European Journal of Applied Mathematics (2026), 1–33
doi:10.1017/S0956792526100527

CAMBRIDGE UNIVERSITY PRESS

PAPERS

On the existence of optimal set-valued decoders and their accuracy bounds for ill-posed inverse problems

Nina Maria Gottschling[1], Paolo Campodonico[2], Vegard Antun[3] and Anders C. Hansen[4]

[1]CCSD, Oak Ridge National Laboratory, USA
[2]DAMTP, University of Cambridge, UK
[3]University of Oslo Faculty of Mathematics and Natural Sciences, Norway
[4]University of Cambridge, USA
Corresponding author: Nina Maria Gottschling; Email: gottschlinnm@ornl.gov

Received: 13 May 2025 UTC; **Revised:** 09 August 2026 UTC; **Accepted:** 10 August 2026 UTC

Keywords: Inverse problems; Bayesian inverse problems; optimal decoders; compressed sensing; learning in inverse problems

2020 Mathematics Subject Classification: 41A46, 65J22, 15A29 (Primary); 94A08, 68T07, 28B05, 26E25, 28B20, 35R30 (Secondary)

Abstract

Ill-posed inverse problems occur everywhere in the sciences, including medical imaging, radar, astronomy, etc., yielding underdetermined or ill-posed linear (non-linear) reconstruction problems. There are now a myriad of techniques to design decoders/reconstruction methods that can tackle such problems, ranging from optimisation-based approaches, such as compressed sensing, to data-driven techniques such as deep learning (DL) and variants in between the two techniques. The variety of methods begs for a unifying approach to determine the existence of optimal decoders and fundamental accuracy bounds in order to facilitate a theoretical and empirical understanding of the performance of existing and future methods. Such a theory must allow for both single-valued and set-valued decoders, as underdetermined and ill-posed inverse problems typically have multiple solutions. Indeed, set-valued decoders arise due to non-uniqueness of minimisers in optimisation problems, such as in compressed sensing, and for DL-based decoders in generative adversarial models, such as diffusion models and ensemble models. In this work, we provide a framework for assessing the lowest possible reconstruction accuracy in terms of worst-case and average errors. The universal bounds only depend on the measurement model F, the model class $\mathscr{M}_1$ and the noise model $\mathscr{E}$. For linear F, these bounds depend on its kernel, and in the non-linear case, the concept of kernel is generalised for undersampled and ill-posed settings. Additionally, we provide set-valued variational solutions that obtain the lowest possible reconstruction error.

1. Introduction

Finite-dimensional inverse problems are ubiquitous in the computational sciences, as they naturally appear in a plethora of applications. An incomplete list of examples includes most types of computational imaging [1, 16, 30], matrix completion [34], parametric PDEs/system identification [2], phase retrieval [36, 52, 97], quantised sampling [29], etc. Most often, we can express these problems using the following general model:

$$\text{Given noisy measurements } y = F(x, e) \text{ of } x \in \mathscr{M}_1 \quad \text{and} \quad e \in \mathscr{E}, \text{ recover } x. \tag{1}$$

This manuscript has been co-authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Here, x represents the signal of interest, while e represents the noise. The sets $\mathscr{M}_1$ and $\mathscr{E}$ describe the signals of interest and the potential noise, respectively. The function $F\colon \mathscr{M}_1 \times \mathscr{E} \to \mathscr{M}_2^{\mathscr{E}}$ models the forward process and is deliberately kept general to encompass many known models. Examples include

$$y = G(x) + e, \qquad \text{(additive noise)} \tag{2}$$

$$y = G(x) \odot e, \qquad \text{(multiplicative noise)} \tag{3}$$

$$y = G(x) \odot e_1 + e_2, \qquad \text{(mixture of multiplicative and additive noise)} \tag{4}$$

where $G\colon \mathscr{M}_1 \to \mathscr{M}_2$ is a linear or non-linear forward map, and the dimension of the set $\mathscr{E}$ depends on the considered model in such a way that the point-wise multiplication $\odot$ is defined.

In most instances of interest, the problem of recovering x given y is ill-posed or undersampled, unless further assumptions are made. An inverse problem is ill-posed if it is not well-posed in the sense of Hadamard [63, 64]. In particular, it is ill-posed if either there does not exist a solution x for every y, or the solution is not unique, or the solution does not depend continuously on y. We consider the case when the solution is not unique. This can be the case if the forward model F is either poorly conditioned or dimensionality reducing. A standard assumption for making the problem (1) tractable is to assume that the noiseless forward model is injective when restricted to the model class $\mathscr{M}_1$. This assumption is a cornerstone for models with additive noise, as it allows for accurate reconstruction up to the noise level in both the linear [31, 56, 104] and non-linear [73] settings. Traditionally, the set $\mathscr{M}_1$ has been given a precise mathematical description, and reconstruction methods have been designed for the given choice of $\mathscr{M}_1$. Examples of sets $\mathscr{M}_1$ include sparse vectors [35], unions of subspaces [26], manifolds [15], sparsity in a given basis [4] or frame [92], matrices with low rank [37, 70], etc.

More recently, data-driven approaches [12, 69] have emerged as an alternative to many of these standard methods, often promising superior performance [112]. Methods based on data typically do not specify the solution set $\mathscr{M}_1$, but aim to learn the reconstruction mapping $\Psi\colon \mathscr{M}_2 \to \mathscr{M}_1$ given a finite number of training data $\mathscr{T} \subset \mathscr{M}_1$ and access to the forward model F [12]. This approach has the advantage that the learned set $\mathscr{M}_1$ might provide a better approximation of the underlying data stemming from a real-world process, rather than a potentially simplistic abstract mathematical modelling of the set $\mathscr{M}_1$. However, the challenge with these methods is that the learned mapping Ψ tends to produce accurate reconstructions of all elements in $\mathscr{T}$, regardless of whether the noiseless problem is injective on $\mathscr{M}_1$ or not. This has made these methods susceptible to both hallucinations [62, 85] and instabilities [10].

The purpose of this work is to develop a mathematical framework for inverse problems which provides a notion of optimality when F is not necessarily injective on $\mathscr{M}_1$, and to provide a variational expression for the optimal mappings in this setting. In summary, this work answers the question posed by M. Burger and T. Roith in *Learning in Image Reconstruction: A Cautionary Tale* [33],

> *Can we ever quantify the reconstruction error [of ill-posed inverse problems]?*

Our main contributions are the following:

1. *We provide a framework which allows for the classification of the lowest achievable reconstruction accuracy for* (1). Note that many such frameworks exist in the literature. For example, Gelfand widths in the noiseless linear setting [89], best k-term approximation [41], and to obtain the worst-case bounds set-valued decoders on Banach spaces have been considered in [11] and then extended to metric spaces in [81], and when noise is included, one has the notion of optimal learning [25] and generalised instance optimality [31]. See also [83] for a survey of early works on optimal recovery. In many of these works, the underlying spaces are infinite- or finite-dimensional Banach spaces, and the reconstruction error is measured by the given norm. A contribution of this work is to consider the general setting of metric spaces, to allow for set-valued decoders which naturally arise as solutions of undersampled inverse problems and to extend previous studies on worst-case scenarios to the average error scenario.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

2. *We derive explicit formulas for set-valued reconstruction mappings which achieve the lowest possible reconstruction error, both in a worst-case sense and in a probabilistic sense.* Finding and analysing reconstruction mappings which achieve optimal or near-optimal recovery – given some notion of optimality – is an essential question in inverse problems, and many mappings exist [25, 31, 56, 90, 104]. In this work, we provide explicit formulas for possibly set-valued reconstruction mappings that achieve the lowest possible reconstruction error. Moreover, a contribution of this work is to prove that the optimal mappings in the average error scenario are measurable, compact valued and admit a measurable single-valued selector by utilising the Measurable Maximum Theorem, [6, Theorem 18.19] and Hausdorff distance. As such, this work provides a first and necessary step before any numerical procedure can provide statistical approximations to these mappings.
3. *We provide lower and upper bounds on the lowest achievable reconstruction accuracy described in Section 1.* The bounds solely depend on the forward operator, signal and noise class of the inverse problem and not on the method or decoder used to solve (1). Hence, the performance of any – possibly set-valued – method can be evaluated with respect to the problem's intrinsic optimal accuracy. However, obtaining lower bounds comes at a price: in contrast to compressed sensing results, which typically provide upper bounds (see [31, 56]), fewer assumptions are required, and for a non-zero lower bound, the reconstruction error can not vanish. Specifically, the trade-off is as follows: lower bounds can be derived without assuming that the forward model F is injective on $\mathscr{M}_1$, at the cost that the reconstruction error admits a non-zero limit. A key point in order to obtain lower bounds on reconstruction accuracy is to consider worst-case and average errors instead of point-wise approximation errors. Such worst-case error bounds can be related to the Chebyshev radius, as in [25], and analogous worst-case and average error bounds are established in [90] in the case of normed spaces and single-valued decoders.

Remark. *Extending previous work from normed spaces to metric spaces is of particular relevance in imaging applications. In fact, the ℓ^2 or ℓ^1 norm are not the only candidates considered for image quality assessment. A prominent example is the structural similarity index (SSI), firstly proposed to assess image quality in [108], does not constitute a norm but can rather be related to a metric [32]. Note that the proposed framework encompasses all metrics, except trivial degenerate metrics.*

Remark. (**Set-valued decoders, computability and randomised algorithms**). *Many of the decoders provided in theory will not be computable, as the phenomenon of generalised hardness of approximation [9, 17, 45, 53, 61] – within the Solvability Complexity Index (SCI) hierarchy [20, 21, 43, 44, 65] – happen in many inverse problems. This includes a wide range of neural network decoders, as well as compressed sensing decoders. Typically, they can only be computed to a certain accuracy $\epsilon_0 > 0$, referred to as the approximation threshold, and it is the size of the approximation threshold that determines if the decoder can be used in practice. However, as demonstrated in [17], when the decoder is set-valued, randomised algorithms may actually help mitigate the non-computability. This is only possible for set-valued decoders, as single-valued decoders that are non-computable will have no help from randomised algorithms [17, 47].*

2. Preliminaries

2.1. The forward map and the model class

In the following, we introduce the main assumptions and setup. Let $\mathscr{X}$, $\mathscr{Y}$ and $\mathscr{Z}$ be non-empty sets. Furthermore, consider subsets $\mathscr{M}_1 \subset \mathscr{X}$ and $\mathscr{E} \subset \mathscr{Z}$, referred to as the *model class* and the *noise class*, respectively. As outlined in the introduction, we consider inverse problems with a given forward map $F\colon \mathscr{M}_1 \times \mathscr{E} \to \mathscr{M}_2^{\mathscr{E}} \subset \mathscr{Y}$, with

$$\mathscr{M}_2^{\mathscr{E}} = F(\mathscr{M}_1 \times \mathscr{E}) = \{y \in \mathscr{Y} \colon \exists (x, e) \in \mathscr{M}_1 \times \mathscr{E} \text{ s.t. } y = F(x, e)\}$$

denote the image of $F\colon \mathscr{X} \times \mathscr{Z} \to \mathscr{Y}$ on the class $\mathscr{M}_1 \times \mathscr{E}$. We consider inverse problems as:

$$\text{Given noisy measurements } y = F(x, e) \text{ of } x \in \mathscr{M}_1 \quad \text{and} \quad e \in \mathscr{E}, \text{ recover } x, \tag{5}$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

where e represents noise and x is the signal of interest. The goal of solving the inverse problem is to produce an approximation $\hat{x}$ of the true solution x. In order to quantify the error of the approximation $\hat{x} \in \mathscr{X}$ to the sample signal $x \in \mathscr{M}_1$, it is common to assume [41, 56, 89] that $\mathscr{X}$ is a normed space, or even a Banach or Hilbert space. In the current work, we extend the scope to the more general setting of metric spaces. The use of metrics – rather than norms – allows for a more general theory, in which normed vector spaces are a special case. Thus, we equip the sets $\mathscr{X}$, $\mathscr{Y}$ and $\mathscr{Z}$ with metrics $d_{\mathscr{X}}$, $d_{\mathscr{Y}}$ and $d_{\mathscr{Z}}$, respectively, and on each set, we consider the topology induced by the respective metric. To avoid certain pathological cases for metric spaces, we make assumptions on the choice of metrics involved. Our first assumption is the following.

Assumption 1.

(1) We assume that the metrics $d_{\mathscr{X}}$, $d_{\mathscr{Y}}$, $d_{\mathscr{Z}}$ are chosen so that the topologies induced by the metrics are second countable (they admit a countable base).

(2) We assume that the metrics $d_{\mathscr{X}}$ and $d_{\mathscr{Z}}$ satisfy the Heine-Borel property, i.e., that all closed and bounded sets are compact.

Here, the first assumption ensures that the topologies induced by the metrics are separable, which is a convenient assumption for both theoretical and practical purposes. In particular, this assumption excludes metrics such as the discrete metric over an uncountable set, as one cannot find a countable base for an uncountable set using this metric.

For the second assumption, we recall that in metric spaces, a compact subset is automatically closed and bounded; however, the converse does not necessarily hold, unless the metrics satisfy the Heine-Borel property. Note that this 2 is not implied by 1, as for example the bounded metric $d_{\mathscr{X}}(x, y) = \min\{\|x - y\|_{\ell^2}, 1\}$ on $\mathscr{X} = \mathbb{C}^N$ satisfies assumption 1 (since $d_{\mathscr{X}}$ induces the same topology as the Euclidean topology, which is second countable), but not assumption 2 (since $\mathbb{C}^N$ is closed and bounded but not compact with respect to $d_{\mathscr{X}}$). Finally, note that together with second countability, the Heine-Borel property implies in particular that $\mathscr{X}$ is a complete separable metric space, usually referred to as a *Polish space*.

2.2. Set-valued reconstruction maps

Now that the preliminaries have been clarified, we return to the problem of solving the inverse problem in (5). A *reconstruction mapping* for (5) is a mapping that takes a measurement $y = F(x, e)$ and returns one or more approximations $\hat{x}$ to the unknown signal x. That is, a *reconstruction mapping* is a set-valued function $\varphi\colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$, where φ takes values in the power set $2^{\mathscr{X}}$ of the set $\mathscr{X}$. Note that it is necessary to consider set-valued maps, as model-based methods often write the reconstruction mapping as an optimisation problem. These problems might not have unique solutions; see, e.g., [45, SI, Section 3.B] for a few standard examples. Finally, we mention that given a set-valued mapping $\varphi\colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$, we call a function $f\colon \mathscr{M}_2^{\mathscr{E}} \to \mathscr{X}$ a *selector* for φ if it satisfies $f(y) \in \varphi(y)$ for each $y \in \mathscr{M}_2^{\mathscr{E}}$.

Distances between sets in $\mathscr{X}$ will be measured via the Hausdorff distance. Given subsets $A, B \subseteq \mathscr{X}$, their *Hausdorff distance* is

$$d_{\mathscr{X}}^H(A, B) := \max\left\{ \sup_{a\in A} \inf_{b\in B} d_{\mathscr{X}}(a, b), \sup_{b\in B} \inf_{a\in A} d_{\mathscr{X}}(a, b) \right\}.$$

The Hausdorff distance satisfies all the properties of a metric only when restricted to the subsets of $\mathscr{X}$ that are bounded (which ensures $d_{\mathscr{X}}^H$ is finite) and closed (which ensures that $d_{\mathscr{X}}^H(A, B) = 0$ only if $A = B$). With a slight abuse of notation, we will denote $d_{\mathscr{X}}^H(\{a\}, B)$ by $d_{\mathscr{X}}^H(a, B)$. The Hausdorff distance between a point and a set should not be confused with the usual distance between a point and a set, $\operatorname{dist}_{\mathscr{X}}(a, B) := \inf_{b\in B} d_{\mathscr{X}}(a, b)$. Indeed, note that $\operatorname{dist}_{\mathscr{X}}(a, B) = \inf_{b\in B} d_{\mathscr{X}}(a, b) \le \sup_{b\in B} d_{\mathscr{X}}(a, b) = d_{\mathscr{X}}^H(a, B)$.

We require an extra assumption that relates the forward map F and the model class $\mathscr{M}_1$. In particular, we require that for every y, the set of possible true solutions x's is bounded. Throughout the text, we denote by $\pi_1\colon \mathscr{X} \times \mathscr{Z} \to \mathscr{X}, (x, e) \mapsto x$ the projection on the first component.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Assumption 2. *We assume that for every $y \in \mathscr{M}_2^{\mathscr{E}}$ the* feasible set

$$F_y := \pi_1(F^{-1}(y)) = \{x \in \mathscr{M}_1 : \exists e \in \mathscr{E} \text{ s.t. } F(x,e) = y\} \tag{6}$$

is bounded.

The feasible set F_y consists of all the candidate solutions x's that are consistent with the measurement y, for some realisation of the noise e. Note that the definition of $\mathscr{M}_2^{\mathscr{E}}$ implies that the feasible set is non-empty. Additionally, in Remark 3.2, we outline assumptions on the forward map such that the feasible set contains more than one element.

The previous condition is satisfied, for example, if the model class $\mathscr{M}_1$ is compact, as assumed in e.g., [25].

Finally, we clarify our notation for objects in a metric space (X, d). We denote by $B_d(x, r)$ the *closed* ball of centre $x \in X$ and radius $r \geq 0$. If $A \subseteq X$, then the closed ball around A of radius $r \geq 0$ is

$$B_d(A, r) = \{x \in X : \operatorname{dist}_d(x, A) \leq r\} = \bigcup_{x \in A} B_d(x, r)$$

where we recall that $\operatorname{dist}_d(x, A) = \inf_{a \in A} d(x, a)$.

Finally, the diameter of $A \subseteq X$ is denoted by $\operatorname{diam}_d(A) := \sup\{d(a, a') : a, a' \in A\}$. The subscript d is omitted whenever the metric is unambiguous from context.

3. Main results

This section introduces the inverse problem framework studied in this paper, highlighting the key concepts of an *optimal mapping* and the *kernel size*, which serve as bounds on the reconstruction error. The section is divided into two subsections that both introduce the definitions of an *optimal mapping* and the *kernel size*, but for different measures of accuracy. In particular, Section 3.1 presents an analysis based on the *worst-case* error, whereas Section 3.2 considers a probabilistic model class, where an *average* error is analysed. The main results, given in Theorems 3.1 and 3.2 for Sections 3.1 and 3.2, respectively, bound the *optimality constant* of an optimal mapping in terms of the *kernel sizes* and provide an explicit construction of an optimal reconstruction mapping. Finally, in Section 4, we elaborate on how these quantities relate to common frameworks in the literature. The proofs of the main results are referred in Section 7.

3.1. Worst-case optimality bounds and an optimal map with worst-case noise

What is the best possible reconstruction error one can achieve for a given model class $\mathscr{M}_1$? This is an old question [83] that has been asked many times and in different settings [25, 31, 41, 50, 58, 73, 104, 105]. In the definition below, we investigate this question for measurements contaminated by worst-case noise. This definition has appeared in [50, 58] for the linear model with additive noise in a normed vector space and is here generalised to metric spaces.

Definition 1 (Optimality constant with worst-case noise). *The optimality constant with worst-case noise of the inverse problem (1) is*

$$c_{\text{opt}}^{w}(F, \mathscr{M}_1, \mathscr{E}) = \inf_{\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}} \sup_{x \in \mathscr{M}_1} \sup_{e \in \mathscr{E}} d_{\mathscr{X}}^{H}(x, \varphi(F(x,e))).$$

A mapping $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ that attains such an infimum is called an optimal map with worst-case noise.

