

**JYX**



**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Brander, Tommi; Ilmavirta, Joonas; Piironen, Petteri; Tyni, Teemu

**Title:** Optimal recovery of a radiating source with multiple frequencies along one line

**Year:** 2020

**Version:** Accepted version (Final draft)

**Copyright:** © 2020 American Institute of Mathematical Sciences

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Brander, T., Ilmavirta, J., Piironen, P., & Tyni, T. (2020). Optimal recovery of a radiating source with multiple frequencies along one line. *Inverse Problems and Imaging*, 14(6), 967-983.  
<https://doi.org/10.3934/ipi.2020044>

# Optimal recovery of a radiating source with multiple frequencies along one line

Tommi Brander<sup>1</sup>, Joonas Ilmavirta<sup>2</sup>, Petteri Piiroinen<sup>3</sup>, and Teemu Tyni<sup>4</sup>

<sup>1</sup>Norwegian University of Science and Technology, Department of Mathematical Sciences; tommi.brande@ntnu.no

<sup>1</sup>Technical University of Denmark, Department of Applied Mathematics and Computer Science

<sup>2</sup>University of Jyväskylä, Department of Mathematics and Statistics; joonas.ilmavirta@jyu.fi

<sup>3</sup>University of Helsinki, Department of Mathematics and Statistics; petteri.piiroinen@helsinki.fi

<sup>4</sup>University of Oulu, Research Unit of Mathematical Sciences; teemu.tyni@helsinki.fi

25th May 2020

## Abstract

We study an inverse problem where an unknown radiating source is observed with collimated detectors along a single line and the medium has a known attenuation. The research is motivated by applications in SPECT and beam hardening. If measurements are carried out with frequencies ranging in an open set, we show that the source density is uniquely determined by these measurements up to averaging over levelsets of the integrated attenuation. This leads to a generalized Laplace transform. We also discuss some numerical approaches and demonstrate the results with several examples.

# 1 Introduction

We consider the following one-dimensional inverse problem: The intensity of radiation from an unknown source in a known medium is measured at multiple frequencies. How uniquely does this determine the density of the source? A more detailed description of the model is given in section 1.1 below.

This physical problem boils down to the following mathematical question on the interval  $I = (a, b)$ : Given a function  $p: I \rightarrow \mathbb{R}$ , does the knowledge of the function

$$D(\lambda) = \int_a^b \lambda^{p(x)} \rho(x) dx \tag{1}$$

for  $\lambda$  in an open set  $U \subset (0, 1)$  determine the function  $\rho: I \rightarrow \mathbb{R}$  uniquely? Uniqueness of  $\rho$  depends on the properties of  $p$  in a peculiar way.

We denote by  $P: L^2(I) \rightarrow L^2(I)$  the unique projection onto  $L^2(I, \sigma(p))$ , the subspace of  $L^2(I)$  consisting of  $\sigma(p)$ -measurable functions. Here  $\sigma(p)$  is the smallest sigma-algebra on  $I$  which makes  $p$  measurable and contains sets of zero Lebesgue measure. This operator satisfies for any  $\rho \in L^2(I)$  that the function  $P\rho$  is  $\sigma(p)$ -measurable and for every  $\sigma(p)$ -measurable  $A \subset I$  we have

$$\int_A P\rho(x) dx = \int_A \rho(x) dx.$$

If  $I$  has measure one, the operator is the conditional expectation  $\mathbb{E}[\rho|\sigma(p)]$ ; for more on conditional expectation we refer to the books [7, chapter 2][15, chapter 5]. For more details on the operator, see section 2.1.

The conclusion is that the data  $D$  does not in general determine  $\rho$ , but it does determine the projection  $P\rho$  and nothing more. We have unique determination precisely when  $\sigma(p)$  is the whole Lebesgue algebra. This is formulated precisely in our main theorem, which we prove in two different languages; by a measure-theoretical approach in section 2 and a probabilistic one in section 3:

**Theorem 1.** *Suppose  $\rho \in L^2(I)$  and  $p \in L^\infty(I)$ . Let  $U \subset (0, 1)$  be a nonempty open set. Then the following are equivalent:*

1. *The function  $D(\lambda)$  defined in (1) vanishes for all  $\lambda \in U$ .*
2.  *$P\rho = 0$ , that is, the function  $\rho$  is orthogonal to  $L^2(I, \sigma(p))$ .*
3. *For all  $A \in \sigma(p)$  it holds that  $\int_A \rho(x) dx = 0$ .*

The probabilistic approach also works with  $\rho \in L^1(I)$  and with  $U$  a sequence with a cluster point. The next corollary follows immediately.

**Corollary 2.** *If the functions  $D_1$  and  $D_2$  arise from the functions  $\rho_1$  and  $\rho_2$  by (1), then  $D_1 = D_2$  if and only if  $P\rho_1 = P\rho_2$ .*

*The linear map  $\rho \mapsto D$  is injective if and only if  $\sigma(p)$  is the whole Lebesgue algebra. This happens, in particular, when  $p$  is injective.*

The function  $D: U \rightarrow \mathbb{R}$  defined by (1) is the data, and it depends linearly on the unknown function  $\rho$ . Therefore it follows from the theorem that a function  $\rho \in L^2(I)$  is determined by  $D$  up to an element of the space  $L^2(I, \sigma(p))^\perp \subset L^2(I)$ , which is the kernel of the linear operator  $\rho \mapsto D$ . Another way of viewing this is that the push-forward of  $\rho dx$  under  $p$  is recovered.

The theorem completely characterizes what can be said about  $\rho$ , given  $p$  and  $D$ . If the attenuation coefficient is strictly positive (see section 1.1), then  $p$  is strictly increasing and the source density  $\rho$  is determined fully uniquely.

For a concrete situation where the result applies, consider a medium composed of a single material. That material has been CT scanned so as to determine its density  $\beta(x)$ . Then suppose another substance is added, and it emits radiation on a broad spectrum. The density of this new material on a single line can be monitored by measuring intensity of radiation coming along that line at multiple frequencies with a collimated detector. If  $\beta > 0$ , this information determines the density  $\rho$  of the source uniquely. To monitor the intensity of the source as a function of time after the initial CT scan, it is sufficient to use a single line for measurements. This idea is similar to that used in SPECT and PET where a CT scan is first needed to map the attenuation before something else is used to image the radiating source. For more on related imaging modalities, see section 1.2.

## 1.1 The model

Suppose the source of radiation is  $f(\omega, x)$  and the attenuation coefficient is  $\mu(\omega, x)$ . The frequency  $\omega$  takes values in an open set  $U \subset (0, \infty)$  and  $x \in I = (a, b)$ . The measurement is the intensity of radiation at  $a$ , which is given by

$$M(\omega) = \int_a^b e^{-\int_a^x \mu(\omega, y) dy} f(\omega, x) dx$$

according to Beer-Lambert’s law [14]. The physical problem is to recover the source  $f$  using the measurement  $M(\omega)$  for a large number of frequencies  $\omega$ .

