GENERALIZED REGULARIZATION TECHNIQUES WITH CONSTRAINTS FOR THE ANALYSIS OF SOLAR BREMSSTRAHLUNG X-RAY SPECTRA

EDUARD P. KONTAR[1], MICHELE PIANA[2], ANNA MARIA MASSONE[3],
A. GORDON EMSLIE[4] and JOHN C. BROWN[1]

[1]*Department of Physics and Astronomy, The University of Glasgow, G12 8QQ, U.K.*
(e-mail: eduard@astro.gla.ac.uk)
[2]*Dipartimento di Matematica, Università di Genova, via Dodecaneso 35, I-16146 Genova, Italy*
[3]*INFM, UdR di Genova, via Dodecaneso 33, I-16146 Genova, Italy*
[4]*Department of Physics, The University of Alabama in Huntsville, Huntsville, AL 35899, U.S.A.*

Abstract. Hard X-ray spectra in solar flares provide knowledge of the electron spectrum that results from acceleration and propagation in the solar atmosphere. However, the inference of the electron spectra from solar X-ray spectra is an ill-posed inverse problem. Here, we develop and apply an enhanced regularization algorithm for this process making use of physical constraints on the form of the electron spectrum. The algorithm incorporates various features not heretofore employed in the solar flare context: Generalized Singular Value Decomposition (GSVD) to deal with different orders of constraints; rectangular form of the cross-section matrix to extend the solution energy range; regularization with various forms of the smoothing operator; and "preconditioning" of the problem. We show by simulations that this technique yields electron spectra with considerably more information and higher quality than previous algorithms.

1. Introduction

To address fundamental questions on electron propagation and acceleration in solar flares, it is necessary to infer as much quantitative information as possible on the electron spectrum in the solar plasma. A long-standing method of doing this involves the analysis of the emitted hard X-ray (HXR) bremsstrahlung spectrum, in particular the inversion of the integral equation (Brown, 1971) relating the two spectra. This task is particularly challenging since even the most accurate photon spectra are contaminated by noise, which is dramatically amplified in any unconstrained attempt to extract the electron flux spectrum. Traditional approaches to the determination of the electron spectrum sidestep this problem by assuming simple (e.g., isothermal + power-law) forms and adjusting their parameters to achieve the best fit to the HXR data (e.g., Holman *et al*., 2003). However, such algorithms, by their very nature, cannot detect features in the electron spectrum that, although real, were not included into the prescribed empirical form. Indeed, the analysis of high-resolution RHESSI (Ramaty High-Energy Solar Spectroscopic Imager) (Lin *et al*., 2002) spectra shows substantial deviations from simple models (Kontar *et al*., 2003).

While this was adequate for earlier low-resolution data, the goal in high-resolution photon spectrum analysis should be the suppression of noise-induced unphysical behavior in the electron flux, while maintaining maximum ability to recover faithfully real features. Various algorithms have been employed with this intent (see, e.g., Johns and Lin, 1992; Thompson *et al.*, 1992; Piana, 1994; Piana and Brown, 1998: Piana *et al.*, 2003) all belonging to the wide class of regularization methods for linear ill-posed inverse problems, though differing in method of regularization. For example, Johns and Lin (1992) sum over energy intervals to obtain sufficiently good statistical accuracy (regularization by coarse energy binning). Their results suggested downturn of the electron spectra below 50 keV, though the results were too uncertain to be conclusive. Piana *et al.* (2003) have detected, through Tiknonov regularized inversion (Tikhonov, 1963), a feature at $E \simeq 55$ keV in the mean source electron spectrum for the July 23, 2002 solar flare, that has been impossible to detect through a forward-fitting algorithm involving power-law functions such as those used by Holman *et al.* (2003). Although it has not been possible so far to establish the origin of this particular feature,[1] nevertheless the $\bar{F}(E)$ form obtained by Piana *et al.* (2003) is a faithful description of the $\bar{F}(E)$ corresponding to the photon spectrum used.

An observed hard X-ray spectrum $I(\epsilon)$ is related, through a bremsstrahlung cross-section $Q(\epsilon, E)$, to the *mean electron flux spectrum* $\bar{F}(E)$ in the source, through the relation (Brown, 1971; Johns and Lin, 1992; Brown, Emslie, and Kontar, 2003)

$$I(\epsilon) = \frac{1}{4\pi R^2} \, \bar{n}V \int_{\epsilon}^{\infty} \bar{F}(E) \, Q(\epsilon, E) \, dE, \tag{1}$$

where R is the distance to the observer, V is the emitting volume and $\bar{n} = V^{-1} \int n(\mathbf{r})dV$ is the mean target density. The problem of determining $\bar{F}(E)$ from $I(\epsilon)$ when they are related by a Volterra-type Equation such as (1) is an ill-posed problem in the sense of Hadamard (1923). As a result, every experimental problem described by an equation like (1) is affected by a numerical pathology termed *ill-conditioning* whereby, when an unconstrained solution procedure is followed, the presence of measurement noise is reflected in unphysical oscillations in the reconstructed solution for the source function. To obtain a meaningful solution $\bar{F}(E)$ one needs to avoid noise amplification (e.g., Craig and Brown, 1986) by means of regularization methods applying physical constraints to the electron spectrum. The algorithm looks for an approximate least-squares solution of the integral equation relating the photon and the electron spectra, but subject to inclusion of additional information based on physically meaningful constraints or assumptions on the solution. This leads to the formulation of a family of regularization methods exploiting different possible *a priori* information on the electron flux coming from solar

[1]It is possible that this particular feature has non-solar origin and is a result of the effects of pulse pile-up – Smith *et al.* (2002).

physics. The aim of the present paper is to introduce this generalized constraint approach into the field of solar HXR spectrum analysis, together with various other new features (physical, mathematical and numerical) yielding a more effective algorithm, based on Generalized Singular Value Decomposition, for determination of $\bar{F}(E)$. More precisely, in this paper we will discuss the three major problems concerning the application of a regularization approach to the analysis of HXR data, that is: how to effectively introduce physically meaningful constraints into the inversion procedure; how to tune the stability requirement avoiding artificial clustering in the residuals, and, finally, how to adapt the regularization techniques to solar data characterized by huge dynamical ranges.