Given the optimality constant with worst-case noise of an inverse problem (1), the question arises on how to upper and lower bound such an optimality constant. As shown in Theorem 3.1, the upper and lower bounds can be obtained by referring to a constant intrinsic to the problem, its *kernel size with worst-case noise*, as defined below.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Definition 2 (Kernel size with worst-case noise). *The kernel size with worst-case noise of the problem (1) is*

$$\operatorname{kersize}^{w}(F, \mathscr{M}_1, \mathscr{E}) = \sup_{\substack{(x,e),(x',e')\in \mathscr{M}_1\times\mathscr{E} \text{ s.t.} \\ F(x,e)=F(x',e')}} d_{\mathscr{X}}(x, x'). \tag{7}$$

Remark. *The kernel size with worst-case noise has been considered under different names in previous work. This includes, but is not limited to, the supremum taken over the measurements y of the diameter of the Chebyshev balls of the feasible sets F_y, similar to [25], or the diameter of information in [90]. However, as this work focuses on characterising accuracy bounds for undersampled and ill-posed inverse problems, we opt for referring to the above-defined constant as kernel size. In Example 1 in Section 4, more details on this relation are presented.*

The kernel size with worst-case noise gives the maximum distance between any two points $x, x' \in \mathscr{M}_1$ with identical measurement $y = F(x, e) = F(x', e')$ for some noise vectors $e, e' \in \mathscr{E}$. It is worth observing that the optimality constant and the kernel size with worst-case noise also can be expressed as follows.

$$\begin{aligned} c^{\mathrm{w}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}) &= \inf_{\varphi\colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}} \sup_{y\in \mathscr{M}_2^{\mathscr{E}}} \sup_{x\in F_y} d^H_{\mathscr{X}}(x, \varphi(y)) \\ \operatorname{kersize}^{\mathrm{w}}(F, \mathscr{M}_1, \mathscr{E}) &= \sup_{y\in \mathscr{M}_2^{\mathscr{E}}} \operatorname{diam}_{d_{\mathscr{X}}}(F_y). \end{aligned} \tag{8}$$

This establishes a connection of the optimality constant and the kernel size with the worst-case to the feasible sets F_y. In our first theorem, we provide an upper and lower bound on the optimality constant in terms of the kernel size with worst-case noise. Moreover, we provide a variational expression for an optimal map with worst-case noise.

Theorem 3.1 (Worst-case optimality bounds). *Under Assumptions 1 and 2, the following holds.*

(1) We have that

$$kersize^{w}(F, \mathscr{M}_1, \mathscr{E})/2 \le c^{w}_{opt}(F, \mathscr{M}_1, \mathscr{E}) \le kersize^{w}(F, \mathscr{M}_1, \mathscr{E}). \tag{9}$$

(2) The map

$$\Psi(y) = \operatorname*{argmin}_{z\in\mathscr{X}} \sup_{x\in F_y} d_{\mathscr{X}}(x, z) = \operatorname*{argmin}_{z\in\mathscr{X}} d^H_{\mathscr{X}}(z, F_y), \tag{10}$$

has non-empty, compact values and it is an optimal map with worst-case noise.

The previous theorem illustrates a fundamental limit for the inverse problem (5). Indeed, for (5), one would hope to find a solution whose error is as close to zero as possible. However, the lower bound in (9) shows that there is a fundamental constant intrinsic to the problem – the kernel size with worst-case noise – such that no worst-case reconstruction error can be made smaller than this constant for all possible choices of $x \in \mathscr{M}_1$. Moreover, note that in the definition of the worst-case optimality constant, all potentially set-valued maps are included – in particular, no single-valued selector can obtain a smaller worst-case reconstruction error. Additionally, we can apply the axiom of choice to select a single-valued map $\bar{\Psi}$ from the set-valued optimal map in (10). By the definition of the Hausdorff distance, we have that this single-valued selection satisfies $d_{\mathscr{X}}(x, \bar{\Psi}(F(x, e))) \le \sup_{z\in\Psi(F(x,e))} d_{\mathscr{X}}(x, z) = d^H_{\mathscr{X}}(x, \Psi(F(x, e)))$ for all $x \in \mathscr{M}_1$ and $e \in \mathscr{E}$. Hence, by the minimisation property of the optimal map in (10), the bounds also hold for any single-valued selector.

Remark. *Note that analogous bounds as in (9) under slightly different assumptions have been obtained in a variety of previous works, including but not limited to [11, 25, 81, 90]. However, to the best of our knowledge, the characterisation of the set-valued optimal map (10) as compact-valued has not been considered under the same assumptions.*

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

3.2. Average optimality bounds and optimal map with average error

The previous theorem provides fundamental limits when performance is assessed by considering worst-case reconstruction error for a given inverse problem of the form (5). Such bounds naturally lead to very pessimistic estimates of performance. This motivates the need for a model that considers the average error of a reconstruction mapping, given a probabilistic model of the data. Below, we will adapt the worst-case setup considered in the previous section to a probabilistic model. Note that in the machine learning literature, this notion of average error is often referred to as the *risk* of a given reconstruction mapping [96, Chapter 3] [57, Chapter 1].

We start by introducing the notation. For a topological space (X, τ), the *Borel σ-algebra* of X, denoted by $\mathscr{B}(X)$, is the σ-algebra generated by the family τ of open sets. Elements of $\mathscr{B}(X)$ are called *Borel sets*. If X is a metric space, then $\mathscr{B}(X)$ is the smallest family of sets containing all the open sets that are closed under countable intersections and countable disjoint unions. See, e.g., [6, Cor. 4.16] for further reference. A *Borel measure* on X is a measure defined on the Borel σ-algebra $\mathscr{B}(X)$. Borel measurable functions are defined analogously.

Assumption 3. *We consider the measurable spaces $(\mathscr{X} \times \mathscr{Z}, \mathscr{B}(\mathscr{X} \times \mathscr{Z}))$, $(\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}))$, $(\mathscr{M}_2^{\mathscr{E}}, \mathscr{B}(\mathscr{M}_2^{\mathscr{E}}))$ and assume that $F\colon \mathscr{M}_1 \times \mathscr{E} \to \mathscr{M}_2^{\mathscr{E}}$ is Borel measurable. We equip $\mathscr{X} \times \mathscr{Z}$ with a finite Borel measure μ supported on $\mathscr{M}_1 \times \mathscr{E}$, such that $\mu(\mathscr{M}_1 \times \mathscr{E}) > 0$. We equip $\mathscr{M}_2^{\mathscr{E}}$ with the pushforward measure $F_*\mu$ given by $(F_*\mu)(E) = \mu(F^{-1}(E))$ for every $E \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$.*

Next, we need to define conditional probabilities on $\mathscr{M}_1 \times \mathscr{E}$ for the considered measure μ for different values of $y \in \mathscr{M}_2^{\mathscr{E}}$. To compute such conditional probabilities, we consider a *disintegration* of the measure μ. Intuitively, a disintegration of the measure μ given the mapping F, is a family of probability measures $\{\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ such that for each $y \in \mathscr{M}_2^{\mathscr{E}}$, μ^y is concentrated on the set $F^{-1}(y)$, and which allows us to reconstruct the original measure μ by integration. Note that these sets $\{F^{-1}(y)\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ might have measure zero with respect to μ. The concept of disintegration is introduced in the following definition, which is taken from [40] and adapted to our Assumption 3. See also [27, Chapter 10] or [72, Chapter 1] for more on disintegration. Moreover, note that in [90] a similar tool – that of a *regular conditional probability density* – is used in order to prove analogous bounds when restricted to normed spaces and single-valued decoders.

Definition 3. *A disintegration of the measure μ along the measurable function $F : \mathscr{M}_1 \times \mathscr{E} \to \mathscr{M}_2^{\mathscr{E}}$ is a family $\{\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ of probability measures on the measurable space $(\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}))$ such that*

(1) for $F_*\mu$-almost every $y \in \mathscr{M}_2^{\mathscr{E}}$, μ^y is a probability measure concentrated on $F^{-1}(y)$, i.e., $\mu^y(\mathscr{M}_1 \times \mathscr{E} \setminus F^{-1}(y)) = 0$ for $F_*\mu$-almost all $y \in \mathscr{M}_2^{\mathscr{E}}$.

and such that, for each non-negative Borel measurable function f on $(\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}))$,

(2) the function

$$y \mapsto \int_{\mathscr{M}_1 \times \mathscr{E}} f(x, e)\, d\mu^y(x, e), \quad \text{with } y \in \mathscr{M}_2^{\mathscr{E}},$$

is Borel measurable,

(3) and

$$\int_{\mathscr{M}_1 \times \mathscr{E}} f(x, e)\, d\mu(x, e) = \int_{\mathscr{M}_2^{\mathscr{E}}} \left(\int_{F^{-1}(y)} f(x, e)\, d\mu^y(x, e) \right) d(F_*\mu)(y).$$

Remark. Example of a disintegration along a projection. *Let μ be Lebesgue measure on $\mathbb{R}^2$ and let $\pi_1(x, y) = x$ be the projection onto the first coordinate. Then, μ admits a disintegration with respect to π_1 given by the family of measures $\{\mu^x\}_{x \in \mathbb{R}}$, where μ^x is Lebesgue measure on the fibre $\{x\} \times \mathbb{R}$. In this*

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

case, the disintegration formula

$$\int_{\mathbb{R}^2} f(x,y)\,d\mu(x,y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,d\mu^x(y) \right) d\pi_{1*}\mu(x,y)$$

for a non-negative, measurable and integrable function f is the Fubini-Tonelli's theorem (see [55, Theorem 2.37]).

Note that while we assume that μ is a finite measure (and not necessarily a probability measure), the definition above requires that μ^y's are probability measures. In Section 7.1, we show that a disintegration of μ always exists and that it is essentially unique in our setting given Assumption 4, stated below.

Assumption 4. *We assume that $\mathscr{M}_1$ is compact, and we assume that $\mathscr{E}$ is a Borel set.*

Note that Assumption 4 implies Assumption 2, which will therefore be omitted from now on. Additionally, since $\mathscr{M}_1 \times \mathscr{E}$ is a Borel subset of $\mathscr{X} \times \mathscr{Z}$, the Borel σ-algebra $\mathscr{B}(\mathscr{M}_1 \times \mathscr{E})$, that is induced by the subspace topology on $\mathscr{M}_1 \times \mathscr{E}$, coincides with the trace σ-algebra $\mathscr{B}(\mathscr{X} \times \mathscr{Z})_{|\mathscr{M}_1 \times \mathscr{E}} = \{B \cap (\mathscr{M}_1 \times \mathscr{E}) : B \in \mathscr{B}(\mathscr{X} \times \mathscr{Z})\}$, see e.g., [27, Section 1.1].

Remark. *As an example for an inverse problem as in (5) satisfying the Assumptions 2 and 4, we consider a linear forward model with additive noise $F(x,e) = Ax + e$ for $x \in \mathscr{M}_1 \subset \mathscr{X} = \mathbb{C}^N$ and $e \in \mathscr{E} \subset \mathscr{Z} = \mathbb{C}^m$ with $m, N \in \mathbb{N}$ and $A \in \mathbb{C}^{m \times N}$. For many relevant applications the linear map $A \in \mathbb{C}^{m \times N}$ has a non-trivial null space $\mathscr{N}(A) := \{x \in \mathbb{C}^N : Ax = 0\} \neq \{0\}$ or is ill-conditioned. The linear map typically has a non-trivial null space for undersampled inverse problems. In the case that A is ill-conditioned, numerically the condition $\mathscr{N}(A) \neq \{0\}$ can in principle be approximated by setting the singular values around or below machine precision to zero, which results in $\mathscr{N}(A) \neq \{0\}$. If the null space is non-trivial, the feasible sets F_y can be non-empty. As a subset of the bounded set $\mathscr{M}_1$, the feasible sets are bounded, and by the Heine-Borel property of $d_{\mathscr{X}}$, in Assumption 1, are also compact. In this case, the feasible sets satisfy Assumption 2. Thus, the kernel size with worst-case noise can be non-zero. Assumption 4 is likely to be satisfied in applications. The condition that data sets or model classes, $\mathscr{M}_1 \subseteq \mathbb{C}^N$, are bounded, ensures that the set of admissible signals remains within physically or numerically reasonable limits. Additionally, if $\mathscr{M}_1$ is considered to be data, it can be assumed to be closed, since they include boundary points as part of the domain. Thus, it is natural to treat $\mathscr{M}_1$ as closed, since excluding boundary points would not align with how physical data is collected or interpreted. Finally, many applications involve quantities that vary continuously (or piecewise continuously) with respect to their underlying parameters. This continuity can imply that boundary points do not represent any abrupt change in the physical state and in such cases can be seamlessly included in $\mathscr{M}_1$. This supports modelling $\mathscr{M}_1$ as a closed set, as no artificial exclusion of boundary points is required. Hence, assuming that $\mathscr{M}_1$ is closed and bounded is consistent with typical modelling assumptions. By the Heine-Borel property of $d_{\mathscr{X}}$, from Assumption 1, the set $\mathscr{M}_1$ is compact, as it is closed and bounded, satisfying Assumption 4. In probabilistic settings, the set of noise vectors $\mathscr{E}$ is assumed to be Borel measurable, satisfying Assumption 4.*

Next, let

$$r_\varphi : \mathscr{M}_1 \times \mathscr{E} \to [0, +\infty], \quad r_\varphi(x,e) = d^H_{\mathscr{X}}(x, \varphi(F(x,e)))$$

denote the *residual map* of a given reconstruction mapping $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ and let

$$\mathscr{C} := \{\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X} \colon r_\varphi \text{ is Borel measurable}\}. \tag{11}$$

denote the set of all reconstruction mappings with a Borel measurable residual map. For $p \in [1, \infty]$ and a given reconstruction map $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$, we denote its pth order error by

$$\mathrm{Err}^{\mathrm{a}}(\varphi, p) = \begin{cases} \left(\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}\big(x, \varphi(F(x,e))\big)^p \, d\mu(x,e) \right)^{\frac{1}{p}} & \text{for } 1 \leq p < \infty \\ \operatorname{esssup}_{(x,e) \in \mathscr{M}_1 \times \mathscr{E}} \; d^H_{\mathscr{X}}(x, \varphi(F(x,e))) & \text{for } p = \infty \end{cases} \tag{12}$$

where the essential supremum for $p = \infty$ is taken with respect to the measure μ. Concretely, for $p \in [1, \infty]$, this quantity is the $L^p(\mathscr{M}_1 \times \mathscr{E}, \mu)$ norm of the residual function r_φ.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

This definition of reconstruction error generalises the usual L^p-norms to include a set-valued mapping and a metric $d_{\mathscr{X}}$ that is not necessarily induced by a norm in the single-valued case. However, the proposed setup includes many standard expectations as special cases. Indeed, let φ be a single-valued map and let $d_{\mathscr{X}}$ be given by an ℓ^q norm. Then, if we choose $p = q = 1$, we recover the expected absolute error, whereas for $p = q = 2$, we find the expected root squared error. Moreover, for $p = \infty$, the choice of metric $d_{\mathscr{X}}$ corresponds to the choice of the loss function in a machine learning setting [96, Chapter 3] [57, Chapter 1]. One can also view the case where $p = \infty$ as a weak form of worst-case error where all irregularities on sets of measure zero are ignored.

With the error term in (12) defined, we can now establish the notion of optimality constant with average error and kernel constant with average error.

Definition 4 (Optimality constant with average error). *For $p \in [1, \infty]$, the optimality constant with average error of order p for the inverse problem (5) is*

$$c^a_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, p) = \inf_{\varphi \in \mathscr{C}} \mathrm{Err}^a(\varphi, p). \tag{13}$$

A map $\varphi \in \mathscr{C}$ attaining the infimum in (13) for a given p is called an optimal map with average error of order p.

Definition 5 (Average kernel size). *The average kernel size of the inverse problem (1) for $p \in [1, \infty)$ is given by*

$$\mathrm{kersize}^a(F, \mathscr{M}_1, \mathscr{E}, p) = \left(\int_{\mathscr{M}_2^{\mathscr{E}}} \int_{F^{-1}(y)} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, x')^p \, d\mu^y(x, e) \, d\mu^y(x', e') \, d(F_* \mu)(y) \right)^{\frac{1}{p}},$$

and for $p = \infty$ it is

$$\mathrm{kersize}^a(F, \mathscr{M}_1, \mathscr{E}, \infty) = \operatorname*{esssup}_{y \in \mathscr{M}_2^{\mathscr{E}}} \operatorname*{esssup}_{\substack{(x,e) \in F^{-1}(y) \\ (x',e') \in F^{-1}(y)}} d_{\mathscr{X}}(x, x'), \tag{14}$$

where the left and right essential suprema are taken with respect to $F_* \mu$ on $\mathscr{M}_2^{\mathscr{E}}$ and μ^y on $F^{-1}(y)$, respectively.

Intuitively, the average kernel size measures the average distance between the x-components of $(x, e), (x', e') \in \mathscr{M}_1 \times \mathscr{E}$ that have the same measurement $F(x, e) = F(x', e') = y$ in the case of $p \in [1, \infty)$, and the maximum of such distances (up to negligible sets) when $p = \infty$.

Our main theorem in this section is stated below. It bounds the average optimality constant in terms of the average kernel size. It also ensures the existence of an optimal map, provides a variational expression for one such map and establishes some regularity properties of this mapping.

Theorem 3.2 (Average optimality bounds). *Under Assumptions 1, 3, 4, the following holds for every $p \in [1, \infty]$,*

(1) We have that

$$\mathrm{kersize}^a(F, \mathscr{M}_1, \mathscr{E}, p)/2 \leq c^a_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, p) \leq \mathrm{kersize}^a(F, \mathscr{M}_1, \mathscr{E}, p).$$

(2) Consider $p \in [1, \infty)$, and let $\Psi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ be given by

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, z)^p \, d\mu^y(x, e). \tag{15}$$

Then we have the following: Ψ has compact values, it is measurable, and it admits a measurable selector. Moreover, Ψ is an optimal map with average error of order p.

(3) Consider $p = \infty$, and let $\Psi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ be given by

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} \operatorname*{esssup}_{(x,e) \in F^{-1}(y)} d_{\mathscr{X}}(x, z). \tag{16}$$

Then we have the following: Ψ has compact values, it is measurable, and it admits a measurable selector. Moreover, Ψ is an optimal map with average error of order p.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Remark. Measurability of set-valued mappings. *There exist several notions of measurability for set-valued mappings. The precise definition of the term used in the theorem above can be found in Definition 7. A measurable selector from a set-valued mapping* $\Phi : \mathscr{Y} \rightrightarrows \mathscr{X}$ *between measurable spaces is a measurable function* $\phi : \mathscr{Y} \to \mathscr{X}$ *such that* $\phi(y) \in \Phi(y)$ *for all* $y \in \mathscr{Y}$*, [6, §. 18.3].*

A key finding in Theorem 3.2 is that the possibly set-valued Ψ admits a measurable selector, i.e., a *single-valued* function $\psi : \mathscr{M}_2^{\mathscr{E}} \to \mathscr{X}$ which is Borel measurable and such that $\psi(y) \in \Psi(y)$ for every y. Thus, Theorem 3.2 not only provides fundamental limits for the reconstruction error of an inverse problem but also gives a (variational) expression for a mapping that is optimal. Moreover, such a reconstruction mapping satisfies some regularity properties, which one usually hopes for when solving an inverse problem. One can argue that measurability is still a weak condition, but in order to obtain stronger conditions (such as continuity of the reconstruction), one would need to impose stronger assumptions on the problem. Moreover, note that in the definition of the average optimality constant, all potentially set-valued maps are included. Hence, the bounds also hold for any measurable single-valued selector.

4. Examples

In this section, we present three examples from research areas in inverse problems and briefly illustrate their connection to the framework developed in this paper. Additionally, we provide examples where the upper and lower bounds in Theorem 3.1 (*i*) and the upper bound in Theorem 3.2 (*i*) are attained and, hence, show that the bounds are sharp.

Example 1. *Linear inverse problems with the robust null-space property*

The robust null space property from compressed sensing was first introduced as a necessary condition for uniform recovery of sparse vectors from linear measurements. See [56] for a historical overview. Later it has been extended to a wide range of model classes $\mathscr{M}_1$, including low-rank [70] and sparsity in levels [18] models. A very general version of the property appears in [31], which considers general sets $\mathscr{M}_1$. We now recall a specialised version of the property from [31], and then we relate this to kernel size with worst-case noise.