To do this, structural assumptions on  $\mu$  and  $f$  are needed. We assume that they both factor:  $\mu(\omega, x) = \alpha(\omega)\beta(x)$  and  $f(\omega, x) = \epsilon(\omega)\rho(x)$ . The functions  $\beta$  and  $\rho$  can be regarded as the spatial densities of the absorbent and the source. The functions  $\alpha$  and  $\epsilon$  correspond to “spectral densities” and depend on the physical process behind attenuation and emission. We note that similar assumptions have been made in an earlier study [10, remark 1].

If both the absorbent and the source are composed of a single material, this factorization is well-justified. The rate of absorption or emission is directly proportional to the density, and the functions  $\alpha$  and  $\epsilon$  are simply the coefficients of proportionality which may well — and generally do — depend on frequency. If there are multiple materials, the frequency dependence can be different for the different materials, and the overall attenuation coefficient and source no longer factorize.

We introduce two auxiliary functions,  $\phi(\omega) = e^{-\alpha(\omega)}$  and  $p(x) = \int_a^x \beta(y)dy$ . We work on a frequency range where  $\alpha(\omega) > 0$ , and so  $\phi(\omega) = \lambda \in (0, 1)$ . We assume that there is a function  $\eta: U \rightarrow \mathbb{R}$  so that  $\phi(\eta(\lambda)) = \lambda$  for all  $\lambda \in U$ . That is,  $\eta$  is a right inverse of  $\phi$ , and it exists for some interval  $U$  if, for example,  $\alpha$  is continuously differentiable and non-constant.

With these assumptions the measurement  $M(\omega)$  may be processed to yield the data

$$D(\lambda) = \frac{M(\eta(\lambda))}{\epsilon(\eta(\lambda))}$$

used in the equation (1) as a function of  $\lambda$ .

## 1.2 Background

In the inverse problem presented above, the goal is to reconstruct the source when the attenuation is known. Well-studied imaging modalities with the same goal are instances of emission computer tomography. We mention single-photon emission computed tomography (SPECT) and positron emission tomography (PET).

In both SPECT and PET a radioactive substance is injected into the target. The substance decays and emits gamma radiation, which is detected outside the target. From these measurements, one tries to reconstruct the location of the radioactive substance. In SPECT, the substance emits single gamma-ray photons, which are then detected. In PET, it emits positrons,

which soon combine with an electron, shooting two gamma ray photons to opposite directions; this pair is then detected [18, 10, 31]. In SPECT, some radioactive substances radiate at several frequencies [31].

In both PET and SPECT the reconstruction is improved if the anatomy of the target is known a priori [10]. A CT scan is a common approach. Once the anatomy is known, the values of attenuation can be recovered based on known values. We likewise assume a known attenuation and an unknown source; hence, we are investigating a model of one-dimensional multispectral SPECT/PET.

When a multispectral X-ray beam passes through a material, low-energy photons are typically attenuated more strongly than high-energy ones, which changes the frequency profile of the beam. This is called beam hardening [4]. Our model is consistent with this phenomenon – we assume the factorization  $\mu(\omega, x) = \alpha(\omega)\beta(x)$ , where the  $\alpha$  is the dependency of the attenuation on the frequency. Our model can not only take beam hardening into account but make use of it.

Multispectral (also multi-energy, multichromatic, spectral, spectroscopic, energy-selective, energy-sensitive, energy discrimination, or colour) X-ray tomography started with the work of Alvarez and Macovski [1]. Incoming photons are classified in a number, often two, of energy bins according to their energy by the photon-counting detectors, after which one can make separate reconstructions at different energy levels or attempt a joint reconstruction from all the available information. The two energy levels are especially natural due to Compton and photoelectric effects [1]. The main challenges are that the measurement devices are expensive and that the smaller amount of radiation leads to worse reconstructions at every energy level [30]. For more on multispectral X-ray tomography we refer to the reviews [17, 19].

Recovering source terms in an attenuating medium can be formulated as the attenuated or exponential X-ray or Radon transform [6, section 8.8][13][18, section VI-C][29, chapter 8], which is a multidimensional theory that only uses measurements at a single frequency. The attenuated Radon transform has an explicit inversion formula [24, 23, 9]. The present reconstruction theory is one-dimensional, and the X-ray transform with or without attenuation is never injective in one dimension. The key is to use several different attenuations or weights along the fixed line, and in our case this is achieved by using a large number of frequencies.

A different way of using several weights is to study the momentum ray transform [29, 16]. In that problem, one defines the momentum ray transform

for points  $(x, \xi) \in TS^{d-1}$  by

$$(I^k f)(x, \xi) = \int_{-\infty}^{\infty} t^k \langle f(x + t\xi), \xi^m \rangle dt$$

for suitable tensor fields  $f$ , and tries to recover  $f$  from the integrals indexed by all points  $(x, \xi)$  and sufficient number of powers  $k$ . The usual X-ray transform  $I^0$  of a tensor field only determines the field up to a gauge, but using moments up to the order of the tensor field determines it uniquely. For recent results in tensor tomography, we refer to [25, 26, 13], and we also mention the classical book of Sharafutdinov [29]. In the same spirit of using moments, in the works [5, 20, 21] the moments of noisy measurements are related to the moments of the unknown density function.

The Hausdorff moment problem [28, 32] asks: Given a sequence  $(s_n)$ , does there exist a measure  $\mu$  such that, for all  $n \in \mathbb{N}$ ,

$$s_n = \int_0^1 x^n d\mu?$$

If it exists, is it unique? In our problem, we know that the measure  $\rho(x)dx$  exists and want to understand its uniqueness, or, in the language of moment problem literature, determinacy. More fundamentally, whereas in the moment problem one couples the measure with polynomials, we use polynomials of a function  $p$ , which might not be continuous or injective.

Our main theorem turns out to be similar to an inverse problems result for the variable exponent  $p(\cdot)$ -Laplacian [3]. The methods are quite similar, but in the variable exponent case the equivalent of the measurements  $\lambda$  are explicitly related to a quantity  $K_\lambda$ , which cannot be expressed analytically as a function of  $\lambda$ , and the measurements are, using the notations in this paper,

$$\int_I \rho(x) K_\lambda^{p(x)/(p(x)-1)}.$$

The intermediate quantity  $K_\lambda$  causes complications in the proofs, but the final results are similar to the present ones. In another result the exponent  $p$  has been recovered [2], which suggest that it might be possible to recover the attenuation from known sources in the present model.

Our problem can also be seen as inverting a generalized Laplace transform [32]. Namely, if  $\rho$  is continuously differentiable and satisfies  $\rho' > 0$ ,

then after changing the variable of integration in (1) from  $x$  to  $y = p(x)$  and writing  $\lambda = e^{-\alpha}$  (see section 1.1), the data can be rewritten as

$$D(\lambda) = \tilde{D}(\alpha) = \int_{\tilde{a}}^{\tilde{b}} e^{-\alpha y} \tilde{\rho}(y) dy,$$

where  $\tilde{a} = p(a)$ ,  $\tilde{b} = p(b)$ , and  $\tilde{\rho}(y) = \rho(p^{-1}(y))/p'(p^{-1}(y))$ . Therefore  $\tilde{D}$  is the Laplace transform of  $\tilde{\rho}$ . However, when  $p$  is less well-behaved, such reduction to Laplace transform does not work. We also point out that putting  $p(x) = x$  in theorem 1 implies that if  $\rho$  is a compactly supported  $L^2$  function whose Laplace transform vanishes on an interval, then  $\rho = 0$ . This corollary is of course not new [32, chapter 2, section 6].

