

Regularization Methods for the Solution of Inverse Problems in Solar X-ray and Imaging Spectroscopy

Marco Prato

Received: 13 November 2008 / Accepted: 13 November 2008 / Published online: 13 February 2009
© CIMNE, Barcelona, Spain 2009

Abstract Astronomical practice often requires addressing remote sensing problems, whereby the radiation emitted by a source far in the sky and measured through ‘ad hoc’ observational techniques, contains very indirect information on the physical process at the basis of the emission. The main difficulties in this investigations rely on the poor quality of the measurements and on the ill-posedness of the mathematical model describing the relation between the measured data and the target functions. In the present paper we consider a set of problems in solar physics in the framework of the NASA Ramaty High Energy Solar Spectroscopic Imager (RHESSI) mission. The data analysis activity is essentially based on the regularization theory for ill-posed inverse problems and a review of the main regularization methods applied in this analysis is given. Furthermore, we describe the main results of these applications, in the case of both synthetic data and real observations recorded by RHESSI.

1 Introduction

A typical problem in applied sciences is the description of a physical system having on our hands just indirect information about it. This happens, for example, when we want to recover some properties of a very far source from the knowledge of its emitted radiation, or even when we use an acoustic or electromagnetic radiation as a probe, in order

to make it interact with a physical system which cannot be explored directly and get information about its state. In all these cases, the detectors measure physical quantities which are related to the emitted or diffracted radiation; the result of the experiment is a function (or, more realistically, a vector) g which depends on one or more variables (space, time, energy, temperature) and the aim is the reconstruction, through the elaboration of g , of a function f which describes an unknown geometrical or physical property of the system.

When we have to deal with this kind of data, a first method of analysis is a direct approach: we describe the unknown property of the examined system as a function and we assume several hypotheses about its explicit form. Then we simulate the action of the physical system on these expressions of the unknown parameters and we identify the ones which provide simulated data which are very close to the real data. Unfortunately, this method can lead to very unreliable results; in fact, due to reasons related to intrinsic mathematical properties of the problem, it may happen that among all the functions which reproduce faithfully the data we find some of them which are very different one from the other. A physical problem which presents such features is said an ill-posed problem in the sense of Hadamard [35]. An effective way to solve an ill-posed problem is looking at the data analysis as an inversion problem, taking into account the pathological nature of most of inverse problems which are numerically unstable. Reliable solutions can be accomplished by applying a so-called regularization method, which looks for an optimal trade-off between stability and data reproducibility.

In addition to the inverse problems in classical frameworks like medical imaging, optics or geophysics, an increasing attention has been recently addressed to the application of inversion methods in astronomy and plasma physics. Indeed, astronomy is based on the observation of

M. Prato (✉)
Dipartimento di Matematica Pura e Applicata, Università di
Modena e Reggio Emilia, Via Campi 213/b, Modena, 41100, Italy
e-mail: marco.prato@unimore.it

M. Prato
CNR-INFM LAMIA, Via Dodecaneso 33, Genova, 16146, Italy

phenomena which are not reproducible and in this sense it represents an example of an observational rather than experimental science. Moreover, these phenomena take place at great distances and concern physical systems with a high number of degrees of freedom so that generally very different theoretical models are formulated in order to describe them. In other terms, the typical inverse problem in astronomy is based on the reconstruction of a source function which, through an unknown physical process, produces a radiation which propagates in the interstellar medium whose properties are unknown, goes through the atmosphere whose properties are doubtful and finally is intercepted by a detector in order to provide the data. In a situation of such a great ambiguity it becomes crucial to take into account the intrinsic mathematical properties of the inversion process, describing rigorously the problem in an analytical way and applying properly the inversion techniques of the regularization theory in order to obtain an approximated solution of the inverse problem which is numerically stable.

We find a very significant example of this kind of problem in plasma physics when we deal with solar flares. Solar flares [55, 71, 77] are the most dramatic and mysterious events in the solar system. These transient phenomena are characterized by a sudden release of huge amounts of energy and their typical manifestations are the acceleration of electrons in the solar plasma, a notable heating of the solar atmosphere and a significant electromagnetic emission particularly in the X-ray range. All the theoretical modelizations of the physics of the solar flares formulated in the last thirty years are essentially based on the well-established equations of solar plasma physics and magnetohydrodynamics [27]. Common ground for these models is the assumption that the X-ray emission during flares is the consequence of a collisional interaction between the accelerated electrons and the ions of the plasma. Such an event, known as bremsstrahlung collision, is fully described by a quantity, named bremsstrahlung cross section, which represents the probability that an X-ray photon of given energy is produced by an electron in the plasma of given (bigger) energy. However, for most flares, none of these theoretical models is able to fully quantitatively predict the amounts of energy release observed by the detectors. As clearly explained by Brown *et al.* [17], the only way to bypass this puzzling situation is to focus on the analysis of the observed X-ray data to determine the mean electron spectrum associated to the electrons accelerated in the solar plasma. This function represents the electron flux that would be required to produce the observed X-ray flux in a homogeneous plasma source of given ion density and volume. The importance of this function relies on the fact that its determination from the observed X-rays depends only on the bremsstrahlung cross section and does not require any modeling assumption on the acceleration or

propagation mechanisms. Two ingredients are necessary in order to infer information on the mean electron flux from the observed X-ray spectra, the first one being the availability of a notable amount of high resolution X-ray measurements. On February 5, 2002 the NASA Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) mission has been launched with the precise intent of providing X-ray data of unprecedented spectral resolution and of combining them with X-ray 2D images of unprecedented spatial resolution [57]; RHESSI is currently operating and its spectra are at disposal of the scientific community. The second issue is concerned with the mathematical aspects of the equation relating the X-ray flux to the unknown mean electron flux. In fact, such equation is a linear Volterra equation of the first kind whose integral kernel is given by the bremsstrahlung cross section. In this paper the bremsstrahlung Volterra equation is studied within a functional analytic setting and an approach based on regularization theory to its solution is discussed. Such an approach has been applied in the case of very general forms of the bremsstrahlung cross section and validated with both synthetic and real spectra measured by RHESSI [59, 60].

Further interpretation of the mean electron spectrum requires model dependent assumptions. For example, it is possible that bremsstrahlung arises not from energetic electrons moving in a cool background plasma but rather from an inhomogeneous hot plasma with electrons locally Maxwellian everywhere but with the combined distribution of plasma density and temperature (the emission measure differential in temperature) governing the photon spectrum. This assumption leads to a Fredholm integral equation with a Laplace kernel and consequently to an inverse problem whose severe ill-posedness is well known [7]. This is due to the very broad filtering action of the negative exponential kernel (compared to that in the basic bremsstrahlung inverse problem which is of Volterra type and not severely filtering). The huge ill-posedness of the Laplace inversion problem requires again the use of regularization methods to avoid the solution being swamped by amplified noise. In the paper the solution of this thermal bremsstrahlung inverse problem is addressed by means of an ‘ad hoc’ formulation of the Tikhonov regularization method in which the choice of the function space where the solution is reconstructed and smoothed (i.e., the form of the *a priori* information) plays a key role. It must be pointed out that for the present application the destructive effects of ill-posedness are even more dramatic, due to the huge dynamic range of the input data and the complexity of typical source functions. In this context a particular implementation of the method is proposed which allows to account for the variability of the data preserving the reliability of the reconstructions. In this case, one finds that the usual choice of the space of the square summable functions (zero-order regularization) does not provide

a reliable approximation of the true solution. Better reconstructions can be achieved by choosing a first-order regularization approach, implemented by the assumption that the first derivative of the solution belongs to the space of the square summable functions (i.e. by the choice of an appropriate inner product in the source space) and by the addition of some boundary conditions [69].

The availability of spectroscopic methods is only one step toward the investigation of solar flares. The other step is developing X-ray imaging methods which provide spatial information on the flaring mechanisms. Combining imaging and spectroscopy is one of the main goals of the RHESSI mission; the only practical method of obtaining \sim arcsecond angular resolution in hard X-rays and gamma-rays within the cost, mass, and launch constraints of a small satellite is to use Fourier transform imaging [70]. One of the most powerful Fourier techniques is rotational modulation synthesis, first proposed by Mertz [62] and implemented by Schnopper *et al.* [74]. This idea is at the basis of RHESSI, which uses nine bi-grid collimators, each consisting of a pair of widely separated grids (each pair with a different pitch) in front of a Germanium detector, to modulate the X-ray emission during solar flares [40]. Thanks to this hardware, RHESSI is able to perform, in particular, hard X-ray imaging at an angular resolution in the range 2–7 arcseconds, a temporal resolution of tens of milliseconds in the energy range from 3 keV to 400 keV and hard X-ray spectroscopy with a spectral resolution from 0.5 keV to 2 keV, in the same energy range. So far, typical imaging spectroscopy techniques have combined these tools according to the following scheme: a) build a set of count images of the source at different count energies; b) extract count flux spectra from specific regions of the count maps; c) reconstruct the corresponding electron flux spectra through regularized inversion. What we want to do now is to go one step further: we propose a method which uses the same regularization technique in order to produce spatial maps of the electron flux spectrum. Once these electron maps are available, electron flux spectra from local regions can be straightly extracted and compared. The method involves three steps [68]: first, for each count energy channel, a set of count visibilities (i.e., of calibrated measurements of spatial Fourier components of the source distribution) is extracted from RHESSI data in correspondence of different spatial frequencies; then, Tikhonov method is applied to obtain a set of regularized electron visibilities for each electron energy channel; finally, Fourier-based imaging techniques are applied to these reconstructed electron visibilities to obtain the two-dimensional electron flux maps at different energies. This technique has three main advantages: first, to provide information on the spatial distribution of the electron flux, which is a quantity of greater interest than the spatial distribution of the count flux; second, to utilize as input data for the inversion count visibilities, which

are the best measurements available in the RHESSI framework; finally, to impose a certain level of smoothness in the electron energy direction, thus avoiding the unphysical artifacts produced by traditional imaging spectroscopy methods which build up each count image independently and without any spectral correlation.

This paper is divided in eight sections. The first two sections are devoted to the theoretical background which will be used in the whole paper. In the first one a rigorous mathematical formulation of a linear inverse problem is given and the existence and uniqueness of its solution are discussed. A particular attention is given to the concepts of generalized inverse operator and generalized solution, both in the case of problems formulated in functional spaces (essentially Hilbert spaces) and in the case of problems with discrete data, which are more significant in applications. In the second section we describe detailedly the Tikhonov regularization method, which represents the fundamental tool that we will use in the numerical applications.

The next three sections are dedicated to the analysis of the linear Volterra equation of the first kind which relates the X-ray flux to the unknown mean electron flux. In Sect. 3 the inverse problem is described providing an analytical study of the bremsstrahlung equation with three different cross sections; in Sect. 4 the regularization method described in the previous section is applied to both simulated and real data in order to test the efficiency of the method itself (in the simulated cases) and to show the reconstructed solutions (in the real cases). In all the reconstructions of the mean electron spectrum a solid-angle-averaged form [37] for the bremsstrahlung cross section has been used. However, for a given photon emission, the cross section is in general a function not only of the photon energy and electron energy, but also of the incoming and outgoing electron directions and of the polarization state of the emitted photon [32, 33]. In Sect. 5 the angle-dependency of the bremsstrahlung cross section is considered and the consequences of this are investigated by comparing the solutions in the case of a real spectrum.

Section 6 is devoted to the thermal problem and namely to the inversion of the Fredholm integral equation which relates the mean electron spectrum to the differential emission measure. In particular, the inverse problem is described introducing and justifying the regularization method that is used in order to find numerically stable solutions. Then, as in the case of the non-thermal problem, the regularization method is applied to both simulated and real data.

Finally, in the last part of the paper, the imaging spectroscopy topic is analyzed. In particular, in Sect. 7 we describe the new method for the imaging spectroscopy of RHESSI and we validate it both on synthetic and on real flares. Some comments and conclusions are offered in Sect. 8.

2 Linear Inverse Problems and Regularization Theory

2.1 Ill-posedness

In 1976 Professor J.B. Keller gave the following definition of what an inverse problem means: one calls *two* problems *inverse to each other* if the formulation of one problem involves the solution of the other one [43]. What one usually does is to call one of these problems (usually the simpler one or the one which was studied earlier) the *direct* problem, while the other one is the *inverse* problem. This choice is less arbitrary if we have to deal with a mathematical problem applied to a concrete problem; in this case, there is a quite natural distinction between the direct and the inverse problem. For example, if we know a particular state of a physical system and the physical laws that rule it and we want to predict its future behavior, we are in front of a direct problem. On the contrary, if we want to calculate the evolution of the system backwards in time or we want to identify some physical parameter from observations of the evolution of the system we are dealing with inverse problems.

In other words, we might say that inverse problems regard the study of the causes for a specific effect.

Such inverse problems most often do not fulfill Hadamard's postulates of well-posedness [35].

Definition 1 A problem is said well-posed in the sense of Hadamard if all the following properties hold:

1. for all admissible data, a solution exists;
2. for all admissible data, the solution is unique;
3. the solution depends continuously on the data.

A problem is said ill-posed if at least one of the three well-posedness conditions is not satisfied.

Let us consider an ill-posed problem. If we are handling exact data, the existence of a solution is an important requirement and its violation can usually be repaired by relaxing the notion of a solution. In the case of perturbed data, the problem has to be "regularized" and hence changed anyway.

If the condition of uniqueness of the solution is not fulfilled, we have to decide in some way which one among all the solutions is of interest, typically through the assumption of additional information about the solution itself. We observe that, when we work with real data, non-uniqueness is usually introduced by the need for discretization.

Finally, a solution of an inverse problem which does not depend continuously on the data causes serious numerical problems. In fact, if we face the problem by using a "traditional" numerical method (as in the case of well-posed problems), then we have to expect that the numerical solution becomes unstable. We can partially repair this problem with the use of "regularization methods", although we

have to keep in mind that no mathematical trick can make an inherently unstable problem stable. All that a regularization method can do is to recover partial information about the solution as stably as possible. Our goal when we apply regularization methods must always be to find the right compromise between accuracy and stability.

Definition 2 A general *linear inverse problem* can be formulated according to a general scheme. First the corresponding direct problem is defined; the solution of the direct problem introduces a linear (continuous) operator A whose domain is in the Hilbert space X of the solutions and range in the Hilbert space Y of the functions representing the measured data: the inverse problem consists in determining $f \in X$ from the knowledge of $g \in Y$ when g and f are related by the equation

$$g = Af. \quad (1)$$

In the following, we assume that X and Y are Hilbert spaces and we denote with $\mathcal{L}(X, Y)$ the space of the linear and continuous operators between X and Y .

In the case of a linear inverse problem (i.e., when $A \in \mathcal{L}(X, Y)$), the well-posedness conditions can be synthetically written as:

1. $D(A^{-1}) = Y$;
2. $N(A) = 0$;
3. A^{-1} is continuous

where $N(A)$ is the kernel of A and $D(A^{-1})$ is the domain of its inverse operator.

Ill-posedness is a typical feature common to most inverse problems. For example, in the case of linear inverse problems with discrete data, i.e. in the case where Y is a finite dimensional Euclidean space, uniqueness is not verified. Existence does not occur in many linear integral equation of the first kind. In fact, when $X = L^2(a, b)$ and $Y = L^2(c, d)$,

$$g(x) = (Af)(x) = \int_a^b K(x, y)f(y)dy, \quad x \in [c, d] \quad (2)$$

is an analytical function of x if the integral kernel $K(x, y)$ is an analytical function of x . This implies that the set of functions defined by (2) is a proper subset of Y and existence does not occur for all g in $L^2(c, d)$. Finally, let us consider the case where $X = L^1(E)$ and $Y = C(F)$, E and F are compact sets and A is the linear integral operator

$$(Af)(x) = \int_E K(x, y)f(y)dy, \quad x \in F, \quad (3)$$

with $K(x, y)$ continuous in x and y . It can be proved that in this case A^{-1} is unbounded.

Well-posedness is not a sufficient condition for the stability of the solution of a linear inverse problem. Indeed, let

us assume that A^{-1} is well-defined and continuous. Then, if in (1) δg is a small variation of the datum and δf is the corresponding variation on the solution, the continuity of A^{-1} implies

$$\|\delta f\|_X \leq \|A^{-1}\| \|\delta g\|_Y. \tag{4}$$

On the other hand the continuity of A implies

$$\|f\|_X \geq \frac{\|g\|_Y}{\|A\|} \tag{5}$$

so that

$$\frac{\|\delta f\|_X}{\|f\|_X} \leq \|A\| \|A^{-1}\| \frac{\|\delta g\|_Y}{\|g\|_Y}. \tag{6}$$

Definition 3 The real positive number

$$C(A) = \|A\| \|A^{-1}\| \tag{7}$$

is said *condition number* and provides an estimate of the instability of the problem.

From the inequality (6) we see that, if the condition number $C(A)$ is much greater than 1, then a small relative variation on the data can produce very dramatic oscillations on the solution. It follows that, in general, the presence of even a small error on the data of an ill-posed inverse problem (with an elevate condition number) may make its solution extremely unstable.

2.2 Pseudosolutions and Generalized Inverse Operator

Definition 4 Let $A \in \mathcal{L}(X, Y)$, $g \in Y$. A function $u \in X$ is called *normal solution* or *pseudosolution* of the linear inverse problem (1) if

$$\|Au - g\|_Y = \inf\{\|Af - g\|_Y : f \in X\}. \tag{8}$$

The following theorem gives a complete description of the pseudosolutions of a linear inverse problem:

Theorem 1 *If $A \in \mathcal{L}(X, Y)$, $u \in X$, $g \in Y$, A^* is the adjoint operator of A and $P : Y \rightarrow \overline{R(A)}$ is the linear projection onto the closure of the range of A , then the following conditions are equivalent:*

- (1) $Au = Pg$;
- (2) $\|Au - g\|_Y \leq \|Af - g\|_Y \quad \forall f \in X$;
- (3) $A^*Au = A^*g$.

Proof (1) \Rightarrow (2) from the decomposition $Y = \overline{R(A)} \oplus R(A)^\perp$ we have that $Au - g = Pg - g \in R(A)^\perp$; moreover, given $f \in X$, we have that $Af - Pg \in R(A)$ and then from (1) it follows that

$$\|Af - g\|^2 = \|Af - Pg\|^2 + \|Au - g\|^2 \geq \|Au - g\|^2. \tag{9}$$

(2) \Rightarrow (3) since $Pg \in \overline{R(A)}$ and $\overline{R(A)}$ is closed, there exists a sequence $\{f_n\}_{n=1}^\infty \subseteq R(A)$ such that $Pg = \lim_{n \rightarrow \infty} Af_n$. Then

$$\begin{aligned} \|Au - g\|^2 &= \|Au - Pg\|^2 + \lim_{n \rightarrow \infty} \|Af_n - g\|^2 \\ &\geq \|Au - Pg\|^2 + \|Au - g\|^2. \end{aligned} \tag{10}$$

It follows that $Au - Pg = 0$ and then $Au - g = Pg - g$. Since $Pg - g \in R(A)^\perp = N(A^*)$, (3) is true.

(3) \Rightarrow (1) if (3) holds, then $Au - g \in N(A^*) = R(A)^\perp$ and (1) follows. \square

From Theorem 1 it follows that a pseudosolution of the linear inverse problem (1) exists if and only if the datum g belongs to $R(A) \oplus R(A)^\perp$ (which is dense in Y). In this case, the set of pseudosolution is convex and closed in the Hilbert space X , and therefore there always exists a unique pseudosolution u^\dagger of the linear inverse problem (1) with minimum norm. This property leads us to the following definition:

Definition 5 Let $A \in \mathcal{L}(X, Y)$, $g \in R(A) \oplus R(A)^\perp$. A function $u^\dagger \in X$ is called *generalized solution* of the linear inverse problem (1) if u^\dagger is the only pseudosolution of (1) such that

$$\|u^\dagger\|_X = \inf\{\|u\|_X : u \text{ is a pseudosolution of (1)}\}. \tag{11}$$

The operator $A^\dagger : R(A) \oplus R(A)^\perp \rightarrow X$ defined by

$$A^\dagger g = u^\dagger \tag{12}$$

is said *generalized inverse operator*.

The generalized solution u^\dagger is the unique pseudosolution in $N(A)^\perp$. In fact, the decomposition

$$u^\dagger = u_1 + u_2 \tag{13}$$

with $u_1 \in N(A)^\perp$ and $u_2 \in N(A)$ implies that u_1 is a pseudosolution too. Then

$$\|u^\dagger\|^2 = \|u_1 + u_2\|^2 = \|u_1\|^2 + \|u_2\|^2 \geq \|u_1\|^2. \tag{14}$$

The definition of generalized solution imply that $u_1 = u^\dagger$ and $u_2 = 0$.

The previous remark allows us to show that the operator A^\dagger is linear. In fact, if g_1 and g_2 belong to $R(A) \oplus R(A)^\perp$, then

$$\begin{aligned} AA^\dagger g_1 + AA^\dagger g_2 &= Pg_1 + Pg_2 \\ &= P(g_1 + g_2) = AA^\dagger(g_1 + g_2). \end{aligned} \tag{15}$$

Therefore $A^\dagger g_1 + A^\dagger g_2 - A^\dagger(g_1 + g_2)$ is in the kernel of A . But $A^\dagger g_1$, $A^\dagger g_2$ and $A^\dagger(g_1 + g_2)$ are the generalized solutions of the linear inverse problems (1) corresponding to the

different data g_1, g_2 and $g_1 + g_2$ and each generalized solution is orthogonal to the kernel of A . Since $N(A)^\perp$ is a linear subspace of X and $N(A) \cap N(A)^\perp = \{0\}$, we have that

$$A^\dagger(g_1 + g_2) = A^\dagger g_1 + A^\dagger g_2. \tag{16}$$

On the other hand

$$AA^\dagger \alpha g = P \alpha g = \alpha P g = \alpha AA^\dagger g = A \alpha A^\dagger g. \tag{17}$$

Therefore $A^\dagger \alpha g - \alpha A^\dagger g$ is in $N(A) \cap N(A)^\perp$, i.e. $A^\dagger \alpha g = \alpha A^\dagger g$.

The relation between the range of the generalized inverse operator and the range of the adjoint operator is described by the following theorem:

Theorem 2 *Let $A \in \mathcal{L}(X, Y)$. Then $R(A^*) \subseteq R(A^\dagger)$. Moreover, if $R(A)$ is closed, then $R(A^*) = R(A^\dagger)$.*

Proof Let assume that $u \in R(A^*)$; then $u \in \overline{R(A^*)} = N(A)^\perp$. If we define $g = Au$, then u is the generalized solution of the linear inverse problem (1) corresponding to the datum g (because u is a pseudosolution and belongs to $N(A)^\perp$). It follows that $u = A^\dagger g$ and therefore $R(A^*) \subseteq R(A^\dagger)$.

Let now assume that $u \in R(A^\dagger)$; we notice that u is a generalized solution, so we have that $u \in N(A)^\perp$. But $R(A)$ closed implies $R(A^*)$ closed and then $N(A)^\perp = \overline{R(A^*)} = R(A^*)$; it follows that $u \in R(A^*)$ and $R(A^\dagger) \subseteq R(A^*)$. \square

The introduction of the generalized solution and the generalized inverse operator allows us to formulate a new inverse problem which is made up of the solution of two successive minimum problems, described by the two equations (8) and (11). This problem is well-posed if, for all function g in Y , the generalized solution exists unique and the generalized inverse operator is continuous. We show now that if the range of the operator A is closed, then the well-posedness of the problem is guaranteed. First of all, if $R(A)$ is closed, then we have that $Y = \overline{R(A)} \oplus R(A)^\perp$; it follows that the projection of g onto $\overline{R(A)}$ belongs to $R(A)$ and consequently the space of the pseudosolutions is not empty. The existence and the uniqueness of the generalized solution are straight consequences of its own definition, while the continuity of the generalized inverse operator is provided by the following lemma and theorem:

Lemma 1 *If $R(A)$ is closed, then there exists a positive constant m such that*

$$\|Af\|_Y \geq m \|f\|_X \tag{18}$$

for all f in $N(A)^\perp$.

Proof We observe that the restriction $A' : N(A)^\perp \rightarrow R(A)$ of A is a bijective operator between two Hilbert spaces; then, for the Open Mapping Theorem, $(A')^{-1}$ is continuous. It follows that, for any function f in $N(A)^\perp$,

$$\|f\|_X = \|A^{-1}Af\|_X \leq \|A^{-1}\| \|Af\|_Y \tag{19}$$

so we can choose $m = 1/\|A^{-1}\|$ and the claim is proved. \square

Theorem 3 *Let $A \in \mathcal{L}(X, Y)$. Then A^\dagger is continuous if and only if $R(A)$ is closed.*

Proof Let A^\dagger be a continuous operator, and let us suppose that $R(A)$ is not closed. Then the domain of A^\dagger is $R(A) \oplus R(A)^\perp$, which is dense in Y ; it follows that, for any function g in $R(A) \oplus R(A)^\perp$, we have

$$AA^\dagger g = P g, \tag{20}$$

where P is the linear projection onto $\overline{R(A)}$. But A^\dagger is linear and so, for the B.L.T. Theorem (see [72]) there exists a continuous linear extension \hat{A}^\dagger of A^\dagger whose domain is the space Y and such that

$$A\hat{A}^\dagger g = P g \tag{21}$$

for all g in Y . This is a contradiction, because we can always find a function g which belongs to $\overline{R(A)} \setminus R(A)$.

Let us suppose now that $R(A)$ is closed. From Lemma 1 we have that, for any function g in Y ,

$$\|g\|_Y \geq \|P g\|_Y = \|AA^\dagger g\|_Y \geq m \|A^\dagger g\|_X \tag{22}$$

and, consequently,

$$\|A^\dagger g\|_X \leq \frac{1}{m} \|g\|_Y. \tag{23} \quad \square$$

Theorem 3 shows that, if the range of the operator A is closed, then the problem of the determination of the generalized solution is well-posed; on the contrary, if $R(A)$ is not closed, the determination of u^\dagger is an ill-posed problem. Among the operators whose range is not closed, we find very significant classes of operators that people frequently use in applications. An important example is given by the compact operators, as we can see in the following theorem:

Theorem 4 *If $A \in \mathcal{L}(X, Y)$ is compact and $R(A)$ is closed, then $R(A)$ is finite-dimensional.*

Proof The operator $A : X \rightarrow R(A)$ is linear, bounded and surjective between two Hilbert spaces. Then, from the Open Mapping Theorem, given $g = Af$, the open unit sphere in X with center f is mapped by A onto an open set of $R(A)$ containing g . Since A is compact, the closure of this open

set (which still belongs to $R(A)$) is compact; this means that $R(A)$ is locally compact and therefore that it has finite dimension [73]. \square

In conclusion we notice that, as we have already observed in the previous section, the closure of the range of the operator A does not guarantee the stability of the generalized solution: the inequality

$$\frac{\|\delta u^\dagger\|_X}{\|u^\dagger\|_X} \leq C(A) \frac{\|\delta g\|_Y}{\|g\|_Y} \tag{24}$$

is true with, this time, $C(A) = \|A\| \|A^\dagger\|$. If $C(A)$ is large and the datum is affected by noise, the generalized solution is numerically unstable.

2.3 Linear Inverse Problems with Discrete Data

The usual linear inverse problem in physics can be formulated in the following way: an unknown property of the system we are studying is described by a function f and is related to another function g through a linear relation A . In actual applications, we do not have the function g for each value of the independent variable x ; what we usually know is a set of numbers $\{g_1, \dots, g_N\}$ which represents the output of the physical system and which are in some way connected to the values of the function $f(x)$ in the set of points $\{x_1, \dots, x_N\}$. In particular, if the response of the system is linear, then each g_n is linearly related to the value of the data function g in the point x_n ; it follows that, if we neglect the constant which depends on the efficiency of the instruments, the relation between the vector of the data and the unknown function is given by

$$g_n = g(x_n) = (Af)_n \tag{25}$$

where A is a linear operator which maps the Hilbert space X into the Euclidean space Y equipped with the inner product

$$(\mathbf{g}, \mathbf{h})_Y = \sum_{m,n=1}^N g_m w_{mn} h_n^* \tag{26}$$

The numbers w_{mn} are the elements of a weight-matrix W , positive definite, whose choice depends on the properties of the physical system.

If the operator A is continuous, the Riesz's Lemma allows us to formulate the inverse problem with discrete data in a particular way: given the set of functions $\{\phi_1, \dots, \phi_N\}$ in the Hilbert space X and the element $\mathbf{g} = \{g_1, \dots, g_N\}$ in the Euclidean space Y , find $f \in X$ such that [8]

$$g_n = (f, \phi_n)_X, \quad n = 1, \dots, N \tag{27}$$

where $(\cdot, \cdot)_X$ is the inner product in X . In this way, the n -th component of the element Af of the Euclidean space Y

corresponds to the value of the bounded linear functional described by

$$(Af)_n = (f, \phi_n)_X. \tag{28}$$

If the functions $\{\phi_1, \dots, \phi_N\}$ are linearly independent, there exists at least one solution of (27) for each $\mathbf{g} \in Y$. In fact, if we suppose that A is not surjective, then $R(A)$ is a closed subspace of Y and, consequently, there exists an element $\mathbf{c} \in (Af)^\perp$ for all $f \in X$, i.e.

$$c_1(f, \phi_1)_X + \dots + c_N(f, \phi_N)_X = 0 \quad \forall f \in X. \tag{29}$$

From this relation it follows that

$$c_1\phi_1 + \dots + c_N\phi_N = 0 \tag{30}$$

and so the functions ϕ_1, \dots, ϕ_N are linearly dependent, which is a contradiction. This solution is obviously not unique because, if $X_N = \text{span}\{\phi_1, \dots, \phi_N\}$, f_0 is a solution of the problem and $f_1 \in (X_N)^\perp$, then $f_0 + f_1$ is still a solution of the problem. But the set of the solutions of the problem is closed and convex, so there exists a unique solution with minimum norm, called again *generalized solution*, denoted by u^\dagger and belonging to X_N . The explicit form of the generalized solution can be easily obtained by writing u^\dagger as a linear combination of the functions ϕ_n , i.e. posing

$$u^\dagger = \sum_{n=1}^N a_n \phi_n. \tag{31}$$

If we insert this expression of u^\dagger in (27) we obtain

$$g_n = \sum_{m=1}^N a_m (\phi_m, \phi_n)_X, \quad n = 1, \dots, N. \tag{32}$$

In a natural way (32) leads to the following definition [8]:

Definition 6 We call the *Gram matrix* and we denote by G the matrix whose entries are given by

$$G_{mn} = (\phi_m, \phi_n)_X, \quad m, n = 1, \dots, N. \tag{33}$$

The Gram matrix G is invertible because the ϕ_n are linearly independent, so the coefficients a_n in (32) become

$$a_n = \sum_{m=1}^N g_m (G^{-1})_{mn}, \quad n = 1, \dots, N. \tag{34}$$

If we insert this expression of the coefficients a_n in (31) we obtain

$$u^\dagger = \sum_{m,n=1}^N g_m (G^{-1})_{mn} \phi_n. \tag{35}$$

Equation (35) shows that the generalized solution depends continuously on the data and so its determination is a well-posed problem.

This property is not true if the functions ϕ_n are linearly dependent. In this case, $X_{N'} = \text{span}\{\phi_1, \dots, \phi_N\}$ and $Y_{N'} = R(A)$ are two subspaces of X and Y respectively whose dimension is $N' < N$; it follows that a solution of problem (27) exists if and only if the datum belongs to $Y_{N'}$. On the contrary, if \mathbf{g} does not belong to the range of A , then it is natural to introduce the least-squares problem

$$\|Af - \mathbf{g}\|_Y = \min \tag{36}$$

whose solutions are called, as in the previous section, *pseudosolutions*.

Theorem 1 holds even in the case of linear inverse problems with discrete data, so the pseudosolutions u satisfy the equations

$$Au = P\mathbf{g}, \tag{37}$$

where P is the projection onto $Y_{N'}$, and

$$A^*Au = A^*\mathbf{g}. \tag{38}$$

We can obtain the explicit form of the adjoint operator A^* if we observe that

$$\begin{aligned} (Af, \mathbf{g})_Y &= \sum_{m,n=1}^N (Af)_m w_{mn} g_n^* \\ &= \sum_{m,n=1}^N (f, \phi_m)_X w_{mn} g_n^* \\ &= \left(f, \sum_{m,n=1}^N \phi_m w_{mn} g_n \right)_X, \end{aligned} \tag{39}$$

so that $A^* : Y \rightarrow X$ is the operator which maps a vector $\mathbf{g} \in Y$ in

$$A^*\mathbf{g} = \sum_{m=1}^N \left(\sum_{n=1}^N w_{mn} g_n \right) \phi_m. \tag{40}$$

The set of the pseudosolutions is still closed and convex, so there exists a unique pseudosolution with minimum norm u^\dagger of problem (36).