The plan of the paper is as follows. In Section 2 we review the mathematical formulation of the problem pointing out its numerical instability. In Section 3 we discuss the relation between regularization and physical constraints, while in Section 4 we describe the regularization and the computational method based on the Generalized Singular Value Decomposition. In Section 5 a solution is proposed, through analysis of the cumulative residuals, to the crucial problem of optimal choice of the regularization parameter which controls the trade-off between stability and loss of information. In Section 6 we perform an analysis of the solution structure. Finally, in Section 7 we present some applications in the case of simulated photon spectra to demonstrate the improvements achievable and the summary is presented in Section 8.

In a subsequent paper (Kontar *et al.*, 2004 (hereafter Paper II)) we will apply this approach to real spectra observed with RHESSI (Lin *et al.*, 2002).

2. Mathematical Formulation

Equation (1) is a Volterra integral equation of the first kind and can be expressed in terms of a linear integral operator $A : \mathbb{X} \to \mathbb{Y}$,

$$(A\bar{F})(\epsilon) \equiv \frac{1}{4\pi R^2} \bar{n} V \int_\epsilon^\infty \bar{F}(E) \, Q(\epsilon, E) \, dE, \tag{2}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are two appropriate functional (Hilbert) spaces. For physical forms of the bremsstrahlung cross-section Q, A is a compact linear operator so that (Bertero, De Mol, and Pike, 1985) every discretization of Equation (2) is characterized by (significant) numerical instabilities. Let us, then, consider some convenient discretized form of (1), viz. the linear system

$$\mathbf{A}\bar{\mathbf{F}} = \mathbf{g}, \tag{3}$$

where

$$A_{ij} = \frac{\bar{n} V}{4\pi R^2} Q((\epsilon_{i+1} + \epsilon_i)/2, (E_{j+1} + E_j)/2) \, \delta E_j,$$
$$i = 1, \ldots, N, \quad j = 1, \ldots, M(> N), \tag{4}$$

$\bar{\mathbf{F}} = (\bar{F}(E_1), \ldots, \bar{F}(E_M))$, $\mathbf{g} = (g(\epsilon_1), \ldots, g(\epsilon_N))$, with $M > N$, and the δE_j and $\delta\epsilon_i$ are appropriate weights. The values $g(\epsilon_i)$ correspond to a set of discrete photon counts in energy bands $\epsilon_i \rightarrow \epsilon_i + \delta\epsilon_i$, while the $\bar{F}(E_j)$ are the corresponding values of the mean electron flux in energy bands $E_j \rightarrow E_j + \delta E_j$. We use a matrix $\mathbf{A}$ corresponding to the bremsstrahlung cross-section due to Haug (1997) with the Elwert (1939) Coulomb correction applied.[2]

It is important to recognize that, since electrons of all energies $E \geq \epsilon$ contribute to the photon emission at energy ϵ, in general the hard X-ray spectrum over a finite range $[\epsilon_1, \epsilon_N]$ of photon energies contains information on the *electron* spectrum over a much wider range, in particular within the range $\epsilon_N < E < E_M$ above the uppermost photon energy. For example, if $\bar{F}(E)$ has an upper energy cutoff at $E = E_{\max} > \epsilon_N$, then this cutoff will affect the observed photon spectrum below ϵ_N, since the photon spectrum must tend to zero at $\epsilon \rightarrow E_{\max}$. Hence, by extending our array of E values to a sufficiently high value, the solution of (3) can potentially reveal evidence of an upper energy cutoff (see Section 7 below).

Since $M > N$, problem (3) is under determined, i.e., there is no unique solution of the linear system. The same holds true if we consider the least-squares problem

$$\|\mathbf{A}\bar{\mathbf{F}} - \mathbf{g}\|^2 = \min, \tag{5}$$

where $\|\cdot\|$ is the canonical Euclidean norm defined by

$$\|f\| \equiv \left(\int_{x_{\min}}^{x_{\max}} f^2(x) \mathrm{d}x \right)^{1/2}. \tag{6}$$

The solutions of (5) are commonly known as pseudo solutions (Bertero, De Mol, and Pike, 1985). Obviously, additional constraints need to be applied to obtain a unique solution.

3. Physical Constraints and Regularization

As a physical quantity $\bar{\mathbf{F}}$ must satisfy various physical conditions such as $\bar{\mathbf{F}} \geq \mathbf{0}$ and any known constraints such as properties which are to be conserved or to be minimized/maximized. Many such properties can be expressed, for a suitable closed operator $\mathbf{L}$, in the form

$$\|\mathbf{L}\bar{\mathbf{F}}\| \leq \text{const}, \tag{7}$$

---

[2]Note that the Elwert correction is not applicable for the high-frequency limit, $\epsilon \simeq E$. However, since photons of energy ϵ are produced, in general, by a wide range of electron energies E, the neglect of a more sophisticated form of the cross-section in this very narrow energy range does not significantly affect our results.

leading to the need to solve least-square problem (5) subject to additional constraint (7). This constrained minimum problem can be solved using the Lagrange multiplier method, namely

$$\mathcal{L}(\bar{\mathbf{F}}) \equiv \|\mathbf{A}\bar{\mathbf{F}} - \mathbf{g}\|^2 + \lambda \|\mathbf{L}\bar{\mathbf{F}}\|^2 = \min, \tag{8}$$

where λ is a Lagrange multiplier. This approach to regularize the problem is known as the Tikhonov regularization (Tikhonov, 1963).

Three possible different choices for $\mathbf{L}$ are discussed in the following.

As a first example, we observe that the density-weighted target-averaged energy-integrated flux of electrons is given by the Euclidean norm $\|\bar{\mathbf{F}}\|$. Physically, this quantity must be fixed by the total flux (the function of the total number of electrons or energy), a requirement that can be met by a suitable choice of the second (constraint) term in Equation (8), the simplest choice being $\|\bar{\mathbf{F}}\|$ or $\mathbf{L} = \mathbf{1}$, the identity matrix, so that the second term of Equation (8) is just the Euclidean norm of the solution $\bar{\mathbf{F}}$. Problem (8) with this constraint operator is also termed *zero order regularization* and was used by Piana *et al.* (2003) for solar data. It defines the non-parametric $\bar{F}(E)$ of target-averaged total electron flux consistent with the data for a given parameter λ to be chosen via some appropriate additional condition.