Let $\mathscr{X}, \mathscr{Y}$ and $\mathscr{Z}$ be vector spaces, with $\mathscr{Y} = \mathscr{Z}$, and let $A \colon \mathscr{X} \to \mathscr{Y}$ be a linear mapping that is onto $\mathscr{Y}$. We consider the additive linear model $F(x, e) = Ax + e$ for $(x, e) \in \mathscr{M}_1 \times \mathscr{E} \subset \mathscr{X} \times \mathscr{Z}$. Let $|||\cdot|||_{\mathscr{X}_1}$ and $|||\cdot|||_{\mathscr{X}_2}$ be norms on $\mathscr{X}$, and let $|||\cdot|||_{\mathscr{Y}}$ be a norm on $\mathscr{Y}$ (slightly weaker conditions are used in [31]). Furthermore, for the model class $\mathscr{M}_1 \subset \mathscr{X}$, we denote the set $\mathscr{M}_1 - \mathscr{M}_1 := \{x - x' : x, x' \in \mathscr{M}_1\}$, and let $\operatorname{dist}_{\mathscr{X}_2}(h, \mathscr{M}_1 - \mathscr{M}_1) = \inf_{z \in \mathscr{M}_1 - \mathscr{M}_1} |||z - h|||_{\mathscr{X}_2}$. Then, the linear mapping A is said to satisfy the *robust null-space property* (rNSP) with constants $D_1, D_2 > 0$ with respect to the set $\mathscr{M}_1$, if

$$|||h|||_{\mathscr{X}_1} \leq D_1 \operatorname{dist}_{\mathscr{X}_2}(h, \mathscr{M}_1 - \mathscr{M}_1) + D_2 |||A(h)|||_{\mathscr{Y}} \quad \text{for all } h \in \mathscr{X}. \tag{17}$$

In Proposition 4.1, we assume that Assumption 4 holds, i.e., that $\mathscr{M}_1$ is compact. For linear inverse problems, this assumption is often used on the noise level, whereas the set $\mathscr{M}_1$ is often an unbounded set (sparse vectors, matrices with low rank, etc.). However, when modelling any practical application, it is not unreasonable to assume that the set of interest is compact or closed and bounded. Now under the assumption that the noise is bounded by $\operatorname{diam}_{d_{\mathscr{Y}}}(\mathscr{E}) = \sup_{e, e' \in \mathscr{E}} |||e - e'|||_{\mathscr{Y}} = \eta$ in Proposition 4.1, we show that the kernel size with worst-case noise, with respect to the metrics induced by $|||\cdot|||_{\mathscr{X}_1}$ and $|||\cdot|||_{\mathscr{Y}}$, is bounded by

$$\operatorname{kersize}^{\mathrm{w}}(F, \mathscr{M}_1, \mathscr{E}) \leq D_2 \eta. \tag{18}$$

In this sense, Proposition 4.1 provides a relation between the rNSP and the kernel size with worst-case noise.

Proposition 4.1. *Let* $\eta \geq 0$*. Let* $\mathscr{X}, \mathscr{Y}$ *be normed vector spaces with norms* $|||\cdot|||_{\mathscr{X}_1}$ *and* $|||\cdot|||_{\mathscr{X}_2}$ *on* $\mathscr{X}$*, and* $|||\cdot|||_{\mathscr{Y}}$ *on* $\mathscr{Y}$*. Let* $\mathscr{M}_1 \subseteq \mathscr{X}$ *satisfy Assumption 4, and let* $\mathscr{E} \subset \mathscr{Y}$ *with* $\operatorname{diam}_{d_{\mathscr{Y}}}(\mathscr{E}) = \eta$*. Let* $F(x, e) = Ax + e$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

for $(x, e) \in \mathcal{M}_1 \times \mathcal{E}$, where $A\colon \mathcal{X} \to \mathcal{Y}$ is a linear map onto $\mathcal{Y}$ and that satisfies the rNSP with constants $D_1, D_2 \geq 0$ with respect to $\mathcal{M}_1$. Then, the following holds:

(1) A is injective on $\mathcal{M}_1$, and this is equivalent to

$$(\mathcal{M}_1 - \mathcal{M}_1) \cap \mathcal{N}(A) = \{0\},$$

where $\mathcal{N}(A) = \{x \in \mathcal{X} : Ax = 0\}$.

(2) The kernel size with worst-case noise can be bounded from above as

$$\operatorname{kersize}^{w}(F, \mathcal{M}_1, \mathcal{E}) \leq D_2 \eta.$$

Note that if we take $\eta = 0$, i.e., a noiseless forward model $\mathcal{E} = \{0\}$, we have that $\operatorname{kersize}^{w}(F, \mathcal{M}_1, \mathcal{E}) = 0$.

Example 2. *Optimal learning, Chebyshev centres and Gelfand widths*

Recently, the notion of optimal learning was introduced in [25]. Herein, one considers the setting in which the set $\mathcal{X}$ is a Banach space with norm $\|\cdot\|_{\mathcal{X}}$ and $\mathcal{M}_1$ is a compact subset of $\mathcal{X}$. To sample an element $x \in \mathcal{M}_1$, one uses m linear functionals $\lambda_1, \ldots, \lambda_m \in \mathcal{X}^*$ from the dual space of $\mathcal{X}$. Now, for a given $x \in \mathcal{M}_1$, let $y = (\lambda_1(x), \ldots, \lambda_m(x)) \in \mathbb{C}^m$ and let

$$K_y = \{x \in \mathcal{M}_1 : \lambda_i(x) = y_i, i \in \{1, \ldots, m\}\}$$

denote the set of solutions which are data-consistent. The above measurement model is noiseless, so define for convenience $\mathcal{M}_2 = \{y = (\lambda_1(x), \ldots, \lambda_m(x)) : x \in \mathcal{M}_1\}$ as the set of noiseless measurements.

Now, let $B_{\mathcal{X}}(z, r)$ denote the closed ball in $\mathcal{X}$ with centre z and radius r. For a compact set $\mathcal{S} \subset \mathcal{X}$, the Chebyshev radius of $\mathcal{S}$ is given by

$$R_{\mathcal{X}}(\mathcal{S}) := \inf\{r > 0 : \mathcal{S} \subset B_{\mathcal{X}}(z, r) \text{ for some } z \in \mathcal{X}\}.$$

In [25], the quantity $R_{\mathcal{X}}(K_y)$ is defined as the *optimal recovery rate* for a given $x \in \mathcal{M}_1$, with $y = (\lambda_1(x), \ldots, \lambda_m(x))$. This quantity is zero if K_y is a singleton; however, here, we follow [25] and focus on the cases for which $R_{\mathcal{X}}(K_y) > 0$.

While the framework developed in this paper does not extend to general Banach spaces, it does cover the special case of $\mathcal{X} = \mathbb{C}^N$ with the metric induced by a norm $\|\cdot\|_{\mathcal{X}}$. Indeed, by letting $A \in \mathbb{C}^{m \times N}$ be the matrix given by the linear functionals λ_i, taking $\mathcal{E} = \{0\}$ and using the linear additive model $F(x, e) = Ax + e$, as $K_y = F_y$, it is straightforward to see that $R_{\mathcal{X}}(K_y) = \frac{1}{2}\operatorname{diam}_{\mathcal{X}}(F_y)$, since the minimal enclosing ball of a bounded set has radius equal to half its diameter when centred at a Chebyshev centre. By taking the supremum over all $y \in \mathcal{M}_2$, we obtain of $\sup_{y \in \mathcal{M}_2} 2R_{\mathcal{X}}(K_y) = \sup_{y \in \mathcal{M}_2} \operatorname{diam}_{\mathcal{X}}(F_y)$, which coincides with the kernel size with worst-case noise, as characterised in (8). To further analyse the noisy setting, we refer to [25].

The framework proposed in this work is closely related to the concept of Chebyshev centres. In fact, the optimal maps Ψ defined in (10) and (15) correspond to the problem of finding the Chebyshev centre or the p-centre (see [87]) of the set F_y of candidate solutions. The setting proposed in this paper guarantees that such centres exist so that the optimal maps are well-defined; however, Chebyshev centres may not exist in general, and other sufficient and necessary conditions for their existence can be found in the literature (see [8], [7], [109], [93]).

In this work, we have kept the forward model F fixed. However, in the special case of a linear forward model, it is reasonable to also consider the question of how well we can perform given a fixed budget of m linear functionals. This question is well studied in the literature [89] and is given by the Gelfand width of $\mathcal{M}_1$,

$$w^m(\mathcal{M}_1) := \inf_{A \in \mathbb{C}^{m \times N}} \sup_{x \in \mathcal{M}_1} R_{\mathcal{X}}(K_{Ax}).$$

We note that this presents a lower bound for $c^{\mathrm{w}}_{\mathrm{opt}}(A, \mathcal{M}_1, \{0\})$ in the noiseless setting. We do not consider the question of what the optimal forward model would be any further in this work but remark that it is an interesting question worth investigating.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Example 3. *Bayesian inverse problems*

Bayesian inverse problems [12, 101] take a probabilistic approach. Instead of using a measurement y to output a *single* candidate solution x, the goal of a Bayesian approach is to use a measurement y to output a *distribution* on the possible solutions x's.

The following proposition of Theorem 3.2 demonstrates how Bayesian inverse problems can be related to the framework presented in this work. The proposition shows that the posterior mean estimator is an optimal map in the average case for $p = 2$ and for metrics $d_{\mathscr{X}}$ that are induced by an inner product.

Proposition 4.2 (Bayesian problems in the current framework). *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $(\mathscr{X}, d_{\mathscr{X}})$, $(\mathscr{Y}, d_{\mathscr{Y}})$, $(\mathscr{Z}, d_{\mathscr{Z}})$ and $\mathscr{M}_1, \mathscr{E}$, $F : (\mathscr{X} \times \mathscr{E}, \mathscr{B}(\mathscr{X} \times \mathscr{E})) \to (\mathscr{M}_2^{\mathscr{E}}, \mathscr{B}(\mathscr{M}_2^{\mathscr{E}}))$ and*

$$(\mathbf{x}, \mathbf{e}) : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathscr{X} \times \mathscr{E}, \mathscr{B}(\mathscr{X} \times \mathscr{E}))$$

be a random variable with (joint) distribution $\mu := \mu_{\mathbf{x},\mathbf{e}} = \mathbb{P}[(\mathbf{x}, \mathbf{e}) \in \cdot]$ satisfying Assumptions 1, 3 and 4. Let $(\mathscr{Y}, d_{\mathscr{Y}})$ be a Polish space. Define the random variable $\mathbf{y} := F(\mathbf{x}, \mathbf{e}) = F \circ (\mathbf{x}, \mathbf{e})$ whose law is given by $\mu_{\mathbf{y}} = F_* \mu = \mathbb{P}[\mathbf{y} \in \cdot]$. Denoting by $\pi_1 : \mathscr{X} \times \mathscr{E} \to \mathscr{X}$ the coordinate projection, the following holds:

(1) The Bayesian prior distribution of $\mathbf{x}$ is the distribution of $\mathbf{x}$ and can be obtained as $\mu_{\mathbf{x}} = \pi_{1*}\mu$;

(2) There exist likelihood distributions $\{\mu_{\mathbf{y}|\mathbf{x}=x}\}_{x \in \mathscr{X}}$ of $\mathbf{y}$ given $\mathbf{x}$;

(3) The Bayesian posterior distributions $\{\mu_{\mathbf{x}|\mathbf{y}=y}\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ can be obtained by the pushforward with respect to π_1 of the disintegration $\{\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ of μ along F as: $\mu_{\mathbf{x}|\mathbf{y}=y} = \pi_{1*}\mu^y$;

(4) If the metric $d_{\mathscr{X}}$ is replaced with a metric induced by an inner product $\langle \cdot, \cdot \rangle$ and $p = 2$, then the optimal map Ψ from Theorem 3.2 part (ii) coincides with the posterior mean and $\Psi(y) = \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]$ for every $y \in \mathscr{M}_2^{\mathscr{E}}$.

As Proposition 4.2 is intended only to be a brief example, the relevant definitions and its proof are presented in Section 7. This result illustrates how the framework introduced in this chapter is capable of capturing not only inverse problems in their classical sense but also in the Bayesian approach.

Example 4. *The kernel sizes are sharp bounds to the optimality constants*

In the following, we provide two examples that show that the upper and lower worst-case bounds in Theorem 3.1 (*i*) can be attained and, hence, are sharp. Then, we provide another example showing that the upper bound is attained in Theorem 3.2 (*i*) and, hence, is also sharp. We first provide two simple examples that obtain the upper and lower worst-case bounds in Theorem 3.1 (*i*). In the following, for fixed $n \in \mathbb{N}$, $\{e_1, \ldots, e_n\}$ denotes the canonical basis of $\mathbb{R}^n$.

(1) *Sharpness of the lower bound in Theorem 3.1 (i)*: Let $(\mathscr{X}, d_{\mathscr{X}}) = (\mathbb{R}^2, \|\cdot\|_2)$, $(\mathscr{Y}, d_{\mathscr{Y}}) = (\mathscr{E}, d_{\mathscr{E}}) = (\mathbb{R}, |\cdot|)$, $\mathscr{M}_1 = \{-e_1, e_1\}$, $\mathscr{E} = \{0\}$ and $F : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ with $F((x_1, x_2), e) = x_2$, so that $\mathscr{M}_2^{\mathscr{E}} = \{0\}$. Then, as $F^{-1}(0) = \mathscr{M}_1$, we have that

$$\operatorname{kersize}^{\mathrm{w}}(F, \mathscr{M}_1, \mathscr{E}) = \|e_1 - (-e_1)\|_2 = 2, \quad c^{\mathrm{w}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}) = 1,$$

as the optimal map with worst-case noise is given by $\Psi(0) = \operatorname{argmin}_{z \in \mathbb{R}^2} \sup_{x \in \mathscr{M}_1} \|z - x\|_2 = 0$.

(2) *Sharpness of the upper bound in Theorem 3.1 (i)*: Consider the same setup as above, but replace $\mathscr{X}$ with $(\mathscr{X}, d_{\mathscr{X}}) = (\mathscr{M}_1, \|\cdot\|_2)$ and as $\mathscr{M}_1 \subset \mathbb{R}^2$ restrict F by $F{:}\mathscr{M}_1 \times \mathbb{R} \to \mathbb{R}$ with $F((x_1, x_2), e) = x_2$. Then, we have that

$$\operatorname{kersize}^{\mathrm{w}}(F, \mathscr{M}_1, \mathscr{E}) = 2 = c^{\mathrm{w}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}),$$

as the optimal maps with $\Psi_i(0) = \operatorname{argmin}_{z \in \mathscr{M}_1} \sup_{x \in \mathscr{M}_1} \|z - x\|_2$ are given by $\Psi_i : \mathscr{M}_2^{\mathscr{E}} \to \mathscr{M}_1$ for $i = 1, 2, 3$, where $\Psi_1(0) = e_1$, $\Psi_2(0) = -e_1$ and $\Psi_3(0) = \{-e_1, e_1\}$.

For the average case by providing examples attaining the bounds, we can also show that the upper bound in Theorem 3.2 (*i*) is sharp. Consider the same setup as in part (1) above. Let $\alpha \in [0, 1]$ be an arbitrary

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

parameter and equip $\mathbb{R}^2 \times \mathscr{E}$ with the measure $\mu = \mu_\alpha$ given by $\mu_\alpha = \big(\alpha\delta_{-e_1} + (1-\alpha)\delta_{e_1}\big) \otimes \delta_0$, which is supported on $\mathscr{M}_1 \times \mathscr{E} = \{-e_1, e_1\} \times \{0\}$. For $p = 1$ and as $F^{-1}(0) = \mathscr{M}_1$, we have that the average kernel size is

$$\operatorname{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, 1) = \alpha^2 \cdot 0 + 2\alpha(1-\alpha)2 + (1-\alpha)^2 \cdot 0 = 4\alpha(1-\alpha),$$

and the optimality constant and the optimal map with average error $\Psi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathbb{R}^2$, obtained by Theorem 3.2 (*ii*) as $\Psi(0) = \operatorname{argmin}_{z\in\mathbb{R}^2} (\alpha\|z + e_1\|_2 + (1-\alpha)\|z - e_1\|_2)$, are

$$c^{\mathrm{a}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, 1) = \begin{cases} 2\alpha, & \text{if } \alpha \in [0, \frac{1}{2}), \\ 1, & \text{if } \alpha = \frac{1}{2}, \\ 2(1-\alpha), & \text{if } \alpha \in (\frac{1}{2}, 1], \end{cases} \quad \text{and, } \Psi(0) = \begin{cases} e_1, & \text{if } \alpha \in [0, \frac{1}{2}), \\ \{(t,0) : t \in [-1, 1]\}, & \text{if } \alpha = \frac{1}{2}, \\ -e_1, & \text{if } \alpha \in (\frac{1}{2}, 1]. \end{cases}$$

For $\alpha = \frac{1}{2}$, any single-valued selector $\psi : \mathscr{M}_2^{\mathscr{E}} \to \mathbb{R}^2$ of Ψ is an optimal map. For $\alpha = \frac{1}{2}$ the upper bound in Theorem 3.2 (*i*) is attained, as $c^{\mathrm{a}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, 1) = \operatorname{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, 1) = 1$.

5. Relation to previous work

As an ongoing topic of research in many areas of mathematics, undersampled and ill-posed inverse problems have been studied in many different areas. As a non-extensive list of examples, they have been studied from a theoretical standpoint [49], from a statistical perspective [88], by using iterative deep neural networks [5] and in applied fields, such as radio tomography of the ionosphere [59]. For solving undetermined and ill-posed inverse problems, often regularisation methods are utilised – here M. Benning et al. give a comprehensive review of modern methods [22]. Different noise models have been studied, and while an assumption of additive noise appears ubiquitously [23, 67, 71], often multiplicative noise models can be relevant in applications [13, 68, 98, 111]. The authors of [94] note that an open challenge in this research field is optimisation guarantees with milder assumptions, which are provided in this work. To the best of the authors' knowledge, fundamental accuracy bounds for undersampled and ill-posed inverse problems with multiplicative noise have not been produced. Moreover, despite the existing extensive amount of research, there is little to be found on fundamental accuracy bounds of approximate set-valued solutions to undersampled and ill-posed inverse problems.

Undersampled inverse problems and AI hallucinations: In [62], we provide sufficient conditions for AI hallucinations to occur due to a non-trivial kernel of the forward operator. For example, instabilities, such as false positives, false negatives, and especially AI hallucinations, have been an issue in the fastMRI challenge [85] and in microscopy [19, 66]. In [33], M. Burger and T. Roith outline that fundamental performance and accuracy limits for data-driven approaches, such as deep learning, for solving ill-posed inverse problems do not exist. Moreover, they outline that the lack of such meaningful error estimates in the context of medical imaging and undetermined inverse problems is demonstrated using AI hallucinations in recent studies [10, 24, 62]. In this work, we provide the missing link between a non-trivial kernel and optimal recovery.

(Set-valued) Decoders arising in deep learning: In the work [46], the ability of deep neural networks to nonlinearly approximate functions, and thus also decoders, is investigated. A range of results on how the resulting decoder may be constituted is presented by M. Unser in [106]. From an application-based perspective, a detailed overview of deep learning in inverse problems and stability of robustness for deep learning is given in [12, 80, 82]. Compared to standard methods for solving inverse problems, partially data-driven approaches, such as from L. Borcea, J. Garnier et al. [28], and fully data-driven approaches using deep learning, such as [19, 100, 112], have reported superior accuracy. This can potentially lead to instabilities, which is highlighted by V. Antun et al. in [10] and in [62]. In fact, a variety of research has established that artificial intelligence techniques based on deep learning are unstable, firstly in image classification [51, 74, 84, 86, 102], and later in applications ranging from audio and speech recognition [38, 39, 110] to natural language processing [78] and automatic diagnosis in medicine [54]. Moreover,

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

many DL-based approaches implicitly include set-valued functions, where examples include, but are not limited to, deep ensembles [75] or model-based probabilistic conditional diffusion models as in [79]. In fact, any probabilistic DL model used to solve an inverse problem that uses sampling from a distribution or predicts parameters of a distribution – see the work of J. Gawlikowski [60] for an introduction on uncertainty quantification and probabilistic DL models – can be considered to be a set-valued decoder.