If the generalized solution u^\dagger belongs to the subspace X_N (or, more precisely, $X_{N'}$ if the ϕ_n are not linearly independent; in the following we will assume for sake of simplicity that $N = N'$) of X , then we can write it as a linear combination of the elements of any basis of X_N . Among all the possible choices, there is one particular basis of X_N which will result very useful in the following. In order to produce this basis, we have to introduce the operators A^*A and AA^* ;

these two operators are linear, bounded, self-adjoint, positive definite and with finite rank N . A^*A is defined on the Hilbert space X and its explicit form, using (40), is given by

$$A^*Af = \sum_{m=1}^N \left(\sum_{n=1}^N w_{mn} (f, \phi_n)_X \right) \phi_m \tag{41}$$

while AA^* acts on the Euclidean space Y and its k -th component is given by

$$(AA^*\mathbf{g})_k = \sum_{m,n=1}^N (\phi_m, \phi_k)_X w_{mn} g_n, \quad k = 1, \dots, N. \tag{42}$$

If we remind the definition (33) of the Gram matrix, we obtain the relation

$$AA^* = G^T W. \tag{43}$$

The eigenvalues of AA^* , each with its multiplicity, are denoted by σ_n^2 and ordered in order to provide the non increasing sequence

$$\sigma_1^2 \geq \dots \geq \sigma_N^2. \tag{44}$$

The corresponding eigenvectors $\{\mathbf{v}_n\}_{n=1}^N$ of AA^* form an orthonormal basis of Y . The operator A^*A has the same eigenvalues σ_n^2 of AA^* with the same multiplicity and the corresponding eigenfunctions $\{u_n\}_{n=1}^N$ form an orthonormal basis in X_N . It can be shown [8] that we can always choose the eigenfunctions u_n and the eigenvectors \mathbf{v}_n such that the shifted-eigenvalues problem

$$Au_n = \sigma_n \mathbf{v}_n, \quad A^*\mathbf{v}_n = \sigma_n u_n \tag{45}$$

holds.

Definition 7 The set of triples $\{\sigma_n; u_n, \mathbf{v}_n\}_{n=1}^N$ which satisfy the shifted-eigenvalues problem (45) is called the *singular system* of the operator A ; the real numbers σ_n are the *singular values*, the functions u_n are the *singular functions* and the vectors \mathbf{v}_n are the *singular vectors*.

The knowledge of the singular system of the operator represents a crucial point in regularization theory. The singular values and the singular vectors can be calculated in a simple way by diagonalizing the matrix (43) while the singular functions can be obtained from the relations (45) and the explicit form (40) of the adjoint operator:

$$u_k = \frac{1}{\sigma_k} \sum_{m=1}^N \left(\sum_{n=1}^N w_{mn} (\mathbf{v}_k)_n \right) \phi_m. \tag{46}$$

From the singular system of the operator A we can obtain a very meaningful expression for the generalized solution of

the inverse problem (25). This function, as we have already observed, has to be a pseudosolution (and so it solves (38)) and has minimum norm among all the pseudosolutions (and so it belongs to X_N , i.e. it has no component in $(X_N)^\perp$). It follows that, if we replace the expansions

$$u^\dagger = \sum_{n=1}^N a_n u_n \tag{47}$$

and

$$g = \sum_{n=1}^N (g, v_n)_Y v_n \tag{48}$$

in (38) and we compare the two sums term by term, we can deduce the following expression for the coefficients a_n :

$$a_n = \frac{1}{\sigma_n} (g, v_n)_Y, \quad n = 1, \dots, N. \tag{49}$$

Consequently, the generalized solution becomes

$$u^\dagger = \sum_{n=1}^N \frac{(g, v_n)_Y}{\sigma_n} u_n. \tag{50}$$

From this formula we can notice that, if we want to show explicitly the generalized solution of an inverse problem with discrete data, it is necessary to start with the choice of the topology on the data space Y . Then, from (50) we can deduce immediately the continuous dependence of the generalized solution from the datum (in accordance with the fact that the range of a linear bounded functional is closed).

The knowledge of the singular system of the operator A gives us also a quantitative analysis of the instability of the generalized solution; in fact, it can be shown that the condition number is given by

$$C(A) = \frac{\sigma_1}{\sigma_N}, \tag{51}$$

where σ_1 is the greatest singular value while σ_N is the smallest one. Indeed we have that (see [72])

$$\|A\|^2 = \|A^* A\| = \sup_{k=1, \dots, N} \sigma_k^2 = \sigma_1^2 \tag{52}$$

while

$$\|A^\dagger\|^2 = \|A^\dagger (A^\dagger)^*\| = \sup_{k=1, \dots, N} \frac{1}{\sigma_k^2} = \frac{1}{\sigma_N^2}. \tag{53}$$

In order to show that the numbers $\frac{1}{\sigma_k}$, $k = 1, \dots, N$ are the eigenvalues of the operator $A^\dagger (A^\dagger)^*$, we observe that

$$A^\dagger v_k = \sum_{n=1}^N \frac{(v_k, v_n)_Y}{\sigma_n} u_n = \frac{1}{\sigma_k} u_k \tag{54}$$

and, since

$$\begin{aligned} ((A^\dagger)^* u_k, v_n)_Y &= (u_k, A^\dagger v_n)_X \\ &= \left(u_k, \frac{1}{\sigma_n} u_n \right)_X = \frac{1}{\sigma_n} \delta_{kn} \end{aligned} \tag{55}$$

we obtain that

$$\begin{aligned} (A^\dagger)^* u_k &= \sum_{n=1}^N ((A^\dagger)^* u_k, v_n)_Y v_n \\ &= \sum_{n=1}^N \frac{1}{\sigma_n} \delta_{kn} v_n = \frac{1}{\sigma_k} v_k. \end{aligned} \tag{56}$$

It follows that the set of triples $\{\frac{1}{\sigma_n}; u_n, v_n\}_{n=1}^N$ satisfies the shifted-eigenvalues problem

$$A^\dagger v_n = \frac{1}{\sigma_n} u_n, \quad (A^\dagger)^* u_n = \frac{1}{\sigma_n} v_n \tag{57}$$

and so it is the singular system of the operator A^\dagger ; from (57) we have that

$$A^\dagger (A^\dagger)^* u_k = \frac{1}{\sigma_k} A^\dagger v_k = \frac{1}{\sigma_k^2} u_k \tag{58}$$

and therefore the numbers $\frac{1}{\sigma_k^2}$, $k = 1, \dots, N$ are the eigenvalues of the operator $A^\dagger (A^\dagger)^*$.

Expression (51) of the condition number shows that, even if the determination of the generalized solution of an inverse problem with discrete data is a well-posed problem, the numerical stability is not guaranteed. In fact, the more the singular values fastly decrease, the more the ratio (51) increases and the propagation of the error from the datum to the generalized solution produces very significant effects.

2.4 Regularization Theory: General Formulation

In general terms, *regularization* is the best approximation of an ill-posed problem by a family of neighboring well-posed problems. We motivate the definition of a *regularization operator* and of a *regularization method* in this way: we want to approximate the best-approximate solution $u^\dagger = A^\dagger g$ of the inverse problem (1) for a specific right-hand side g in the situation where the “exact data” g are not known precisely, but only an approximation $g^{(\delta)}$ with

$$\|g^{(\delta)} - g\|_Y \leq \delta \tag{59}$$

is available; we will call $g^{(\delta)}$ the *noisy data* and δ the *noise level*.

In the ill-posed case, and above all when the condition number is particularly great, $A^\dagger g^{(\delta)}$ is certainly not a good approximation of $A^\dagger g$ due to the high numerical instability.

In order to repair this pathology, we look for some approximation, say $f_\lambda^{(\delta)}$, of u^\dagger which does, on the one hand, depend continuously on the (noisy) data $g^{(\delta)}$, so that it can be computed in a stable way, and has, on the other hand, the property that as the noise level δ decreases to zero and the regularization parameter λ is chosen appropriately (whatever this means), then $f_\lambda^{(\delta)}$ tends to u^\dagger .

These considerations lead to the following definition:

Definition 8 Let A be a linear continuous operator with domain in the Hilbert space X and range in the Hilbert space Y . The one-parameter family of operator $\{R_\lambda\}_{\lambda>0}$ such that

- (1) $R_\lambda : Y \rightarrow X$ is bounded $\forall \lambda$;
- (2) $\lim_{\lambda \rightarrow 0} \|R_\lambda g - u^\dagger\|_X = 0$ for every g in Y such that $Pg \in R(A)$, where P is the linear projection onto $\overline{R(A)}$,

is said a *regularization algorithm*.

The regularization algorithm $\{R_\lambda\}_{\lambda>0}$ is said *linear* if R_λ is linear for each value of the regularization parameter λ .

The operators R_λ represent continuous approximations of the generalized inverse operator of A ; in particular, condition (2) can be replaced by

$$\lim_{\lambda \rightarrow 0} \|R_\lambda g - A^\dagger g\|_X = 0, \tag{60}$$

and this relation is true for every g in $R(A) \oplus R(A)^\perp$, which is dense in Y .

The application of a regularization algorithm provides an approximation of the generalized solution of an ill-posed linear inverse problem in the case of noise-free data. However the use of regularization is crucial in the case where the data g is affected by experimental error. In this case $g^{(\delta)}$ can be represented in the form

$$g^{(\delta)} = Au^\dagger + w^{(\delta)} \tag{61}$$

with

$$\|w^{(\delta)}\|_Y = \delta. \tag{62}$$

Expression (61) may not be explicitly known but it is always possible to assume that it exists. Furthermore, it does not necessarily mean that the noise is additive, since $w^{(\delta)}$ may depend on u^\dagger . In the case of a linear regularization algorithm one easily obtains

$$\|R_\lambda g^{(\delta)} - u^\dagger\|_X \leq \|R_\lambda Au^\dagger - u^\dagger\|_X + \delta \|R_\lambda\|. \tag{63}$$

Equation (63) represents a basic inequality in linear regularization theory. The first term at the right hand side represents the approximation error due to the use of R_λ instead of the generalized inverse operator; from condition (2) it tends to zero when λ tends to zero. On the other hand, the second term measures the error on the regularized solution $R_\lambda g^{(\delta)}$

due to the presence of noise on the data and typically grows up to infinity when λ tends to 0. Every regularization algorithm requires a strategy for choosing the parameter λ in dependence on the error level δ in order to achieve an acceptable total error for the regularized solution. On the one hand, the accuracy of the approximation asks for a small error $\|R_\lambda Au^\dagger - u^\dagger\|_X$, i.e., for a small parameter λ . On the other hand, the stability requires a small $\|R_\lambda\|$, i.e., a large parameter λ . An optimal choice would find a value $\lambda_{opt}(\delta)$ of the regularization parameter such that the right hand side of (63) becomes minimal. This choice of the regularization parameter realizes a compromise between accuracy and stability. For a reasonable regularization strategy we expect the regularized solution to converge to the exact solution when the error level tends to zero. We express this requirement through the following definition:

Definition 9 A regularization algorithm $\{R_\lambda\}_{\lambda>0}$ is said *regular* if, for $\delta \rightarrow 0$, $\lambda_{opt}(\delta) \rightarrow 0$ and $R_{\lambda_{opt}(\delta)} g^{(\delta)} \rightarrow u^\dagger$.

When we deal with a strongly ill-posed linear inverse problem, $\lambda_{opt}(\delta)$ changes very slowly for different values of δ . Moreover, we observe that, even if a function $\lambda_{opt}(\delta)$ exists, this does not mean that it is easy to determine it. In general, the estimate of a reliable $\lambda_{opt}(\delta)$ is the main problem in regularization theory.

2.5 A Regularization Algorithm: The Tikhonov Method

The Tikhonov method is, historically, the first algorithm rigorously described in regularization theory and it has been introduced in order to solve Fredholm integral equations of the first kind [79, 80]. The first step in the definition of such method (in a general context) is the minimization of the functional

$$\Phi_\lambda[f] = \|Af - g\|_Y^2 + \lambda \|f\|_X^2 \tag{64}$$

with λ a real positive number. It is not difficult to prove that for each λ the minimum problem is equivalent to the Euler equation

$$(A^*A + \lambda I)f = A^*g. \tag{65}$$

Indeed, f_λ is a solution of the minimum problem if and only if, for any t in \mathbb{C} and for any function ϕ in the Hilbert space X , we have

$$\begin{aligned} & \|Af_\lambda - g\|_Y^2 + \lambda \|f_\lambda\|_X^2 \\ & \leq \|A(f_\lambda + t\phi) - g\|_Y^2 + \lambda \|f_\lambda + t\phi\|_X^2. \end{aligned} \tag{66}$$

By writing the norms as scalar products one easily obtains

$$|t|^2 (\|A\phi\|_Y^2 + \lambda \|\phi\|_X^2) + t \{ (A\phi, Af_\lambda - g)_Y + \lambda (\phi, f_\lambda)_X \}$$

$$+ \bar{t}\{(Af_\lambda - g, A\phi)_Y + \lambda(f_\lambda, \phi)_X\} \geq 0. \tag{67}$$

From the arbitrariness of the complex number t we have that relation (67) holds if and only if the linear term in λ

$$(Af_\lambda - g, A\phi)_Y + \lambda(f_\lambda, \phi)_X \tag{68}$$

is zero for all ϕ in X and therefore the Euler equation follows.

Since $A^*A + \lambda I$ is (strictly) positive definite, its inverse operator is continuous.¹ It follows that the function f_λ which solves the Euler equation is given by

$$f_\lambda = (A^*A + \lambda I)^{-1} A^*g. \tag{69}$$

If we define the operator

$$R_\lambda = (A^*A + \lambda I)^{-1} A^*, \tag{70}$$

then the following theorem holds:

Theorem 5 *The one-parameter family of operator $\{R_\lambda\}_{\lambda>0}$ defined by*

$$R_\lambda = (A^*A + \lambda I)^{-1} A^* \tag{71}$$

is a linear regularization algorithm. The algorithm is also regular.

Proof First of all, the relation

$$(A^*A + \lambda I)A^* = A^*(AA^* + \lambda I) \tag{72}$$

holds and, if we multiply both members on the left by $(A^*A + \lambda I)^{-1}$ and on the right by $(AA^* + \lambda I)^{-1}$, we obtain

$$R_\lambda = A^*(AA^* + \lambda I)^{-1}. \tag{73}$$

This means that the range of the regularized inverse operator R_λ is contained in the range of A^* ; then, from Theorem 2, we have that the range of R_λ is contained in the range of A^\dagger .

The continuity of R_λ is a consequence of the inequality

$$\|R_\lambda g\|_X^2 \leq \|AA^*(AA^* + \lambda I)^{-1}\| \|(AA^* + \lambda I)^{-1}\| \|g\|_Y^2 \tag{74}$$

obtained by using the Schwartz inequality in the relation

$$\begin{aligned} \|R_\lambda g\|_X^2 &= (A^*(AA^* + \lambda I)^{-1}g, A^*(AA^* + \lambda I)^{-1}g)_X \\ &= (AA^*(AA^* + \lambda I)^{-1}g, (AA^* + \lambda I)^{-1}g)_X. \end{aligned} \tag{75}$$

¹In fact, if $\lambda > 0$ and $f \in X$, then

$$\| \lambda f \|^2 = (\lambda f, f)_X \leq ((A^*A + \lambda I)f, f)_X \leq \|(A^*A + \lambda I)f\|_X \|f\|_X$$

and the claim follows from Theorem 12.12(c) of [73].

Now, if $\Omega(AA^*)$ is the spectrum of AA^* , then

$$\|AA^*(AA^* + \lambda I)^{-1}\| = \sup_{\omega \in \Omega(AA^*)} \frac{\omega}{\omega + \lambda} \leq 1 \tag{76}$$

and

$$\|(AA^* + \lambda I)^{-1}\| = \sup_{\omega \in \Omega(AA^*)} \frac{1}{\omega + \lambda} \leq \frac{1}{\lambda} \tag{77}$$

so that

$$\|R_\lambda\| \leq \frac{1}{\sqrt{\lambda}}, \tag{78}$$

i.e. R_λ is continuous for all $\lambda > 0$.

In order to prove the second property of the regularization algorithms, we need arguments based on the spectral theory of linear continuous operator. Let $f \in N(A)^\perp$; then

$$\|R_\lambda Af - f\|_X = \left\| \int_0^{\|A\|^2} \frac{\lambda}{\omega + \lambda} dE_\omega f \right\|_X \tag{79}$$

where dE_ω is the spectral measure associated to the operator A^*A (which is self-adjoint and positive semidefinite).² The function $\omega \mapsto \frac{\lambda}{\omega + \lambda}$ is integrable with respect to the spectral measure over $[0, \|A\|^2]$ and is bounded by 1, integrable over the same interval. The Dominated Convergence Theorem now implies

$$\lim_{\lambda \rightarrow 0} \|R_\lambda Af - f\|_X = \|E_0 f\| \tag{80}$$

where E_0 is the projection onto $N(A^*A) = N(A)$. It follows that $E_0 f = 0$.

Finally the regularity of the algorithm is guaranteed by the choice of a function $\lambda_{opt}(\delta)$ such that $\frac{\delta}{\sqrt{\lambda}} \rightarrow 0$ since in this case inequality (63) becomes

$$\|R_\lambda g^{(\delta)} - u^\dagger\|_X \leq \|R_\lambda Au^\dagger - u^\dagger\|_X + \frac{\delta}{\sqrt{\lambda}}. \quad \square \tag{81}$$

In the case of linear inverse problems with discrete data, the knowledge of the singular system of the operator A allows us to write the *regularized solution* f_λ (i.e. the function which minimizes the Tikhonov functional defined in (64)) in a very useful way for the applications. In fact, if we write the data vector \mathbf{g} as

$$\mathbf{g} = \sum_{n=1}^N (\mathbf{g}, \mathbf{v}_n)_Y \mathbf{v}_n \tag{82}$$

²From the Spectral Theorem (see [72]) we have that

$$h(A^*A) = \int_{\Omega(A^*A)} h(\omega) dE_\omega.$$

In this case $h(\omega) = |\frac{\omega}{\omega + \lambda} - 1| = \frac{\lambda}{\omega + \lambda}$.

Moreover, since $\sup_{\omega \in \Omega(A^*A)} |\omega| = \|A^*A\| = \|A\|^2$, it follows that $\Omega(A^*A) \subseteq [0, \|A\|^2]$.

and we recall the shifted-eigenvalues problem (45) which defines the singular system of the operator A , then the Euler equation (65) becomes [9]

$$f_\lambda = \sum_{k=1}^N \frac{\sigma_k}{\sigma_k^2 + \lambda} (\mathbf{g}, \mathbf{v}_k)_Y u_k. \tag{83}$$

However, it is interesting to notice that, in the case of a linear inverse problem with discrete data, the knowledge of the singular system of the operator is not necessary if we want to find the function which minimizes the Tikhonov functional (64). In fact we can obtain R_λ also from (73); but $(AA^* + \lambda I)$ assumes values in the Euclidean space Y with finite dimension N , so it can be described by a $N \times N$ matrix. It follows that the determination of R_λ implies the inversion of a matrix followed by the application of the operator A^* .

2.6 Optimal Choice of the Regularization Parameter

The parameter choice rule $\lambda = \lambda(\delta, g^{(\delta)})$ depends explicitly on the noise level δ and on the actual perturbed data $g^{(\delta)}$. Also, it usually depends on every specific g in the domain of the generalized inverse operator A^\dagger ; since g is not known, this dependence can only be on some qualitative a-priori knowledge about g like smoothness properties. Finally, λ depends also on the operator A .

We distinguish between two types of parameter choice rules:

Definition 10 Let $\lambda = \lambda(\delta, g^{(\delta)})$ be a parameter choice rule. If λ does not depend on $g^{(\delta)}$, but only on δ , then we call λ an *a-priori* parameter choice rule. Otherwise, we call λ an *a-posteriori* parameter choice rule.

Thus, an a-priori parameter choice rule depends only on the noise level, not on the actual data and, hence, not on results obtained during the actual computation like the residual $\|Af_\lambda - g^{(\delta)}\|_Y$, where $f_\lambda = R_\lambda g^{(\delta)}$ is the regularized solution. Such a rule may be devised before the actual calculation, hence the name a-priori parameter choice rule.

Let us consider now the specific case of the Tikhonov regularization algorithm. According to Theorem 5, any a-priori choice of the regularization parameter $\lambda = \lambda(\delta)$ satisfying $\delta^2/\lambda(\delta) \rightarrow 0$ as $\delta \rightarrow 0$ leads to a regular algorithm for the solution of the linear inverse problem $Af = g$. Although this asymptotic result may be theoretically satisfying, it would seem that a choice of the regularization parameter that is based on the actual computations performed, that is an a-posteriori choice of the regularization parameter, would be more effective in practice. One such a-posteriori strategy is the *discrepancy principle* of Morozov (see [9]). The idea of the strategy is to choose the regularization parameter so that

the size of the residual $\|Af_\lambda - g^{(\delta)}\|_Y$ is the same as the error level in the data, i.e.,

$$\|Af_\lambda - g^{(\delta)}\|_Y = \delta. \tag{84}$$

In the case of linear inverse problems with discrete data, assuming that the signal-to-noise ratio is larger than one, that is $\|\mathbf{g}^{(\delta)}\| > \delta$, and that $\mathbf{g} \in R(A)$, then it is not hard to see that there is a unique positive parameter λ satisfying (84). To do this, we use the singular value decomposition

$$\|Af_\lambda - \mathbf{g}^{(\delta)}\|_Y^2 = \sum_{k=1}^N \left(\frac{\lambda}{\sigma_k^2 + \lambda} \right)^2 |(\mathbf{g}^{(\delta)}, \mathbf{v}_k)|^2 + \|P\mathbf{g}^{(\delta)}\|^2 \tag{85}$$

where P is the projector of Y onto $R(A)^\perp$. From (85) we see that the real function

$$f(\lambda) = \|Af_\lambda - \mathbf{g}^{(\delta)}\|_Y^2 \tag{86}$$

is a continuous, increasing function of λ satisfying (since $P\mathbf{g} = 0$)

$$\begin{aligned} \lim_{\lambda \rightarrow 0^+} f(\lambda) &= \|P\mathbf{g}^{(\delta)}\| = \|P\mathbf{g}^{(\delta)} - P\mathbf{g}\| \\ &\leq \|\mathbf{g}^{(\delta)} - \mathbf{g}\| \leq \delta \end{aligned} \tag{87}$$

and

$$\lim_{\lambda \rightarrow +\infty} f(\lambda) = \|\mathbf{g}^{(\delta)}\| > \delta. \tag{88}$$

Therefore, by the Intermediate Value Theorem, there is a unique $\lambda = \lambda(\delta, \mathbf{g}^{(\delta)})$ satisfying (84).

We close this section by showing that the choice $\lambda(\delta, g^{(\delta)})$ as given by the discrepancy method (84) leads to a regular scheme for approximating $A^\dagger g$, that is

$$f_{\lambda(\delta, g^{(\delta)})} \rightarrow u^\dagger \quad \text{as } \delta \rightarrow 0. \tag{89}$$

To do this it is sufficient to show that for any sequence $\delta_n \rightarrow 0$ there is a subsequence, which for notational convenience we will denote by $\{\delta_k\}$, such that

$$f_{\lambda(\delta_k, g^{(\delta_k)})} \rightarrow u^\dagger. \tag{90}$$

We are assuming that $g \in R(A)$ and we recall that u^\dagger is the unique vector in X satisfying $Au^\dagger = g$ and $u^\dagger \in N(A)^\perp$. For sake of simplicity we will neglect the dependency of $\lambda(\delta, g^{(\delta)})$ on $g^{(\delta)}$.

From the functional characterization (64) of the Tikhonov approximation we have

$$\Phi_{\lambda(\delta)}[f_{\lambda(\delta)}] \leq \Phi_{\lambda(\delta)}[u^\dagger] \tag{91}$$

that is,

$$\begin{aligned} \delta^2 + \lambda(\delta)\|f_{\lambda(\delta)}\|_X^2 &= \|Af_{\lambda(\delta)} - g^{(\delta)}\|_Y^2 + \lambda(\delta)\|f_{\lambda(\delta)}\|_X^2 \\ &\leq \Phi_{\lambda(\delta)}[u^\dagger] \\ &= \|g - g^{(\delta)}\|_Y^2 + \lambda(\delta)\|u^\dagger\|_X^2 \\ &\leq \delta^2 + \lambda(\delta)\|u^\dagger\|_X^2 \end{aligned} \tag{92}$$

and hence $\|f_{\lambda(\delta)}\|_X \leq \|u^\dagger\|_X$. Therefore, for any sequence $\delta_n \rightarrow 0$ there is a subsequence $\delta_k \rightarrow 0$ with $f_{\lambda(\delta_k)} \rightharpoonup z$ for some $z \in X$ (where the arrow \rightharpoonup denote the weak convergence). Since

$$f_{\lambda(\delta)} = A^*(AA^* + \lambda(\delta)I)^{-1}g^{(\delta)} \in R(A^*) \subseteq N(A)^\perp \tag{93}$$

and $N(A)^\perp$ is weakly closed, we find that $z \in N(A)^\perp$. Also, since

$$\|Af_{\lambda(\delta_k)} - g^{(\delta_k)}\|_Y \rightarrow 0 \tag{94}$$

we see that $Af_{\lambda(\delta_k)} \rightarrow g$. But A is weakly continuous and therefore

$$Af_{\lambda(\delta_k)} \rightharpoonup Az. \tag{95}$$

It follows that $Az = g$ and $z \in N(A)^\perp$, i.e., $z = u^\dagger$.

We observe now that, since

$$f_{\lambda(\delta_k)} \rightharpoonup u^\dagger \tag{96}$$

and the operator $f \rightarrow \|f\|_X$ is continuous and convex, from Corollary 1.8.3 of [2] we have that

$$\|u^\dagger\|_X \leq \liminf_{k \rightarrow +\infty} \|f_{\lambda(\delta_k)}\|_X. \tag{97}$$

But $\|f_{\lambda(\delta_k)}\|_X \leq \|u^\dagger\|_X \forall k$, and therefore

$$\|f_{\lambda(\delta_k)}\|_X \rightarrow \|u^\dagger\|_X. \tag{98}$$

From relations (96) and (98) it follows that

$$f_{\lambda(\delta_k)} \rightarrow u^\dagger, \tag{99}$$

(see Theorem 1.8.3 of [2]) and the proof is complete.

3 Application to Solar Physics: Non Thermal Bremsstrahlung

3.1 The Bremsstrahlung Equation

A solar flare is the rapid release of a large amount of energy stored in the solar atmosphere. During a flare, gas is heated from 10 to 20 million degrees Kelvin and radiates both soft X-rays and longer-wavelength emission (hard X-rays and γ -rays). It is important to notice that the analysis

of the X-ray spectra provides information about the physical processes that take place in the magnetized plasma of the solar atmosphere during a flare, such as impulsive energy release, particle acceleration and particle and energy transport [78]. This high-energy processes play a major role at sites throughout the universe ranging from magnetosphere to active galaxies. Consequently, the importance of understanding these processes transcends the field of solar physics, and represents one of the major goals of space physics and astrophysics.

Hard X-rays are emitted by electrons with relativistic velocities, typically impossible to reach if we take into account simply the agitation due to the temperature of the plasma. Soft X-rays have lower energies and are emitted by electrons with lower or thermal velocities. In the last case the electrons assume shift velocities (about 0.05 times c) that make them leave their original atomic nuclei. These electrons are then attracted by other atomic nuclei which slow them down. The excess energy is released in the form of X-rays in a process known with the German term *bremsstrahlung* (“braking” radiation). The same process is followed also by electrons with relativistic velocities, but in this case the energies involved are much bigger and cause the emission of hard X-rays.

The bremsstrahlung of electrons with the ions of the plasma is then a collisional process and so it is characterized by a cross section which, generally, depends on the energies of the photons and of the X-ray producing electrons. If this cross section is analytically known, the photon spectrum can be directly related to the electron distribution. Generally, the study of the emission process can be made under particular hypotheses about the physical conditions of the source. First of all, bremsstrahlung radiation is considered optically thin [15], that is absorption can be neglected; this implies that the observed X-ray spectra are quite similar to the ones emerging from the emitting region. Moreover the electron velocity distribution is characterized by isotropic conditions. Finally the plasma is assumed to be hydrogen dominated, so that the ions are almost completely protons. Under these hypotheses, the equation linking the distribution function of electrons with the hard X-ray intensity observed at distance R from a source can be written in the following way [20]:

$$I(\epsilon) = \frac{1}{4\pi R^2} \int_V n(\mathbf{r}) \int_\epsilon^\infty F(E, \mathbf{r}) Q(\epsilon, E) dE d\mathbf{r} \tag{100}$$

where V is the source volume, $n(\mathbf{r})$ is the local proton density in the plasma, E is the electron energy, ϵ is the photon energy, $F(E, \mathbf{r})$ is the electron distribution function, $Q(\epsilon, E)$ is the bremsstrahlung cross section, differential in ϵ , $I(\epsilon)$ is the total rate of photon emission measured in photons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$. It must be noted that $I(\epsilon)$ represents the flux at the sun; in practice, measurements regard fluxes at the earth, but here only the shape of the spectrum and the

relative errors in it are considered, so that the multiplicative constants are irrelevant.

Averaging the electron distribution function over the volume of the emitting region and weighting this function with the density of the ions, one obtains the relation [20]:

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \bar{F}(E)Q(\epsilon, E)dE, \quad (101)$$

where the mean electron spectrum $\bar{F}(E)$ is given by

$$\bar{F}(E) = \frac{1}{\bar{n}V} \int_V n(\mathbf{r})F(E, \mathbf{r})d\mathbf{r}, \quad (102)$$

and the mean target ion density in the source is defined as

$$\bar{n} = \frac{1}{V} \int_V n(\mathbf{r})d\mathbf{r}. \quad (103)$$

For sake of simplicity, we define

$$f(E) = \bar{n}V\bar{F}(E), \quad (104)$$

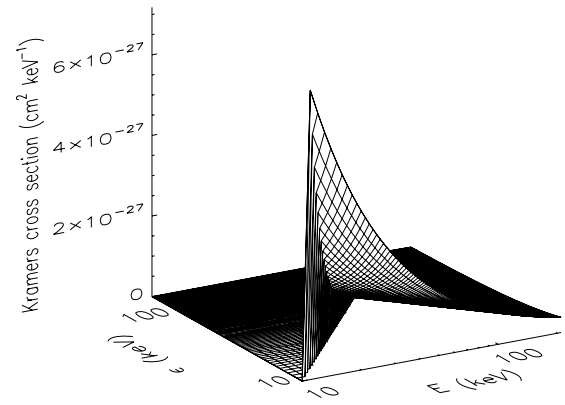
$$g(\epsilon) = 4\pi R^2 I(\epsilon) \quad (105)$$

and the integral equation we are interested in becomes

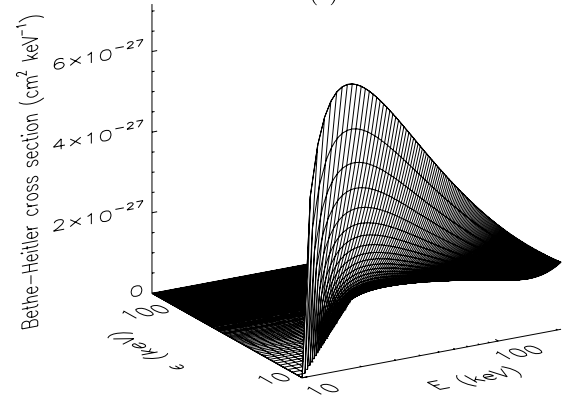
$$g(\epsilon) = \int_{\epsilon}^{\infty} f(E)Q(\epsilon, E)dE. \quad (106)$$

Equation (106) is the bremsstrahlung equation in solar plasma physics: it can be written without any assumption on the physical processes in the source and just for this reason it is the most general equation describing the X-ray emission mechanism during solar flares. We point out that this approach is completely isotropic, i.e. no angular dependency in the mean electron spectrum or in the cross section is considered. First investigations toward anisotropic modelization of the emission will be described in Sect. 5, although further generalizations of (106) are in progress.

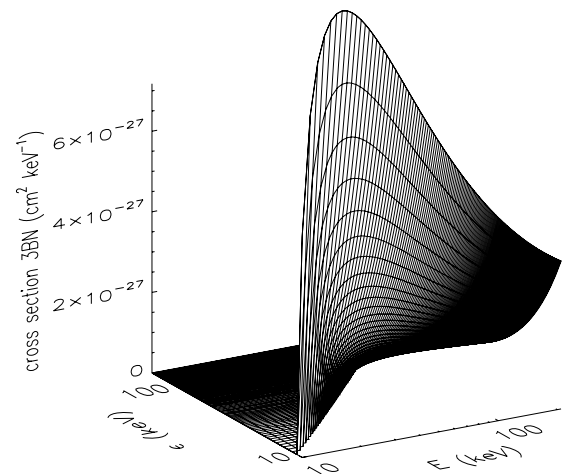
A crucial role in (106) is played by the cross section $Q(\epsilon, E)$ differential in the photon energy ϵ at electron energy E . In practice, $Q(\epsilon, E)$ measures the probability that an X-ray photon of energy ϵ is emitted by an electron of energy $E \geq \epsilon$. This function carries a significant physical meaning, describing the effectiveness of the bremsstrahlung process. Furthermore, on the mathematical side, its analytical shape has important consequences on the numerical stability of the solution process and provides a reliable indication of the information content which can be retrieved from the X-ray data. According to the physical conditions where the emission process takes place, many different forms of such a cross section can be written [44]. Among them, three representatives are particularly meaningful, representing three different amounts of the impact due to the relativistic effects on the emission process. These three cross sections, represented in Fig. 1, are:



(a)



(b)



(c)

Fig. 1 Cross sections: (a) Kramers approximation; (b) Bethe-Heitler approximation; (c) highly relativistic cross section

- The Kramers cross section

$$Q_K(\epsilon, E) = \frac{Q_0}{\epsilon E}. \quad (107)$$

It is a completely classical formula, where the relativistic effects are neglected. In (107) the multiplicative constant

factor Q_0 is defined as

$$Q_0 = \frac{8}{3} \frac{Zr_0^2}{137} mc^2, \tag{108}$$

where r_0 is the classical radius of the electron, m is its rest mass, c is the speed of light in the vacuum and Z is the average atomic number in the plasma (Z is of the order of 1, since the plasma is mostly made of ionized hydrogen).

- The Bethe-Heitler formula

$$Q_{BH}(\epsilon, E) = \frac{Q_0}{\epsilon E} \log \frac{1 + \sqrt{1 - \frac{\epsilon}{E}}}{1 - \sqrt{1 - \frac{\epsilon}{E}}}, \tag{109}$$

where Q_0 is as in (108) and which accounts for mild relativistic effects in the logarithmic factor at the right hand side.

