Learning and Inverse Problems: from Theory to Solar Physics Applications

Sabrina Guastavino

Dipartimento di matematica (DIMA)

Università degli Studi di Genova

Supervisor

Federico Benvenuto

In partial fulfillment of the requirements for the degree of

Ph.D. Defense in Mathematics and Applications

Acknowledgements

First and foremost I would like to express my gratitude to my supervisor Federico Benvenuto. It has been an honor to be his first Ph.D. student. The joy and enthusiasm he has for his research was contagious and motivational for me. I want to thank him for many reasons: at first for his support and guidance throughout my Ph.D. course, for the time that he dedicated to me, for his useful scientific and human advice and finally for his constant encouragement.

I am grateful to Prof. Michele Piana to encourage me to be a member of the MIDA group, to give me the opportunity together with my supervisor to work in very interesting areas. I would like to thanks again them and Anna Maria Massone to introduce me to the RHESSI Community and to offer me the opportunity to spend a period at NASA Goddard space Flight center.

I would thank Richard Schwartz for his hospitality at Goddard and for his useful scientific advice. I thank all those I met at GSFC for their kindness making my abroad period comfortable.

Many thanks to all my colleagues in these three years who have worked in MIDA group for making my Ph.D. period so enjoyable. The group has been a source of friendships as well as good advice and collaboration.

I would like to thank all my friends who encouraged and supported me in all this period.

A special thanks to my family to be always there for me. Words can not express how grateful I am to my mother Rosella and my

father Franco for all of the sacrifices that they have made on my behalf. Thanks to my grandmother Giusy for her affection to me. Thanks to my sister Simona for supporting me in everything, for trusting in me and for her neverending love. Thanks to my little darling niece Agnese.

Last but not the least, I would thank my lovely boyfriend Mattia for his support and for always believing in me by showing every day how proud he is of me!

Contents

List of Figures

List of Tables

Introduction

The problem of approximating a function from a set of discrete measurements has been extensively studied since the seventies [13; 14; 43; 98; 136; 143] and it is so common in applications that it is a cross-cutting theme in many areas including inverse problems and machine learning. Further, this problem has been considered by the mathematical literature under many different formal assumptions, ranging from direct to indirect measurements, from deterministic to statistical noise hypotheses, from known to unknown statistical noise distributions. This Ph.D. Thesis proposes a theoretical analysis of the problem of function approximation, first within a completely general setting and then focusing on a specific class of problems where measurements are distributed according to a Poisson law. Finally, as far as applications are concerned, in this Thesis we consider two problems in solar physics, i.e. a forecasting problem, where the aim is the prediction of solar storms using images of the magnetic field on the sun, and an image reconstruction problem for solar flares based on inverse diffraction and background estimation.

Contribution and related works

Our theoretical analysis proposes a formalization of the function approximation problem which allows dealing with inverse problems and supervised kernel learning as two sides of the same coin. The proposed formalization takes into account arbitrary noisy data (deterministically or statistically defined), arbitrary loss functions (possibly seen as a log-likelihood), handling

both direct and indirect measurements. The core idea of this part relies on the analogy between statistical learning and inverse problems. One of the main evidences of the connection occurring across these two areas is that regularization methods, usually developed for ill-posed inverse problems, can be used for solving learning problems. In particular, when a kernel is given and the loss function is the squared loss, spectral regularization methods have been used [84; 148]. Furthermore, spectral regularization convergence rate analyses provided in these two areas, share the same source conditions but are carried out with either increasing number of samples in learning theory or decreasing noise level in inverse problems. Even more in general, regularization via sparsity-enhancing methods is widely used in both areas and it is possible to apply well-known ℓ_1-penalized methods for solving both learning and inverse problems (see e.g. [62; 68; 131]). Therefore, the fact that learning and inverse problems can be solved using analogous regularization methods, the sparsity concept can be applied to both problems with different purposes, and similar convergence rate analyses are provided in the literature, beg the question of *to what extent these two problems are similar* and which are the key points of the connection. In this work, we analyze such a connection at three levels: (1) at an infinite dimensional level, we define an abstract function approximation problem from which the two problems can be derived; (2) at a discrete level, we provide a unified formulation according to a suitable definition of sampling; and (3) at a convergence rates level, we provide a comparison between convergence rates given in the two areas, by quantifying the relation between the noise level and the number of samples.

In the second part of this Thesis, we focus on a specific class of problems where measurements are distributed according to a Poisson law. In this case the loss function is the Kullback Leiber (KL) divergence [129]. As this divergence is not Lipschitz continuous, regularization methods in this case usually require proximal calculus theory and the resulting algorithms need much more computational time with respect to the square loss case. In this part we provide a data-driven, asymptotically unbiased, and globally quadratic

approximation of the Kullback-Leibler divergence. This approximation is inspired by the results in [152] and, roughly speaking, it holds for large count amounts. The main advantage is the possibility to treat the Maximum Likelihood estimation problem for Poisson data as a data-driven reweighted ℓ_2 norm minimization problem. Such a global quadratic approximation of the KL divergence leads to define a new method for solving sparse Poisson regression problems, named PRiL for Poisson Reweighed Lasso, which works as a Lasso-type method [68; 134] with the same computational cost of a standard minimization of an ℓ_2 data fitting term plus an ℓ_1 regularization term. By analyzing the statistical properties of this new method we prove that it is a consistent estimator. Moreover, we propose an adaptive version of this method, named APRiL (Adaptive Poisson Reweighted Lasso), by following the theory of adaptive ℓ_1 methods in [156] and we prove that this adaptive version performs variable selection in a consistent manner. We also assess in applications the theoretical properties of these methods evaluating their performances on both (synthetic) learning and inverse problems.

In the third part of the Thesis, we apply these sparsity-enhancing methods to two problems in solar physics: the problem of forecasting solar flares (learning application) and the desaturation problem of solar flare images (inverse problem application). Solar flares are the most explosive phenomena in the heliosphere, releasing a huge amount of electromagnetic radiation at all wavelengths and, in this way, triggering the whole space weather connection. The full comprehension of solar flare physics is still an open issue. Solar flares originate from magnetically active regions on the Sun. However, not all active regions give rise to solar flares and the nature of the prediction is intrinsically probabilistic.

For the first problem of interest, we apply the ℓ_1-penalized method proposed in this Thesis to predict if an active region originates solar flares. The challenge of solar flare prediction benefits by an intelligent computational analysis of physics-based features extracted from active regions from data provided by Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics

Observatory (SDO). The training phase of the algorithm is therefore based on an historical dataset of magnetic features of active regions labeled with some information about flare occurrences. The central goal of this application is to exploit the above mentioned ℓ_1-penalized algorithm to predict occurrences of solar flares on the basis of how many flares are originated by an active region: in this case the labels are reasonably Poisson distributed. This represents an advantage with respect to using labels with unknown distribution. The use of a sparsity-enhancing method is not only devoted to solar flares prediction but it also permits to identify the most predictive features. Relevant features are most likely associated to crucial physical processes and the knowledge of these features has hardware implications: instruments that accurately observe the most predictive features are probably more worthwhile designing.

The second application concerns the restoration problem of Extreme Ultra-Violet (EUV) solar flare images recorded by a second instrument on board SDO, the Atmospheric Imaging Assembly (AIA). SDO/AIA is probably the most powerful instrument for EUV solar imaging ever conceived, opening new crucial windows on the comprehension of how the solar magnetic fields release the huge amount of energy they store. This telescope has an unprecedented spatial resolution observing a 41 arcmin field of view in ten EUV and UV channels, with 0.6-arcsec pixels and 4096×4096 Charged Coupled Devices (CCDs) [8o]. Such a spatial resolution requires very small pixels, which are more likely affected by saturation effects with increasing incoming photon flux. Saturation includes two phenomena: primary saturation refers to the fact that, for intense incoming flux, CCD pixels lose their ability to accommodate additional charge; blooming, or secondary saturation, names the fact that primary saturation causes charge to spill into their neighbors. The resulting overall artifact appears as a bright region in the image surrounded by diffraction fringes and this phenomenon usually happens when intense solar flares occur and it makes such images unusable for scientific purposes. Image saturation has been an issue for several instruments in solar astronomy, mainly at EUV wavelengths. However, with the launch of AIA, image saturation has become

a big data issue, involving around 10^5 frames of the impressive dataset the telescope has been providing every year since February 2010. In this part of the work, we developed a new method for desaturating AIA images, called Sparsity-Enhancing DESAT (SE-DESAT), based on the ℓ_1-penalized method proposed in this Thesis. Sparsity in this context is conceived in the pixel space: this is for the fact that the saturated region of an image has a relatively small support with respect to the entire image. By promoting sparsity of the solution on the pixel space the proposed method performs segmentation and reconstruction of the saturated region simultaneously. Such a feature, together with the capability to estimate the background, enables the proposed method to desaturate several consecutively deteriorated frames recorded during long-lasting intense solar storms, like the one occurred on September 10, 2017. This makes this method superior to the existing one, developed in [149] and called DESAT, which has the limitation to need an a priori estimate of the background, usually not available when strong saturation effects occur for a whole time series of images.

Plan of the Thesis

In the following we report a summary of each chapter of the Thesis.

In Chapter 2 we analyze the connection between learning and inverse problems at an infinite dimensional level. An inverse problem can be formulated as a function approximation problem given an operator A and a data y. On the other hand, a learning problem consists in finding a function which explains the input-output relation from a given set of samples. Both these problems can be subsumed under an abstract common approximation problem. In particular the key instrument to define such a common approximation problem is the notion of Reproducing Kernel Hilber Space (RKHS). Indeed, the hypothesis space in learning problems is a RKHS and at the same time the range of a bounded operator is provided with a RKHS structure in a natural way.

In Chapter 3 we analyze the connection at a discrete dimensional level

providing a unified formulation of the two problems. Once the common RKHS approximation problem is defined we build a sampling operator which allows us to derive from such an infinite dimensional problem different discrete problems such as learning and discrete inverse problems. The peculiarity of this sampling operator is that it can take into account different natures of samples: in learning problems, data are usually assumed to be given as the result of a *stochastic process* whose underlying distribution is *unknown*, whereas in discrete inverse problems, data are assumed to be given according to a *deterministic scheme*, at least for the independent variables and even when the dependent variables are assumed to be drawn in a stochastic manner, the underlying distribution is supposed to be *known*. Finally, we discuss the conditions for the convergence of the discrete problem formulation (being it either deterministic or stochastic) to the infinite dimensional one.

In Chapter 4 we analyze the convergence rates of spectral regularization providing a comparison between the ones computed with respect to the number of samples and the ones computed with respect to the noise level in order to quantify the differences. In the literature, regularization methods have been studied in these two contexts providing error convergence rates under the same Holder-type source condition: in the context of ill-posed inverse problems, convergence rates for spectral regularization depending on the noise level δ have been known for years [2; 43]; in the context of learning results on optimal convergence rates depending on the number of samples n are more recent [9; 17; 30; 81; 102; 114; 127; 144; 148]. The question that naturally arises is whether the above rates are comparable and, if it is the case, which relation occurs between δ and n for quantifying the difference between optimal rates in the two contexts. We provide a comparison defining the relation between δ and n and considering an hybrid estimator [87; 139] which allows us to compare the rates given in the two settings.

In Chapter 5 we provide sparsity-enhancing methods for Poisson data to use in both learning and inverse problems applications. Lasso-type methods (i.e. ℓ_1-penalized methods) are widely used both in inverse problems as image

reconstruction problems, where usually images are compressed in suitable basis where few coefficients are non zero; and in learning applications, where the most predictive variables have to be selected. Poisson noise is common in both fields especially when data represents counts. The fidelity term characterizing the ℓ_1-penalized methods with Poisson data is usually represented by the KL divergence. Therefore, first we provide an asymptotically unbiased globally quadratic approximation of KL, which leads to the definition of new ℓ_1-penalized methods, called PRiL and APRiL. These novel methods can take advantage from the fast algorithms developed for those ℓ_1-penalized methods that have the least square functional as the fidelity term. We prove theoretical consistent properties of these estimators and we show their effectiveness on both learning and image reconstruction experiments.

