

UNIVERSITY OF CALGARY

Energetic Electron Precipitation Measurements from High Altitude Platforms

by

Matthew Patrick

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

SPACE PHYSICS

CALGARY, ALBERTA

April, 2021

© Matthew Patrick 2021

Abstract

Acknowledgements

Paragraph 1

Paragraph 2

Paragraph 3

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
List of Symbols	ix
1 The Inverse Problem	1
1.1 Definition and Properties	1
1.2 Framing the X-ray Inversion Problem	3
1.3 Creating Synthetic Data	8
1.4 The Least-Squares Regression and the Bias-Variance Trade-off	10
1.5 Biased Estimators and Improving the Condition Number	13
1.6 Tikhonov Regularization and the L-Curve Method	18
1.7 Higher Order Regularization	27
1.8 Preconditioning Techniques	30
1.9 Constrained Optimization	40
1.10 Evaluation on Synthetic Data	42
Bibliography	52
A First Appendix	53

List of Tables

List of Figures and Illustrations

1.1	(Top): Example of three different assumed electron energy spectra (mono-energetic, exponentially decreasing, and gaussian) for a precipitating electron flux of 1×10^9 electrons per cm^2 per second. (Bottom): Resulting simulated X-ray spectra for each electron spectrum, using the methods developed in Chapter 3.	4
1.2	Condition number of mapping \mathbf{G} as a function of X-ray spectrum and electron spectrum bin-widths in KeV. The white region is the domain where the condition number is infinite, and \mathbf{G} is singular.	6
1.3	Direct matrix inversion applied to simulated exponential electron spectra using bin widths of 10 (A), 50 (B), and 100 (C) KeV. The total incident electron count is 1×10^8	7
1.4	Example of over-fit and under-fit models to data. A random data set is generated and polynomial models are fit using least-squares regression. The reduced χ^2 statistic is shown for each model.	12
1.5	Plots of the singular values of the kernel matrix for the X-ray inversion problem for different electron spectrum bin widths, with singular values arranged from biggest to smallest.	15
1.6	Plots of the condition number of the kernel matrix for the X-ray inversion problem as a function of electron spectrum and X-ray spectrum bin width, for 1 (A) through 6 (F) truncated singular values.	16
1.7	Plots of simulated electron spectra for a total electron flux of 1×10^9 electrons/ cm^2 /second, and their reconstruction from the associated simulated X-ray spectra using a truncated singular value approximation of the kernel matrix. Electron bin widths are set to 50 KeV, and X-ray bin widths to 1 KeV. (A), (B), and (C) show reconstructions with the smallest 1, 5, and 10 singular values dropped from the expansion. . .	19
1.8	Application of Tikhonov regularization and the “L-curve” method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).	21
1.9	Random data set (blue) with under-fit (orange), and over-fit (green) representations. A particular data point (red) is removed prior to fitting. The removed data point is far from both the over-fit and under-fit representations of the data.	25
1.10	Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).	26
1.11	Application of second-order Tikhonov regularization to synthetic X-ray measurements corresponding to two mono-energetic electron beams at different X-ray fluxes and noise levels (A - 10^7 electrons / sec / keV, B - 10^6 electrons / sec / keV, C - 5×10^5 electrons / sec / keV)	29
1.12	(left): Graph of the rows of the X-ray response matrix. A high degree of collinearity is apparent, as with a large scale change between rows. (right): Graph of the singular values of the X-ray response matrix, from largest to smallest.	33

1.13	(top): Contour plot of the condition number of the X-ray response matrix as a function of electron spectrum and X-ray spectrum bin widths, with no preconditioner. (bottom): The same contour plot with diagonal scaling preconditioner applied. . .	35
1.14	(top left): Rows of the response matrix with an optimized diagonal preconditioner applied. (top right): corresponding spectra of singular values for the response matrix. (bottom): condition number as a function of iteration number for a conjugate gradient optimization of the preconditioner matrix entries	36
1.15	Response matrix for the X-ray inversion problem. Scale differences across both rows and columns of the matrix contribute to the high condition number. Units are normalized per incident electron.	38
1.16	Response matrix for the X-ray inversion problem with left and right preconditioners applied. Scale differences across both rows and columns of the matrix have been reduced, improving the condition number. The units on the color axis are no longer physical, having been scaled by the preconditioning matrices.	39
1.17	Response matrix condition numbers as a function of X-ray and electron spectrum bin widths for: (A) - no preconditioning, (B) - left scaling preconditioner only, (C) - right preconditioner only, (D) - both left and right preconditioners.	41
1.18	Retrieval of electron spectra from synthetic X-ray data across different simulated count rates. Left and right preconditioners, second-order regularization, and a non-negativity constraint are applied.	43
1.19	Retrieval of exponential electron spectra (100 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.	44
1.20	Retrieval of exponential electron spectra (250 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.	45
1.21	Retrieval of exponential electron spectra (400 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.	46
1.22	Retrieval of gaussian electron spectra (50 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.	47
1.23	Retrieval of gaussian electron spectra (100 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation. . .	48

1.24	Retrieval of gaussian electron spectra (150 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation. . .	49
1.25	Retrieval of monoenergetic electron spectra from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.	50

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
U of C	University of Calgary

Chapter 1

The Inverse Problem

1.1 Definition and Properties

In this chapter the problem of obtaining information about the causative electron precipitation directly from X-ray measurements will be examined. We will term this the “X-ray inversion problem.” This type of analysis belongs to a broad class of what are called inverse problems, which attempt to reconstruct physical events based on limited measurement data. A classic example of an inverse problem is the determination of media velocities within the Earth based on seismic measurements at discrete points. The “forward problem” of predicting instrument responses based on an assumed physical event can be handled by, for example, computer simulations, as was done in the last chapter for the problem of predicting the measured X-ray distributions which result from a given electron precipitation event. Unlike forward problems, inverse problems have mathematical properties which make them difficult or impossible to solve without significant a-priori assumptions. For the X-ray inversion problem, we will show the reasons for this, and demonstrate some measures which can be taken to mitigate these effects.

Define the forward problem by the mapping:

$$\mathbf{F} : \mathbf{M} \rightarrow \mathbf{D} \tag{1.1}$$

where \mathbf{D} is the space containing data from measurements, and \mathbf{M} is the space of models, or solutions. The mapping \mathbf{F} may be either linear, or non-linear. Additionally, the definition of the model \mathbf{M} is generally not unique [3]. For the forward problem, we start with an assumed model and evaluate its predictions against the available data. The metric by which this is done is another

choice which may not be unique for the given problem. The inverse problem is:

$$\mathbf{F}^{-1} : \mathbf{D} \rightarrow \mathbf{M} \quad (1.2)$$

In addition to uncertainty in the measurements \mathbf{D} , knowledge of the mapping \mathbf{F} may be incomplete, or based on uncertain experimental data or simulations. To describe the challenges associated with inverse problems some definitions are needed.

Definition 1.1.1 (ill-posedness). A problem is ill-posed if it fails to have any of the following properties [2] :

1. A solution exists.
2. The solution depends continuously on the input data.
3. The solution is unique.

Inverse problems are ill-posed as a general rule. This requires that we shift the analysis from a search for a particular solution, as with forward problems, towards a search for families of solutions, and methods of determining which produce the best description of the available data, often requiring the use of a-priori information about the particular problem at hand. Based on the results of the last chapter, the X-ray inversion problem certainly satisfies property 1, that a solution exists. The difficulties lie in its failure to satisfy properties 2 and 3. We will show that without special treatment, finding a unique solution to the X-ray inversion problem is, in most realistic cases, not possible. Another definition is needed to see why:

Definition 1.1.2 (ill-conditioning). A problem is ill conditioned if a small change in the input data results in a large change in the solution, that is, the condition number[1]:

$$\kappa = \lim_{\epsilon \rightarrow 0} \sup_{\|\delta x\| \leq \epsilon} \frac{\|\delta f\|}{\|\delta x\|}$$

is large for small changes in the kernel function δf and input data δx .

The X-ray inversion problem is ill-conditioned. Figure 1.1 illustrates this by showing three different model precipitating electron spectra, and a simulation of their resulting X-ray spectra in log-linear space. A small change in the measured X-ray spectrum results in a drastically different model electron spectrum. The treatment of this instability will be the main problem addressed in this chapter.

1.2 Framing the X-ray Inversion Problem

The forward problem of determining expected instrument responses from electron precipitation events is linear, and can be written as:

$$\mathbf{G}\mathbf{m} = \mathbf{d}$$

where \mathbf{G} is a matrix representing the discretization of the kernel function mapping precipitating electron spectra \mathbf{m} to resulting measurements of the X-ray spectra \mathbf{d} . Techniques for computing an approximation of \mathbf{G} were discussed in the previous chapter. There is a result that allows the calculation of the condition number of \mathbf{G} as the ratio of its maximal and minimal singular values. As a general rule, if a given linear problem has condition number κ , then we can expect to lose 10^κ digits of precision in calculating the solution. This is a best-case estimate, since there will also be round-off errors caused by computer floating point representations of the matrix and the data. Techniques to frame the X-ray inversion problem in such a way as to reduce the effect of the ill-conditioning of \mathbf{G} will be discussed in the this section.

A representation of matrix \mathbf{G} was developed in Chapter 3 using monte-carlo simulations. In that analysis, a Green's function approach is used to write the X-ray spectra resulting from arbitrary input electron spectra as a linear combination of the effects of mono-energetic electron beams. The dimensions of matrix G are determined by the binning scheme used to represent electron and X-ray energy spectra. For the forward problem, the particular binning scheme chosen for both spaces is not critical, provided it is fine enough to capture the scale of the details of interest in both spectra,

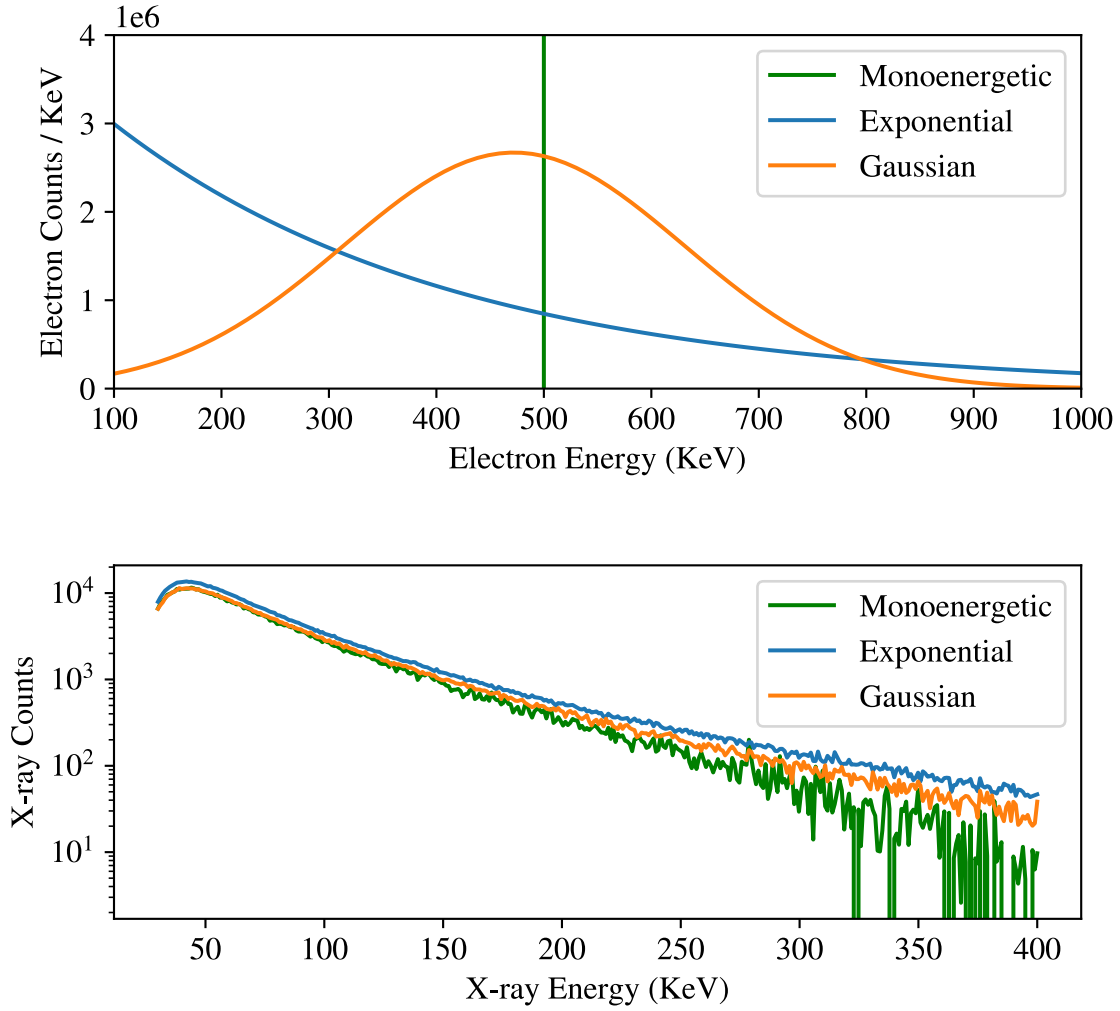


Figure 1.1: (Top): Example of three different assumed electron energy spectra (mono-energetic, exponentially decreasing, and gaussian) for a precipitating electron flux of 1×10^9 electrons per cm^2 per second. (Bottom): Resulting simulated X-ray spectra for each electron spectrum, using the methods developed in Chapter 3.

but coarse enough to support obtaining adequately pure statistics for the resulting X-ray spectra in a practical amount of computing time. For the inverse X-ray problem, however, the binning scheme chosen for \mathbf{G} has a critical effect on the conditioning of the inverse problem, and can determine whether solutions are possible to obtain entirely. To show the effect of choosing different binning schemes for the rows and columns of \mathbf{G} , we plot the condition number as a function of bin widths in KeV for the input electron spectra and output X-ray spectra in Figure 1.2.

Figure 1.2 shows that there is a set of bin widths for the electron and X-ray spectra which make \mathbf{G} singular, which implies that it has no inverse. There are also some general trends apparent. Smaller bin widths for the X-ray spectrum tend to reduce the condition number of the problem, as do larger bin widths for the electron spectra. This can be viewed as the effect of supplying the problem with more information (more X-ray data points), and requiring less information from the output, corresponding to larger electron bin widths. The singular region of Figure 1.2 corresponds to binning schemes where too much information is being sought from a limited input data set for there to be a unique solution.

The matrix \mathbf{G} is not necessarily square. The inverse used in the non-square case is the Moore-Penrose pseudo inverse, which corresponds to solving the linear system in the least-squares sense. Results from attempting to solve the inverse problem by directly inverting \mathbf{G} are shown for simulated X-ray and electron spectra across different binning schemes in Figure 1.2.

The direct inversion attempts in Figure 1.2 are an example of a fundamental trade-off encountered when solving inverse problems. Because the information contained in the available measurements is subset to the information contained in the physical event being studied, the best solution, in terms of minimizing residuals between predictions of the resulting model and the data, often has a high variance and contains little useful information. This corresponds to case (C) in Figure 1.2, where the electron spectrum bin widths are small enough that the model obtained is mostly random noise. On the other hand, choosing wide bin widths as in case (A) can leave relevant information contained in the measurements out of the model, so that while the variance in the solution is low,

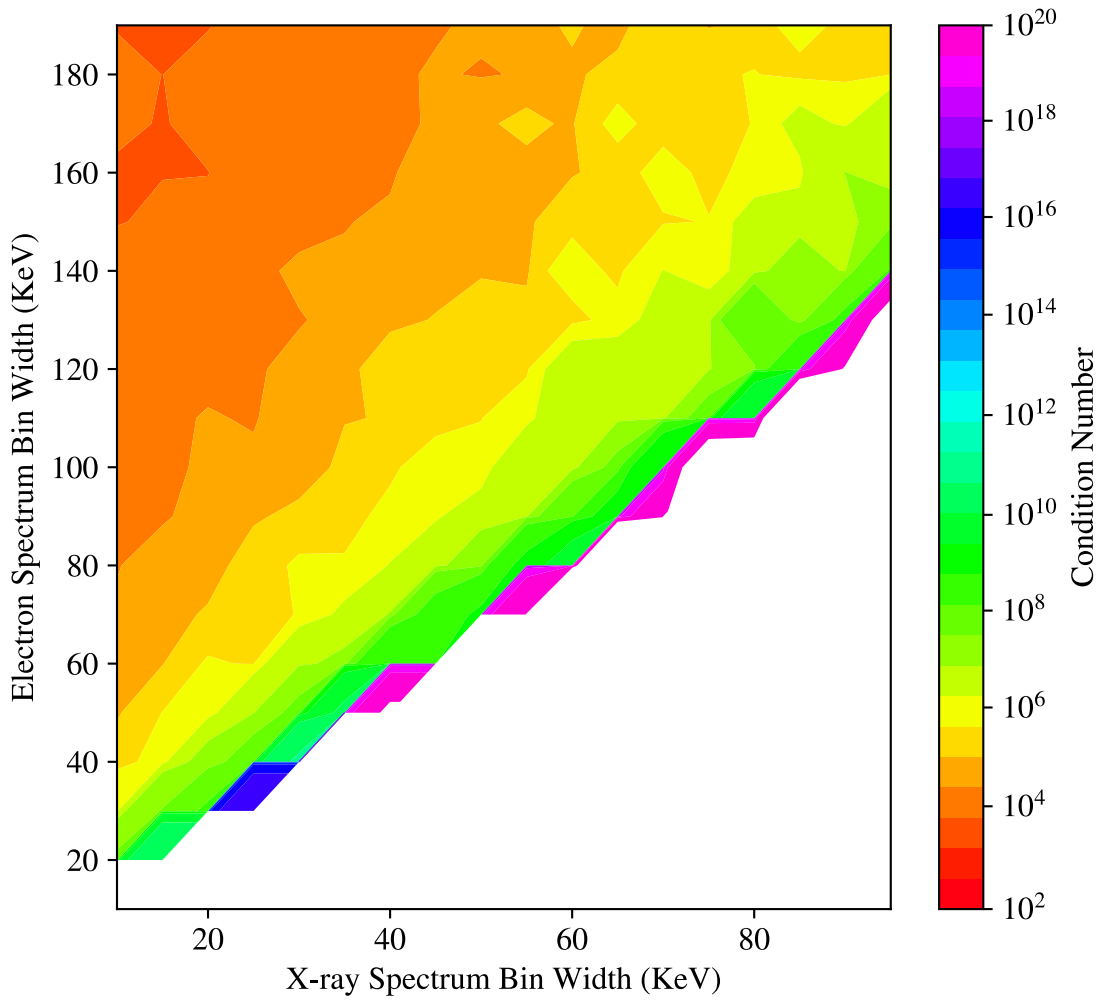


Figure 1.2: Condition number of mapping \mathbf{G} as a function of X-ray spectrum and electron spectrum bin-widths in KeV. The white region is the domain where the condition number is infinite, and \mathbf{G} is singular.