- The highly relativistic formula given by equation (3BN) in [44]. Such cross section is considered the most general one for this kind of process and its analytical form is

$$\begin{aligned} Q_{3BN}(\epsilon, E) &= \frac{Z^2 r_0^2}{137} \frac{p_f}{k p_0} \left\{ \frac{4}{3} - 2E_0 E_f \left(\frac{p_f^2 + p_0^2}{p_f^2 p_0^2} \right) \right. \\ &+ \frac{v_0 E_f}{p_0^3} + \frac{v_f E_0}{p_f^3} - \frac{v_0 v_f}{p_0 p_f} \\ &+ L \left[\frac{8E_0 E_f}{3 p_0 p_f} + \frac{k^2 (E_0^2 E_f^2 + p_0^2 p_f^2)}{p_0^3 p_f^3} \right. \\ &+ \frac{k}{2 p_0 p_f} \left(\left(\frac{E_0 E_f + p_0^2}{p_0^3} \right) v_0 \right. \\ &\left. \left. - \left(\frac{E_0 E_f + p_f^2}{p_f^3} \right) v_f + \frac{2k E_0 E_f}{p_f^2 p_0^2} \right) \right] \Bigg\} \end{aligned} \tag{110}$$

where

$$\begin{aligned} k &= \frac{\epsilon}{mc^2}; & E_0 &= \frac{E}{mc^2} + 1; & E_f &= E_0 - k; \\ p_0 &= \sqrt{E_0^2 - 1}; & p_f &= \sqrt{E_f^2 - 1} \\ L &= 2 \ln \left[\frac{E_0 E_f + p_0 p_f - 1}{k} \right]; \\ v_0 &= \ln \left(\frac{E_0 + p_0}{E_0 - p_0} \right); & v_f &= \ln \left(\frac{E_f + p_f}{E_f - p_f} \right). \end{aligned}$$

There are two reasons why it is interesting to study (106) for different forms of the cross section. The first one is physical and is the fact that the differences in the solution $f(E)$ for the three different forms of $Q(\epsilon, E)$ previously introduced can provide quantitative information on the incidence

of the relativistic corrections on the photon production in the solar plasma. The second reason is computational and is the fact that the numerical solution of (106) is notably less onerous in the case of (107) and (109) and then it is helpful to verify to what extent the highly relativistic effects can be neglected in the solution procedure. In the next two sections an analytical study of the solution of the bremsstrahlung problem will be performed in the case of continuous and discrete photon spectra for the Kramers and Bethe-Heitler approximations. The case concerned with the more complicated cross section (110) will be treated in Sect. 3.4.

3.2 Analytical Study of the Bremsstrahlung Equation

In the case of the Kramers and Bethe-Heitler cross sections, (106) can be analytically solved by applying the theory of Mellin transform [76]. The Mellin transform of a function $f : (0, \infty) \rightarrow \mathbb{C}$ such that

$$\int_0^\infty |f(x)| x^{\sigma-1} dx < \infty \tag{111}$$

for some real number σ in $(0, \frac{1}{2})$ is defined as

$$\tilde{f}(\xi) = \int_0^\infty f(x) x^{-\frac{1}{2}+i\xi} dx, \tag{112}$$

where ξ belongs to $(-\infty, +\infty)$.

Theorem 6 *Let $f : (0, \infty) \rightarrow \mathbb{C}$ be a continuously differentiable function such that condition (111) holds. Then*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\xi) x^{-\frac{1}{2}-i\xi} d\xi \tag{113}$$

for every $x \in (0, \infty)$.

Proof If we do the change of variable $x = e^{-t}$ in (112), we see that the Mellin transform of a function $x \mapsto f(x)$ coincides with the Fourier transform of the function $t \mapsto F(t) = f(e^{-t})e^{-t/2}$. In fact,

$$\begin{aligned} \tilde{f}(\xi) &= \int_0^\infty f(x) x^{-\frac{1}{2}+i\xi} dx \\ &= \int_{-\infty}^{+\infty} f(e^{-t}) e^{(\frac{1}{2}-i\xi)t} e^{-t} dt \\ &= \int_{-\infty}^{+\infty} \left[f(e^{-t}) e^{-\frac{t}{2}} \right] e^{-i\xi t} dt \\ &= \int_{-\infty}^{+\infty} F(t) e^{-i\xi t} dt = \widehat{F}(\xi). \end{aligned} \tag{114}$$

This equality and the inversion formula for the Fourier transform lead to the inversion formula for the Mellin transform.

Indeed

$$f(e^{-t})e^{-\frac{t}{2}} = F(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\xi)e^{i\xi t} d\xi. \tag{115}$$

Coming back to the variable x we have

$$f(x)\sqrt{x} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\xi)x^{-i\xi} d\xi \tag{116}$$

and the explicit form of the Mellin inverse transform follows:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\xi)x^{-\frac{1}{2}-i\xi} d\xi. \quad \square \tag{117}$$

The Mellin transform is particularly helpful in the case of linear integral equations of the first kind when the kernel depends on the variables ratio. Indeed the following theorem holds:

Theorem 7 *We consider the integral equation*

$$g(x) = \int_0^{+\infty} k\left(\frac{x}{y}\right) f(y) \frac{dy}{y}. \tag{118}$$

The application of the Mellin transform to both sides of this equation leads to the diagonalization

$$\tilde{g}(\xi) = \tilde{f}(\xi)\tilde{k}(\xi). \tag{119}$$

Proof We have

$$\begin{aligned} \tilde{g}(\xi) &= \int_0^{+\infty} g(x)x^{-\frac{1}{2}+i\xi} dx \\ &= \int_0^{+\infty} \left(\int_0^{+\infty} k\left(\frac{x}{y}\right) f(y) \frac{dy}{y} \right) x^{-\frac{1}{2}+i\xi} dx \\ &= \left(\int_0^{+\infty} f(y)y^{-\frac{1}{2}+i\xi} dy \right) \\ &\quad \times \left(\int_0^{+\infty} k\left(\frac{x}{y}\right) \left(\frac{x}{y}\right)^{-\frac{1}{2}+i\xi} \frac{dx}{y} \right) \\ &= \left(\int_0^{+\infty} f(y)y^{-\frac{1}{2}+i\xi} dy \right) \left(\int_0^{+\infty} k(z)z^{-\frac{1}{2}+i\xi} dz \right) \\ &= \tilde{f}(\xi)\tilde{k}(\xi). \quad \square \end{aligned} \tag{120}$$

When $Q(\epsilon, E)$ assumes the forms (107) or (109) the bremsstrahlung equation (106) assumes the same form as (118). In fact, if one introduces the new integral kernel

$$K(\epsilon, E) = \begin{cases} 0, & E < \epsilon, \\ \frac{\epsilon E}{Q_0} Q(\epsilon, E), & E \geq \epsilon \end{cases} \tag{121}$$

and defines the new data function

$$J(\epsilon) = \frac{\epsilon g(\epsilon)}{Q_0}, \tag{122}$$

(106) becomes

$$J(\epsilon) = \int_0^{+\infty} f(E)K(\epsilon, E) \frac{dE}{E}. \tag{123}$$

Now let us introduce the functions

$$k(\epsilon/E) = k_K(\epsilon/E) = \begin{cases} 1, & \epsilon/E \leq 1, \\ 0, & \text{otherwise} \end{cases} \tag{124}$$

for the Kramers cross section and

$$\begin{aligned} k(\epsilon/E) &= k_{BH}(\epsilon/E) \\ &= \begin{cases} \log \frac{1+\sqrt{1-\frac{\epsilon}{E}}}{1-\sqrt{1-\frac{\epsilon}{E}}}, & \epsilon/E \leq 1, \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{125}$$

for the Bethe-Heitler cross section, so that (123) can be written as

$$J(\epsilon) = \int_0^{+\infty} f(E)k(\epsilon/E) \frac{dE}{E}. \tag{126}$$

It follows that, if the Mellin transform of the integral kernels (124) and (125) does not vanish anywhere, the analytical solution of (106) when these two cross sections are used can be easily obtained by using Theorems 6 and 7.

Theorem 8 *Let $f : (0, \infty) \rightarrow \mathbb{C}$ be a continuously differentiable function such that condition (111) holds. Then the solution of (106) is given by*

$$f(E) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1+2i\xi}{2} \tilde{J}(\xi) E^{-(\frac{1}{2}+i\xi)} d\xi \tag{127}$$

in the case of the Kramers cross section and by

$$f(E) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1+2i\xi}{2\sqrt{\pi}} \frac{\Gamma(1+i\xi)}{\Gamma(\frac{1}{2}+i\xi)} \tilde{J}(\xi) E^{-(\frac{1}{2}+i\xi)} d\xi \tag{128}$$

in the case of the Bethe-Heitler cross section.

Proof Let us introduce the Beta function $B(p, q)$ defined as

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx, \quad \text{Re}(p) > 0, \text{Re}(q) > 0 \tag{129}$$

and its relation $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ with the Gamma function defined as

$$\Gamma(z) = \int_0^{+\infty} e^{-t} t^{z-1} dt, \quad \text{Re}(z) > 0. \tag{130}$$

We observe that, for the Kramers case,

$$\begin{aligned} \tilde{k}_K(\xi) &= \int_0^1 x^{-\frac{1}{2}+i\xi} dx = B\left(\frac{1}{2} + i\xi, 1\right) \\ &= \frac{\Gamma(\frac{1}{2} + i\xi)\Gamma(1)}{\Gamma(\frac{3}{2} + i\xi)} = \frac{2}{1 + 2i\xi}, \end{aligned} \tag{131}$$

where in the last equality we used the relations $\Gamma(1) = 1$ and $\Gamma(1 + z) = z\Gamma(z)$, $z \in \mathbb{C}$.

For the Bethe-Heitler case, integration by parts leads to

$$\begin{aligned} \tilde{k}_{BH}(\xi) &= \int_0^1 \log \frac{1 + \sqrt{1-x}}{1 - \sqrt{1-x}} x^{-\frac{1}{2}+i\xi} dx \\ &= - \int_0^1 \frac{-1}{x\sqrt{1-x}} \cdot \frac{x^{\frac{1}{2}+i\xi}}{\frac{1}{2} + i\xi} dx \\ &= \frac{1}{\frac{1}{2} + i\xi} \int_0^1 \frac{1}{\sqrt{1-x}} x^{-\frac{1}{2}+i\xi} dx \\ &= \frac{1}{\frac{1}{2} + i\xi} B\left(\frac{1}{2} + i\xi, \frac{1}{2}\right). \end{aligned} \tag{132}$$

Since $\Gamma(1/2) = \sqrt{\pi}$ we have

$$\tilde{k}_{BH}(\xi) = \frac{\sqrt{\pi}}{\frac{1}{2} + i\xi} \frac{\Gamma(\frac{1}{2} + i\xi)}{\Gamma(1 + i\xi)} = \frac{2\sqrt{\pi}}{1 + 2i\xi} \frac{\Gamma(\frac{1}{2} + i\xi)}{\Gamma(1 + i\xi)}. \tag{133}$$

We finally observe that both $\tilde{k}_K(\xi)$ and $\tilde{k}_{BH}(\xi)$ do not vanish anywhere. \square

Even if Theorem 8 formally gives the solutions of (106) in the case of both Kramers and Bethe-Heitler cross sections, we have to point out the intrinsic ill-posedness of this procedure. In fact, \tilde{k}_K and \tilde{k}_{BH} are analytical functions which tend to zero when $|\xi| \rightarrow \infty$. Indeed, while in the Kramers approximation the computation is trivial, in the Bethe-Heitler case we have that

$$\begin{aligned} &\lim_{|\xi| \rightarrow \infty} |\tilde{k}_{BH}(\xi)| \\ &= \lim_{|\xi| \rightarrow \infty} \left| \frac{2\sqrt{\pi}}{1 + 2i\xi} \cdot \frac{\Gamma(\frac{1}{2} + i\xi)}{\Gamma(1 + i\xi)} \right| \\ &= \lim_{|\xi| \rightarrow \infty} \frac{2\sqrt{\pi}}{\sqrt{1 + 4\xi^2}} \cdot \frac{|\Gamma(\frac{1}{2} + i\xi)|}{|\Gamma(1 + i\xi)|} \\ &= \lim_{|\xi| \rightarrow \infty} \frac{2\sqrt{\pi}}{\sqrt{1 + 4\xi^2}} \cdot \frac{|\Gamma(\frac{1}{2} + i\xi)|}{|i\xi\Gamma(i\xi)|} \\ &= \lim_{|\xi| \rightarrow \infty} \frac{2\sqrt{\pi}}{\sqrt{1 + 4\xi^2}} \cdot \sqrt{\frac{\pi}{\cosh(\pi\xi)}} \cdot \frac{1}{|\xi|} \cdot \sqrt{\frac{|\xi| |\sinh(\pi\xi)|}{\pi}} \\ &= \lim_{|\xi| \rightarrow \infty} \frac{2\sqrt{\pi}}{\sqrt{1 + 4\xi^2}} \cdot \sqrt{\left| \frac{\tanh(\pi\xi)}{\xi} \right|} = 0, \end{aligned} \tag{134}$$

where we used the properties of the Gamma function (see [1])

$$\Gamma(i\xi)\Gamma(-i\xi) = |\Gamma(i\xi)|^2 = \frac{\pi}{\xi \sinh(\pi\xi)}; \tag{135}$$

$$\begin{aligned} \Gamma\left(\frac{1}{2} + i\xi\right)\Gamma\left(\frac{1}{2} - i\xi\right) &= \left| \Gamma\left(\frac{1}{2} + i\xi\right) \right|^2 \\ &= \frac{\pi}{\cosh(\pi\xi)}. \end{aligned} \tag{136}$$

Therefore a small perturbation of $\tilde{J}(\xi)$ for large values of ξ produces a completely different solution, since, as follows from (127) and (128), this small perturbation is divided by the vanishing values of \tilde{k}_K and \tilde{k}_{BH} . In general, the integrals (127) and (128) do not converge when (119) is replaced by the more realistic one

$$\tilde{g}(\xi) = \tilde{f}(\xi)\tilde{k}(\xi) + \tilde{n}(\xi) \tag{137}$$

where $\tilde{n}(\xi)$ is a function describing the effect of the noise on the data.

For practical applications, (127) and (128) are of little interest. Indeed, in order to obtain the mean electron spectrum from the analytic solutions of the bremsstrahlung equation provided by the previous theorem, the photon spectrum should be known over the whole energy interval (ideally, up to infinite photon energies) and with an accuracy sufficiently high to make the computation of the Mellin transform of the data possible. However, in real observations, the detectors provide data vectors whose components correspond to noisy values of the X-ray spectra sampled over a bounded range of photon energies. Therefore it is natural to study the functional analytic properties of the operator $A_1 : L^2(\epsilon_{min}, \infty) \rightarrow L^2(\epsilon_{min}, \epsilon_{max})$ defined as

$$(A_1 f)(\epsilon) = \int_{\epsilon_{min}}^{\infty} f(E)K(\epsilon, E) \frac{dE}{E} \tag{138}$$

when $\epsilon_{min} > 0$, $\epsilon_{max} < \infty$ and the integral kernel is given by (121) with $Q(\epsilon, E)$ as in (107) for the Kramers case and as in (109) for the Bethe-Heitler case.

The following theorem proves that such an operator is compact.

Theorem 9 *Let us consider the linear integral operator $A_1 : L^2(\epsilon_{min}, \infty) \rightarrow L^2(\epsilon_{min}, \epsilon_{max})$ defined in (138). For both the Kramers and the Bethe-Heitler integral kernels this operator is compact.*

Proof Let us consider the functions

$$E \mapsto \frac{1}{E^2} K^2(\epsilon, E) \tag{139}$$

and

$$\epsilon \mapsto \int_{\epsilon_{min}}^{\infty} \frac{1}{E^2} K^2(\epsilon, E) dE; \tag{140}$$

for the Fubini’s Theorem, if we show that the first one belongs to $L^1(\epsilon_{min}, \infty)$ and the second one belongs to $L^1(\epsilon_{min}, \epsilon_{max})$, we have that the function

$$(\epsilon, E) \mapsto \frac{1}{E} K(\epsilon, E) \tag{141}$$

is in $L^2([\epsilon_{min}, \epsilon_{max}] \times [\epsilon_{min}, \infty))$ and, therefore, that A_1 is Hilbert-Schmidt. But, in the Kramers case,

$$\int_{\epsilon_{min}}^{\infty} \frac{1}{E^2} K^2(\epsilon, E) dE \leq \frac{1}{\epsilon_{min}} \tag{142}$$

while, in the Bethe-Heitler case,

$$\begin{aligned} & \int_{\epsilon_{min}}^{\infty} \frac{1}{E^2} K^2(\epsilon, E) dE \\ &= \int_{\epsilon}^{\infty} \frac{1}{E^2} \left(\log \frac{1 + \sqrt{1 - \frac{\epsilon}{E}}}{1 - \sqrt{1 - \frac{\epsilon}{E}}} \right)^2 dE \\ &\leq \int_{\epsilon_{min}}^{\infty} \frac{1}{E^2} \left(\log \frac{1 + \sqrt{1 - \frac{\epsilon}{E}}}{1 - \sqrt{1 - \frac{\epsilon}{E}}} \right)^2 dE \\ &\leq \int_{\epsilon_{min}}^{\infty} \frac{1}{E^2} \left(\log \frac{1 + \sqrt{1 - \frac{\epsilon_{min}}{E}}}{1 - \sqrt{1 - \frac{\epsilon_{min}}{E}}} \right)^2 dE < \infty \end{aligned} \tag{143}$$

where we use the fact that the function

$$\epsilon \mapsto \left(\log \frac{1 + \sqrt{1 - \frac{\epsilon}{E}}}{1 - \sqrt{1 - \frac{\epsilon}{E}}} \right)^2$$

is a decreasing function of ϵ .

The proof is completed by observing that $[\epsilon_{min}, \epsilon_{max}]$ is compact. \square

The linear inverse problem

$$J = A_1 f \tag{144}$$

is ill-posed in the sense of Hadamard. In particular, the compactness of A_1 implies that the solution of the problem does not depend continuously on the data. From a practical viewpoint, the impact of ill-posedness on the inversion of real photon spectra is notable. In fact, any discretization of (144) must account for the numerical ill-conditioning consequence of the presence of the measurement noise on the observed spectra. It follows that at some stage of the inversion

process a regularization procedure must be introduced, in order to reduce the numerical instabilities and to provide physically reliable approximate solutions of the inverse problem.

3.3 The Inverse Problem with Discrete Data

During flare observations, X-ray detectors provide sets of numbers which are proportional to the integral of the photon spectrum over small ranges of photon energies. However, in the case of last generation devices such as RHESSI, it is realistic to assume that the difference between these counts and the point values of $g(\epsilon)$ at some energy value in the channel are negligible (for a quantitative discussion of this difference, see [63]). It follows that the equation we have to deal with when treating observed spectra is

$$g(\epsilon_n) = \int_{\epsilon_n}^{\infty} f(E) Q(\epsilon_n, E) dE, \quad n = 1, \dots, N, \tag{145}$$

where ϵ_n denotes the sampled photon energies for $n = 1, \dots, N$. In the case of the Kramers and Bethe-Heitler integral kernels, we want to address the study of this equation by maintaining its solution in an infinite dimensional space. In fact, we still assume that the source space is $L^2(\epsilon_{min}, \infty)$ and we choose as data space the finite dimensional vector space Y equipped with the weighted inner product

$$(\mathbf{g}, \mathbf{h})_Y = \sum_{m,n=1}^N g_m w_{mn} h_n. \tag{146}$$

Then we consider the finite-rank linear operator $A_2 : L^2(\epsilon_{min}, \infty) \rightarrow Y$ defined by

$$(A_2 f)_n = \int_{\epsilon_n}^{\infty} f(E) Q(\epsilon_n, E) dE, \quad n = 1, \dots, N \tag{147}$$

and the inverse problem we are interested in can be described by the equation

$$\mathbf{g} = A_2 f. \tag{148}$$

Equation (148) shows that the problem of the determination of the distribution function of electrons from the knowledge of the photon spectrum is a linear inverse problem with discrete data. In the cases of the Kramers and the Bethe-Heitler cross sections, we can follow the approach described in Sect. 2.3 and compute analytically the functions $\phi_n(E)$ defined in (27) and the entries of the Gram matrix given in (33). If we adopt the Kramers approximation, the explicit form of the functions ϕ_n is given by

$$\phi_n(E) = \begin{cases} 0, & E < \epsilon_n, \\ \frac{1}{\epsilon_n E}, & E \geq \epsilon_n \end{cases} \tag{149}$$

and a straightforward integration shows that the mn entry of the Gram matrix is

$$G_{mn} = \begin{cases} \frac{1}{\epsilon_m^2 \epsilon_n}, & m \geq n, \\ G_{nm}, & m < n. \end{cases} \tag{150}$$

If we adopt the Bethe-Heitler cross section, the explicit form of the functions ϕ_n is given by

$$\phi_n(E) = \begin{cases} 0, & E < \epsilon_n, \\ \frac{1}{\epsilon_n E} \log \frac{1 + \sqrt{1 - \frac{\epsilon_n}{E}}}{1 - \sqrt{1 - \frac{\epsilon_n}{E}}}, & E \geq \epsilon_n \end{cases} \tag{151}$$

and the mn entry of the Gram matrix is

- for $m > n$:

$$G_{mn} = \frac{4}{(\epsilon_n \epsilon_m)^{\frac{3}{2}} \log \frac{1 + \sqrt{\frac{\epsilon_n}{\epsilon_m}}}{1 - \sqrt{\frac{\epsilon_n}{\epsilon_m}}} - \frac{2}{\epsilon_n \epsilon_m^2} \log \frac{\epsilon_n}{\epsilon_m} + \frac{2}{\epsilon_n \epsilon_m^2} \left(1 + \frac{\epsilon_m}{\epsilon_n}\right) \log \left(1 - \frac{\epsilon_n}{\epsilon_m}\right)}; \tag{152}$$

- for $m = n$:

$$G_{mm} = \frac{8 \log 2}{\epsilon_m^3}; \tag{153}$$

- for $m < n$:

$$G_{mn} = G_{nm}. \tag{154}$$

In the case of the relativistic cross section (110), the explicit form of the functions $\phi_n(E)$ is extremely complicated and the entries of the Gram matrix can be computed only through some numerical integration.

3.4 The Relativistic Cross Section

Owing to computational reasons, in the case of the highly relativistic cross section $Q_{3BN}(\epsilon, E)$ we have preferred not to maintain the solution in an infinite dimension space but to fully discretize the integral equation (106) and to study the resulting rectangular linear system

$$\mathbf{g} = \mathbf{A}\mathbf{f}, \tag{155}$$

where \mathbf{g} is the data vector with components

$$(\mathbf{g})_n = g(\epsilon_n), \quad n = 1, \dots, N, \tag{156}$$

\mathbf{f} is the solution vector with components

$$(\mathbf{f})_m = f(E_m), \quad m = 1, \dots, M \tag{157}$$

and \mathbf{A} is the rectangular matrix with entries

$$A_{nm} = Q(\epsilon_n, E_m)\eta_{nm}, \tag{158}$$

η_{nm} being the quadrature coefficients (which depend on the kind of sampling used). Here, for reasons related to the physics of the problem and, in particular, to the characteristics of the acquisition procedure followed by RHESSI, we will always have $M > N$. The determination of the generalized solution of the linear system (155), i.e. the vector solving the constrained least-squares problem

$$\|\mathbf{g} - \mathbf{A}\mathbf{f}\| = \min, \quad \|\mathbf{f}\| = \min, \tag{159}$$

is given by

$$\mathbf{f}^\dagger = \sum_{k=1}^N \frac{1}{\sigma_k} (\mathbf{g}, \mathbf{v}_k) \mathbf{u}_k, \tag{160}$$

where $\{\sigma_k; \mathbf{v}_k, \mathbf{u}_k\}_{k=1}^N$ is the singular system of the matrix \mathbf{A} defined by means of

$$\mathbf{A}\mathbf{u}_k = \sigma_k \mathbf{v}_k, \quad \mathbf{A}^t \mathbf{v}_k = \sigma_k \mathbf{u}_k. \tag{161}$$

Analogously, the Tikhonov regularized solution of (155) is given by

$$\mathbf{f}_\lambda = \sum_{k=1}^N \frac{\sigma_k}{\sigma_k^2 + \lambda} (\mathbf{g}, \mathbf{v}_k) \mathbf{u}_k \tag{162}$$

and the optimal value of the regularization parameter can be fixed by applying the discrepancy principle in its obvious discrete formulation.

4 Reconstruction of the Mean Electron Spectrum

The aim of this section is two-fold. From one side we want to study the effectiveness of the Tikhonov regularization method for solving the general bremsstrahlung equation, by using simulated spectra. On the other hand, we want to describe the sensitivity of this inversion approach to the use of different integration kernels, i.e. for different choices of the bremsstrahlung cross section. In particular, our discussion will be organized into two sections. In the first one we will study the conditioning of the problem by means of the Singular Value Decomposition of the integral operator. In the second one we will use realistic synthetic data to test the effectiveness of the Tikhonov method and of the optimal criterion for choosing the regularization parameter. In the last section, we will give an example of application of the Tikhonov method to real X-ray data recorded by RHESSI.

4.1 Condition Number for the Different Cross Sections

A critical point in the numerical solution of integral equations is the dependence of the numerical instability degree

on the shape of the integral kernel. In the case of the bremsstrahlung equation, such an issue assumes a particularly relevant role, owing to the notable physical meaning of the approximations leading to the different forms of the bremsstrahlung cross section. The key question, here, is to assess if and under which physical conditions, the use of simplified cross sections may increase the ill-conditioning of the inverse problem and, therefore, the reliability of the corresponding regularized solutions. From a computational viewpoint, the use of the singular system of the finite rank operator (147), in the case of the Kramers and Bethe-Heitler cross sections, and of the rectangular matrix (158) in the case of the relativistic cross section, allows to easily determine the condition number

$$C = \frac{\sigma_1}{\sigma_N} \quad (163)$$

for different values of the number of sampling points and for different values of the photon sampled energies.

Let us consider first the computation of the singular system for a fixed range $[\epsilon_{min}, \epsilon_{max}]$ of photon energies and a fixed number N of sampling points. For this application we chose $\epsilon_{min} = 10$ keV, $\epsilon_{max} = 99$ keV, $N = 90$ in the case of a uniform sampling with sampling distance equal to 1 keV (these experimental conditions are compatible with the typical parameters used during the pre-processing of RHESSI observations). Table 1 contains the first ten singular values for the three cross sections while Fig. 2 shows the first four singular functions u_k for the three cross sections. Finally, Table 2 contains the condition numbers corresponding to the three kernels with $\epsilon_{min} = 10$ keV and three different values of the maximum photon energy. Under the assumption that the cross section (110) represents the exact emission probability in the bremsstrahlung process (or, at least, its most accurate approximation), these results point out that:

- when the Kramers form is used, i.e., when all the relativistic effects are neglected, the resulting singular system is significantly different than the exact one. In particular, the Kramers singular spectrum runs first below and then above the exact one with relative differences of even more than 50% for the largest and smallest singular values.
- The singular functions corresponding to the Kramers cross section systematically reaches larger maximum values than the exact ones and, what is more important, their zeros occur at smaller energy values. Since the regularized solution can be expanded on the basis given by the singular functions or vectors (see (83) and (162)), this implies that the spectral resolution achievable when the Kramers approximation is assumed for the inversion deteriorates with energy more rapidly than for the other two cross sections.
- The semi-relativistic Bethe-Heitler approximation provides a singular spectrum which is close to the exact one

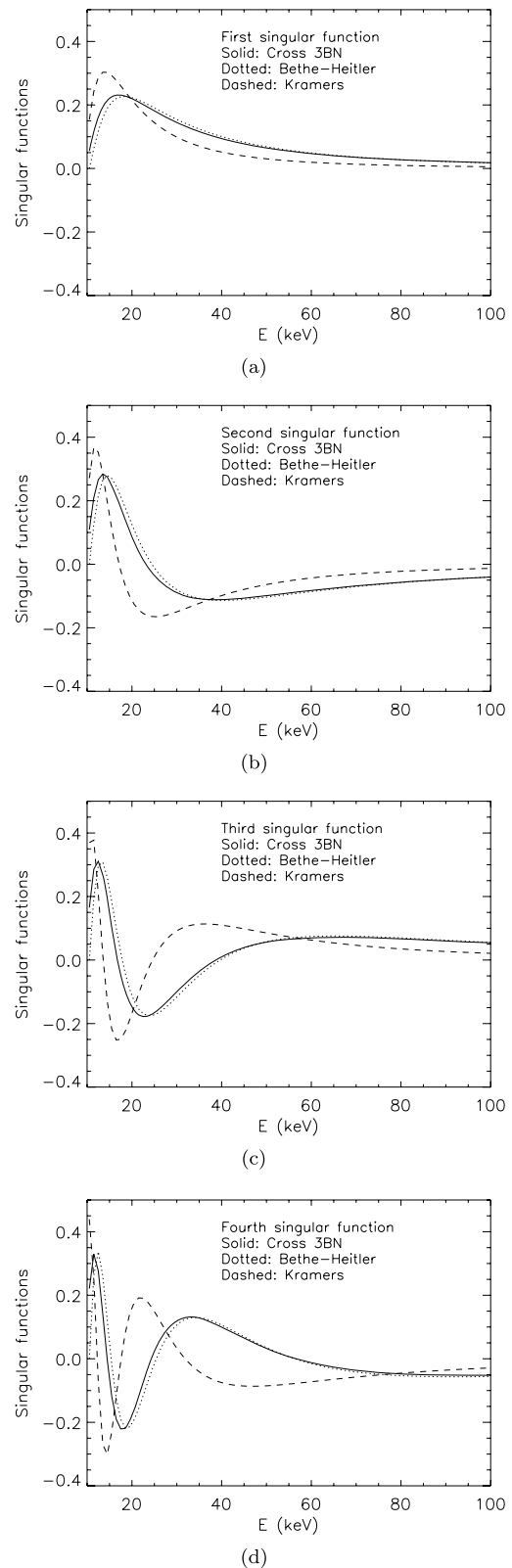


Fig. 2 Singular functions of the finite-rank operator A_2 for the Kramers and Bethe-Heitler integral kernels and of the matrix \mathbf{A} in the case of the relativistic cross section: (a) first singular function; (b) second singular function; (c) third singular function; (d) fourth singular function

Table 1 Singular values of the finite-rank operator A_2 for the Kramers and Bethe-Heitler integral kernels and of the matrix \mathbf{A} in the case of the fully relativistic cross section. The experimental conditions are: $\epsilon_{min} = 10$ keV, $\epsilon_{max} = 99$ keV, $N = 90$, uniform sampling with 1 keV sampling distance

	Kramers	Bethe-Heitler	Cross 3BN
σ_1	0.69675061	1.4327710	1.6616698
σ_2	0.23456303	0.3009761	0.3251901
σ_3	0.14120844	0.1354425	0.1442251
σ_4	0.10133624	0.0817600	0.0865855
σ_5	0.07936335	0.0561889	0.0589741
σ_6	0.06555440	0.0418901	0.0434190
σ_7	0.05611209	0.0329645	0.0335835
σ_8	0.04916533	0.0270082	0.0269361
σ_9	0.04366882	0.0227958	0.0221901
σ_{10}	0.03908788	0.0196456	0.0186748

Table 2 Condition numbers for the linear inverse problem with discrete data when the cross sections are given by the Kramers and Bethe-Heitler formulas and for the fully discretized problem in the case of the relativistic cross section. The minimum sampled photon energy is $\epsilon_{min} = 10$ keV in all cases, while three different values of the maximum sampled photon energy are considered

	Kramers	Bethe-Heitler	Cross 3BN
$\epsilon_{max} = 99$	764.32	9662.26	17572.21
$\epsilon_{max} = 149$	1424.45	28069.19	54625.84
$\epsilon_{max} = 199$	3043.64	60945.02	117049.32

(the relative error is smaller than 15% for all the singular values). The singular functions are very similar too.

- As one may expect, the conditioning of the problem worsens when the sampling range increases. Anyway, the inverse problem corresponding to the Kramers kernel is always better conditioned.

4.2 Reconstructions: Synthetic Data

The inversion of simulated photon spectra allows to validate two important aspects of the regularization approach: the robustness of data reduction to modifications of the integral kernel and the effectiveness of the discrepancy principle for the optimal choice of the regularization parameter. In order to study such issues we have generated simulated data corresponding to three different experimental situations.

Power-law: small solar flares typically generate photon spectra characterized by a monotonically steeply decreasing behavior approximated by

$$g(\epsilon) \sim \epsilon^{-\gamma} \tag{164}$$

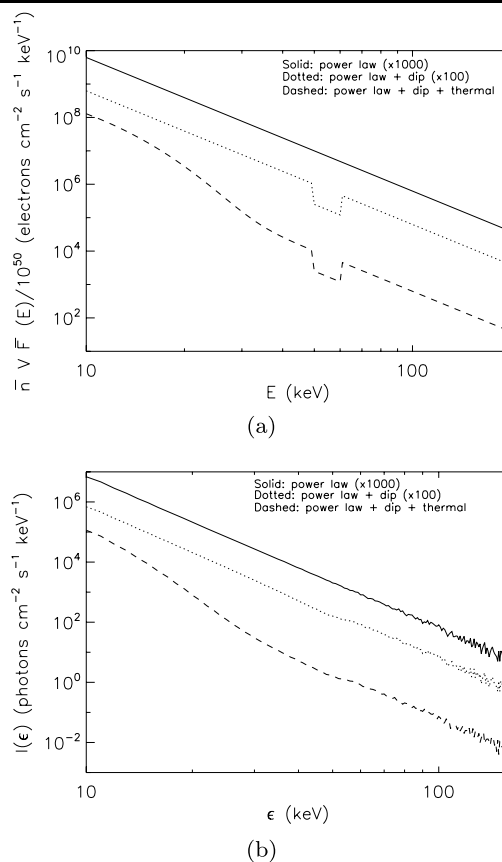


Fig. 3 Synthetic data: (a) the three theoretical electron spectra; (b) the three corresponding photon spectra. In both figures, in order to distinguish the ‘power-law’ case (solid) and the ‘power-law plus dip’ case (dotted) from the ‘power-law plus dip plus thermal component’ case (dashed), a $\times 1000$ and a $\times 100$ artificial scale factors have been used respectively

with, typically, $\gamma \geq 3$. Simple asymptotic considerations on (106) show that this kind of photon spectrum is generated by an averaged electron spectrum of the kind

$$f(E) \sim E^{-\delta} \tag{165}$$

with $\delta \sim \gamma - 1$.