In Chapter 6 we introduce the problem of solar flares forecasting. The prediction of solar flares is one of the key questions of heliophysics since solar flares are the primary drivers of space weather. Many observational and machine learning studies confirmed the important role that magnetic field properties in active regions play for the prediction of solar flares. In particular, we addressed the issue of both predicting the occurrence of solar flares and identifying the most predictive features using Lasso-type methods as PRiL and APRiL. Since these are supervised learning algorithms they need a training phase. We trained the methods on datasets where the labeling is not only the occurrence of solar flares but also other tasks such as the number of the originated flares which are reasonably assumed to be affected by Poisson noise.

Chapter 7 is devoted to an image reconstruction problem and presents a new method for restoring solar images affected by saturation and diffraction effects. We propose a formalization of the saturation process which takes into account both primary saturation and blooming effects and we formulate the desaturation problem as a linear inverse problem between Hilbert spaces in which the forward operator encodes both the diffraction effects of light rays and the conservation of the photon-induced charge in the CCD. The fact that the diffraction effects visible over the background solar activity come from a subset

of saturated pixels (i.e. the primary saturated ones) is translated in a sparsity constraint in the pixel domain. Moreover, the solar activity background is estimated iteratively by means of an alternating minimization algorithm. We compare the results of our algorithm with the ones of the existing DESAT method [137], showing the performances of the two approaches for both synthetic data and strongly saturated real observations. Furthermore, we apply the new method to desaturate images related to the solar storm on September 10, 2017: in this case DESAT cannot be used for the lack of a reliable estimate of the background. The new method works without any need of a priori information on the image background and therefore can be applied even for desaturation of several consecutively deteriorated frames recorded during long-lasting intense solar storms. This peculiar methodological property could make this algorithm a possible tool for the realization of an automatic pipeline for the processing of the whole AIA data archive.

List of publications

- Peer reviewed journals
  - *A consistent and numerically efficient variable selection method for sparse Poisson regression with applications to learning and signal recovery*,
    Guastavino S. & Benvenuto F.,
    Statistics and Computing 29 (2019).
  - *Flare forecasting and feature ranking using SDO/HMI data*,
    Piana M., Campi C. Benvenuto F., Guastavino S., Massone A.M.,
    IL NUOVO CIMENTO 42 C (2019).
  - *Desaturating SDO/AIA observations of solar flaring storms*,
    Guastavino S., Piana M., Massone A.M., Schwartz R., Benvenuto F.,
    The Astrophysical Journal 882 (2019).
- Pre print

  - *On the connection between supervised learning and linear inverse problems*,
    Guastavino S. & Benvenuto F.,
    arXiv:1807.11406 (2018).

- Submitted

  - *Convergence rates of spectral regularization methods: a comparison between ill-posed inverse problems and statistical kernel learning*,
    Guastavino S. & Benvenuto F.,
    submitted to SIAM Journal on Numerical Analysis.
  - *Restoration of solar images affected by diffraction and saturation effects via sparsity in the pixel space*,
    Guastavino S. & Benvenuto F.,
    submitted to Inverse Problems.

Chapter 1

Learning and inverse problems from a function approximation point of view

Inverse problems are typically ill-posed in the sense of Hadamard [64] and the regularization theory has been developed in order to provide a family of approximated solutions of an inverse problem. Formally, the goal of solving a linear inverse problem is to recover a function f such that

$$y = Af \tag{1.1}$$

given the data y and A a linear operator. For estimating f one can consider to have noisy infinite dimensional data, e.g. y^δ such that $\|y^\delta - y\| \leq \delta$, or, more realistically, a finite set of noisy samples $\{y_1, \dots y_n\}$ taken at points $\{x_1, \dots x_n\}$.

On the other hand, in supervised learning we have a quantitative (or categorical) outcome which we want to predict from a set of features. We have a disposal a training set of data in which we know the outcome and the features: from the training set we want to build an estimator or a prediction model which will be able to predict the outcome when a new feature is given. In other words, the aim of supervised learning is to find a function g from a set of examples $\{(X_i, Y_i)\}_{i=1}^n$ randomly drawn from an unknown probability

distribution ρ, such that g has to explain the relationship between input and output, i.e.

$$Y_i \sim g(X_i) \tag{1.2}$$

for all $i = 1, \dots, n$ and $g(x)$ has to be a good estimate of the output when a new input x is given. Learning algorithm such as the regularized least square algorithm is used to avoid overfitting and infer stability in the solution in order to assure the generalization property [22; 36].

One of the main evidences of the connection occurring across supervised learning and inverse problems is the conceptual analogy between learning algorithms and regularization ones. In this Chapter, first we recall the main ingredients of linear inverse problems (section 1.1.1) and supervised learning (section 1.1.2) and we provide an overview about the main works concerning the connection between these two fields (section 1.1.3). After that, in section 1.2 we provide the connection at the infinite dimensional level from a function approximation point of view: the fact that the range of a bounded linear operator is provided with a Reproducing Kernel Hilbert Space (RKHS) structure in a natural way (see [78]) allows us to describe the two problems as the same approximation problem in functional spaces. Here, we define the approximation in a RKHS as an optimization problem. In particular, by introducing a non-linear generalization of the Moore-Penrose inverse, we prove that the solution of an approximation problem in a RKHS can always be associated with a solution of a certain inverse problem. Conversely, we prove that the set of solutions of a class of inverse problems corresponds to the solution of a certain approximation problem in a RKHS. This set is defined up to the action of the unitary group. The same relation applies between Tikhonov-type solutions of approximation and inverse problems.

1.1 Introduction to learning and inverse problems

The concept of inverse problems has to come after the definition of the direct problem, which has to be thought of as a mathematical model describing

a particular process. Inverse problems [13; 43; 58] consist in reconstructing causes from observed effects: they have wide application in many fields as in medical imaging, signal processing, geophysics and they are common in astrophysics since the quantities of interest cannot be observed directly.

On the other hand, statistical learning [32; 51; 140] has conquered a central role in many areas of sciences, finance and industry and, as the name suggests, it is based on a phase of learning, called also training, in order to be able to generalize and so to predict from new inputs.

In the following we formally introduce inverse problems and learning problems.

1.1.1 Linear inverse problems

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces and A be a bounded linear operator $A : \mathcal{H}_1 \to \mathcal{H}_2$.

Definition 1. *The inverse problem associated with the operator A consists in finding $f \in \mathcal{H}_1$ satisfying the equation*

$$Af = y \tag{1.3}$$

given $y \in \mathcal{H}_2$.

Usually the inverse problem is ill-posed in the sense of Hadhamard, which means that the solution could not exist, could not be unique or could not depend continuously on the data. Therefore, the problem is addressed by searching for the Moore-Penrose generalized solution, denoted as $f^\dagger$. Formally, we have the following definition

Definition 2. *Let $P_{\Im(A)} y \in \Im(A)$, where $\Im(A)$ denotes the range of the operator A and $P_{\Im(A)}$ the projection on the closure of the range of A. Let $\mathfrak{M}_A$ be the set of the least-square solutions, i.e.*

$$\mathfrak{M}_A := \arg\min_{f \in \mathcal{H}_1} \|y - Af\|^2_{\mathcal{H}_2}. \tag{1.4}$$

The Moore-Penrose generalized solution $f^\dagger$ is defined as

$$f^\dagger := \arg\min_{f \in \mathfrak{M}_A} \|f\|_{\mathcal{H}_1}. \tag{1.5}$$

With the introduction of the Moore-Penrose generalized solution, the existence is restored by solving the least square problem $\min_{f \in \mathcal{H}_1} \|y - Af\|^2_{\mathcal{H}_2}$ (provided that the projection on the closure of the range of A of the data y belongs to the range of A) and the uniqueness is restored by taking the minimal norm solution of the least square problem. However, the generalized solution does not depend continuously on the data and this is a problem since only a noisy version y^δ of the data is available, where $\delta > 0$ represents the noise level. Such a problem is addressed by using some regularization methods such as Tikhonov-type regularization methods, which lead to minimize the following functional

$$\|y^\delta - Af\|^2_{\mathcal{H}_2} + \lambda \Omega(f), \tag{1.6}$$

where λ is the regularization parameter and $\Omega(f)$ is the penalty term needed to regularize the solution. In the case of the Tikhonov-type regularization we take the penalty term with the following form

$$\Omega(f) := \psi(\|f\|_{\mathcal{H}_1}), \tag{1.7}$$

where $\psi : [0, +\infty) \to \mathbb{R}_+$ is a continuous convex and strictly monotonically increasing real-valued function. The usual Tikhonov regularization is presented with the special choice $\psi(\|f\|_{\mathcal{H}_1}) = \|f\|^2_{\mathcal{H}_1}$. Therefore, the usual Tikhonov regularization leads to the following optimization problem

$$\min_{f \in \mathcal{H}_1} \|y^\delta - Af\|^2_{\mathcal{H}_2} + \lambda \|f\|^2_{\mathcal{H}_1}. \tag{1.8}$$

The regularization parameter λ has to create a trade off between the residual term (or fidelity term, i.e. the first term of the minimization functional in equation (1.6)) and the penalty term (the second term of the minimization functional in equation (1.6)). The parameter λ has to be chosen such that the

reconstruction error given by

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} \tag{1.9}$$

is small, where f_δ^λ represents the regularized solution. In detail, λ (which depends on the noise level δ and the data y^δ) is selected in such a way that the convergence of the regularized solution to the generalized solution holds, i.e.

$$\lim_{\delta \to 0} \|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} = 0, \tag{1.10}$$

for any data y^δ.

1.1.2 Learning from examples

The aim of supervised learning is to find a function g from a set of examples which are randomly drawn from a fixed but unknown probability distribution such that g explains the relationship between input and output and it satisfies the generalization property which means that it has to provide a good estimate of the output when a new input is given. Formally, we give the following definition.

Definition 3. *Let*

$$\mathcal{Z}_n := \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \tag{1.11}$$

be a finite set of samples, which are drawn independently identically distributed (i.i.d.) according to a given (but unknown) probability distribution ρ on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X} \subseteq \mathbb{R}^p$, with $p > 0$, and $\mathcal{Y} \subseteq \mathbb{R}$: $\mathcal{X}$ and $\mathcal{Y}$ represent the so-called input and output spaces, respectively. $\mathcal{X}$ and $\mathcal{Y}$ can be assumed to be compact spaces and ρ admits the following factorization

$$\rho(X, Y) = \rho(Y|X)\nu(X) \tag{1.12}$$

where ν is the marginal distribution on $\mathcal{X}$ and $\rho(\cdot|X = x)$ is the conditional distribution on $\mathcal{Y}$ for almost all $x \in \mathcal{X}$. Given the set of samples $\mathcal{Z}_n$, the aim is to find a function $\hat{g} : \mathcal{X} \to \mathbb{R}$, called estimator, such that $\hat{g}(X)$ is a good estimate of the output

when a new input X is given.

Given a measurable function g, the ability of g to describe the distribution ρ is measured by the expected risk defined as

$$R_\rho(g) = \int_{\mathcal{X}\times\mathcal{Y}} V(Y, g(X))\, d\rho(X, Y)\ , \tag{1.13}$$

where V is called loss function and $V(Y, g(X))$ measures the cost paid by replacing the true label Y with the estimate $g(X)$. A common choice of loss function is the square loss $V(Y, g(X)) = (Y - g(X))^2$. In such a case the regression function, defined as

$$g_\rho(X) = \int_{\mathcal{Y}} Y d\rho(Y|X) \tag{1.14}$$

is the minimizer of the expected risk in equation (1.13) (over all measurable functions), i.e. it can be seen as an ideal estimator of the unknown distribution ρ . However only the set $\mathcal{Z}_n$ is available and therefore learning is performed by minimizing over an hypothesis space $\mathcal{H}_K$ (which is usually a Reproducing Kernel Hilbert Space (RKHS) [29]) the empirical risk given by

$$R_{\mathcal{Z}_n}(g) = \frac{1}{n}\sum_{i=1}^{n} V(Y_i, g(X_i))\ . \tag{1.15}$$

Therefore, in the case of square loss the empirical risk minimizer is the least square estimator, defined as follows.