Optimal (set-valued) decoders for undersampled inverse problems: The study of optimal and near-optimal recovery in inverse problems has always been a central question, and many single-valued near-optimal mappings have been proposed and analysed [25, 31, 56, 90, 104]. A survey of early works on optimal recovery is given in [83]. For Compressed Sensing (CS) C. Poon provides sampling conditions and error bounds for near-optimal recovery in [91]. The case of set-valued optimal decoders has been considered for obtaining the worst-case bounds on Banach spaces in [11] and extended to metric spaces in [81]. As presented in Example 3, the Bayesian approach to inverse problems aims at recovering a distribution-valued decoder [12, 101]. Here, there exist a posteriori accuracy bounds in the case of normed spaces [76, 77, 107].

Accuracy bounds in approximation theory: Approximation theory aims to provide accuracy bounds and theoretical guarantees for function approximation tasks under specific assumptions. However, in [3], B. Adcock and N. Dexter state that despite the ability of deep neural networks to approximate functions relevant to scientific computing, a large gap between expressivity and practical performance remains. An in-depth treatment of sparse polynomial approximations of high-dimensional functions is given by B. Adcock, S. Brugiapaglia and C. Webster in [2]. Fundamental accuracy bounds, the Gelfand Widths, have been established in [89] for the noiseless linear setting. A framework for obtaining the best k-term approximation and corresponding accuracy bounds is established [41]. In [42], the authors propose a framework to measure the optimal performance for nonlinear methods of approximation. The notion of optimal learning in a noisy setting and upper worst-case accuracy bounds based on the Chebyshev radius are presented in [25]. B. Adcock et al. [4] propose a framework for accuracy bounds extending the classical assumption of CS. Generalised instance optimality and the corresponding accuracy bounds have been presented in [31]. Related to the CS framework, the Restricted Isometry Property (RIP) is generalised in [104], and the null space property is generalised in [103]. However, A. Bastounis et al. establish that the RIP is absent in real-world CS applications in [18] and, hence, propose an extension thereof, the RIP in levels, in order to explain the success of the method. Related to our work, analogous worst-case and average error bounds for normed spaces and single-valued decoders are established in [90].

6. Discussion and outlook

A fundamental limitation of this work is that the optimal maps achieving the smallest worst-case and average error in (10), (15) and (16), respectively, are generally not analytically computable. Moreover, these optimal maps are defined via embedded optimisation problems. While, in principle, such problems could be approximated numerically, doing so may be computationally infeasible for inverse problems involving high-dimensional datasets. In the average-case setting, the optimal map admits a measurable selector; however, it remains open to determine conditions under which it also admits a continuous selector or a continuous approximation. This is closely related to identifying structural conditions under which the set-valued map $y \mapsto F_y$ is continuous (in an appropriate sense), which in turn depends on properties of the forward map F, the model class $\mathscr{M}_1$, the noise set $\mathscr{E}$, and the underlying metrics. Another limitation is that the worst-case and average kernel sizes are not readily computable. Developing computable and provably convergent approximations to these quantities would enable quantitative accuracy guarantees in applications. In Section 4, we presented illustrative examples highlighting connections between this framework and other approaches to inverse problems. In particular, Proposition 4.2 shows that the optimal map in the average-case setting coincides with the posterior mean estimator in Bayesian inverse

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

problems. A natural direction for future work is to investigate how other Bayesian estimators, such as the maximum a posteriori estimator, relate to the present framework. Finally, extending Proposition 4.1 to generalised rNSP conditions such as in [31] and identifying the corresponding optimal maps in compressed sensing remains an interesting direction.

The proposed theoretical framework for undetermined and ill-posed inverse problems can be seen as a generalisation of a variety of previous frameworks to set-valued decoders. Additionally, commonly used assumptions, such as the convexity of the set $\mathscr{M}_1$ and a linear forward operator A and additive noise, as well as the condition $(\mathscr{M}_1 - \mathscr{M}_1) \cap \mathcal{N}(A) = \{0\}$, which is for example implied by the RIP, are extended or are shown to be unnecessary. Under general assumptions, our work provides relevant accuracy bounds for set- and single-valued decoders and a variational expression for optimal decoders. Due to the generality of the assumptions, these bounds provide a means to bridge the gap between theory and practice. The lower bounds can be used for assessing accuracy and performance of a wide range of models, including ensemble models, diffusion models and sampling-based approaches that have set-valued solutions. Moreover, based on our previous work [62], these bounds can be used in future research to mitigate AI hallucinations in DL-based reconstructions for ill-posed inverse problems.

7. Proofs

This section contains the proofs of our results and propositions required to prove these. Additionally, where necessary, we include and cite results from prior work that are used in our proofs.

7.1. Existence and uniqueness of a disintegration of a measure given Assumptions 1, 3 and 4

In this section, we prove the following proposition, which ensures that our setting guarantees the existence of a disintegration of the measure μ.

Proposition 7.1. *Under Assumptions 1, 3 and 4, there exists a disintegration $\{\mu^y\}_{y\in\mathscr{M}_2^{\mathscr{E}}}$ of the measure μ along F. Moreover, such disintegration is essentially unique: if $\{\tilde{\mu}^y\}_{y\in\mathscr{M}_2^{\mathscr{E}}}$ is another family satisfying $(i)-(iii)$ in Definition 3, then $\tilde{\mu}^y = \mu^y$ for almost every y.*

The proof of Proposition 7.1 is a special case of what is found in [40]. Before jumping to the proof, in this section, we start by recalling some well-known definitions for measures and stating the relevant theorems from [40]. Let $(X, \mathscr{B}(X))$ be the Borel measurable space associated with the topological space X. A measure μ on $(X, \mathscr{B}(X))$ is said to be a *Radon measure* (sometimes also called a regular measure [6]) if

(1) $\mu(K) < +\infty$ for each compact $K \in \mathscr{B}(X)$,

(2) $\mu(B) = \inf\{\mu(V) : V \in \mathscr{B}(X), V \text{ open, and } B \subset V\}$, for every $B \in \mathscr{B}(X)$ (μ is outer regular), and

(3) $\mu(B) = \sup\{\mu(K) : K \in \mathscr{B}(X), K \text{ compact, and } K \subset B\}$, for every $B \in \mathscr{B}(X)$ (μ is tight).

Moreover, a measure μ is said to *dominate* a measure ν if $\mu(B) = 0$ implies $\nu(B) = 0$ for every $B \in \mathscr{B}(X)$. Suppose that X can be covered by at most countably many Borel measurable sets $\{B_i\}_{i\in I}$, $I \subset \mathbb{N}$, and that $\mu(B_i) < \infty$ for each $i \in I$. Then, μ is said to be a σ-finite measure. In particular, every finite measure is σ-finite. We also have the following results for finite measures that are used in the proof of Proposition 7.1.

Theorem 7.2 ([6, Theorem 12.7]). *A finite measure on a Polish space (i.e., a complete separable metric space) is Radon.*

In Definition 3, one may replace the measure $F_*\mu$ with another measure ρ, in which case $\{\mu^y\}_y$ would be called a (F, ρ) disintegration of μ. We refer the reader to [40, Definition 1] for the detailed statement.

Theorem 7.3 ([40, Theorem 1]). *Let μ be a σ-finite Radon measure on a metric space X and let F be a measurable map from $(X, \mathscr{B}(X))$ to the measurable space (Y, Σ). Let ρ be a σ-finite measure on Σ that*

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

dominates the pushforward measure $F_*\mu$. If Σ is countably generated and contains all the singleton sets $\{y\}$, then μ has a (F, ρ)-disintegration. The μ^y measures are uniquely determined up to an almost sure equivalence: if $\{\mu^y_*\}$ is another (F, ρ)-disintegration, then $\rho(\{y \in \mathscr{Y}\colon \mu^y_* = \mu^y\}) = 0$.

Theorem 7.4 ([40, Theorem 2 (iii)]). *Let μ have a (F, ρ)-disintegration $\{\mu^y\}$, with μ and ρ each σ-finite. Then, the measures $\{\mu^y\}$ are probabilities for ρ-almost all $y \in \mathscr{Y}$ if and only if $\rho = F_*\mu$.*

Proof of Proposition 7.1. Let $d := \max\{d_{\mathscr{X}}, d_{\mathscr{Z}}\}$ be one of the standard metrics on the product $\mathscr{X} \times \mathscr{Z}$. The spaces $(\mathscr{X}, d_{\mathscr{X}})$ and $(\mathscr{Z}, d_{\mathscr{Z}})$ are separable, as they are second countable by Assumption 11, and complete by Assumption 12. Thus the product space $(\mathscr{X} \times \mathscr{Z}, d)$ is complete and separable, [99] (pg. 26). Thus, $(\mathscr{X} \times \mathscr{Z}, d)$ is a Polish space. Since by Assumption 3, μ is a finite measure on the Polish space $\mathscr{X} \times \mathscr{Z}$, Theorem 7.2 guarantees that μ is Radon. Moreover, since by Assumption 3, μ is finite, it is in particular σ-finite. Taking $\rho = F_*\mu$, clearly ρ dominates $F_*\mu$ trivially, as they are the same measure. By Assumption 11, $(\mathscr{M}_2^{\mathscr{E}}, d_{\mathscr{Y}})$ is second countable; hence, its topology is countably generated, and so is the corresponding Borel σ-algebra $\mathscr{B} = \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$. Furthermore, since singletons $\{y\} \subseteq \mathscr{M}_2^{\mathscr{E}}$ are closed, the Borel σ-algebra contains all singletons. Lastly, as the mapping F is measurable by Assumption 3 and as by Assumption 4 $\mathscr{M}_1 \times \mathscr{E}$ is a Borel subset of $\mathscr{X} \times \mathscr{Z}$, we extend F trivially to be a measurable function on $\mathscr{X} \times \mathscr{Z}$, for instance by defining it to be constant outside $\mathscr{M}_1 \times \mathscr{E}$. Thus, all the conditions of Theorem 7.3 are satisfied. Hence, there exists a disintegration of the measure μ along F. Such a disintegration is essentially unique in the sense of Theorem 7.3. Finally, since μ and $\rho = F_*\mu$ are both finite and in particular σ-finite, Theorem 7.4 guarantees that $\rho = F_*\mu$ implies that the measures μ^y are probability measures for $F_*\mu$-almost every $y \in \mathscr{M}_2^{\mathscr{E}}$. □

7.2. Proof of Theorem 3.1

Proof of Theorem 3.1. We begin with the proof of (ii). First, let us introduce some notation: for fixed $y \in \mathscr{M}_2^{\mathscr{E}}$, define the function

$$f_y\colon \mathscr{X} \to [0, +\infty), \quad f_y(z) = \sup_{x \in F_y} d_{\mathscr{X}}(x, z) = d^H_{\mathscr{X}}(z, F_y)$$

and notice that f_y appears in the definition of Ψ in (10). Note that f_y does not attain the value $+\infty$, as by Assumption 2 the feasible set F_y is bounded. That is, we have the relation,

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} f_y(z). \tag{19}$$

Furthermore, for each $y \in \mathscr{M}_2^{\mathscr{E}}$ let

$$r_y = \sup_{\substack{x \in F_y \\ x' \in F_y}} d_{\mathscr{X}}(x, x') = \operatorname{diam}_{d_{\mathscr{X}}}(F_y).$$

Next, we make a claim which, once it is established, we use to prove that Ψ has non-empty, compact values. This ensures that Ψ is well-defined.

Claim. *For every $y \in \mathscr{M}_2^{\mathscr{E}}$, we have that,*

(1) f_y is continuous,

(2) for any $x \in F_y$, we have that

$$\operatorname*{argmin}_{z \in \mathscr{X}} f_y(z) = \operatorname*{argmin}_{z \in B(x, r_y)} f_y(z),$$

(3) $\Psi(y)$ is closed.

We proceed to prove the claim and start by considering (I). We consider a general setting, and let (X, d) be a metric space, $A \subseteq X$ be a bounded subset and $g(x) := d^H(x, A) = \sup_{a \in A} d(x, a)$. We claim that $|g(x) - g(y)| \leq d(x, y)$ for all $x, y \in X$. To see this, consider $a, x, y \in X$ and note that

$$d(x, a) \leq d(x, y) + d(y, a) \quad \Longrightarrow \quad \sup_{a \in A} d(x, a) \leq d(x, y) + \sup_{a \in A} d(y, a).$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Switching the roles of x and y, leads to the desired inequality. It follows that g is continuous. Moreover, F_y is bounded by Assumption 2. Thus, letting $(X, d) = (\mathscr{X}, d_{\mathscr{X}})$, $A = F_y$, $g = f_y$ above, proves (I).

To prove (II), we show that no point outside of the closed ball $B_{d_{\mathscr{X}}}(x, r_y)$ can be a minimiser of f_y. First, note that $r_y = \operatorname{diam}_{d_{\mathscr{X}}}(\pi_1(F_y)) < \infty$ as $\pi_1(F_y)$ is bounded. Moreover, by definition, we have that

$$r_y = \sup_{x \in F_y} \sup_{x' \in F_y} d_{\mathscr{X}}(x, x') = \sup_{x \in F_y} f_y(x),$$

which implies that $r_y \geq f_y(x)$ for all $x \in F_y$. Now, pick an $\hat{x} \in F_y$. If $z \in \mathscr{X} \setminus B(\hat{x}, r_y)$ is a point outside of the ball, then $d_{\mathscr{X}}(z, \hat{x}) > r_y$, and

$$f_y(z) = \sup_{x' \in F_y} d_{\mathscr{X}}(x', z) \geq d_{\mathscr{X}}(\hat{x}, z) > r_y \geq f_y(\hat{x}).$$

It follows that any minimiser of f_y must lie in the ball $B(\hat{x}, r_y)$. This proves our claim in (II).

Part (III) of the claim follows directly from the continuity of f_y. Indeed, let $\alpha_y := \min f_y \in [0, +\infty)$. Then, since $\{\alpha_y\} \subset [0, \infty)$ is closed and f_y is continuous, the preimage $f_y^{-1}(\{\alpha_y\}) = \Psi(y)$ is closed. This proves (III) and concludes our proof of the claim.

Next, we prove the following properties of Ψ. These properties finalise the proof of statement (ii) of the theorem.

(1) Ψ *has non-empty and compact values*;

(2) Ψ *is an optimal map*;

We start with the proof of (a). Let $y \in \mathscr{M}_2^{\mathscr{E}}$ and consider $x \in F_y$. From (II) and (19), we see that

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} f_y(z) = \operatorname*{argmin}_{z \in B(x, r_y)} f_y(z). \tag{20}$$

By Assumption 12, $d_{\mathscr{X}}$ satisfies the Heine-Borel property, which implies that the closed and bounded ball $\mathscr{B}(x, r_y)$ is compact. Moreover, from (I), we know that the objective function f_y is continuous. Thus, it follows from the Extreme Value Theorem that f_y attains its minimum in (20). This implies that Ψ has non-empty values. To see that Ψ has compact values, observe that $\Psi(y)$ is closed by (III) and that $\Psi(y) \subseteq B(x, r_y)$, where $B(x, r_y)$ is compact. Since a closed subset of a compact set is compact, we conclude that $\Psi(y)$ is compact.

Next, we prove (b). We will show that the minimum worst-case reconstruction error of Ψ equals the optimality constant as given in Definition 1. To this end, let $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ and fix $y \in \mathscr{M}_2^{\mathscr{E}}$. Recall from above that Ψ is non-empty. By construction, we have that

$$\sup_{x \in F_y} d_{\mathscr{X}}^H(\Psi(y), x) \leq \sup_{x \in F_y} d_{\mathscr{X}}(x, z), \quad \text{for all } z \in \varphi(y).$$

It follows that

$$\sup_{x \in F_y} d_{\mathscr{X}}^H(\Psi(y), x) \leq \sup_{x \in F_y} \sup_{z \in \varphi(y)} d_{\mathscr{X}}(x, z) = \sup_{x \in F_y} d_{\mathscr{X}}^H(x, \varphi(y)).$$

By taking the supremum with respect to $y \in \mathscr{M}_2^{\mathscr{E}}$ on both sides, we obtain

$$\sup_{y \in \mathscr{M}_2^{\mathscr{E}}} \sup_{x \in F_y} d_{\mathscr{X}}^H(\Psi(y), x) \leq \sup_{y \in \mathscr{M}_2^{\mathscr{E}}} \sup_{x \in F_y} d_{\mathscr{X}}^H(\varphi(y), x).$$

As $F \colon \mathscr{M}_1 \times \mathscr{E} \to \mathscr{M}_2^{\mathscr{E}}$ is surjective, the previous inequality can be rewritten as

$$\sup_{(x,e) \in \mathscr{M}_1 \times \mathscr{E}} d_{\mathscr{X}}^H(\Psi(F(x, e)), x) \leq \sup_{(x,e) \in \mathscr{M}_1 \times \mathscr{E}} d_{\mathscr{X}}^H(\varphi(F(x, e)), x).$$

Now, since $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ was arbitrary, the above inequality holds for any $\varphi \colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$. Taking the infimum over all mappings on the right-hand side, we obtain the optimality constant:

$$\sup_{(x,e) \in \mathscr{M}_1 \times \mathscr{E}} d_{\mathscr{X}}^H(\Psi(F(x, e)), x) \leq c_{\text{opt}}^{\text{w}}(F, \mathscr{M}_1, \mathscr{E}).$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

The opposite inequality is trivial, since $\Psi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ is one of the reconstruction mappings over which the infimum is taken. Therefore,

$$\sup_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} d_{\mathscr{X}}^H(\Psi(F(x,e)),x) = c_{\text{opt}}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}),$$

and we conclude that Ψ is an optimal map. This concludes the proof of statement (ii) in the theorem.

We proceed with the proof of (i) and start with the lower bound in (9). Let $\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ and $y \in \mathscr{M}_2^{\mathscr{E}}$. Then

$$\operatorname{diam}_{d_{\mathscr{X}}}(F_y) = \sup_{x,x'\in F_y} d_{\mathscr{X}}(x,x') \le \sup_{x\in F_y} 2d_{\mathscr{X}}^H(x,\varphi(y)).$$

Now, taking the supremum over all $y \in \mathscr{M}_2^{\mathscr{E}}$ gives the inequality

$$\operatorname{kersize}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}) \le \sup_{y\in\mathscr{M}_2^{\mathscr{E}}} \sup_{x\in F_y} 2d_{\mathscr{X}}^H(x,\varphi(y)). \tag{21}$$

Finally, since φ was arbitrary, taking the infimum over all $\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ in (21) gives

$$\operatorname{kersize}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}) \le 2c_{\text{opt}}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}).$$

which proves the lower bound in (9).

For proving the upper bound in (9), we will make use of the mapping Ψ in (10), which we know from statement (ii) is an optimal map. Let $y \in \mathscr{M}_2^{\mathscr{E}}$ and $x' \in F_y$. Then by construction of Ψ, we have that

$$\sup_{x\in F_y} d_{\mathscr{X}}^H(\Psi(y),x) \le \sup_{x\in F_y} d_{\mathscr{X}}(x,x') \le \operatorname{diam}(F_y).$$

Taking the supremum over all $y \in \mathscr{M}_2^{\mathscr{E}}$ on both sides above, gives

$$\sup_{y\in\mathscr{M}_2^{\mathscr{E}}} \sup_{x\in F_y} d_{\mathscr{X}}^H(\Psi(y),x) = c_{\text{opt}}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}) \le \operatorname{kersize}^{\text{w}}(F,\mathscr{M}_1,\mathscr{E}),$$

where we used the fact that Ψ is an optimal map in the first equality. This establishes statement (i). □

7.3. Preliminaries from measure theory

In order to prove Theorem 3.2, we need to state and cite some existing results from measure theory and prove a preliminary lemma.