In our simulations we have assumed

$$f(E) = 10^4 \left(\frac{E}{50} \right)^{-\delta}, \tag{166}$$

where the numerical constants have been introduced for physical reasons.

Dip: in [67] a rather surprising small wavelength structure has been reconstructed in the mean electron spectrum corresponding to a photon spectrum emitted during the July 23, 2002 flare. The nature and origin of such an intermediate energy dip are still under investigation. Here we mimic

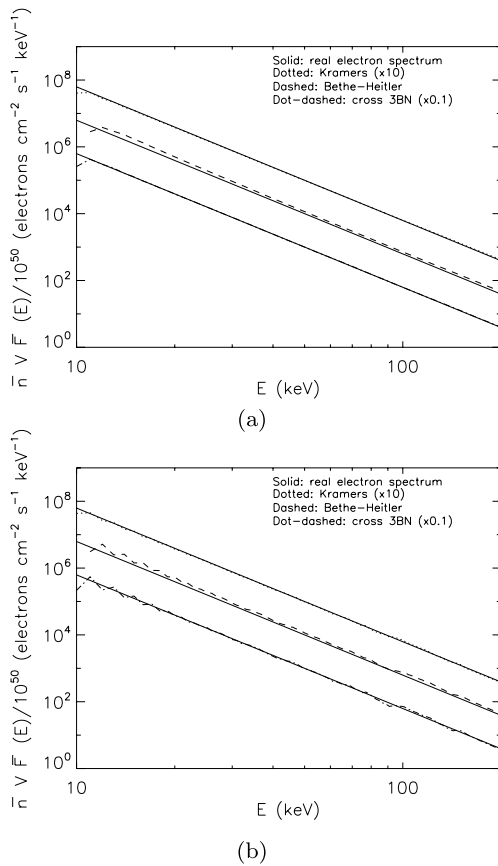


Fig. 4 Reconstruction of three different forms of the averaged electron spectrum for the three cross sections considered: **(a)** the power law when λ is chosen by minimizing the distance between the theoretical and regularized solutions; **(b)** as in **(a)**, with λ chosen by the discrepancy principle. Artificial scale factors have been used to better distinguish the three cases

its presence by means of the analytical form

$$f(E) = \begin{cases} 0.25 \cdot 10^4 \left(\frac{E}{50}\right)^{-\delta}, & 50 \leq E \leq 60, \\ 10^4 \left(\frac{E}{50}\right)^{-\delta}, & E < 50, E > 60. \end{cases} \quad (167)$$

Thermal component: particularly in the case of intense flares, part of the low energy X-ray emission is not due to the injection of accelerated electrons into a cold plasma but directly comes from a thermal emission corresponding to peak temperatures of some million degrees. In the analytical form

$$f(E) = \begin{cases} 0.25 \cdot 10^4 \left(\frac{E}{50}\right)^{-\delta}, & 50 \leq E \leq 60 \\ 10^4 \left[\left(\frac{E}{50}\right)^{-\delta} + 5 \cdot 10^5 e^{-E/2.7}\right], & \text{otherwise,} \end{cases} \quad (168)$$

the thermal component is represented by the negative exponential function at $E < 50$ keV.

In all simulations, we have assumed the realistic value $\delta = 4$. The photon spectra have been computed by inserting the

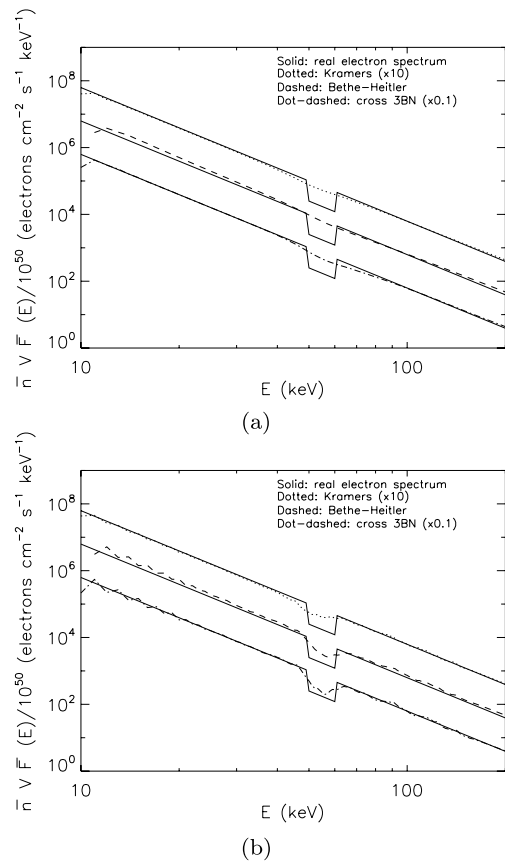


Fig. 5 Reconstruction of three different forms of the averaged electron spectrum for the three cross sections considered: **(a)** the power law plus an intermediate energy dip when λ is chosen by minimizing the distance between the theoretical and regularized solutions; **(b)** as in **(a)**, with λ chosen by the discrepancy principle. Artificial scale factors have been used to better distinguish the three cases

functions (166), (167) and (168) into the bremsstrahlung equation by using the relativistic formula Q_{3BN} as integral kernel. The photon data have been sampled from $\epsilon_{min} = 10$ keV to $\epsilon_{max} = 99$ keV with a uniform sampling of 1 keV bin. Finally, realistic Poisson noise has been added to the photon data vectors. The three theoretical input mean electron spectra and the corresponding photon spectra are plotted in Fig. 3. This plot clearly shows that, due to the blurring action played by the integral kernel, slightly distinguishable differences in the output data correspond to notably different input functions. The regularized solutions have been computed by means of (83) for the Kramers and Bethe-Heitler cases and of (162) for the highly relativistic cross section. The regularization parameter has been fixed according to two different approaches: 1) by minimizing the Euclidean distance between the regularized solution and the theoretical input function (criterion I); 2) by applying the discrepancy principle (criterion II). The results of this computation are described in Figs. 4–6 and Table 3. Figures 4–6 contain the reconstructions of the three theoretical averaged electron

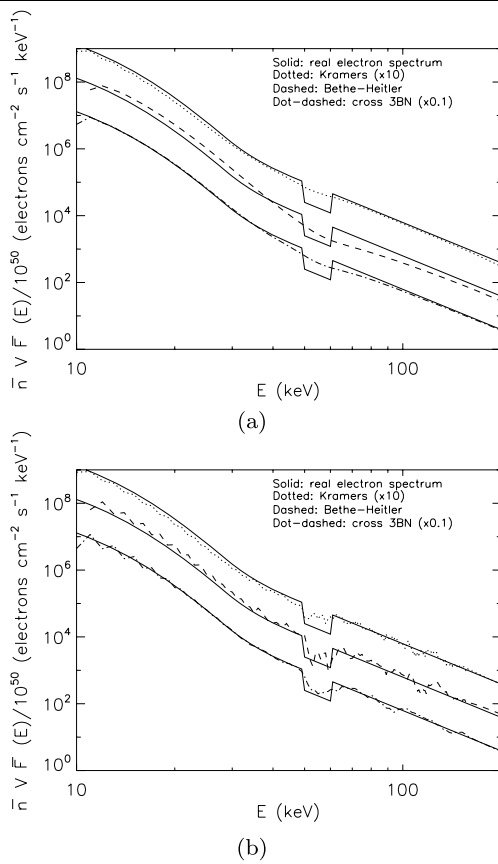


Fig. 6 Reconstruction of three different forms of the averaged electron spectrum for the three cross sections considered: (a) the power law plus dip plus a low energy thermal component, when λ is chosen by minimizing the distance between the theoretical and regularized solutions; (b) as in (a), with λ chosen by the discrepancy principle. Artificial scale factors have been used to better distinguish the three cases

spectra provided by the Tikhonov regularization method for the three cross sections; in the top panels the regularization parameter has been fixed by minimizing the distance between the theoretical and regularized solutions while in the bottom panels the discrepancy principle has been applied. Table 3 contains the reconstruction errors defined as

$$\rho = \frac{\|f - f_\lambda\|_2}{\|f\|_2}, \tag{169}$$

where $\|\cdot\|_2$ is the L^2 norm in the Kramers and Bethe-Heitler cases and the canonical Euclidean norm in the highly relativistic case, while λ is fixed with the two criteria. These results point out that:

- in general Tikhonov regularization provides satisfactory reconstructions of the theoretical mean electron spectra. The regularized solutions are sufficiently stable and the overall behavior is always correctly reproduced.
- The reconstructions given by criterion I are more stable but cannot reproduce the intermediate dip. This is due to

Table 3 Reconstruction errors provided by Tikhonov regularization while reproducing three different forms of the averaged electron spectrum for the three cross sections adopted. The regularization parameter is fixed by determining the minimum distance to the theoretical form (criterion I) and by applying the discrepancy principle (criterion II)

	P.I.		P.I.+dip		P.I.+dip+thermal	
	I	II	I	II	I	II
K	0.045	0.082	0.045	0.081	0.136	0.144
BH	0.328	0.439	0.330	0.319	0.344	0.419
3BN	0.019	0.257	0.019	0.257	0.021	0.291

the fact that in minimizing the overall error a more significant role is played by the low energy part of the spectrum, where the actual values of the function are significantly bigger. The use of the discrepancy principle focuses on the residuals and provides smaller λ : therefore the reconstructions are more unstable (see, in particular, the low energy parts) but more reliable in reproducing small features.

- If λ is chosen by means of criterion I, the inversion with the highly relativistic formula is the most reliable (as one may expect, since the data have been simulated by using just that cross section). If λ is chosen with criterion II, the minimum restoration error is obtained by using Kramers approximation. This is again reasonable since the discrepancy principle provides smaller λ and, as pointed out in the previous section, the conditioning is smaller when the Kramers kernel is adopted.

4.3 Reconstructions: Application to RHESSI Data

As an example of application of the Tikhonov method to real X-ray data recorded by RHESSI, we consider the photon spectrum corresponding to emission peak during the August 21, 2002 flare. In the restoration of real data two further issues must be accounted for, both concerning the physical interpretation of the regularized reconstructions. The first item is the determination of the uncertainty on the reconstruction. This assessment can be accomplished by producing the so-called confidence strip of the regularized solution [3]. Such a construction is performed by generating different random realizations of the photon data, produced by modifying each X-ray vector component according to a Gaussian distribution of zero mean and standard deviation equal to the experimental one. The regularized solutions for all these realizations are determined and superimposed in order to evaluate the robustness of the regularization algorithm to data instability. The height of the confidence strip at each value of the electron energy provides the propagation error on the regularized solution.

The second item is concerned with the assessment of the energy resolution achievable by the method. Our approach

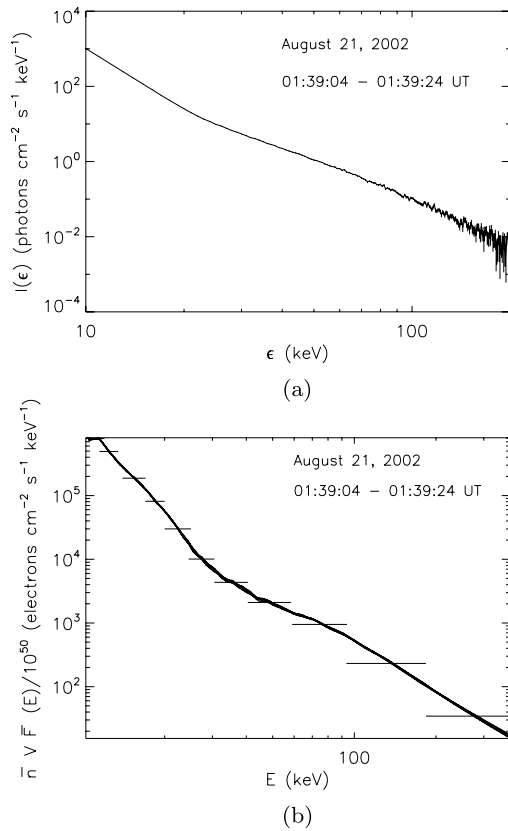


Fig. 7 Inversion of a real spectrum emitted during the August 21, 2002 flare: **(a)** the photon spectrum at the emission peak; **(b)** the reconstructed mean electron spectrum provided by the Tikhonov method. The regularization parameter has been fixed by means of the discrepancy principle

follows the method introduced in [4] and is based on the observation that in the expansions (83) and (162) not all the singular components actually contribute to the form of the regularized solution with the same significance. In fact, for increasing k values, the contribution of the corresponding singular functions is increasingly filtered out by the coefficient $\frac{\sigma_k}{\sigma_k + \lambda}$. We use the truncation rule

$$\sigma_k \geq \sqrt{\lambda} \quad (170)$$

in these expansions and observe that, since the singular values are decreasingly ordered, there will be an index \bar{k} for which condition (170) is satisfied for the last time. Therefore a reasonable estimate of the resolution achievable by the method is given by the distance between the successive zeros of $u_{\bar{k}}$.

Figure 7 contains the result of the application of Tikhonov regularization to the inversion of the real spectrum recorded during the August 21, 2002 flare. The photon spectrum corresponding to the emission peak is given in Fig. 7(a) while Fig. 7(b) presents the corresponding reconstruction provided by the Tikhonov method. In this plot the confi-

dence strip is represented together with the horizontal resolution bars. In this reconstructed spectrum the thermal component at low energies is fairly visible, together with a spectral ‘knee’ at energies in the range 20–40 keV. The asymptotic electron spectral index (i.e., the exponent of the power law best fitting the spectrum at high energies) is $\delta \simeq 2.45$ while the photon spectral index is $\gamma \simeq 3.54$, and this is in accordance with the ‘rule of thumb’ $\gamma \simeq \delta + 1$ based on asymptotic approximations.

5 Anisotropic Bremsstrahlung Emission

5.1 Angular Dependency of the Bremsstrahlung Equation

In the previous sections we showed how to apply regularization techniques to the inversion of high resolution X-ray spectra from solar flares with the aim of inferring information on the electron flux spectrum in the source. In our analysis we considered particular forms [37] for the bremsstrahlung cross section $Q(\epsilon, E)$ which depend only on the photon energy ϵ and the electron energy E . However, the correct cross section to use must in practice take into account two important aspects of the geometrical and physical environment. First, the direction of the precollision electron is not, in general, univocally definite, but there will be a significant spread in the incoming directions of the electrons (e.g., [45, 46]). Second, the guiding magnetic field may be inclined away from the vertical toward the observer, and a similar inclination in guiding field direction angle may be appropriate for the bremsstrahlung-producing electron beam [75]. It follows that the probabilistic kernel which describes the bremsstrahlung phenomenon is, in general, a function also of the incoming and outgoing electron directions and of the polarization state³ of the emitted photon [33]. In principle, also components due to both electron-ion and electron-electron bremsstrahlung have to be considered, but it can be shown that the latter is quite negligible except at mildly or extremely relativistic energies [50]. Several authors (e.g., [13, 25, 26, 36, 38, 53, 54]) showed how the X-ray emission from a particular electron source changes significantly for different viewing directions. In this section we consider the expression for the electron-photon bremsstrahlung cross section, integrated over the direction of the outgoing electron and summed over the polarization states of the emitted photon, provided by Gluckstern & Hull [32]. This kernel $Q(\epsilon, E; \theta)$ is a function of three variables: besides the energies E of the electron (keV) and ϵ of the photon (keV), also the dependency on the angle θ between the directions

³Because propagating light consists of a transverse electric and magnetic field, a single photon will oscillate on a line perpendicular to the propagation direction.

of the incoming electron and the emitted photon is kept in consideration.

Let us generalize now the concept of mean source electron flux spectrum given in Sect. 3.1 [12, 17] to the case of an anisotropic cross section and/or source electron distribution function. The hard X-ray intensity $I(\epsilon)$ observed at distance R from a source can be written as

$$I(\epsilon) = \frac{1}{4\pi R^2} \int_{\epsilon}^{\infty} \int_{\Omega} \int_V Q(\epsilon, E; \theta) \times F(E, \mathbf{r}, \Omega) n(\mathbf{r}) d\mathbf{r} d\Omega dE, \tag{171}$$

where $F(E, \mathbf{r}, \Omega)$ is the electron flux differential in electron energy E , position \mathbf{r} , and solid angle of the incoming electron direction Ω , and $Q(\epsilon, E; \Omega)$ is the cross section for bremsstrahlung emission in the direction of the observer. If we define

$$\widehat{F}(E, \Omega) = \frac{\int_V F(E, \mathbf{r}, \Omega) n(\mathbf{r}) d\mathbf{r}}{\int_V n(\mathbf{r}) d\mathbf{r}} = \frac{\int_V F(E, \mathbf{r}, \Omega) n(\mathbf{r}) d\mathbf{r}}{\bar{n}V}, \tag{172}$$

where we recall that $\bar{n} = (1/V) \int_V n(\mathbf{r}) d\mathbf{r}$, then (171) can be written as

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \int_{\Omega} Q(\epsilon, E; \theta) \widehat{F}(E, \Omega) d\Omega dE. \tag{173}$$

If $\widehat{F}(E, \Omega)$ is isotropic, then we define $\overline{F}(E) \equiv \widehat{F}(E)$ and let

$$\overline{Q}(\epsilon, E) = \int_{\Omega} Q(\epsilon, E; \theta) d\Omega. \tag{174}$$

Then we can write

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \overline{F}(E) \overline{Q}(\epsilon, E) dE. \tag{175}$$

This is the expression we used to define the ‘‘mean electron spectrum’’ $\overline{F}(E)$ in (101) and (102). It should be noted that (175) also applies in the (somewhat hypothetical) case where $Q(\epsilon, E; \theta)$ is independent of θ , with $\overline{Q} \equiv Q$ and

$$\overline{F}(E) = \int_{\Omega} \widehat{F}(E, \Omega) d\Omega. \tag{176}$$

In the physically realistic case, neither $Q(\epsilon, E; \theta)$ nor $\widehat{F}(E, \Omega)$ is isotropic. To make progress, therefore, requires some further assumptions on the form of $\widehat{F}(E, \Omega)$. In this section we restrict ourselves to the simplest assumption, namely that $\widehat{F}(E, \Omega)$ is separable in E and Ω , i.e.,

$$\widehat{F}(E, \Omega) = \overline{F}(E) \frac{h(\Omega)}{\int_{\Omega} h(\Omega) d\Omega}. \tag{177}$$

With such an assumption, (173) can be written as

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \overline{F}(E) dE \left[\frac{\int_{\Omega} Q(\epsilon, E; \Omega) h(\Omega) d\Omega}{\int_{\Omega} h(\Omega) d\Omega} \right]. \tag{178}$$

If we define

$$\overline{Q}(\epsilon, E) = \frac{\int_{\Omega} Q(\epsilon, E; \Omega) h(\Omega) d\Omega}{\int_{\Omega} h(\Omega) d\Omega}, \tag{179}$$

then (178) is formally identical to (175) and can be solved for any adopted form of $h(\Omega)$ once $\overline{Q}(\epsilon, E)$ is evaluated using (179). We remark that in practice, e.g., in a collisional thick target, the E and Ω dependencies of $\widehat{F}(E, \Omega)$ are *not* separable and further vary along the electron paths. To deal with that situation more properly requires explicit modeling of the electron propagation, i.e., of electron scattering and energy losses; in general, this can probably be done only by forward modeling. Nevertheless, the results of our separable inversion formulation will provide a better starting point than the assumption of isotropy used hitherto.

5.2 Anisotropic Form of the Cross Section

The form of the angle-dependent cross section $Q(\epsilon, E; \theta)$ has been given by Gluckstern & Hull [32] and Koch & Motz ([44], formula 2BN). Here we reprint this result in our notation (and in units of $\text{cm}^{-2} \text{keV}^{-1} \text{sr}^{-1}$),⁴ and we also present a polar diagram of the angular dependence for various electron and photon energies in order to make some of the discussion in the following sections more comprehensible.

Formally, the cross section $Q(\epsilon, E; \theta)$ for electron-ion bremsstrahlung, differential in photon energy ϵ , electron energy E , and the angle between the incoming electron and the emitted photon (but integrated over the direction of the emergent electron and summed over the polarization states of the emitted photon) is

$$Q(\epsilon, E; \theta) = Z^2 \frac{\alpha}{2} \frac{r_0^2}{m_e c^2} \left(\frac{1}{\tilde{\epsilon}} \right) \frac{\sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}}{\sqrt{\tilde{E}^2 - 1}} \times \left\{ 8 \frac{2\tilde{E}^2 + 1}{(\tilde{E}^2 - 1)\Delta^4} \sin^2 \theta - 2 \frac{5\tilde{E}^2 + 2\tilde{E}(\tilde{E} - \tilde{\epsilon}) + 3}{(\tilde{E}^2 - 1)\Delta^2} - 2 \frac{\tilde{E}^2 - \tilde{\epsilon}^2 - 1}{T^2 \Delta^2} + 4 \frac{\tilde{E} - \tilde{\epsilon}}{(\tilde{E}^2 - 1)\Delta} \right.$$

⁴The steradian (sr) is defined as the solid angle subtended at the center of a sphere of radius r by a portion of the surface of the sphere having an area r^2 . Since the surface area of this sphere is $4\pi r^2$, then the definition implies that a sphere measures 4π steradians.

$$\begin{aligned}
 & + \frac{L}{\sqrt{((\tilde{E} - \tilde{\epsilon})^2 - 1)(\tilde{E}^2 - 1)}} \\
 & \times \left[\frac{4\tilde{E}(3\tilde{\epsilon} - (\tilde{E}^2 - 1)(\tilde{E} - \tilde{\epsilon}))}{(\tilde{E}^2 - 1)\Delta^4} \sin^2 \theta \right. \\
 & + \frac{4\tilde{E}^2(\tilde{E}^2 + (\tilde{E} - \tilde{\epsilon})^2)}{(\tilde{E}^2 - 1)\Delta^2} \\
 & + \frac{-2(7\tilde{E}^2 - 3\tilde{E}(\tilde{E} - \tilde{\epsilon}) + (\tilde{E} - \tilde{\epsilon})^2) + 2}{(\tilde{E}^2 - 1)\Delta^2} \\
 & \left. + 2\tilde{\epsilon} \frac{\tilde{E}^2 + \tilde{E}(\tilde{E} - \tilde{\epsilon}) - 1}{(\tilde{E}^2 - 1)\Delta} \right] + \frac{\gamma_T}{T\sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}} \\
 & \times \left[\frac{4}{\Delta^2} - \frac{6\tilde{\epsilon}}{\Delta} - \frac{2\tilde{\epsilon}(\tilde{E}^2 - \tilde{\epsilon}^2 - 1)}{T^2\Delta} \right] \\
 & - \left. \frac{4\gamma}{\Delta\sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}} \right\} \\
 & \times \frac{\beta(1 - e^{-2\pi Z\alpha/\beta})}{\beta'(1 - e^{-2\pi Z\alpha/\beta'})},
 \end{aligned}$$

where Z is the atomic number of the ion, $\alpha \simeq 1/137$ is the fine-structure constant, $r_0 \simeq 2.8 \times 10^{-13}$ cm is the classical electron radius, $m_e c^2 \simeq 511$ keV is the electron rest energy, and

$$\begin{aligned}
 \tilde{E} &= 1 + \frac{E}{m_e c^2}, & \tilde{\epsilon} &= \frac{\epsilon}{m_e c^2}, \\
 \beta &= \sqrt{1 - \frac{1}{\tilde{E}^2}}, & \beta' &= \sqrt{1 - \frac{1}{(\tilde{E} - \tilde{\epsilon})^2}}, \\
 \Delta &= \tilde{E} - \sqrt{\tilde{E}^2 - 1} \cos \theta, \\
 T &= \left(\tilde{E}^2 - 1 + \tilde{\epsilon}^2 - 2\tilde{\epsilon}\sqrt{\tilde{E}^2 - 1} \cos \theta \right)^{1/2}, \\
 L &= \ln \left\{ \frac{\tilde{E}(\tilde{E} - \tilde{\epsilon}) - 1 + \sqrt{(\tilde{E}^2 - 1)[(\tilde{E} - \tilde{\epsilon})^2 - 1]}}{\tilde{E}(\tilde{E} - \tilde{\epsilon}) - 1 - \sqrt{(\tilde{E}^2 - 1)[(\tilde{E} - \tilde{\epsilon})^2 - 1]}} \right\}, \\
 \gamma_T &= \ln \left[\frac{T + \sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}}{T - \sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}} \right], \\
 \gamma &= \ln \left[\frac{(\tilde{E} - \tilde{\epsilon}) + \sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}}{(\tilde{E} - \tilde{\epsilon}) - \sqrt{(\tilde{E} - \tilde{\epsilon})^2 - 1}} \right].
 \end{aligned}$$

The last factor in the definition of $Q(\epsilon, E; \theta)$ —involving the velocity β (in units of the speed of light c)—is the Elwert [24] Coulomb correction and does not appear in the

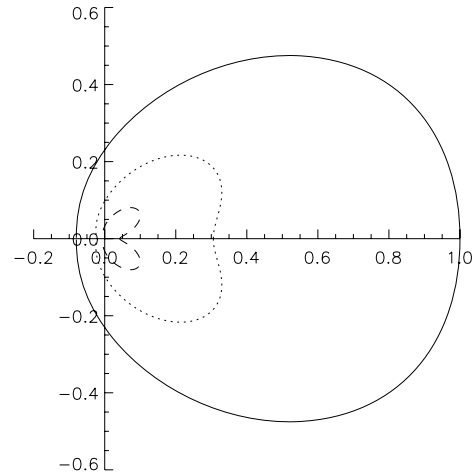


Fig. 8 Polar diagram of the bremsstrahlung cross section for $E = 100$ keV and photon energies $\epsilon = 30$ keV (solid line), $\epsilon = 50$ keV (dotted line), and $\epsilon = 80$ keV (dashed line); the radial coordinate is proportional to the size of the cross section and the angle from the x -axis corresponds to the angle between the incoming electron direction and the line to the observer. Note that at energies $E \gg \epsilon$ the cross section peaks at $\theta = 0^\circ$, while for $E \simeq \epsilon$ the cross section peaks at $\theta \simeq 30^\circ\text{--}40^\circ$

expressions in [32]. This correction is sufficiently accurate (to within a few percent) except at electron energies above ~ 100 keV and approaching the “high-frequency limit” $\epsilon \rightarrow E$; in such regimes more elaborate expressions are appropriate. For more details, see [44]. It should be noted that the expression of $Q(\epsilon, E; \theta)$ is indeterminate at $\epsilon = E$; in practice this can be handled in numerical computation by setting E slightly higher than ϵ .

Figure 8 shows the angular dependency of such a cross section. As we can see, for electron energies $E \gg \epsilon$, the form of $Q(\epsilon, E; \theta)$ is a decreasing function of θ ; there is a preference for photons to be emitted in the direction of the incoming electron. However, at electron energies comparable to the photon energy there must be a substantial scattering angle between the incoming and outgoing electrons, and hence the photons tend to be emitted preferentially at a significant angle relative to the incoming electron velocity. Hence, at electron energies E slightly greater than ϵ , the cross section peaks not in the forward direction but rather at a modest angle ($30^\circ\text{--}40^\circ$ —see Fig. 8).

5.3 Recovering the Electron Spectrum as a Function of Viewing- Angle

In order to analyse the effects of the anisotropic cross section with respect to the results obtained in Sect. 3 with several angle-averaged kernels, we keep on considering the photon spectrum corresponding to emission peak during the August 21, 2002 flare (represented in Fig. 7(a)). The location of the flare on the solar disk is $x = 696''$, $y = -248''$ (with respect to a fixed heliocentric coordinate system whose origin

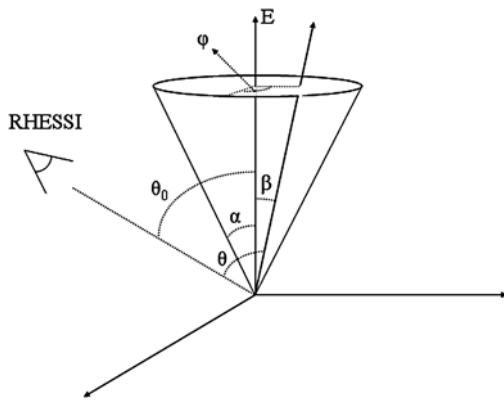


Fig. 9 The uniform distribution of the target-averaged incoming electron velocities. The z-axis (labeled with the electron energy E) represents the mean direction of the electron

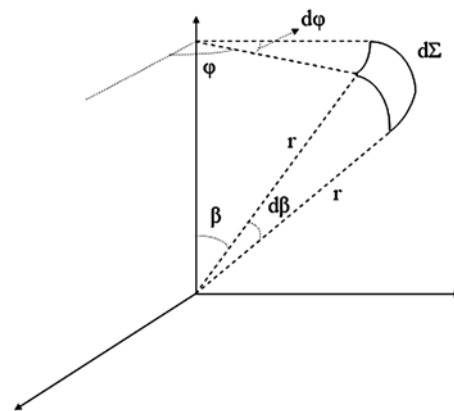


Fig. 10 The infinitesimal solid angle $d\Omega$

is placed in the center of the Sun), which corresponds to a heliocentric angle of $\sim 50^\circ$.

We assume that, for each energy, all the possible directions of the precollision electrons are uniformly distributed over a solid angle within a cone of half-angle α centered on a direction that makes an angle θ_0 relative to the direction of photon emission, i.e. (with β as the polar angle relative to the axis of the cone—see Fig. 9),

$$h(\Omega) = \begin{cases} 1, & \beta \leq \alpha, \\ 0, & \text{otherwise.} \end{cases} \quad (180)$$

With this form of $h(\Omega)$ the average cross section $\overline{Q}(\epsilon, E)$ defined in (179) is

$$\overline{Q}(\epsilon, E; \theta_0, \alpha) = \frac{1}{2\pi(1 - \cos \alpha)} \int_0^{2\pi} \int_0^\alpha Q(\epsilon, E; \theta) \sin \beta d\beta d\phi, \quad (181)$$

where θ is the angle between the observer and an elementary electron beam oriented at polar coordinates (β, ϕ) relative to the axis of the cone, viz.,

$$\cos \theta = \cos \theta_0 \cos \beta + \sin \theta_0 \sin \beta \cos \phi. \quad (182)$$

We observe that in (181) we used the fact that

$$d\Omega = \frac{d\Sigma}{r^2} = \frac{(r \sin \beta d\phi)(r d\beta)}{r^2} = \sin \beta d\beta d\phi \quad (183)$$

as we can see from Fig. 10.

Expression (181) for $\overline{Q}(\epsilon, E; \theta_0, \alpha)$ has been used to invert (178) (with the atomic number Z set equal to 1.2) and regularized electron spectra $\overline{F}_{\theta_0, \alpha}(E)$ have been recovered for mean angles θ_0 over the range from 0° (photon emission parallel to the direction of the incoming electron) to 180° (photon emission in the anti-parallel direction). Several values for the spread angle α ($10^\circ, 30^\circ, 60^\circ, 90^\circ$, and

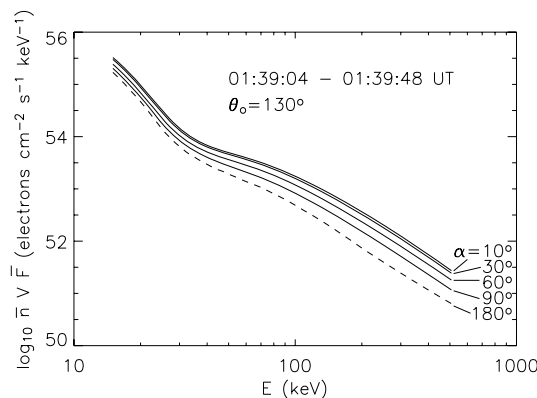


Fig. 11 $\overline{F}_{\theta_0, \alpha}(E)$ for $\theta = 130^\circ$ and the values of α (spread in incoming electron directions) shown. The dashed curve (labeled $\alpha = 180^\circ$) is the spectrum obtained using the angle-averaged cross section $Q(\epsilon, E)$

180°) were used; from our reconstructions we found that angles α up to 10° provide essentially the same results. Moreover, we remark that the maximum value we chose (namely, $\alpha = 180^\circ$) corresponds to an integration over the entire sphere, i.e., to the angle-averaged cross section $Q(\epsilon, E)$ described in Sects. 3 and 4. Following what we have done in Sect. 4.3, a “confidence strip” of $\overline{F}_{\theta_0, \alpha}(E)$ forms based on different realizations of the (noisy) data was produced for each photon spectrum $I(\epsilon)$ and the mean of this confidence strip was used for further analysis.

The August 21, 2002 flare was located at a heliocentric angle of approximately 50° . If we assume that the mean direction of the incoming electrons was vertically downward at this location, then the corresponding value of θ_0 is 130° . Figure 11 shows the reconstructed $\overline{F}_{\theta_0, \alpha}(E)$ corresponding to $\theta_0 = 130^\circ$ and various values of α , including $\alpha = 180^\circ$, i.e., the solid-angle-averaged cross section.