Definition 4. *Under the same hypothesis in Definition 3, the least square estimator $\hat{g}^\dagger$, is defined as follows*

$$\hat{g}^\dagger = \arg\min_{g\in\mathcal{H}_K} \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))^2. \tag{1.16}$$

From a numerical point of view the solution of the minimization problem (1.16) is not stable and therefore, following the approach of Tikhonov regularization, it is useful to introduce a penalty term in order to stabilize

the solution. Therefore, the regularized problem consists in minimizing the following penalized functional

$$R_{\mathcal{Z}_n}(g) + \lambda\Omega(g) \ , \tag{1.17}$$

where λ is the regularization parameter and Ω represents the penalty term. In the case of the Tikhonov-type regularization we take the penalty term with the following form

$$\Omega(g) := \psi(\|g\|_{\mathcal{H}_K}), \tag{1.18}$$

where $\psi : [0, +\infty) \to \mathbb{R}_+$ is a non-decreasing convex function. The usual Tikhonov regularization is presented with the special choice $\psi(\|g\|_{\mathcal{H}_K}) = \|g\|^2_{\mathcal{H}_K}$. The algorithm generated by using Tikhonov regularization in this context of learning theory is the well-known regularized least-square algorithm, which consists in the following optimization problem

$$\min_{g\in\mathcal{H}_K} \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))^2 + \lambda\|g\|^2_{\mathcal{H}_K}. \tag{1.19}$$

The regularization parameter λ has to yield a trade off between the fitting term (to avoid overfitting) and the stabilizer term. In detail, λ has to be chosen such that the learning algorithm is consistent, which means that the discrepancy, measured as

$$R_\rho(\hat{g}^\lambda_n) - \inf_{g\in\mathcal{H}_K} R_\rho(g), \tag{1.20}$$

is small in probability, where $\hat{g}^\lambda_n$ represents the estimator, that is the minimizer of the penalized functional in equation (1.17). In the case of square loss, simple computations show that

$$R_\rho(\hat{g}^\lambda_n) - \inf_{g\in\mathcal{H}_K} R_\rho(g) = \|\hat{g}^\lambda_n - P_{\mathcal{H}_K} g_\rho\|^2_{L^2(\mathcal{X},\nu)}, \tag{1.21}$$

where $P_{\mathcal{H}_K}$ denotes the projection on the closure of $\mathcal{H}_K$ in $L^2(\mathcal{X}, \nu)$. As we have seen that in inverse problems setting the convergence results are given in

terms of the reconstruction error, in this case the convergence results are given in terms of the prediction error $\|\hat{g}_n^\lambda - P_{\mathcal{H}_K} g_\rho\|^2_{L^2(\mathcal{X},\nu)}$. This error (which is a random variable since it depends on observations) is estimated in probability or in expectation: in detail, the regularization parameter λ (which depends on n and $\mathcal{Z}_n$) is selected in such a way that

$$\lim_{n\to\infty} \mathbb{E}_{\rho^{\otimes n}}(\|\hat{g}_n^\lambda - P_{\mathcal{H}_K} g_\rho\|^2_{L^2(\mathcal{X},\nu)}) = 0, \tag{1.22}$$

where $\rho^{\otimes n}$ indicates the distribution tensor product (see [17] and references therein).

1.1.3 Overview of the connection between learning and inverse problems

One of the main evidences of the connection occurring across learning and inverse problems is that regularization methods, such as Tikhonov regularization (as we have already mentioned in the previous sections) or spectral regularization (as we will see in details in Chapter 3), which have been developed in the inverse problems theory can be used for solving learning problems [84; 148]. More in general, a classical approach relies on the concept of variational regularization [11, 24]. It combines knowledge about how data is generated in the forward operator with a regularization functional that encodes prior knowledge about the solution to be reconstructed. As we will see in Chapter 4, when solutions admit a sparse representation it is possible to apply variational regularization, e.g. ℓ_1-penalized methods, for solving both learning and inverse problems [62]. On the other hand, a recent trend is to use neural networks, a common tool for learning problems, for solving inverse problems, in particular in imaging applications [1; 92].

In the literature several authors proposed to solve learning problems by using regularization techniques originally developed for inverse problems, offering a glimpse of the connection between supervised learning and inverse problems [33; 56; 79; 94; 126; 128; 144]. In recent years, a rigorous formalization

of this connection between supervised learning and linear inverse problems has been proposed according to two strategies: the first considers the learning problem as an instance of an inverse one (see e.g. [38; 84]) whereas the second introduces a bounded operator in the model equation of the statistical learning and it is known as inverse learning (see e.g. [17; 86; 114]). The first strategy interprets a learning problem as an inverse one in which the forward operator is an inclusion of the hypothesis space $\mathcal{H}_K$ into the Hilbert space of square integrable functions $L^2(\mathcal{X}, \nu)$. Its main objective is to draw a connection between consistency in kernel learning and regularization in inverse problems offering a full connection in the case of square loss. On the other hand, the second strategy considers inverse problems from a statistical estimation perspective highlighting the fact that statistical inverse problems can be thought of as learning problems starting from indirect data. In particular, in this case the observations are modeled as follows

$$Y_i = g_\rho(X_i) + \epsilon_i, \quad i = 1, \dots, n \quad \text{with} \quad g_\rho = A f_\rho \tag{1.23}$$

where A is a uniformly bounded operator and ϵ_i are independent centered noise variables. Furthermore, under appropriate probabilistic source conditions, error rates are provided for both the predictive error $\|A\hat{f}_n^\lambda - Af_\rho\|^2_{L^2(\mathcal{X},\nu)}$ and the estimation (or reconstruction) error $\|\hat{f}_n^\lambda - f_\rho\|^2_{\mathcal{H}_1}$. The latter one is studied in inverse problems theory, especially in the case that $\hat{f}_n^\lambda$ is a spectral regularized estimator. The operator A encodes the information contained in the *feature map* introduced in supervised kernel learning. Indeed, when A is uniformly bounded the range of A ia a RKHS [78]. This is the key point to interpret learning and inverse problems from a common function approximation perspective.

1.2 A common function approximation problem

In order to outline the connection at the infinite dimensional level two ingredients are necessary: the definition of a suitable function approximation problem in Reproducing Kernel Hilbert Spaces (RKHSs) and the definition of a non-linear generalization of the Moore-penrose solution in the context of inverse problems between Hilbert spaces. Once these two ingredients are provided we show the connection between approximation problems in RKHSs and classes of inverse problems (up to the action of the unitary group).

1.2.1 Approximation problem in RKHSs

RKHSs arise in a number of areas, including statistical machine learning theory, approximation theory, generalized spline theory and inverse problems [33; 75; 107]. The general theory of RKHSs was developed by [6]. The usual definition of a RKHS is given for a Hilbert space of functions, as follows:

Definition 5. *Let $\mathcal{H}$ be a Hilbert space of real valued functions on a non-empty set $\mathcal{X}$. $\mathcal{H}$ is said a reproducing kernel Hilbert space if for all $x \in \mathcal{X}$ the evaluation functional $L_x : f \in \mathcal{H} \to L_x(f) := f(x)$ is continuous.*

An important characterization of RKHSs, which can be even considered as an alternative definition, is the following:

Definition 6. *$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a reproducing kernel of a Hilbert space $\mathcal{H}$ if for all $f \in \mathcal{H}$, $f(x) = \langle f, K_x \rangle_{\mathcal{H}}$, where $K_x := K(x, \cdot) \in \mathcal{H}$, $\forall\, x \in \mathcal{X}$.*

We recall some well known facts. The kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a symmetric positive definite function, where positive definite means that for each set of points $\{x_i\}_{i=1}^n$ in $\mathcal{X}$ and set of real numbers $\{a_i\}_{i=1}^n$,

$$\sum_{i,j=1}^{n} a_i a_j K(x_i, x_j) \geq 0. \tag{1.24}$$

If K is continuous then each $g \in \mathcal{H}_K$ is continuous. If it is assumed further that

$$\kappa := \sup_{x \in \mathcal{X}} \sqrt{K(x,x)} < \infty, \tag{1.25}$$

then, for the reproducing property, for each $x \in \mathcal{X}$ and for each $g \in \mathcal{H}_K$

$$|g(x)| = |\langle g, K_x \rangle_{\mathcal{H}_K}| \leq \|g\|_{\mathcal{H}_K} \|K_x\|_{\mathcal{H}_K} = \|g\|_{\mathcal{H}_K} \sqrt{K(x,x)}, \tag{1.26}$$

which implies that

$$\|g\|_\infty \leq \kappa \|g\|_{\mathcal{H}_K}. \tag{1.27}$$

This means that convergence in $\|\cdot\|_{\mathcal{H}_K}$ implies the uniform convergence.

The definition of RKHS is not restricted to function spaces but allows us to consider reproducing kernels K defined on $\mathcal{X} \times \mathcal{X}$, where $\mathcal{X}$ is a Borel set. For function spaces $\mathcal{X}$ shall be $\mathbb{R}$ or $\mathbb{C}$, but in general it can be a countable set or a finite set [6] (e.g. a pixel space) . This perspective takes to see the reproducing kernel K as function of two variables (x, x'), which can be continuous variables, e.g. $x, x' \in \mathbb{R}$, or can be represented by indexes (i, j), e.g. countable variables $i, j \in \mathbb{N}$ or finite discrete variables $i, j \in \{1, \ldots, n\}$. In the latter case, the kernel K is an infinite or finite matrix.

We now define the approximation problem in a RKHS as the problem of finding the closest element of the RKHS to a given one.

Definition 7. *Let y be the element to approximate in a given Hilbert space $\mathcal{H}_2$ and let $\mathcal{H}_K \subseteq \mathcal{H}_2$ be a RKHS with reproducing kernel K. We define the solution of the approximation problem as the minimizer of a functional $R_y : \mathcal{H}_2 \to \mathbb{R}$ over the RKHS $\mathcal{H}_K$, i.e.*

$$g_{R_y} := \arg\min_{g \in \mathcal{H}_K} R_y(g). \tag{1.28}$$

The idea is that $R_y(g)$ measures the discrepancy between y and g. We require that $R_y(g) \geq 0$ for all $g \in \mathcal{H}_2$, and $R_y(g) = 0$ iff $g = y$. Under these hypotheses, if $y \in \mathcal{H}_K$ the existence and uniqueness are assured by requiring that R_y is strictly convex. Otherwise, if $y \notin \mathcal{H}_K$ the existence and uniqueness are assured either by requiring that

a) R_y is lower semicontinuous, strictly convex and coercive with respect to the norm $\|\cdot\|_{\mathcal{H}_2}$ and $\mathcal{H}_K \subseteq \mathcal{H}_2$ is closed, or

b) R_y is lower semicontinuous, strictly convex and coercive with respect to the norm $\|\cdot\|_{\mathcal{H}_K}$.

A typical example is $R_y(g) = \|y - g\|^2_{\mathcal{H}_2}$ with $\mathcal{H}_K$ closed in $\mathcal{H}_2$.