7.3.1. Carathéodory functions

Let (S,Σ_S), (X,Σ_X) and (Y,Σ_Y) be measurable spaces. For functions $f : S\times X \to Y$, whose domain is a product of two measurable spaces, several notions of measurability exist. The function f could be *jointly measurable*, i.e., measurable with respect to the product σ-algebra $\Sigma_S \otimes \Sigma_X$. For fixed $s \in S$ or $x \in X$, the functions $f^s = f(s,\cdot) : X \to Y$ and $f^x = f(\cdot,x) : S \to Y$, could be measurable with respect to Σ_X and Σ_S, respectively. A function f which is measurable in one variable for each fixed $s \in S$ and for each fixed $x \in X$ is said to be *separately measurable*. In general, joint measurability implies separate measurability, but the converse is not true [6, p. 152].

A class of functions which are jointly measurable in many important cases are Carathéodory functions.

Definition 6 ([6, Definition 4.50]). *Let (S,Σ) be a measurable space, and let X and Y be topological spaces. A function $f : S\times X \to Y$ is a Carathéodory function if:*

(1) for each $x \in X$, the function $f^x = f(\cdot,x) : S \to Y$ is $(\Sigma,\mathscr{B}(Y))$-measurable, and

(2) for each $s \in S$, the function $f^s = f(s,\cdot) : X \to Y$ is continuous.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

In particular, if X is a separable metric space and Y is a metric space, then every Carathéodory function $f\colon S \times X \to Y$ is jointly measurable [6, Lemma 4.51]. It is also straightforward to see that if f is continuous in both arguments separately, then f is a Carathéodory function.

7.3.2. Measurability of set-valued functions

For a single-valued function $f\colon Y \to X$, the inverse image of a set $A \subset X$ is $f^{-1}(A) := \{y \in Y\colon f(y) \in A\}$. For a set-valued function $\varphi\colon Y \rightrightarrows X$, there are different ways to generalise the concept of an inverse image: Given a set $A \subset X$, we say that the *upper inverse* of A is $\varphi^{\mathrm{u}}(A) := \{y \in Y : \varphi(y) \subset A\}$ and we say that the *lower inverse* of A is

$$\varphi^{\ell}(A) := \{y \in Y : \varphi(y) \cap A \neq \emptyset\}.$$

Naturally, the notion of continuity of set-valued mappings between topological spaces depends on the definition of inverse. Interested readers are referred to [6, Chapter 17]. In this work, we shall be most concerned with the lower inverse of a set-valued mapping, since this notion of inverse allows us to pick a measurable selector. Next, we define two different notions of measurability for set-valued mappings based on the concept of lower inverse.

Definition 7 ([6, Definition 18.1]). *Let (Y, Σ) be a measurable space and X a topological space. We say that a set-valued mapping $\varphi : Y \rightrightarrows X$ is:*

- *weakly measurable, if $\varphi^{\ell}(V) \in \Sigma$ for each open subset V of X,*
- *measurable, if $\varphi^{\ell}(A) \in \Sigma$ for each closed subset A of X.*

Importantly, one needs to remark here that the definition of weakly measurable for set-valued mappings in [6, Definition 18.1] is equivalent to that of measurable set-valued mappings [14, Definition 8.1.1. (i), (ii)]. Additionally, we have that in some cases measurability implies weak measurability for set-valued mappings.

Lemma 7.5 ([6, Lemma 18.2]). *Let $\varphi : Y \rightrightarrows X$ be a set-valued mapping from a measurable space (Y, Σ) into a metric space. Then the following hold:*

(1) If φ is measurable, then it is also weakly measurable.

(2) If φ is compact-valued and weakly measurable, then it is measurable.

Note that if the topology of X above is induced by a metric, then every set-valued measurable mapping is also weakly measurable [6, Lemma 18.2]. Our main tool for proving Theorem 3.2 is the following result.

Theorem 7.6 (Measurable maximum theorem, [6, Theorem 18.19]). *Let X be a separable metric space and (S, Σ) a measurable space. Let $\varphi\colon S \rightrightarrows X$ be weakly measurable with non-empty compact values, and suppose $f\colon S \times X \to \mathbb{R}$ is a Carathéodory function. Define the value function $m\colon S \to \mathbb{R}$ by*

$$m(s) = \max_{x \in \varphi(s)} f(s, x),$$

and the correspondence $\Phi\colon S \rightrightarrows X$ of maximisers by

$$\Phi(s) = \{x \in \varphi(s) : f(s, x) = m(s)\} = \operatorname*{argmax}_{x \in \varphi(s)} f(s, x).$$

Then m is measurable, Φ has non-empty and compact values, and Φ is measurable and admits a measurable selector.

Note that for real-valued functions $\min_{x \in X} f(x) = \max_{x \in X} (-f(x))$ and the Measurable Maximum Theorem can be applied to minimisation problems as well. The Measurable Maximum Theorem will be applied in two ways. The first can be found in Proposition 7.7, which gives sufficient conditions for a

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

map $\phi\colon \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ to be in the set $\mathscr{C}$ given by (11). The second use of the theorem above is found in the proof of part (ii) and (iii) of Theorem 3.2.

Proposition 7.7. *If $\phi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ is measurable and has non-empty compact values, then $\phi \in \mathscr{C}$.*

Proof. By definition of $\mathscr{C}$, we need to prove that the function

$$r_\phi : \mathscr{M}_1 \times \mathscr{E} \to \mathbb{R}, \quad r_\phi(x,e) = d_{\mathscr{X}}^H(x, \phi(F(x,e))) = \sup_{z\in\phi(F(x,e))} d_{\mathscr{X}}(x,z)$$

is measurable. Denote $S := \mathscr{M}_1 \times \mathscr{E}$, $X := \mathscr{X}$, and

$$\varphi := \phi \circ F : \mathscr{M}_1 \times \mathscr{E} \rightrightarrows \mathscr{X}, \qquad \varphi(x,e) = \phi(F(x,e)),$$
$$f := d_{\mathscr{X}} \circ (\pi_1, \mathrm{id}_{\mathscr{X}}) : (\mathscr{M}_1 \times \mathscr{E}) \times \mathscr{X} \to \mathbb{R}, \qquad f((x,e),z) = d_{\mathscr{X}}(x,z).$$

We now show that the set-valued mapping $\varphi := \phi \circ F : \mathscr{M}_1 \times \mathscr{E} \rightrightarrows \mathscr{X}$ is weakly measurable. Let $V \subset \mathscr{X}$ be an open set. Then, by the definition [6, Section 17.1] of the lower inverse, φ is given by

$$\varphi^\ell(V) = \{(x,e) \in \mathscr{M}_1 \times \mathscr{E} \colon \varphi(x,e) \cap V \neq \emptyset\} = \{(x,e) \in \mathscr{M}_1 \times \mathscr{E} \colon \phi(F(x,e)) \cap V \neq \emptyset\}.$$

Observe that this set can be written as $\varphi^\ell(V) = F^{-1}(\phi^\ell(V))$, where $\phi^\ell(V) = \{y \in \mathscr{M}_2^{\mathscr{E}} : \phi(y) \cap V \neq \emptyset\}$ denotes the lower inverse of V under ϕ. Since ϕ is a set-valued mapping that is measurable, by Lemma 7.5, it follows that ϕ is also weakly measurable. As $\phi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ is weakly measurable and as $V \subset \mathscr{X}$ is an open set, we have $\phi^\ell(V) \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$. Moreover, $F\colon (\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E})) \to (\mathscr{M}_2^{\mathscr{E}}, \mathscr{B}(\mathscr{M}_2^{\mathscr{E}}))$ is Borel measurable by Assumption 3. Therefore, the preimage of the Borel set $\phi^\ell(V) \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$ under F is Borel measurable

$$\varphi^\ell(V) = F^{-1}(\phi^\ell(V)) \in \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}).$$

Hence, the set-valued mapping $\varphi : \mathscr{M}_1 \times \mathscr{E} \rightrightarrows \mathscr{X}$ is weakly measurable. Moreover, $\varphi : \mathscr{M}_1 \times \mathscr{E} \rightrightarrows \mathscr{X}$ has non-empty compact values because $\phi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ has non-empty compact values by assumption. By Assumption 3, $(\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}), \mu)$ is a $\sigma-$finite measure space. The space $(\mathscr{X}, d_{\mathscr{X}})$ is separable, as it is second countable by Assumption 11, and together with Assumption 12, it is also complete. Finally, the function f is continuous in both arguments and hence Carathéodory. Thus, the assumptions in Theorem 7.6 are satisfied, and its application implies that the value function $m : \mathscr{M}_1 \times \mathscr{E} \to \mathbb{R}$

$$m(x,e) = \max_{z\in\phi(F(x,e))} d_{\mathscr{X}}(x,z)$$

is measurable. It is immediate to notice that $m = r_\phi$, so we conclude that r_ϕ is measurable. Hence we have proven that $\phi \in \mathscr{C}$. □

7.3.3. Disintegration of measures

Next, we prove a useful proposition that will be applied in the proof of Theorem 3.2. For completeness, we start by recalling that every time we write an expression of the form

$$\operatorname*{esssup}_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} f(x,e), \qquad \operatorname*{esssup}_{(x,e)\in F^{-1}(y)} f(x,e), \qquad \operatorname*{esssup}_{y\in\mathscr{M}_2^{\mathscr{E}}} f(y).$$

Whenever it is clear from context, it is implicit that we are taking the essential supremum with respect to μ on $\mathscr{M}_1 \times \mathscr{E}$, with respect to μ^y on F_y, and with respect to $F_*\mu$ on $\mathscr{M}_2^{\mathscr{E}}$. We also recall that $\pi_1\colon \mathscr{X} \times \mathscr{Z} \to \mathscr{X}$ refers to the projection on the first component $\pi_1(x,e) = x$.

Proposition 7.8. *Given Assumptions 1, 3 and 4, we have the following:*

(1) For $A \in \mathscr{B}(\mathscr{M}_1 \times \mathscr{E})$, $\mu(A) = 0$ if and only if $\mu^y(A) = 0$ for almost every $y \in \mathscr{M}_2^{\mathscr{E}}$.

(2) Let $f : \mathscr{M}_1 \times \mathscr{E} \to [0, +\infty)$ be Borel measurable. Then the function

$$m\colon \mathscr{M}_2^{\mathscr{E}} \to \mathbb{R}, \quad m(y) = \operatorname*{esssup}_{(x,e)\in F^{-1}(y)} f(x,e)$$

is Borel measurable.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

(3) Let $f : \mathscr{M}_1 \times \mathscr{E} \to [0, +\infty)$ be Borel measurable. Then

$$\operatorname*{esssup}_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} f(x,e) = \operatorname*{esssup}_{y\in\mathscr{M}_2^{\mathscr{E}}} \operatorname*{esssup}_{(x,e)\in F^{-1}(y)} f(x,e).$$

(5) Let $g : \mathscr{M}_1 \to [0, +\infty)$ be Borel measurable. Then

$$\int_{F^{-1}(y)} g(\pi_1(x,e))d\mu^y(x,e) = \int_{F_y} g(x)d\big(\pi_{1*}\mu^y\big)(x),$$

and

$$\mu^y - \operatorname*{esssup}_{(x,e)\in F^{-1}(y)} g(\pi_1(x,e)) = \pi_{1*}\mu^y - \operatorname*{esssup}_{x\in F_y} g(x),$$

where $\mu^y - \operatorname{esssup}$, respectively, and $\pi_{1*}\mu^y - \operatorname{esssup}$, denote the essential supremum with respect to the disintegration of measure μ^y and, respectively, the pushforward of the disintegration of measure $\pi_{1*}\mu^y$.

Proof. We start with (*i*). Let 1_A denote the indicator function on $A \in \mathscr{B}(\mathscr{M}_1 \times \mathscr{E})$. Suppose that $\mu(A) = 0$, then by (iii) in Definition 3, we have that

$$0 = \mu(A) = \int_{\mathscr{M}_1\times\mathscr{E}} 1_A \, d\mu = \int_{\mathscr{M}_2^{\mathscr{E}}} \left(\int_{F^{-1}(y)} 1_A \, d\mu^y \right) d(F_*\mu)(y).$$

The integral of a positive function with respect to a positive measure is zero if and only if the integrand function is zero almost everywhere, so the previous equality is equivalent to

$$\int_{F^{-1}(y)} 1_A \, d\mu^y = \mu^y(A) = 0 \quad \text{for almost every } y \in \mathscr{M}_2^{\mathscr{E}}.$$

The reverse argument can be made with the same steps. Let us now prove (*ii*). Since the Borel σ-algebra on $\mathbb{R}$ is generated by sets of form $(a, +\infty)$, we proceed to show that $m^{-1}((a, +\infty))$ is Borel measurable for any $a \in \mathbb{R}$. We have

$$\begin{aligned} m^{-1}((a,+\infty)) &= \big\{y \in \mathscr{M}_2^{\mathscr{E}} : m(y) > a\big\} \\ &= \big\{y \in \mathscr{M}_2^{\mathscr{E}} : \mu^y(\{(x,e) \in \mathscr{M}_1 \times \mathscr{E} : f(x,e) > a\}) > 0\big\} \\ &= \Big\{y \in \mathscr{M}_2^{\mathscr{E}} : \int_{\mathscr{M}_1\times\mathscr{E}} 1_{f^{-1}((a,\infty))} \, d\mu^y > 0\Big\} \\ &= \big\{y \in \mathscr{M}_2^{\mathscr{E}} : h_a(y) > 0\big\} = h_a^{-1}((0,+\infty)), \end{aligned}$$

where $h_a(y) = \int_{\mathscr{M}_1\times\mathscr{E}} 1_{f^{-1}((a,\infty))} \, d\mu^y$. In particular, since f is Borel measurable, $f^{-1}((a,\infty))$ is a measurable set, and hence, $1_{f^{-1}((a,\infty))}$ is a measurable function on $(\mathscr{M}_1 \times \mathscr{E}, \mathscr{B}(\mathscr{M}_1 \times \mathscr{E}))$. Now, from (ii), in Definition 3, we know that the function $y \mapsto h_a(y) = \int_{\mathscr{M}_1\times\mathscr{E}} 1_{f^{-1}((a,\infty))} \, d\mu^y$ is measurable, and so $h_a^{-1}((0,+\infty)) = m^{-1}((a,+\infty))$ is a Borel measurable subset of $\mathscr{M}_2^{\mathscr{E}}$. As $a \in \mathbb{R}$ was arbitrary, this proves that m is Borel-measurable. This concludes (*ii*). Next we consider (*iii*), and let m be as in (*ii*). We start by showing that

$$\operatorname*{esssup}_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} f(x,e) \geq \operatorname*{esssup}_{y\in\mathscr{M}_2^{\mathscr{E}}} m(y). \tag{22}$$

Let $K = \operatorname{esssup}_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} f(x,e)$. By definition, we have $\mu(\{(x,e) \in \mathscr{M}_1 \times \mathscr{E} : f(x,e) > K\}) = 0$. Hence, by Proposition 7.8, (*i*), and the fact that μ^y is concentrated on $F^{-1}(y)$, we have

$$0 = \mu^y(\{(x,e) \in \mathscr{M}_1 \times \mathscr{E} : f(x,e) > K\}) = \mu^y(\{(x,e) \in F^{-1}(y) : f(x,e) > K\})$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

for almost every $y \in \mathscr{M}_2^{\mathscr{E}}$. So $m(y) = \operatorname{esssup}_{(x,e)\in F^{-1}(y)} f(x,e) \leq K = \operatorname{esssup}_{\mathscr{M}_1\times\mathscr{E}} f$ for almost every $y \in \mathscr{M}_2^{\mathscr{E}}$. This proves (22). Next, we consider the reverse inequality. By definition of $\operatorname{esssup}_{\mathscr{M}_2^{\mathscr{E}}} m$, we have

$$m(y) = \operatorname*{esssup}_{(x,e)\in F^{-1}(y)} f(x,e) \leq \operatorname*{esssup}_{\mathscr{M}_2^{\mathscr{E}}} m$$

for almost every $y \in \mathscr{M}_2^{\mathscr{E}}$. Hence, as μ^y is concentrated on $F^{-1}(y)$

$$\mu^y(\{(x,e) \in F^{-1}(y) : f(x,e) > \operatorname*{esssup}_{\mathscr{M}_2^{\mathscr{E}}} m\}) = \mu^y(\{(x,e) \in \mathscr{M}_1 \times \mathscr{E} : f(x,e) > \operatorname*{esssup}_{\mathscr{M}_2^{\mathscr{E}}} m\}) = 0$$

for almost every $y \in \mathscr{M}_2^{\mathscr{E}}$. By Proposition 7.8, (*i*), this is equivalent to $\mu(\{(x,e) \in \mathscr{M}_1 \times \mathscr{E} : f(x,e) > \operatorname{esssup}_{\mathscr{M}_2^{\mathscr{E}}} m\}) = 0$, which implies $\operatorname{esssup}_{(x,e)\in\mathscr{M}_1\times\mathscr{E}} f(x,e) \leq \operatorname{esssup}_{y\in\mathscr{M}_2^{\mathscr{E}}} m(y)$, as desired. This concludes the proof of (*iii*).

Finally, let us prove (*iv*). Equality between integrals

$$\int_{F^{-1}(y)} (g \circ \pi_1)\, d\mu^y = \int_{F_y} g d\pi_{1*}\mu^y$$

comes directly from the definition of pushforward measure and can be found in detail [6, Theorem 13.46]. To see the equality between essential suprema, take $M \in [0, +\infty)$, then note that

$$(\pi_{1*}\mu^y)(g^{-1}(M, +\infty)) = \mu^y(\pi_1^{-1}(g^{-1}(M, +\infty))) = \mu^y((g \circ \pi_1)^{-1}(M, +\infty)).$$

Taking $M = \operatorname{esssup}_{F^{-1}(y)} (g \circ \pi_1)$, then the definition of essential supremum gives $\mu^y((g \circ \pi_1)^{-1}(M, +\infty)) = 0$, which by the previous chain of equalities implies that $0 = (\pi_{1*}\mu^y)(\{x \in F_y : g(x) > \operatorname{esssup}_{F^{-1}(y)} (g \circ \pi_1)\})$, so that $g(x) \leq \operatorname{esssup}_{F^{-1}(y)} (g \circ \pi_1)$ for $\pi_{1*}\mu^y$-almost every $x \in F_y$. This gives $\operatorname{esssup}_{x\in F_y} g(x) \leq \operatorname{esssup}_{F^{-1}(y)} (g \circ \pi_1)$, by taking instead $M = \operatorname{esssup}_{x\in F_y} g(x)$, the same reasoning proves the reverse inequality between essential suprema. Hence equality is proven, concluding the proof of (*iv*). □

7.4. Proof of Theorem 3.2

Having established the above preliminaries, we now proceed to proving Theorem 3.2. In order to do so, we will split the Theorem up into two results, Propositions 7.9 and 7.10, and prove these. These results combined provide the proof of Theorem 3.2.

Proposition 7.9 (Lower bound of part (*i*), Theorem 3.2). *Given the assumptions in Theorem 3.2, then the following holds for every $p \in [1, \infty]$:*

$$\operatorname{kersize}^a(F, \mathscr{M}_1, \mathscr{E}, p) \leq 2c^a_{\text{opt}}(F, \mathscr{M}_1, \mathscr{E}, p).$$

Proof of Proposition 7.9. Let $\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ be an arbitrary reconstruction mapping. Fix $y \in \mathscr{M}_2^{\mathscr{E}}$ and consider $(x,e), (x',e') \in \mathscr{M}_1 \times \mathscr{E}$ such that $F(x,e) = F(x',e') = y$. Then, by the triangle inequality, we deduce

$$d_{\mathscr{X}}(x,x') \leq d^H_{\mathscr{X}}(x, \varphi(F(x,e))) + d^H_{\mathscr{X}}(x', \varphi(F(x',e'))). \tag{23}$$

Let us now distinguish the two cases where $p = \infty$ and $p \in [1, \infty)$. The structure of the proof will be very similar in the two cases, with the main difference that the case $p = \infty$ involves essential suprema, while the case $p \in [1, +\infty)$ involves integrals. We will fully prove the case $p \in [1, \infty)$ and provide a sketch of the proof of the case $p = \infty$, as it is virtually identical. In both cases, however, Equation (23) will play a crucial role. To ensure that the essential suprema and integrals are well-defined, we assume that $\varphi \in \mathscr{C}$.