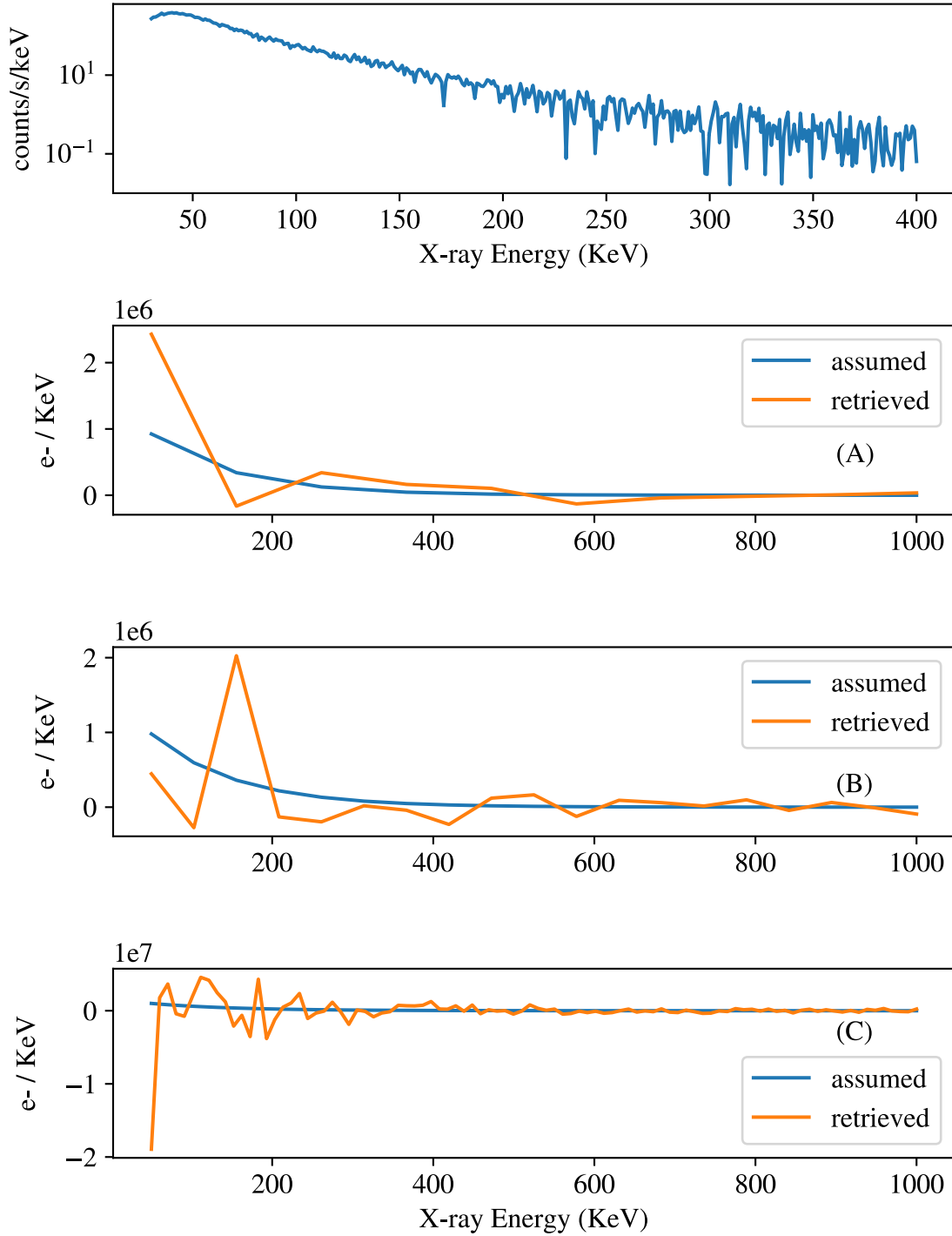


Figure 1.3: Direct matrix inversion applied to simulated exponential electron spectra using bin widths of 10 (A), 50 (B), and 100 (C) KeV. The total incident electron count is 1×10^8 .

the model is still not descriptive of the event being studied. This trade-off can be handled through optimization techniques, and the application of these will be the main focus of this chapter.

1.3 Creating Synthetic Data

The evaluation of techniques to solve inverse problems often depends on the use of synthetically generated, or simulated, data which correspond to solutions which are already known. For the X-ray inversion problem, synthetic data are generated using the monte-carlo techniques discussed in Chapter 3. Since the forward problem is linear, X-ray spectra can be generated which correspond to arbitrary electron spectra using a Greens function approach at little computational expense.

Realistic synthetic data sets always have an error term added, which must be consistent with the statistics of the experiment being simulated. Measuring X-ray spectra is a counting experiment, and therefore the random error associated with the recorded count for a given energy bin has a Poisson distribution. This distribution can be treated as approximately normal for total counts greater than around 10 in each energy bin due to the central limit theorem. This property can be used to simulate an X-ray spectrum which corresponds to an assumed count rate, by adding appropriate noise to the signal produced using the techniques of Chapter 3.

In a real experiment, a measured X-ray spectrum must be background subtracted. The background subtraction is difficult to simulate and is a source of systematic error. Natural sources such as the 511 KeV annihilation line couple with noise inherent to the detector to make a background energy spectrum with a characteristic shape. We will use a sample drawn from experimental data sets to generate a characteristic background spectrum, and then add an appropriately scaled version to synthetic data generated for a given electron distribution and count rate. This approach neglects any systematic error caused by, for example, detector calibration drifting due to temperature changes.

Due to the sensitivity of the X-ray inversion problem to small perturbations in the measured spectrum, care must be taken when applying tests using synthetic data so as not to introduce a sam-

pling bias in the measured performance of different inversion techniques. It is possible to combine a particular background subtraction with a simulated spectrum and corresponding measurement noise to obtain a solution to the inverse problem which very nearly matches the assumed causative electron spectrum through random chance. This may not occur when the same test is subsequently run with a different sample of the background spectrum, or a different random seed for the simulated detector noise. For this reason, single tests of an inversion method on one synthetically generated spectrum have no significance. Evaluating a solution technique to the inverse problem on synthetic data is itself an experiment, and must be treated using a large enough sample size.

The data available for this project was obtained using multiple detectors which all had an identical design (discussed in Chapters 2 and 3). Since the data set is large, many background samples can be generated for use when creating synthetic data. An example of a “quiet time” X-ray spectrum is shown in Figure ???. Over 20 of these background spectra are collected across the available data set, and used to create the average background spectrum shown in Figure ??.

Simulated X-ray spectra which are generated using the techniques of Chapter 3 are created without a background component. To account for this in the generation of synthetic data, the average of the available measured background spectra is added to the simulated spectrum, and then a randomly chosen background spectrum is subtracted. We will assume that the differences across all of the measured background spectra can be used as a proxy for the systematic error expected when a particular background subtraction is applied.

There needs to be an objective measure for the relative performance of a given inversion method. The metric should be a scalar, and should increase with the amount of detail that the inversion method captures about the assumed electron spectrum when it is applied to synthetic data. The normalized mean squared error (NMSE) is a simple metric which satisfies these properties. The NMSE is the squared total difference between the retrieved electron spectrum and the assumed electron spectrum, but normalized to the total signal magnitude to account for the fact that it is applied across different count rates. The definition follows.

Definition 1.3.1 (NMSE). The normalized mean squared error between assumed spectrum \mathbf{y}' and expected spectrum \mathbf{y} is:

$$\text{NMSE}(\mathbf{y}, \mathbf{y}') = \frac{\|\mathbf{y} - \mathbf{y}'\|}{\|\mathbf{y}\|}$$

The NMSE is zero for a perfect agreement between a simulated and retrieved electron spectrum, and increases as the agreement becomes worse. Tests on different inversion techniques will produce different distributions of the NMSE.

1.4 The Least-Squares Regression and the Bias-Variance Trade-off

Due to the ill-posed nature of the X-ray inversion problem, solutions often take the form of estimators. These estimators may be either unbiased, as in the case of least-squares regression, or contain a bias term towards solutions with certain properties. In Figure 1.1, it is apparent that directly solving the X-ray inversion problem using least-squares regression can be ineffective. This is because the problem has a high condition number (Figure 1.2). There is a trade-off apparent: the condition number can be reduced by selecting coarse binning in model space, which reduces the variance in the solution, but also reduces the amount of information about the data the solution represents. The optimization of this trade-off for a given data set is a basic and well-understood problem in data analysis. A solution is said to be over-fit when it aligns well with one data set, but fails to usefully describe the data under small perturbations or additional measurements. On the other hand, a solution is under-fit when it fails to reflect significant changes in measurement data.

Determining the balance between over-fitting and under-fitting a given data set is a basic problem in data analysis. Roughly speaking, the variance in the data set is used as a proxy for how much variance is expected in well-fit models which describe the data. The χ^2 statistic is a standard descriptor for this relative variance.

Definition 1.4.1 (χ^2 Statistic). The χ^2 statistic is the quantity

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{\sigma_i^2}$$

where \mathbf{O}_i are the observed, or measured quantities, \mathbf{E}_i are the expected, or model, quantities, and σ_i^2 are the variances in the measured quantities.

The χ^2 statistic is usually normalized to the number of degrees of freedom which a particular problem has, which gives the reduced χ^2 statistic.

Definition 1.4.2 (Reduced χ^2 Statistic). The reduced χ^2 statistic is the χ^2 per degree of freedom, v :

$$\chi_v^2 = \frac{\chi^2}{v}.$$

If the reduced χ^2 statistic is much greater than 1, then it indicates a poor model fit. When the reduced χ^2 statistic is much less than 1, it indicates over-fitting to the data. An example of this is shown in Figure 1.4 for polynomial models fit to a randomly generated data set.

Managing the balance over-fitting and under-fitting models is related to the condition number of the problem. If a problem is badly conditioned, then it may be necessary to somewhat under-fit the data to generate a solution with a sufficiently small variance. Another problem which frequently occurs for the X-ray inversion problem is the existence of negative particle fluxes in predicted electron spectra, an example of which can be seen in Figure 1.1. There is no clear way to interpret models with these features. Eliminating them from predicted models is possible using constrained optimization techniques, and will be discussed in the next section. Constrained optimization, combined with other modifications to the least-squares estimator to improve conditioning of the X-ray inversion problem, will be the topic of the remainder of this chapter.

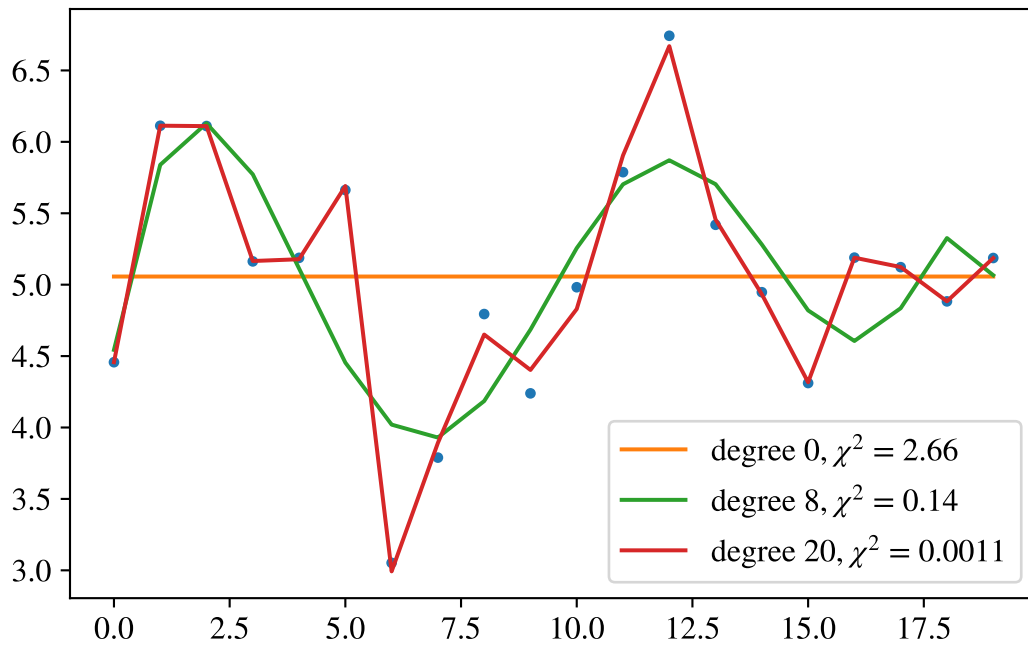


Figure 1.4: Example of over-fit and under-fit models to data. A random data set is generated and polynomial models are fit using least-squares regression. The reduced χ^2 statistic is shown for each model.

1.5 Biased Estimators and Improving the Condition Number

The condition number for a given linear problem can be expressed as the ratio of the maximum and minimum singular values in the kernel matrix. Using the singular value decomposition, it is possible to write an alternative version of the problem with smaller singular values suppressed or removed. The aim is to exchange the original problem for a similar one which is better conditioned. The singular value decomposition writes the kernel matrix \mathbf{G} as:

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where \mathbf{U} and \mathbf{V} are unitary, and $\mathbf{\Sigma}$ is a diagonal matrix with the singular values on the diagonal. Since the decomposition can be chosen such that the singular values are in descending order, an approximation and reduction in dimensions to the original kernel matrix can be generated by truncating the expansion. Processes like this produce biased estimators, or imperfect representations of the kernel matrix, in exchange for more stable solutions. In this section we will examine the effects of these approximations on the X-ray inversion problem using synthetically generated data from the techniques in Chapter 3.

Figure 1.2 shows that the binning scheme for the X-ray inversion problem with the smallest condition number minimizes the bin sizes for the data (X-ray spectra), and maximizes the bin size for the models (electron spectra). This choice must be made within the limits of the statistics of the data set at hand, and the amount of detail desired in the model electron spectrum. As discussed in the previous section, the χ^2 statistic is used to avoid over or under-fitting the problem. Unfortunately, this test is valid only for unbiased estimators, such as the least-squares regression. For a biased estimator, determining the right solution size becomes a much more complicated problem that will be addressed in the next section.

Figure 1.5 shows the distribution of singular values across different electron spectrum bin widths. The X-ray spectrum bin widths are left at 1 KeV. Most of the information contained in the kernel matrix is captured in the first few terms of the singular value decomposition. This sug-

gests that reducing the dimensions of the problem by truncating the singular value decomposition might improve its condition.

To justify the use of a biased estimator in solving the X-ray inversion problem, two questions need to be answered. The first is whether or not the use of the biased estimator significantly improves the conditioning of the problem. The second is whether the bias significantly degrades the amount of information contained in the solutions. The first question is easily answered by graphing the condition number of the kernel matrix as a function of truncation index. This is shown as a function of binning scheme in Figure 1.6.

Compared to Figure 1.2, the plots in Figure 1.6 show that truncating the smallest singular values of the kernel matrix has a positive effect on the conditioning of the X-ray inversion problem. The number of binning schemes which produce a singular kernel matrix, with no inverse, is reduced with increasing truncation index. In addition, the regions with low condition numbers where the inverse problem is tractable (less than 10^3) are expanded.