The source averaged electron flux $\bar{F}(E)$ physically results from a combination of the injected electron flux $F_0(E_0)$ spectrum and the physics of particle transport in the radiating source. Under a broad, but not comprehensive range of conditions, these flux spectra are related by (Brown and McKinnon, 1985)

$$\bar{\mathbf{F}}(E) \sim \frac{1}{|dE/dN|} \int_E^\infty F_0(E_0) dE_0, \tag{9}$$

where dE/dN is the rate of energy loss per unit column density. For example, for collisional energy losses in a cold target, $dE/dN \sim -1/E$ and therefore (Brown and Emslie, 1988)

$$F_0(E_0) \sim -\frac{d}{dE} \left[\frac{\bar{\mathbf{F}}(E)}{E} \right]_{E=E_0}. \tag{10}$$

Equation (10) shows that, if the purpose of finding a solution $\bar{F}(E)$ is to subsequently use that solution to infer the injected electron flux spectrum $F_0(E_0)$, then the mean electron flux should be differentiable, a requirement which can be incorporated in (8) by adopting for $\mathbf{L}$ the differentiation operator $\mathbf{L} \sim \mathbf{D}^1$ which is termed *first order regularization*.

Physically, if the electron acceleration can be described deterministically, e.g., the injection function resulted from acceleration can be presented similar to an integral (9) with systematic acceleration instead of deceleration, then the injected spectrum should be a differential function. For example, acceleration by an electric field leads to differentiable spectrum of accelerated (injected) electrons. If we believe the injection function is differentiable, then $\bar{F}(E)$ should not only be differentiable,

but have a differentiable first derivative. This corresponds to the requirement of a bounded second order derivative, hence to *second order regularization* $\mathbf{L} \sim \mathbf{D}^2$.

It is informative to consider the effect of applying regularization methods of very high order, since they can obscure physical features in $\bar{F}(E)$ that do not comply with the imposed smoothness constraint. The k-order derivative $\mathbf{D}^{(k)}\bar{\mathbf{F}}$ cannot be defined over less than k points, which reduces the dimension of the solution space to $N - k$ for N data points. This makes higher-order solutions more restrictive and potentially less precise though in practice the high-resolution data now available have such large $N \sim 300$ that this is not an issue for any reasonable order of regularization. However, in the case of forward-fitting procedures, the natural $N \sim 300$ dimensional space of possible solutions is forcefully squeezed into the space of only a few dimensions (5 in the case of power-law + isothermal components – Holman *et al.*, 2003) which corresponds to very-high-order regularization and hence is very restrictive.

4. Regularized Solution and Generalized Singular Value Decomposition

Provided that the null spaces of the matrices $\mathbf{A}$ and $\mathbf{L}$ intersect trivially (i.e., $\mathbf{AF} = \mathbf{0}$ and $\mathbf{LF} = \mathbf{0}$ have no common solutions other than $\mathbf{F} = \mathbf{0}$), the formal solution of the minimum problem can be shown to be (Hansen, 1992)

$$\bar{\mathbf{F}}_\lambda = (\mathbf{A}^*\mathbf{A} + \lambda \mathbf{L}^*\mathbf{L})^{-1}\mathbf{A}^*, \tag{11}$$

where $\mathbf{A}^*$ is adjoint of an operator $\mathbf{A}$. From the point of view of applications, this formula is not helpful since truncation errors imply a notable loss of information in forming the cross-product matrix $\mathbf{A}^*\mathbf{A}$ and, furthermore, the computational effort required is significant. Computational heaviness, together with the presence of local minima, affect also the use of quadratic programming for convex functional minimization, which is a typical strategy for computing the solution of (8). A more effective approach is to use the Generalized Singular Value Decomposition (GSVD) algorithm. Following van Loan (1976), we consider an $M \times N$ matrix $\mathbf{A}$ and a $P \times N$ matrix $\mathbf{L}$ ($M \geq N \geq P$). Then for each pair of real matrices $(\mathbf{A}, \mathbf{L})$

$$\mathbf{A} \in \mathbb{R}^{M \times N}, \quad \mathbf{L} \in \mathbb{R}^{P \times N}, \tag{12}$$

it can be shown that there exists a set of singular values σ_k^A, σ_k^L satisfying the relation $(\sigma_k^A)^2 + (\sigma_k^L)^2 = 1$, and singular vectors $\tilde{\mathbf{u}}_k$, $\tilde{\mathbf{v}}_k$, $\tilde{\mathbf{w}}_k$, where the first two sets are orthogonal and the third one is a set of linearly independent vectors satisfying the simultaneous equations

$$\mathbf{A} = \tilde{\mathbf{U}} \begin{pmatrix} \mathrm{diag}(\sigma_k^A), & 0 \\ 0 & \mathbf{1}_{N-P} \\ 0 & 0 \end{pmatrix} \tilde{\mathbf{W}}^{-1}, \quad \mathbf{L} = \tilde{\mathbf{V}}(\mathrm{diag}(\sigma^L)\,0)\tilde{\mathbf{W}}^{-1}. \tag{13}$$

Here the $M \times M$ matrix $\tilde{\mathbf{U}}$ is formed from the M column vectors $\tilde{\mathbf{u}}_k$, $k = 1, \ldots, M$, with similar definitions for the $P \times P$ matrix $\tilde{\mathbf{V}}$ and the $N \times N$ matrix $\tilde{\mathbf{W}}$. The generalized singular values are defined as the ratios $\sigma_k = \sigma_k^A / \sigma_k^L$.

