catana. “The Dawn of a New Era in Low-Dose PET Imaging”. In: *Radiology* 290.3 (Mar. 2019), pp. 657–658. ISSN: 0033-8419. DOI: 10.1148/radiol.2018182573.
- [97] Riccardo Barbano et al. “Chapter 26 - Uncertainty Quantification in Medical Image Synthesis”. In: *Biomedical Image Synthesis and Simulation*. Ed. by Ninon Burgos and David Svoboda. The MICCAI Society Book Series. Academic Press, Jan. 2022, pp. 601–641. ISBN: 978-0-12-824349-7. DOI: 10.1016/B978-0-12-824349-7.00033-5.
- [98] *Quantitative Compton Imaging in 3D — INMM Resources*.
- [99] George Sgouros and Robert F. Hobbs. “Dosimetry for Radiopharmaceutical Therapy”. In: *Seminars in nuclear medicine* 44.3 (May 2014), pp. 172–178. ISSN: 0001-2998. DOI: 10.1053/j.semnuclmed.2014.03.007.
- [100] Antje-Christin Knopf and Antony Lomax. “In Vivo Proton Range Verification: A Review”. In: *Physics in Medicine and Biology* 58.15 (Aug. 2013), R131–160. ISSN: 1361-6560. DOI: 10.1088/0031-9155/58/15/R131.
- [101] *4-Pi Compton Imaging with Single 3D Position-Sensitive CdZnTe Detector*.
- [102] Daniel Hellfeld et al. “A Spherical Active Coded Aperture for 4 π Gamma-Ray Imaging”. In: *IEEE Transactions on Nuclear Science* 64.11 (Nov. 2017), pp. 2837–2842. ISSN: 1558-1578. DOI: 10.1109/TNS.2017.2755982.
- [103] Daniel Hellfeld et al. “Real-Time Free-Moving Active Coded Mask 3D Gamma-Ray Imaging”. In: *IEEE Transactions on Nuclear Science* 66.10 (Oct. 2019), pp. 2252–2260. ISSN: 1558-1578. DOI: 10.1109/TNS.2019.2939948.
- [104] Andrew Gordon Wilson and Hannes Nickisch. *Kernel Interpolation for Scalable Structured Gaussian Processes (KISS-GP)*. Mar. 2015. DOI: 10.48550/arXiv.1503.01057. arXiv: 1503.01057 [cs, stat].
- [105] Finn Lindgren, Håvard Rue, and Johan Lindström. “An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4 (2011), pp. 423–498. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2011.00777.x.
- [106] Andrew Gordon Wilson and Ryan Prescott Adams. *Gaussian Process Kernels for Pattern Discovery and Extrapolation*. Dec. 2013. DOI: 10.48550/arXiv.1302.4245. arXiv: 1302.4245 [cs, stat].
- [107] Andrew Gordon Wilson et al. *Deep Kernel Learning*. Nov. 2015. DOI: 10.48550/arXiv.1511.02222. arXiv: 1511.02222 [cs, stat].

- [108] Frederik Kunstner, Philipp Hennig, and Lukas Balles. “Limitations of the Empirical Fisher Approximation for Natural Gradient Descent”. In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc., 2019.
- [109] Ryan Pavlovsky et al. *3-D Radiation Mapping in Real-Time with the Localization and Mapping Platform LAMP from Unmanned Aerial Systems and Man-Portable Configurations*. <https://arxiv.org/abs/1901.05038v1>. Dec. 2018.
- [110] Jaewon Lee et al. “Radiation Image Reconstruction and Uncertainty Quantification Using a Gaussian Process Prior”. In: *Scientific Reports* 14.1 (Oct. 2024), p. 22958. ISSN: 2045-2322. DOI: 10.1038/s41598-024-71336-z.
- [111] Glenn F. Knoll. *Radiation Detection and Measurement*. Wiley, 1979. ISBN: 978-0-471-49545-1.
- [112] Huamin Zhang and Feng Ding. “On the Kronecker Products and Their Applications”. In: *Journal of Applied Mathematics* 2013.1 (2013), p. 296185. ISSN: 1687-0042. DOI: 10.1155/2013/296185.
- [113] Jingu Kim, Yunlong He, and Haesun Park. “Algorithms for Nonnegative Matrix and Tensor Factorizations: A Unified View Based on Block Coordinate Descent Framework”. In: *Journal of Global Optimization* 58.2 (Feb. 2014), pp. 285–319. ISSN: 1573-2916. DOI: 10.1007/s10898-013-0035-4.