The improved conditioning of the inverse problem by truncating small singular values comes at a significant cost by introducing solution bias. There is no simple method to describe the amount of bias this process introduces, or the effect it will have on a particular data set, however, we can apply the least squares inversion to sample data, as in 1.1, using the truncated kernel and see what effect the bias has.

Before applying a biased estimator to the inversion problem, there is an important fact which must be made explicit. The main symptom of an ill-conditioned problem is a high sensitivity to small perturbations on the input data. Because of this, one obtains almost no information about the overall behaviour of a given estimator on the X-ray inversion problem by running single tests on synthetic data (which can be generated using the techniques of Chapter 3.) X-ray measurements are counting experiments, and they are subject to statistical noise in their measurements. The statistical noise will follow a Poisson distribution in a well-designed experiment without significant uncorrected detector bias. Since the data have a random component, it would be an error

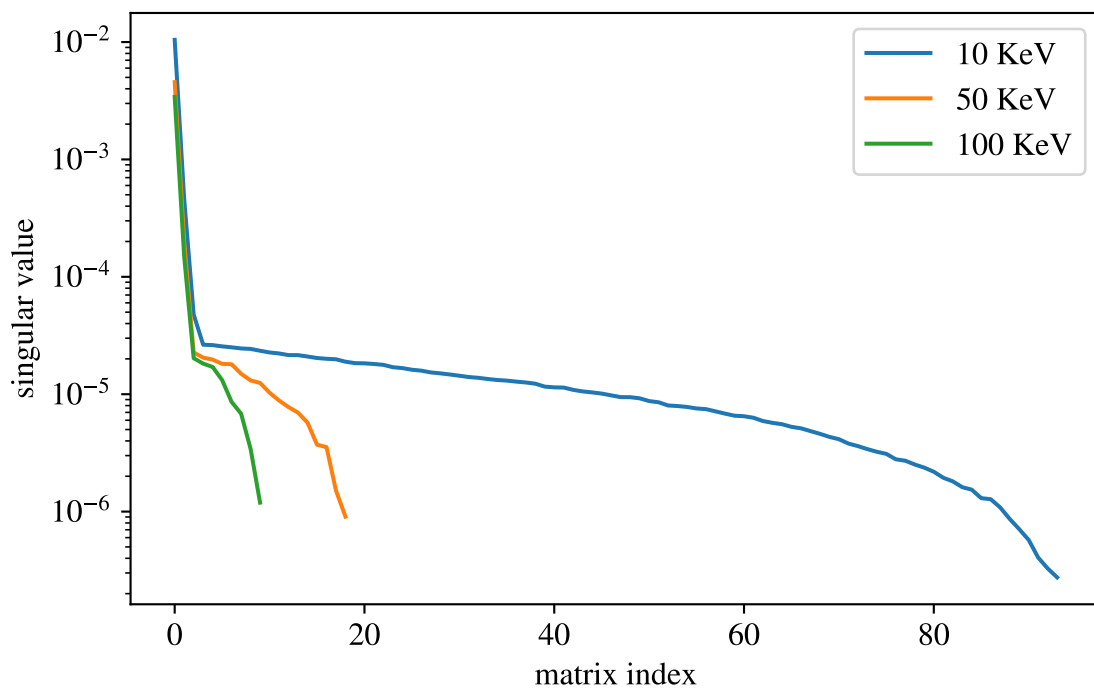


Figure 1.5: Plots of the singular values of the kernel matrix for the X-ray inversion problem for different electron spectrum bin widths, with singular values arranged from biggest to smallest.

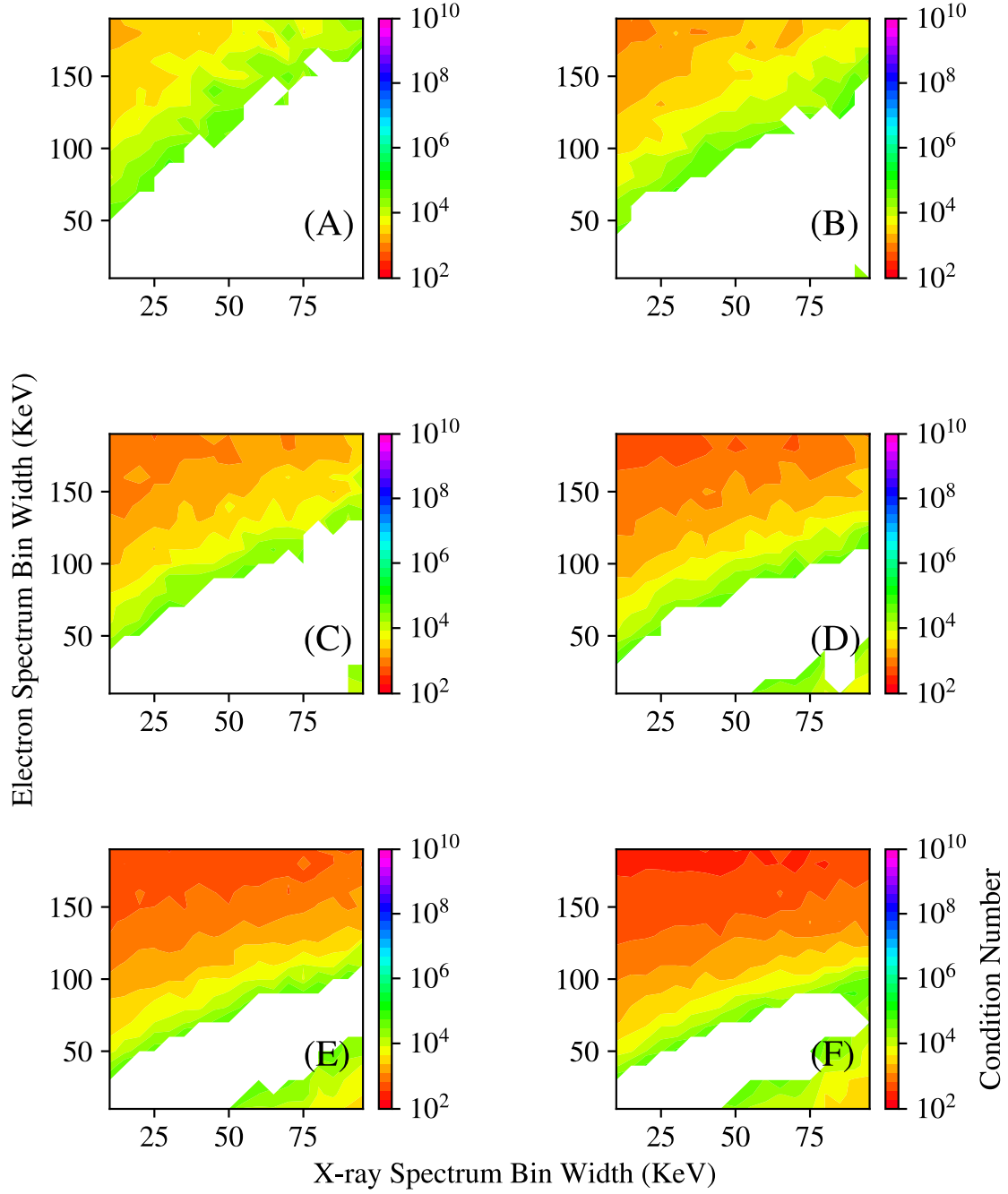


Figure 1.6: Plots of the condition number of the kernel matrix for the X-ray inversion problem as a function of electron spectrum and X-ray spectrum bin width, for 1 (A) through 6 (F) truncated singular values.

in methodology to attempt to solve the X-ray inversion problem on one specific instance of synthetically generated data. Rather, one should sample a given synthetic data set many times, and then use the inversion technique of choice to see the resulting distribution of computed electron spectra.

This technique is applied to synthetic electron spectra in Figure 1.7. Exponential electron distributions and their resulting X-ray spectra are simulated, and then the X-ray inversion problem is solved using the truncated singular value decomposition and least-squares regression for different truncation indices. The stabilizing effect the truncated SVD approximation has on solutions to the X-ray inversion problem is apparent, though, it is important to note that the spread in the reconstructed spectra due to statistical noise in the simulated X-ray measurements remains significant. In a real experiment, one obtains only one measurement of the X-ray spectrum and the underlying statistical distribution in the noise. This presents a significant problem for reconstructions based on biased estimators, since there is no a-priori way to predict how their characteristic bias will map the noise statistics in measurement space to model space. Despite this problem, and the associated difficulty assigning meaningful error estimates to reconstructed spectra, Figure 1.7 motivates the search among biased estimators for solutions to the X-ray inversion problem. It is entirely possible, for example, that a measured X-ray spectrum has statistics which only support a binning scheme which makes the inverse problem impossible. A biased estimator, such as truncated singular value decomposition, which effectively improves solution resistance to noise, can be the only way to recover a useful model in these scenarios.

The use of the truncated singular value decomposition as a biased estimator introduces the truncation index as a free parameter. This leads to an important question: when is the problem well-conditioned enough for the data set being analysed? Higher degrees of truncation improve the conditioning of the X-ray inversion problem (Figure 1.5), but only at the cost of less information being represented in the inverse mapping. The introduction of a free parameter which controls the bias-variance trade-off is a general feature of biased estimators used to solve inverse problems.

This too represents a cost, since the choice of this free parameter must be somehow justified. The question then becomes whether there is a method to choose this parameter based on data from a given experiment. This question is central to the X-ray inversion problem, and is the focus of the remainder of this chapter.

1.6 Tikhonov Regularization and the L-Curve Method

The truncated singular value decomposition explored in the last section is not the only method which has been developed to recondition inverse problems. For the optimization problems that arise from the introduction of a free parameter with biased estimators, it is desirable to have a continuous variable which controls the degree of reconditioning, rather than a discrete truncation index. Further, it is useful to have an estimator which exhibits a bias towards certain families of solutions, using a-priori knowledge we might have about the shape of the expected solution to the inverse problem. The method of Tikhonov regularization has these properties.

In Tikhonov regularization, the unbiased least-squares regression:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|$$

is replaced with the modified version:

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\| + \alpha \|\mathbf{x}\|.$$

The parameter α is a positive real number which controls the strength of the regularization. When α is zero, the least-squares solution is recovered. The addition of $\alpha \|\mathbf{x}\|$ to the least-squares minimization represents a penalty term on the size of the solution norm. There is a close relationship between the effect of adding this penalty term and the suppression of small singular values in the kernel matrix. Write the singular value decomposition of the kernel matrix as:

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

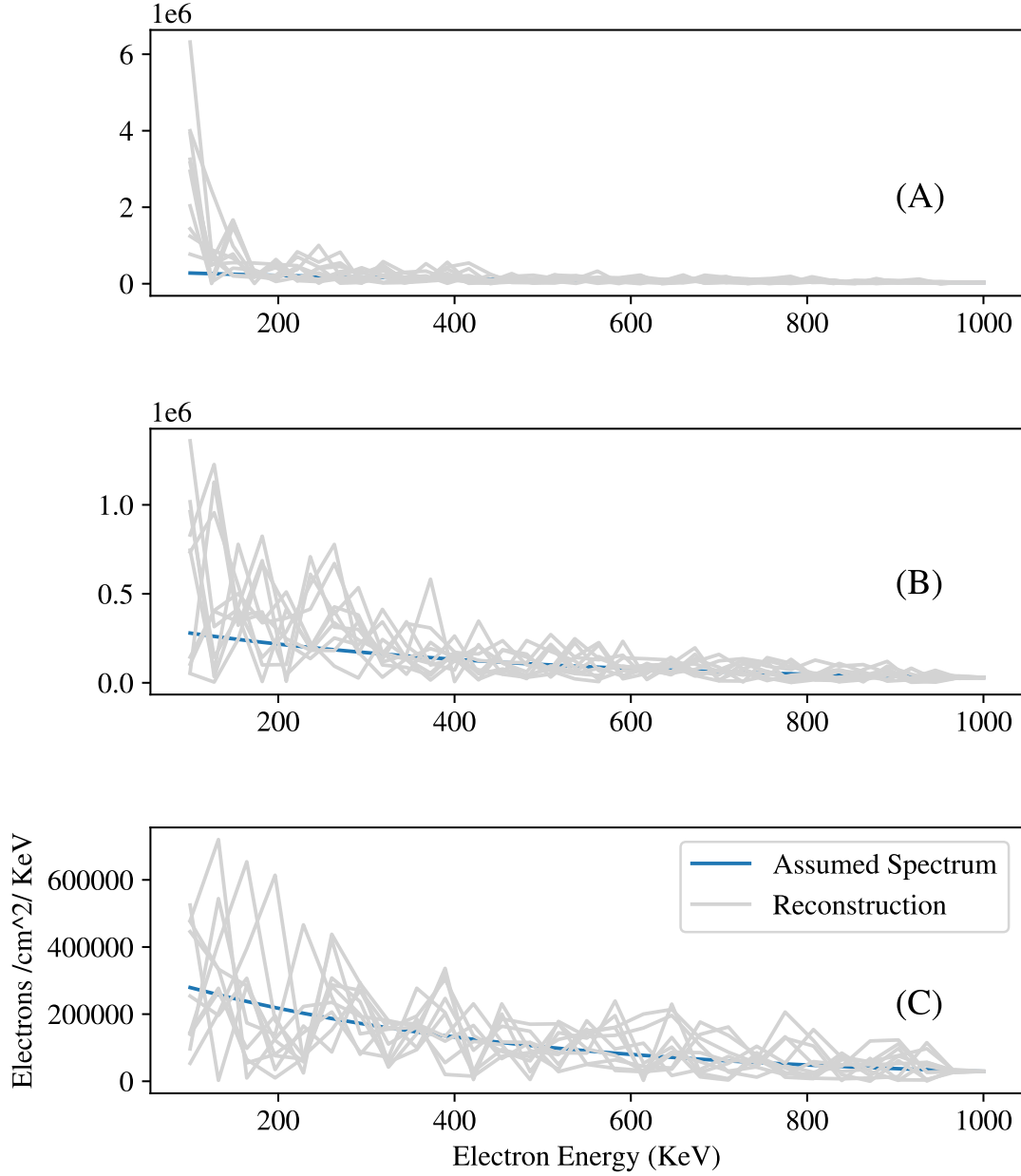


Figure 1.7: Plots of simulated electron spectra for a total electron flux of 1×10^9 electrons/cm²/second, and their reconstruction from the associated simulated X-ray spectra using a truncated singular value approximation of the kernel matrix. Electron bin widths are set to 50 KeV, and X-ray bin widths to 1 KeV. (A), (B), and (C) show reconstructions with the smallest 1, 5, and 10 singular values dropped from the expansion.

It can be shown that the solution to the Tikhonov regularization can be written as:

$$\mathbf{x} = \mathbf{V}\mathbf{D}\mathbf{U}^T \mathbf{b}$$

where the diagonal values of \mathbf{D} are:

$$\mathbf{D}_{ii} = \frac{\sigma_i}{\sigma_i^2 + \alpha^2}$$

where σ_i are the singular values of \mathbf{G} . The regularization term applies a filtering effect which smoothly suppresses the smallest singular values in the kernel matrix. This effect improves the condition number of the kernel matrix, and stabilizes solutions to the inverse problem. This is accomplished at the cost of a bias towards solutions with a small norm, which scales with the degree of regularization applied.

Tikhonov regularization has a heuristic interpretation. The deviation from a solution with a small norm can be viewed as a measure of its complexity, and the amount of information which it contains. The trade-off between solutions to inverse problems with stability but high bias, and solutions with small residual errors but high variance, is controlled by the regularization parameter α . There is no single method to choose an appropriate value for α for a given problem. Techniques to select α can include heuristic arguments, a-priori knowledge based on the properties of the particular problem at hand, and statistical arguments. Combinations of these methods are also possible, but all of them represent an implicit cost to the analysis of the problem at hand, namely, that a choice in α ultimately needs to be made, and that choice will always depend, to an extent, on the preferences of the analyst. This is a danger to the analysis of the X-ray inversion problem, which we will mitigate through careful testing on synthetic data across multiple techniques, and objective scoring of different techniques using clearly defined metrics.

The classic method for choosing α is the so-called “L-curve”. This heuristic method works by plotting the solution norm $\|\mathbf{x}\|$ vs the error term $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ in log-log space. Typically, but not always, this plot has a characteristic L shape, with an inflection point which locates the value for

α that produces the most significant reduction in solution norm for the smallest error term. If the error term is thought of as the cost of solution stability, then the inflection point locates the most economical value of α .

An example of this applied to the X-ray inversion problem is shown in Figure 1.8. We use a synthetic X-ray spectrum based on an exponential beam of electrons with a folding energy of 300 KeV. The L-curve is shown and used to select $\alpha = 2.0 \times 10^{-5}$. The assumed electron spectrum is plotted along with spectra determined using direct matrix inversion, and Tikhonov regularization for the selected value of α . Multiple runs across synthetic data generated with different random seeds are used to provide an accurate representation of the variance in each solution.