The solution to this generalized minimization problem (8) can be shown to be (Hansen, 1992)

$$\bar{\mathbf{F}}_\lambda = \sum_{k=1}^{P} \left(\frac{\sigma_k^2}{\sigma_k^2 + \lambda} \frac{(\mathbf{g} \cdot \tilde{\mathbf{u}}_k)}{\sigma_k^A} \right) \tilde{\mathbf{w}}_k + \sum_{k=P+1}^{N} (\mathbf{g} \cdot \tilde{\mathbf{u}}_k) \tilde{\mathbf{w}}_k. \tag{14}$$

In the particular case when $\mathbf{L} = \mathbf{1}$ (zero-order regularization), $P = N$ and Equation (13) shows that $\tilde{\mathbf{u}}_k = \mathbf{u}_k$, $\tilde{\mathbf{v}}_k = \mathbf{v}_k$. Hence the second term in the solution (14) vanishes identically and the first term reduces to the one for the zero-order regularized solution.

5. Choice of the Regularization Parameter

It has to be recalled that the choice of the regularization parameter λ, and indeed of $\mathbf{L}$, has to be made independently of the equation and of the data, using prior knowledge or prejudice, since there is no unique solution to the Equation (5) itself. As mentioned, the integral properties of the electron flux (7), if they were known, would unambiguously determine the Lagrange multiplier λ in our problem (8). Unfortunately, we do not know *a priori* the total flux of X-ray producing electrons or other integral quantity. Therefore it is advantageous to use knowledge of the errors in the recovered solution to choose the regularization parameter. Several criteria for determining the optimal λ in Equation (14) have been introduced. A general property of them is that the optimal λ tends to zero when the noise level tends to zero. For example, according to the *discrepancy principle* (Tikhonov *et al.*, 1995), the best value of λ is given by the solution of the equation

$$\|\mathbf{A}\bar{\mathbf{F}}_\lambda - \mathbf{g}\|^2 = \|\delta\mathbf{g}\|^2, \tag{15}$$

where δg is some measure of the noise affecting the data (essentially, typically the canonical norm of the error vector). The discrepancy principle is typically rather robust in that it yields stable values but empirical tests show that the parameter it provides is often too large, and the corresponding regularized solution oversmoothed.

Here, we propose a procedure for the choice of the regularization parameter based on the analysis of the residuals $r_k = ((\mathbf{A}\bar{\mathbf{F}})_k - \mathbf{g}_k)/\delta g_k$. Then the deviation weighted by the error

$$\|(\mathbf{A}\bar{\mathbf{F}}_\lambda - \mathbf{g})(\delta\mathbf{g})^{-1}\|^2 \simeq 1 \tag{16}$$

accounts more accurately for point-to-point error variation than (15). Indeed, λ defined by Equation (16) has accounted for detailed structure of errors, while the discrepancy principle uses only total error.

Ideally, the normalized residuals should be consistent with statistical deviations in the data, and should therefore form a Gaussian distribution with zero mean; the cumulative normalized residual

$$C_j = \frac{1}{j} \sum_{k=1}^{j} r_k \qquad (17)$$

should be mostly within $\pm 1/\sqrt{j}$. A too-small set of values for the cumulative residuals indicates insufficient smoothing, whereas a set of values that consistently exceed $\pm 1/\sqrt{j}$ (especially if the sign of the residuals cluster) indicates too great a regularizing parameter. Therefore, we start with the value of λ given by (16) and then reduce λ until the average residual in the photon spectrum over the energy range from ϵ_1 to ϵ_j as a function of j is mostly within the $\pm \sigma$ limits expected if the residuals were purely statistical, drawn from a normal distribution. This technique is similar to requiring $\chi^2 \simeq 1$ in hypothesis testing situations.

6. Analysis of the Solution Structure

A key issue in achieving the most meaningful construction of a regularized solution lies in analysis of the quantities

$$c_k = \left| \frac{(\sigma_k)^2}{(\sigma_k)^2 + \lambda} \frac{(\mathbf{g} \cdot \tilde{\mathbf{u}}_k)}{\sigma_k^A} \right| \qquad k = 1, \ldots, N \qquad (18)$$

i.e., the absolute value of the coefficients which multiply the singular vectors $\mathbf{v}_k$ in the solution (14). To obtain a meaningful solution represented by Equation (14) the singular values σ_k should decrease faster on average than the coefficients $(\mathbf{g}, \mathbf{u}_k)$ (the *Picard condition*; Groetsch, 1984). Figure 1 shows a typical behavior of the coefficients c_k.

The first singular vector is always either positive or negative definite (so that $c_1 \tilde{\mathbf{w}}_1$ is always positive), while the $\tilde{\mathbf{w}}_k$ always have oscillatory behavior for $k > 1$. Therefore, for sufficiently small regularization parameters λ, the regularized solution may exhibit negative values when the coefficients c_k are not decreasing fast enough, so that relatively large values of the regularization parameter λ are necessary to guarantee a positive definite solution.

Inasmuch as typical solar flare hard X-ray spectra are sufficiently steep, it is advantageous to transform the fundamental problem (1) (or, equivalently, (8)) to a form in which the c_k are decreasing faster with k, so that smaller values of λ, which preserve more fidelity in the recovered solution, can be used. This will also avoid errors connected with finite precision arithmetics in machine calculations.
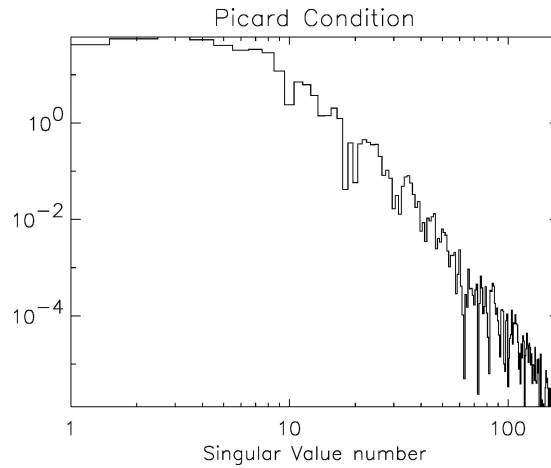
Figure 1. Variation of the coefficients c_k in Equation (18) as a function of the vector number k for the simulated data set with $\delta = 2$.