One can also think of equation (1) as a Fredholm integral equation of the first kind for the compact operator  $\tilde{D}: L^2(I) \rightarrow L^2(U)$  given by  $\tilde{D}(\rho)(\lambda) := \int_a^b \lambda^{p(x)} \rho(x) dx$ .

### 1.3 Discussion

If  $p$  is piecewise constant — which corresponds to the attenuation being a sum of delta functions corresponding to thin absorbent films — then the data  $D$  determines the average of  $\rho$  on every piece. Nothing else is determined, and the averages of  $\rho$  over the levelsets of  $p$  is optimal information.

If  $p$  is strictly increasing — which corresponds to strictly positive attenuation — then heuristically the levelsets are points and the averages should determine  $\rho$  uniquely. This is indeed the case, as the sigma-algebra generated by  $p$  is the full Lebesgue algebra.

Non-negativity of the attenuation coefficient implies that  $p$  is increasing. The model can be extended to the case where  $p: (a, b) \rightarrow [0, \infty]$ , where  $p = \infty$  corresponds to the intensity being attenuated all the way to zero. If  $p = \infty$  on some subinterval  $[b', b) \subset (a, b)$ , then the data tells nothing about  $\rho$  on this subinterval as  $\lambda^p = 0$  on this set. Therefore we may restrict the problem to the interval  $(a, b')$  with no loss of data. It is thus reasonable to assume that on the interval of interest  $p < \infty$ .

Very weak regularity assumptions on the attenuation coefficient  $\beta$  are sufficient. If  $\beta \in L^1(I)$ , then  $p$  is absolutely continuous, which is more than enough for our theorem.

Physically  $p$  is increasing (since  $\beta \geq 0$ ), but for mathematical purposes it may be any bounded measurable function. If  $I = [-1, 1]$  and  $p(x) = x^2$ ,



then the sigma-algebra  $\sigma(p)$  generated by  $p$  consists of subsets of  $I$  that are symmetric with respect to reflection up to an error of measure zero. Then the data  $D$  determines the symmetric part of  $\rho$  uniquely but does not constrain the antisymmetric part at all.

The rough idea of the proof is that linear combinations of functions  $x \mapsto \lambda^{p(x)}$  indexed by  $\lambda \in U$  can be used to approximate any  $\sigma(p)$ -measurable function  $I \rightarrow \mathbb{R}$ . Thus  $D = 0$  is equivalent with  $\rho$  being perpendicular to the subspace of these functions. This implies that variations of  $\rho$  within levelsets of  $p$  are undetectable. This phenomenon is well-illustrated by a simple observation which we provide next.

**Proposition 3.** *Suppose  $c \in \mathbb{R}$  and let  $\bar{\rho} \in L^2(I)$  be such that  $\text{supp}(\bar{\rho}) \subseteq p^{-1}(\{c\})$  and*

$$\int_{p^{-1}(\{c\})} \bar{\rho}(x) dx = 0.$$

*Then, for all  $\lambda \geq 0$ ,*

$$\int_I (\bar{\rho}(x) + \rho(x)) \lambda^{p(x)} dx = \int_I \rho(x) \lambda^{p(x)} dx.$$

The proof is a straightforward calculation.

## Acknowledgements

T.B. was partially funded by grant no. 4002-00123 from the Danish Council for Independent Research — Natural Sciences and partially by the Research Council of Norway through the FRIPRO Toppforsk project ”Waves and nonlinear phenomena”. J.I. was supported by the Academy of Finland (decision 295853). T.T. was supported by the Academy of Finland (application number 312123, Centre of Excellence of Inverse Modelling and Imaging 2018–2025).

We would like to thank Alexander Meaney for a discussion and valuable references and the anonymous referees for insightful comments and suggestions.

## 2 Main results

In this section we specify notation and prove the main theorem. An alternative stochastic proof can be found in section 3.

## 2.1 Definitions and notation

We denote by  $\mathcal{L}$  the Lebesgue sigma-algebra on  $I = (a, b)$ . We always use the Lebesgue measure  $dx$ . The sigma-algebra  $\sigma(p)$  generated by a function  $p: I \rightarrow \mathbb{R}$  is also a sigma-algebra on  $I$ , and we define it as the smallest sigma-algebra so that  $p$  is measurable and sets of zero Lebesgue outer measure are measurable. For more on sigma-algebras generated by sets and functions, see the books [7, chapter 1, definition 5][15, chapter 1].

The following lemma states that  $\sigma(p)$  does not depend on the representative of  $p$ .

**Lemma 4.** *If  $g = h$  almost everywhere, then  $\sigma(g) = \sigma(h)$ .*

*Proof.* Write as  $\tilde{\sigma}(g)$  the preimage  $g^{-1}(\sigma(\mathbb{R}))$ , which is a sigma-algebra [15, lemma 1.3]. Suppose  $A \in \tilde{\sigma}(g)$ . There then exists a measurable  $B \subset \mathbb{R}$  such that  $A = g^{-1}(B)$ . Now

$$(A \setminus h^{-1}(B)) \cup (h^{-1}(B) \setminus A) \subseteq \{x \in I; g(x) \neq h(x)\},$$

which is a null set. Because  $A = h^{-1}(B)$  up to null sets, we have  $A \in \sigma(h)$ .

If  $A' \in \sigma(g)$ , then it is equal to some  $A \in \tilde{\sigma}(g)$  up to null sets, and by the above reasoning  $A' \in \sigma(h)$ .  $\square$

The space  $L^q(I, \sigma(p))$  is the space of  $\sigma(p)$ -measurable functions in  $L^q(I)$  up to almost everywhere equality. If  $\sigma(p) = \mathcal{L}$  (the Lebesgue sigma-algebra), then  $L^q(I, \sigma(p)) = L^q(I)$ .

Since  $L^2(I, \mathcal{L})$  is a complete Hilbert space and  $L^2(I, \sigma(p))$  a closed convex set, we can define the following unique orthogonal projection.

**Definition 5.** The mapping  $P$  is the orthogonal projection  $P: L^2(I, \mathcal{L}) \rightarrow L^2(I, \sigma(p))$ .

## 2.2 Lemmas

**Lemma 6.** *For all  $A \in \sigma(p)$  the projection  $P$  satisfies*

$$\int_A f dx = \int_A P f dx.$$

*Proof.* Since  $A \in \sigma(p)$ , the characteristic function  $\chi_A$  is in  $L^2(I, \sigma(p))$ . On the other hand,  $f - Pf \in (L^2(I, \sigma(p)))^\perp$ . Hence,

$$\int_A f dx = \int_I \chi_A (Pf + (f - Pf)) dx = \int_I \chi_A Pf dx = \int_A Pf dx.$$

□

**Lemma 7.** *If a mapping  $Q: L^2(I, \mathcal{L}) \rightarrow L^2(I, \sigma(p))$  satisfies, for all  $A \in \sigma(p)$ ,*

$$\int_A f dx = \int_A Qf dx,$$

*then it is the orthogonal projection onto  $L^2(I, \sigma(p))$ .*

*Proof.* The map  $Q$  is a linear projection by definition.