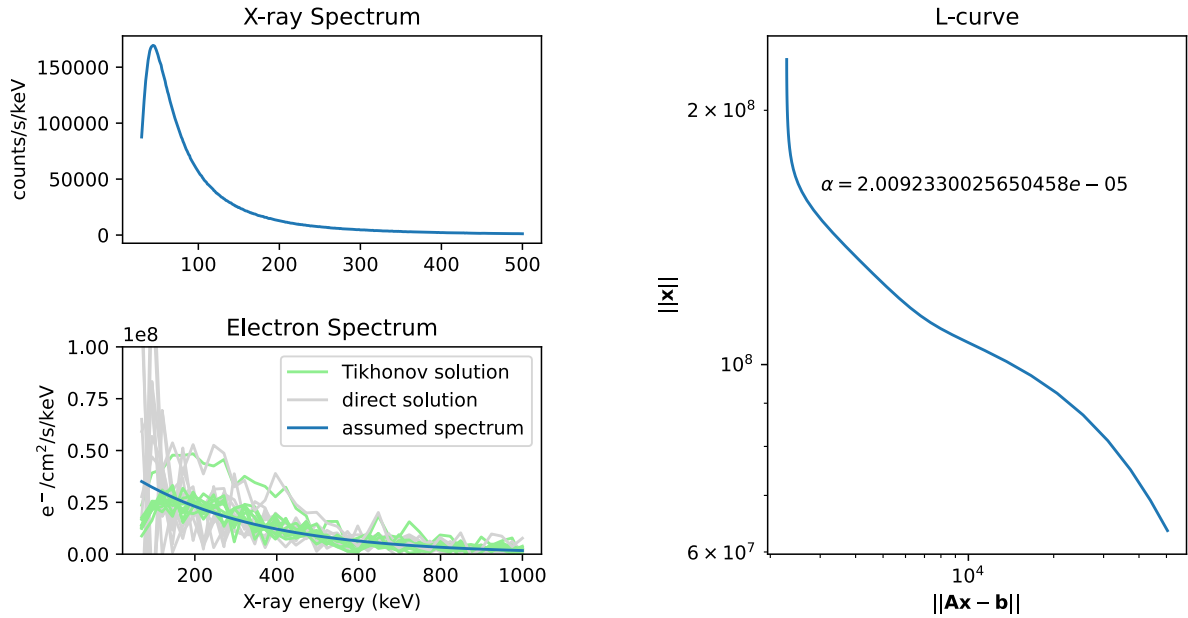


Figure 1.8: Application of Tikhonov regularization and the “L-curve” method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).

The reduction in variance gained from the regularization process is evident in Figure 1.8. The spread in solutions using direct inversion is large enough, that besides a general trend towards a decrease with lower energies, few other features are readily apparent. Additionally, some of the direct solutions show negative particle fluxes in the lower energy bins. Without careful analysis of

the variation in solutions caused by the counting statistics, inversions attempted using experimental data could even be misleading, showing apparently significant features which are in fact only caused by the mapping of random noise through the ill-conditioned kernel matrix of the problem.

The bias introduced by regularization is also apparent in Figure 1.8. Towards lower energies the solution trends downward, which is a feature the assumed electron spectrum does not have. The spectrum retrieved using regularization is also not fully exponential, and a folding energy retrieved by fitting an exponential curve would have a non-zero error when compared with the assumed data. Although the variance in the direct solution is large enough to be untenable when applied to experimental data, it remains exponential, and on average, the folding energy will match the assumed model.

Unfortunately, there is no general method to assign useful confidence intervals to solutions from biased estimators, including those generated from Tikhonov regularization. The effect of the variance in solutions is simple enough to capture, as done in Figure 1.8 by running the inversion multiple times under different random seeds, but a true confidence interval would include the effects of both the bias and the variance in the solution. The use of a biased estimator such as Tikhonov regularization makes an otherwise unsolvable problem tractable, but only at the cost of the certainty and bounds we can place on the solutions. In a way, an ill-posed problem represents a mismatch between the amount of information contained in the solution, and the amount of information in the available data. Tikhonov regularization allows us to control the manifestation of this fundamental lack of information, but there is no process by which it can be eliminated.

For all the caveats and drawbacks which come with applying Tikhonov regularization to a problem, it has a critical property which the unbiased methods lack - namely that it is *adaptive*. The method works by balancing the variance in the solution against the norm, which means that as data become noisy and harder to resolve, the bias is increased to compensate and maintain stability. This is in sharp contrast to, for example, choosing a binning scheme or parameterization which remains fixed for a given problem. Some data sets are “easier” to invert through a given

kernel matrix than others, which can be for a multitude of reasons. For example, the data set could represent a monochromatic solution, which incurs the least variance-inducing resistance in the inversions. Tikhonov regularization is sensitive to this fact, and the L-curve will have a shape which corresponds to the particular problem and particular data set at hand.

The L-curve method is computationally expensive, since solutions need to be generated for every value of α being examined. For relatively small matrices, say, less than 1000 by 1000 rows and columns, this is a small problem, but it becomes significant as the problem being examined scales. Because of this, when using Tikhonov regularization, the most sensible binning scheme for the X-ray inversion problem is not necessarily the one which provides the finest resolution in X-ray spectra, as Figure 1.2 would suggest. Practical considerations based on computational limits need to be taken into account.

The fact that the L-curve method is somewhat heuristic motivates us to look for another way to select a value for α . The “corner” of the L-curve plot, if it appears, is not always sharp. This represents a way that the expectations of the human doing the analysis can find their way into the solutions created using Tikhonov regularization. There is a family of related methods which can be used to select α based on the statistical properties of the data alone. The cross-validation methods use the idea that the artificial removal of parts of the input data produce different solutions and can give information about the stability of the problem being analysed. In particular, under the hypothesis that the effect of the removal of a single data point, such as a single energy bin in the X-ray inversion problem, should have only a small effect on the retrieved electron spectrum. Iterating over the available energy bins, and suppressing each one from the problem in turn, will produce a spread of solutions, whose variation gives a picture of the overall stability of the inversion. If this entire process is repeated over different values of α , then the total mean squared error between the suppressed data points and the regularized solutions is can be compared and minimized. This process is motivated schematically in Figure 1.9. In both the over-fit and under-fit cases, the missing data point is fit poorly. Minimizing this effect represents a balance between these cases

and produces a corresponding value for α .

Iteration over one suppressed data point is called leave one out cross-validation. It is also possible to leave out any number of data points up to $n - 1$, where n is the total number of points in the data set. For leave n out cross validation, one evaluates the solutions generated by every possible way to suppress n data points. This is incredibly computationally expensive.

In Figure ??, the synthetic data set from Figure 1.8 is reproduced, but α is selected using leave-one-out cross validation. Minimizing the mean-squared error using the cross-validation method suggests $\alpha = 7.8 \times 10^{-2}$. This is a different value from what was used in Figure ??, but the retrieved spectrum is approximately the same. This suggests a degree of stability across values of α for the application of Tikhonov regularization to the X-ray inversion problem, but testing across many data sets is required for verification. This will be attempted towards the end of this chapter, where the different regularization methods are compared across a large set of synthetic data sets.

Unlike the “L-curve” method, cross-validation leaves no room for the human analysing the problem to have input on the regularization parameter α . Instead of looking for an inflection point on a log-log graph, which has some room for interpretation, the minimization of squared errors across values of α for cross-validation is carried out numerically. The minimum located in Figure ??, for example, is quite sharply defined. This lack of interpretation is a strength of the cross-validation method. Further, since the minimization process requires no human intervention, it can be applied to large bodies of synthetic data automatically. This is useful both for running on synthetic data for testing, as well as different sets of experimental data. Since the objective characterization of the performance of an inversion technique to the X-ray inversion problem will depend on its application across to synthetic data, this strength is important.

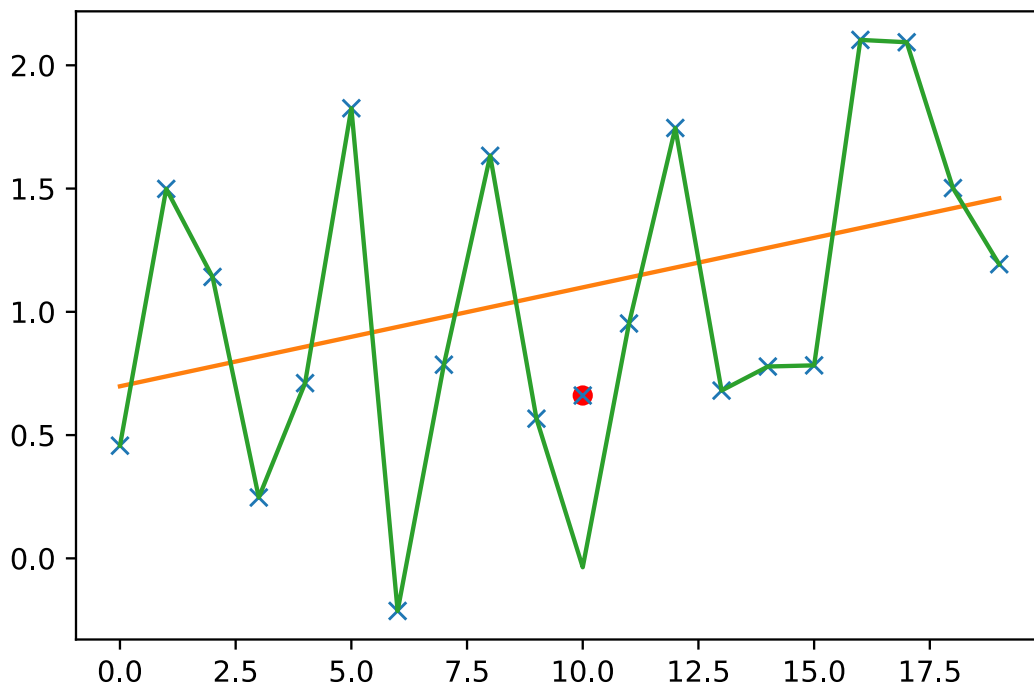


Figure 1.9: Random data set (blue) with under-fit (orange), and over-fit (green) representations. A particular data point (red) is removed prior to fitting. The removed data point is far from both the over-fit and under-fit representations of the data.

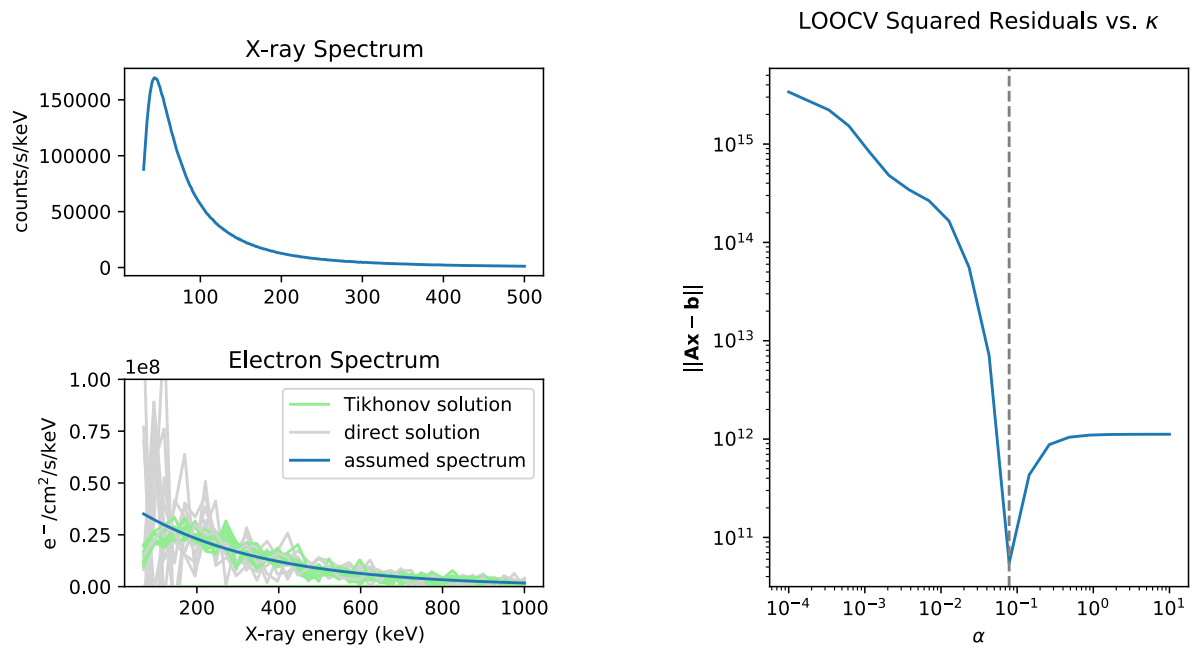


Figure 1.10: Application of Tikhonov regularization and the leave-one-out cross-validation method to synthetic exponential electron spectrum. The assumed electron spectrum is shown in blue (bottom, left), and the retrieved spectra are shown for direct inversion (grey) and Tikhonov regularization (green).

1.7 Higher Order Regularization

The process of Tikhonov regularization can be interpreted as the addition of a-priori information to the problem. In Tikhonov regularization, the modified least-squares problem:

$$\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}|| + \alpha ||\mathbf{x}||$$

has a penalty term $||\alpha||\mathbf{x}||$ introduces a preference for solutions with a smaller norm. This was motivated by the fact that more complex solutions typically have a larger number of non-zero terms than simpler solutions, and, to introduce stability, the strength of the regularization should increase with solution complexity. This was indicated graphically through the L-curve method. There are other forms for the penalty term, which can be chosen to align with the properties of physically reasonable solutions to the problem at hand. This section will examine these higher-order regularization terms as applied to the X-ray inversion problem.

Write a new least-squares problem as:

$$\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}|| + \alpha ||\mathbf{Lx}||$$

where \mathbf{L} is a linear operator. If \mathbf{L} is the identity matrix, then the ordinary regularization is recovered. This is termed zeroth-order tikhonov regularization. In first-order tikhonov regularization, \mathbf{L} is chosen to be the matrix which approximates the first derivative operator:

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & \dots \\ 0 & 1 & -1 & \dots \\ \vdots & \ddots & & \\ 0 & \dots & 1 & -1 \end{bmatrix}$$

Unlike zeroth-order regularization, which prefers small solutions, this choice of \mathbf{L} penalizes solutions which change - the preference is for constant solutions. For problems where we have

an a-priori expectation that solutions are constant, this operator would be the clear choice. A useful and interesting feature of first-order regularization is that no advance knowledge of the best constant solution needs to be applied.

For the X-ray inversion problem, we do not have any strong expectation that solutions will be constant. In fact, it is more reasonable that electron fluxes decrease as a function of energy. This is motivated by both satellite observations, and balloon flights where models of exponentially decreasing electron spectra tend to best describe measured X-ray data. This motivates the search for a different penalty term for the regularization problem, which better matches this expectation.

Second-order Tikhonov regularization uses a linear operator which approximates the second derivative:

$$\mathbf{L} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ \vdots & \ddots & & & \\ 0 & \dots & 1 & -2 & 1 \end{bmatrix}$$

This operator introduces a preference for smooth solutions. There is an intuitive reason why using this operator for the X-ray inversion problem might be useful - as the noise in the measured X-ray spectrum increases, the number of features detectable in the model electron spectrum will go down. A relevant example would be an electron spectrum which is well-described by two beams at different energies. For low noise levels in the corresponding measured X-ray spectrum, the two beams can be resolved through direct inversion, without applying regularization. As the noise in the measured spectrum is increased, second-order regularization will dull and blend the two sharp peaks until eventually they are not distinguishable. This is a desirable property, since the bias tends towards solutions which reflect the lack of information a high noise measurement represents. This is illustrated in Figure 1.11.