Case $p \in [1, \infty)$; here, we use that for $a \in [0, \infty)$ the map $a \mapsto a^p$ is convex, and by the definition of convexity, we have that for $a, b \in [0, \infty)$ that $(\frac{1}{2}a + \frac{1}{2}b)^p \leq \frac{1}{2}a^p + \frac{1}{2}b^p$. Replacing a, b with $2a, 2b$ yields $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ for $a, b \geq 0$. Now integrating (23) twice with respect to μ^y and using that

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

$(a+b)^p \le 2^{p-1}(a^p+b^p)$ for $a, b \ge 0$, we obtain

$$\begin{aligned}
&\int_{F^{-1}(y)} \int_{F^{-1}(y)} d_{\mathscr{X}}(x,x')^p \, d\mu^y(x,e) \, d\mu^y(x',e') \\
&\le \int_{F^{-1}(y)} \int_{F^{-1}(y)} \left(d^H_{\mathscr{X}}(x, \varphi(F(x,e))) + d^H_{\mathscr{X}}(x', \varphi(F(x',e'))) \right)^p \, d\mu^y(x,e) \, d\mu^y(x',e') \\
&= 2^p \int_{F^{-1}(y)} d^H_{\mathscr{X}}(x, \varphi(F(x,e)))^p \, d\mu^y(x,e)
\end{aligned}$$

where in the last step we also used the fact that μ^y is a probability measure and the integrals are well-defined as $\varphi \in \mathscr{C}$. Now, integrating the above inequality with respect to $F_*\mu$ on $\mathscr{M}_2^{\mathscr{E}}$, by (iii) in Definition 3, and raising to the power $\frac{1}{p}$ gives that

$$\begin{aligned}
\text{kersize}^{\text{a}}(F, \mathscr{M}_1, \mathscr{E}, p) &= \left(\int_{\mathscr{M}_2^{\mathscr{E}}} \int_{F_y} \int_{F_y} d_{\mathscr{X}}(x,x')^p \, d\mu^y(x,e) \, d\mu^y(x',e') \, d(F_*\mu)(y) \right)^{\frac{1}{p}} \\
&\le \left(2^p \int_{\mathscr{M}_2^{\mathscr{E}}} \int_{F_y} d^H_{\mathscr{X}}(x, \varphi(F(x,e)))^p \, d\mu^y(x,e) \, d(F_*\mu)(y) \right)^{\frac{1}{p}} \\
&= 2 \left(\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(x, \varphi(F(x,e)))^p \, d\mu(x,e) \right)^{\frac{1}{p}}.
\end{aligned}$$

Since $\varphi \in \mathscr{C}$ was arbitrary, by taking the infimum over $\varphi \in \mathscr{C}$ we obtain:

$$\begin{aligned}
\text{kersize}^{\text{a}}(F, \mathscr{M}_1, \mathscr{E}, p) &\le 2 \inf_{\varphi \in \mathscr{C}} \left(\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(x, \varphi(F(x,e)))^p \, d\mu(x,e) \right)^{\frac{1}{p}} \\
&= 2c^{\text{a}}_{\text{opt}}(F, \mathscr{M}_1, \mathscr{E}, p).
\end{aligned}$$

The proposition is therefore also proven in the case $p \in [1, +\infty)$.

Case $p = \infty$; the proof follows the basic structure of the case $p \in [1, \infty)$. In (23), instead of integrating, we take the essential supremum respect to the measure μ^y. Then in the next step, instead of integrating, we take the essential supremum in y with respect to the measure $F_*\mu$ on $\mathscr{M}_2^{\mathscr{E}}$ and apply Proposition 7.8. Finally, as $\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ was arbitrary, taking the infimum over $\varphi \in \mathscr{C}$ concludes the proof. □

Proposition 7.10 (Part (*ii*) and (*iii*) and upper bound in (i) of Theorem 3.2). *The following holds for every* $p \in [1, \infty]$*:*

$$c^{a}_{\text{opt}}(F, \mathscr{M}_1, \mathscr{E}, p) \le \text{kersize}^{a}(F, \mathscr{M}_1, \mathscr{E}, p),$$

and the map $\Psi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ *given by,*

$$\Psi(y) = \text{argmin}_{z \in \mathscr{X}} \, \text{esssup}_{(x,e) \in F^{-1}(y)} \, d_{\mathscr{X}}(x,z) \qquad (p = \infty) \tag{24}$$

$$\Psi(y) = \text{argmin}_{z \in \mathscr{X}} \int_{F^{-1}(y)} d_{\mathscr{X}}(x,z)^p \, d\mu^y(x,e) \qquad (p \in [1, \infty)) \tag{25}$$

is an optimal map with average error of order p. Moreover, Ψ *has non-empty compact values, is measurable and it admits a measurable selector.*

Proof of Proposition 7.10. We distinguish the cases $p = \infty$ and $p \in [1, \infty)$. In both cases, the structure of the proof consists in proving the following steps:

(1) Ψ, *defined either by* (24) *or* (25), *has non-empty values;*

(2) Ψ *is measurable, has compact values and admits a measurable selector;*

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

(3) $\Psi \in \mathscr{C} = \{\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X} : (x,e) \mapsto d_{\mathscr{X}}^H(x, \varphi(F(x,e))\text{is measurable}\}$;

(4) Ψ *is an optimal map with average error of order* p;

(5) *Upper bound:* $c_{\text{opt}}^{\text{a}}(F, \mathscr{M}_1, \mathscr{E}, p) \leq \text{kersize}^{\text{a}}(F, \mathscr{M}_1, \mathscr{E}, p)$ *with* $p = \infty$ *or* $p \in [1, +\infty)$.

First, we consider the case $p \in [1, \infty)$. Again, let us first introduce some notation, similar to above. Fix $y \in \mathscr{M}_2^{\mathscr{E}}$. Define $f_y : \mathscr{X} \to [0, \infty)$,

$$f_y(z) = \int_{F^{-1}(y)} d_{\mathscr{X}}(x,z)^p \, d\mu^y(x,e)$$

for $z \in \mathscr{X}$. Define also

$$\begin{aligned} r_y &= \operatorname*{esssup}_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_{\mathscr{X}}(x,x'), \\ E_y^{(x,e)} &= \{(x',e') \in F^{-1}(y) : d_{\mathscr{X}}(x,x') > r_y\} \quad \text{for } (x,e) \in F^{-1}(y), \\ G_y &= \{(x,e) \in F^{-1}(y) : \mu^y(E_y^{(x,e)}) = 0\}. \end{aligned}$$

Note that f_y is Borel measurable. The function $((x,e),z) \mapsto 1_{F^{-1}(y)}(x,e)\, d_{\mathscr{X}}(x,z)$ is non-negative and Borel measurable. By the Fubini-Tonelli [55, Theorem 2.37] f_y, as $z \mapsto \int_{F^{-1}(y)} d_{\mathscr{X}}(x,z)^p \, d\mu^y(x,e)$, is Borel measurable. Additionally, by a similar line of reasoning, one can show that the function $y \mapsto r_y$ is Borel measurable. Additionally, as the metric space $(\mathscr{M}_1, d_{\mathscr{X}})$ is compact by Assumption 4, the set $\mathscr{M}_1$ is bounded. Thus, as $F^{-1}(y) \subset \mathscr{M}_1 \times \mathscr{E}$, we have that

$$r_y \leq \sup_{\substack{(x,e)\in F^{-1}(y)\\(x',e')\in F^{-1}(y)}} d_{\mathscr{X}}(x,x') \leq \operatorname{diam}(\mathscr{M}_1) < \infty.$$

We now show that the set G_y is non-empty and $\mu^y(G_y) = 1$. Fix $y \in \mathscr{M}^{\mathscr{E}}$. By definition of the essential supremum, it follows that

$$(\mu^y \otimes \mu^y)\big(\{((x,e),(x',e')) \in F^{-1}(y) \times F^{-1}(y) : d_{\mathscr{X}}(x,x') > r_y\}\big) = 0.$$

Then, by Fubini-Tonelli [55, Theorem 2.37],

$$\begin{aligned} 0 &= (\mu^y \otimes \mu^y)\left(\{((x,e),(x',e')) \in F^{-1}(y) \times F^{-1}(y) : d_{\mathscr{X}}(x,x') > r_y\}\right) \\ &= \int_{F^{-1}(y)} \mu^y\big(E_y^{(x,e)}\big)\, d\mu^y(x,e). \end{aligned}$$

Since the integrand is nonnegative, we conclude that

$$\mu^y\big(E_y^{(x,e)}\big) = 0 \quad \text{for } \mu^y\text{-a.e. } (x,e) \in F^{-1}(y).$$

Recall the definition $G_y = \{(x,e) \in F^{-1}(y) : \mu^y(E_y^{(x,e)}) = 0\}$. As the disintegration μ^y is a probability measure that is concentrated on $F^{-1}(y)$, Definition 3 (i), we have that $\mu^y(G_y) = 1$ and, in particular, $G_y \neq \emptyset$.

Claim. *We claim that the following holds: for every* $y \in \mathscr{M}_2^{\mathscr{E}}$

(I) f_y *is continuous,*

(II) $\operatorname*{argmin}_{\mathscr{X}} f_y = \operatorname*{argmin}_{B(x,2r_y)} f_y$ *for* μ^y*-almost every* $(x,e) \in F^{-1}(y)$.

We proceed to prove the claim and start by considering (I). We consider a general setting and let (X,d) be a metric space equipped with a probability measure ν concentrated on a bounded subset $A \subseteq X$, and define

$$g(x) := \Big(\int_A d(x,a)^p \, d\nu(a)\Big)^{\frac{1}{p}} = \|d(x,\cdot)\|_{L^p(X,\nu)} = \|d(x,\cdot)\|_p.$$

We claim that $|g(x) - g(z)| \leq d(x,z)$ for all $x, y \in X$. To see this, consider $a, x, z \in X$ and note that by the triangle inequality $d(x,a) \leq d(x,z) + d(z,a)$. Taking $L^p(X,\nu)$-norms in the variable a and applying

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Minkowski's inequality gives

$$g(x) = \|d(x,\cdot)\|_p \le \|d(x,z) + d(z,\cdot)\|_p \le \|d(x,z)\|_p + \|d(z,\cdot)\|_p = d(x,z) + g(z),$$

where in the last passage we used that ν is a probability measure. Switching the roles of x and z, leads to the desired inequality. It follows that g is continuous. By assumption 4, since $\mathscr{M}_1$ is compact and $F_y \subseteq \mathscr{M}_1$, then F_y is bounded. Thus, letting $(X,d) = (\mathscr{X}, d_{\mathscr{X}})$, $\nu = \pi_{1*}\mu^y$, $A = F_y$ above, and recalling Proposition 7.8, (iv), then $g^p = f_y$, which proves (I).

To prove (II), we will show that for μ^y-almost every $(x,e) \in F^{-1}(y)$ we have

$$\Psi(y) = \operatorname{argmin}_{z\in\mathscr{X}} \int_{F^{-1}(y)} d_{\mathscr{X}}(x',z)^p \, d\mu^y(x',e') = \operatorname{argmin}_{z\in B_{d_{\mathscr{X}}}(x,2r_y)} f_y(z). \tag{26}$$

Fix $(x,e) \in G_y$. If $z \in \mathscr{X} \setminus B_{d_{\mathscr{X}}}(x,2r_y)$, then for μ^y-almost every $(x',e') \in F^{-1}(y)$,

$$d_{\mathscr{X}}(z,x') \ge d_{\mathscr{X}}(z,x) - d_{\mathscr{X}}(x,x') > 2r_y - r_y = r_y.$$

Thus, the previous inequality holds for any $z \in \mathscr{X} \setminus B_{d_{\mathscr{X}}}(x,2r_y)$. Thus,

$$\begin{aligned} f_y(z) &= \int_{F^{-1}(y)} d_{\mathscr{X}}(z,x')^p \, d\mu^y(x',e') \\ &> \int_{F^{-1}(y)} r_y^p \, d\mu^y(x',e') = r_y^p. \end{aligned}$$

On the other hand,

$$\begin{aligned} f_y(x) &= \int_{F^{-1}(y)} d_{\mathscr{X}}(x,x')^p \, d\mu^y(x',e') \\ &\le \int_{F^{-1}(y)} r_y^p \, d\mu^y(x',e') = r_y^p. \end{aligned}$$

Therefore, $f_y(z) > f_y(x)$ whenever $z \notin B(x,2r_y)$, which implies that points z outside of such a ball cannot be minimisers, hence proving (26). This concludes part (II) of the claim.

With the claim, we can prove the required (a)–(e) properties of Ψ. First, let us prove (a), namely that Ψ in (25) has non-empty values. By the Heine-Borel property of the metric $d_{\mathscr{X}}$, the set $B_{d_{\mathscr{X}}}(x,2r_y) \subset \mathscr{X}$ is compact, since it is a closed ball with respect to the metric $d_{\mathscr{X}}$. Hence, the function f_y is continuous by (I) and its minimisers are by (II) restricted to the compact set $B_{d_{\mathscr{X}}}(x,2r_y)$. Therefore, the minimum in (25) is attained by the Extreme Value Theorem. This shows that the argmin is non-empty, and hence that Ψ has non-empty values on $\mathscr{M}_2^{\mathscr{E}}$.

We now proceed to prove (b), namely that Ψ is measurable and has compact values. Recall that above we have shown that $\operatorname{diam}(\mathscr{M}_1) < \infty$. We will apply the Maximum Measurable Theorem, 7.6, with $S = \mathscr{M}_2^{\mathscr{E}}$, $X = \mathscr{X}$, and

$$\begin{aligned} &\varphi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}, \qquad \varphi(y) = B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\operatorname{diam}(\mathscr{M}_1)), \\ &f : \mathscr{M}_2^{\mathscr{E}} \times \mathscr{X} \to \mathbb{R}, \; f(y,z) = f_y(z) = \int_{F^{-1}(y)} d_{\mathscr{X}}(x,z)^p \, d\mu^y(x,e). \end{aligned}$$

In order to apply the theorem, we need to verify that the assumptions are satisfied. We start by proving that φ is weakly measurable with non-empty compact values. Firstly, it is clear that, since φ is constant, then φ is weakly measurable and has non-empty values. Moreover, the only value φ takes is $B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\operatorname{diam}(\mathscr{M}_1))$, which we now prove to be compact. Recall that the distance is $\operatorname{dist}_{d_{\mathscr{X}}}(x,A) = \inf_{a\in A} d_{\mathscr{X}}(x,a)$. Note that, since $\mathscr{M}_1$ is compact by Assumption 4, then $\mathscr{M}_1$ is bounded and hence $B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\operatorname{diam}(\mathscr{M}_1)) = \{x \in \mathscr{X} \colon \operatorname{dist}_{d_{\mathscr{X}}}(x,\mathscr{M}_1) \le 2\operatorname{diam}(\mathscr{M}_1)\}$ is bounded too.

We now show that the function $\operatorname{dist}_{d_{\mathscr{X}}}(\cdot,\mathscr{M}_1)$ is continuous. We consider a general setting, and let (X,d) be a metric space, $A \subseteq X$ be a bounded subset and $g(x) := \operatorname{dist}_d(x,A) = \inf_{a\in A} d(x,a)$. We claim that $|g(x) - g(y)| \le d(x,y)$ for all $x, y \in X$. To see this, consider $a, x, y \in X$ and note that

$$d(x,a) \le d(x,y) + d(y,a) \quad \Longrightarrow \quad \inf_{a\in A} d(x,a) \le d(x,y) + \inf_{a\in A} d(y,a).$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Switching the roles of x and y, leads to the desired inequality. It follows that g is continuous. Moreover, since the function $\mathrm{dist}_{d_{\mathscr{X}}}(\cdot, \mathscr{M}_1)$ is continuous, the set

$$B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\,\mathrm{diam}\,(\mathscr{M}_1)) = \mathrm{dist}_{d_{\mathscr{X}}}(\cdot, \mathscr{M}_1)^{-1}([0, 2\,\mathrm{diam}\,(\mathscr{M}_1)])$$

is closed. Hence, we have proven that $B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\mathrm{diam}(\mathscr{M}_1))$ is closed and bounded, and by the Heine Borel property of $d_{\mathscr{X}}$ granted by Assumption 4 it follows that $B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\mathrm{diam}(\mathscr{M}_1))$ is compact. This proves that φ is weakly measurable with non-empty compact values. Secondly, to prove that f is Carathéodory, we need to show that $f(y, \cdot) = f_y$ is continuous for every fixed $y \in \mathscr{M}_2^{\mathscr{E}}$ and that $f(\cdot, z)$ is measurable for every fixed $z \in \mathscr{M}_1$. On the one hand, for every fixed $y \in \mathscr{M}_2^{\mathscr{E}}$, the function f_y is continuous on $\mathscr{X}$ as proven in claim (I). On the other hand, for every fixed z, the function $f(\cdot, z)$: $y \mapsto \int_{F^{-1}(y)} d_{\mathscr{X}}(x, z)^p \, d\mu^y(x, e)$ is Borel measurable, described by Definition 3 (ii). Then, by Theorem 7.6, the possibly set-valued function $\Phi : \mathscr{M}_2^{\mathscr{E}} \rightrightarrows \mathscr{X}$ given by

$$\Phi(y) = \operatorname*{argmin}_{z \in B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\mathrm{diam}(\mathscr{M}_1))} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, z)^p \, d\mu^y(x, e)$$

is measurable and has non-empty, compact values. Moreover, by combining (26) and the fact that for $(x, e) \in G_y$, $B_{d_{\mathscr{X}}}(x, 2r_y) \subseteq B_{d_{\mathscr{X}}}(\mathscr{M}_1, 2\mathrm{diam}(\mathscr{M}_1))$, we deduce that $\Phi = \Psi$. Hence, Ψ is measurable and has non-empty, compact values.

We now prove (c), namely that $\Psi \in \mathscr{C}$. This follows directly from Proposition 7.7.