For the characteristic function  $\chi_A$  of any  $A \in \sigma(p)$  we have

$$\int_I \chi_A (f - Qf) dx = 0,$$

whence  $f - Qf$  is orthogonal to  $L^2(I, \sigma(p))$ , since any function there can be approximated by measurable step functions. Because the range of  $Q$  is  $L^2(I, \sigma(p))$  and  $f$  is arbitrary, this shows orthogonality. □

Recall the data  $D: \mathbb{R}_+ \rightarrow \mathbb{R}$ ,

$$D(\lambda) = \int_a^b \lambda^{p(x)} \rho(x) dx.$$

**Lemma 8.** *Suppose  $\lambda_0$  is an interior point of an open set  $U \subset [0, 1]$ . If  $D(\lambda) = 0$  for all  $\lambda \in U$ , then*

$$\int_a^b r(p(x)) \lambda_0^{p(x)} \rho(x) dx = 0 \tag{2}$$

*for all polynomial functions  $r$ .*

*Proof.* The function  $D$  is smooth. In fact, it is complex analytic in a neighborhood of  $U$  as a consequence of Morera's theorem [27, theorem 10.17].

For any natural number  $k$  applying the operator  $\lambda \frac{d}{d\lambda}$  to  $D$  a total of  $k$  times gives

$$\left( \lambda \frac{d}{d\lambda} \right)^k D(\lambda) = \int_a^b \lambda^{p(x)} \rho(x) (p(x))^k dx.$$

Derivatives of all orders vanish at  $\lambda = \lambda_0$ , whence

$$\int_a^b \lambda_0^{p(x)} \rho(x) (p(x))^k dx = 0$$

for all  $k$ . Finite linear combinations of these integrals give equation (2).  $\square$

The identity (2) holds for any function  $r$  that can be approximated by polynomials in a suitable sense. We will next study this approximation.

The following multiplicative system theorem is a version of the monotone class theorem [15, theorem 1.1] written in terms of functions, rather than sets [7, chapter 1, theorems 19–21].

**Lemma 9** (Multiplicative system theorem, [7, chapter 1, theorem 21]). *Suppose  $H$  is a vector space of real-valued bounded measurable functions on a measurable space  $X$ . Suppose  $H$  contains constant functions and is closed under the pointwise convergence of uniformly bounded increasing sequences of functions. Let  $E \subseteq H$  be closed under pointwise multiplication, and let  $\mathcal{G}$  be the sigma-algebra generated by  $E$ .*

*Then  $H$  contains all bounded  $\mathcal{G}$ -measurable functions.*

We note that the multiplicative system theorem considers a vector space of functions  $H$  and a subset  $E$ , whereas we operate with Lebesgue spaces of equivalence classes of functions.

We consider the set of functions

$$\tilde{E} = \{\lambda_0^{p(\cdot)} r(p(\cdot)); r \text{ polynomial}\} \subset L^\infty(I).$$

Recall that  $\sigma(p)$  is the smallest sigma-algebra that makes  $p$  measurable and contains sets of measure zero.

**Lemma 10.** *The closure of  $\tilde{E}$  in the  $L^2(I)$ -norm is  $L^2(I, \sigma(p))$ .*

The proof is by Nathaniel Eldredge [8], and similar to proofs in variable exponent Calderón’s problem [3, lemmas 18 and 27]. We omit the space  $L^2$  from the notation of the closure.

*Proof of lemma 10.* The function  $\lambda_0^{p(\cdot)}$  is  $\sigma(p)$ -measurable and bounded away from zero and infinity. Therefore it may be “divided out” and it suffices to prove the lemma in the case  $\lambda_0 = 1$ . We write as  $E$  the space of polynomials of  $p$ .

Every continuous function of  $p$  is  $\sigma(p)$ -measurable, so the equivalence classes of the functions in  $E$  form a subspace of  $L^2(I, \sigma(p))$ . The space  $L^2(I, \sigma(p))$  is closed in  $L^2(I)$ , since a converging sequence in  $L^2$  has a pointwise almost everywhere converging subsequence [27, theorem 3.12], and the limit of such a subsequence is, as the limit of measurable functions, measurable with respect to the same sigma-algebra. Hence, we have  $\overline{E} \subseteq L^2(I, \sigma(p))$ , where we understand  $\overline{E}$  as a space of equivalence classes of functions.

For the other direction,  $L^2(I, \sigma(p)) \subseteq \overline{E}$ , we start by considering the vector space  $H = \{f; [f] \in \overline{E} \cap L^\infty(I)\}$ , which consists of all functions whose equivalence classes are in  $\overline{E} \cap L^\infty(I)$ . It satisfies all the assumptions of the multiplicative system theorem:

- Constant functions are bounded and polynomials of the function  $p$ .
- If a uniformly bounded sequence converges pointwise, then the sequence of the squared absolute values of the functions also does so, as do their equivalence classes. By monotone convergence, the  $L^p$  norms of the sequence converge.

We have that all representatives of  $E$  are in  $H$  and  $E$  is closed under pointwise multiplication. By multiplicative system theorem (lemma 9),  $H$  contains all bounded  $\sigma(p)$ -measurable functions, whence  $L^2(I, \sigma(p)) \cap L^\infty(I) \subseteq \overline{E} \cap L^\infty(I)$ .

Consider a (not necessarily bounded) equivalence class of functions  $h \in L^2(I, \sigma(p))$  and define

$$h_n(x) = \max(-n, \min(h, n)).$$

Then, for all  $n \in \mathbb{N}$ ,  $h_n \in H \subset \overline{E}$  and  $h_n \rightarrow h$  in  $L^2(I)$  as  $n \rightarrow \infty$ , so  $h \in \overline{E}$ .  $\square$

We note that we use a slightly different formalism than the paper of Brander and Winterrose [3]; here,  $\sigma(p)$  is a completion of a sigma-algebra, whereas they instead use a mapping  $[g]_{\sigma_p} \rightarrow [g]_{\mathcal{L}}$ , which maps equivalence classes without the completion into equivalence classes with it. This leads to superficial differences in the proof of lemma 10 when compared to the similar proofs of [3, lemmas 18 and 27].

## 2.3 Proof of theorem 1

We are now ready to prove our main result. Let us first recall equation (1):

$$D(\lambda) = \int_a^b \lambda^{p(x)} \rho(x) dx.$$

We also recall for convenience the main theorem, which is stated as:

**Theorem** (Theorem 1). *Suppose  $\rho \in L^2(I)$  and  $p \in L^\infty(I)$ . Let  $U \subset (0, 1)$  be a nonempty open set. Then the following are equivalent:*

1. *The function  $D(\lambda)$  vanishes for all  $\lambda \in U$ .*
2.  *$P\rho = 0$ , that is, the function  $\rho$  is orthogonal to  $L^2(I, \sigma(p))$ .*
3. *For all  $A \in \sigma(p)$  it holds that  $\int_A \rho(x) dx = 0$ .*

Now we are ready to prove it.