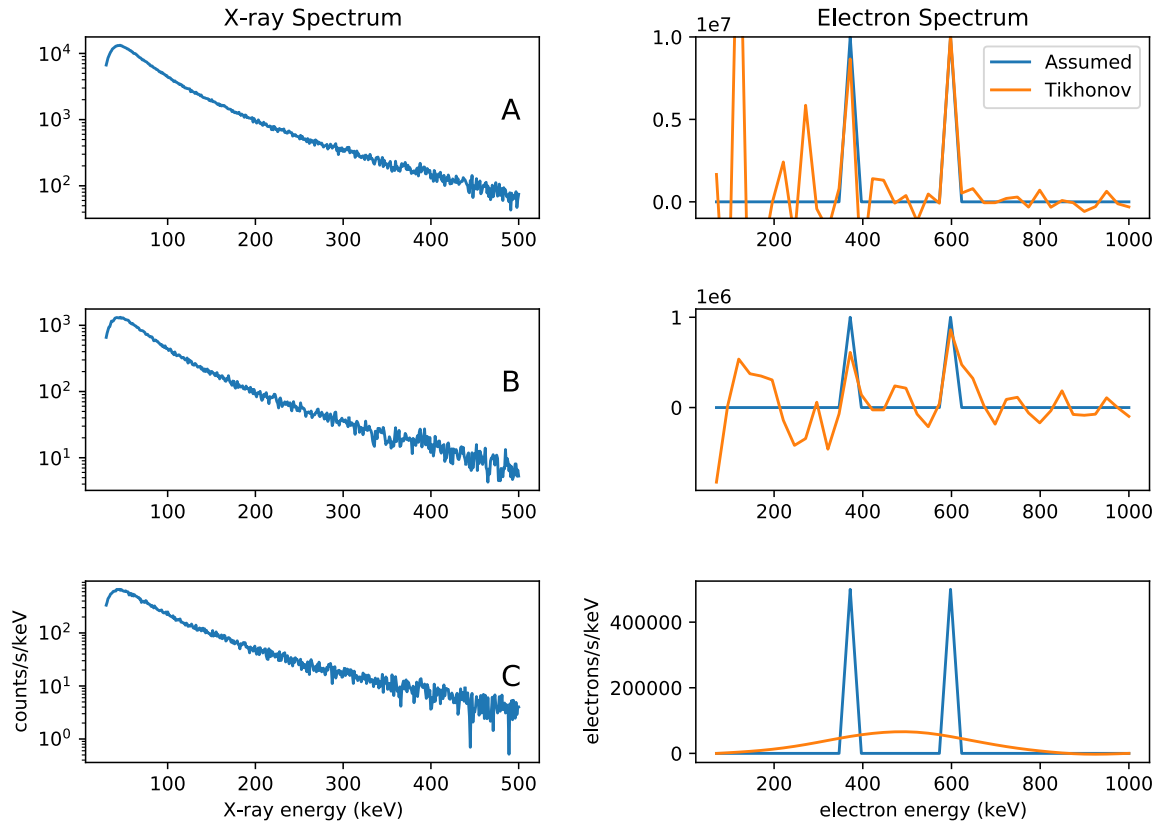


Figure 1.11: Application of second-order Tikhonov regularization to synthetic X-ray measurements corresponding to two mono-energetic electron beams at different X-ray fluxes and noise levels (A - 10^7 electrons / sec / keV, B - 10^6 electrons / sec / keV, C - 5×10^5 electrons / sec / keV)

An additional argument for applying second-order Tikhonov regularization to the X-ray inversion problem is that completely sharp, mono-energetic features aren't realistic in natural electron spectra. Thermal processes tend to relax such distributions to have a characteristic width, and over a long enough time period, the central limit theorem tends them towards Gaussian distributions regardless. Since sharp solutions aren't generally expected, it makes sense to choose \mathbf{L} such that they are penalized. This represents further injection of a-priori information to the problem.

1.8 Preconditioning Techniques

The application of regularization techniques makes the X-ray inversion problem tractable, with realistic and imperfectly measured input data. In using a biased estimator, we have replaced an impossible linear problem with a condition number approaching 10^{10} , with a nearby problem that is better behaved. The cost of this is an artificial tendency towards families of solutions influenced by a-priori information, such as smooth solutions, for second-order Tikhonov regularization. The effect of this bias increases with noise in the input data. In this section, we explore methods which can be used to apply simplifying transformations to the X-ray inversion problem prior to the regularization process. Provided that the transformations applied have an inverse, they can be used to re-cast the X-ray inversion problem as an equivalent linear problem in a different space. If the transformations applied are chosen carefully, the equivalent problem can have a condition number which is significantly reduced. This reduces the amount of regularization that needs to be applied for stability, or equivalently, allows more information to be obtained from noisy measurement data.

As a motivating example, take the linear problem $\mathbf{Ax} = \mathbf{b}$, with \mathbf{A} defined as:

$$\mathbf{A} = \begin{bmatrix} .0001 & 1.0001 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The condition number of \mathbf{A} is approximately 2.0×10^4 . Instead of calculating \mathbf{A}^{-1} , first left

and right multiply by linear transformation:

$$\mathbf{L} = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The new matrix is:

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 100.01 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with a condition number of 1.0×10^4 . The corresponding transformation can be applied to \mathbf{b} to obtain the solution, \mathbf{x} , of the original problem. This preconditioning can be applied before using regularization methods, reducing the amount of regularization required for stability with a particular problem and data set.

In general the left and right-multiplication can be done using different matrices. Label the left preconditioner \mathbf{L} and the right preconditioner \mathbf{R} . The linear problem:

$$\mathbf{Ax} = \mathbf{b}$$

can then be recast as:

$$(\mathbf{LAR}^{-1}\mathbf{R})\mathbf{x} = \mathbf{Lb}.$$

Since the X-ray inversion problem is linear, there is no inherent cost to applying preconditioning methods before solutions are found. Provided matrices \mathbf{L} and \mathbf{R} are not singular, the solution to the original problem can be recovered.

There is no known general method to determine optimal values for \mathbf{L} and \mathbf{R} . There are some specific cases, such as for symmetric matrices, where techniques exist, but the matrix which de-

scribes the X-ray inversion problem does not fall into any of these categories. This problem makes finding an appropriate preconditioner for the X-ray inversion problem non-trivial. We will try different combinations of known preconditioners, and stop when a selection is found that significantly reduces the condition number of the problem.

From a computational perspective, the “ideal” matrix representing the X-ray inversion problem would be the identity, with a condition number of one. Compared to the identity matrix, the real matrix representing the X-ray inversion problem has two main undesirable properties. The first is that the rows, which represent X-ray spectra from mono-energetic beams of electrons, have a high degree of col-linearity. This is the main cause of the high condition number of the matrix. The second problem is that the rows and columns have vastly different norms. This is because low energy beams of electrons do not penetrate far into the atmosphere, and their impact in terms of measurable X-rays is small. On the other hand, high energy beams of electrons need only have a small total particle flux to create a very large measurable X-ray response. Figure 1.12 (top) shows this for a particular binning scheme on the response matrix through graphs of its rows. The spectrum of singular values is also shown. We will show that preconditioning can have a significant effect on the singular values of the resulting matrix.

One classic left-preconditioner is called diagonal scaling. For this preconditioner, a diagonal matrix is constructed such that the entries are equal to the inverse of the norm of the rows of the target matrix. This has the effect of making the row norms of the transformed matrix closer together. Figure 1.12 (bottom) shows the effect of applying this preconditioner. The row norms and the spectra of singular values span fewer orders of magnitude. For this example, diagonal scaling has improved the response condition number from a value of 1.6×10^4 to 1.3×10^4 . This is not a big change, at least not large enough to see a significant difference in the degree of regularization required to stabilize the X-ray inversion problem, but it is a reduction, and it motivates a search for other, perhaps more effective, preconditioners to use.

The effect of the diagonal scaling preconditioner can also be seen across different binning

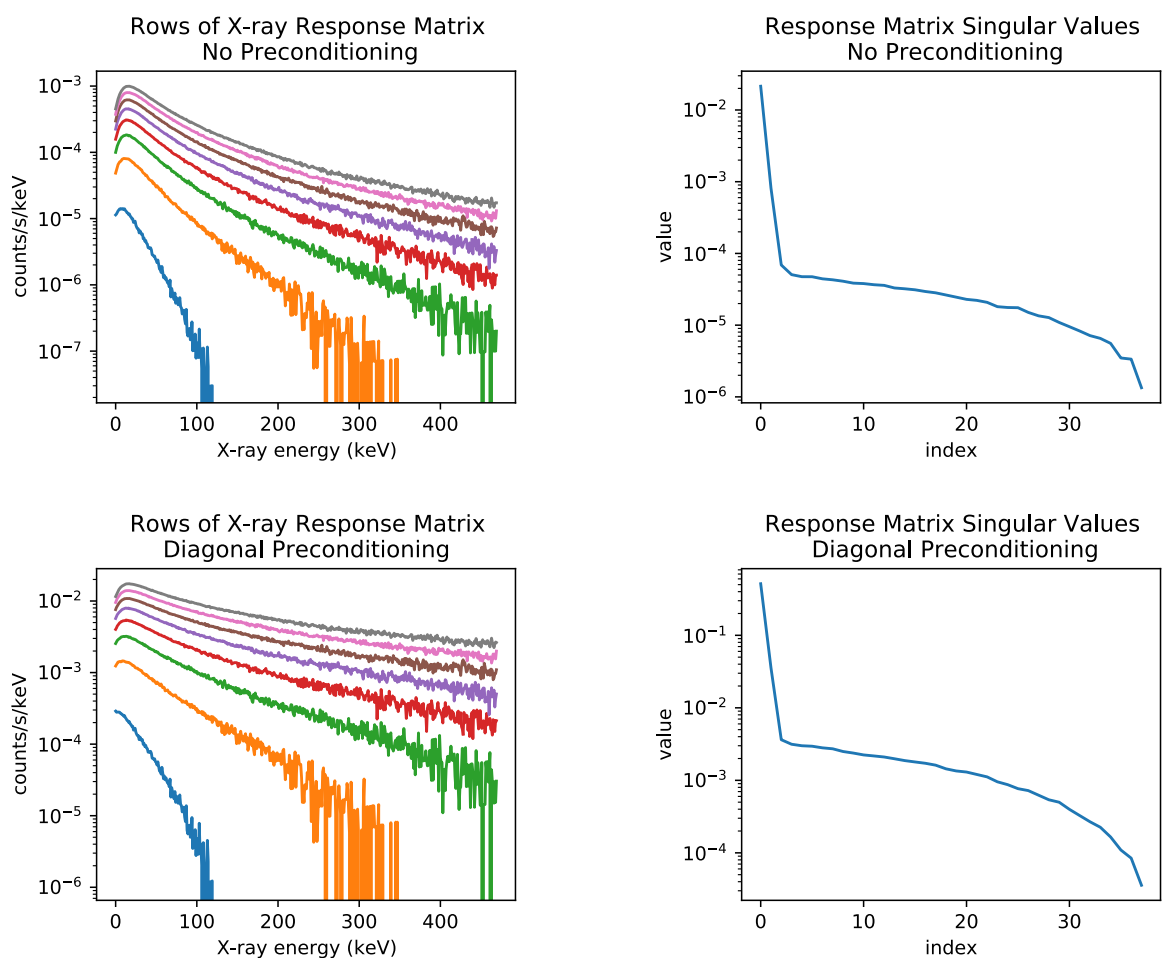


Figure 1.12: (left): Graph of the rows of the X-ray response matrix. A high degree of co-linearity is apparent, as with a large scale change between rows. (right): Graph of the singular values of the X-ray response matrix, from largest to smallest.

schemes on the X-ray response matrix. The effect is not constant across binning schemes. This is shown in Figure 1.13.

Calculating the condition number of a matrix requires computing the singular value decomposition, which is a well known and optimized process and can be done very quickly on modern computers. As an experiment, we can think of the condition number as a function of the diagonal entries of the preconditioning matrix and attempt to minimize it, using, for example, a conjugate gradient search. Despite the fact that this represents a problem with many dimensions, this simple idea can produce a right preconditioner which is effective. An example of the application of this process is shown in Figure 1.14.

It is clear that there are different matrices that can be selected to improve the condition of the X-ray inversion problem. Since one can generally expect to lose k digits of precision for a linear problem with a condition number of 10^k , the search for an optimal preconditioner for the problem at hand isn't very sensitive. Significant benefits will only be returned if order of magnitude improvements in the condition number can be found. This can be achieved through regularization, but comes at the cost of a bias which is difficult to quantify. Because of this, the best approach to the X-ray inversion problem is a combination of left and right preconditioners, chosen optimally for a given selected binning scheme, combined with a regularization method. This dual approach obtains the full benefit of any preconditioners, with the stability, at a cost, which regularization provides. Since the regularization techniques are adaptive to both the noise in the measurement data and the inherent conditioning of the problem, the result is that regularization is used "as little as possible", to obtain stability in the solution.

Choosing left and right preconditioning through optimization techniques generally works to reduce the condition number of the problem, but there are opportunities for it to fail, such as the numerical optimizer failing to converge. Additionally, there is a risk that the optimization adjusts preconditioning based on noise in the response matrix, which is generated through simulations, rather than the inherent shape of the response curves the matrix contains. This can be seen to an

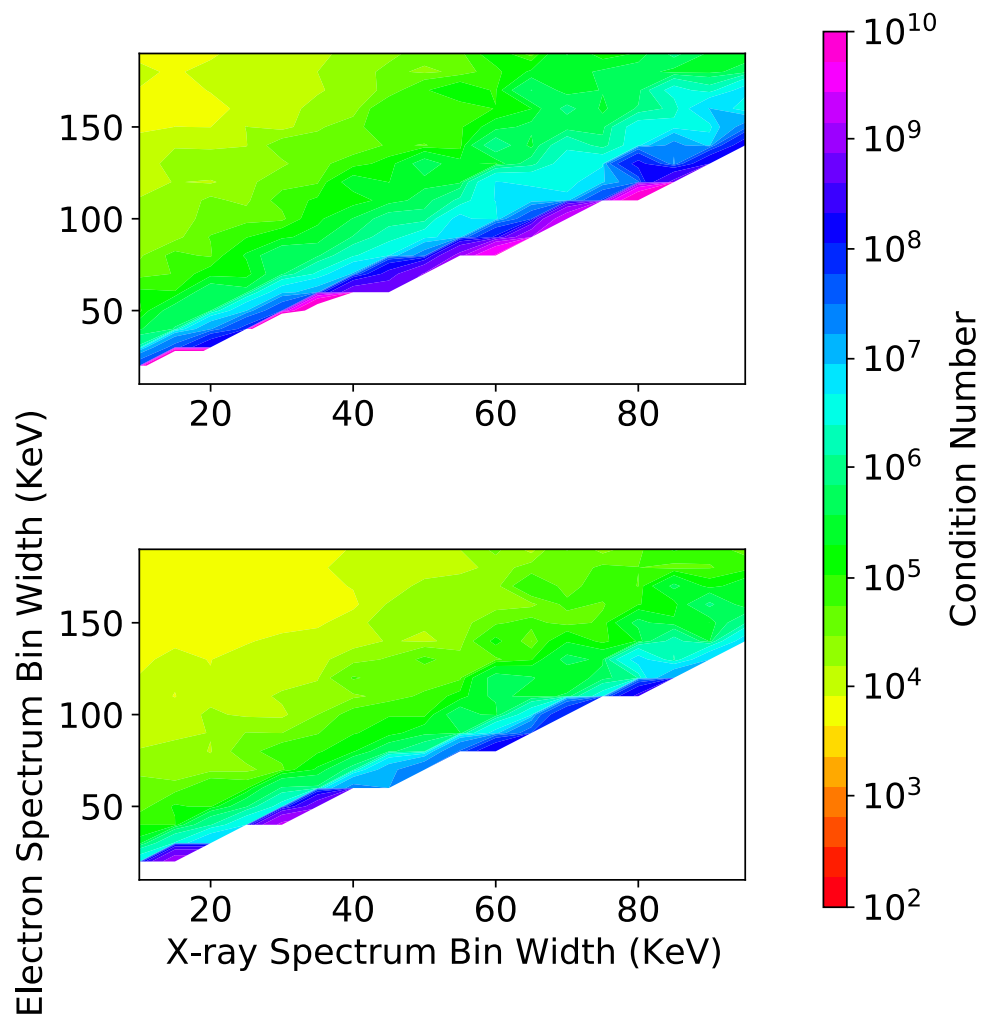


Figure 1.13: (top): Contour plot of the condition number of the X-ray response matrix as a function of electron spectrum and X-ray spectrum bin widths, with no preconditioner. (bottom): The same contour plot with diagonal scaling preconditioner applied.