Two strategies can be followed to overcome this difficulty. In the first one, instead of considering Equation (3) we consider the new linear system

$$\mathbf{A}\delta\bar{\mathbf{F}} = \delta\mathbf{g} \tag{19}$$

with

$$\delta\bar{\mathbf{F}} = \bar{\mathbf{F}} - \bar{\mathbf{F}}_*, \quad \delta\mathbf{g} = \mathbf{g} - \mathbf{A}\bar{\mathbf{F}}_*, \tag{20}$$

where $\bar{\mathbf{F}}_*$ is an adopted form (often, but not necessarily, a closed parametric expression) for $\bar{\mathbf{F}}$. Now $\delta\mathbf{g}$ represents the deviations from the reference spectrum. Hence, if this reference spectrum is chosen appropriately (say from a forward-fit solution – e.g., Holman *et al.*, 2003), then these deviations $\delta\bar{\mathbf{F}}$ vary around zero, and the function $\delta\mathbf{g}$ will be significantly flatter than $\mathbf{g}$, so that the behavior of the c_k becomes more monotonic.

In addition to facilitating the calculation of a smooth (but not unnecessarily over-smoothed) solution of the minimization problem, the quantity $\delta\bar{\mathbf{F}}$ is interesting in its own right. It represents the deviation of the actual electron spectrum $\bar{\mathbf{F}}(E)$ from the assumed reference spectrum $\bar{\mathbf{F}}_*(E)$ and hence is an "adjustment" to this assumed (e.g., forward-fitted) form. This "initial guess" approach is the one adopted in the present paper.

A second possible strategy to constrain the behavior of the c_k is to consider the change of variables,

$$g(\epsilon_i) \rightarrow \epsilon_i^p \, g(\epsilon_i), \quad \bar{F}(E_j) \rightarrow E_j^q \, \bar{F}(E_j), \quad \mathbf{A}_{ij} \rightarrow \frac{\epsilon_i^p}{E_j^q} \, \mathbf{A}_{ij}, \tag{21}$$

with p, q positive real numbers. Then the basic form of the solution (14) is unaltered, but the forms of the matrix and its associated singular system are altered which

leads to modified behavior of the coefficients c_k. If a scaling (21) can be found that makes the c_k decreasing functions of k, then this transformed solution will have more desirable properties. We have found through experimentation that a judicious choice of scaling is $p = q = (\gamma - 1)/2$, where γ is the best-fit power-law spectral index to the array of g values. For all values of γ (from 2 to ~ 20), this scaling drives the coefficients c_k toward a rapidly decreasing form. Note that while re-scaling with $p = \gamma$ produces a much flatter data function $\mathbf{g}$, such a steeper scaling concomitantly leads (Equation (21)) to the matrix $\mathbf{A}$ becoming less diagonal, thereby increasing the ill-conditioning of the system and hence the rate of decay of the σ_k with k. The choice $p = (\gamma - 1)/2$ hence lies at the "middle ground" between steepness of the input data vector and ill-conditioning of the transformation matrix; similar arguments apply to the choice of q.

It should also be noted that "rescaling" is equivalent to constructing the *ratio* (rather than the *difference*) of $\bar{F}(E)$ relative to a reference $\bar{F}_*(E)$ form.

7. Application of the Algorithm to Simulated Data

In this section we explore the application of the above techniques to simulated data, in order to demonstrate some of the effects of various features.

7.1. NON-SQUARE MATRIX

To demonstrate the benefits of using a non-square $\mathbf{A}$ extending to higher electron energies than the highest observed photon energy, we assumed an *actual* $\bar{F}(E)$ of the form $\bar{F}(E) \sim E^{-\delta}$, $10 < E < E_{\max}$, with $\delta = 2$ and $E_{\max} = 300, 400, 500$ keV. We then used Equation (1) and the electron-proton bremsstrahlung cross-section from Haug (1997) to calculate the photon spectrum $I(\epsilon)$, in the range $10 \leq \epsilon \leq 200$ keV, from each electron spectrum. (It should be noted that the upper limit to the "observed" photon spectrum is in all cases less than the upper-energy cutoff in the electron spectrum.) Figure 6 shows the calculated photon spectra and the corresponding local spectral index $\gamma = -d \log I(\epsilon)/d \log \epsilon$. At low energies, the spectrum is well represented by a power-law form with $\gamma = \delta + 1 = 3$, but even at energies as low as ~ 50 keV, the deviation from a power-law behavior induced by the upper-energy cutoff in $\bar{F}(E)$ is already evident. For the case $E_{\max} = 300$ keV, by $\epsilon = 100$ keV the spectrum has steepened from its low-energy form ($\sim \epsilon^{-3}$) sufficiently that the spectral index has increased by as much as 0.5.

Such a deviation in local hard X-ray spectral index γ is clear evidence for a significant deviation from the power-law behavior of the generating $\bar{F}(E)$ spectrum at higher energies, although the exact nature of this high-energy deviation is not immediately obvious from the form of the photon spectrum (or even from a $\gamma(\epsilon)$ plot). We therefore now explore the ability of our technique to uncover the actual nature of this deviation (namely, in this case the upper energy cutoff

at $E = E_{max}$). We performed a regularized inversion of the photon spectra of Figure 6 by using solution (14) under the following conditions: zero order regularization; the simulated photon "data" used covered the range $10 < \epsilon < 200$ keV and the electron upper energy limits E_{upper} used in the inversion were 400, 500, and 600 keV.