We now proceed to prove (d), namely that Ψ is an optimal map. Let $\varphi \in \mathscr{C}$. By the minimising definition of Ψ,

$$\int_{F^{-1}(y)} d^H_{\mathscr{X}}(\Psi(y), x)^p \, d\mu^y(x, e) \leq \int_{F^{-1}(y)} d_{\mathscr{X}}(z, x)^p \, d\mu^y(x, e)$$

for every $z \in \varphi(y)$. In particular, taking the supremum with respect to $z \in \varphi(y)$, which coincides with considering the Hausdorff distance, and using Fatou's Lemma yields

$$\begin{aligned}\int_{F^{-1}(y)} d^H_{\mathscr{X}}(\Psi(y), x)^p \, d\mu^y(x, e) &\leq \sup_{z \in \varphi(y)} \int_{F^{-1}(y)} d_{\mathscr{X}}(z, x)^p \, d\mu^y(x, e) \\ &\leq \int_{F^{-1}(y)} \sup_{z \in \varphi(y)} d_{\mathscr{X}}(z, x)^p \, d\mu^y(x, e) \\ &\leq \int_{F^{-1}(y)} d^H_{\mathscr{X}}(\varphi(y), x)^p \, d\mu^y(x, e).\end{aligned}$$

By integrating with respect to $y \in \mathscr{M}_2^{\mathscr{E}}$, we obtain

$$\begin{aligned}&\int_{y \in \mathscr{M}_2^{\mathscr{E}}} \int_{F^{-1}(y)} d^H_{\mathscr{X}}(\Psi(y), x)^p \, d\mu^y(x, e) \, d(F_* \mu)(y) \\ &\quad \leq \int_{y \in \mathscr{M}_2^{\mathscr{E}}} \int_{F^{-1}(y)} d^H_{\mathscr{X}}(\varphi(y), x)^p \, d\mu^y(x, e) \, d(F_* \mu)(y).\end{aligned}$$

Due to Definition 3 (iii) of the disintegration of measure, the above integrals can be rewritten as

$$\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(\Psi(F(x, e)), x)^p \, d\mu(x, e) \leq \int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(\varphi(F(x, e)), x)^p \, d\mu(x, e).$$

Now, as $\varphi \in \mathscr{C}$ was arbitrary and by raising both sides to the power $\frac{1}{p}$, we obtain

$$\left(\int_{(x,e) \in \mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(\Psi(F(x, e)), x)^p \, d\mu(x, e) \right)^{\frac{1}{p}} \leq c^{\mathrm{a}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, p).$$

The opposite inequality holds trivially, as $\Psi \in \mathscr{C}$. Therefore, Ψ is an optimal map.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Finally, we proceed to prove (e), namely the upper bound $c^{\mathrm{a}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, p) \leq \mathrm{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, p)$. By the minimisation property of Φ, hence also of Ψ, for every $(x', e') \in F^{-1}(y)$ and for $F_*\mu$-a.e. $y \in \mathscr{M}_2^{\mathscr{E}}$:

$$\int_{F^{-1}(y)} d^H_{\mathscr{X}}(x, \Psi(y))^p \, d\mu^y(x, e) \leq \int_{F^{-1}(y)} d_{\mathscr{X}}(x, x')^p \, d\mu^y(x, e). \tag{27}$$

Integrating (27) with respect to μ^y in the (x', e') variables yields,

$$\int_{F^{-1}(y)} d^H_{\mathscr{X}}(x, \Psi(y))^p \, d\mu^y(x, e) \leq \int_{F^{-1}(y)} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, x')^p \, d\mu^y(x, e) \, d\mu^y(x', e').$$

where we used that μ^y is a probability measure. Integrating both sides over $y \in \mathscr{M}_2^{\mathscr{E}}$ with respect to $F_*\mu$ we obtain

$$\begin{aligned}&\int_{\mathscr{M}_2^{\mathscr{E}}} \int_{F_y} d^H_{\mathscr{X}}(x, \Psi(y))^p \, d\mu^y(x, e) \, d(F_*\mu)(y) \\ &\quad \leq \int_{\mathscr{M}_2^{\mathscr{E}}} \int_{F^{-1}(y)} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, x')^p \, d\mu^y(x, e) \, d\mu^y(x', e') \, d(F_*\mu)(y).\end{aligned}$$

Using the definition of disintegration of the measure μ on the left hand side of the above inequality yields

$$\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(x, \Psi(F(x, e)))^p \, d\mu(x, e) \leq \mathrm{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, p)^p.$$

Then, raising both sides to the $\frac{1}{p}$-th power, gives

$$\left(\int_{\mathscr{M}_1 \times \mathscr{E}} d^H_{\mathscr{X}}(x, \Psi(F(x, e)))^p \, d\mu(x, e) \right)^{\frac{1}{p}} \leq \mathrm{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, p).$$

And finally, since $\Psi \in \mathscr{C}$, we conclude

$$c^{\mathrm{a}}_{\mathrm{opt}}(F, \mathscr{M}_1, \mathscr{E}, p) \leq \mathrm{kersize}^{\mathrm{a}}(F, \mathscr{M}_1, \mathscr{E}, p).$$

This concludes the case $p \in [1, +\infty)$. Hence, the proof is complete.

Now we consider the case $p = \infty$. The proof only requires minor modifications to the case $p \in [1, \infty)$, that we will describe in the following. The objective function for $y \in \mathscr{M}_2^{\mathscr{E}}$ and $z \in \mathscr{X}$ is

$$f_y(z) = \operatorname*{esssup}_{(x,e) \in F^{-1}(y)} d_{\mathscr{X}}(x, z).$$

where the essential supremum is taken with respect to μ^y. Then $\Psi(y) = \operatorname{argmin}_{\mathscr{X}} f_y$. Define, as before,

$$r_y = \operatorname*{esssup}_{\substack{(x,e) \in F^{-1}(y) \\ (x',e') \in F^{-1}(y)}} d_{\mathscr{X}}(x, x').$$

Claim. *For every* $y \in \mathscr{M}_2^{\mathscr{E}}$*:*

(I) f_y *is continuous;*

(II) $\operatorname*{argmin}_{\mathscr{X}} f_y = \operatorname*{argmin}_{B(x, 2r_y)} f_y$ *for* μ^y*-almost every* $(x, e) \in F^{-1}(y)$.

The claims are proven analogously to the case $p \in [1, \infty)$, except that to prove (I), Minkowski's inequality is replaced by the following. We consider a general setting, and let (X, d) be a metric space, $A \subseteq X$ be a bounded subset, ν be a probability measure on X concentrated on A and $g(x) := \operatorname{esssup}_{a \in A} d(x, a)$. We claim that $|g(x) - g(z)| \leq d(x, z)$ for all $x, z \in X$. To see this, consider $a, x, z \in X$ and notice that

$$d(x, a) \leq d(x, z) + d(z, a) \quad \Longrightarrow \quad \operatorname{esssup}_{a \in A} d(x, a) \leq d(x, z) + \operatorname{esssup}_{a \in A} d(z, a).$$

Switching the roles of x and z, leads to the desired inequality. It follows that g is continuous. Claim (II) follows the same line of reasoning as in the case $p \in [1, \infty)$ by replacing the integration with the essential supremum with respect to μ^y.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

With the claim, we can show the properties given by the list (a)-(e) stated at the beginning of the proof, similarly to the case $p \in [1, \infty)$. For the sake of brevity, we provide a sketch of the proof.

Part (a) follows the same line of reasoning as in the case $p \in [1, \infty)$. In part (b) we again apply the Maximum Measurable Theorem 7.6 with $f(y, z) = f_y(z) := \operatorname{esssup}_{(x,e)\in F^{-1}(y)} d_{\mathscr{X}}(x, z)$ and apply Proposition 7.8 to show that $f(\cdot, z) : y \mapsto \operatorname{esssup}_{(x,e)\in F^{-1}(y)} d_{\mathscr{X}}(x, z)$ is Borel measurable. Part (c) is again a direct consequence of (b) and Proposition 7.7. Parts (d) and (e) follows the same line of reasoning as in the case $p \in [1, \infty)$ by replacing the integration with the essential supremum. This concludes the proof of the proposition in the case $p = \infty$. □

7.5. Proof of Proposition 4.1

Proof. We begin with part (1) by showing that if A satisfies the rNSP, (17), with respect to the set $\mathscr{M}_1$, then A is injective on $\mathscr{M}_1$. Let $x, x' \in \mathscr{M}_1$ such that $Ax = Ax'$, then $A(x - x') = 0$. Now we apply the rNSP, (17), to $h := x - x'$, and since $x - x' \in \mathscr{M}_1 - \mathscr{M}_1$, this yields

$$\left|\!\left|\!\left| x - x' \right|\!\right|\!\right|_{\mathscr{X}_1} \leq D_1 \operatorname{dist}_{\mathscr{X}_2}(x - x', \mathscr{M}_1 - \mathscr{M}_1) + 0 = 0.$$

This implies that $h = x - x' = 0$. Hence A is injective on $\mathscr{M}_1$.

Additionally, if A is injective on $\mathscr{M}_1$, then for $x, x' \in \mathscr{M}_1$ such that $Ax = Ax'$, then $x - x' = 0$ and we have that

$$(\mathscr{M}_1 - \mathscr{M}_1) \cap \mathscr{N}(A) = \{0\}.$$

For the converse direction, let $x, x' \in \mathscr{M}_1$ such that $Ax = Ax'$. This is equivalent to $A(x - x') = 0$ and as $(\mathscr{M}_1 - \mathscr{M}_1) \cap \mathscr{N}(A) = \{0\}$, we have that $x - x' = 0$. Thus, $x = x'$ and A is injective.

For part (2), let $y \in \mathscr{M}_2^{\mathscr{E}}$ and $x, x' \in \mathscr{M}_1$ and $e, e' \in \mathscr{E}$ be such that $y = F(x, e) = Ax + e = Ax' + e' = F(x', e')$. From the rNSP (17), as $\operatorname{diam}_{d_{\mathscr{Y}}}(\mathscr{E}) = \sup_{e,e'\in\mathscr{E}} |\!|\!| e - e' |\!|\!|_{\mathscr{Y}} = \eta$ and since $x - x' \in \mathscr{M}_1 - \mathscr{M}_1$, we then have that

$$\begin{aligned}\left|\!\left|\!\left| x - x' \right|\!\right|\!\right|_{\mathscr{X}_1} &\leq D_1 \operatorname{dist}_{\mathscr{X}_2}(x - x', \mathscr{M}_1 - \mathscr{M}_1) + D_2 \left|\!\left|\!\left| A(x - x') \right|\!\right|\!\right|_{\mathscr{Y}} \\ &= D_2 \left|\!\left|\!\left| e - e' \right|\!\right|\!\right|_{\mathscr{Y}} \leq D_2 \eta.\end{aligned}$$

Taking supremum over all $x, x' \in F_y$ and, then, over $y \in \mathscr{M}_2^{\mathscr{E}}$ yields

$$\operatorname{kersize}^{\mathrm{w}}(F, \mathscr{M}_1, \mathscr{E}) = \sup_{y\in\mathscr{M}_2^{\mathscr{E}}} \sup_{x,x'\in F_y} d_{\mathscr{X}}(x, x') \leq D_2 \eta.$$

□

7.6. Proof of Proposition 4.2

Before introducing the Bayesian framework for inverse problems, we recall the concept of regular conditional distribution.

Definition 8 (Regular conditional distribution, [48, Section 10.2]). *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $\mathbf{x} : (\Omega, \mathscr{F}) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and $\mathbf{y} : (\Omega, \mathscr{F}) \to (\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ be random variables. The regular conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is a function $\mathbb{P}_{\mathbf{x}|\mathbf{y}=\cdot} = \mathbb{P}_{\mathbf{x}}[\cdot \mid \mathbf{y} = \cdot] : \mathscr{B}(\mathscr{X}) \times \mathscr{Y} \to [0, 1]$ such that:*

(1) For every $y \in \mathscr{Y}$, the function $\mathbb{P}_{\mathbf{x}|\mathbf{y}=y} : \mathscr{B}(\mathscr{X}) \to [0, 1]$ is a probability measure;

(2) For every $A \in \mathscr{B}(\mathscr{X})$, the function $\mathbb{P}_{\mathbf{x}|\mathbf{y}=\cdot}(A) : \mathscr{Y} \to [0, 1]$ is $(\mathscr{B}(\mathscr{Y}), \mathscr{B}([0, 1]))$-measurable and

$$\mathbb{P}_{\mathbf{x}}(A) = \int_{\mathscr{Y}} \mathbb{P}_{\mathbf{x}|\mathbf{y}=y}(A)\, d\mathbb{P}_{\mathbf{y}}(y), \tag{3.5.1}$$

where $\mathbb{P}_{\mathbf{x}} = \mathbb{P} \circ \mathbf{x}^{-1}$ and $\mathbb{P}_{\mathbf{y}} = \mathbb{P} \circ \mathbf{y}^{-1}$ are the laws of x on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and of y on $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$, respectively.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

Regular conditional distributions do not always exist. The following positive result guarantees that a regular conditional distribution exists when the random variable that is being conditioned takes values in a completely metrisable and separable topological space, also referred to as a Polish space.

Theorem 7.11 (Existence and uniqueness of regular conditional distributions [48, Theorem 10.2.2] or [95, Theorem B.32 and Lemma B.40]). *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space and let* $\mathscr{Y}$ *be a Polish space. Let* $\mathbf{x} : (\Omega, \mathscr{F}) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$ *and* $\mathbf{y} : (\Omega, \mathscr{F}) \to (\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ *be random variables. Then, there exists a regular conditional distribution* $\mathbb{P}_{\mathbf{x}|\mathbf{y}=\cdot} = \mathbb{P}_{\mathbf{x}}[\cdot \mid \mathbf{y} = \cdot] : \mathscr{B}(\mathscr{X}) \times \mathscr{Y} \to [0, 1]$ *of* $\mathbf{x}$ *given* $\mathbf{y}$. *Moreover, the regular conditional distribution is unique up to* $\mathbb{P}$*-almost sure equivalence.*

Remark. Disintegrations of measure are regular conditional probabilities. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space and let* $\mathscr{X}$ *be a Polish space. Let* $\mathbf{x} : (\Omega, \mathscr{F}) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$ *be a random variable. Equip* $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ *with a probability measure given by the law of* $\mathbf{x}$, *as* $\mu_{\mathbf{x}} = \mathbb{P}_{\mathbf{x}}(\cdot)$. *Let* $F : (\mathscr{X}, \mathscr{B}(\mathscr{X})) \to (\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ *be measurable. Define the random variable* $\mathbf{y} := F \circ \mathbf{x}$ *on* $\mathscr{Y}$. *Then, a disintegration* $\{\mu_{\mathbf{x}}^y\}_{y \in \mathscr{Y}}$ *of* $\mu_{\mathbf{x}}$ *along* F *defines a regular conditional distribution of* $\mathbf{x}$ *given* $\mathbf{y}$. *The reverse, that a regular conditional distribution is a disintegration of measure, in general does not hold; as in Definition 3 (i) of the disintegration of measure, the support is required to concentrate on the preimages* $F^{-1}(y)$.

Proof. We start with part (1). Starting from the joint random variable $(\mathbf{x}, \mathbf{e}) \sim \mu_{\mathbf{x},\mathbf{e}} = \mathbb{P}[(\mathbf{x}, \mathbf{e}) \in \cdot]$, the random variable $\mathbf{x}$ can be obtained by the composition $\mathbf{x} = \pi_1 \circ (\mathbf{x}, \mathbf{e})$. Then, its distribution is the Bayesian prior of the inverse problem, and is given by the marginal distribution $\mu_{\mathbf{x}} = \pi_{1*}\mu_{\mathbf{x},\mathbf{e}} = \mathbb{P}[\mathbf{x} \in \cdot]$.

For part (2) apply Proposition 7.1, but consider a different measurable function and different spaces. We disintegrate μ along the projection $\pi_1 : (\mathscr{X} \times \mathscr{E}, \mathscr{B}(\mathscr{X} \times \mathscr{E})) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$. By Proposition 7.1 there exists a disintegration $\{\mu^x\}_{x \in \mathscr{X}}$ of μ along π_1 and each measure μ^x is a probability measure that is concentrated on $\{x\} \times \mathscr{E}$. Then, this disintegration μ^x can be pushed forward via F, as $\mu_{\mathbf{y}|\mathbf{x}=x} := F_*\mu^x$. Now we show that $\mu_{\mathbf{y}|\mathbf{x}=\cdot}$ is a regular conditional distribution of $\mathbf{y}$ given $\mathbf{x}$. As μ^x is a probability measure, for $A \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$ we have that $\mu_{\mathbf{y}|\mathbf{x}=x}(A) = \mu^x(F^{-1}(A)) \in [0, 1]$. Additionally, we have that $\mu_{\mathbf{y}|\mathbf{x}=x}(\mathscr{M}_2^{\mathscr{E}}) = \mu^x(F^{-1}(\mathscr{M}_2^{\mathscr{E}})) = 1$, since μ^x is a probability measure and as μ is supported on $\mathscr{M}_1 \times \mathscr{E}$ by Assumption 3. Thus $\mu_{\mathbf{y}|\mathbf{x}=x}$ is a probability measure. For all $V \in \mathscr{B}(\mathscr{X} \times \mathscr{E})$ the function $x \mapsto \mu^x(V)$ is measurable. For $A \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$ as F is measurable, $V = F^{-1}(A)$ is measurable, and we have that $x \mapsto \mu_{\mathbf{y}|\mathbf{x}=x}(A) = \mu^x(F^{-1}(A))$ is measurable. Now recall that by Definition 3 (iii), we have that the disintegration of measure satisfies

$$\mu(A) = \int_{\mathscr{X} \times \mathscr{E}} 1_A(x, e) d\mu(x, e) = \int_{\mathscr{X}} \left(\int_{\mathscr{E}} 1_A(x, e) d\mu^x(x, e) \right) d\pi_{1*}\mu(x) = \int_{\mathscr{X}} \mu^x(A) d\pi_{1*}\mu(x).$$

Now by Definition 3 (iii) and as $\mu_{\mathbf{y}} = F_*\mu_{\mathbf{x},\mathbf{e}}$, the property (2) in Definition 8 of conditional distributions is verified:

$$\begin{aligned} \mu_{\mathbf{y}}(A) &= F_*\mu(A) = \mu_{\mathbf{x},\mathbf{e}}(F^{-1}(A)) \\ &= \int_{\mathscr{X}} \mu^x(F^{-1}(A))\, d\pi_{1*}\mu(x) = \int_{\mathscr{X}} F_*\mu^x(A)\, d\pi_{1*}\mu(x) \\ &= \int_{\mathscr{X}} \mu_{\mathbf{y}|\mathbf{x}=x}(A)\, d\mu_{\mathbf{x}}(x), \end{aligned}$$

for every $A \in \mathscr{B}(\mathscr{M}_2^{\mathscr{E}})$. This construction proves that $\{\mu_{\mathbf{y}|\mathbf{x}=x}\}_{x \in \mathscr{X}}$ is a likelihood distribution in the Bayesian sense. For part (3), we apply Proposition 7.1 to disintegrate μ along the function $F : (\mathscr{X} \times \mathscr{E}, \mathscr{B}(\mathscr{X} \times \mathscr{E})) \to (\mathscr{M}_2^{\mathscr{E}}, \mathscr{B}(\mathscr{M}_2^{\mathscr{E}}))$ and, then, pushforward the resulting disintegration $\{\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ with respect to π_1 to obtain the Bayes posterior $\{\pi_{1*}\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$. The Bayes posterior is a family of regular conditional distributions $\{\mu_{\mathbf{x}|\mathbf{y}=y}\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ that is uniquely defined (up to $\mu_{\mathbf{y}}$-almost sure equivalence) by satisfying the condition

$$\mu_{\mathbf{x}}(A) = \int_{\mathscr{M}_2^{\mathscr{E}}} \mu_{\mathbf{x}|\mathbf{y}=y}(A)\, d\mu_{\mathbf{y}}(y), \tag{28}$$

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

for all $A \in \mathscr{B}(\mathscr{X})$. By Proposition 7.1, there exists a disintegration $\{\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ of μ along F and each measure μ^y is a probability measure that is concentrated on $F^{-1}(y) \subset \mathscr{X} \times \mathscr{E}$. We now prove that the family $\{\pi_{1*}\mu^y\}_{y \in \mathscr{M}_2^{\mathscr{E}}}$ satisfies the condition (28) of a regular conditional distribution. For $A \in \mathscr{B}(\mathscr{X})$, by Definition 3 (iii), we have

$$\mu_{\mathbf{x}}(A) = \big(\pi_{1*}\mu\big)(A) = \mu\big(\pi_1^{-1}(A)\big) = \int_{\mathscr{M}_2^{\mathscr{E}}} \mu^y\big(\pi_1^{-1}(A)\big)\, dF_*\mu(y) = \int_{\mathscr{M}_2^{\mathscr{E}}} \pi_{1*}\mu^y(A)\, d\mu_{\mathbf{y}}(y).$$

Therefore, the defining condition for the posterior is verified. Since $(\mathscr{Y}, d_{\mathscr{Y}})$ is a Polish space, the regular conditional distribution $\mu_{\mathbf{x}|\mathbf{y}=\cdot}$ exists and is unique up to $\mu_{\mathbf{y}}$-almost sure equivalence by [48, Theorem 10.2.2]. We conclude that $\pi_{1*}\mu^y = \mu_{\mathbf{x}|\mathbf{y}=y}$ for $\mu_{\mathbf{y}}$-a.e. $y \in \mathscr{M}_2^{\mathscr{E}}$.