It should be recognized that the assumption of a vertical mean incoming electron direction, and so the choice of $\theta_0 = 130^\circ$ is not rigorously justified. For comparison, therefore,

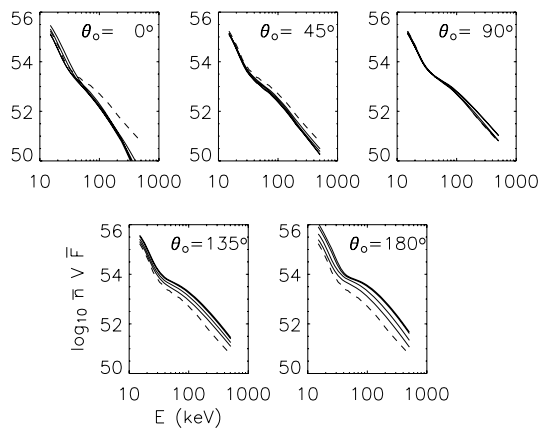


Fig. 12 As for Fig. 11, but for values of $\theta_0 = 0^\circ, 45^\circ, 90^\circ, 135^\circ,$ and 180° . The curves have the same significance as in Fig. 11, with the uppermost solid curves corresponding to the lowest values of α

Fig. 12 shows the same results for values of θ_0 ranging from 0° to 180° in 45° steps.

5.4 Results and Discussion

Let us consider the reconstructed $\overline{F}_{\theta_0, \alpha}(E)$ for the fixed value $\theta_0 = 130^\circ$ of Fig. 11 and for different spread angles α . The case $\alpha = 0^\circ$ corresponds to all the emission concentrated at $\theta = \theta_0$. On the other hand, for values of α which are different from zero, the X-ray emission is spread over a range of θ from $\theta_0 - \alpha$ to $\theta_0 + \alpha$. Owing to the asymmetry of the emission polar diagram which grows for increasing electron energies (see Fig. 8), the enhanced emission in the $\theta_0 - \alpha$ direction more than compensates for the decreased emission in the $\theta_0 + \alpha$ direction, so that fewer total electrons are required to produce a given photon flux than for the unidirectional ($\alpha = 0^\circ$) case. Figure 11 shows that the magnitude of the reconstructed $\overline{F}_{\theta_0, \alpha}(E)$ does indeed depend quite strongly on the range of incoming electron directions α , especially at high electron energies E (at 500 keV, the required electron flux is less by a factor of ~ 4). Since this effect becomes more important with increasing energy, the reconstructed $\overline{F}_{\theta_0, \alpha}(E)$ also becomes *steeper* with increasing α . Recognizing that the extreme maximum case $\alpha = 180^\circ$ corresponds to the angle-averaged cross section used in the previous sections, we see that the use of a cross section that more realistically reflects the range of incoming electron velocity vectors always *flattens* the high-energy part of the inferred electron spectrum relative to that found using the angle-averaged cross section $Q(\epsilon, E)$.

We discuss now the different forms of $\overline{F}_{\theta_0, \alpha}(E)$ when several viewing angles θ_0 are used. For $\theta_0 = 180^\circ$ (Fig. 12), corresponding to vertically downward electrons in a disk-center flare, the recovered flux $\overline{F}_{\theta_0, \alpha}(E)$ for modest values of α is, at high energies, significantly (an order of magnitude or so) greater than the value of $\overline{F}_{\theta_0, 180^\circ}(E)$, i.e., to the

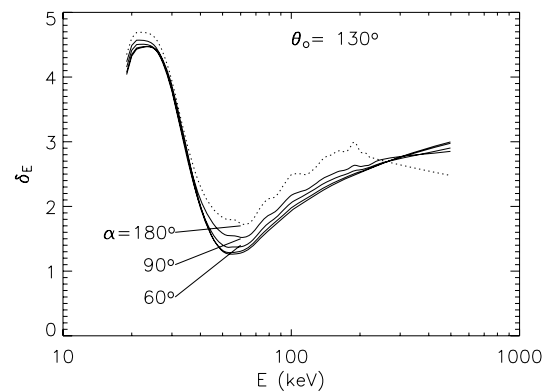


Fig. 13 Spectral index variation for $\theta = 130^\circ$ and values of α indicated

result using the angle-averaged cross section $Q(\epsilon, E)$. This is due to the very low values of the normalized cross section $Q(\epsilon, E; \theta)$ appropriate to this viewing angle (Fig. 8) and hence the inefficiency of photon production in such a direction. Such large fluxes may introduce issues of beam stability. Conversely, for $\theta = 90^\circ$ (corresponding to vertically downward electrons in a limb flare; see Fig. 12), the enhancement over the angle-averaged ($\alpha = 180^\circ$) case is less pronounced; indeed the recovered spectra are remarkably similar to that derived using the angle-averaged cross section, particularly at energies slightly greater than 30 keV.

Values of $\theta_0 < 90^\circ$ correspond to the case in which the mean velocity of the electrons has a component *toward* the observer, and therefore *away* from the Sun. In general, θ_0 values in this first quadrant lead to a *decreased* value of $\overline{F}_{\theta_0, \alpha}(E)$ relative to the angle-averaged result $\overline{F}_{\theta_0, 180^\circ}(E)$, because of the preferential tendency for photons to be emitted in the forward hemisphere (relative to the incoming electron velocity) and hence the smaller number of electrons needed. However, at very low values of θ_0 (slightly smaller than 30°), the required $\overline{F}_{\theta_0, \alpha}(E)$ is, for low energies, *greater* than both the $\theta_0 = 45^\circ$ case and the angle-averaged case (*dashed line*); this is a consequence of the angular behavior of $Q(\epsilon, E; \theta)$ (in particular the low value near $\theta = 0^\circ$) shown in Fig. 8 and noted at the end of Sect. 5.2. The greatest deviations between the correct spectrum and the one deduced using the angle-averaged cross section are thus achieved for $\theta_0 \simeq 45^\circ$ (correct spectrum steeper) and for $\theta_0 \simeq 180^\circ$ (correct spectrum flatter).

Figure 13 shows the variation of the local spectral index δ_E as a function of E , for the $\theta_0 = 130^\circ$ spectra of Fig. 11. Compared to the results for the isotropic case ($\alpha = 180^\circ$), the spectral indices for the anisotropic electron distributions are substantially smaller (flatter spectrum) at low energies (between about 40 and 200 keV), and larger (steeper spectrum) at high energies (greater than about 200 keV). In all cases the value of δ_E increases with decreasing energy below ~ 50 keV, indicative of the transition to a softer,

more thermal, spectrum. At low energies (smaller than about 25 keV), the value of δ_E changes to a form that increases with E , indicative of the general steepening trend associated with thermal spectra.

In conclusion, in determining the mean source electron population $\bar{F}(E)$ responsible for a given hard X-ray spectrum, it is very important to use a bremsstrahlung cross section Q that accurately represents the geometric relationship between the source and the observer. As we have shown in our analysis, the use of the correct, direction-dependent cross section can yield recovered mean source electron spectra of significantly different shape than the results using the usual angle-averaged cross section.

5.5 Open Problems

The results obtained in the simplified cases of the previous sections are the first attempt to face the problem of deeply understanding the angular dependency of the reconstructed electron spectrum. In the general case, the inversion of the anisotropic bremsstrahlung equation

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \int_{\Omega} Q(\epsilon, E; \theta) \widehat{F}(E, \Omega) d\Omega dE, \quad (184)$$

where all the functions and constants are defined in Sect. 5.1, leads to the reconstruction of the function of two variables $\widehat{F}(E, \Omega)$ from the knowledge of noisy measurements of the function of one variable $I(\epsilon)$. This kind of problem is known as *bivariate problem* and its resolution implies notable difficulties.

As described before, our results have been obtained under the simple assumption that $\widehat{F}(E, \Omega)$ is separable in E and Ω —see (177); furthermore, we chose a particular form for the function describing the angular dependence of $\widehat{F}(E, \Omega)$ and we reconstructed the mean electron spectrum $\bar{F}(E)$, which is the part of $\widehat{F}(E, \Omega)$ that depends only on the electron energy E . The results that we provided clearly pointed out that the angular dependency of $\widehat{F}(E, \Omega)$ cannot be neglected.

A distinct but important problem to study is now to assume the E -dependence of $\widehat{F}(E, \Omega)$ as known and still separable and to see if we can recover the θ -dependence from the photon spectrum.

From (184) and (183) we can write

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_{\epsilon}^{\infty} \int_0^{\pi} \int_0^{2\pi} f(E, \beta) \times Q(\epsilon, E; \theta(\beta, \varphi, \theta_0)) \sin \beta d\varphi d\beta dE \quad (185)$$

where $f(E, \beta) = \widehat{F}(E, \Omega)$ (we have not any dependence on φ since symmetries). If we define

$$K(\epsilon, E, \beta) = \frac{\bar{n}V}{4\pi R^2} \int_0^{2\pi} Q(\epsilon, E; \theta(\beta, \varphi, \theta_0)) d\varphi, \quad (186)$$

(185) becomes

$$I(\epsilon) = \int_{\epsilon}^{\infty} \int_0^{\pi} f(E, \beta) K(\epsilon, E, \beta) \sin \beta d\beta dE. \quad (187)$$

The problem is now univariate in β if we assume for example $f(E, \beta) = E^{-\delta} G(\beta)$ (i.e., a power law behavior for the E -dependence of f). It follows that

$$\begin{aligned} I(\epsilon) &= \int_{\epsilon}^{\infty} \int_0^{\pi} E^{-\delta} G(\beta) K(\epsilon, E, \beta) \sin \beta d\beta dE \\ &= \int_0^{\pi} G(\beta) \left[\int_{\epsilon}^{\infty} E^{-\delta} K(\epsilon, E, \beta) dE \right] \sin \beta d\beta \\ &= \int_0^{\pi} G(\beta) H(\epsilon, \beta) \sin \beta d\beta, \end{aligned} \quad (188)$$

where

$$H(\epsilon, \beta) = \int_{\epsilon}^{\infty} E^{-\delta} K(\epsilon, E, \beta) dE. \quad (189)$$

The inversion of (188) is a Fredholm problem to find $G(\beta)$ from $I(\epsilon)$.

Both the problem univariate in E and the problem univariate in β have the drawbacks that they are not the right way to invert (184) and they need strong assumptions about the form of $\widehat{F}(E, \Omega)$. On the other hand, they have the great advantage to reduce the bivariate problem to simpler univariate problems whose resolution can be easily performed. Anyway, the main goal of the anisotropic problem is to invert (184) directly. To this aim, our idea is to perform a lexicographical re-arrangement of the variables followed by a standard zero-order regularization inversion. In other terms, we consider the discretized form of (187) (with weights included in K)

$$I_i = \sum_{j=0}^{n_{el}-1} \sum_{k=0}^{n_{\beta}-1} K_{ijk} f_{jk} \quad (190)$$

where n_{el} and n_{β} are the number of the electron energies sampled and the number of the angles β considered respectively. Then we merge indexes j and k in a single one ($l = n_{\beta} \cdot j + k, l = 0, \dots, n_{\beta}n_{el} - 1$), converting the matrix f_{jk} in a vector f'_l and the tensor K_{ijk} in a matrix K'_{il} . It follows that (190) becomes

$$I_i = \sum_{l=0}^{n_{el}n_{\beta}-1} K'_{il} f'_l \quad (191)$$

so that we can recover f'_l and then deduce f_{jk} from the knowledge of the one-to-one relationship between l and (j, k) .

6 Application to Solar Physics: Thermal Bremsstrahlung

6.1 The Differential Emission Measure

In the previous sections we proposed some regularization techniques to infer information on the electron distribution in the source starting from its measured X-ray emission, given a certain form for the probabilistic kernel which describes the emission process. Once that realistic electron spectra have been provided, the next step is to investigate to what extent the electron distribution responsible for the emission comprises

- (a) non-thermal particles trapped in a low density plasma (thin-target);
- (b) particles “injected into” and stopped in a dense plasma (thick-target);
- (c) a spatial distribution of locally Maxwellian electrons with a location-dependent temperature (T),

or some mixture of these three situations.

Under a purely thermal interpretation of (100) [14], the electron distribution is assumed to be locally Maxwellian, *i.e.* (with T in energy units)

$$F(E, \mathbf{r}) = \frac{2^{3/2}}{(\pi m_e)^{1/2}} \frac{n(\mathbf{r})E}{[T(\mathbf{r})]^{3/2}} e^{-E/T(\mathbf{r})}, \tag{192}$$

so that (100) becomes

$$I(\epsilon) = \frac{1}{4\pi R^2} \cdot \frac{2^{3/2}}{(\pi m_e)^{1/2}} \int_0^\infty \int_\epsilon^\infty n^2(T) \frac{E}{T^{3/2}} \times e^{-E/T} Q(\epsilon, E) dE \frac{d\mathbf{r}}{dT}, \tag{193}$$

and the photon spectrum then provides information on the differential emission measure loosely defined, for stratified structures, by

$$\xi(T) = n^2(T) \frac{d\mathbf{r}}{dT}. \tag{194}$$

A direct connection between $I(\epsilon)$ and $\xi(T)$ can be established by inserting expression (192) for the local electron distribution into the model-independent equation (100), with a Kramers approximation used for $Q(\epsilon, E)$ (see (107)). Because of the extreme simplicity of the Kramers form, the result is a Laplace-transform-like integral equation relating the photon spectrum ($I(\epsilon)$) directly to the differential emission measure ($\xi(T)$). This equation has been studied by Piana, Brown and Thompson [66] in the framework of regularization theory for inverse problems and applications to high-resolution balloon data [56] have been considered. Craig & Brown [20] noticed that an analogous approximate equation could be obtained for the more complex Bethe-Heitler form

of the cross section $Q(\epsilon, E)$ (see (109)). However, both of these use a rather coarse approximation to the true cross section, which as well as being quite smooth, has a much more complex analytic form, like for example the highly relativistic formula given by equation (3BN) in [44] and reported in Sect. 3.1 (see (110)). As we showed in Sect. 4, the ill-posedness of the integral inversion problem implies that small changes in the kernel (Q here) can result in significant changes in the solution [5, 42, 51] so that results using approximate representations of Q may not be reliable. It follows that a completely rigorous description of the thermal model requires the introduction of two integral equations. First, the (isotropic) source-averaged, effective electron flux spectrum $\bar{F}(E)$, defined in (102) is related to the photon spectrum $I(\epsilon)$ by means of the Volterra integral equation

$$I(\epsilon) = \frac{\bar{n}V}{4\pi R^2} \int_\epsilon^\infty \bar{F}(E) Q(\epsilon, E) dE, \tag{195}$$

with a fully correct form of $Q(\epsilon, E)$. Then, (192), (194) and (102) lead to the Fredholm integral equation relationship between $\bar{F}(E)$ and $\xi(T)$ [15]

$$\bar{F}(E) = \frac{1}{\bar{n}V} \frac{2^{3/2}E}{(\pi m_e)^{1/2}} \int_0^\infty \frac{\xi(T)}{T^{3/2}} e^{-E/T} dT. \tag{196}$$

The aim of this section is to address the following two basic questions concerning the thermal model: 1) is the available photon spectrum compatible with a thermal interpretation, *i.e.*, can the observed $I(\epsilon)$ be fully explained by a non-negative $\xi(T)$? And 2) if the answer to 1) is yes, what is the actual form of $\xi(T)$ for that particular form of $I(\epsilon)$? One way to see whether an entire $I(\epsilon)$, or even part of it, is compatible with a thermal model for the emission process, is to test whether the corresponding $\bar{F}(E)$ obtained by solving (195) satisfies criteria arising from (196), making allowances for the data-induced noise. One such test is the “derivative test” for thermality found by Brown and Emslie [15]. This follows directly by differentiating (196) (with both sides divided by E) i times and states that an electron spectrum ($\bar{F}(E)$) is compatible with a purely-thermal interpretation if and only if the quantity $\bar{F}(E)/E$ is “completely monotonic”, *i.e.* its i -th derivative has sign $(-)^i$ at all E . This approach has a technical limitation. Equation (195) can be solved by using regularization techniques (see Sects. 3 and 4) but derived electron spectra are affected by noise in the photon spectra used. Successive derivatives in the thermality test therefore have rapidly escalating errors, due to the instability of numerical differentiation. It follows that the computation of only the first two or three orders of derivative is reliable [28], with the higher- i terms in the “derivative test” too noisy to be useful. On the other hand, we can be confident that any $\bar{F}(E)$ clearly failing the “derivative test” at a high confidence level, for given noise,

can be ruled out as entirely due to a thermal distribution with an everywhere non-negative $\xi(T)$. However, even if the $\overline{F}(E)$ does pass the “derivative” test, this itself does not tell us the (non-negative) form of $\xi(T)$ to which $\overline{F}(E)$ corresponds. Therefore, in principle a much more effective technique would be to solve (196), where $\overline{F}(E)$ is obtained by solving (195), thus describing the $\xi(T)$ corresponding to an observed photon spectrum: if $\xi(T) \geq 0$ for all T , then the photon spectrum can be reliably interpreted according to a thermal model for the bremsstrahlung emission; if $\xi(T) < 0$ for some temperature interval, then at least part of the emission is certainly non-thermal. Furthermore the knowledge of possible features in such reconstructed form could yield important information on plasma heating and conduction processes [14, 16]. In recent years this inversion problem has achieved an unprecedented level of importance because of the high-resolution photon spectra ($\Delta\epsilon \simeq 1$ keV) obtained from the RHESSI mission [57]. Combined with optimization of computational methods for regularized solutions of the ill-posed inverse problem involved [19, 20, 41, 48, 49, 58, 63, 67, 78] it is now possible to infer mean source electron spectra with which specific physical source models can be compared [17].

A basic technical difficulty in the reconstruction of $\xi(T)$ is due to the fact that solving the second inverse problem (196) is extremely problematic. Simple changes of variables reduce this problem to a Laplace transform inversion problem with noisy data. There is a vast literature [22, 31, 81] showing that this problem is intrinsically highly pathological, due to the very broad filtering action of the Fredholm-Laplace integral kernel (compared to that in the basic bremsstrahlung inverse problem (195) which is of Volterra type and not severely filtering). Several regularization methods [7, 11, 61] have been introduced to handle this inversion by reducing the unphysical oscillations due to the presence of noise. For all of them two considerations are mandatory: first, that, as stated by Davies and Martin [22] “[in the Laplace inversion problem with noisy data] no single method gives optimum results for all purposes...”, and therefore no general method exists which is effective at the highest level for all physical situations and all kinds of data; second, that, whatever method is applied, even with very accurate data, only a coarse resolution will be achieved in the recovered solution [6].

Most inversion methods for the real Laplace transform have been formulated within the framework of regularization theory for ill-posed inverse problems [3]. At the core of these approaches there is the search for an optimal trade-off between stability against unphysical oscillations and accurate reproducibility of the data. Such an optimization result is obtained either by fixing a real positive regularization parameter in Tikhonov-like methods (see Sect. 2.5) or by

applying some stopping rule to iterative procedures. However, the present application is particularly challenging owing to the particular nature of the solar spectral data involved here. Typical solar $\overline{F}(E)$ are characterized by a large dynamic range (at least three orders of magnitude for around one order of magnitude in the E range) and, more significantly, the corresponding $\xi(T)$ have completely different forms at low and high T : at small T , a near-thermal (δ function) component which differs from zero only in a small T range (narrow support); at high T , a monotonic component spread over a large interval. A consequence of this complexity in the source function is that regularization approaches may lose some (or most) of their effectiveness. For example, the reconstruction of $\xi(T)$ at low T with classical Tikhonov regularization may correctly reproduce the location of the temperature peak but typically presents ringing effects whose negative components, which are numerical artefacts, might suggest that the spectrum is not thermally interpretable. Negative ringing can be eliminated by applying a reconstruction method with a positivity constraint. However such an approach is not effective at recovering the high-temperature part of $\xi(T)$ which has a power-law-like behavior and requires regularization methods with more smoothing power. To deal with these kinds of difficulty, here we utilize the following approach: an iterative scheme with a positivity constraint is applied for the inversion of the low-energy part of $\overline{F}(E)$, in order to eliminate unphysical ringing effects with negative oscillations in the reconstruction of the part of $\xi(T)$ characterized by a narrow support; then, a first-order Tikhonov regularization method is applied for the inversion of the high-energy part of $\overline{F}(E)$, where an appropriate boundary condition constrains the reconstructed $\xi(T)$ to behave well (*i.e.*, with a slope compatible with the spectral index of the photon data) at high T . The two reconstructed $\xi(T)$ are then connected together noting that the connection temperature is easily determined by the T value where the thermal $\xi(T)$ goes to zero, *i.e.* the high- T limit of the narrow support of the thermal $\xi(T)$. We observe that, as far as the inversion of the low energy part of $\overline{F}(E)$ is concerned, the use of the positivity constraint in the inversion makes the thermality test based on the verification that the reconstructed $\xi(T)$ is positive at all T , inappropriate, since positivity is forcefully imposed in the inversion procedure. Therefore for this inverse problem the compatibility between the data and the thermal model is tested by checking whether the residuals in $\overline{F}(E)$ corresponding to the $\xi(T)$ recovered by exploiting the positivity constraint are statistically acceptable.

6.2 Regularization Methods

We could in principle proceed directly from (196) to see whether some $\overline{F}(E)$ could be wholly thermal in origin if we had a completely reliable inversion method: given a data

vector $\overline{F}(E)$, a wholly thermal interpretation of it is possible if and only if the $\xi(T)$ obtained by the inversion method has no statistically significant negative values over any range of temperature (T). Our aim here is to address this problem by means of two numerical algorithms based on regularization theory for ill-posed inverse problems, keeping in mind that the effectiveness of any regularized inversion approach in the present case is much weaker than for most other linear inverse problems due to the extreme numerical instability of the Laplace problem, with its very broad kernel. A quantitative estimate [34] for the instability of linear equations like (196) is given in terms of the condition number C of the kernel (cross section) matrix (see Sect. 2.1). It can be shown [20, 66] that, for typical solar data parameters, the condition number associated with the (Fredholm) equation (196) is of the order of 10^{10} , which is much bigger, for similar parameters, than the condition number associated with the (Volterra) bremsstrahlung spectrum to electron spectrum inversion problem (195) (see Table 2). The actual consequences of ill-conditioning are highly significant. A regularization algorithm essentially expresses the approximate smoothed solution as a truncated linear sum of some basis functions. In the basic bremsstrahlung spectrum inversion problem (195), $\overline{F}(E)$ can be expressed in terms of around ten basis functions for typical noise in the case of a data vector with around 100 points, while in the differential emission measure inversion problem (196) we find that only two, or at most three, basis functions can be meaningfully included in the expansion of $\xi(T)$. Therefore, in the recovery of $\xi(T)$ it is necessary to introduce much more severe constraints than the one adopted in the $\overline{F}(E)$ inversion procedure described in Sects. 3 and 4. Even incorporating these constraints, it will be impossible to achieve a temperature resolution anywhere nearly comparable with the spectral resolution with which $\overline{F}(E)$ can be reconstructed through the solution of the bremsstrahlung equation (195) (cf. the analysis of the temperature resolution problem in [20]).

Adopting the change of variable $y = 1/T$, (196) becomes

$$K \frac{\overline{n}V\overline{F}(E)}{E} = \int_0^\infty f(y) \exp(-Ey)dy, \tag{197}$$

where $K = \sqrt{\pi m_e/8} = 1.89 \times 10^{-14} \text{ gm}^{1/2} = 4.73 \times 10^{-10} \text{ keV}^{1/2} \text{ cm}^{-1} \text{ s}$ and $f(y) \text{ (cm}^{-3} \text{ keV}^{-1/2})$ is defined as

$$f(y) = \frac{\xi(1/y)}{y^{1/2}}, \tag{198}$$

with $\xi(T)$ in units of $\text{cm}^{-3} \text{ keV}^{-1}$. Equation (197) involves a continuous representation of the model ($f(y)$) and of the data ($\overline{F}(E)$), while real data are discrete, truncated and affected by measurement and systematic noise. In reality,

therefore, the situation is described by the (finite rank linear) operator $L : X \rightarrow Y$ such that

$$(Lf)_n = \int_0^\infty f(y) \exp(-E_n y)dy, \quad n = 1, \dots, N, \tag{199}$$

where the $\{E_n\}_{n=1}^N$ are the sampled electron energies, X is the functional space containing the solution and Y is the Euclidean space containing the data. Then our problem is to solve

$$Lf = \mathbf{g}, \tag{200}$$

with data vector \mathbf{g} in Y having components

$$g_n = K \frac{\overline{n}V\overline{F}(E_n)}{E_n}, \quad n = 1, \dots, N. \tag{201}$$

As already stated, (200) is a strongly ill-conditioned linear problem and the only way to obtain a realistic approximate solution in the presence of noise is some reconstruction technique based on regularization theory for linear inverse problems. One approach is the first order Tikhonov method, which solves the minimization over f of

$$\|Lf - \mathbf{g}\|_Y^2 + \lambda \|f'\|_X^2 = \text{minimum}, \tag{202}$$

where λ is the (real positive) regularization parameter. It can be proved [66] that under boundary conditions

$$f(0) = 0 \tag{203}$$

and

$$\lim_{y \rightarrow \infty} f'(y) = 0 \tag{204}$$

the analytical solution of (202) is

$$f_\lambda(y) = \sum_{k=1}^N \frac{\sigma_k}{\sigma_k^2 + \lambda} (\mathbf{g}, \mathbf{v}_k)_Y u_k(y), \tag{205}$$

where the σ_k and \mathbf{v}_k are respectively the eigenvalues and eigenvectors of the Gram matrix (see Definition 6)

$$G_{nm} = \int_0^\infty \phi'_n(y)\phi'_m(y)dy, \tag{206}$$

$$\phi_n(y) = \frac{1}{E_n^2} (1 - e^{-E_n y}) \tag{207}$$

and

$$u_k(y) = \frac{1}{\sigma_k} \sum_{n=1}^N (\mathbf{v}_k)_n \phi_n(y). \tag{208}$$

For this problem, first-order regularization is more effective than zero-order regularization for two basic reasons.

First of all, it prescribes a bound on the first derivative of the regularized solution which, in this case of large numerical oscillations, is a sensible thing to do. Second, in this particular implementation, condition (203) constrains the regularized solution to behave well at $y = 0$ ($T \rightarrow \infty$), thus improving the restoration accuracy for $\xi(T)$ at high T . It is also true, however, that condition (204) has no physical basis, and hence may yield artefacts at low T .

The main disadvantage of using Tikhonov regularization is that solutions with negative components can result from noisy data. In particular, in the reconstruction of the low temperature component of $\xi(T)$, typically characterized by a very narrow support, an effective method would allow us to constrain the restored solution to be positive, thus avoiding unphysical ringing effects due to the presence of noise on the data. The introduction of such a constraint [64] has the effect of increasing the resolution power of the inversion approach, allowing reconstruction of more details in the source function. The method with positivity applied in this section is the projected Landweber method, first formulated by Lagendijk, Biemond and Boeke [52] for the image restoration problem. The mathematical properties of this method are discussed, for example, by Eicke [23] and an accelerated version has been provided by Piana and Bertero [65]. We first consider the discretized version of (200)

$$\mathbf{g} = \mathbf{L}\mathbf{f} \tag{209}$$

where \mathbf{f} comes from the sampling of (198) and \mathbf{L} is the matrix with entries

$$L_{mn} = \exp(-E_n y_m) \delta y \tag{210}$$

where the $y_m, m = 1, \dots, M$ are uniformly sampled and δy is an appropriate integration weight. The projected Landweber method provides reconstructions of $f(y)$ (and therefore of $\xi(T)$) by optimally stopping the iteration

$$\mathbf{f}_{k+1} = P_+(\mathbf{f}_k + \tau \mathbf{L}^T(\mathbf{g} - \mathbf{L}\mathbf{f}_k)), \quad \mathbf{f}_0 = 0, \tag{211}$$

where τ is a relaxation parameter, \mathbf{L}^T is the transpose matrix of \mathbf{L} , and P_+ sets to zero all the negative components at each iteration.

As already stated in the previous section, the regularization effects on the approximate solutions provided by Tikhonov first-order regularization and by the projected Landweber method can be obtained by fixing λ in (205) and the iteration number in (211). To this purpose many criteria have been introduced [21]; here we adopted the same approach as in Piana *et al.* [67], based on the analysis of the regularized cumulative residuals. For example, in the case of first-order Tikhonov method we consider the function

$$S^\lambda(k) = \frac{1}{k} \sum_{i=1}^k r_i^\lambda, \quad k = 1, \dots, N \tag{212}$$

where r_i^λ is the i -th normalized regularized residual corresponding to the regularized $\xi(T)$. For completely uncorrelated noise, the normalized cumulative residuals exhibit a random walk with expected deviation $1/\sqrt{k}$. In (212) the presence of the regularization parameter increasingly correlates the r_i^λ for increasing values of λ . Therefore an optimal criterion to fix λ is to look for the largest value of λ such that $|S^\lambda(k)|$ is bounded by $3/\sqrt{k}$. An analogous procedure is followed for stopping the projected iterations, whereby, in this case, the regularization parameter is represented by the iteration number.

6.3 Simulations

In this section we wish to test the effectiveness of the regularization approach as introduced before. In particular, we describe the case of a power law with a low-energy cut-off showing that if the mean electron spectrum is sampled starting from energies bigger than the cutoff, the reconstruction is rather accurate (in fact, the problem becomes that of recovering a pure power law in a limited domain) while the reconstruction dramatically fails if the minimum sampled energy is smaller than the energy cut-off, in accordance with the fact that a power law with a low-energy cut-off is *not* compatible with thermal bremsstrahlung emission. Then, the temperature resolution achievable by the method is discussed, the performance of the method in reconstructing power laws is tested and, finally, a realistic form of $\bar{F}(E)$ obtained by regularized inversion of a synthetic photon spectrum is considered.

- **Compatibility Test**

We want to verify whether an $\bar{F}(E)$ reconstructed from photon data $I(\epsilon)$ can be interpreted as consistent with a purely thermal model. The “derivative test” of Brown & Emslie [15] provides a possible approach, but does not yield information on the temperature structure of the source. A more informative approach is to apply a reconstruction method and to check if the reconstructed $\xi(T)$ is non-negative for all T . As an example, let us consider the case of a mean source electron spectrum

$$\bar{F}(E) \propto \begin{cases} E^{-\delta}, & E \geq E_c, \\ 0, & E < E_c, \end{cases} \tag{213}$$

with E_c a low-energy cutoff. Before performing the inversion, however, we discuss some informative analytic aspects of (213) in relation to the general expression (196) for $\bar{F}(E)$ from a purely thermal source, which we rewrite, ignoring constant factors, as

$$\bar{F}(E) \propto E \int_0^\infty \frac{\xi(T)}{T^{3/2}} e^{-E/T} dT. \tag{214}$$

First it is obvious that if $\xi(T)$ is greater than zero over any T interval then the corresponding $\bar{F}(E)$ is never zero at

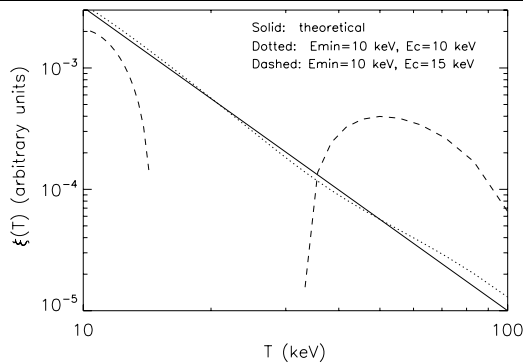


Fig. 14 Reconstruction of the differential emission measure corresponding to an electron spectrum in the form of a power law with a low-energy cutoff, for two different values of the cutoff energy (E_c) and of the minimum sampled energy (E_{min}). The reconstruction method is first-order Tikhonov regularization with boundary conditions. If $E_{min} \geq E_c$, the sampled electron spectrum does not include a cutoff and $\xi(T)$ is faithfully recovered (dotted line). If $E_{min} < E_c$, the sampled electron spectrum does include a cutoff and so is not compatible with a thermal interpretation. In this case the reconstruction of $\xi(T)$ is unphysical (dashed line). The solid line is proportional to $T^{-5/2}$ in the range 10–100 keV

any E . Thus the $\overline{F}(E)$ in (213) cannot be purely thermal (it clearly fails the derivative test at $E = E_c$). Second, we note that for a pure power law $\xi(T) \propto T^{-\alpha}$ at all T with α constant, the resulting $\overline{F}(E)$ is proportional to $E^{-\alpha + \frac{1}{2}}$ at all E . Consequently, for $\alpha = \delta + \frac{1}{2}$ a pure (untruncated) power-law $\xi(T)$ predicts $\overline{F}(E)$ in (213) perfectly in the range $E \geq E_c$ but completely contradicts it in the range $E < E_c$. Thus a thorough thermality test must be applied to all E ; failure (within the allowed uncertainties) at even one value of E is enough to rule out a purely thermal model.

A somewhat surprising result here is that a wholly-thermal model is ruled out by the form of $\overline{F}(E)$ at low rather than at high energies. We also emphasize that the power-law relation between $\xi(T)$ and $\overline{F}(E)$ only holds (at $E \geq E_c$) for a complete power-law $\xi(T)$. If $\xi(T)$ is only a power law over some finite range, say (T_1, T_2) , the corresponding $\overline{F}(E)$ is not a power law at any E but rather, with $x = T/E$,

$$\overline{F}(E) \propto E^{-\alpha + 1/2} \int_{T_1/E}^{T_2/E} x^{-\alpha - 3/2} e^{-1/x} dx. \quad (215)$$

We now show that application of our inversion method to (213) agrees well with these analytic results.