In all cases the recovered electron spectra quite faithfully reproduce the actual high-energy cutoffs $E_{max} = 300$ keV (Figure 2). The higher energy cut-off is seen as substantial steepening in the reconstructed spectra. However, we should note
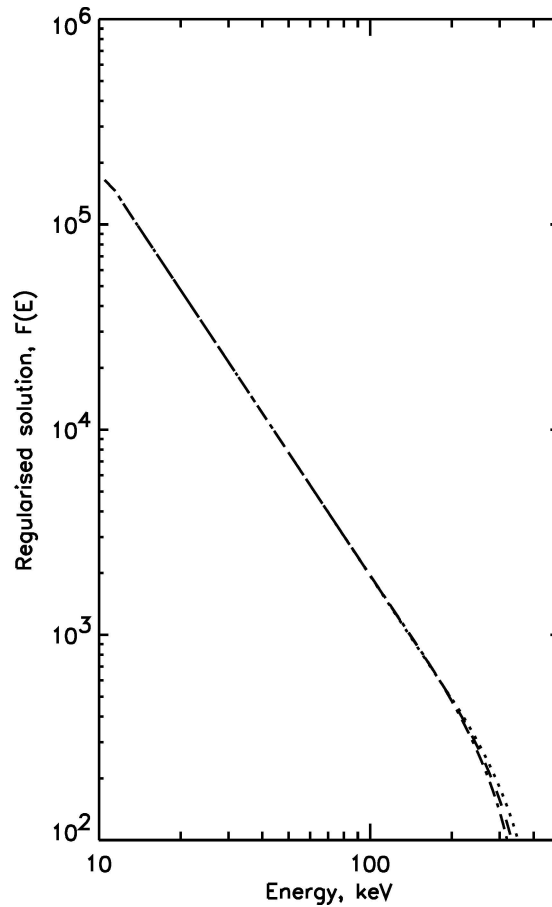
Figure 2. Mean source electron spectra $\bar{F}(E)$ recovered from the photon spectrum $I(\epsilon; E_{max} = 300$ keV) (dot-dashed spectrum in Figure 6), using the GSVD technique and zero order regularization with photon "data" in the range $10 < \epsilon < 150$ keV and electron energy ranges 10 keV $< E < E_{upper}$, where $E_{upper} = 400$ keV (*dot-dashed curve*), 500 keV (*dashed curve*) and 600 keV (*dotted curve*). The reference spectrum (Equation [20]) was of the form $\bar{F}_*(E) \sim E^{\gamma_*}$. Note that the high-energy behavior of $\bar{F}(E)$ (in particular the high-energy cutoff at $E_{max} = 300$ keV) is quite faithfully reproduced.

somewhat obvious limitations: if the true spectra has some variations above the range we are given, the reconstructed solution does not display these features due to the lack of information given (Figure 3).

7.2. HIGHER-ORDER REGULARIZATION WITH GSVD

To illustrate the benefits of GSVD and higher-order regularization we consider the reconstruction of an electron spectrum which is the superposition of a power-law plus an oscillatory trigonometric function. In this case, the use of a first-order penalty term governed by $\mathbf{L} \sim \mathbf{D}^1$ is more effective than the zero-order
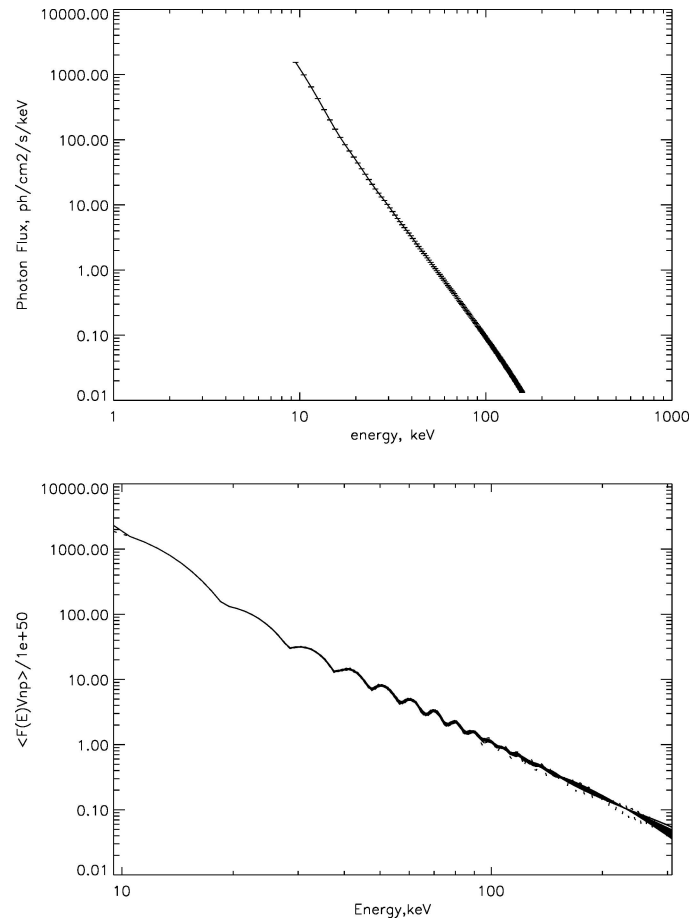
Figure 3. Reconstruction from simulated data. *Upper panel*: the simulated photon spectrum; *bottom panel*: the reconstruction obtained by using first order regularization. The *dotted line* shows the input electron spectrum and *solid lines* presents 30 realizations of the solution with the data randomly perturbed within error bars.

regularization algorithm, since the bound on the first derivative of the regularized solution assures a sufficiently correct behavior (in terms of residuals) of the solution. In Figure 4 (upper panel) the simulated photon spectrum is obtained by inserting the theoretical electron spectrum into Equation (3) while Figure 4 (lower panel) shows the reconstruction obtained by using formula (14) when $\mathbf{L} \sim \mathbf{D}^1$ and λ is chosen by using the cumulative residuals analysis criterion (in this figure we also superimposed the theoretical electron spectrum in order to point out the reliability of our approach). In Figure 5 (upper panel) and Figure 5 (lower panel) the behavior of the normalized and cumulative residuals shows that the fitting performance of the regularized solution is extremely accurate in the range below 150 keV.
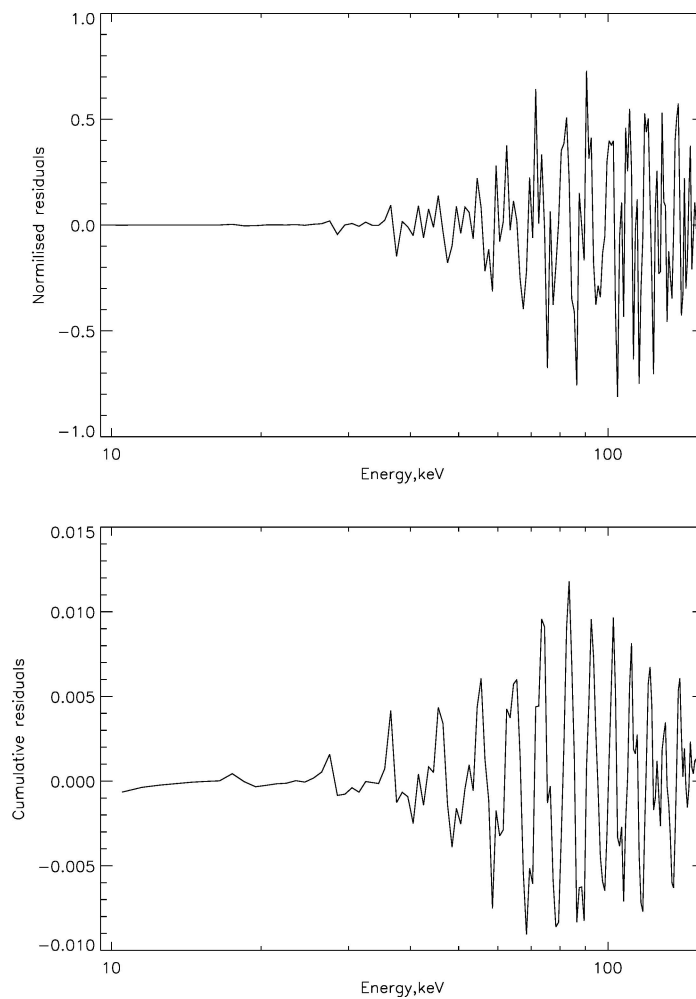