*Proof of theorem 1.* 1  $\implies$  2: By lemma 8 the  $L^2$  inner product between  $\rho$  and all polynomials of  $p$  is zero. This implies that  $\rho$  is orthogonal to the  $L^2$  closure of the space of polynomials of  $p$ , which, by lemma 10, is  $L^2(I, \sigma(p))$ . Hence the projection is also zero.

2  $\implies$  1: Let  $\lambda \geq 0$ . Since  $\lambda^{p(x)}$  is  $\sigma(p)$ -measurable and bounded, we have  $\lambda^{p(x)} \in L^2(I, \sigma(p))$ . By orthogonality the  $L^2$ -inner product in equation (1) is zero.

2  $\iff$  3: This is lemmas 6 and 7.

If  $p$  injective, then  $\sigma(p) = \mathcal{L}$  and injectivity of  $\rho \mapsto D$  follows.  $\square$

## 3 Probabilistic interpretation

In this section we prove the main theorem in probabilistic language. We make frequent use of the basic properties of conditional measures [15, section 5, especially theorem 5.1] [7, chapter 2, number 41]. Let  $\mathcal{F}$  be the sigma-algebra of Lebesgue-measurable sets on  $\mathbb{R}$  (sometimes restricted to  $I$  without changing the notation) and let  $\mathbb{P}(A) = \frac{dx}{b-a}$  be the Lebesgue measure rescaled into a probability measure. Then  $(I, \mathcal{F}, \mathbb{P})$  is a probability space. We use the Iverson bracket notation

$$[p \in A] = \begin{cases} 1, & \text{if } p(x) \in A \\ 0, & \text{if } p(x) \notin A. \end{cases}$$

**Theorem 11.** Suppose  $\rho \in L^1(I, \mathcal{F}, \mathbb{P})$  and  $p \in L^\infty(I, \mathcal{F}, \mathbb{P})$ . Let  $U \subset \mathbb{R}_+$  be a nonempty open set and write  $D(\lambda) = \mathbb{E}\lambda^p\rho$ .

Then the following are equivalent:

1. For all  $\lambda \in U$  we have  $D(\lambda) = 0$ .
2.  $\mathbb{E}(\rho|p) = 0$  holds  $\mathbb{P}$ -a.s.
3. For all  $A \in \mathcal{F}$  we have  $\mathbb{E}[p \in A]\rho = 0$ .

*Remark 12.* In 1 in the theorem, it is sufficient that there exists a countably infinite set of frequencies  $\lambda$  with a cluster point  $\lambda_0$  such that  $D(\lambda) = 0$ .

We prove the theorem with the remark included, as the remark is stronger than the theorem.

*Proof.* We show that the first and the second condition are equivalent and that the second and the third condition are equivalent.

**2  $\implies$  1:** For all  $\lambda \in \mathbb{R}_+$

$$\begin{aligned} D(\lambda) &= \mathbb{E}\lambda^p\rho = \mathbb{E}(\mathbb{E}(\lambda^p\rho|p)) \\ &= \mathbb{E}(\lambda^p\mathbb{E}(\rho|p)) = 0. \end{aligned}$$

**2  $\iff$  3:** Define  $\xi = \mathbb{E}(\rho|p)$ . Then, by definition,  $\xi$  is the unique  $\sigma(p)$ -measurable random variable in  $L^1(I, \mathcal{F}, \mathbb{P})$  satisfying  $\mathbb{E}[p \in A]\xi = \mathbb{E}[p \in A]\rho$ . Now, 2 implies 3:

$$\mathbb{E}([p \in A]\rho) = \mathbb{E}([p \in A]\mathbb{E}(\rho|p)) = 0.$$

On the other hand, 3 implies 2. For all  $A \in \mathcal{F}$

$$\mathbb{E}[p \in A]\rho = 0 = \mathbb{E}[p \in A] \cdot 0.$$

Thus, by uniqueness of  $\xi$ , we have  $\xi = 0$  as a random variable, which is claim 2.

**1**  $\implies$  **2**: Let  $V$  be a countably infinite set of points  $\mu$  with cluster  $\mu_0$ . Write  $\mu_j = \log \lambda_j$  and suppose they converge to  $\mu_0$  as a strictly increasing or a strictly decreasing sequence (by taking a subsequence without changing the notation). Since  $p$  is bounded, we can define  $\widetilde{D}$  by  $\widetilde{D}(\mu) = D(e^\mu) = 0$  for all  $\mu \in V$ .

By using  $j + 1$  consecutive points from the sequence  $(\mu_k)$ , that is the points  $\mu_n, \mu_{n+1}, \dots, \mu_{n+j}$ , we can construct the Lagrange polynomial  $L_{j,n}$  approximation of order  $j$  for a function  $f$ , namely

$$L_{j,n}(\mu) = \sum_{i=0}^j l_{j,n,i}(\mu) f(\mu_{n+i})$$

where the function  $l_{j,n,i}$  is the Lagrange basis polynomial

$$l_{j,n,i}(\mu) = \prod_{k=0, k \neq i}^j \frac{\mu - \mu_k}{\mu_i - \mu_k}.$$

These are defined for every function  $f$ , for every  $n = 1, 2, \dots$  and for every  $j = 1, 2, \dots$ .

The  $j$ th order Lagrange polynomial  $L_{j,n}$  coincides with the function  $f$  (at least) in  $j + 1$  points in the interval  $I_n$  between  $\mu_n$  and  $\mu_{n+j}$ , so by Rolle's Theorem the  $j$ th order derivative  $L_{j,n}^{(j)}$  coincides with the  $f^{(j)}$  at least in one point in the interval  $I$ . Therefore, by the Lagrange approximation, the derivatives of order  $j$  of the error

$$R_{j,n}(\mu) = f(\mu) - L_{j,n}(\mu)$$

for a smooth function  $f \in C^{j+1}$  is bounded in the interval  $I_n$  by

$$\sup_{\mu \in I_n} |\partial_\mu^{(j)} R_{j,n}(\mu)| \leq \sup_{\mu \in I} |f^{(j+1)}(\mu)| |\mu_n - \mu_{n+j}|.$$

Since the distance  $|\mu_n - \mu_{n+j}| \rightarrow 0$  as  $n \rightarrow \infty$ , the  $j$ th derivative of the Lagrange polynomial  $\partial_\mu^{(j)} L_{j,n}$  evaluated at  $\mu_n$  is an approximation of the  $j$ th derivative of the function  $f$  evaluated at  $\mu_n$ . Let  $a_{n,j,i} = \partial_\mu^{(j)} l_{n,j,i}(\mu_n)$  be the corresponding coefficient in the expansion

$$(\partial_\mu^{(j)} L_{j,n})(\mu_n) = \sum_{i=0}^j \partial_\mu^{(j)} l_{n,j,i}(\mu_n) f(\mu_{n+i}) = \sum_{i=0}^j a_{n,j,i} f(\mu_{n+i})$$



Note that the coefficients don't depend on the function  $f$ , but only on the indices  $n, i, j$  and the sequence  $(\mu_n)$ . Hence, we assume that the function  $f(\mu) = e^{\mu p}$ , which is a smooth function with random coefficient  $p$  in the exponent. Moreover, the  $j$ th derivative of  $f$  is  $f^{(j)}(\mu) = p^j f(\mu)$ .