1.2.2 Non-linear generalization of the Moore-Penrose solution

We consider the setting of inverse problems described in section 1.1.1. Therefore, given a data $y \in \mathcal{H}_2$ and a bounded linear operator $A : \mathcal{H}_1 \to \mathcal{H}_2$ the aim is to find a function f such that the equation (1.3) is satisfied. We recall that the ill-posedness of inverse problems leads to the definition of the generalized solution, denoted by $f^\dagger$, which can be seen, from a variational point of view, as the minimal norm solution of the least squares problem (see Definition 2). This variational form can be generalized by minimizing the functional R_y as follows

$$\mathfrak{M}_{A,R_y} := \arg\min_{f \in \mathcal{H}_1} R_y(Af) \tag{1.29}$$

and take the minimum norm solution. Hereafter, we refer to $\mathfrak{M}_{A,R_y}$ as the set of the R_y-minimum solutions. When at least an R_y-minimum solution f_{R_y} exists, $\mathfrak{M}_{A,R_y}$ is the affine subspace given by $f_{R_y} + Ker(A)$, where $Ker(A)$ denotes the nullspace of A.

Definition 8. *Consider the inverse problem in equation* (1.3). $f^\dagger_{R_y} \in \mathcal{H}_1$ *is called the* R_y*-generalized solution of the inverse problem* (1.3) *if it is the* R_y*-minimum solution (see equation* (1.29)*) with minimum norm, i.e.*

$$f^\dagger_{R_y} = \arg\min_{f \in \mathfrak{M}_{A,R_y}} \|f\|_{\mathcal{H}_1}. \tag{1.30}$$

As in section 1.2.1 we require that $R_y(g) \geq 0$ for all $g \in \mathcal{H}_2$, and $R_y(g) = 0$ iff $g = y$. We discuss some hypotheses which assure the existence and uniqueness of the R_y-generalized solution. Under these hypotheses, if $y \in$

$\Im(A)$ the existence and uniqueness are assured by requiring that R_y is strictly convex. Otherwise, if $y \notin \Im(A)$ the existence and uniqueness are assured either by requiring that

a) R_y is lower semicontinuous, strictly convex and coercive with respect to the norm $\|\cdot\|_{\mathcal{H}_2}$ and $\Im(A) \subseteq \mathcal{H}_2$ is closed, or

b) $f \in \mathcal{H}_1 \mapsto R_y(Af)$ is lower semicontinuous, strictly convex and coercive with respect to the norm $\|\cdot\|_{\mathcal{H}_1}$.

When R_y is different from the least squares functional, this procedure provides a generalization of the so-called Moore-Penrose generalized solution. Such a generalization is needed to develop the equivalence between approximation problems in RKHSs and classes of linear inverse problems. We introduce it in the next paragraph.

1.2.3 Approximation in RKHSs and inverse problems

We show the equivalence between an approximation problem in a RKHS and an inverse problem by proving that there is a natural correspondence of the solutions of the two problems. We make use of the following:

Assumption 1. *Let $\mathcal{H}_1$ be a real separable Hilbert space and $\mathcal{H}_2$ be a real Hilbert space on a Borel space $\mathcal{X}$. For all $x \in \mathcal{X}$ and for all $f \in \mathcal{H}_1$ there exists a constant $c > 0$ such that*

$$|Af(x)| \leq c\|f\|_{\mathcal{H}_1}. \tag{1.31}$$

The assumption 1 together with the Riesz's representation theorem implies that for all x there exists an element $\phi_x \in \mathcal{H}_1$ such that

$$(Af)(x) = \langle f, \phi_x \rangle_{\mathcal{H}_1} \tag{1.32}$$

and

$$\|\phi_x\|_{\mathcal{H}_1} = \|A_x\|_{\mathcal{H}_1^*} \leq c, \tag{1.33}$$

where $A_x : \mathcal{H}_1 \to \mathbb{R}$ has to be intended as $A_x(f) = (Af)(x)$ for each $f \in \mathcal{H}_1$ and $\|\cdot\|_{\mathcal{H}_1^*}$ represents the norm in the dual space $\mathcal{H}_1^*$, i.e. $\|A_x\|_{\mathcal{H}_1^*} := \sup_{\|f\|_{\mathcal{H}_1} \leq 1} |Af(x)|$. Moreover, it is well known that the range of the operator A is a RKHS (e.g. see [6; 78; 132]). The following proposition is an adaptation of this result to our context.

Proposition 1. *$\Im(A)$ equipped with the norm*

$$\|g\|_{\mathcal{H}_K} = \min\{\|w\|_{\mathcal{H}_1} : w \in \mathcal{H}_1 \; s.t \; g(x) = \langle w, \phi_x \rangle_{\mathcal{H}_1}, \; x \in \mathcal{X}\}$$

is a RKHS with kernel

$$\begin{aligned} K : &\mathcal{X} \times \mathcal{X} \to \mathbb{R} \\ &(x, r) \to K(x, r) := \langle \phi_x, \phi_r \rangle_{\mathcal{H}_1}. \end{aligned} \tag{1.34}$$

We remark that K by definition is a positive semi-definite kernel over $\mathcal{X}$ and ϕ represents the feature map on the feature space $\mathcal{H}_1$. Furthermore, we have

$$\Im(A) = \overline{span\{K_x, \; x \in \mathcal{X}\}}.$$

Moreover, we emphasise that conditions usually required on a reproducing kernel and on its associated RKHS are satisfied: $\mathcal{H}_K$ is separable since $\mathcal{H}_1$ is separable and A is a partial isometry from $\mathcal{H}_1$ to $\Im(A)$, and for all $x \in \mathcal{X}$ $K(x, x) \leq c^2$ since $K(x, x) = \langle \phi_x, \phi_x \rangle_{\mathcal{H}_1} = \|\phi_x\|^2_{\mathcal{H}_1}$ and inequality (1.33) applies.

Now we introduce the restriction of A to the space orthogonal to its nullspace and we prove the main result of this section which identifies the solutions of the two problems g_{R_y} and $f^\dagger_{R_y}$ as defined in equations (1.28) and (1.30), respectively. We denote with $\tilde{A}$ the restriction operator, i.e.

$$\tilde{A} := A_{|Ker(A)^\perp} : Ker(A)^\perp \to \Im(A). \tag{1.35}$$

By definition, $\tilde{A}$ admits the inverse operator $\tilde{A}^{-1}$.

Theorem 1. *Let g_{R_y} be the solution of the approximation problem in the RKHS $\mathcal{H}_K$ with kernel K defined in equation* (1.28)*. Let $f^\dagger_{R_y}$ be the solution of the inverse*

problem defined in equation (1.30) *with the operator A defined in equation* (1.32). *If* $\forall$ $x, x' \in \mathcal{X}$ $K(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1}$, *we have*

$$f^{\dagger}_{R_y} = \tilde{A}^{-1} g_{R_y} \, . \tag{1.36}$$

Proof. By hypothesis we have the following identification $\Im(A) = \mathcal{H}_K$ intended as RKHSs. Thanks to this identification the hypotheses on R_y in problems (1.28) and (1.30) (see sections 1.2.1 and 1.2.2) are exactly the same: the equivalence of hypotheses a) is straightforward; the hypotheses b) are equivalent as the coercivity of R_y with respect to the norm $\| \cdot \|_{\mathcal{H}_K}$ corresponds to the coercivity of $f \mapsto R_y(Af)$ with respect to the norm $\| \cdot \|_{\mathcal{H}_1}$. Let g_{R_y} be the solution of the problem (1.28) and let $\tilde{f} := \tilde{A}^{-1} g_{R_y}$. Then for all $f \in \mathcal{H}_1$ we have

$$R_y(Af) \geq \min_{g \in \Im(A)} R_y(g) = R_y(g_{R_y}) = R_y(A\tilde{f}) \, , \tag{1.37}$$

i.e. $\tilde{f}$ is solution of problem (1.29). Furthermore, by definition of $\tilde{A}^{-1}$, $\tilde{f} \in Ker(A)^{\perp}$ and therefore $\tilde{f}$ is the solution of (1.30), that is $\tilde{f} = f^{\dagger}_{R_y}$. □

Two remarks about this result are mandatory.

1) Under assumption 1, given an inverse problem described by a linear operator A (characterized by a map ϕ), it is always possible to associate with it an approximation problem in the RKHS $\mathcal{H}_K$ with kernel K defined by the map ϕ, i.e. $K(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1}$ for all $x, x' \in \mathcal{X}$.

2) Given an approximation problem in the RKHS $\mathcal{H}_K$ with kernel K, it is always possible to associate with it a feature map $\phi : x \in \mathcal{X} \to \phi_x \in \mathcal{H}_1$, where $\mathcal{H}_1$ is a Hilbert space and such that $K(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1}$ for all $x, x' \in \mathcal{X}$. In such a way we define $\mathcal{F} = \overline{span\{\phi_x \, , \, x \in \mathcal{X}\}}$, which is the feature space, and an inverse problem whose operator A is given in equation (1.32). By construction we have the identification between the feature space and the orthogonal of the null space of the operator, i.e. $\mathcal{F} = Ker(A)^{\perp}$. In the case that K is a continuous reproducing kernel, the Mercer theorem [95; 96; 97] gives us the way to describe the feature map

ϕ and the feature space is ℓ_2, whereas in the general case (when K is not necessarily continuous) we can consider the canonical feature map, that is $\phi : x \in \mathcal{X} \to \phi_x \in \mathcal{H}_K$ where $\forall\ x \in \mathcal{X}\ \phi_x = K_x$.

From the second remark the feature map associated with a given kernel K is determined up to the action of unitary group on $\mathcal{H}_1$, i.e.

$$K(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1} = \langle U\phi_x, U\phi_{x'} \rangle_{\mathcal{H}_1}, \tag{1.38}$$

for each unitary operator U acting on $\mathcal{H}_1$. In particular, we can define an equivalence relation $\sim$ on $\mathcal{H}_1$ using the left action of the unitary group $\mathcal{U}$. Let $f, f' \in \mathcal{H}_1$

$$f \ \sim\ f' \quad \iff \quad \exists\, U \in \mathcal{U} \ \mid\ f' = Uf. \tag{1.39}$$

We can also define an equivalence $\sim^{\mathcal{X}}$ between feature maps. Let $\phi, \phi' \in \mathcal{H}_1^{\mathcal{X}}$

$$\phi \ \sim^{\mathcal{X}}\ \phi' \quad \iff \quad \phi_x \sim \phi'_x\ ,\ \forall\ x \in \mathcal{X}. \tag{1.40}$$