Figure 4. Reconstruction from simulated data. *Upper panel*: normalized residuals; *bottom panel*: normalized cumulative residuals.
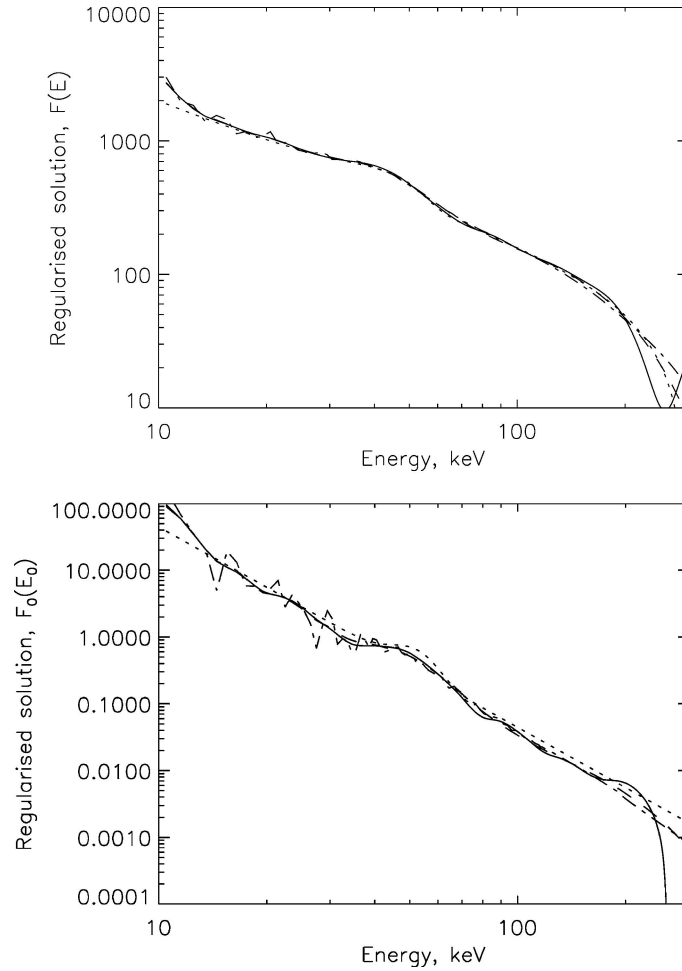
Figure 5. Reconstruction from simulated data. Recovered mean flux $\bar{F}(E)$ (*upper panel*) and re-covered injected flux $F_0(E_0)$ (*lower panel*) given by Equation (10). The *dotted lines* show the true solution, while three various orders of regularization are shown: second order regularization (*solid line*), first order (*dashed line*), zero (*dash-dot*).

The importance of higher-order regularization using GSVD also becomes explicit when one wants to derive the injected electron distribution $F_0(E_0)$. Figure 5 shows both the mean source and injected electron spectra obtained using different orders of regularization, compared with their true forms. As a true form of $F_0(E_0)$ we took a power law with a bump simulated by exponent as can be seen in Figure 5. The simulated photon spectrum is obtained by inserting the theoretical electron spectrum with the high energy cut-off at 300 keV into Equation (3) and 5% noise has been added. The reconstructed spectra and input spectra are shown in Figure 5. Note, that in the reconstruction we used data only up to 150 keV.
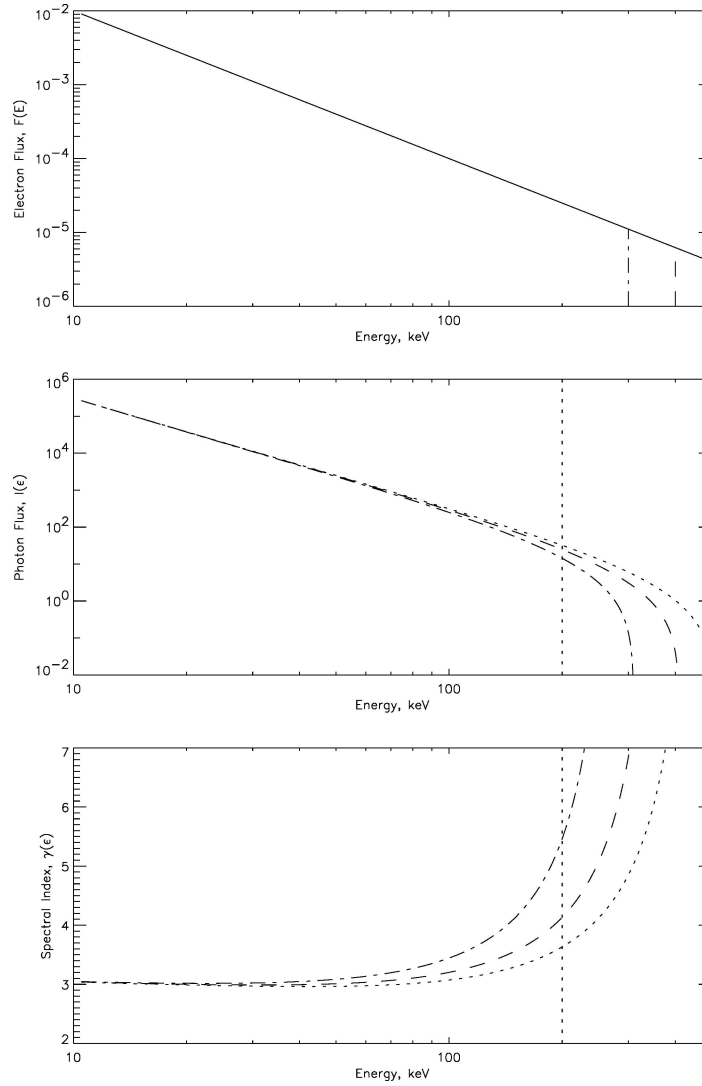
Figure 6. Upper panel: Simulated mean source electron spectrum $\bar{F}(E)$, which has the form of a power-law (index $\delta = 2$), with three different upper energy cutoffs $E_{\max}$. *Middle panel:* Corresponding photon spectra $I(\epsilon)$ for each of the $\bar{F}(E)$ forms. *Lower panel:* Local spectral index $\gamma = d \log I(\epsilon)/d \log \epsilon$ for each photon spectrum. Note the evidence for the high-energy cutoff in the hard X-ray spectrum and its local spectral index at much lower energies than $E_{\max}$. *The vertical lines* at $\epsilon = 200$ keV represent the upper energy limit of the spectral data used for subsequent analysis.