- [114] Daniel Lee and H. Sebastian Seung. “Algorithms for Non-negative Matrix Factorization”. In: *Advances in Neural Information Processing Systems*. Vol. 13. MIT Press, 2000.
- [115] Daniel D. Lee and H. Sebastian Seung. “Learning the Parts of Objects by Non-Negative Matrix Factorization”. In: *Nature* 401.6755 (Oct. 1999), pp. 788–791. ISSN: 1476-4687. DOI: 10.1038/44565.
- [116] Crow. *Lognormal Distributions: Theory and Applications*. New York: Routledge, May 2018. ISBN: 978-0-203-74866-4. DOI: 10.1201/9780203748664.
- [117] John S. Bridle. “Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition”. In: *Neurocomputing*. Ed. by Françoise Fogelman Soulié and Jeanny Hérault. Berlin, Heidelberg: Springer, 1990, pp. 227–236. ISBN: 978-3-642-76153-9. DOI: 10.1007/978-3-642-76153-9_28.
- [118] Atılım Günes Baydin et al. “Automatic Differentiation in Machine Learning: A Survey”. In: *J. Mach. Learn. Res.* 18.1 (Jan. 2017), pp. 5595–5637. ISSN: 1532-4435.
- [119] Pierre-Simon de (1749-1827) Auteur du texte Laplace. *Oeuvres Complètes de Laplace. Tome 6 / Publiées Sous Les Auspices de l’Académie Des Sciences, Par MM. Les Secrétaires Perpétuels*. 1878.
- [120] *David MacKay: Information Theory, Inference, and Learning Algorithms: The Book*. <https://www.inference.org.uk/mackay/itila/book.html>.

- [121] Bradley P. Carlin and Thomas A. Louis. “Empirical Bayes: Past, Present and Future”. In: *Journal of the American Statistical Association* 95.452 (2000), pp. 1286–1289. ISSN: 0162-1459. DOI: 10.2307/2669771. JSTOR: 2669771.
- [122] Aravind Rajeswaran et al. *Meta-Learning with Implicit Gradients*. Sept. 2019. DOI: 10.48550/arXiv.1909.04630. arXiv: 1909.04630.
- [123] Jonathan Lorraine, Paul Vicol, and David Duvenaud. *Optimizing Millions of Hyperparameters by Implicit Differentiation*. Nov. 2019. DOI: 10.48550/arXiv.1911.02590. arXiv: 1911.02590.
- [124] Boyan Gao et al. “Loss Function Learning for Domain Generalization by Implicit Gradient”. In: *Proceedings of the 39th International Conference on Machine Learning*. PMLR, June 2022, pp. 7002–7016.
- [125] Jie Ren et al. *TorchOpt: An Efficient Library for Differentiable Optimization*. Nov. 2022. DOI: 10.48550/arXiv.2211.06934. arXiv: 2211.06934.
- [126] Joshua J. Bon et al. “Being Bayesian in the 2020s: Opportunities and Challenges in the Practice of Modern Applied Bayesian Statistics”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381.2247 (Mar. 2023), p. 20220156. DOI: 10.1098/rsta.2022.0156.
- [127] Haitao Liu et al. “When Gaussian Process Meets Big Data: A Review of Scalable GPs”. In: *IEEE Transactions on Neural Networks and Learning Systems* 31.11 (Nov. 2020), pp. 4405–4423. ISSN: 2162-2388. DOI: 10.1109/TNNLS.2019.2957109.
- [128] Riley Murray et al. *Randomized Numerical Linear Algebra : A Perspective on the Field With an Eye to Software*. Apr. 2023. DOI: 10.48550/arXiv.2302.11474. arXiv: 2302.11474.
- [129] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Dec. 2019. DOI: 10.48550/arXiv.1912.01703. arXiv: 1912.01703.
- [130] Martín Abadi et al. *TensorFlow: A System for Large-Scale Machine Learning*. May 2016. DOI: 10.48550/arXiv.1605.08695. arXiv: 1605.08695.
- [131] MatricesZhaojun Bai and G. Golub. “Bounds for the Trace of the Inverse and the Determinant of Symmetric Positive Definite”. In: 1996.
- [132] Zhaojun Bai, Gark Fahey, and Gene Golub. “Some Large-Scale Matrix Computation Problems”. In: *Journal of Computational and Applied Mathematics* 74.1 (Nov. 1996), pp. 71–89. ISSN: 0377-0427. DOI: 10.1016/0377-0427(96)00018-0.