Then, we define the map

$$\begin{aligned} \mathcal{K} : \mathcal{H}_1^{\mathcal{X}} &\longrightarrow \mathbb{R}^{\mathcal{X}\times\mathcal{X}} \\ \phi &\longmapsto K_\phi \end{aligned}$$

with $K_\phi(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1}$ and where $\mathcal{H}_1^{\mathcal{X}}$ denotes functions $\mathcal{X} \to \mathcal{H}_1$ and $\mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ denotes functions $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Therefore, from equations (1.38) and (1.40) we have a bijection

$$\begin{aligned} \mathcal{H}_1^{\mathcal{X}} / \sim^{\mathcal{X}} &\longleftrightarrow \Im(\mathcal{K}) \subset \mathbb{R}^{\mathcal{X}\times\mathcal{X}} \\ \bar{\phi} &\longleftrightarrow K_\phi\ , \end{aligned}$$

where $\bar{\phi}$ is the class induced by the equivalence relation $\sim^{\mathcal{X}}$ in (1.40). We denote with A_ϕ the operator defined in equation (1.32). We have

$$g_{R_y} = A_\phi f^\dagger_{R_y} = A_{\phi'}(f^\dagger_{R_y})',$$

where $\phi \sim^{\mathcal{X}} \phi'$ and $f^{\dagger}_{R_y} \sim (f^{\dagger}_{R_y})'$. Then we also have a bijection

$$\begin{aligned} \mathcal{H}_1/\sim &\longleftrightarrow \mathcal{H}_K \\ \overline{f^{\dagger}_{R_y}} &\longleftrightarrow g_{R_y} \end{aligned}$$

stating that, for any R_y satisfying conditions of problem (1.28) (or equivalently (1.30)) and for any $y \in \mathcal{H}_2$, the class of R_y-generalized solutions $\overline{f^{\dagger}_{R_y}}$ corresponds to the solution g_{R_y} of the approximation problem in the RKHS $\mathcal{H}_K$ defined in equation (1.28). Let us now fix an element $y \in \mathcal{H}_2$ and a functional R_y. For each $K \in \Im(\mathcal{K})$ we define the function $T_{R_y}(K) := g_{R_y}$ which maps the kernel K to the solution of the approximation problem in a RKHS defined in equation (1.28). In the same way, for each $\phi \in \mathcal{H}_1^{\mathcal{X}}$ we define the function $H^{\dagger}_{R_y}(\phi) := f^{\dagger}_{R_y}$ which maps the feature map ϕ to the R_y-generalized solution of the inverse problem defined in (1.30). Then, for each class $\overline{\phi}$, we can define a map $\overline{H^{\dagger}_{R_y}} : \mathcal{H}_1^{\mathcal{X}}/\sim^{\mathcal{X}} \to \mathcal{H}_1/\sim$ as follows

$$\overline{H^{\dagger}_{R_y}}(\overline{\phi}) := \pi(H^{\dagger}_{R_y}(\phi)) \,, \tag{1.41}$$

where ϕ is a representer of $\overline{\phi}$ and π is the quotient map with respect to the equivalence relation $\sim$ in (1.39). Furthermore, we denote with $\pi^{\mathcal{X}}$ the quotient map with respect to the equivalence relation $\sim^{\mathcal{X}}$ defined in (1.40). This definition is well-posed since it does not depend on the choice of the representer ϕ. We can summarize this discussion with the commutative diagram in Figure 1.1. In synthesis, when an approximation problem in a RKHS is provided with a feature map, it is equivalent to a linear inverse problem. If a feature map is not given, we can associate with the approximation problem in a RKHS as many inverse problems as feature maps (and so features spaces) which give rise to the same kernel.

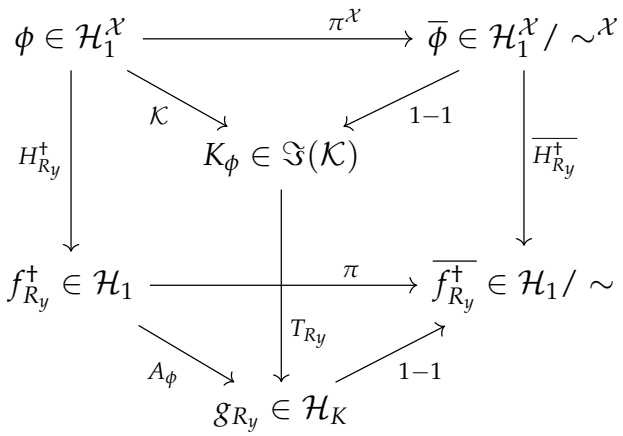


Figure 1.1: Commutative diagram summarizing the equivalence between approximation in a RKHS and linear inverse problems.

1.2.4 Tikhonov-type solutions of approximation in RKHSs and inverse problems

When y is corrupted by noise, the inverse problem needs to be addressed in a different way as the R_y-generalized solution $f^{\dagger}_{R_y}$ may not exist or it may not depend continuously on the data. A well-known strategy common to both approximation and inverse problems is Tikhonov regularization [43]. It allows us to find solutions of the problem which depend continuously on the data by re-stating the approximation problem in RKHS $\mathcal{H}_K$ defined in equation (1.28) as follows

$$\hat{g}_{R_y,\lambda} = \arg\min_{g \in \mathcal{H}_K} R_y(g) + \lambda\Omega(g), \tag{1.42}$$

and the inverse problem associated to the operator A given data y defined in equation (1.30) as follows

$$\hat{f}_{R_y,\lambda} = \arg\min_{f \in \mathcal{H}_1} R_y(Af) + \lambda\Omega(f)\ . \tag{1.43}$$

In these generalized Tikhonov regularization schemes R_y is usually called the data fidelity term, Ω is the penalty term and $\lambda > 0$ is the regularization parameter. The purpose of the penalty term is to induce stability and to allow

the incorporation of a priori information about the desired solution according to the magnitude of the parameter λ. In this context we assume that the penalty term has the following form

$$\Omega(h) := \psi(\|h\|_{\mathcal{H}}) \,, \tag{1.44}$$

where $\psi : [0, +\infty) \to \mathbb{R}_+$ is a continuous convex and strictly monotonically increasing real-valued function, h is an element of a Hilbert space $\mathcal{H}$ and $\|\cdot\|_{\mathcal{H}}$ denotes its norm. Now we show that the result of Theorem 1 can be extended to the case of Tikhonov regularized solutions $\hat{f}_{R_y,\lambda}$ and $\hat{g}_{R_y,\lambda}$.

Theorem 2. *Under the same assumptions of Theorem 1 we have*

$$\hat{f}_{R_y,\lambda} = \tilde{A}^{-1}\hat{g}_{R_y,\lambda} \,. \tag{1.45}$$

Proof. As in the proof of the Theorem 1 we have the identification $\Im(A) = \mathcal{H}_K$ as RKHSs and the hypotheses on functionals to minimize in equations (1.42) and (1.43) are the same. Let $\tilde{f} := \tilde{A}^{-1}\hat{g}_{R_y,\lambda}$. By definition of $\tilde{A}^{-1}$, $\tilde{f} \in Ker(A)^{\perp}$ and so $\|\hat{g}_{R_y,\lambda}\|_{\mathcal{H}_K} = \|\tilde{f}\|_{\mathcal{H}_1}$. For all $f \in \mathcal{H}_1$ we have

$$\begin{aligned} R_y(Af) + \lambda\psi(\|f\|_{\mathcal{H}_1}) &\geq \min_{g\in\Im(A)} R_y(g) + \lambda\psi(\|g\|_{\mathcal{H}_k}) \\ &= R_y(\hat{g}_{R_y,\lambda}) + \lambda\psi(\|\hat{g}_{R_y,\lambda}\|_{\mathcal{H}_k}) \\ &= R_y(A\tilde{f}) + \lambda\psi(\|\tilde{f}\|_{\mathcal{H}_1}) \end{aligned} \tag{1.46}$$

i.e. $\tilde{f}$ is solution of problem (1.43). This concludes the proof. □

As in the case of R_y-generalized solutions, we have a commutative diagram for Tikhonov regularized solutions. The diagram has exactly the same shape of the one shown in Figure 1.1 but arrows and nodes refer to the solution of problems in equations (1.42) and (1.43). In particular, we have to replace: $T^{\dagger}_{R_y}$ with the function $T_{R_y,\lambda}(K) := \hat{g}_{R_y,\lambda}$ which maps the kernel K to the Tikhonov solution in equation (1.42); $H^{\dagger}_{R_y}$ with the function $H_{R_y,\lambda}(\phi) := \hat{f}_{R_y,\lambda}$ which maps the feature map ϕ to the Tikhonov solution in equation (1.43); $\overline{H^{\dagger}_{R_y}}$ with

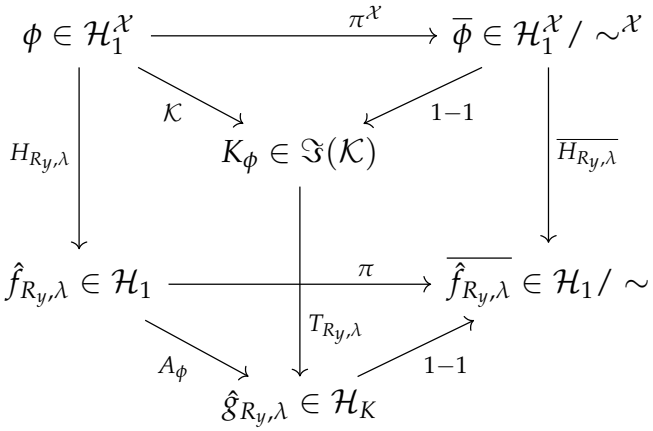


Figure 1.2: Commutative diagram summarizing the connection between Tikhonov-type solutions of the approximation in a RKHS and linear inverse problems.

the map $\overline{H_{R_y,\lambda}}$ defined as in equation (1.41) by substituting $H_{R_y}^\dagger$ with $H_{R_y,\lambda}$; $\overline{f_{R_y}^\dagger}$ with $\overline{\hat{f}_{R_y,\lambda}}$, which is the class of Tikhonov solutions corresponding to the Tikhonov solution of the approximation problem in the RKHS represented by K_ϕ.

Chapter 2

A unified formulation for learning and inverse problems

In this Chapter we provide a uniform formulation for applied inverse problems such as supervised learning or discrete inverse problems. Once the connection between approximation problems in RKHSs and inverse problems at the infinite dimensional level is established, the unified formulation needs the definition of a general sampling operator which allows us to derive different applied estimation problems from the same abstract approximation problem in RKHSs by suitably choosing the parametrization of the sampling operator. Therefore a general sampling operator has to take into account both deterministic and stochastic samples (section 2.1), providing an extension of the sampling operator defined in [126]. By means of this sampling operator, we show that supervised learning and discrete inverse problems are closely related to each other, both of which being able to be subsumed under the same infinite dimensional approximation problem.

Finally, for the sake of completeness, in section 2.3 we provide a discussion of the arguments used to prove convergence of discrete solutions to the ideal solutions in both statistical and deterministic settings. Finally, we make use of the equivalence between approximation in RKHSs and inverse problems (provided in Chapter 1) to show that the Representer Theorem [118], a well-

known result in statistical learning theory applies to inverse problems, too.

2.1 Sampling operator

The purpose of this section is to show that applied problems, such as discrete inverse problems, interpolation problems and statistical (inverse) learning, despite appearing different, can be thought of as instances of the approximation problem in a RKHS in Definition 7. To this end, we introduce a suitable discretization operator mapping the infinite dimensional data y to a finite number of samples together with a specific form of the functional R_y. The idea of the discretization operator is to consider, in place of the data y, a set of samples $\{(X_i, Y_i)\}_{i=1}^n$ statistically or deterministically related to y. In this way we will retrieve the formulation of various applied problems by minimizing the empirical form of the ideal functional R_y. To realize the discretization operator, i.e. a map from $\mathcal{H}_2$ to a sample space, we proceed as follows.

Definition 9 (V-characteristic of a distribution $\tilde{\rho}$). *Let us consider the set $\mathcal{P}$ of all possible Borel probability distributions over a compact space $\mathcal{Y} \subseteq \mathbb{R}$. We define the function $F_V : \mathcal{P} \to \mathbb{R}$ as follows*

$$F_V(\tilde{\rho}) := \arg\min_{w \in \mathbb{R}} \int_{\mathcal{Y}} V(Y, w)\, d\tilde{\rho}(Y), \tag{2.1}$$

where V is called loss function in the statistical learning terminology [116].

The function F_V is defined provided that $V : \mathcal{Y} \times \mathbb{R} \to [0, +\infty)$ is measurable and integrable with respect to the first variable and $V(Y, \cdot)$ is lower semicontinuous, strictly convex and coercive $\forall\ Y \in \mathcal{Y}$. Given a function V, $F_V(\tilde{\rho})$ can represent a characteristic of the distribution $\tilde{\rho}$, we show in the following some examples. Let Z be a random variable with probability distribution $\tilde{\rho}$,

- if V is the square loss usually used in regression problems, i.e. $V(Y, w) = (w - Y)^2$, or V is the Kullback-Leibler divergence, then $F_V(\tilde{\rho}) = \mathbb{E}(Z)$, i.e. is the expected value;

- if V is the square loss usually used in classification problems, i.e. $V(Y, w) = (1 - Yw)^2$ then $F_V(\tilde{\rho}) = \mathbb{E}(Z)/\mathbb{E}(Z^2)$;
- if V is the absolute value loss, i.e. $V(Y, w) = |w - Y|$ then $F_V(\tilde{\rho})$ is the median of the distribution $\tilde{\rho}$.