The second-order regularized solution shows the closest solution for $F_0(E_0)$ in Equation (5). On the other hand first- and zero-order regularizations show results with smaller χ^2 in recovering the spatially integrated spectra (5). Indeed, second- and first-order regularization preserves information on small scale (closer reproduction of a hump in $F_0(E_0)$ around 30 keV), while zero-order regularization is better

in global properties of the solution. The first- and zero order regularization show lack of sufficient smoothness that is displayed in oscillations in $F_0(E_0)$ (Figure 5). The main deviations from the true solutions are observed above 200 keV, where we have only approximate solution and near low energy cut-off due to boundary effects.

8. Conclusions

In the paper, we have summarized the essential mathematics associated with application of an Generalized Singular Value Decomposition (GSVD) technique to the solution of Volterra integral equations arising in solar X-ray spectroscopy, and in particular to the inference of mean source electron spectra $\bar{F}(E)$ and injected (accelerated) electron spectrum $F_0(E_0)$ from observations of solar flare hard X-ray spectra $I(\epsilon)$. Judicious use of this methodology can recover forms of $\bar{F}(E)$ that are not only relatively free from the effects of data noise amplification, but which also recover features that are not realizable using more traditional (e.g., forward-fitting) methods. Further, they can reveal approximate behavior in the electron spectrum well above the range of photon energies observed.

In the companion paper (Paper II) we will make use of these techniques in the analysis of high-resolution solar flare spectra observed by RHESSI.

Acknowledgement

References

Bertero, M., De Mol, C., and Pike, E. R.: 1985, *Inverse Prob.* **1**, 300.
Bertero, M., De Mol, C., and Pike, E. R.: 1988, *Inverse Prob.* **4**, 573.
Brown, J. C.: 1971, *Solar Phys.* **18**, 489.
Brown, J. C. and MacKinnon, A. L.: 1985, *Astrophys. J.* **292**, L31.
Brown, J. C. and Emslie, A. G.: 1988, *Astrophys. J.* **331**, 554.
Brown, J. C., Emslie, A. G., and Kontar, E. P.: 2003, *Astrophys. J. Lett.* **595**, L115.
Craig, I. J. D. and Brown, J. C.: 1986, *Inverse Problems in Astrophysics*, Bristol, Adam-Hilger.
Elwert, G.: 1939, *Ann. Physik* **34**, 178.
Emslie, A. G., Kontar, E. P., Krucker, S., and Lin, R. P.: 2003, *Astrophys. J. Lett.* **595**, L107.
Groetsch, C. W.: 1984, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, (Pitman: Boston).
Hadamard, J.: 1923, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale University Press: New Haven.
Hansen, P. C.: 1992, *Inverse Problems* **8**, 849.

Haug, E.: 1997, *Astron. Astrophys.* **326**, 417.

Holman, G. D., Sui, L., Schwartz, R. A., and Emslie, A. G.: 2003, *Astrophys. J. Lett.* **595**, L97.

Johns, C. and Lin, R. P.: 1992, *Solar Phys.* **137**, 121.

Kontar, E. P., Brown, J. C., Emslie, A. G., Schwartz, R. A., Smith, D. M., and Alexander, R. C.: 2003, *Astrophys. J. Lett.* **595**, L123.

Kontar, E. P., Piana, M., Emslie, A. G., and Brown, J. C.: 2004, *Solar Phys.*, in press (Paper II).

Lin, R. P., *et al.*: 2002, *Solar Phys.* **210**, 3.

Lin, R. P., *et al.*: 2003, *Astrophys. J. Lett.* **595**, L69.

Piana, M.: 1994, *Astron. Astrophys.* **288**, 949.

Piana, M. and Brown, J. C.: 1998, *Astron. Astrophys.* **132**, 291.

Piana, M., Massone, A. M., Kontar, E. P., Emslie, A. G., Brown, J. C., and Schwartz, R. A.: 2003, *Astrophys. J. Lett.* **595**, L127.

Smith, D. M., *et al.*: 2002, *Solar Physics* **210**, 33.

Thompson A. M., Brown J. C., Craig, I. J. D., and Fulber, C.: l992, *Astron. Astrophys.* **265**, 278.

Tikhonov, A. N.: 1963, *Sov. Math. Dokl.* **4**, 1035.

Tikhonov, A. N., Goncharsky, A., Stepanov, V., and Yagola, A.: 1995, *Numerical methods for the solution of ill-posed problems*, Kluwer: Dordrecht.

van Loan, C. F.: 1976, *SIAM J. Num. Anal.* **13**, 76.