Therefore, we have for every  $n$  and every  $j$  that

$$\mathbb{E}(\rho L_{j,n}(\mu_n)) = \mathbb{E}\left(\rho \sum_{i=0}^j a_{n,j,i} e^{p\mu_{n+1}}\right) = \sum_{i=0}^j a_{n,j,i} \widetilde{D}(\mu_{n+i}) = 0$$

since  $\widetilde{D}(\mu) = 0$  for every  $\mu \in V$ . Since the random variable  $p$  is bounded, we may apply dominated convergence and we obtain

$$\mathbb{E}(\rho p^j e^{\mu_0 p}) = \lim_{n \rightarrow \infty} \mathbb{E}(\rho L_{j+1,n}(\mu_n)) = 0$$

for every  $j = 0, 1, \dots$  that implies that with linearity that we have  $\mathbb{E}r(p)e^{\mu_0 p} \rho = 0$  for all polynomials  $r$ .

The same approximation would be obtained with iterated differences, when these are defined recursively for sequences  $\lambda = (f(\mu_n))_n$  as

$$\begin{aligned} d(1, \mu, \lambda)_n &= \frac{\lambda_{n+1} - \lambda_n}{\mu_{n+1} - \mu_n} \\ d(j+1, \mu, \lambda)_n &= \frac{d(j, \mu, \lambda)_{n+1} - d(j, \mu, \lambda)_n}{\mu_{n+j+1} - \mu_n} \end{aligned}$$

where the high-order differences use the right scaling. If the differences would form a grid, these would be the usual higher order stencil approximations. However, showing that this works in general is easiest to show via Rolle's Theorem and therefore, leads directly to the Lagrange approximation.

Now, by standard argument, since  $p$  has compact range, the boundedness of  $p$  implies by Stone-Weierstrass that for all continuous functions  $f$  we have  $\mathbb{E}f(p)e^{\mu_0 p} \rho = 0$ . Monotone convergence then implies that

$$\begin{aligned} &\mathbb{E}[p \in J] e^{\mu_0 p} \rho = 0 \text{ for all intervals } J \\ \implies &\mathbb{E}(e^{\mu_0 p} \rho | p) = 0 \text{ holds } \mathbb{P}\text{-a.s.} \\ \iff &e^{\mu_0 p} \mathbb{E}(\rho | p) = 0 \text{ holds } \mathbb{P}\text{-a.s.,} \end{aligned}$$

which implies claim 2 in the theorem. □

## 4 Numerical demonstration

To generate the synthetic measurement data  $D_j$ , we compute the integrals

$$\int_0^1 \lambda_j^{p(x)} \rho_0(x) dx = D_j \quad (3)$$

by using the Simpson  $(2n + 1)$ -point quadrature rule, where  $n = 64$ . In all of our examples, the measurements are corrupted by small additive Gaussian white noise with standard deviation  $\sigma$  equal to 0.5% of the maximum of absolute value of the measurements.

Next, to solve the inverse problem of recovering  $\rho_0$  in (3) from the measurements  $D_j$  we first substitute instead of  $\rho_0$  the piecewise linear form

$$\rho_0(x) \approx \sum_{k=1}^N f_k \phi_k(x),$$

where  $\phi_k$  are the hat-functions

$$\phi_k(x) = \begin{cases} \frac{x - x_{k+1}}{x_k - x_{k+1}}, & x_k \leq x \leq x_{k+1} \\ \frac{x - x_{k-1}}{x_k - x_{k-1}}, & x_{k-1} \leq x < x_k, \\ 0, & \text{otherwise.} \end{cases}$$

Then (3) becomes

$$\sum_{k=1}^N f_k \int_0^1 \lambda_j^{p(x)} \phi_k(x) dx = D_j + E_j, \quad (4)$$

where  $E_j$  is an error term which we ignore in the sequel. In practise, the integrals appearing in (4) can be numerically precomputed for a given function  $p$ . Finally equation (4) can be written as the linear system  $Af = D$  and solved by some regularization method.

We avoid committing inverse crime [22] by generating the measurements  $D_j$  independently of the theory matrix  $A$ . Indeed, the measurements are obtained by integrating the precise model (3) and adding noise, while the theory matrix  $A$  is computed by only integrating the known function  $\lambda^p(\cdot)$  against  $\phi_k$ ,  $k = 1, \dots, N$ , while avoiding the use of the unknown  $\rho_0$ . Further, we shall use smaller number of measurements  $D_j$  than the number of unknowns  $f_k$ .

In our case the number of measurements and unknowns will be 300 and 400, respectively.

The computation is not very demanding and can be expected to work on a modern computer. The numerical work is done in MATLAB.

## 4.1 Regularization

Since the properties of this problem depend quite drastically on the functions  $p$  and  $\rho$ , at this point we opt to not propose any universal solution to the problem. Rather, this part should be regarded as a demonstration that the problem can in principle be solved numerically. As is usual in inverse problems, the linear problem  $Af = D$  is rather unstable and regularization methods are necessary. We have tested Tikhonov, total variation (TV), and conjugate gradient least squares (CGLS) regularization methods.

In Tikhonov regularization [11, 22] one solves the minimization problem

$$\arg \min \left\{ \|Af - D\|_2^2 + \alpha \|f\|_2^2 \right\}.$$

This classical regularization method is very simple to implement and in our tests works well with low noise-levels when the unknown  $\rho_0$  is reasonably smooth. The down-side to the  $L^2$ -penalty is that it promotes smoother solutions, failing to recover discontinuous and irregular functions.

The TV-regularization instead aims to minimize the expression

$$\arg \min \left\{ \|Af - D\|_2^2 + \alpha \|f'\|_1 \right\}.$$

The  $L^1$ -penalty term allows some steep gradients and thus can be used when  $\rho$  is more irregular [11, 22].

Finally, the CGLS method is an iterative regularization procedure, where one solves the least-squares problem

$$\arg \min \left\{ \|Af_k - D\|_2^2 \right\}$$

subject to the condition that

$$f_k \in \text{span} \left\{ A^T D, (A^T A) A^T D, \dots, (A^T A)^{k-1} A^T D \right\}, \quad k = 0, 1, 2, \dots,$$

that is, the iterates belong to a Krylov subspace. Heuristically, this method attempts to pick only the significant singular components of the solution  $f_k$ , because the Krylov subspaces take into account the (noisy) data  $D$ . Using only the significant singular components allows one to reduce the effect of noise. We use the CGLS-algorithm of Hestenes and Stiefel; see e.g. [11, 12].