For the inversion, the data (213) is discretized according to uniform sampling starting from a minimum sampled energy (E_{min}), realistic Poisson noise is added to the corresponding photon spectrum and errors on $\overline{F}(E)$ are generated by inverting the noisy $I(\epsilon)$. We applied the first-order Tikhonov inversion method for two possible experimental situations concerning the relative values of

the pair E_c, E_{min} , and for $\delta = 2$ with the results shown in Fig. 14. When $E_c \leq E_{min}$ (i.e., the cutoff is not sampled), a stable differential emission measure is restored. There are some slight, long-wavelength oscillations in the recovered $\xi(T)$ of roughly the width of the kernel but the mean temperature spectral index is close to the theoretical value $\alpha = 2.5$. On the other hand, when $E_{min} < E_c$ (and so the cutoff is sampled), the reconstruction contains large negative ranges and is absolutely unphysical as expected. This behavior is consistent with the fact that a mean source electron spectrum with any cutoff is incompatible with a purely thermal interpretation of the emission (since any Maxwellian contains electrons of all E).

• Temperature Resolution.

Heuristically, the effective temperature resolution achievable by our inversion method can be assessed by reconstructing $\xi(T)$ forms using the $\overline{F}(E)$ corresponding to input δ functions $\xi(T) \sim \delta(T - T_0)$. The resulting reconstructed forms of $\xi(T)$ are characterized by finite Full Widths at Half Maximum (FWHM) which estimate the resolution achievable around $T = T_0$. Therefore for inverse problems the resolution power depends on the reconstruction method. In Table 4 the FWHM values realized through application of the two reconstruction methods discussed in Sect. 6.2 to $\overline{F}(E)$ spectra are given for different values of T_0 . The averaged electron spectra are obtained by inverting the corresponding photon spectra (affected by realistic Poisson noise) and contain 50 points in the energy range 1–50 keV. Table 4 also contains values for the “centroid” temperature of the reconstructed distributions, defined by $\langle T \rangle = \int T \xi(T) dT / \int \xi(T) dT$. In the case of first-order regularization these $\langle T \rangle$ are 10% or so higher than the T_{max} at which the recovered $\xi(T)$ peaks because the $\xi(T)$ are skewed, and they compare well with the single input T_0 of the originally-assumed δ functions. In the case of the iterative projected Landweber method, $\langle T \rangle$ and T_{max} coincide in most cases and are very close to the theoretical T_0 .

When first-order regularization is applied in the case of multi-thermal sources, the FWHM values given in Table 4 may be overly optimistic estimates of the temperature resolution particularly when trying to separate narrow features. As an example, we consider reconstructions of two δ functions with both methods, where the first is peaked at $T_1 = 2.5$ keV and the second is peaked at $T_2 = 10, 7, 5, 4.5$ keV, respectively—see Fig. 15.

Also in this case for the reconstruction we considered $\overline{F}(E)$ sampled in the energy range 1–50 keV. We note that the use of the positivity constraint increases the resolution limit as explained for example in Piana and Bertero [64] by means of arguments based on the analytic continuation principle. Furthermore, the $\xi(T)$ reconstructed

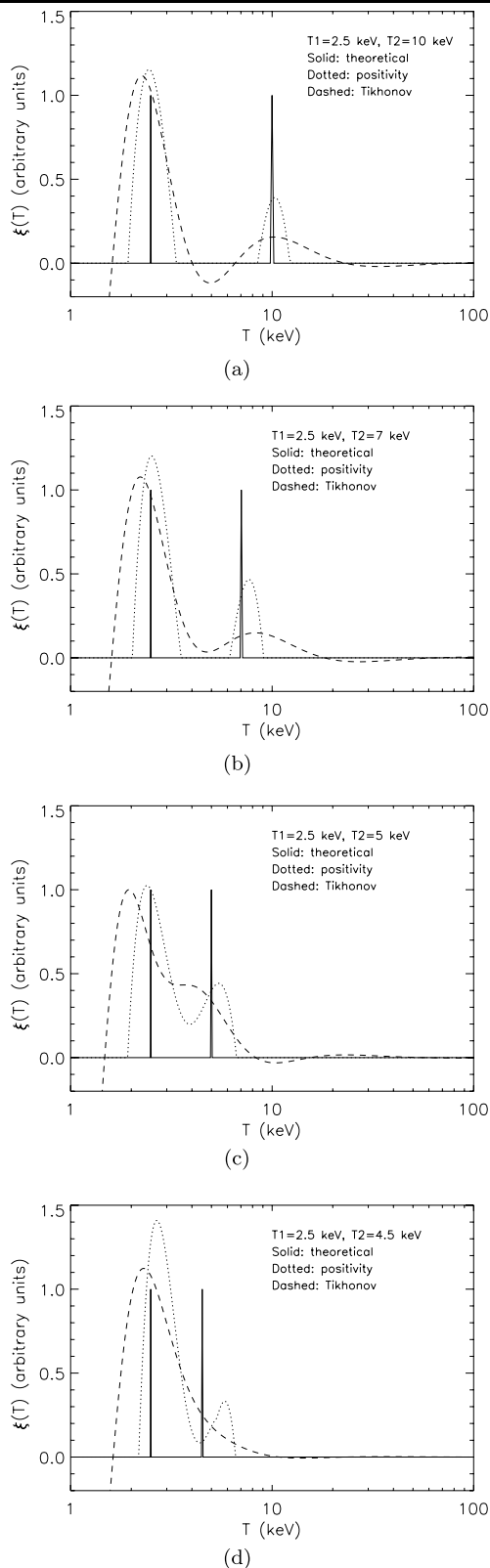


Fig. 15 Reconstructions of two δ functions (solid line) by means of first-order regularization (dashed) and projected-Landweber method (dotted). The averaged electron spectrum contains 50 points uniformly sampled in the range 1–50 keV and is obtained by inverting the corresponding photon spectrum with realistic Poisson noise added: (a) $T_1 = 2.5$ keV, $T_2 = 10$ keV; (b) $T_1 = 2.5$ keV, $T_2 = 7$ keV; (c) $T_1 = 2.5$ keV, $T_2 = 5$ keV; (d) $T_1 = 2.5$ keV, $T_2 = 4.5$ keV

Table 4 Full Widths at Half Maximum (FWHM) for the reconstruction of δ functions peaked at different temperatures. The maximum and the centroid values of T corresponding to the reconstructed $\xi(T)$ are also given. The inversion methods are first-order regularization with boundary conditions and projected-Landweber method with positivity

Input T_0 (keV)	2.5	3.5	4.5	5.5	6.5
T_{max} (Tikhonov)	2.3	2.8	3.4	3.9	4.5
$\langle T \rangle$ (Tikhonov)	2.6	3.3	4.6	5.7	7.1
FWHM (Tikhonov)	1.5	2.0	3.3	4.2	5.6
T_{max} (positivity)	2.5	3.6	4.6	5.6	6.6
$\langle T \rangle$ (positivity)	2.5	3.6	4.6	5.6	6.6
FWHM (positivity)	0.8	0.9	1.1	0.9	1.5

Input T_0 (keV)	7.5	8.5	9.5	10.5	11.5
T_{max} (Tikhonov)	5.2	6.1	7.3	7.9	8.6
$\langle T \rangle$ (Tikhonov)	8.5	9.0	9.6	10.5	11.6
FWHM (Tikhonov)	7.3	9.4	11.8	12.3	13.2
T_{max} (positivity)	7.6	8.3	9.6	10.6	11.6
$\langle T \rangle$ (positivity)	7.6	8.5	9.7	10.8	11.6
FWHM (positivity)	1.3	2.6	2.3	2.3	2.4

by means of the Tikhonov method have unphysical negative components. We conclude that for the recovery of the low-temperature part of the differential emission measure the projected algorithm is significantly more effective. We also observe that the energy range 1–50 keV, where $\bar{F}(E)$ was sampled for this inversion, is in some sense optimal, since it always includes the peak temperatures to be recovered. In real $\bar{F}(E)$, energies up to typically 10 keV must be avoided owing to the presence of (or problematical correction for) lines of non-bremsstrahlung origin or to systematic errors introduced by the hardware. In other words, a typical experimental situation is that $\bar{F}(E)$ is inverted from electron energies larger than the temperatures involved in the thermal process. In order to study the effect of this on the inversion method, we considered the test shown in Fig. 16.

The electron spectrum corresponding to an isothermal $\xi(T)$ with $T_0 = 7$ keV is inverted for different electron energy sampling ranges: 2–20 keV (solid), 2–7 keV (dashed), 7–20 keV (dotted) and 20–70 keV (dot-dashed). We found that if T_0 is higher than the energy range considered, the reconstruction preserves the symmetry of the δ function (so that T_{max} and $\langle T \rangle$ more or less coincide) but the peak temperature is notably overestimated (almost 20%). If T_0 is smaller than the sampled energies (which is the realistic situation), the reconstruction is rather skewed (in such a way that T_{max} is bigger than $\langle T \rangle$, as opposed to the case of Tikhonov regularization), presents a widened

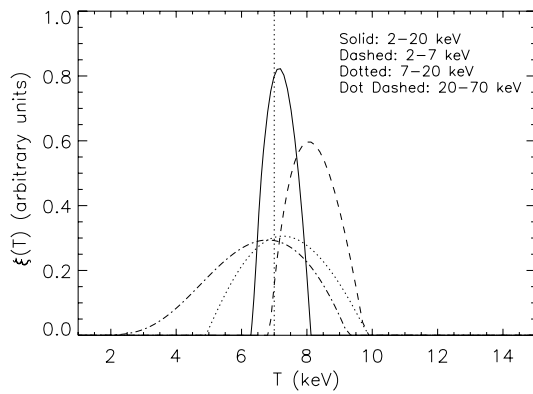


Fig. 16 Reconstruction of a δ function peaked at $T_0 = 7$ keV when the corresponding $\bar{F}(E)$ is sampled over different electron energy ranges: 2–20 keV (solid); 2–7 keV (dashed); 7–20 keV (dotted); 20–70 keV (dot-dashed). The reconstruction method is the projected-Landweber method with positivity

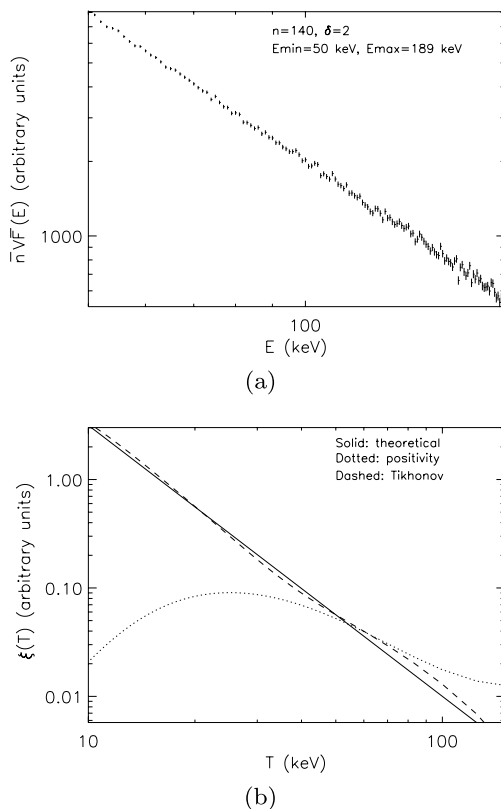


Fig. 17 Inversion of $\bar{F}(E)$ corresponding to the case $\xi(T) \sim T^{-5/2}$: (a) $\bar{F}(E)$ uniformly sampled with $N = 140$ points in the energy range 50–189 keV; (b) theoretical $\xi(T)$ (solid), reconstruction given by first order regularization (dashed), reconstruction given by the projected-Landweber method (dotted)

FWHM and the peak temperature is slightly underestimated: for example, if the selected range is 20–70 keV, the reconstructed T_{max} is $\sim 5\%$ smaller than the true one.

• Power Laws

The situation is notably different for the recovery of the high temperature part of $\xi(T)$. In this range the typical behavior is close to a power law $T^{-\alpha}$, which generates a power-law shape $E^{-\delta}$ in the corresponding high energy part of $\bar{F}(E)$ (the approximate relation $\alpha \sim \delta + 1/2$ has been already discussed before). In this case, the small y (high T) boundary condition (203) plays a constructive role in the recovery of $\xi(T)$ and makes first-order regularization more effective than the projected iterative scheme (in this case the action of the positivity constraint is insignificant, since the $\xi(T)$ to be recovered has a wide support *i.e.* is everywhere far from zero and possible residual oscillations do not induce negative components). A test example is represented in Fig. 17, where we invert the spectrum $\bar{F}(E)$ corresponding to the input form

$$\xi(T) \sim T^{-5/2}. \quad (216)$$

This electron spectrum has been obtained by inverting the corresponding photon spectrum given by (195) with the addition of realistic Poisson noise. $\bar{F}(E)$ in Fig. 17(a) has been uniformly sampled with $N = 140$ points in the range 50–189 keV (in the case of power laws the sensitivity of the reconstruction qualities on the energy range adopted for the inversion is not very significant) and inverted in Fig. 17(b) by means of the first-order regularization method and for the projected Landweber method. The results of this computation clearly show that first-order regularization with the boundary condition (203) is particularly effective in this case. We finally note that for notably larger values of δ a certain deterioration of the reconstructions may occur, due to the fact that λ is a global regularization parameter which works in a less effective way when the function to reconstruct is steep. However this deterioration can be reduced by means of an appropriate rescaling of the power law (see [49]).

• More Realistic Spectra

A mean electron flux spectrum reconstructed from a real photon spectrum is often assumed [39] to comprise an isothermal component at low electron energies plus a power-law behavior at high energies. A simple example is given by the model electron spectrum (with T, E in keV)

$$\bar{F}(E) = 100T_0^{-3/2} E e^{-E/T_0} + \frac{\Gamma(\delta+1)}{100} \left(\frac{E}{50}\right)^{-\delta}; \quad (217)$$

application of the derivative test [15] shows that this spectrum is consistent with a wholly thermal source; indeed the corresponding differential emission measure is

$$\xi(T) = 100\delta(T - T_0) + 0.5(50)^{\delta-1} T^{-(\delta+0.5)}. \quad (218)$$

We discretized the $\bar{F}(E)$ form (217) with a uniform 1 keV sampling from 1 keV to 250 keV with $T_0 = 4$ keV and

Fig. 18 Inversion of $\bar{F}(E)$ corresponding to $\xi(T)$ given in (218) for $T_0 = 4$ keV and $\delta = 2$: (a) electron spectrum with $N = 250$ sampled energies in the range 1–250 keV; (b) theoretical $\xi(T)$ (solid) with the reconstructions given by first-order regularization (dashed) and projected-Landweber method (dotted); (c) theoretical $\xi(T)$ (solid) and reconstruction (dashed) obtained by inverting the low-energy part of the electron spectrum with the projected-Landweber method and the high-energy part with first-order Tikhonov regularization and by connecting the two restorations; (d) cumulative regularized residuals (solid) for the method with positivity (upper panel) and Tikhonov regularization (lower panel) compared to the statistical bound $\pm 3/\sqrt{k}$ (dashed)

$\delta = 2$. We then generated the corresponding photon spectrum ($I(\epsilon)$) using an exact (isotropic) cross section and added random Poisson noise, resulting in corresponding noise in $\bar{F}(E)$. Figure 18(a) shows the resulting simulated $\bar{F}(E)$ while Fig. 18(b) contains the restorations provided by the two methods. Both reconstructions present notable unphysical artefacts which are essentially due to the fact that neither method is able to fully restore the two completely different behaviors of the source function at low and high T . Therefore we considered an approach whereby the two different inversion methods are applied one to the low- and one to the high-energy part of the electron spectrum separately. More precisely, the projected Landweber method is applied to $\bar{F}(E)$ in the low-energy range (here we used 2–36 keV, which approximately corresponds to the range where the spectrum is optimally fitted by the isothermal component). On the other hand, first-order Tikhonov regularization with boundary conditions is applied to $\bar{F}(E)$ in the high-energy range (here we used 55–204 keV, which approximately corresponds to the range where the spectrum is optimally fitted by a power-law). The two reconstructed $\xi(T)$ are connected together at the temperature where the thermal peak goes to zero and plotted in Fig. 18(c) while Fig. 18(d) shows that the regularized cumulative residuals (212) are statistically reliable for the chosen value of the iteration number and of the Tikhonov regularization parameter.

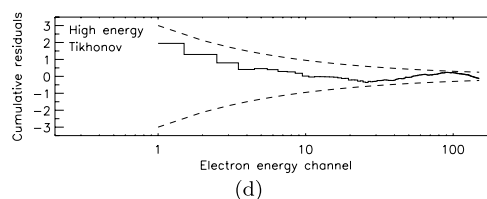
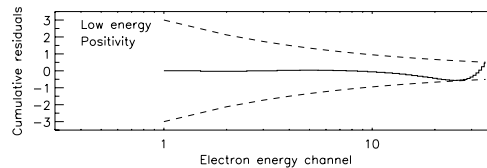
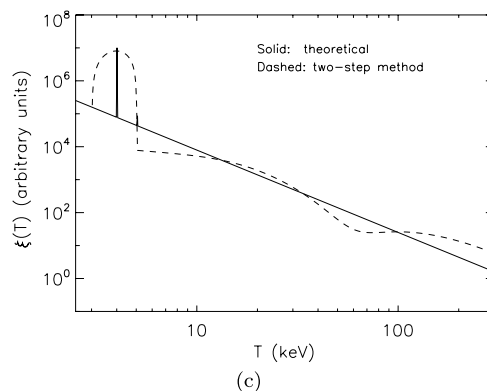
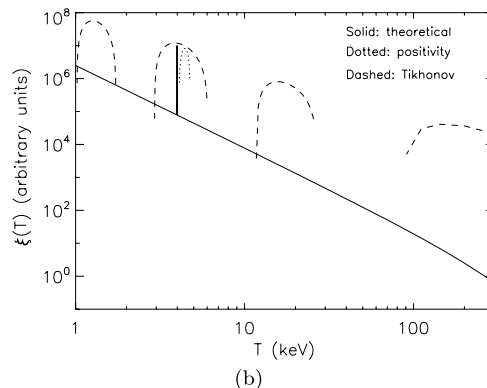
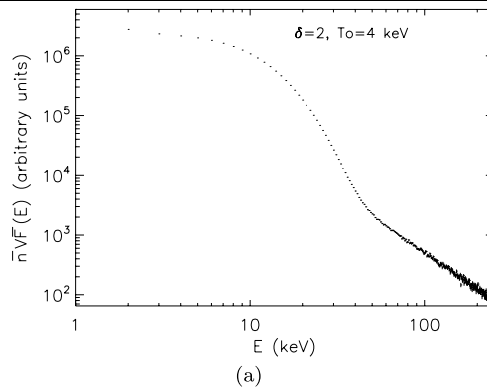
6.4 Application to RHESSI Data

In order to address the analysis of real spectra provided by RHESSI, we first need to check the compatibility between condition (203) and the asymptotic behavior of the recorded photon spectrum at high energies. Such an issue can be addressed by simple integral computations showing that, if a function $F(t)$, for $t \rightarrow 0$, is

$$F(t) \sim At^\beta \tag{219}$$

with $\beta > -1$, then its Laplace transform $(\mathcal{L}F)(s)$, for $t \rightarrow \infty$, is

$$(\mathcal{L}F)(s) \sim A \frac{\Gamma(\beta + 1)}{s^{\beta+1}}, \tag{220}$$



with

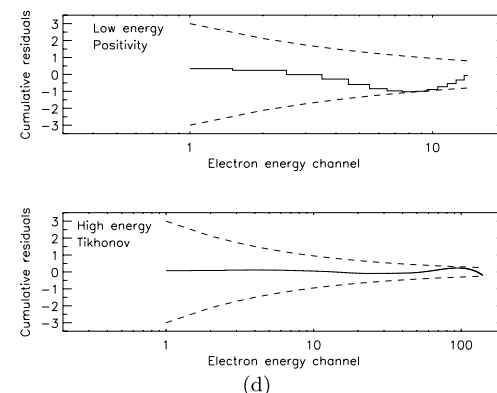
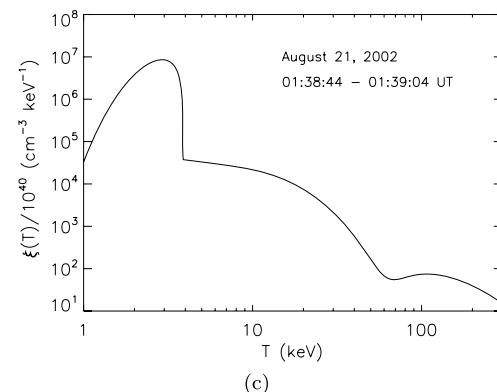
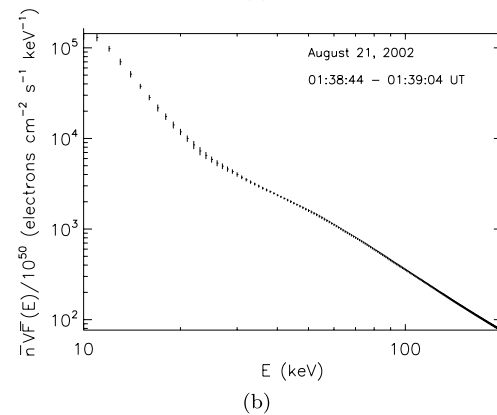
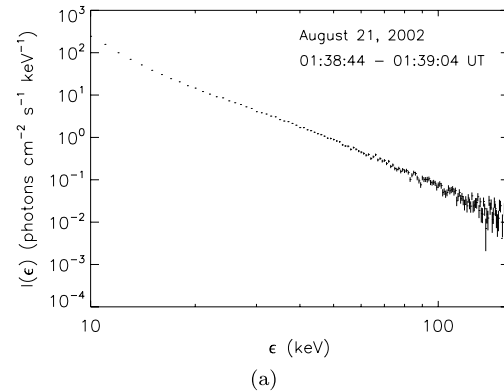
$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt. \tag{221}$$

Fig. 19 August 21 2002 flare recorded by RHESSI in the time interval 01:38:44–01:39:04 UT: **(a)** photon spectrum; **(b)** mean source electron spectrum reconstructed by zero-order Tikhonov regularization; **(c)** reconstruction obtained by inverting the low-energy part of the electron spectrum with the projected-Landweber method and the high-energy part with first-order Tikhonov regularization and by connecting the two restorations; **(d)** cumulative regularized residuals (*solid*) for the method with positivity (*upper panel*) and Tikhonov regularization (*lower panel*) compared to the statistical bound $\pm 3/\sqrt{k}$ (*dashed*)

Therefore condition (203) is compatible with the mean source electron spectrum with an asymptotic ($E \rightarrow \infty$) electron spectral index $\delta > 0$ (corresponding to a photon spectral index $\gamma > 1$).

The reconstruction procedure described in the previous section has been applied to three photon spectra observed by RHESSI corresponding to three different flares. Figure 19(a) shows the photon spectrum corresponding to the August 21 2002 flare in the time interval 01:38:44–01:39:04 UT, while Fig. 19(b) shows the corresponding averaged electron spectrum obtained by using zero-order Tikhonov regularization.

The low energy part of this spectrum (11–24 keV) has been inverted by means of the Landweber iterative scheme with positivity for 10^5 iterations while the high energy part (50–189 keV) has been inverted by using first-order Tikhonov regularization with boundary conditions (again the two electron energy ranges correspond to the intervals where $\bar{F}(E)$ is optimally fitted by a thermal component and a power-law respectively). The two reconstructed $\xi(T)$ are connected together and plotted in Fig. 19(c) while the cumulative residuals contained in Fig. 19(d) show that the reconstruction is statistically reliable. At small T , $\xi(T)$ presents a peak at $T \sim 2.9$ keV, FWHM ~ 1.5 keV and $\langle T \rangle \sim 2.8$ keV (the temperature provided by best-fitting $\bar{F}(E)$ is 2.6 keV). In order to study the compatibility of this spectrum with a single-temperature thermal interpretation we produced a synthetic $\bar{F}(E)$ corresponding to a δ -function peaked at 2.9 keV and inverted it with the same projected Landweber method applied to the same electron energy range. The restoration presented a FWHM of around 1.5 keV showing that this flare can be reliably interpreted according to an isothermal model. At higher temperatures, $\xi(T)$ presents a dip between 60 and 70 keV and an asymptotic power-law-like behavior with $\alpha \sim 2.8$ (this value is in accordance with the fact that the asymptotic electron spectral index is $\delta \sim 2.3$). In order to study the statistical relevance of the non-monotonic structure in the 60–70 keV temperature range, in Fig. 20 we constructed the confidence strip for the regularized $\xi(T)$ [63, 67] by means of repeated inversions using different realizations of the data set and by superimposing the corresponding regularized solutions. The strip results to be notably large in correspondence with the dip, thus allowing to interpret this structure in terms of a ‘plateau’, a broken-power-law or even a simple power-law behavior.



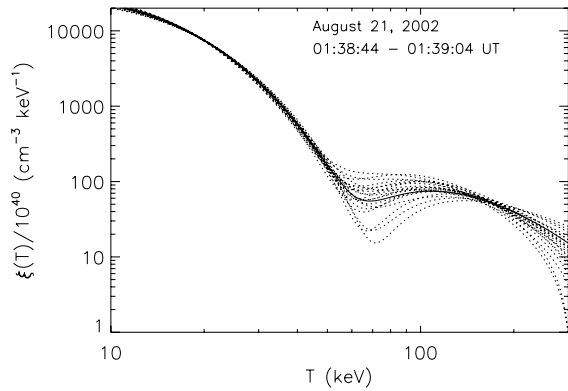


Fig. 20 The confidence strip for the regularized $\xi(T)$ at high T , corresponding to the August 21 2002 flare in the time interval 01:38:44–01:39:04 UT. The strip has been obtained by repeated inversions of $\bar{F}(E)$ in Fig. 19(b) using 20 different random realizations of this data set. The inversion method is first-order Tikhonov regularization

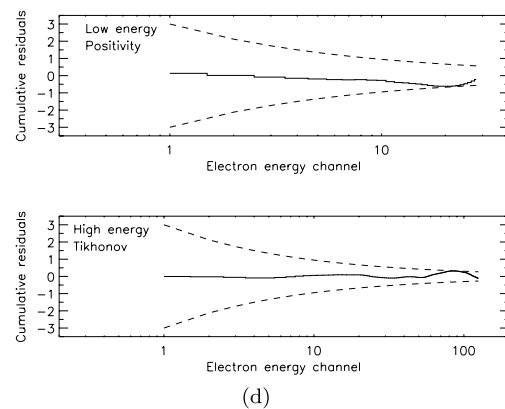
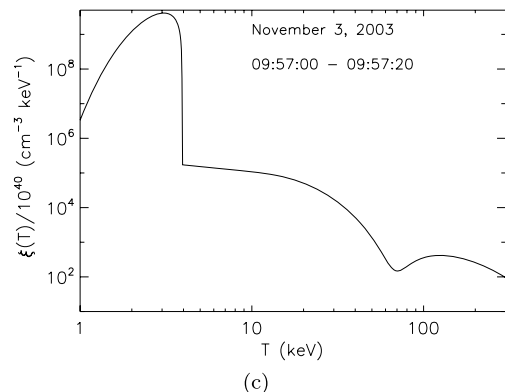
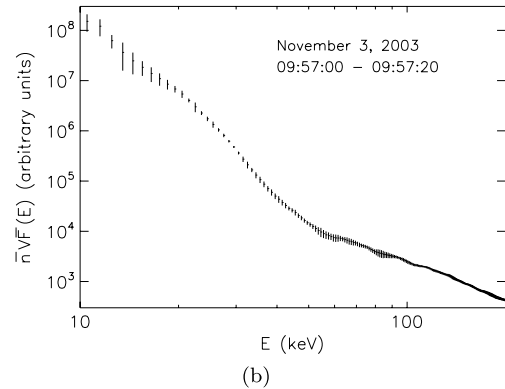
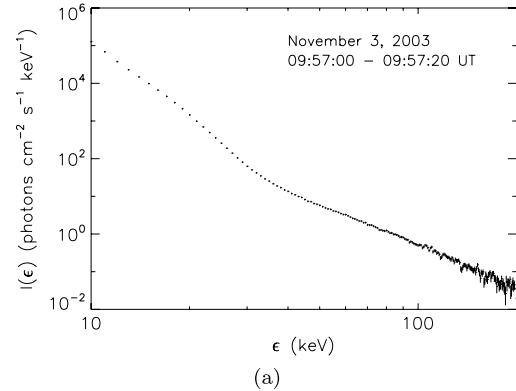


Fig. 21 November 3 2003 flare recorded by RHESSI in the time interval 09:57:00–09:57:20 UT: **(a)** photon spectrum; **(b)** mean source electron spectrum reconstructed by zero-order Tikhonov regularization; **(c)** reconstruction obtained by inverting the low-energy part of the electron spectrum with the projected-Landweber method and the high-energy part with first-order Tikhonov regularization and by connecting the two restorations; **(d)** cumulative regularized residuals (solid) for the method with positivity (upper panel) and Tikhonov regularization (lower panel) compared to the statistical bound $\pm 3/\sqrt{k}$ (dashed)

An analogous procedure has been applied for the analysis of the photon spectrum in the time interval 09:57:00–09:57:20 UT of the November 3 2003 flare (see Fig. 21(a) for the photon spectrum and Fig. 21(b) for the inverted averaged electron spectrum). $\bar{F}(E)$ has been inverted with the positivity method in the 13.5–40.5 keV range and with first-order regularization in the 56.5–180.5 keV range. The reconstructed $\xi(T)$ in Fig. 21(c) presents a peak at $T \sim 3.1$ keV with FWHM ~ 1.4 keV and $\langle T \rangle \sim 2.8$ keV (the best-fitting temperature is 3.2 keV). As for the previous flare, in this case a single-temperature interpretation of this part of the spectrum is acceptable. At higher T there is a feature in the range 70–80 keV, which is more pronounced than the one in the August 21 2002 flare (although, also in this case, the confidence strip at these temperatures is very wide). The asymptotic α is around 3.1, which must be compared with an asymptotic δ in $\bar{F}(E)$ of around 2.7 (once more, the asymptotic relation $\alpha \sim \delta + 0.5$ is satisfied). The cumulative residuals in Fig. 21(d) show that these results are statistically reliable.

Things are notably different in the case of the photon and electron spectra in Fig. 22(a) and (b) respectively, corresponding to the time interval 00:30:00–00:30:20 UT of the July 23 2002 flare. This $\bar{F}(E)$ fails the derivative test at several points in the low-energy range. We have computed

the first five derivatives of $\bar{F}(E)/E$ and found failures of the test for different points in the second, third, fourth, and fifth derivative. Figure 22(c) contains, for example, the third

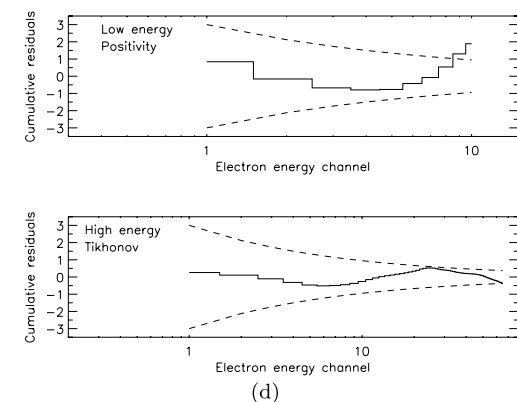
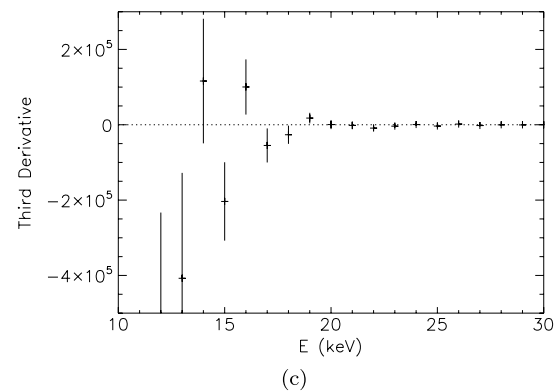
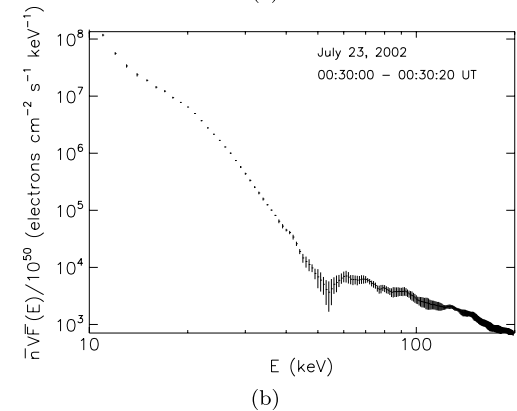
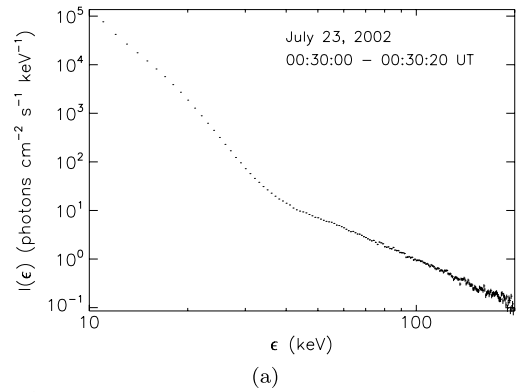
Fig. 22 July 23 2002 flare recorded by RHESSI in the time interval 00:30:00–00:30:20 UT: **(a)** photon spectrum; **(b)** mean source electron spectrum reconstructed by zero-order Tikhonov regularization; **(c)** numerical third derivative of $\bar{F}(E)/E$ with corresponding statistical errors; **(d)** cumulative regularized residual (*solid*) for the method with positivity (*upper panel*) in the case of 10^6 iterations and Tikhonov regularization (*lower panel*) compared to the statistical bound $\pm 3/\sqrt{k}$ (*dashed*)

derivative which should be negative for a thermal spectrum and is in fact positive (with statistical significance) at 16 and 19 keV. By applying the constrained-Landweber method to $\bar{F}(E)$ at low energies (for example between 12 and 21 keV) we found that the cumulative residuals never present the expected random walk, even for huge numbers of iterations (the residuals in Fig. 22(d), upper panel, correspond to 10^6 iterations). This behavior seems to suggest that a thermal interpretation of this photon spectrum could be problematic, although we also observe that this photon data set probably suffers a notable pulse pile-up, which may imply artefacts in the reconstruction of $\bar{F}(E)$. For the high-energy part of the spectrum, first-order regularization provides a power-law-like $\xi(T)$ with $\alpha \sim 2.7$ at high T (δ for this spectrum is around 2).