We now want to define a map from $\mathbb{R}$ to $\mathcal{P}$, roughly speaking an inverse of F_V.

Definition 10 (Distributions with a given V-characteristic). *We introduce an application*

$$\begin{aligned} \vartheta : \mathbb{R} &\to \mathcal{P} \\ z &\to \tilde{\rho}_z, \end{aligned} \tag{2.2}$$

mapping $z \in \mathbb{R}$ in a distribution $\tilde{\rho}_z$ such that $F_V \circ \vartheta = id$. Given a function y, ϑ maps $y(x)$ to a distribution $\tilde{\rho}_{y(x)}$ such that $y(x)$ is the characteristic of $\tilde{\rho}_{y(x)}$ for each $x \in \mathcal{X}$.

Therefore we define the following sampling operator.

Definition 11 (Sampling operator). *Let $S^{(n)}_{\bar{x},\vartheta}$ be defined as follows*

$$\begin{aligned} S^{(n)}_{\bar{x},\vartheta} : \mathcal{H}_2 &\to \mathcal{Y}^n \\ y &\to (Y_i)_{i=1,\dots,n} \end{aligned} \tag{2.3}$$

where each Y_i is drawn from the distribution $\tilde{\rho}_{y(x_i)} := \vartheta(y(x_i))$ and the set of points $\bar{x} = \{x_1, \dots, x_n\} \subset \mathcal{X}$ can be either given a priori (in a deterministic manner) or drawn from a probability distribution ν over $\mathcal{X}$.

Once V is fixed, for any chosen sampling $S^{(n)}_{\bar{x},\vartheta}$ let us consider the functional defined as

$$R_y(g) := \int_{\mathcal{X}\times Y} V(Y, g(X)) d\tilde{\rho}_{y(X)}(Y) d\nu(X) \tag{2.4}$$

which depends on ν and ϑ as well as on y and on V.

Henceforth, we consider the approximation problem in a RKHS (see Definition 7) with functional R_y given in equation (2.4). By applying $S^{(n)}_{\bar{x},\vartheta}$ to the

data y, we now show that we can retrieve the formulation of different applied problems according to whether ρ and ν are known or not and, if they are known, according to their specific explicit form. In general, when just a finite set of sample is known, $\mathcal{Z}_n = \{(x_i, Y_i)\}_{i=1}^n$, all these problems are addressed by minimizing the following empirical form of the functional, i.e.

$$R_{\mathcal{Z}_n}(g) := \frac{1}{n} \sum_{i=1}^{n} V((S^{(n)}_{\bar{x},\vartheta}(y))_i, g(x_i)) \,. \tag{2.5}$$

We summarize the construction of this section in the following

Proposition 2. *Consider the following equivalent minimization problems*

$$\arg\min_{g \in \mathcal{H}_K} R_y(g) \qquad \textit{and} \qquad \arg\min_{f \in \mathfrak{M}_{A,R_y}} \|f\|_{\mathcal{H}_1}, \tag{2.6}$$

where A is a linear operator satisfying assumption 1 (see Chapter 1), R_y is the functional defined in equation (2.4) *and $\mathfrak{M}_{A,R_y}$ is defined in equation* (1.29)*. Consider the sampling operator $S^{(n)}_{\bar{x},\vartheta}$ (see Definition 11), where ϑ is given in Definition 10. The two problems in equation* (2.6) *reduce to the following two empirical problems*

$$\hat{g}^{(n)}_R = \arg\min_{g \in \mathcal{H}_K} R_{\mathcal{Z}_n}(g) \quad \textit{and} \quad (\hat{f}^{(n)}_R)^\dagger = \arg\min_{f \in \mathfrak{M}_{A,R_{\mathcal{Z}_n}}} \|f\|_{\mathcal{H}_1}. \tag{2.7}$$

where $R_{\mathcal{Z}_n}$ is defined as

$$R_{\mathcal{Z}_n}(g) = \frac{1}{n} \sum_{i=1}^{n} V(Y_i, g(X_i)), \tag{2.8}$$

$\mathfrak{M}_{A,R_{\mathcal{Z}_n}}$ is the set of solutions minimizing the functional $f \mapsto R_{\mathcal{Z}_n}(Af)$ and $\mathcal{Z}_n$ is the set of samples generated by applying the sampling operator $S^{(n)}_{\bar{x},\vartheta}$ to y.

The advantage of this result consists in the following:

Remark 1. *According to the choice of the parameters $\bar{x}$ and ϑ in the sampling operator we retrieve the following cases.*

1) Statistical kernel learning:

- *the elements x_i of $\bar{x}$ are given at random according to a distribution ν;*
- *$\tilde{\rho}_{y(x_i)} = \vartheta(y(x_i)) = \rho(\cdot|X = x_i)$, i.e. it is the conditional distribution with respect to $X = x_i$.*

In this case ν and $\rho(\cdot|X)$ are considered to be unknown.

2) *Statistical inverse problems with random matrix design:*

- *same hypotheses of the first case with the substantial difference that ν and $\rho(\cdot|X)$ are considered to be (at least partially) known.*

3) *Statistical inverse problems with fixed matrix design:*

- *the elements x_i of $\bar{x}$ are given not at random;*
- *$\tilde{\rho}_{y(x_i)} = \vartheta(y(x_i)) = \rho(\cdot|X = x_i)$, i.e. it is the conditional distribution with respect to $X = x_i$.*

In this case a natural choice of the loss V is given by the Maximum Likelihood approach according to ρ (see section 2.2).

4) *Deterministic inverse problems:*

- *the elements x_i of $\bar{x}$ are given not at random;*
- *$\tilde{\rho}_{y(x_i)} = \vartheta(y(x_i)) = \delta(\cdot - y(X)|X = x_i)$.*

In this case the samples Y_i are the values of the function y, or a noisy version y^δ, at the points x_i.

In a machine learning problem the samples can be view as the result of a sampling process which takes place upstream of the definition of the problem itself, or in any way, independently of the will of the learner. It is indeed formalized as an empirical process in accordance with an unknown distribution. On the contrary, in an inverse problem the discretization usually takes place downstream of the problem: for example, in the case of an industrial device, it can be defined during the design phase or determined even later, after the signal acquisition, as a variable to be optimized in the inversion process.

Incidentally, we notice that in learning problems a given point can be sampled more than once whereas in inverse problems each sample $x \in \mathcal{X}$ is usually taken once.

We remark that for learning problems formulation given in Proposition 2 differs from the classical one where the samples are given without any discretization process. In the classical formulation the crucial hypothesis is that the samples are drawn independently and identically distributed according to a distribution $\rho(\cdot,\cdot)$ and there is no need to introduce from the beginning ν and the conditional distribution $\rho(\cdot|\cdot)$. However, these last two distributions are the result of the factorization of ρ. In the classical version, y is introduced after ρ, it depends on the choice of V and it is the V-characteristic of the distribution ρ (in the sense of Definition 9) representing the parameter to learn. We remark that, in this case, thanks to the factorization property $\rho(X,Y) = \rho(Y|X)\nu(X)$ the functional R_y in equation (2.4) coincides with the expected risk R_ρ defined in equation (1.13), and the V-characteristic to learn y, given by $y(x) = F_V(\rho(\cdot|X = x))$, is the minimizer of the expected risk (over all measurable functions). The crucial point is that ρ is unknown. In contrast, when ρ is known, which is usually the case of discrete inverse problems (with both random and fixed matrix design), the loss function V can be chosen in a natural way by means of the Maximum Likelihood approach. Table 2.1 summarizes the main sampling schemes corresponding to different applications.

Table 2.1: Discretization schemes.

	sampling $S^{(n)}_{\bar{x},\vartheta}$	
	$\rho(\cdot\|\cdot)$ and ν unknown	$\bar{x}$ given and $\rho(\cdot\|\cdot)$ (partially) known
direct	learning	interpolation
inverse	inverse learning	discrete inverse problems

From a numerical point of view, solving the empirical problems in (2.7)

needs regularization. This is achieved by adding a penalty term as follows

$$\hat{g}_{R,\lambda}^{(n)} := \arg\min_{g \in \mathcal{H}_K} R_{\mathcal{Z}_n}(g) + \lambda\psi(\|g\|_{\mathcal{H}_K}) \tag{2.9}$$

and

$$\hat{f}_{R,\lambda}^{(n)} := \arg\min_{f \in \mathcal{H}_1} R_{\mathcal{Z}_n}(Af) + \lambda\psi(\|f\|_{\mathcal{H}_1}) \,, \tag{2.10}$$

where $\lambda > 0$ is the regularization parameter and the function $\psi : [0, +\infty) \to \mathbb{R}_+$ is a non-decreasing convex function (see Chapter 1). This formulation is known as Tiknonov regularization in inverse problems and structural risk minimization in statistical learning [43; 140]. The general discrete minimization problem in equation (2.5), as well as problems in equation (2.7), depends on the set of points $\mathcal{Z}_n$ but not on their statistical or deterministic origin, i.e. it does not depend on the specific choice of $\bar{x}$ and ϑ. For this reason, the solutions of the discretized problems have been indicated with the notation $\hat{g}_R^{(n)}$, $(\hat{f}_R^{(n)})^\dagger$, $\hat{g}_{R,\lambda}^{(n)}$ and $\hat{f}_{R,\lambda}^{(n)}$ regardless the nature of the samples $\mathcal{Z}_n$. In this respect, we give the following:

Corollary 1. *Given $\mathcal{Z}_n$ a set of samples, let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator between Hilbert spaces, which satisfies assumption 1. Let ϕ the feature map of A (see equation* (1.32)*). By assuming that $\forall\, x, x' \in \mathcal{X}$ $K(x, x') =< \phi_x, \phi_{x'} >_{\mathcal{H}_1}$, where K is the reproducing kernel of $\mathcal{H}_K$ we have that*

$$(\hat{f}_R^{(n)})^\dagger = \tilde{A}^{-1}\hat{g}_R^{(n)} \qquad \text{and} \qquad \hat{f}_{R,\lambda}^{(n)} = \tilde{A}^{-1}\hat{g}_{R,\lambda}^{(n)} \,, \tag{2.11}$$

where $\tilde{A}^{-1}$ is the inverse of the restriction of A in equation (1.35)*. Furthermore, the solutions $\hat{g}_R^{(n)}$ and $\hat{g}_{R,\lambda}^{(n)}$ correspond to the set of solutions $\{U(\hat{f}_R^{(n)})^\dagger \mid U \in \mathcal{U}\}$ and $\{U\hat{f}_{R,\lambda}^{(n)} \mid U \in \mathcal{U}\}$, respectively, where we remind that $\mathcal{U}$ is the set of unitary operators on $\mathcal{H}_1$.*

This result is valid for any choice of $\bar{x}$ and ϑ, i.e. independently of the discretization scheme. The proof is omitted since it is a straightforward application of Theorems 1 and 2.

2.2 Maximum likelihood approach

We consider discretized problems , as instance a discretized inverse problem characterized by a linear bounded operator A, as follows

$$Y_i = (Af)(x_i), \quad i = 1, \dots, n, \tag{2.12}$$

where $x_1, \dots, x_n$ are n points deterministically identified, and for each $i \in \{1, \dots, n\}$ we know the sample Y_i from a given probability distribution $\rho(Y|X = x_i)$. The main difference with respect to the learning framework is that here the probability distribution $\rho(\cdot|\cdot)$ is known and the quantity to be determined is the parameter f which characterizes the distribution $\rho(\cdot|\cdot)$. For this kind of problem the standard approach is the so-called Maximum Likelihood (ML) estimation [13; 73; 133; 142]. We denote with $p_{Af}(\cdot|\cdot)$ the conditional density. In the Maximum Likelihood approach, the likelihood is introduced, which is given by

$$L_{Af}(\mathbf{Y}, \mathbf{x}) = \prod_{i=1}^{n} p_{Af}(Y_i|x_i), \tag{2.13}$$

where $\mathbf{Y}$ and $\mathbf{x}$ denote the n-dimensional vectors which contain the samples Y_i and x_i, respectively. Maximize the likelihood is equivalent to minimize the following negative log-likelihood

$$J_{Af}(\mathbf{Y}, \mathbf{x}) = -\alpha_1 \log(L_{Af}(\mathbf{Y}, \mathbf{x})) + \alpha_2, \tag{2.14}$$

where $\alpha_1 > 0$ and α_2 are two suitable constants. Therefore, the problem is addressed by solving

$$\arg\min_{f \in \mathcal{H}_1} J_{Af}(\mathbf{Y}, \mathbf{x}). \tag{2.15}$$

The knowledge of the noise on data allows us to have an explicit form of J_{Af}. We show two examples.