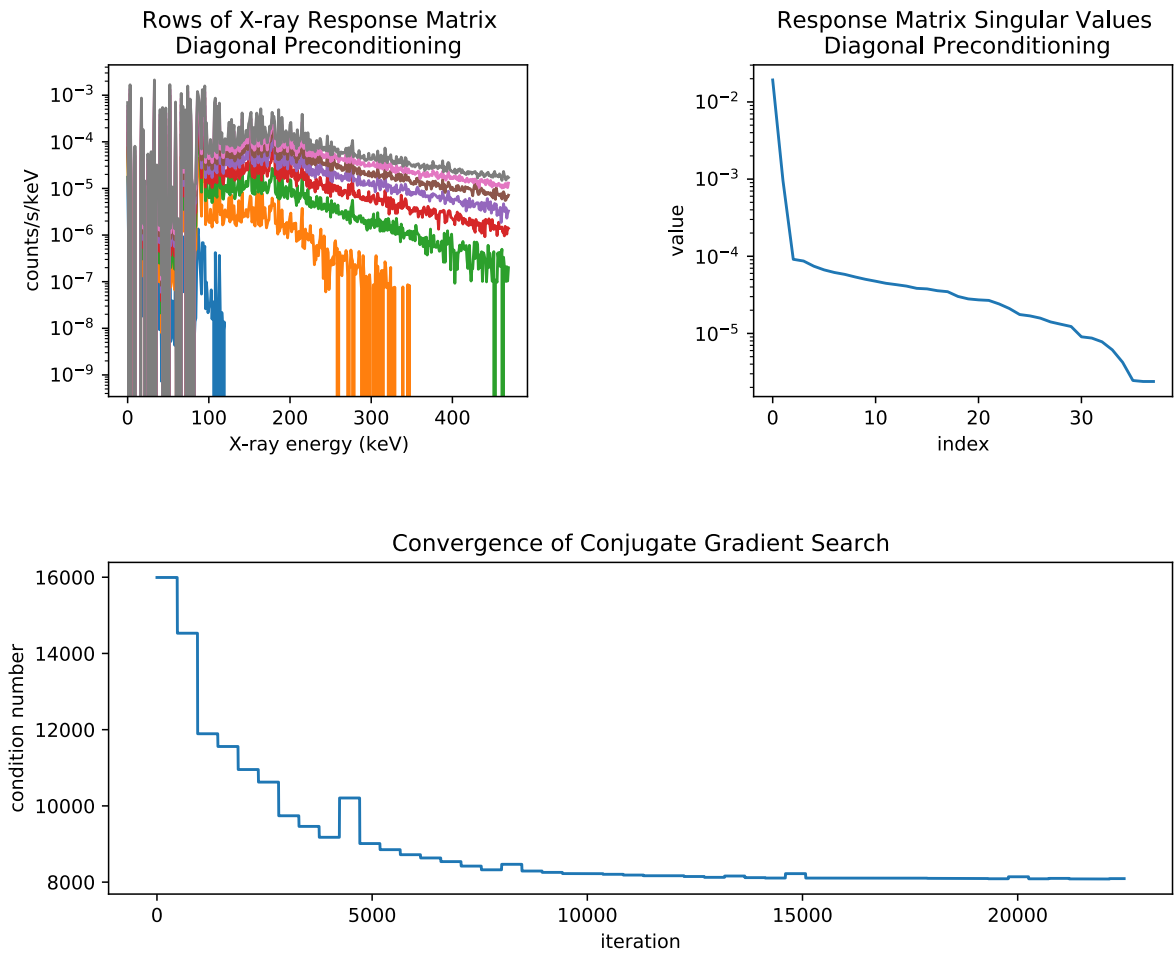


Figure 1.14: (top left): Rows of the response matrix with an optimized diagonal preconditioner applied. (top right): corresponding spectra of singular values for the response matrix. (bottom): condition number as a function of iteration number for a conjugate gradient optimization of the preconditioner matrix entries

extent in the lower energy limit in Figure 1.14. The danger with this effect is that the collinearity of the component vectors of the response matrix is reduced only for synthetic data generated with the same response matrix. Real experimental data will not benefit from this reduced collinearity, since their noise statistics are independent of the response matrix.

An intuitive and useful pair of left and right preconditioners for the X-ray inversion problem can be created based on the observation that the response matrix is poorly scaled in both rows and columns. This is illustrated in Figure 1.15, which shows that the row and column norms differ through orders of magnitude across the response matrix. Motivated by this, we will try a left preconditioner which re-scales the rows of the response matrix, and a right-preconditioner which re-scales the columns. Both will be constructed as diagonal matrices, with dimensions compatibly with the response matrix for a given binning scheme. The diagonal entries of the left and right preconditioning matrices will be the inverse of the row, and column norms of the response matrix, respectively. The effect of left and right preconditioning by these matrices on the response matrix is shown in Figure 1.16.

The effect of this preconditioning is dramatic. In Figure 1.2, the condition number of the response matrix was shown as a function of X-ray and electron spectrum bin-width. This is reproduced in Figure 1.17 for different combinations of these left and right preconditioners. As with the case with no preconditioning, there is a set of binning schemes where the response matrix becomes singular. When preconditioning is applied, however, the sensitivity of the problem to the binning scheme chosen is reduced. This effect is strongest when both the left and right preconditioners are applied. Surprisingly, even though the preconditioners chosen are simple diagonal matrices formed based on intuition, for sensible binning schemes, the X-ray inversion problem becomes completely tractable. This is in contrast to the situation before, where even the coarsest binning schemes produced an inverse problem so ill-conditioned that only the strongest regularization methods could provide stability. As before, this reconditioning comes without introduced bias, and numerically, costs only calculating multiplications and inverses of the preconditioning matrices to transform

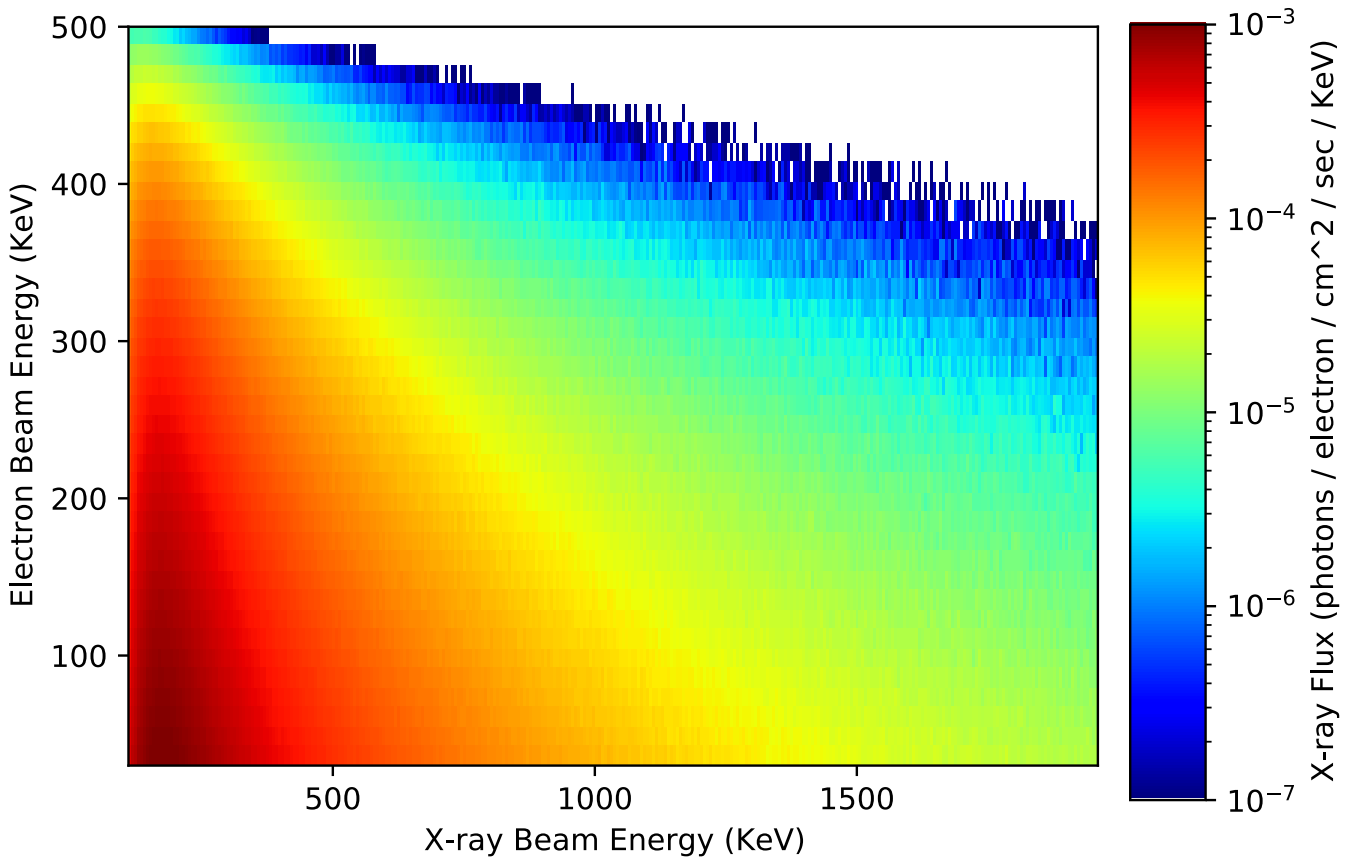


Figure 1.15: Response matrix for the X-ray inversion problem. Scale differences across both rows and columns of the matrix contribute to the high condition number. Units are normalized per incident electron.

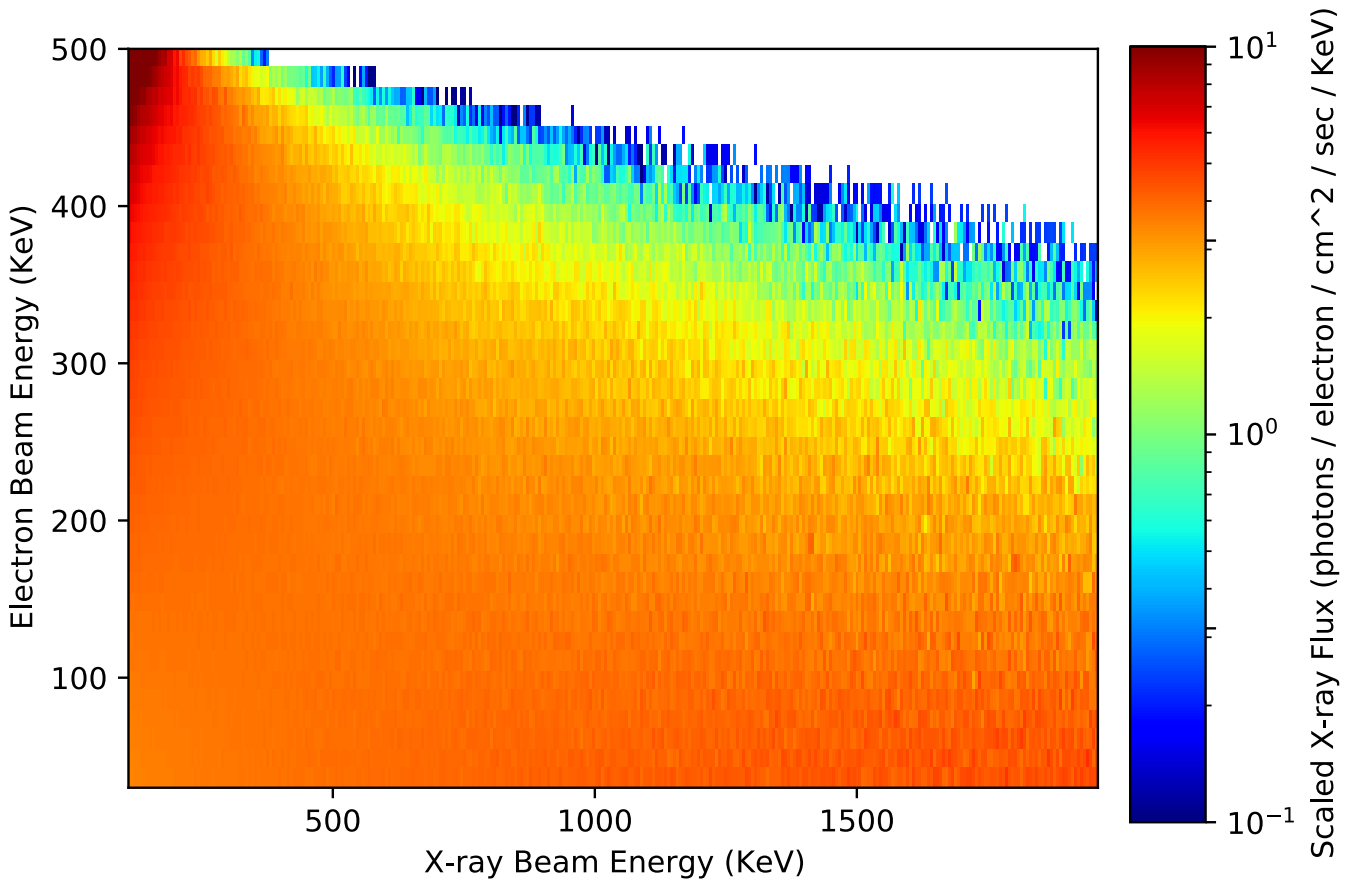


Figure 1.16: Response matrix for the X-ray inversion problem with left and right preconditioners applied. Scale differences across both rows and columns of the matrix have been reduced, improving the condition number. The units on the color axis are no longer physical, having been scaled by the preconditioning matrices.

back to physical units.

1.9 Constrained Optimization

The fact that physical spectra are everywhere positive can easily be applied as a constraint to reduce the solution space of the X-ray inversion problem. Direct inversion, truncated singular value decomposition, and Tikhonov regularization can all be framed as a least-squares optimization problem. Numerical constraints on these problems, especially positivity constraints, are well-known and solves are included in most modern computer algebra systems. There are two primary reasons for imposing positivity constraints on solutions:

1. Since spectra with negative components are non-physical, the interpretation of the negative regions is unclear. Simple approaches, such as setting them to zero, or taking the absolute value, are almost always misleading.
2. Reducing the solution space improves the conditioning of the overall problem.

The first reason exposes a hidden danger when applying preconditioning transformations. The response matrix has a simple physical interpretation, since it maps from the space of electron spectra to the space of X-ray spectra. This interpretation is lost after preconditioning transformations are applied. Although all the same information is contained in the better-conditioned, transformed, matrix, it is stretched and distorted to a degree which is difficult to visualize. This means that a “small” negative region in the preconditioned solution might be significant when mapped back to physical units.

Fast, constrained optimization is achievable through many techniques. For the particular case of non-negative least-squares optimization, the venerable LAPACK FORTRAN routines have been wrapped and made available in most modern programming languages. These routines are extremely fast, and since the X-ray inversion problem has a small (less than a million) entries in the matrices which encode it, the addition of a non-negativity constraint adds almost no computation

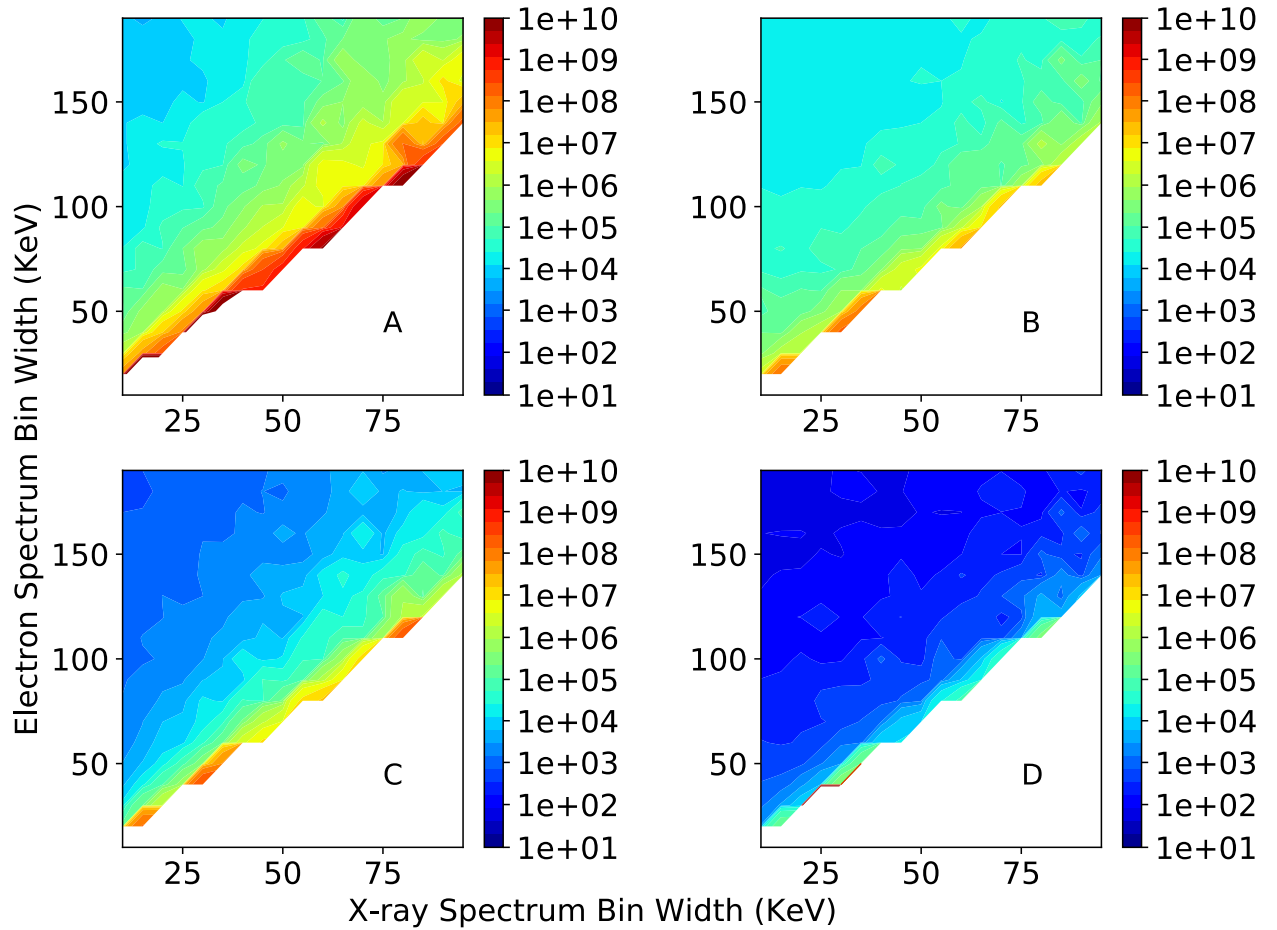


Figure 1.17: Response matrix condition numbers as a function of X-ray and electron spectrum bin widths for: (A) - no preconditioning, (B) - left scaling preconditioner only, (C) - right preconditioner only, (D) - both left and right preconditioners.

time. Figure 1.18 shows an example of synthetic X-ray spectrum corresponding to an exponential electron distribution. The X-ray spectrum is generated at noise levels consistent with different electron fluxes, and an inversion is attempted in each case with the non-negativity constraint applied. The left and right-preconditioners from the last section are used with second order tikhonov regularization to compute the assumed electron spectrum from the data.