6.5 Conclusions

The inference of differential emission measure functions $\xi(T)$ from observed photon spectra $I(\epsilon)$ with a realistic bremsstrahlung cross section is substantially more difficult than the single-step inversion analysis of Piana, Brown, and Thompson [66] based on an approximate Q . A proper procedure involves two inverse problems. The first of these is the inversion of $I(\epsilon)$, through an exact solid-angle-averaged bremsstrahlung cross section kernel and a zero-order Tikhonov regularization method, to obtain the mean source electron spectrum $\bar{F}(E)$. The second uses an approach for inverting $\bar{F}(E)$ which involves the application of a projected algorithm with positivity constraint in the inversion of the low energy part of the spectrum and of a first-order regularization method with boundary conditions in the inversion of the high energy part of the spectrum. The main findings are:

- the approach correctly identifies certain properties of $\bar{F}(E)$ (such as bumps or energy cut-offs) as being inconsistent with any physical $\xi(T) \geq 0$;
- the use of the positivity constraint allows us to obtain a satisfactory temperature resolution in the recovery of δ functions while the use of first-order regularization with boundary conditions provides reliable reconstructions for smooth forms such as power-laws;
- application of the method to observed RHESSI photon spectra has revealed two cases in which the recovered $\xi(T)$ is spectrally consistent with a roughly isothermal low-temperature plasma plus a very broad form of $\xi(T)$



at high temperatures. In a third case, a spectrum from the July 23 2002 flare, the reconstruction method at low temperatures produces unacceptable large residuals. This re-

sult is in accordance with the fact that the same spectrum fails to satisfy the derivative test which verifies the compatibility with a purely thermal interpretation. Possible physical motivations for this behaviour are still unclear and, for example, may be related to the fact that this flare produced spectra which deviate from a power-law behaviour in a manner consistent with non-uniform ionization [47]. However we also observe that the spectrum used in our analysis suffers a notable pulse pile-up which may imply artefacts in the analysis results.

The availability of a reconstruction approach for addressing the difficult inverse problem of restoring $\xi(T)$ from reconstructions of $\bar{F}(E)$ may have important consequences in the analysis and interpretation of RHESSI spectra. In future research we will apply the method to study the influence of albedo effects on the modification of the differential emission measure and to deduce important physical properties on the thermal plasma from the reconstructed $\xi(T)$.

7 Application to Solar Physics: Imaging-Spectroscopy

7.1 From Spectroscopy to Imaging-Spectroscopy

The high X-ray energy resolution achievable with RHESSI's hardware and the consequent measurements of the precise shape of the X-ray continuum, together with suitable mathematical tools to stably invert the bremsstrahlung equation, allow to provide unique information on the spectrum of the accelerated electrons and on the heated plasma. However, the purposes of the RHESSI mission go further: the new approach is to combine, for the first time, high-resolution spectroscopy in X-ray and γ -ray with high-resolution images, so that a detailed energy spectrum can be obtained at each point of the map. In fact, RHESSI produces hard X-ray and γ -ray images with the finest angular and spectral resolutions ever achieved [57]; imaging spectroscopy analysis of this data is a powerful tool with which to explore the underlying physics of particle acceleration and transport in solar flares. Traditional imaging spectroscopy methods (e.g., [30]) start by constructing two-dimensional maps of the source at different count energies by applying image processing algorithms (e.g., back projection, CLEAN, Maximum Entropy, Pixon). This results in a series of images which are consistent both with the broad assumptions of the particular algorithm used and with the imaging information contained in the data. Spatially resolved count spectra are then obtained from this set of images by selecting particular regions in the field of view and comparing the intensity in those regions as a function of count energy. Finally, the corresponding spatially resolved electron spectra are constructed by applying regularized spectral inversion methods (e.g., [18]) to the spatially resolved count spectra.

In this section we introduce a new approach to imaging spectroscopy which is optimized to the distinctive way in which spatial information is encoded in the RHESSI data. The RHESSI instrument employs a rotation modulation collimator (RMC) imaging technique, in which rapid time variations of the detected counts are effected by the placement of a set of RMCs, each with a different pitch, in front of each detector. Spatial information is encoded in the temporal modulation of the detected flux [40]. As the RMC rotates, the amplitude and phase of this pseudo-periodic modulation over a limited range of angles provides a direct, calibrated measurement of a single Fourier component of the source distribution. Such a Fourier component is termed a *visibility* [70], and is the same quantity provided by the correlated signal from a pair of antennas in a radio interferometer. In this case, the spatial frequency of the measured visibility is determined by the angular resolution of the RMC and its instantaneous orientation. Combining data from multiple RMC's at a variety of orientations, the set of visibilities can then be used to reconstruct the spatial distribution of the source. Since visibilities can be summed linearly, this perspective on the data provides a convenient basis for combining data from multiple rotations into a tractable number of visibility measurements with well-defined statistical errors.

The "traditional" approach to imaging spectroscopy, in which images at different count energies are "stacked" and compared, not only fails to take full advantage of the particular nature in which spatial information is contained in the RHESSI data, it also has two significant drawbacks:

- while imaging algorithms can reduce statistical and point-response artifacts in each image, they are completely ineffective in smoothing along the count energy direction, so that recovered images corresponding to adjacent energy bins can exhibit substantial differences;
- owing to these energy-dependent fluctuations, the determination of the count spectrum at a particular point (x, y) (or, more accurately, a particular region $[x \pm \Delta x, y \pm \Delta y]$) in the source image can be problematic, as is the determination of the statistical error on the count flux. As is well-known (e.g., [20]), such noise-related spectral variations are greatly magnified upon performing a spectral inversion to obtain the corresponding electron spectrum.

In addition, as with any indirect imaging technique, the incomplete spatial frequency sampling results in spatial side-lobes in the point response function which can cause contamination by neighboring sources. Further, the statistical noise from all source components contributes to the noise in each selected region of the source.

It is crucial to recognize that it is not the observed counts (or even photons) that are of interest *per se*, but rather the electrons that produce them: the real science goal is to obtain physically plausible (i.e., "sufficiently smooth") electron

spectra throughout the source. Our new method of imaging spectroscopy analysis therefore involves an interchange of two steps in the data processing chain. First, one applies a count to electron inversion algorithm to obtain smoothed electron spectra at each point in the *spatial frequency domain*. Once such electron flux visibility spectra have been obtained, they can be processed using standard image reconstruction techniques to yield electron flux images for the entire field of view. Since the electron flux visibility spectra are regularized, so also are the corresponding electron flux spectra at each location in the image. This renders these spatially-resolved spectra more suitable for further analysis.

We perform the count to electron inversion step using the familiar Tikhonov regularization technique that has proven so effective (see Sects. 3 and 4) in the inference of spatially-integrated electron spectra $\bar{F}(E)$ from observations of spatially-integrated count (or photon) spectra $I(\epsilon)$. Applied to visibilities, the Tikhonov regularization method forces smoothness in the inferred electron visibility spectra at each point in the spatial frequency domain and thus enhances real features that persist over a relatively wide energy band, while suppressing noise-related features that show up only over a narrow range of energies. The combination of visibility data and Tikhonov regularization methodology therefore allows us to derive the most robust information on the spatial structure of the electron flux spectrum image, the key quantity of physical interest.

7.2 Methodology

Define a Cartesian coordinate system (x, y, z) such that (x, y) (in units of arcseconds) represents a location in the image plane and z (cm) represents distance along the line of sight into the source. Let the local density of target particles along the line-of-sight depth $\ell(x, y)$ (cm) be $n(x, y, z)$ (cm^{-3}) and let the differential electron flux spectrum (electrons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$) at the point (x, y, z) in the source be $F(x, y, z; E)$.

Since the source is optically thin, the relation between $F(x, y, z; E)$ and the corresponding observed *photon spectrum image* $I(x, y; \epsilon)$ (photons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1} \text{arcsec}^{-2}$) is

$$I(x, y; \epsilon) = \frac{a^2}{4\pi R^2} \int_{\epsilon}^{\infty} \int_0^{\ell(x, y)} n(x, y, z) \times F(x, y, z; E) Q(\epsilon, E) dz dE, \quad (222)$$

where $a = 7.25 \times 10^7 \text{ cm arcsec}^{-1}$ at $R = 1 \text{ AU}$ and $Q(\epsilon, E)$ ($\text{cm}^2 \text{keV}^{-1}$) is the cross section for emission of a photon at energy ϵ . Here we consider the isotropic bremsstrahlung cross section $Q(\epsilon, E)$ given by equation (3BN) in [44] and reported in Sect. 3.1 (see (110)).

We define the *mean electron flux spectrum* $\bar{F}(x, y; E)$ (electrons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$ at the Sun) by

$$\bar{F}(x, y; E) = \frac{1}{N(x, y)} \int_0^{\ell(x, y)} n(x, y, z) F(x, y, z; E) dz, \quad (223)$$

where the column density (cm^{-2}) at each point (x, y) in the image is given by $N(x, y) = \int_0^{\ell(x, y)} n(x, y, z) dz$. Then, by (222) and (223), we may write

$$I(x, y; \epsilon) = \frac{a^2}{4\pi R^2} \int_{\epsilon}^{\infty} N(x, y) \bar{F}(x, y; E) Q(\epsilon, E) dE. \quad (224)$$

Next we introduce spatial frequencies u and v in the x - and y -directions, respectively, and define the *count visibility spectrum* $V(u, v; q)$ (counts $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$) as the two-dimensional spatial Fourier transform of the *count spectrum image* $J(x, y; q)$ (counts $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1} \text{arcsec}^{-2}$):

$$V(u, v; q) = \int_x \int_y J(x, y; q) e^{2\pi i(ux+vy)} dx dy. \quad (225)$$

The count spectrum and photon spectrum images are related by the instrument's detector response matrix. Hence we may write

$$V(u, v; q) dq = \int_x \int_y \int_q D(q, \epsilon) I(x, y; \epsilon) \times e^{2\pi i(ux+vy)} d\epsilon dx dy, \quad (226)$$

where the dimensionless quantity $D(q, \epsilon)$ is the differential element of the detector response matrix⁵ corresponding to the generation of a count with energy in the energy range $[q, q + dq]$ from a photon in the energy range $[\epsilon, \epsilon + d\epsilon]$.

Combining (224) and (226) gives the rather formidable expression

$$V(u, v; q) dq = \frac{a^2}{4\pi R^2} \int_x \int_y \int_q \int_{\epsilon}^{\infty} [N(x, y) \bar{F}(x, y; E)]$$

⁵The range of ϵ corresponding to a count of energy q is taken to be $[q, \infty)$; only photons of energy $\geq q$ can generate a count of energy q . This (linear) formalism therefore ignores the possibility of the creation of a count of energy q from the arrival at the same detector of two (or more) photons of energy $< q$ within a very short time interval. This "pulse pileup" process is intrinsically nonlinear (the detector response matrix depends on the incoming photon flux) and so cannot be readily accommodated within the present formalism. Our analysis will therefore be restricted to medium-flux events for which pileup is not likely to be significant.

$$\times D(q, \epsilon) Q(\epsilon, E) e^{2\pi i(ux+vy)} dE d\epsilon dx dy, \quad (227)$$

which provides the formal relationship between $[N(x, y) \bar{F}(x, y; E)]$, the quantity of most direct physical interest, and the observed count visibility spectra $V(u, v; q)$. We now introduce the *count cross section* $K(q, E)$ ($\text{cm}^2 \text{keV}^{-1}$) through the expression

$$K(q, E) dq = \int_q^\infty D(q, \epsilon) Q(\epsilon, E) d\epsilon \quad (228)$$

and the *electron flux visibility spectrum* (electrons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$)

$$W(u, v; E) = a^2 \int_x \int_y N(x, y) \bar{F}(x, y; E) e^{2\pi i(ux+vy)} dx dy. \quad (229)$$

With these definitions, and reversing the order of integration with respect to ϵ and E , (227) can be written as the straightforward integral equation

$$V(u, v; q) = \frac{1}{4\pi R^2} \int_q^\infty W(u, v; E) K(q, E) dE. \quad (230)$$

Equation (230) is formally identical to the relation (101) between the spatially-integrated photon spectrum and the source-integrated electron spectrum and so can be solved for the visibilities $W(u, v; E)$ from the observed count visibility spectra $V(u, v; q)$ by applying Tikhonov regularization, that have proven so effective in the solution of (101) for $\bar{n}V\bar{F}(E)$ given $I(\epsilon)$ (see Sects. 3 and 4).

To briefly summarize the Tikhonov methodology in the visibility case, (230) is first discretized in both count and electron energy spaces to yield, at each sampled point (u, v) in the spatial frequency domain, the *data visibility vector* $\mathbf{V}_{[u,v]}$ (the elements of which depend on count energy q) and the *source visibility vector* $\mathbf{W}_{[u,v]}$ (the elements of which depend on electron energy E). These are related through the matrix equation

$$\mathbf{V}_{[u,v]} = \mathbf{K} \cdot \mathbf{W}_{[u,v]}, \quad (231)$$

where \mathbf{K} is the kernel matrix, the elements of which are formed from the values of $K(q, E)$ at the discretized count and electron energy points. Then the *zero-order regularization problem*

$$\|\mathbf{V}_{[u,v]} - \mathbf{K} \cdot \mathbf{W}_{[u,v]}\|^2 + \lambda_{[u,v]}\|\mathbf{W}_{[u,v]}\|^2 = \text{minimum} \quad (232)$$

is solved for $\mathbf{W}_{[u,v]}$ given the prescribed visibility vector $\mathbf{V}_{[u,v]}$ at each sampled point in (u, v) space, using an appropriate value (see below) of the regularization parameter $\lambda_{[u,v]}$. This results in electron visibility spectra that are “smooth” in the sense that the large variations in $W(u, v; E)$

from energy bin to energy bin are suppressed. This technique therefore enhances spatial features (Fourier components) that persist over a wide range of energies, and suppresses (noise) features that exist over only a limited subset of energy bins. Once the electron visibility spectra have been determined, the electron flux spectral image may be determined through the inverse Fourier transform of (229).

7.3 Application to Data

We illustrate the method by applying it to data obtained near the peak of the C7.5 flare of February 20, 2002 (11:06:02–11:06:34 UT), using visibilities from RHESSI RMCs 3 through 9, corresponding to spatial resolutions from ~ 7 to ~ 183 arcseconds. For comparison purposes, we first construct count images by means of the visibility technique and apply the “traditional” imaging spectroscopy approach, in which count images in different energy bands are compared. After discussing the drawbacks of this “traditional” method, we use our new method to obtain more physically useful electron flux maps of the flare.

The “traditional” method begins by converting the X-ray count rate data to a set of visibilities. This requires pre-selecting the number of angular intervals (roll bins) per rotation. The number of roll bins should be maximized to avoid degradation of sensitivity near the edge of the field of view. However, for this application, each roll bin must contain at least one complete modulation cycle to enable the visibility to be well-measured. Using an iterative technique, we maximized the number of roll bins for each detector subject to this constraint and then used a χ^2 analysis to determine statistically acceptable visibilities. Then, since $V(u, v)$ and $V(-u, -v)$ are complex conjugates (see (225)), the visibilities measured at angles separated by 180 degrees are combined to improve the signal-to-noise ratio. Finally, the error bars on the real and imaginary parts of each visibility for each energy channel are computed by propagating the statistical error in the counts through to the calculation of each visibility. The resulting visibilities are used as input to the Maximum Entropy (“MEM-NJIT”) algorithm [10] as implemented in the Solar SoftWare (SSW) package, to produce $80 \text{ arcsec} \times 80 \text{ arcsec}$ maps with 0.4 arcsec pixels. This was done for 16 4-keV wide energy intervals from 10 to 74 keV.

Figure 23 shows some of these count-based images. Two bright features, which we interpret as emission from chromospheric footpoints, are apparent. In addition, there is some evidence for a “strand” of emission linking the two bright features; this we interpret as emission from the coronal region of the magnetic loop linking the footpoints. The lower left panel of Fig. 23 shows the areally-averaged count spectrum (counts $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1} \text{arcsec}^{-2}$) for the northern footpoint region highlighted by the square in each image; this spectrum has been constructed by averaging, for

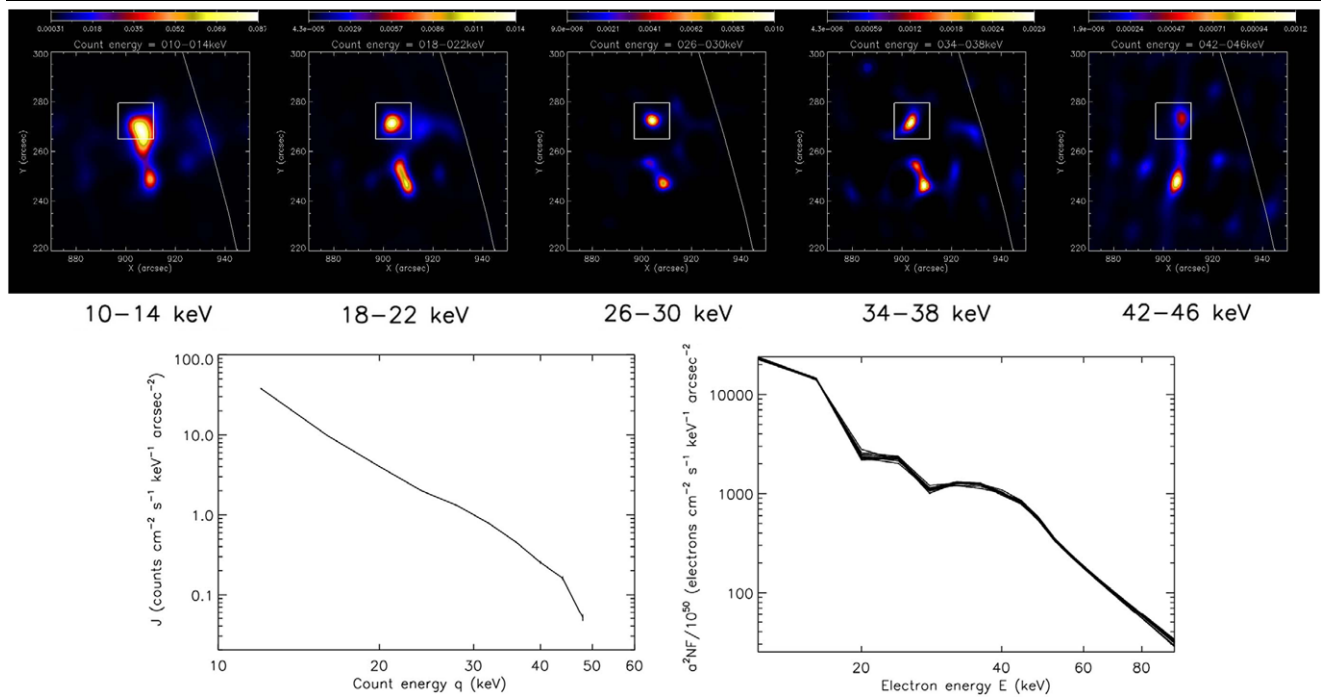


Fig. 23 *Top panels:* Count images for the 20 February, 2002 (11:06:02–11:06:34 UT) event, for the energy intervals shown, produced using the MEM-NJIT algorithm. *Lower panels:* Areal-averaged count spectrum (counts cm⁻² s⁻¹ keV⁻¹ arcsec⁻²;

left) and electron flux confidence strip spectra (electrons cm⁻² s⁻¹ keV⁻¹ arcsec⁻²; *right*) for the footprint region highlighted in the images

each energy channel, the intensities of the pixels that constitute the highlighted region (to get the total count spectrum for the region [counts cm⁻² s⁻¹ keV⁻¹], simply multiply by the area of the region, in this case 14.4 × 14.4 = 207.36 arcsec²). The error bars have been computed as the combination of a count (Poisson) error plus background noise. The lower right panel shows the recovered electron flux spectrum *confidence strip* (i.e., a series of realizations of the electron flux spectrum, each based on a different noisy realization of the data; see Sect. 4.3) for this feature. Each electron spectrum realization was obtained by inverting the count spectrum using the zero-order regularization method applied in Sects. 3 and 4 for spatially integrated spectroscopy.

The count spectrum is conspicuously weak in the 26–30 keV image. This leads to a relatively flat count spectrum in this range (lower left panel of Fig. 23) and hence (since the electron spectrum is, crudely, related to the derivative of the photon spectrum—[12]) to a dip in the recovered electron spectrum (lower right panel) for this feature. Although such spectral dips have been inferred for *spatially-integrated* electron spectra (e.g., [67]), the spatially-integrated count spectra on which such spatially-integrated electron spectra are based are not subject to the imaging artifacts that render suspect the count spectra determined for a particular spatial region. Therefore it is possible, or indeed likely, that the feature in Fig. 23 is *not* real, but rather an artifact imposed by

isolating attention on a limited range of spatial coordinates, rather than on the overall patterns (Fourier components) of emission present in the spatially-integrated emission.

As discussed in Sect. 7.1, because spatial information is fundamentally encoded by RHESSI in Fourier components, rather than in “pixels,” a more cogent approach to imaging spectroscopy involves performing the count to electron inversion step in the *spatial frequency domain*, i.e., on the visibility data. By focusing on the information in distinct Fourier components, we remove the deleterious effects of imaging artifacts that are evident in the more “traditional” approach to imaging spectroscopy.

Figure 24 shows the amplitude (upper panels) and phase (lower panels) of the count visibilities $V(u, v; q)$ (counts cm⁻² s⁻¹ keV⁻¹) for the same event and time interval as Fig. 23, for three count energy ranges. The amplitude of the visibilities generally increases with increasing grid pitch (decreasing value of the corresponding spatial frequency $\sqrt{u^2 + v^2}$). Highlighted by a red star in each plot in Fig. 24 is the (somewhat arbitrary) point $(u^* = -0.0042 \text{ arcsec}^{-1}, v^* = -0.0422 \text{ arcsec}^{-1})$; this point corresponds to a spatial periodicity $1/\sqrt{u^{*2} + v^{*2}} = 23.6 \text{ arcsec}$, which is the spatial periodicity corresponding to (i.e., twice the angular resolution of) RHESSI grid 4. The top panels of Fig. 25 show the amplitude $|V(u^*, v^*; q)|$ and phase $\text{Arg}(V[u^*, v^*; q])$ of the differential count visibility spectrum for this representative point in the spatial frequency domain.

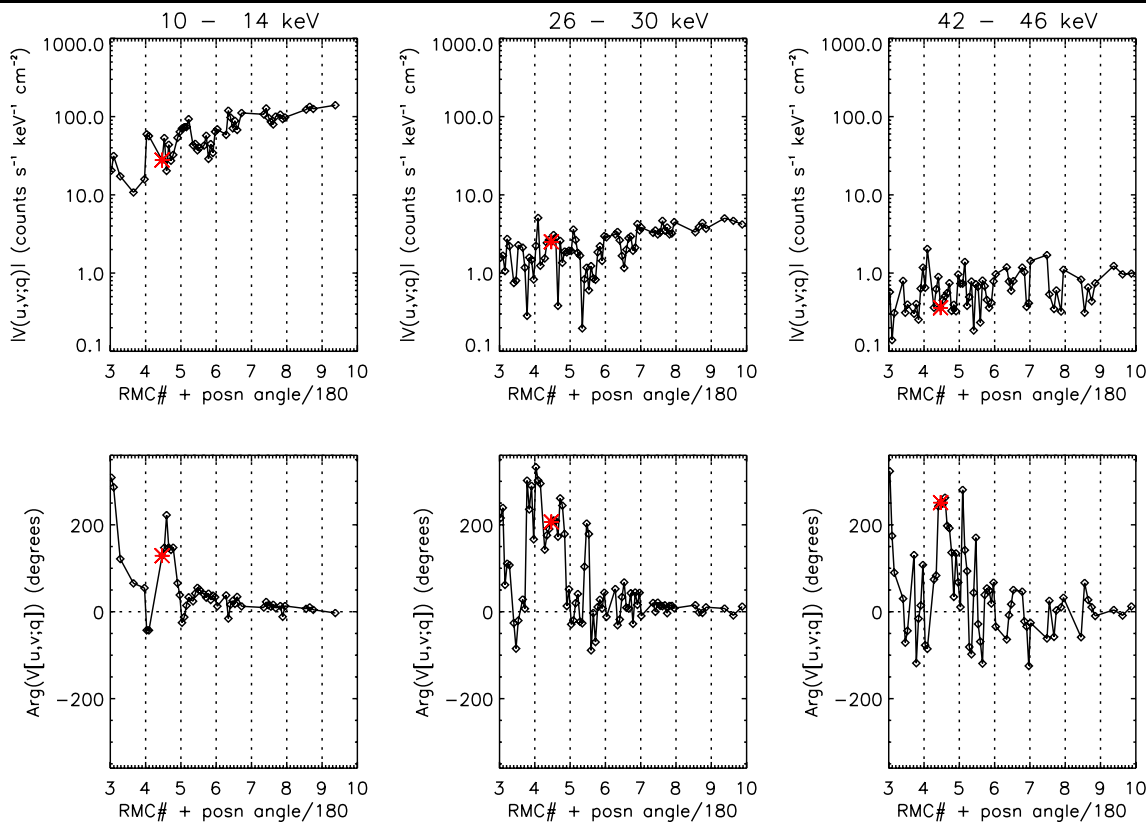


Fig. 24 (Color online) The observed count visibilities (amplitude and phase) for three representative energy bands. In each panel, the region between each pair of dotted vertical lines represents measurements with a single RMC at a fixed spatial period, with the orientation of the measurement increasing from 0 to 180 degrees; successive regions correspond to different RMC's. The angular resolutions for the re-

gions increase in a geometric progression (with a ratio of $\sqrt{3}$ between successive regions), and span 7 arcseconds on the left to 183 arcseconds on the right. The point ($u^* = -0.0042 \text{ arcsec}^{-1}, v^* = -0.0422 \text{ arcsec}^{-1}$), highlighted with a *red star* in each plot, is used in the illustrative spectral plots of Fig. 25

In order to preserve the inherent linearity of the process, a polar to rectangular transformation was performed to convert the amplitude and phase information into real and imaginary components $\text{Re}\{V(u^*, v^*; q)\}$ and $\text{Im}\{V(u^*, v^*; q)\}$. Each of these components was then subjected to the regularized inversion analysis of (232) to obtain the real and imaginary parts of the corresponding (regularized) electron visibility spectrum $W(u^*, v^*; E)$ at the point (u^*, v^*) . Through an inverse rectangular to polar transform, we then recover the amplitude and phase of the electron visibility spectrum $W(u^*, v^*, E)$ at this particular point, as shown in the bottom panels of Fig. 25.

Repeating this regularized inversion process for each sampled point in the (u, v) plane (using a value of the regularization parameter $\lambda[u, v]$ appropriate⁶ to each sampled

(u, v) point), we arrive at complete information on the electron flux visibility spectrum. This information is presented in Fig. 26 in the same format of Fig. 24.

We can now use the set of electron flux visibility spectra to construct electron spectral flux images in each energy range. Images of the electron flux spectrum $\bar{F}(x, y; E)$, recovered by applying the MEM-NJIT algorithm, are shown in Fig. 27. These images represent the quantity of key physical interest.

Figure 27, like Fig. 23, shows evidence for two footpoints, again connected by a “strand” of coronal flux. To the extent that variations in count intensity are a consequence of data noise, the regularization algorithm used to develop the electron flux images of Fig. 27 removes such irregularities, resulting in a more coherent variation of source structure with energy. Consequently, the electron flux images vary much more smoothly with energy, and the coronal “strand” is more persistent at low energies.

The footpoints in the electron images are seen to persist up to electron energies $\sim 75 \text{ keV}$, an energy significantly greater than the maximum photon energy used. As pointed

⁶The value of the regularization parameter $\lambda_{[u^*, v^*]}$ was chosen using the “ 3σ cumulative residual criterion” approach discussed in detail in Sect. 6.2; in general, such a procedure for determining $\lambda_{[u, v]}$ results in more faithful representations of electron flux spectra than the commonly-used “discrepancy principle”.

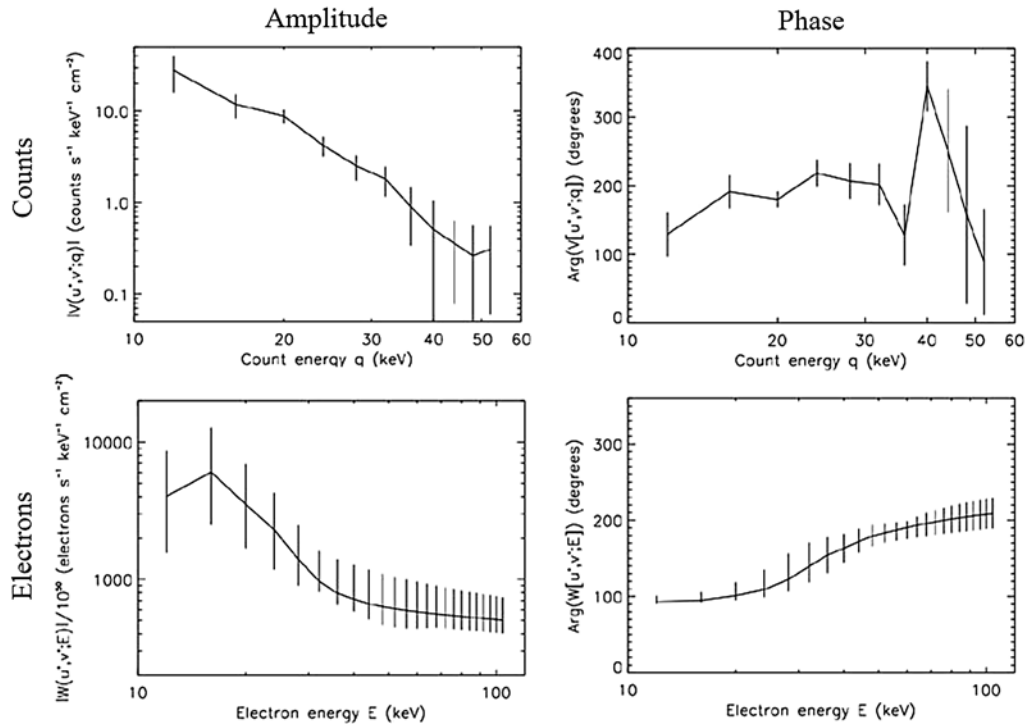


Fig. 25 Top panels: Amplitude (left; counts $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$) and phase (right; degrees) of the count visibility spectrum $V(u^*, v^*; q)$ at the point ($u^* = -0.0042$, $v^* = -0.0422$) in the spatial frequency domain. Bottom panels: Amplitude (left; in units of 10^{30} electrons

$\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$) and phase (right; degrees) of the corresponding electron spectrum visibilities $W(u^*, v^*; E)$ at the same (u, v) point, obtained through regularized inversion of (230) using the zero-order Tikhonov method

out by Kontar *et al.* [48], information on the electron spectrum at high energies is indeed contained in the photon spectrum at lower energies, and can be faithfully extracted using the Tikhonov regularization procedure.

It is instructive to reconstruct the count images corresponding to the regularized electron spectral flux images in Fig. 27 and compare them with the original spatial images obtained through processing of the raw count visibility data using the MEM algorithm. This comparison is presented in Fig. 28. The top row of figures shows the recovered count images at the energies shown, while the bottom row reproduces the original count-based images from Fig. 23. The original count-based images (lower row of images in Fig. 28) show evidence principally of a double-footpoint structure, with some additional evidence for an extension of the emission into the region between the footpoints (see, e.g., the 18–22 keV and 42–46 keV images). However, there is no clear systematic variation with count energy q , either of the intensity of this “strand” emission or of the relative intensity of the two footpoints. By contrast, the count images deduced from the regularized electron flux images (upper row of images in Fig. 28) show much more clearly the evolution of the spatial structure with energy. The “strand” of emission between the footpoints is clearly evident up to 30 keV, but diminishes rapidly at higher energies, and the

relative intensity of the two footpoint sources is more independent of count energy q . These physically plausible enhancements in the image structure are recovered through use of the visibility-based regularized inversion technique, because of its inherent requirement that the source structure vary smoothly from one electron energy E to the next. This requirement in turn forces the count images to change more smoothly with count energy q than do the images deduced directly from the (noisy) data.

7.4 Physical Implications of the Results

The electron spectral flux images of Fig. 27 are quite plausibly interpreted in terms of the collisional thick target model [12] of hard X-ray emission in solar flares.