- Gaussian additive noise: $J_{Af}(\mathbf{Y}, \mathbf{x}) = \sum_{i=1}^{n} (Y_i - (Af)(x_i))^2$ by a suitable choice of α_1 and α_2. The ML coincides with the usual least square minimization problem.

- Poisson noise: $J_{Af}(\mathbf{Y}, \mathbf{x}) = \sum_{i=1}^{n} Y_i \log(\frac{Y_i}{(Af)(x_i)}) + (Af)(x_i) - Y_i$ by a suitable choice of α_1 and α_2, i.e. J_{Af} is the Kullback-Leibler divergence.

This problem can be formalized by means of the sampling operator $S^{(n)}_{\bar{x},\vartheta}$ which yields the sample set

$$\mathcal{Z}_n := \{(x_1, Y_1), \ldots, (x_n, Y_n)\} , \tag{2.16}$$

where $\vartheta(Af(x)) := \rho_{Af}(\cdot|X = x)$, which is the conditional probability distribution (associated with the conditional density $p_{Af}(\cdot|x)$). In the Maximum Likelihood approach the choice of V is the following

$$V(Y, Af(x)) = -\log p_{Af}(Y|X = x) \tag{2.17}$$

less than suitable constants. With this choice equation (2.5) takes the form

$$R_{\mathcal{Z}_n}(g) = \frac{1}{n} \sum_{i=1}^{n} -\log(p_{Af}(Y_i|x_i)) , \tag{2.18}$$

which corresponds to the negative-log formulation of the Maximum Likelihood approach.

2.3 Convergence

In the previous part we considered the ideal functional, defined as in equation (2.4), and its empirical form, in equation (2.5), and we considered the variational problems consisting in minimizing the two functionals in order to define the solution of the approximation/inverse problem and the one of the associated discrete problem. For the sake of completeness we discuss the convergence of the empirical functional defined in equation (2.5) to the ideal one defined in equation (2.4) and the convergence of their respective minimizers in the deterministic and statistical setting. In the case $\mathcal{Z}_n$ is randomly drawn the convergence is defined in terms of probabilities and the conditions

are well established [102; 140]. However, if $\mathcal{Z}_n$ is assumed to be generated in a deterministic manner, the convergence is defined in terms of norms and the theoretical tools for proving the convergence are slightly different. Indeed, whereas in the statistical framework convergence is a consequence of a straightforward application of the argmax continuous theorem [138], we show that in the deterministic framework we need a result relying on the notion of Γ-convergence [24].

2.3.1 Statistical setting

Consider a set of samples $\mathcal{Z}_n = \{(X_i, Y_i)\}_{i=1}^n$ drawn from a probability distribution as described in previous sections. We recall a classical theorem ensuring the consistency of a sequence of arg max-estimators in an *argmin* version suitable for our framework [138]. Let (H, d) be a metric space and (F_n) be a sequence of random functions over H given a probability distribution ν.

Theorem 3 (Argmax continuous theorem). *Let us suppose*

$$\sup_{h \in H} |F_n(h) - F(h)| \to^{\mathbb{P}} 0 \,, \tag{2.19}$$

where F is a fixed function over H and for each $\epsilon > 0$

$$\inf_{h \in H : d(h, h^*) \geq \epsilon} F(h) > F(h^*), \tag{2.20}$$

where h^* is the minimizer of F. Moreover, if $F_n(h^{(n)}) \leq F_n(h^*) + o_{\mathbb{P}}(1)$, we have

$$h^{(n)} \to^{\mathbb{P}} h^* \tag{2.21}$$

where $h^{(n)}$ is the minimizer of F_n.

Whereas the second hypothesis is a property of the limit function F at its minimum point h^*, which is assured when F is strictly convex, coercive and lower semi-continuous, the first hypothesis in equation (2.19) requires the

uniform convergence of (F_n). When F_n takes the form of the empirical risk (equation (2.5)) and F is given by equation (2.4) the condition in equation (2.19) is satisfied if H is a uniform Glivenko-Cantelli class (uGC) [40], provided that V has some Lipschitz property [99]. Then, we have the following

Corollary 2. *Let $\mathcal{H}_K$ be uGC. Let R_y be defined in equation* (2.4) *and let V be a loss function as in section 2.1 with the additional Lipschitz property described in [99]. Assume that V satisfies the following coercivity property: for each sequence $(g_k) \subseteq \mathcal{H}_K$ such that $\|g_k\|_{\mathcal{H}_K} \to \infty$, as $k \to \infty$ then $V(Y, g_k(X)) \to \infty$, as $k \to \infty$, for each $Y \in \mathcal{Y}$ and $X \in \mathcal{X}$. Then as $n \to +\infty$,*

$$\hat{g}_R^{(n)} \longrightarrow^{\mathbb{P}} g_{R_y} \quad \textit{and} \quad (\hat{f}_R^{(n)})^\dagger \longrightarrow^{\mathbb{P}} f_{R_y}^\dagger \ , \tag{2.22}$$

where $\hat{g}_R^{(n)}$ and $(\hat{f}_R^{(n)})^\dagger$ are defined in equation (2.7)*, g_{R_y} is the minimizer of R_y over $\mathcal{H}_K$ and $f_{R_y}^\dagger$ is the R_y-generalized solution in according to the definition in equation* (1.30)*, respectively.*

Proof. Let us take $F_n := R_{\mathcal{Z}_n}$ (where $R_{\mathcal{Z}_n}$ is defined in equation (2.5)) and $F := R_y$ in Theorem 3. Condition in equation (2.19) is verified for the uGC hypothesis on $\mathcal{H}_K$. Condition in equation (2.20) is verified thanks to the hypothesis of uniqueness of the minimizer of R_y. Moreover, the sequence $\hat{g}_R^{(n)}$ satisfies $R_{\mathcal{Z}_n}(\hat{g}_R^{(n)}) \leq R_{\mathcal{Z}_n}(g_{R_y}) + o_{\mathbb{P}}(1)$ as $\hat{g}_R^{(n)}$ is the minimizer of $R_{\mathcal{Z}_n}$. Using the connection between direct and inverse problems, (see Corollary 1) we have the following equalities

$$\|\hat{g}_R^{(n)} - g_{R_y}\|_{\mathcal{H}_K} = \|A(\hat{f}_R^{(n)})^\dagger - A f_{R_y}^\dagger\|_{\mathcal{H}_K} = \|(\hat{f}_R^{(n)})^\dagger - f_{R_y}^\dagger\|_{\mathcal{H}_1} \ . \tag{2.23}$$

This completes the proof. □

Remark 2. *The same convergence result of Corollary 2 applies for Tikhonov-type regularized solutions, i.e. fixed $\lambda > 0$ we have that $\hat{g}_{R,\lambda}^{(n)}$ and $\hat{f}_{R,\lambda}^{(n)}$ defined in* (2.9) *and* (2.10) *converge in probability to $\hat{g}_{R_y,\lambda}$, and $\hat{f}_{R_y,\lambda}$, defined in* (1.42) *and* (1.43)*, respectively.*

2.3.2 Deterministic setting

Consider a set of samples $\mathcal{Z}_n = \{(x_i, y_i)\}_{i=1}^n$ deterministically given, where $y_i = y(x_i)$ $i = 1, \dots, n$ with y the infinite dimensional data. The convergence in the deterministic case needs the use of the fundamental theorem of Γ-convergence [24]. First, we recall the Γ-convergence definition for a given sequence (F_n) of functions on a metric space (H, d) with respect to the distance d.

Definition 12. *The sequence (F_n) Γ-converges in H to a fixed function F if for all $h \in H$ the lim inf inequality holds, i.e. for all sequence h_n such that $d(h_n, h) \to 0$, as $n \to +\infty$*

$$F(h) \leq \liminf_n F_n(h_n) \tag{2.24}$$

and the lim sup inequality holds, i.e. there exists a sequence h_n such that $d(h_n, h) \to 0$, as $n \to +\infty$ such that

$$F(h) \geq \limsup_n F_n(h_n). \tag{2.25}$$

In order to prove the Γ-convergence of a sequence we use the following characterization of the equi-coerciveness of a sequence [34].

Lemma 1. *(F_n) is an equi-coercive sequence $\iff$ there exists a lower semicontinuous coercive function G such that $F_n \geq G$ on H, for each $n \in \mathbb{N}$.*

We also exploit the following result which is a consequence of the fundamental theorem of Γ-convergence (see [25] for details).

Proposition 3. *Let (F_n) be an equi-coercive sequence Γ-converging to F. Let h_n be a minimizer of F_n, and we assume F admits a unique point of minimum h. Then $h_n \to h$, as $n \to +\infty$, i.e. $d(h_n, h) \to 0$, as $n \to +\infty$.*

We now prove the convergence of the minimizer of $R_{\mathcal{Z}_n}$ to the one of R_y over $\mathcal{H}_K$, where R_y is defined in equation (2.4) and V is strictly convex, Lipschitz continuous with respect to the second variable and coercive in the

sense of the definition given in Corollary 2. We remark that in the deterministic case, assuming $dv(x) = dx$, the functional R_y reduces to

$$R_y(g) = \int_{\mathcal{X}} V(y(x), g(x))dx. \tag{2.26}$$

Results apply by considering the minimization problems over the RKHS $\mathcal{H}_K$ or the inverse problems with the operator A which satisfies assumption 1. In both cases assumption 1 assures that the feature map associated to A or the reproducing kernel K associated to $\mathcal{H}_K$ is bounded (see Chapter 1).