1.10 Evaluation on Synthetic Data

The techniques developed so far have taken a difficult inverse problem and transformed it into a more tractable version. Preconditioning through linear transformations reduces the condition number of the problem dramatically, and the subsequent application of second-order regularization provides the rest of the needed stability to the problem. The bias introduced by regularization techniques, which is difficult to quantify, has been kept to a minimum through preconditioning and cross-validation, and chosen so as to gradually relax to smooth solutions in scenarios with high measurement noise. The addition of a non-negativity constraint ensures that solutions are always physical, by construction. What remains is to generate a body of test data and demonstrate that the combination of these techniques generates models which describe the assumed spectra in the test data well. The approach taken will be two-fold. First, simulated data for a set of different plausible forms for precipitating electron spectra will be generated, and then iterated over their respective parameter spaces. Results will be shown in plots to motivate an understanding of the behaviour of the inversion method, and different properties, such as non-negativity constraints, will be turned off and on to see changes in relative performance. Second, a large set of simulated data will be randomly generated, and the inversion method will be run and scored in aggregate based on the normalized mean squared error. The average mean-squared error in inversions across this data set will be used as a metric to show the positive effects of different properties of the inversion method.

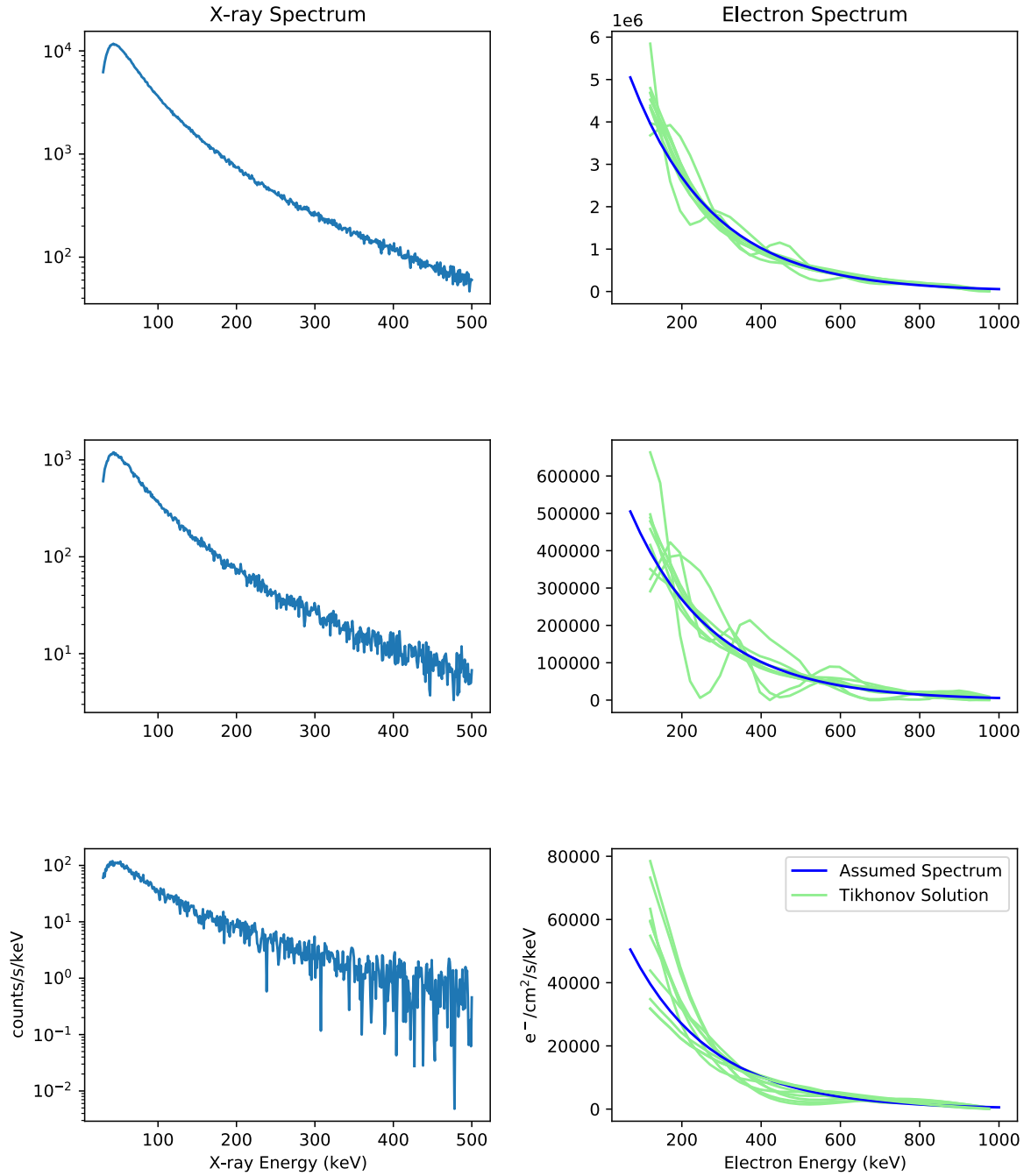


Figure 1.18: Retrieval of electron spectra from synthetic X-ray data across different simulated count rates. Left and right preconditioners, second-order regularization, and a non-negativity constraint are applied.

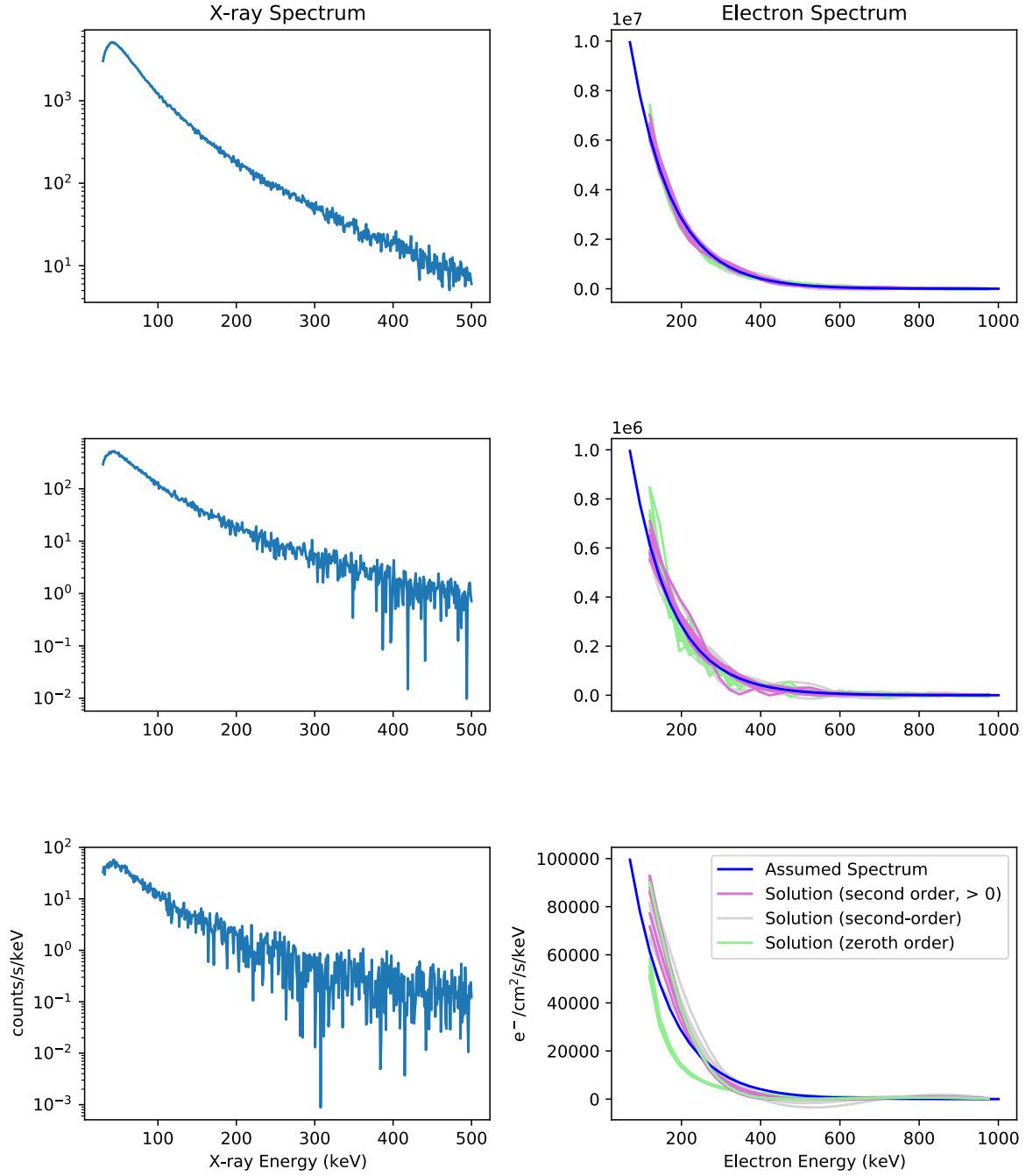


Figure 1.19: Retrieval of exponential electron spectra (100 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

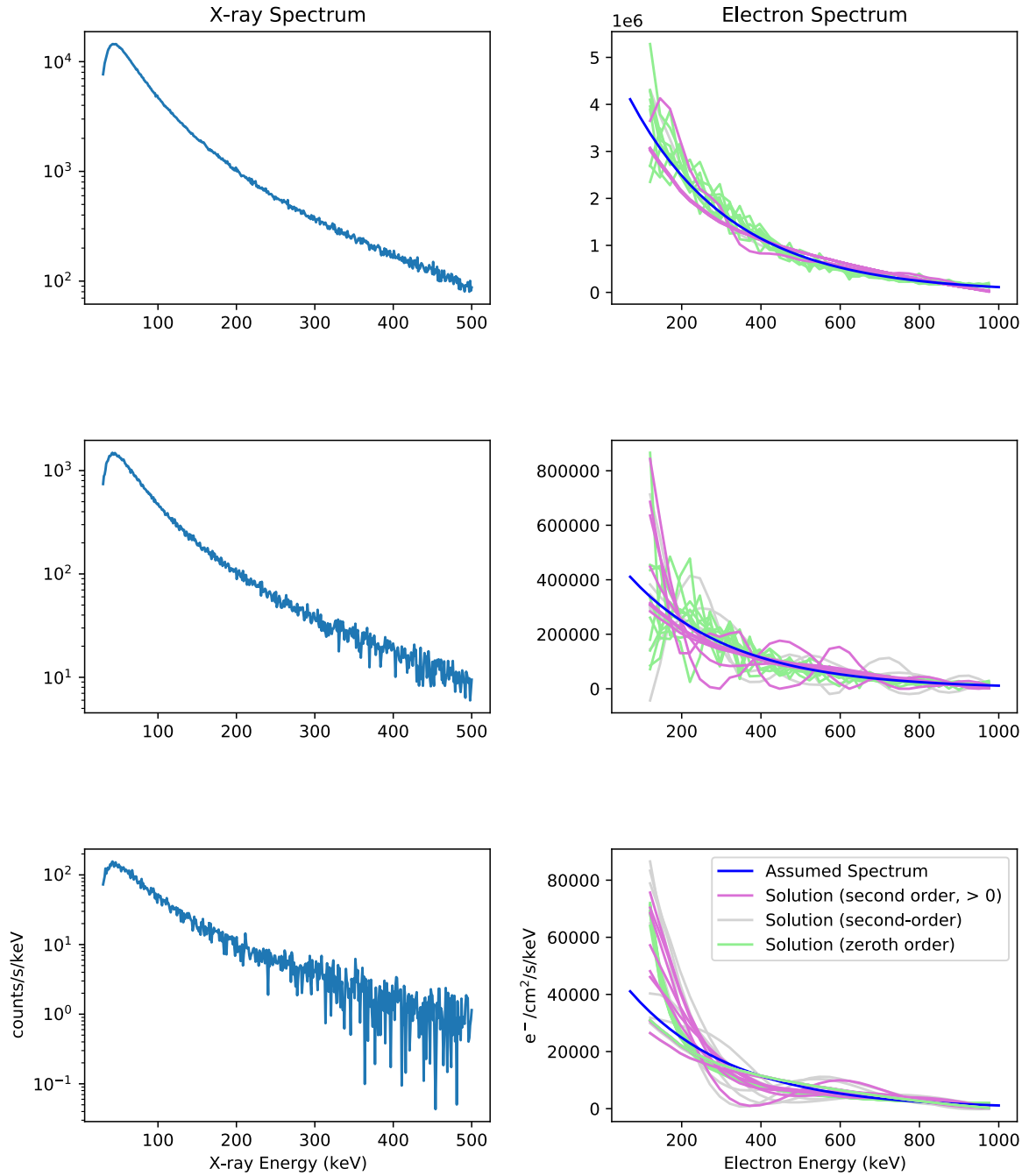


Figure 1.20: Retrieval of exponential electron spectra (250 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

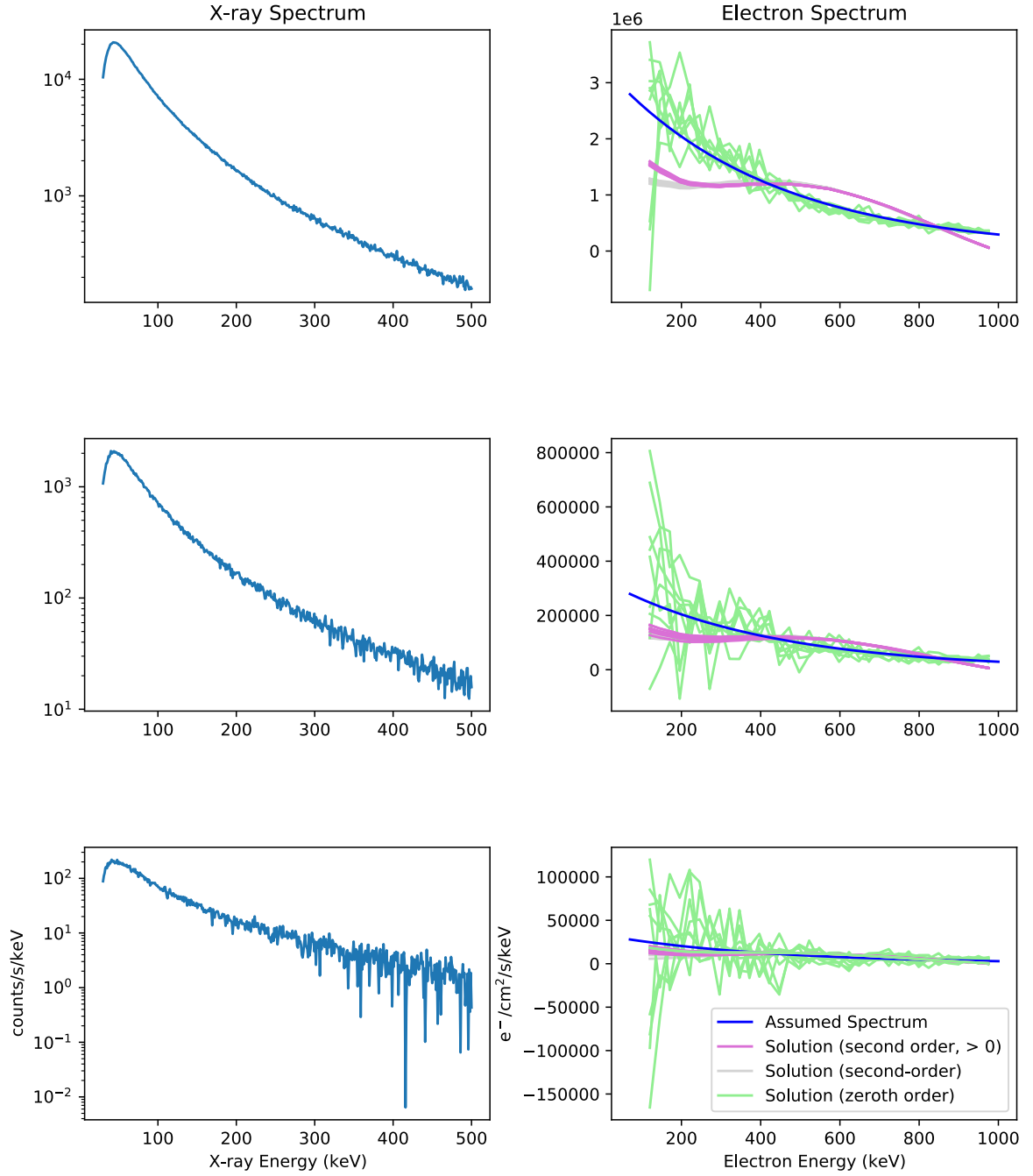


Figure 1.21: Retrieval of exponential electron spectra (400 KeV folding factor) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