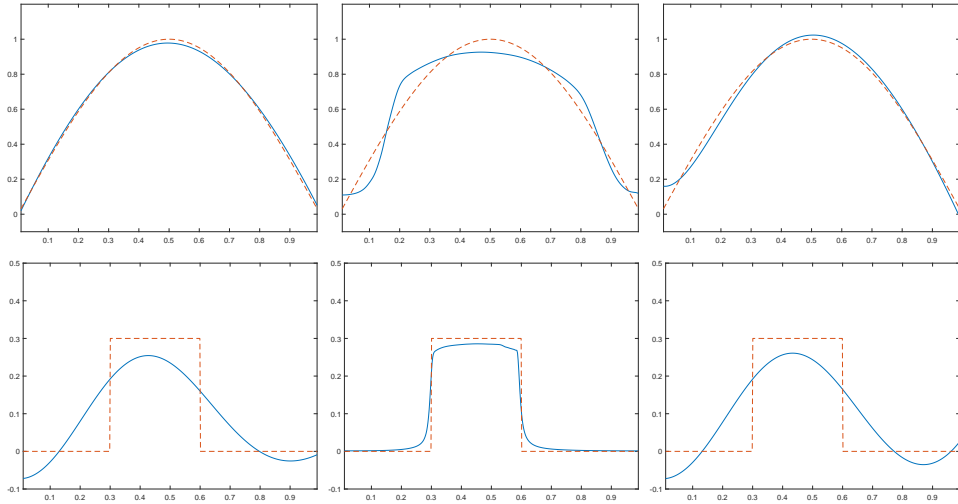


Figure 1: Unknown  $\rho_0$  (dashed red line) and the numerical solution  $\rho$  (solid blue line) with 0.5% noise level. Example 1 (above) and example 2 (below) with Tikhonov-solution (left), TV-solution (middle) and CGLS-solution (right).

## 4.2 Examples

Let the number of measurements be  $M = 300$  and the number of unknown coefficients of the hat functions be  $N = 400$ . In these examples the measurement points  $\lambda_j$  are uniformly distributed on the interval  $(0, 1)$ . We also tested different distributions of  $\lambda_j$  on the interval  $(0, 1)$ , but often a good result was obtained with uniform distribution. Certainly this choice depends on the function  $p$  and can be tailored for applications separately. The functions used in the examples are as follows:

- Example 1:  $p(x) = x$  and  $\rho_0(x) = \sin(\pi x)$ .
- Example 2:  $p(x) = x$  and  $\rho_0(x) = 0.3\chi(x)$ , where  $\chi(x)$  is the characteristic function of the interval  $(0.3, 0.6)$ .

The results are presented in Figure 1, where the precise unknown  $\rho_0$  is depicted with the red dashed line and the numerically computed solutions  $\rho$  are shown as the solid blue line. Apparently Tikhonov- and CGLS-solutions work rather well with smooth  $\rho_0$ , while in the discontinuous case the TV-regularized solution is considerably better.

### 4.3 The sets where $p$ is constant

If the function  $p$  is constant in some set of positive measure, the theory predicts that we can only hope to solve (3) up to the average of  $\rho_0$  in that set. Let  $M = 300$ ,  $N = 400$ , and  $\rho_0(x) = \sin(\pi x)$ . The function  $p$  is given in the piecewise form

$$p(x) = \begin{cases} 5/3x, & x \in (0, 0.2], \\ 1/3, & x \in (0.2, 0.4], \\ 5/3x - 1/3, & x \in (0.4, 0.6], \\ 2/3, & x \in (0.6, 0.8], \\ 5/3x - 2/3, & x \in (0.8, 1]. \end{cases}$$

Here the optimal information to recover is  $\sin(\pi x)$  on the intervals  $(0, 0.2]$ ,  $(0.4, 0.6]$ , and  $(0.8, 1]$  and only the average value of  $5/2\pi \approx 0.796$  on the intervals  $(0.2, 0.4]$  and  $(0.6, 0.8]$ . The results are depicted in Figure 2, where the unknown  $\rho_0$  is shown in red dashed line, the projection  $P\rho_0$  in black dot-dash line and the numerical solution in solid blue line.

The numerical method is expected to produce a function whose projection (averages over sets where  $p$  is constant) is  $P\rho_0$ . There are many such functions, and the choice depends on regularization. With Tikhonov one expects to find the function with minimal  $L^2$  norm – which is precisely  $P\rho_0$  – but with other regularizations something else.

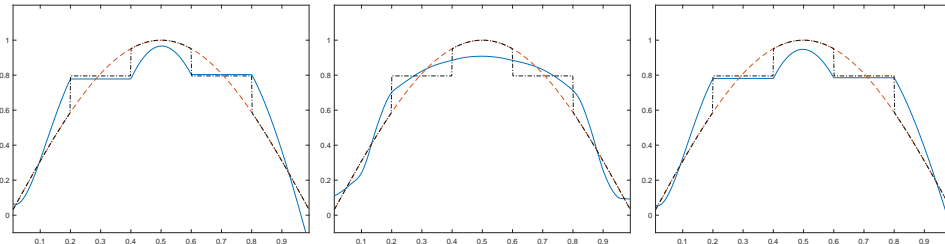


Figure 2: Unknown  $\rho_0$  (dashed red line),  $\rho_0$  averaged over regions where  $p$  is constant (black dot-dash line) and the numerical solution  $\rho$  (solid blue line) with 0.5% noise level. Tikhonov-solution (left), TV-solution (middle) and CGLS-solution (right).

## 4.4 Limited data

The theory states that it suffices to have measurements for  $\lambda_j$ 's in some open interval  $(a, b) \subset (0, 1)$ . Let  $M = 300$ ,  $N = 400$ , the source density  $\rho_0(x) = -0.5 \sin(2\pi x) + 0.5$ , and  $p(x) = e^x - 1$ . To test the solutions, we used several smaller intervals, solved the problem with Tikhonov regularization one hundred times, and collected the averaged  $l^2$ -relative errors in the solutions to Table 1. The respective solutions  $\rho$  are depicted in Figure 3. These solutions are relatively good for all but the smallest interval  $(0.4, 0.5)$ , where the numerical solution is rather unstable. Note that the number of measurements is kept at constant 300 for all solutions. We state without details that similar results are obtained if the measurements are made in some union of open intervals (with large enough measure) contained in  $(0, 1)$ .

Intervals	$(0, 1)$	$(0.2, 0.8)$	$(0.3, 0.6)$	$(0.4, 0.5)$
$\bar{\epsilon}_{\text{rel}}$	0.117	0.170	0.186	0.320
var	$1.88 \cdot 10^{-3}$	$1.27 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$6.65 \cdot 10^{-3}$

Table 1: Averaged relative errors and variances of one hundred solutions on smaller intervals with noise level 0.5%.

## References

- [1] Robert E Alvarez and Albert Macovski. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine & Biology*, 21(5):733, 1976.
- [2] Tommi Brander and Jaakko Siltakoski. Recovering a variable exponent. February 2020. Preprint arXiv:2002.06076.
- [3] Tommi Brander and David Winterrose. Variable exponent Calderón's problem in one dimension. *Annales Academiae Scientiarum Fennicae, Mathematica*, 44:925–943, 2019. DOI 10.5186/aasfm.2019.4459.
- [4] Rodney A Brooks and Giovanni Di Chiro. Beam hardening in X-ray reconstructive tomography. *Physics in Medicine & Biology*, 21(3):390, May 1976.