Now for part (4), we apply Theorem 3.2 part (ii). As the integrand does not depend on e, the optimal map can be rewritten depending on the posterior distribution from part (3)

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} \int_{F^{-1}(y)} d_{\mathscr{X}}(x, z)^p \, d\mu^y(x, e), \tag{29}$$

$$= \operatorname*{argmin}_{z \in \mathscr{X}} \int_{\pi_1(F^{-1}(y))} d_{\mathscr{X}}(x, z)^p \, d(\pi_{1*}\mu^y)(x) \tag{30}$$

Recall that $\pi_{1*}\mu^y = \mu_{\mathbf{x}|\mathbf{y}=y}$. As $d_{\mathscr{X}}$ is induced by an inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$ induced by the inner product, for $p = 2$, and the measure $\pi_{1*}\mu^y$ is supported on $\pi_1(F^{-1}(y))$, the optimal map reduces to

$$\Psi(y) = \operatorname*{argmin}_{z \in \mathscr{X}} \int_{\mathscr{X}} \|x - z\|^2 \, d\mu_{\mathbf{x}|\mathbf{y}=y}(x) = \arg\min_{z \in \mathscr{X}} \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - z\|^2\big],$$

which is the Minimum Mean Squared Error (MMSE) for the posterior distribution $\pi_{1*}\mu^y = \mu_{\mathbf{x}|\mathbf{y}=y}$. Using the bilinearity of the inner product and the linearity of expectation, the objective function to be minimised in the expression for the optimal map Ψ can be rewritten as:

$$\begin{aligned}
\mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - z\|^2\big] &= \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] + \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\|^2\big] \\
&= \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]\|^2\big] + \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\|^2\big] \\
&\quad + 2\mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big\langle \mathbf{x} - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}], \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\big\rangle \\
&= \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]\|^2\big] + \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\|^2\big] \\
&\quad + 2\big\langle \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}], \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\big\rangle \\
&= \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbf{x} - \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]\|^2\big] + \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}\big[\|\mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}] - z\|^2\big]
\end{aligned}$$

which is the sum of two non-negative quantities and is thus minimised when $z = \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]$. Thus, the optimal map is given by $\Psi(y) = \mathbb{E}_{\mu_{\mathbf{x}|\mathbf{y}=y}}[\mathbf{x}]$ and coincides with the posterior mean. □

Funding statement. The work of the first author was supported by the UK EPSRC Centre for Doctoral Training. The work of the last author was supported by the Royal Society University Research Fellowship and the Leverhulme Prize 2017. The first author also acknowledges research sponsored by the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. DOE.

Competing interests. The authors declare no competing interests.

References

[1] Adcock, B. & Hansen, A. C. (2021) *Compressive Imaging: Structure, Sampling, Learning*, Cambridge University Press.
[2] Adcock, B., Brugiapaglia, S. & Webster, C. G. (2022) *Sparse Polynomial Approximation of High-Dimensional Functions*, Vol. 25, SIAM.
[3] Adcock, B. & Dexter, N. (2021) The gap between theory and practice in function approximation with deep neural networks. *SIAM J. Math. Data Sci.* **3** (2), 624–655.
[4] Adcock, B., Hansen, A. C., Poon, C. & Roman, B. (2017) Breaking the coherence barrier: A new theory for compressed sensing. *Forum Math. Sigma*. **5**, e4. Cambridge University Press.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

[5] Adler, J. & Öktem, O. (2017) Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl.* **33** (12), 124007.
[6] Aliprantis, C. D. & Border, K. C. (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer.
[7] Amir, D. & Mach, J. (1984) Chebyshev centers in normed spaces. *J. Approx. Theory* **40** (4), 364–374.
[8] Amir, D., Mach, J. & Saatkamp, K. (1982) Existence of Chebyshev centers, best n-nets and best compact approximants. *Trans. Amer. Math. Soc.* **271** (2), 513–524.
[9] Antun, V., Colbrook, M. J. & Hansen, A. C. (2022) Proving existence is not enough: Mathematical paradoxes unravel the limits of neural networks in artificial intelligence. *SIAM News* **55** (04), 1–4.
[10] Antun, V., Renna, F., Poon, C., Adcock, B. & Hansen, A. C. (2020) On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc. Natl. Acad. Sci. USA* **117** (48), 30088–30095.
[11] Arestov, V. (1986) Optimal recovery of operators and related problems. Collection of Papers from the All-Union School on the Theory of Functions, pp. 3–20.
[12] Arridge, S., Maass, P., Öktem, O. & Schönlieb, C.-B. (2019) Solving inverse problems using data-driven models. *Acta Numer.* **28**, 1–174.
[13] Aubert, G. & Aujol, J.-F. (2008) A variational approach to removing multiplicative noise. *SIAM J. Appl. Math.* **68** (4), 925–946.
[14] Aubin, J.-P., Frankowska, H., et al. (2009) Set-valued analysis. *Modern Birkhäuser Classics*, Birkhäuser Boston.
[15] Baraniuk, R. G. & Wakin, M. B. (2009) Random projections of smooth manifolds. *Found. Comput. Math.* **9** (1), 51–77
[16] Barrett, H. H. & Myers, K. J. (2013) *Foundations of Image Science*, John Wiley & Sons.
[17] Bastounis, A., Hansen, A. C. & Vlačić, V. (2021) The extended Smale's 9th problem – On computational barriers and paradoxes in estimation, regularisation, computer-assisted proofs and learning. arXiv:2110.15734.
[18] Bastounis, A. & Hansen, A. C. (2017) On the absence of uniform recovery in many real-world applications of compressed sensing and the restricted isometry property and nullspace property in levels. *SIAM J. Imaging Sci.* **10** (1), 335–371.
[19] Belthangady, C. & Royer, L. A. (2019) Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16** (12), 1215–1225.
[20] Ben-Artzi, J., Colbrook, M. J., Hansen, A. C., Nevanlinna, O. & Seidel, M. (2020) Computing spectra – On the solvability complexity index hierarchy and towers of algorithms. arXiv:1508.03280v5.
[21] Ben-Artzi, J., Marletta, M. & Rösler, F. (2023) Computing scattering resonances. *J. Eur. Math. Soc*. **25** (9), 3633–3663.
[22] Benning, M. & Burger, M. (2018) Modern regularization methods for inverse problems. *Acta Numer*. **27**, 1–111.
[23] Bertero, M., Boccacci, P. & De Mol, C. (2021) *Introduction to Inverse Problems in Imaging*, CRC press.
[24] Bhadra, S., Kelkar, V. A., Brooks, F. J. & Anastasio, M. A. (2021) On hallucinations in tomographic image reconstruction. *IEEE Trans. Med. Imaging* **40**, 3249–3260.
[25] Binev, P., Bonito, A., DeVore, R. & Petrova, G. (2022) Optimal learning. arXiv preprint arXiv:2203.15994.
[26] Blumensath, T. (2011) Sampling and reconstructing signals From a union of linear subspaces. *IEEE Trans. Inform. Theory* **57**, 4660–4671.
[27] Bogachev, V. I. (2007) *Measure Theory*, Vol. 1, Berlin Heidelberg, Springer.
[28] Borcea, L., Garnier, J., Mamonov, A. V. & Zimmerling, J. (2023) Waveform inversion with a data driven estimate of the internal wave. *SIAM. J. Imaging Sci*. **16**, 280–312.
[29] Boufounos, P. T. & Baraniuk, R. G. (2008) 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pp. 16–21, IEEE.
[30] Bouman, C. A. (2022) *Foundations of Computational Imaging: A Model-Based Approach*, SIAM.
[31] Bourrier, A., Davies, M. E., Peleg, T., Pérez, P. & Gribonval, R. (2014) Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems. *IEEE Trans. Inform. Theory* **60**, 7928–7946.
[32] Brunet, D., Vrscay, E. R. & Wang, Z. (2011) On the mathematical properties of the structural similarity index. *IEEE Trans. Image Process* **21**, 1488–1499.
[33] Burger, M. & Roith, T. (2024) Learning in image reconstruction: A cautionary tale. *SIAM News*, *Collections* **57** (08).
[34] Candes, E. & Recht, B. (2009) Exact matrix completion via convex optimization. *Found. Comput. Math*. **9**, 717–772.
[35] Candes, E. J. (2008) The restricted isometry property and its implications for compressed sensing. *C. R. Math*. **346**, 589–592.
[36] Candes, E. J., Strohmer, T. & Voroninski, V. (2013) Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pur. Appl. Math*. **66**, 1241–1274.
[37] Candès, E. J. & Plan, Y. (2011) Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inform. Theory* **57**, 2342–2359.
[38] Carlini, N. & Wagner, D. (2018) Audio adversarial examples: Targeted attacks on speech-to-text. In *2018 IEEE Security and Privacy Workshop*, pp. 1–7, IEEE.
[39] Carlini, N., Mishra, P., Vaidya, T., Zhang, Y., Sherr, M., Shields, C., Wagner, D. & Zhou, W. (2016) Hidden voice commands. In *25th USENIX Security Symposium*, pp. 513–530, USENIX security 16.
[40] Chang, J. T. & Pollard, D. (1997) Conditioning as disintegration. *Stat. Neerl*. **51**, 287–317.
[41] Cohen, A., Dahmen, W. & DeVore, R. (2009) Compressed sensing and best k-term approximation. *J. Am. Math. Soc*. **22**, 211–231.
[42] Cohen, A., Devore, R., Petrova, G. & Wojtaszczyk, P. (2022) Optimal stable nonlinear approximation. *Found. Comput. Math*. **22**, 607–648.
[43] Colbrook, M. & Hansen, A. C. (2022) The foundations of spectral computations via the solvability complexity index hierarchy. *J. Eur. Math. Soc*. **25**(12), 4639–4712.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

[44] Colbrook, M. (2024) On the computation of geometric features of spectra of linear operators on hilbert spaces. *Found. Comput. Math*. **24** (3), 723–804.
[45] Colbrook, M. J., Antun, V. & Hansen, A. C. (2022) The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proc. Natl. Acad. Sci. USA* **119**, e2107151119.
[46] Daubechies, I., DeVore, R., Foucart, S., Hanin, B. & Petrova, G. (2022) Nonlinear approximation and (Deep) ReLU neural networks. *Constr. Approx*. **55**, 127–172.
[47] de Leeuw, K., Moore, E. F., Shannon, C. E. & Shapiro, N. (1956) Computability by probabilistic machines. In *Automata Studies, no. 34 Ann. of Math. Stud.*, Princeton, NJ, Princeton University Press, pp. 183–212.
[48] Dudley, R. M. (2018) *Real Analysis and Probability*, Chapman and Hall/CRC.
[49] Engl, H. W. & Groetsch, C. W. (2014) *Inverse and Ill-Posed Problems*, Vol. 4, Elsevier.
[50] Ettehad, M. & Foucart, S. (2021) Instances of computational optimal recovery: Dealing with observation errors. *SIAM/ASA J. Uncertain. Quantif*. **9**, 1438–1456.
[51] Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T. & Song, D. (2018) Robust physical-world attacks on deep learning visual classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1625–1634, IEEE.
[52] Fannjiang, A. & Strohmer, T. (2020) The numerics of phase retrieval. *Acta Numer*. **29**, 125–228.
[53] Fefferman C., Hansen A. C. & Jitomirskaya S. (2022) *Computational Mathematics in Computer Assisted Proofs*, American Institute of Mathematics Workshops. American Institute of Mathematics. https://aimath.org/pastworkshops/compproofsvrep.pdf.
[54] Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L. & Kohane, I. S. (2019) Adversarial attacks on medical machine learning. *Science* **363**, 1287–1289.
[55] Folland, G. B. (1999) *Real Analysis: Modern Techniques and Their Applications*, Vol. 40, John Wiley & Sons.
[56] Foucart, S. & Rauhut, H. (2013) *A Mathematical Introduction to Compressive Sensing*, Birkhäuser.
[57] Foucart, S. (2022) *Mathematical Pictures at a Data Science Exhibition*, Cambridge University Press.
[58] Foucart, S., Liao, C., Shahrampour, S. & Wang, Y. (2022) Learning from non-random data in Hilbert spaces: An optimal recovery perspective. *Sampl. Theory Signal Process. Data Anal*. **20**, 5.
[59] Garcia, R. & Crespon, F. (2008) Radio tomography of the ionosphere: Analysis of an underdetermined, ill-posed inverse problem, and regional application. *Radio Sci*. **43**, 1–13.
[60] Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R. & Zhu, X. X. (2023) A survey of uncertainty in deep neural networks. *Artif. Intell. Rev*. **56** (S1), 1–77.
[61] Gazdag, L. E. & Hansen, A. C. (2022) Generalised hardness of approximation and the SCI hierarchy – On determining the boundaries of training algorithms in AI. arXiv:2209.06715.
[62] Gottschling, N. M., Antun, V., Hansen, A. C. & Adcock, B. (2025) The troublesome kernel: On hallucinations, no free lunches, and the accuracy-stability tradeoff in inverse problems. *SIAM Rev*. **67**, 73–104.
[63] Hadamard, J. (1902) Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull*. **13**, 49–52.
[64] Hadamard, J. (2003) *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Courier Corporation.
[65] Hansen, A. C. (2011) On the Solvability Complexity Index, the n-pseudospectrum and approximations of spectra of operators. *J. Am. Math. Soc*. **24**, 81–124.
[66] Hoffman, D. P., Slavitt, I. & Fitzpatrick, C. A. (2021) The promise and peril of deep learning in microscopy. *Nat. Methods* **18**, 131–132.
[67] Hofinger, A. & Pikkarainen, H. K. (2009) Convergence rates for linear inverse problems in the presence of an additive normal noise. *Stoch. Anal. Appl*. **27**, 240–257.
[68] Huang, Y.-M., Ng, M. K. & Wen, Y.-W. (2009) A new total variation method for multiplicative noise removal. *SIAM. J. Imaging Sci*. **2**, 20–40.
[69] Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. (2017) Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process* **26**, 4509–4522.
[70] Kabanava, M., Kueng, R., Rauhut, H. & Terstiege, U. (2016) Stable low-rank matrix recovery via null space properties. *Inf. Inference J. IMA* **5**, 405–441.
[71] Kaipio, J. P., Kolehmainen, V., Vauhkonen, M. & Somersalo, E. (1999) Inverse problems with structural prior information. *Inverse Probl*. **15**, 713.
[72] Kallenberg, O. (2021) *Foundations of Modern Probability*, 3rd ed., Springer.
[73] Keriven, N. & Gribonval, R. (2018) Instance optimal decoding and the restricted isometry property. *J. Phys.: Conf. Ser.* **1131** (1), 012002.
[74] Kurakin, A., Goodfellow, I. & Bengio, S. (2017) Adversarial examples in the physical world. In *International Conference on Learning Representations*. 5th ICLR, Workshop Track.
[75] Lakshminarayanan, B., Pritzel, A. & Blundell, C. (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst*. **30**, 6405–6416.
[76] Leonov, A. S. (2012) A posteriori accuracy estimations of solutions to ill-posed inverse problems and extra-optimal regularizing algorithms for their solution. *Numer. Anal. Appl*. **5**, 68–83.
[77] Leonov, A. S. (2016) Locally extra-optimal regularizing algorithms and a posteriori estimates of the accuracy for ill-posed problems with discontinuous solutions. *Comp. Math. Math. Phys*. **56**, 1–13.

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527

[78] Liang, B., Li, H., Su, M., Bian, P., Li, X. & Shi, W. (2017) Deep text classification can be fooled. In *The 27th International Joint Conference on Artificial Intelligence*. arXiv:1704.08006.

[79] Liu, J., Anirudh, R., Thiagarajan, J.J., He, S., Mohan, K.A., Kamilov, U.S. & Kim, H. (2023) DOLCE: A model-based probabilistic diffusion framework for limited-angle CT reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10498–10508.

[80] Lucas, A., Iliadis, M., Molina, R. & Katsaggelos, A. K. (2018) Using deep neural networks for inverse problems in imaging: Beyond analytical methods. *IEEE Signal Proc. Mag*. **35**, 20–36.

[81] Magaril-Il'yaev, G. G. & Osipenko, K. Y. (1991) Optimal recovery of functionals based on inaccurate data. *Mat. Zametki* **50**, 85–93.

[82] McCann, M. T., Jin, K. H. & Unser, M. (2017) Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Proc. Mag*. **34**, 85–95.

[83] Micchelli, C. A. & Rivlin, T. J. (1977) A survey of optimal recovery. *Optim. Estim. Approx. Theory*, pp. 1–54, lenum Press.

[84] Moosavi-Dezfooli, S., Fawzi, A. & Frossard, P. (2016) DeepFool: A simple and accurate method to fool deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2574–2582, IEEE.

[85] Muckley, M.J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M. & Arberet, S. (2020) State-of-the-art machine learning MRI reconstruction in 2020: Results of the second fastMRI challenge. arXiv preprint arXiv:2012.06318.

[86] Nguyen, A., Yosinski, J. & Clune, J. (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE Conference on Computer Vision and Pattern Recognition 2015*, pp. 427–436, IEEE.

[87] Novak, E. & Ritter, K. (1989) A stochastic analog to Chebyshev centers and optimal average case algorithms. *J. Complex.* **5**, 60–79.

[88] O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems. *Stat. Sci.* **1** (4), 502–518.

[89] Pinkus, A. (2012) *N-Widths in Approximation Theory*, Vol. 7, Springer Science & Business Media.

[90] Plaskota, L. (1996) *Noisy Information and Computational Complexity*, Vol. 95, Cambridge University Press.

[91] Poon, C. (2015) On the role of total variation in compressed sensing. *SIAM. J. Imaging Sci*. **8**, 682–720.

[92] Poon, C. (2017) Structure dependent sampling in compressed sensing: Theoretical guarantees for tight frames. *Appl. Comput. Harmon. Anal*. **42**, 402–451.

[93] Rao, T. (2002) Chebyshev centres and centrable sets. *Proc. Am. Math. Soc*. **130**, 2593–2598.

[94] Scarlett, J., Heckel, R., Rodrigues, M. R., Hand, P. & Eldar, Y. C. (2022) Theoretical perspectives on deep learning methods in inverse problems. *IEEE J. Sel. Areas Inf. Theory* **3**, 433–453.

[95] Schervish, M. J. (2012) *Theory of Statistics*, Springer Science & Business Media.

[96] Shalev-Shwartz, S. & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.

[97] Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J. & Segev, M. (2015) Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Proc. Mag*. **32**, 87–109.

[98] Shi, J. & Osher, S. (2008) A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM. J. Imaging Sci*. **1**, 294–321.

[99] Steen, L. A., Seebach, J. A. & Steen, L. A. (1978) *Counterexamples in Topology*, Vol. 1, Springer.

[100] Strack, R. (2018) Imaging: AI transforms image reconstruction. *Nat. Methods* **15**, 309.

[101] Stuart, A. M. (2010) Inverse problems: A Bayesian perspective. *Acta Numer*. **19**, 451–559.

[102] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. & Fergus, R. (2014) Intriguing properties of neural networks. In *International Conference on Learning Representations*. arXiv:1312.6199.

[103] Tran, H. & Webster, C. (2019) A class of null space conditions for sparse recovery via nonconvex, non-separable minimizations. *Results Appl. Math*. **3**, 100011.

[104] Traonmilin, Y. & Gribonval, R. (2018) Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all. *Appl. Comput. Harmon. Anal*. **45**, 170–205.

[105] Traonmilin, Y., Gribonval, R. & Vaiter, S. (2021) A theory of optimal convex regularization for low-dimensional recovery.arXiv preprint arXiv:2112.03540.

[106] Unser, M. (2020) A unifying representer theorem for inverse problems and machine learning. *Found Comput. Math.* **21** (4), 941–960.

[107] Wang, Y., Leonov, A. S., Lukyanenko, D. V. & Yagola, A. G. (2020) General Tikhonov regularization with applications in geoscience. *CSIAM Trans. Appl. Math*. **1**, 53–85.

[108] Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004) Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612.

[109] Ward, J. (1974) Chebyshev centers in spaces of continuous functions. *Pac. J. Math*. **52**, 283–287.

[110] Zhang, G., Yan, C., Ji, X., Zhang, T., Zhang, T. & Xu, W. (2017) DolphinAttack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 103–117, ACM.

[111] Zhao, X.-L., Wang, F. & Ng, M. K. (2014) A new convex optimization model for multiplicative noise and blur removal. *SIAM. J. Imaging Sci.* **7**, 456–475.

[112] Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. (2018) Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487.

Cite this article: Gottschling N. M., Campodonico P., Antun V. and Hansen A. C. On the existence of optimal set-valued decoders and their accuracy bounds for ill-posed inverse problems. *European Journal of Applied Mathematics*, https://doi.org/10.1017/S0956792526100527

Downloaded from https://www.cambridge.org/core. IP address: 18.97.9.172, on 10 Nov 2026 at 03:21:46, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0956792526100527