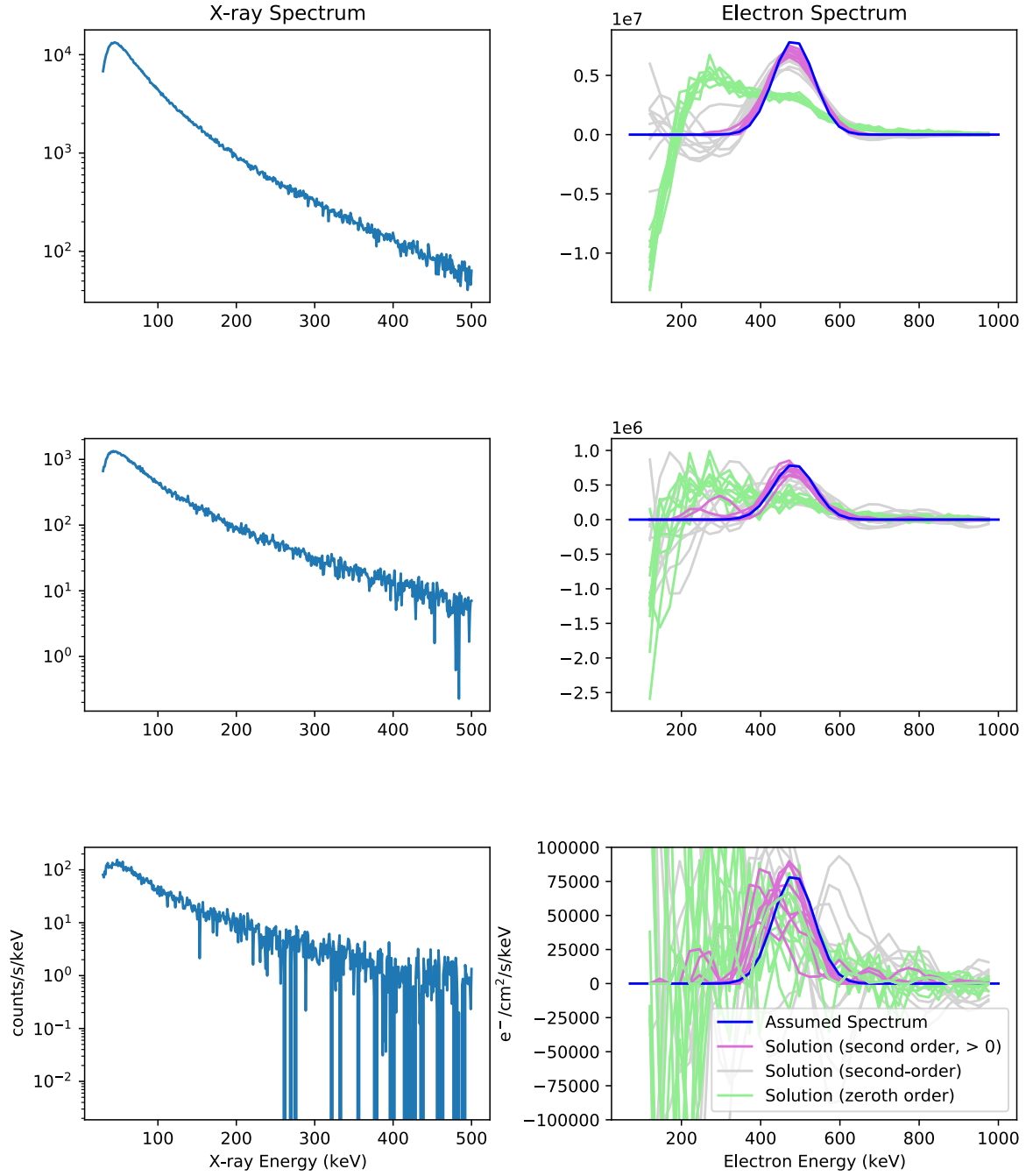


Figure 1.22: Retrieval of gaussian electron spectra (50 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

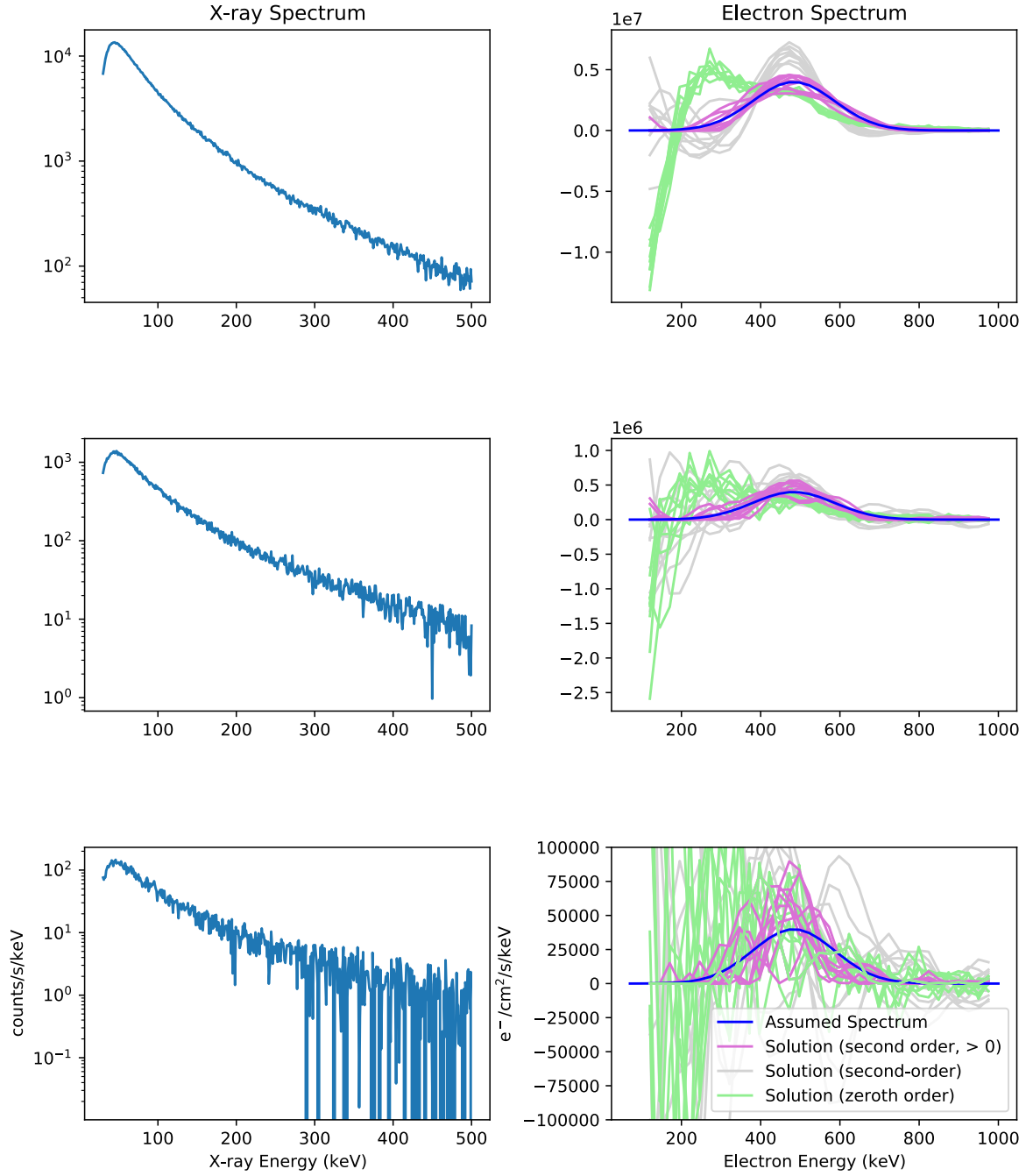


Figure 1.23: Retrieval of gaussian electron spectra (100 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

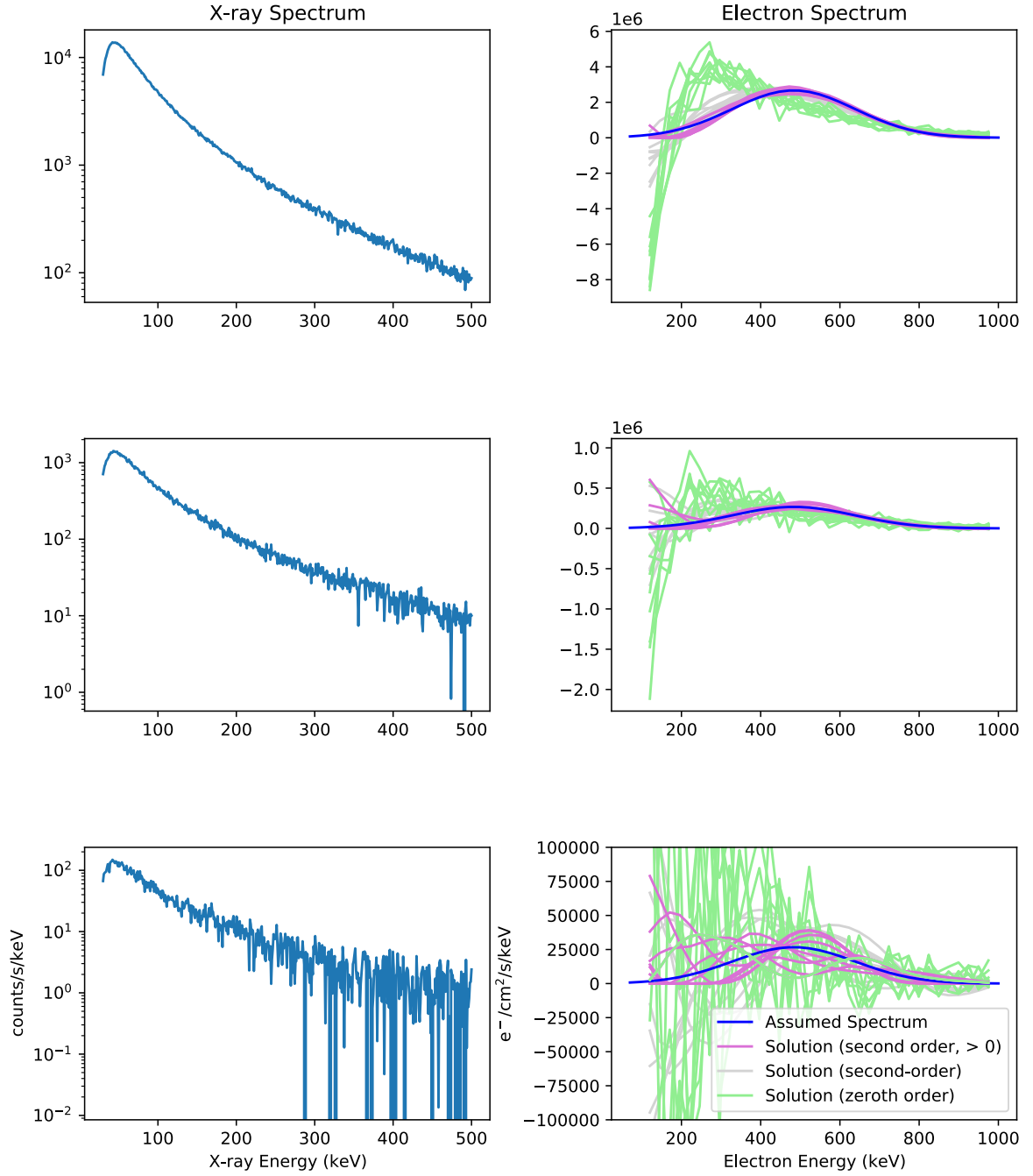


Figure 1.24: Retrieval of gaussian electron spectra (150 KeV variance) from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

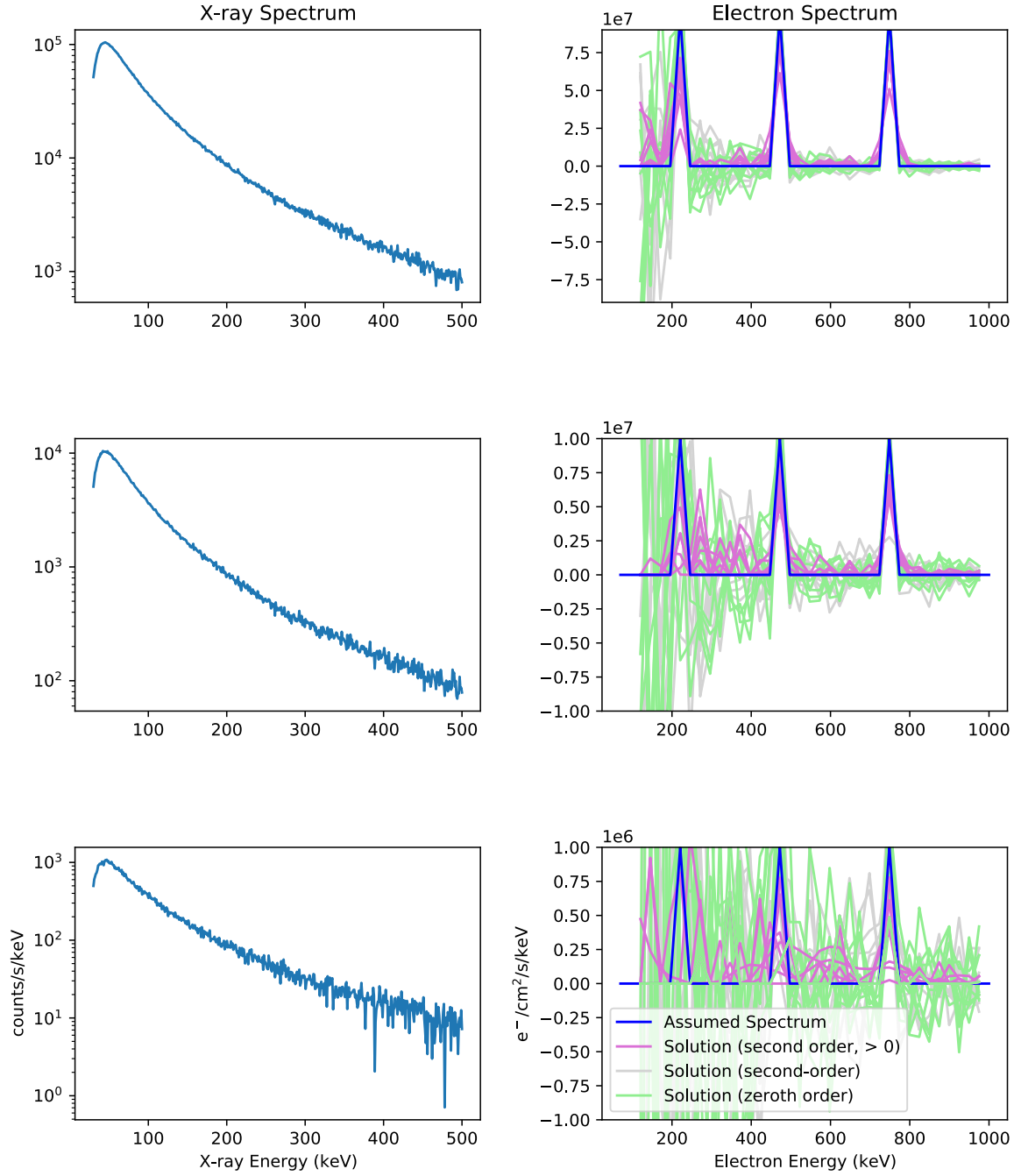


Figure 1.25: Retrieval of monoenergetic electron spectra from synthetic X-ray data across different simulated count rates. Second order Tikhonov regularization, with and without non-negativity constraints, and zeroth order Tikhonov regularization are applied. Regularization strengths are chosen through cross validation.

The data evaluated so far are based on a hypothesis about the types of electron spectra which might exist in nature. In reality, outside the basic constraints which have been discussed so far (positivity, smoothness), and in the absence of additional information, the spectra can have any form. In order to see how the inversion method designed in this chapter performs in general, a large body of synthetic data needs to be created, which extends the test models evaluated in the preceding figures.

There is a decision which has to be made about the amount of structure contained in the test spectra. A test electron spectra which consists entirely of random noise contains little useful information, and the methods used to retrieve it from simulated data will reflect this. On the other hand, a test spectrum with little structure and very smooth features will represent an “easy” test case for the inversion methods. Since the aim is to evaluate inversion methods with the smallest possible number of assumptions on the source electron spectra, we will assign a parameter to the body of generated test data which will correspond to the amount of structure contained.

A set of random electron spectra is generated, and a low-pass filter is applied with a randomly selected cutoff frequency. The distributions for each bin in the random spectra are uniform, as is the distribution for the selected cutoff frequency. Measurement noise which is consistent with a randomly selected total count rate is added to the synthetic X-ray data. The total integrated counts chosen are uniformly distributed in powers of 10, from 10^5 to 10^9 , which roughly represent the possible measurement range of the real data.

Bibliography

- [1] Naomi Altman and Martin Krzywinski. *Regression diagnostics*, volume 13. 2016.
- [2] J Hadamard. *Sur les problems aux derivees patielles et leur signification physique*, volume 13. 1902.
- [3] Albert Tarantola. *Inverse problem theory*. 2004.

Appendix A

First Appendix