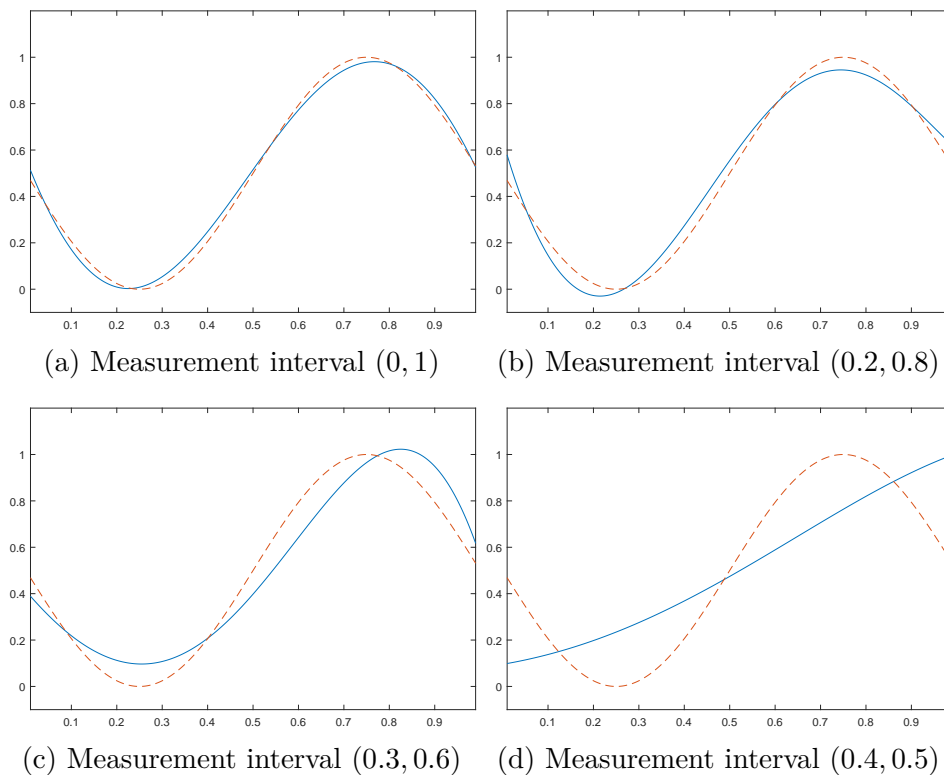


Figure 3: Unknown  $\rho_0$  (dashed red line) and the Tikhonov-solution  $\rho$  (solid blue line) with 0.5% noise level and smaller measurement intervals.

- [5] Hayoung Choi, Victor Ginting, Farhad Jafari, and Robert Mnatsakanov. Modified Radon transform inversion using moments. *arXiv e-prints*, page arXiv:1809.09673, March 2019.
- [6] Stanley R. Deans. *The Radon Transform and Some of Its Applications*. Dover Books on Mathematics Series. Dover Publications, 2007.
- [7] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam – New York, 1978.
- [8] Nathaniel Eldredge. Closure of polynomials of a function in  $L^2$ . Math-Overflow, February 2018. <https://mathoverflow.net/a/292978/1445>.

- [9] David V Finch. The attenuated X-ray transform: recent developments. In Gunther Uhlmann, editor, *Inside Out: Inverse Problems and Applications*, number 47 in Mathematical sciences research institute publications, pages 47–66. Cambridge university press, 2003.
- [10] Daniel Gourion and Dominikus Noll. The inverse problem of emission tomography. *Inverse Problems*, 18(5):1435, 2002.
- [11] Per Christian Hansen. *Discrete inverse problems: Insight and algorithms*. SIAM, USA, 2010.
- [12] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, 49:409–436, 1952.
- [13] Joonas Ilmavirta and François Monard. Integral geometry on manifolds with boundary and applications. In Ronny Ramlau and Otmar Scherzer, editors, *The Radon Transform: The First 100 Years and Beyond*. de Gruyter, to appear.
- [14] James D Ingle Jr and Stanley R Crouch. *Spectrochemical analysis*. Prentice Hall College Book Division, Old Tappan, NJ, USA, January 1988.
- [15] Olav Kallenberg. *Foundations of modern probability*. Probability and its applications. Springer, 1997.
- [16] Venkateswaran P Krishnan, Ramesh Manna, Suman Kumar Sahoo, and Vladimir A. Sharafutdinov. Momentum ray transforms. *Inverse problems and imaging*, 13(3):679–701, 2019.
- [17] L. A. Lehmann and R. E. Alvarez. Energy-selective radiography a review. In James G. Kereiakes, Stephen R. Thomas, and Colin G. Orton, editors, *Digital Radiography: Selected Topics*, pages 145–188. Springer US, Boston, MA, 1986.
- [18] Robert M Lewitt and Samuel Matej. Overview of methods for image reconstruction from projections in emission computed tomography. *Proceedings of the IEEE*, 91(10):1588–1611, September 2003.
- [19] Cynthia H. McCollough, Shuai Leng, Lifeng Yu, and Joel G. Fletcher. Dual- and multi-energy CT: Principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, September 2015.



- [20] Peyman Milanfar. *Geometric estimation and reconstruction from tomographic data*. PhD thesis, Massachusetts Institute of Technology, June 1993.
- [21] Peyman Milanfar, William Clement Karl, and Alan S Willsky. A moment-based variational approach to tomographic reconstruction. *IEEE Transactions on Image Processing*, 5(3):459–470, March 1996.
- [22] Jennifer L. Mueller and Samuli Siltanen. *Linear and nonlinear inverse problems with practical applications*, volume 10 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012.
- [23] Frank Natterer. Inversion of the attenuated Radon transform. *Inverse Problems*, 17(1):113–119, January 2001.
- [24] Roman G Novikov. An inversion formula for the attenuated X-ray transformation. *Arkiv för matematik*, 40(1):145–167, April 2002.
- [25] Gabriel P. Paternain, Mikko Salo, and Gunther Uhlmann. Tensor tomography: Progress and challenges. *Chinese Annals of Mathematics. Series B*, 35(3):399–428, 2014.
- [26] Gabriel P. Paternain, Mikko Salo, and Gunther Uhlmann. Invariant distributions, Beurling transforms and tensor tomography in higher dimensions. *Mathematische Annalen*, 363(1-2):305–362, 2015.
- [27] Walter Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.
- [28] Konrad Schmüdgen. *The Moment Problem*, volume 277 of *Graduate Texts in Mathematics*. Springer International Publishing, 2017.
- [29] Vladimir Altafovich Sharafutdinov. *Integral geometry of tensor fields*. Inverse and Ill-Posed Problems series. VSP, Utrecht, 1994.
- [30] Katsuyuki Taguchi and Jan S. Iwaczyk. Vision 20/20: Single photon counting x-ray detectors in medical imaging. *Medical Physics*, 40(10):100901, September 2013.

- [31] Andy Welch, Grant T Gullberg, Paul E Christian, Jia Li, and Benjamin MW Tsui. An investigation of dual energy transmission measurements in simultaneous transmission emission imaging. *IEEE Transactions on Nuclear Science*, 42(6):2331–2338, December 1995.
- [32] David Vernon Widder. *Laplace Transform*. Number 6 in Princeton Mathematical Series. Princeton University Press, 1946.