Consider three different spatial subregions in the source, labelled in Fig. 29. Two of these regions correspond to the footpoint sources visible at higher energies and the other one to similarly-sized regions located approximately midway between the two footpoints. The lower panel of Fig. 29 shows the areally-averaged⁷ electron flux spectra (electrons

⁷To get the total count spectrum for each region [counts $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1}$], simply multiply the areally-averaged spectrum by the area of that region, viz. $14.4 \times 14.4 = 207.36 \text{ arcsec}^2$ (Footpoint 1),

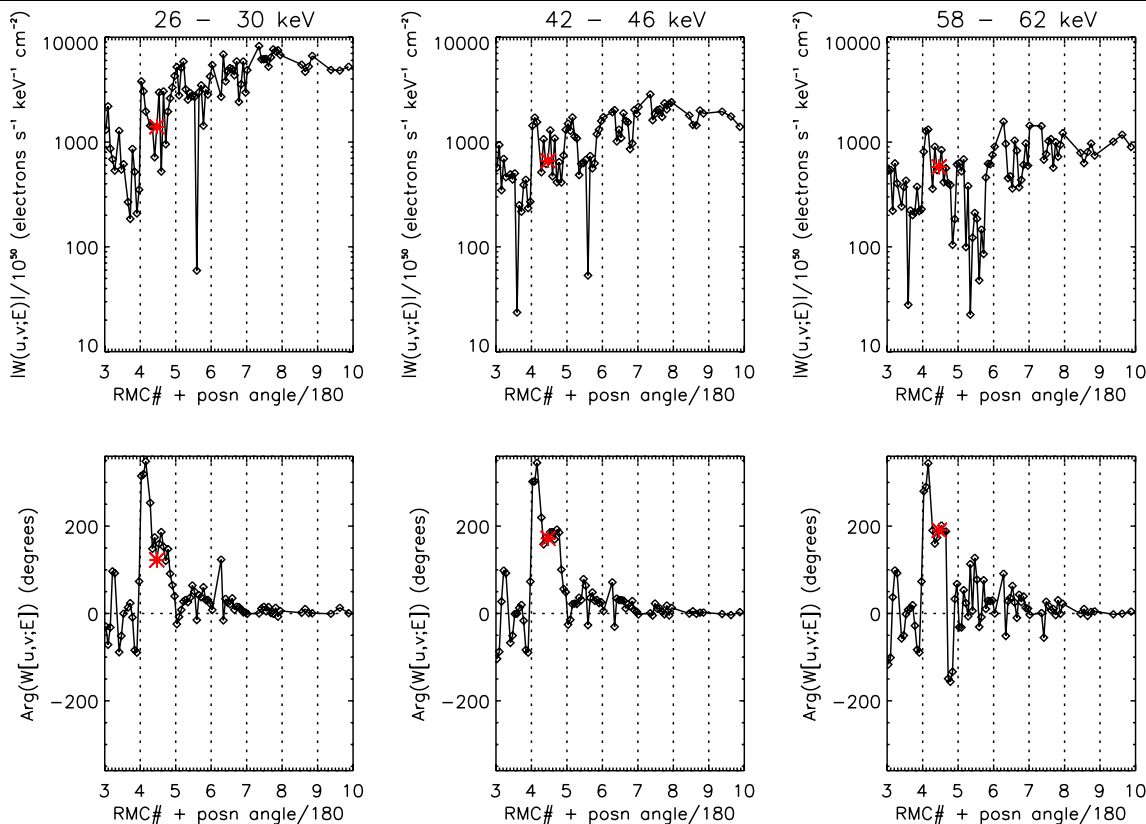


Fig. 26 Recovered regularized electron flux visibilities (scaled by 10^{-50}), presented as in Fig. 24. The sharp minima in the amplitude plots are real, and reflect the two-component nature of the source. The y-intercept to the right is determined by the total flux, while the rate

at which the amplitudes fall off (to the left) is determined by the size of the sources, and reflects, for example, the larger source size at 26–30 keV. The broad similarity of the phase plots reflects the broadly similar location of the sources at various electron energies

$\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1} \text{arcsec}^{-2}$), for each of these three subregions. The region labeled “Footpoint 1” is identical to the region highlighted in Fig. 23; comparison of the electron flux spectra for this region (Fig. 23 and 29) show that the “dip” at $\sim (26\text{--}30)$ keV obtained using the “traditional” approach to imaging spectroscopy is indeed an artifact of the data truncation and overspill issues associated with identification of the flux in a local spatial region; the real electron flux spectrum in this region is smooth and monotonically decreasing with energy E .

At low energies $E \approx 60$ keV, the electron flux at the more southern footpoint (Footpoint 2) is much smaller than that at the more northern footpoint (Footpoint 1). However, the spectrum of Footpoint 2 is very hard ($\delta \approx 1$) and by ~ 60 keV the electron flux at each footpoints has become roughly equal, as is apparent from the spatial images.

Above $E \sim 40$ keV, the spectra corresponding to the two footpoint regions are visibly flatter (harder) than that corresponding to the region between these footpoints. Such a re-

sult is qualitatively consistent with the acceleration of electrons in a source midway between the footpoints, and the subsequent propagation of these electrons to the footpoints. To concentrate the observed degree of spectral hardening in the footpoints constrains the intervening column density to an upper limit $N < E^2/2K \sim 2 \times 10^{17} [E(\text{keV})]^2 \sim 3 \times 10^{20} \text{cm}^{-2}$ (here $K = 2\pi e^4 \Lambda \approx 2.6 \times 10^{-18} \text{cm}^2 \text{keV}^2$, where e is the electronic charge and Λ the Coulomb logarithm). This in turn establishes an upper limit on the coronal density $n \sim N/d$, where d is the distance between the coronal source and the footpoint parallel to the guiding magnetic field. The plane-of-sky projected distance between the “middle” and “footpoint” sources is $\sim 10 \text{arcsec} \sim 7 \times 10^8 \text{cm}$. Assuming a semicircular geometry for the loop connecting the footpoints, $d \sim (\pi/2)$ times this projected distance, i.e., $\sim 10^9 \text{cm}$. We hence infer that the coronal density $n < 3 \times 10^{11} \text{cm}^{-3}$, an entirely reasonable constraint.

7.5 Summary

We have developed, and illustrated the effectiveness of, a new approach to solar hard X-ray imaging spectroscopy. In this approach, two-dimensional Fourier transforms of the

$22.8 \times 9 = 205.2 \text{arcsec}^2$ (Middle), and $14.4 \times 14.4 = 207.36 \text{arcsec}^2$ (Footpoint 2), respectively.

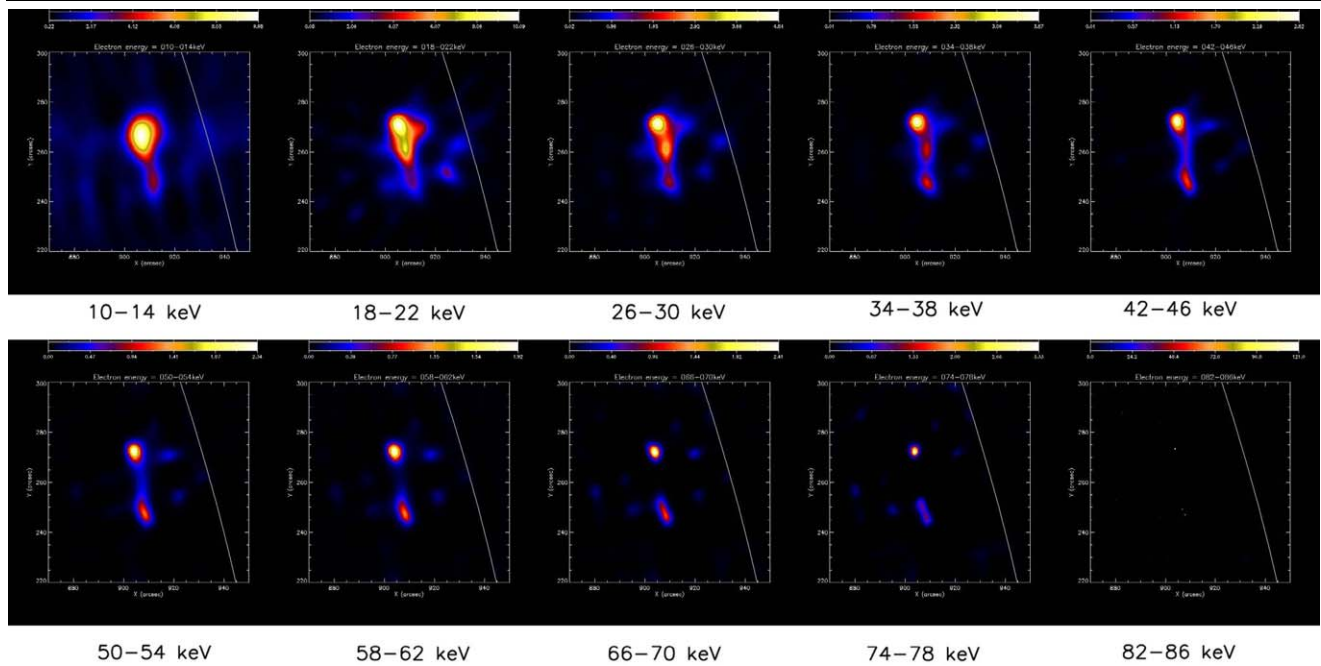


Fig. 27 Electron flux spectral images corresponding to the regularized electron flux spectral visibilities of Fig. 26, obtained through application of the MEM-NJIT algorithm [10]

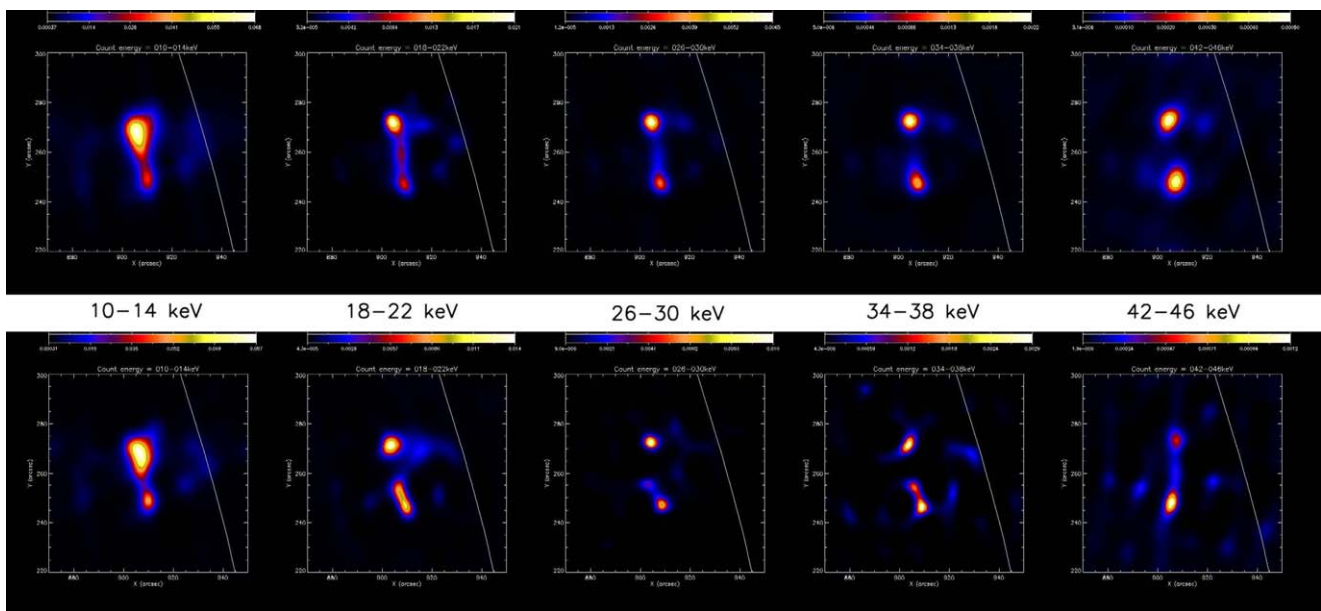
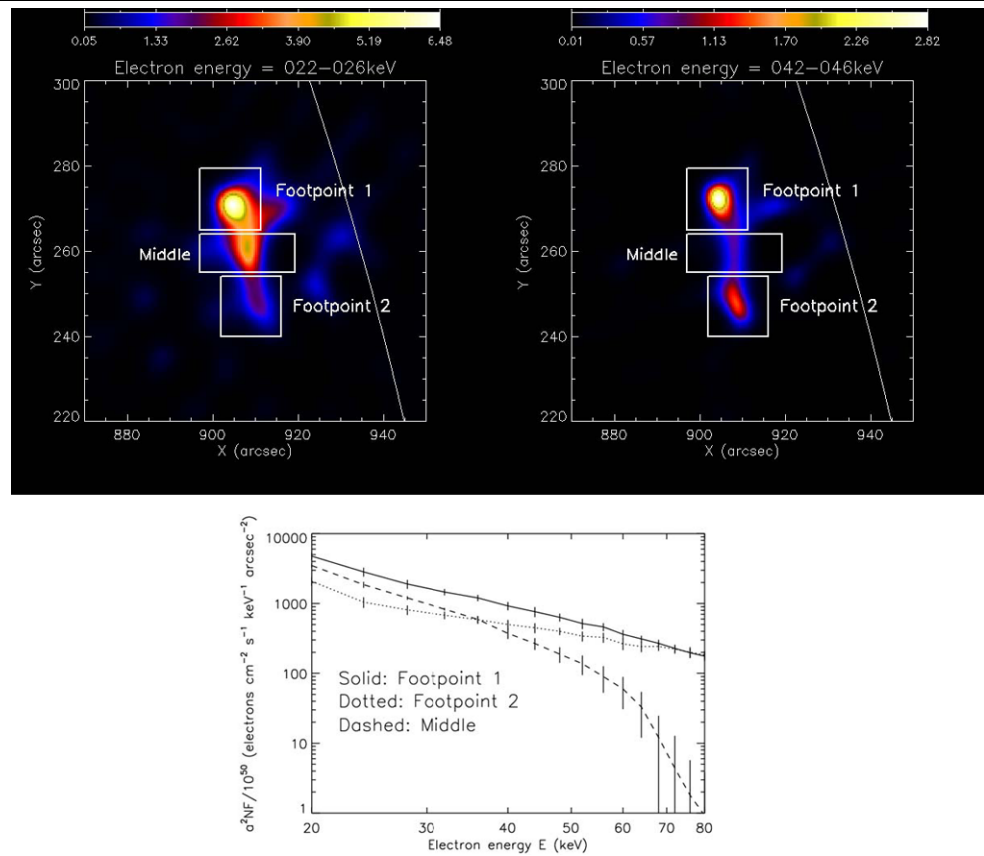


Fig. 28 *Top panels*: Regularized count-based images corresponding to the electron flux spectral images of Fig. 27, compared with (*bottom panels*) the original images from Fig. 23

image in the count domain are transformed, through a regularized inversion procedure that enhances features that persist over a range of energy channels, into Fourier transforms of the electron flux maps. A final image reconstruction based on an inverse Fourier transform then gives the electron flux maps themselves. Because data obtained through rotating modulation collimator instruments such as RHESSI are con-

centrated into a relatively small number of discrete Fourier components (“visibilities”), this approach is highly effective at analyzing such data, and results in recovered spectra that are determined more precisely than with a method that involved regularized inversion of the count spectrum within a spatial subregion of the source (which necessarily involves a combination of spatial Fourier components). Application

Fig. 29 *Top panels:* Electron images in the energy ranges 22–26 keV and 42–46 keV, respectively. Three sub-regions of interest are labeled on each image. Two of these correspond to bright footpoint-like sources and one to a region midway between the footpoints. *Bottom panel:* Areal-averaged electron flux spectra (electrons $\text{cm}^{-2} \text{s}^{-1} \text{keV}^{-1} \text{arcsec}^{-2}$) for each of the three subregions shown



of the method to a flare on February 20, 2002 yielded a series of electron flux images. Varying smoothly with energy, these images in turn permitted recovery of smooth, regularized, electron flux spectra at different regions in the source. Such smooth, regularized, electron flux spectra contrast with those obtained using the more “traditional” approach to imaging spectroscopy (e.g., Fig. 23), in which unphysical features may result from focusing on a particular spatial region.

For the illustrative event studied, the electron spectra at the two bright chromospheric footpoints evident in the images were systematically harder than the spectrum obtained at similarly-sized region between the footpoints. Such a spectral hardening is broadly consistent with collisional modification of an accelerated electron beam if the intervening density is less than $3 \times 10^{11} \text{ cm}^{-3}$.

In future papers, we intend to apply our new technique to a variety of flare events. The resulting sample of electron flux spectrum images provides the required input to the next stage of inquiry, wherein the nature of the physical processes affecting the bremsstrahlung-producing electrons is determined through analysis (e.g., [29]) of variation in the electron flux spectrum throughout the source.

8 Conclusions

This paper wants to be a collection of recent results which show how the regularization theory for linear inverse problems can be applied to recover stable and physically meaningful solutions of real problems in solar X-ray and imaging spectroscopy. The leading wire of this kind of problems is the ill-posedness which characterizes them and the consequent inefficiency of any mathematical tool which does not account for this pathology. The regularization algorithms considered proved to be an effective way to clash the combined effect of the ill-posed nature of the models and the noise which affects the data measurements.

The opportunity to study the problems we faced in the paper has been given by the RHESSI mission, launched by NASA in 2002 with the aim of providing X-ray imaging and spectroscopy with unprecedented spatial and spectral resolution. The X-ray spectra provided by the spacecraft and the electron distribution in the Sun from which they have been produced via a collisional bremsstrahlung process are linked through an integral equation of Volterra type. Several probabilistic kernel have been analyzed, also accounting for the influence of anisotropies in the emission mechanism, and the differences in the corresponding reconstructed electron spectra have been discussed both in simulated and in real

cases to investigate the effects of a more or less realistic (and complicated) cross section.

The output of this model-independent problem has been used as input for a following Fredholm integral equation which, under a purely thermal interpretation of the bremsstrahlung process, describes the electron distribution in the source as a combination of Maxwellian functions weighted by the differential emission measure. This problem can be traced to the inversion of the Laplace transform, which is an extremely challenging problem due to the huge filtering effect of the exponential kernel. In the paper we proposed to split the input electron spectrum in two parts and to apply to each of them a different regularization algorithm which better perform on recovering the corresponding solutions. As in the previous case, tests on both synthetic and real spectra have been provided.

In the last part of the paper we discussed the imaging spectroscopy item, which is one of the main goals of the RHESSI mission as it allows to obtain detailed energy spectrum for different portions of the solar chromosphere. While traditional imaging spectroscopy methods typically require notable computational efforts and affect the reconstructed local electron spectra with unphysical artifacts, we proposed a new method which avoids both drawbacks. The algorithm involves the regularized inversion of spectra of count visibilities, which are calibrated measurements of spatial Fourier components of the source distribution, followed by the application of a Fourier-based imaging reconstruction technique.

The computational methods described in the present paper, all based on the regularization theory for ill-posed problems, represent one possible approach to data analysis in solar X-ray imaging and spectroscopy. However, other techniques can in fact be applied. Inversion by means of statistical methods, for example, may explicitly account for the statistical properties of the noise affecting the data; or, more in general, the use of a Bayesian framework may allow encoding a priori information on the solution or on the model in the probability density functions introduced in the game. In the specific case of RHESSI measurements, a realistic strategy for imaging, could be the use of multiple deconvolution approaches, whereby the data provided by the nine RHESSI detectors are simultaneously processed to produce reconstructions of the source with enhanced spatial resolution. Finally, from the physical viewpoint, RHESSI data, characterized by such a notable observational quality, can be used to investigate more sophisticated astrophysical processes like free-bound recombination or electron propagation in the plasma during flares. The results of these investigations, based on the application of computational tools like the ones described in this paper, may provide new fundamental insights toward the comprehension of the physical processes in the solar atmosphere and may represent the

conceptual basis of new solar missions like Solar Orbiter, in preparation by ESA for the near future.

Acknowledgements The author is most grateful to Prof. Michele Piana and Dr. Anna Maria Massone for the scientific collaboration. This work is partially supported by the Italian national research project Inverse Problems in Medicine and Astronomy, under contract PRIN 2006018748, the grant I/015/07/0 of the Italian ASI/INAF, the Gruppo Nazionale di Calcolo Scientifico (GNCS) and the International Space Science Institute (ISSI) in Bern, Switzerland.

References

1. Abramowitz M, Stegun IA (1965) Handbook of mathematical functions. Dover, New York
2. Balakrishnan AV (1976) Applied functional analysis. Springer, New York
3. Bertero M (1989) Linear inverse and ill-posed problems. In: Hawkes PW (ed) Advances in electronics and electron physics. Academic, New York
4. Bertero M, De Mol C (1996) Super-resolution by data inversion. In: Wolf E (ed) Prog optics XXXVI. Elsevier, Amsterdam
5. Bertero M, De Mol C, Viano GA (1980) The stability of inverse problems. In: Baltes HP (ed) Inverse scattering problems in optics. Topics in current physics, vol 20. Springer, Berlin, pp 161–214
6. Bertero M, Boccacci P, Pike ER (1982) On the recovery and resolution of exponential relaxation rates from experimental data: a singular-value analysis of the Laplace transform inversion in the presence of noise. Proc R Soc Lond A Mat 383:15–29
7. Bertero M, Brianzi P, Pike ER (1985) On the recovery and resolution of exponential relaxation rates from experimental data: Laplace transform inversion in weighted spaces. Inverse Probl 1:1–15
8. Bertero M, De Mol C, Pike ER (1985) Linear inverse problem with discrete data. I: General formulation and singular system analysis Inverse Probl 1:301–330
9. Bertero M, De Mol C, Pike ER (1988) Linear inverse problem with discrete data. II: Stability and regularisation. Inverse Probl 4:573–594
10. Bong SC, Lee J, Gary DE, Yun HS (2006) Spatio-spectral maximum entropy method I: Formulation and test. Astrophys J 636:1159–1165
11. Brianzi P, Frontini M (1991) On the regularized inversion of the Laplace transform. Inverse Probl 7:355–368
12. Brown JC (1971) The deduction of energy spectra of non-thermal electrons in flares from the observed dynamic spectra of X-ray bursts. Sol Phys 18:489–502
13. Brown JC (1972) The directivity and polarisation of thick target X-ray bremsstrahlung from solar flares. Sol Phys 26:441–459
14. Brown JC (1974) On the thermal interpretation of hard X-ray bursts from solar flares. In: Newkirk G (ed) Proc IAU symp on coronal disturbances, vol 57, pp 395–412
15. Brown JC, Emslie AG (1988) Analytic limits on the forms of spectra possible from optically thin collisional bremsstrahlung source models. Astrophys J 331:554–564
16. Brown JC, Melrose DB, Spicer DS (1979) Production of a collisionless conduction front by rapid coronal heating and its role in solar hard X-ray bursts. Astrophys J 228:592–597
17. Brown JC, Emslie AG, Kontar EP (2003) The determination and use of mean electron flux spectra in solar flares. Astrophys J 595:L115–L117
18. Brown JC, Emslie AG, Holman GD, Johns-Krull CM, Kontar EP, Lin RB, Massone MA, Piana M (2006) Evaluation of algorithms for reconstructing electron spectra from their bremsstrahlung hard X-ray spectra. Astrophys J 643:523–531

19. Conway AJ, Brown JC, Eves BA, Kontar EP (2003) Implications of solar flare hard X-ray “knee” spectra observed by RHESSI. *Astron Astrophys* 407:725–734
20. Craig IJD, Brown JC (1986) Inverse problems in astronomy: a guide to inversion strategies for remotely sensed data. Adam Hilger, Bristol
21. Davies AM (1992) Optimality in regularization. In: Bertero M, Pike ER (eds) *Inverse problems in scattering and imaging*. Adam Hilger, Bristol, pp 393–410
22. Davies B, Martin B (1979) Numerical inversion of the Laplace transform: a survey and comparison of methods. *J Comput Phys* 33:1–32
23. Eicke B (1992) Iteration methods for convexly constrained ill-posed problems in Hilbert spaces. *Numer Funct Anal Opt* 13:413–429
24. Elwert G (1939) Accurate calculation of intensity and polarisation in continued X-ray spectra. *Ann Phys-Berlin* 34:178–208
25. Elwert G, Haug E (1970) On the polarization and anisotropy of solar X-radiation during flares. *Sol Phys* 15:234–248
26. Elwert G, Haug E (1971) Anisotropy of solar hard X-radiation during flares. *Sol Phys* 20:413–421
27. Emslie AG (1992) Overview of solar flares. In: Schmelz JT, Brown JC (eds) *The Sun: a laboratory for astrophysics*. Dordrecht
28. Emslie AG, Coffey VN, Schwartz RA (1989) Is the ‘superhot’ hard X-ray component in solar flares consistent with a thermal source? *Sol Phys* 122:313–317
29. Emslie AG, Barrett RK, Brown JC (2001) An empirical method to determine electron energy. *Astrophys J* 557:921–929
30. Emslie AG, Kontar EP, Krucker S, Lin RP (2003) RHESSI hard X-ray imaging spectroscopy of the large Gamma-ray flare of 2002 July 23. *Astrophys J* 595:L107–L110
31. Essah WA, Delves LM (1988) On the numerical inversion of Laplace transform. *Inverse Probl* 4:705–724
32. Gluckstern RL, Hull MH (1953) Polarization dependence of the integrated bremsstrahlung cross section. *Phys Rev* 90:1030–1035
33. Gluckstern RL, Hull MH, Breit G (1953) Polarization of bremsstrahlung radiation. *Phys Rev* 90:1026–1029
34. Golub G, van Loan C (1996) *Matrix computations*. Johns Hopkins, London
35. Hadamard J (1923) *Lectures on Cauchy’s problem in linear partial differential equations*. Yale University Press, New Haven
36. Haug E (1972) Polarization of hard X-rays from solar flares. *Sol Phys* 25:425–434
37. Haug E (1997) On the use of nonrelativistic bremsstrahlung cross sections in astrophysics. *Astron Astrophys* 326:417–418
38. Héroux JC (1975) Anisotropy and polarization of solar X-ray bursts. *Sol Phys* 42:219–233
39. Holman GD, Sui L, Schwartz RA, Emslie AG (2003) Electron bremsstrahlung hard X-ray spectra, electron distributions, and energetics in the 2002 July 23 solar flare. *Astrophys J* 595:L97–L102
40. Hurford GJ, Schmahl EJ, Schwartz RA, Conway AJ, Aschwanden MJ, Csillaghy A, Dennis BR, Johns-Krull C, Krucker S, Lin RP, McTiernan J, Metcalf TR, Sato J, Smith DM (2002) The RHESSI imaging concept. *Sol Phys* 210:61–86
41. Johns CM, Lin RP (1992) The derivation of parent electron spectra from bremsstrahlung hard X-ray spectra. *Sol Phys* 137:121–140
42. Kato T (1980) *Perturbation theory for linear operators*. Springer, New York
43. Keller J (1976) Inverse problems. *Am Math Mon* 83:107–118
44. Koch HW, Motz JW (1959) Bremsstrahlung cross section formulas and related data. *Rev Mod Phys* 31:920–955
45. Kontar EP (2001) Dynamics of electron beams in the inhomogeneous solar corona plasma. *Sol Phys* 202:131–149
46. Kontar EP, Pecseli HL (2002) Nonlinear development of electron beam driven weak turbulence in an inhomogeneous plasma. *Phys Rev E* 65:066408
47. Kontar EP, Brown JC, Emslie AG, Schwartz RA, Smith DM, Alexander RC (2003) An explanation for nonpower-law behavior in the hard X-ray spectrum of the 2002 July 23 solar flare. *Astrophys J* 595:L123–L126
48. Kontar EP, Piana M, Massone AM, Emslie AG, Brown JC (2004) Generalized regularization techniques with constraints for the analysis of solar bremsstrahlung X-ray spectra. *Sol Phys* 225:293–309
49. Kontar EP, Emslie AG, Piana M, Massone AM, Brown JC (2005) Determination of electron flux spectra in a solar flare with an augmented regularization method: application to RHESSI data. *Sol Phys* 226:317–325
50. Kontar EP, Emslie AG, Massone AM, Piana M, Brown JC, Prato M (2007) Electron-electron bremsstrahlung emission and the inference of electron flux spectra in solar flares. *Astrophys J* 670:857–861
51. Kress R (1989) *Linear integral equations*. Springer, New York
52. Legendijk R, Biemond J, Boeckee D (1988) Regularized iterative image restoration with ringing reduction. *IEEE Trans Acoust Speech Signal Process* 36:1874–1888
53. Langer SH, Petrosian V (1977) Impulsive solar X-ray bursts. III—Polarization, directivity, and spectrum of the reflected and total bremsstrahlung radiation from a beam of electrons directed toward the photosphere. *Astrophys J* 215:666–676
54. Leach J, Petrosian V (1983) The impulsive phase of solar flares. II—Characteristics of the hard X-rays. *Astrophys J* 269:715–727
55. Lin RP (1974) The flash phase of solar flares—Satellite observations of electrons. *Space Sci Rev* 16:201–221
56. Lin RP, Schwartz RA (1987) High spectral resolution measurements of a solar flare hard X-ray burst. *Astrophys J* 312:462–474
57. Lin RP, Dennis BR, Hurford GJ, Smith DM, Zehnder A, Harvey PR, Curtis DW, Pankow D, Turin P, Bester M, Csillaghy A, Lewis M, Madden N, van Beek HF, Appleby M, Raudorf T, McTiernan J, Ramaty R, Schmahl E, Schwartz R, Krucker S, Abiad R, Quinn T, Berg P, Hashii M, Sterling R, Jackson R, Pratt R, Campbell RD, Malone D, Landis D, Barrington-Leigh CP, Slassi-Sennou S, Cork C, Clark D, Amato D, Orwig L, Boyle R, Banks IS, Shirey K, Tolbert AK, Zarro D, Snow F, Thomsen K, Henneck R, Mchedlishvili A, Ming P, Fivian M, Jordan J, Wanner R, Crubb J, Preble J, Matrangola M, Benz A, Hudson H, Canfield RC, Holman GD, Cranell C, Kosugi T, Emslie AG, Vilmer N, Brown JC, Johns-Krull C, Aschwanden M, Metcalf T, Conway A (2002) The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI). *Sol Phys* 210:3–32
58. Massone AM, Piana M, Conway AJ, Eves B (2003) A regularization approach for the analysis of RHESSI X-ray spectra. *Astron Astrophys* 405:325–330
59. Massone AM, Emslie AG, Kontar EP, Piana M, Prato M, Brown JC (2004) Anisotropic bremsstrahlung emission and the form of regularized electron flux spectra in solar flares. *Astrophys J* 613:1233–1240
60. Massone AM, Piana M, Prato M (2008) Regularized solution of the solar bremsstrahlung inverse problem: model dependence and implementation issues. *Inverse Probl Sci Eng* 16:523–545
61. McWhirter JG, Pike ER (1978) On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind. *J Phys A, Math Gen* 11:1729–1745
62. Mertz LN (1967) A dilute image transform with application to an X-ray star camera. In: Fox J (ed) *Modern optics, proc symp modern optics, March 22–24 1967*, pp 787–791
63. Piana M (1994) Inversion of bremsstrahlung spectra emitted by a solar plasma. *Astron Astrophys* 288:949–959
64. Piana M, Bertero M (1996) Regularized deconvolution of multiple images of the same object. *J Opt Soc Am A* 13:1516–1523
65. Piana M, Bertero M (1997) Projected Landweber method and preconditioning. *Inverse Probl* 13:441–463

66. Piana M, Brown JC, Thompson AM (1995) Thermal bremsstrahlung hard X-rays and primary energy release in flares. *Sol Phys* 156:315–335
67. Piana M, Massone AM, Kontar EP, Emslie AG, Brown JC, Schwartz RA (2003) Regularized electron flux spectra in the July 23, 2002 solar flare. *Astrophys J* 595:L127–L130
68. Piana M, Massone AM, Hurford GJ, Prato M, Emslie AG, Kontar EP, Schwartz RA (2007) Electron flux spectral imaging of solar flares through regularized analysis of hard x-ray source visibilities. *Astrophys J* 665:846–855
69. Prato M, Piana M, Brown JC, Emslie AG, Kontar EP, Massone AM (2006) Regularized reconstruction of the differential emission measure from solar hard X-ray spectra. *Sol Phys* 237:61–83
70. Prince TA, Hurford GJ, Hudson HS, Crannell CJ (1988) Gamma-ray and hard X-ray imaging of solar flares. *Sol Phys* 118:269–290
71. Ramaty R, Paizis C, Colgate SA, Dulk GA, Hoyng P, Knight JW, Lin RP, Melrose DB, Orrall F, Shapiro PR (1980) Energetic particles in solar flares. In: *Solar flares: a monograph from Skylab Solar Workshop II*. Colorado Associated University Press, Boulder, pp 117–185
72. Reed M, Simon B (1972) *Methods of modern mathematical physics, vol I: Functional analysis*. Academic, New York
73. Rudin W (1991) *Functional analysis*, 2nd edn. McGraw-Hill, New York
74. Schnopper HW, Thompson RI, Watt S (1968) Predicted performance of a rotating modulation collimator for local celestial X-ray sources. *Space Sci Rev* 8:534–542
75. Smith DM, Share GH, Murphy RJ, Schwartz RA, Shih AY, Lin RP (2003) High-resolution spectroscopy of gamma-ray lines from the X-class solar flare of 2002 July 23. *Astrophys J* 595:L81–L84
76. Sneddon IN (1951) *Fourier transforms*. McGraw-Hill, New York
77. Sweet PA (1969) Mechanisms of solar flares. *Annu Rev Astron Astrophys* 7:149–176
78. Thompson AM, Brown JC, Craig IJD, Fulber C (1992) Inference of non-thermal electron energy distributions from hard X-ray spectra. *Astron Astrophys* 265:278–288
79. Tikhonov AN (1963) Regularization of incorrectly posed problems. *Sov Math Dokl* 4:1624–1627
80. Tikhonov AN (1963) Solution of incorrectly formulated problems and the regularization method. *Sov Math Dokl* 4:1035–1038
81. Varah JM (1983) Pitfalls in the numerical solution of linear ill-posed problems. *SIAM J Sci Comput* 4:164–176