- [133] Gene H. Golub and Zdeněk Strakoš. “Estimates in Quadratic Formulas”. In: *Numerical Algorithms* 8.2 (Sept. 1994), pp. 241–268. ISSN: 1572-9265. DOI: 10.1007/BF02142693.
- [134] Gene H. Golub and Gérard Meurant. *Matrices, Moments and Quadrature with Applications*. Princeton University Press, 2010. ISBN: 978-0-691-14341-5. JSTOR: j.ctt7tbvs.

- [135] Gene Howard Golub and Charles F. Van Loan. *Matrix Computations*. JHU Press, Feb. 2013. ISBN: 978-1-4214-0794-4.
- [136] Shashanka Ubaru, Jie Chen, and Yousef Saad. “Fast Estimation of $\text{tr}(f(A))$ via Stochastic Lanczos Quadrature”. In: *SIAM Journal on Matrix Analysis and Applications* 38.4 (Jan. 2017), pp. 1075–1099. ISSN: 0895-4798. DOI: 10.1137/16M1104974.
- [137] Jane K. Cullum and Ralph A. Willoughby. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan. 2002. ISBN: 978-0-89871-523-1. DOI: 10.1137/1.9780898719192.
- [138] Yousef Saad. *Numerical Methods for Large Eigenvalue Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Jan. 2011. ISBN: 978-1-61197-072-2. DOI: 10.1137/1.9781611970739.
- [139] J. A. Hartigan. “Asymptotic Normality of Posterior Distributions”. In: *Bayes Theory*. Ed. by J. A. Hartigan. New York, NY: Springer, 1983, pp. 107–118. ISBN: 978-1-4613-8242-3. DOI: 10.1007/978-1-4613-8242-3_11.
- [140] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. *Variational Inference: A Review for Statisticians*. May 2018. DOI: 10.48550/arXiv.1601.00670. arXiv: 1601.00670.
- [141] Thomas P. Minka. *Expectation Propagation for Approximate Bayesian Inference*. Jan. 2013. DOI: 10.48550/arXiv.1301.2294. arXiv: 1301.2294.
- [142] Alexandros Beskos et al. *Geometric MCMC for Infinite-Dimensional Inverse Problems*. Jan. 2017. DOI: 10.48550/arXiv.1606.06351. arXiv: 1606.06351.
- [143] Shotaro Yamaguchi et al. “Bayesian Penalized-Likelihood Reconstruction Algorithm Suppresses Edge Artifacts in PET Reconstruction Based on Point-Spread-Function”. In: *Physica medica: PM: an international journal devoted to the applications of physics to medicine and biology: official journal of the Italian Association of Biomedical Physics (AIFB)* 47 (Mar. 2018), pp. 73–79. ISSN: 1724-191X. DOI: 10.1016/j.ejmp.2018.02.013.
- [144] Sangtae Ahn et al. “Quantitative Comparison of OSEM and Penalized Likelihood Image Reconstruction Using Relative Difference Penalties for Clinical PET”. In: *Physics in Medicine and Biology* 60.15 (Aug. 2015), pp. 5733–5751. ISSN: 1361-6560. DOI: 10.1088/0031-9155/60/15/5733.
- [145] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, May 2015. ISBN: 978-1-4987-1217-0.
- [146] Yi-Ting Guo, Qin-Qin Li, and Chun-Sheng Liang. “The Rise of Nonnegative Matrix Factorization: Algorithms and Applications”. In: *Information Systems* 123 (July 2024), p. 102379. ISSN: 0306-4379. DOI: 10.1016/j.is.2024.102379.

- [147] Laurent Valentin Jospin et al. “Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users”. In: *IEEE Computational Intelligence Magazine* 17.2 (May 2022), pp. 29–48. ISSN: 1556-6048. DOI: 10.1109/MCI.2022.3155327.
- [148] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. *Deep Image Prior*. May 2020. DOI: 10.48550/arXiv.1711.10925. arXiv: 1711.10925.
- [149] Razvan V. Marinescu, Daniel Moyer, and Polina Golland. *Bayesian Image Reconstruction Using Deep Generative Models*. Dec. 2021. DOI: 10.48550/arXiv.2012.04567. arXiv: 2012.04567.
- [150] Berthy T. Feng et al. *Score-Based Diffusion Models as Principled Priors for Inverse Imaging*. Aug. 2023. DOI: 10.48550/arXiv.2304.11751. arXiv: 2304.11751.