Proposition 4. *Let $x_1, \dots, x_n \in \mathcal{X}$ such that the sequence of points (x_n) is dense in $\mathcal{X}$. Let (y_n) be the set of samples, taken as the points $y_i = y(x_i)$ $i = 1, \dots, n$. Let R_y be defined in equation* (2.26) *and let $R_{\mathcal{Z}_n}$ be defined in equation* (2.8)*, with V a Lipschitz continuous with respect to the second variable and coercive function. Then the sequence $(R_{\mathcal{Z}_n})$ is an equi-coercive sequence and it Γ-converges to R_y taking the metric space $(\mathcal{H}_K, \|\cdot\|_{\mathcal{H}_K})$.*

Proof. To prove the equi-coerciveness of the sequence $(R_{\mathcal{Z}_n})$, it is sufficient to observe that $R_{\mathcal{Z}_n} \geq R_{\mathcal{Z}_1}$ for all $n \in \mathbb{N}$ where $R_{\mathcal{Z}_1}(g) = V(y_1, g(x_1))$ and then $R_{\mathcal{Z}_1}$ is coercive and continuous for the hypothesis on V. Now we prove that $(R_{\mathcal{Z}_n})$ Γ-converges to R_y. Without loss of generality we assume $\mathcal{X} = [0,1]^p$. Let $g \in \mathcal{H}_K$ and let (g_n) be a sequence converging to g, i.e. $\|g_n - g\|_{\mathcal{H}_K} \to 0$, then we have the following inequality

$$\left|R_{\mathcal{Z}_n}(g_n) - R_y(g)\right| \leq \left|R_{\mathcal{Z}_n}(g) - R_y(g)\right| + \left|R_{\mathcal{Z}_n}(g_n) - R_{\mathcal{Z}_n}(g)\right|. \tag{2.27}$$

The first term in the r.h.s. of equation (2.27) converges to 0 as $n \to +\infty$ for the definition of the Riemann integral and for the density of the points x_i in $\mathcal{X}$. Now we prove that the second term in the r.h.s. of equation (2.27) converges to 0. Under the assumption 1 we have that $\|K_{x_i}\|_{\mathcal{H}_K} \leq c$, $\forall$ x_i, where c is a fixed constant (see section 1.2.3). By using the Lipschitz continuity of V and

the reproducing property of K we have the following inequalities

$$\begin{aligned} |R_{\mathcal{Z}_n}(g_n) - R_{\mathcal{Z}_n}(g)| &\leq \frac{1}{n}\sum_{i=1}^{n} |V(g_n(x_i), y_i) - V(g(x_i), y_i)| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \sigma |g_n(x_i) - g(x_i)| \leq c\sigma \|g_n - g\|_{\mathcal{H}_K}, \end{aligned} \tag{2.28}$$

where σ is the Lipschitz constant of V. Therefore, for each sequence $(g_n)_n$ converging to g there exists $\lim_{n\to+\infty} R_{\mathcal{Z}_n}(g_n) = R(g)$. Then $(R_{\mathcal{Z}_n})$ Γ-converges to R_y. □

Corollary 3. *Under the assumptions of Proposition 4 and requiring that V is strictly convex with respect to the second variable we consider $\hat{g}_R^{(n)}$ and $(\hat{f}_R^{(n)})^\dagger$ defined in equation* (2.7)*, g_{R_y} defined in equation* (1.28) *and $f_{R_y}^\dagger$ defined in equation* (1.30)*. Then, as $n \to \infty$*

$$\hat{g}_R^{(n)} \to g_{R_y}\,, \quad \textit{and} \quad (\hat{f}_R^{(n)})^\dagger \longrightarrow f_{R_y}^\dagger\,, \tag{2.29}$$

where the convergence is uniform in $\mathcal{H}_K$ and $\mathcal{H}_1$, respectively.

Proof. The convergence in $\mathcal{H}_K$ follows from Propositions 3 and 4, by observing that R_y admits a unique minimizer. The convergence in $\mathcal{H}_1$ follows from the equality in equation (2.23). □

Remark 3. *The same convergence result of Corollary 3 applies for Tikhonov-type regularized solutions, i.e. $\hat{g}_{R_y,\lambda}^{(n)}$ and $\hat{f}_{R_y,\lambda}^{(n)}$ converge to $\hat{g}_{R_y,\lambda}$ and $\hat{f}_{R,\lambda}$, respectively. Such a result follows from the fact that $(R_{\mathcal{Z}_n} + \lambda\psi(\|\cdot\|_{\mathcal{H}_K}))$ is equi-coercive and Γ-converges to $R_y + \lambda\psi(\|\cdot\|_{\mathcal{H}_K})$ which is a straightforward consequence of Proposition 4 and the fact that $\lambda\psi(\|\cdot\|_{\mathcal{H}_K})$ is continuous (see [24]).*

Finally, as the convergence property of the $R_{\mathcal{Z}_n}$-generalized solution holds regardless the discretization scheme we can summarize functionals, solutions, convergence and discretization with the commutative diagrams shown in Figure 2.1.

The vertexes of the rear side of the cube represent the four minimizing functionals and the vertexes of the front side represent the corresponding

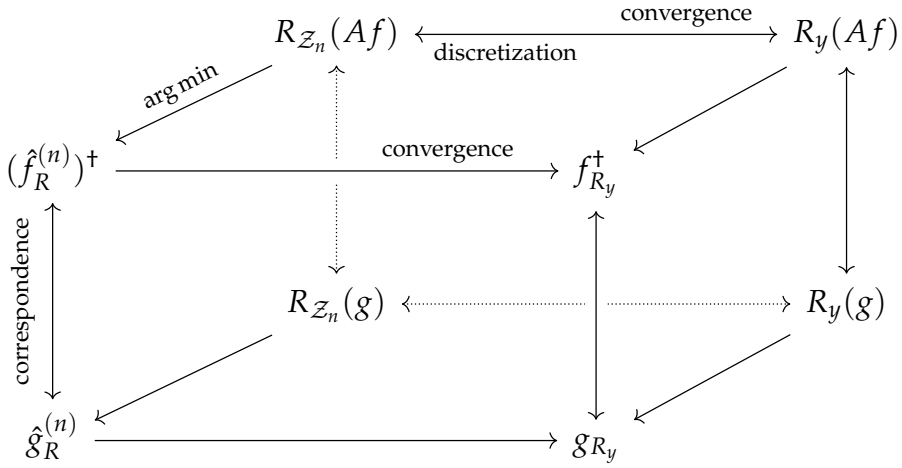


Figure 2.1: A summary of the discretization and convergence results applied to the approximation problems in a RKHS. Arrows indicate: from left to right convergence processes; from right to left discretization processes in the rear panel; from rear to front optimization processes; from top to bottom (and viceversa) the correspondence between inverse and direct problems.

solutions. The empirical and ideal cases are shown on the left and right sides, respectively. The arrows from left to right represent the convergence, whereas the arrows from right to left, on the rear side, represent the discretization. The arrows from rear to front show the minimizing process. In particular, along horizontal arrows of the front side of the cube we show the convergence of the empirical solutions to the ideal ones (Corollaries 2 and 3); along vertical arrows we show the correspondence between solutions of approximation problems in a RKHS and inverse problems (Theorem 1 and Corollary 1).

2.3.3 Representer theorem

The representer theorem and its generalizations prove that the regularized solution $\hat{g}_{R,\lambda}^{(n)}$ defined in equation (2.9) belongs to a finite dimensional subspace of $\mathcal{H}_K$ [118]. Under the assumption 1 on the linear operator $A : \mathcal{H}_1 \to \mathcal{H}_2$, let

$$\mathcal{H}_K^{(n)} := span\{K_{x_1}, \ldots, K_{x_n}\} \tag{2.30}$$

and

$$\mathcal{H}_1^{(n)} := span\{\phi_{x_1}, \dots, \phi_{x_n}\}, \tag{2.31}$$

be two finite dimensional subspaces $\mathcal{H}_K^{(n)} \subset \mathcal{H}_K$ and $\mathcal{H}_1^{(n)} \subset \mathcal{H}_1$, where ϕ and K are related by the equation (1.34). Under the aforementioned conditions on the loss function V and ψ (on which depends the penalty term), in the statistical learning setting the representer theorem allows us to write

$$\hat{g}_{R,\lambda}^{(n)} = \sum_{i=1}^{n} \beta_i K_{x_i}, \tag{2.32}$$

where $\beta_i \in \mathbb{R}$ for all $i \in \{1, \dots, n\}$ are appropriate coefficients. Thus, the problem (2.9) can be re-formulated as follows

$$\hat{g}_{R,\lambda}^{(n)} := \arg \min_{g \in \mathcal{H}_K^{(n)}} R_{\mathcal{Z}_n}(g) + \lambda \psi(\|g\|_{\mathcal{H}_K}), \tag{2.33}$$

where the optimization is performed on the finite dimensional subspace $\mathcal{H}_K^{(n)}$. Clearly, Corollary 1 can be exploited to provide a representer theorem for $\hat{f}_{R,\lambda}^{(n)}$.

Proposition 5. *The regularized solution $\hat{f}_{R,\lambda}^{(n)}$ defined in equation* (2.10) *admits the following representation*

$$\hat{f}_{R,\lambda}^{(n)} = \sum_{i=1}^{n} \beta_i \phi_{x_i}, \tag{2.34}$$

where $\beta_i \in \mathbb{R}$, for all $i \in \{1, \dots, n\}$ are the same coefficients of equation (2.32). *Finally the problem* (2.10) *can be re-formulated as follows*

$$\hat{f}_{R,\lambda}^{(n)} := \arg \min_{f \in \mathcal{H}_1^{(n)}} R_{\mathcal{Z}_n}(Af) + \lambda \psi(\|f\|_{\mathcal{H}_1}), \tag{2.35}$$

where $\mathcal{H}_1^{(n)}$ is defined in equation (2.31).

The major consequence of this result is that it is sufficient to determine coefficients $\{\beta_j\}_{j=1}^n$ in order to solve both problems (2.33) and (2.35). For the

sake of completeness, we report the explicitly computation of the coefficients β_j in the classical Tikhonov regularization case.

Example 1. *Let us consider the Tikhonov regularization for a linear inverse problem which is known as penalized least square approach in supervised learning. Under the usual assumptions, we write the problem* (2.9) *as*

$$\hat{g}_\lambda^{(n)} = \arg\min_{g \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(X_i))^2 + \lambda \|g\|_{\mathcal{H}_K}^2 , \tag{2.36}$$

and the problem (2.10) *as*

$$\hat{f}_\lambda^{(n)} = \arg\min_{f \in \mathcal{H}_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - Af(X_i))^2 + \lambda \|f\|_{\mathcal{H}_1}^2 . \tag{2.37}$$

Using the representer theorem and equation (2.11) *the solution of the two problems* (2.36) *and* (2.37) *is given by solving the following*

$$\hat{\beta}_\lambda^{(n)} = \arg\min_{\beta \in \mathbb{R}^n} \frac{1}{n} \|\boldsymbol{Y} - \boldsymbol{K}\beta\|_2^2 + \lambda \beta^T \boldsymbol{K} \beta, \tag{2.38}$$

where $\boldsymbol{Y}$ is the n-dimensional vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, and $\boldsymbol{K}$ is the $n \times n$ dimensional matrix with entries $\boldsymbol{K}_{ij} := K(X_i, X_j)$, for each $i, j \in \{1, \ldots, n\}$. In such a case the solution $\hat{\beta}_\lambda^{(n)}$ is given by

$$\hat{\beta}_\lambda^{(n)} = (\boldsymbol{K} + \lambda n \boldsymbol{I})^{-1} \boldsymbol{Y}, \tag{2.39}$$

where $\boldsymbol{I}$ is the $n \times n$ dimensional identity matrix. Therefore, solutions of problems (2.36) *and* (2.37) *are given respectively by*

$$\hat{g}_\lambda^{(n)} = \boldsymbol{k}^T (\boldsymbol{K} + \lambda n \boldsymbol{I})^{-1} \boldsymbol{Y}, \tag{2.40}$$

where $\boldsymbol{k} = (K_{X_1}, \ldots, K_{X_n})^T$, and

$$\hat{f}_\lambda^{(n)} = \boldsymbol{\Phi}^T (\boldsymbol{K} + \lambda n \boldsymbol{I})^{-1} \boldsymbol{Y}, \tag{2.41}$$

where $\Phi = (\phi_{X_1}, \ldots, \phi_{X_n})^T$.

Analogously, the solutions $\hat{g}_R^{(n)}$ and $(\hat{f}_R^{(n)})^\dagger$ defined in equation (2.7) admit a finite representation. This follows from the fact that $\hat{g}_R^{(n)}$ can be seen as the minimizer of the problem (2.33) with $\psi = 0$. Hence, at least a minimizer has a finite representation as ψ is non-decreasing and it is unique as $R_{\mathcal{Z}_n}$ is strictly convex [5; 150]. In the next proposition we give a simple alternative proof of the fact that $\hat{g}_R^{(n)}$ and $(\hat{f}_R^{(n)})^\dagger$ admit a finite representation based on Γ-convergence.

Proposition 6. *Let* $R_{\mathcal{Z}_n}$ *be defined in equation* (2.8)*, with* V *strictly convex, coercive (as the definition in Corollary 2) and Lipschitz continuous function. The solution* $\hat{g}_R^{(n)}$ *defined in equation* (2.7) *admits the following representation*

$$\hat{g}_R^{(n)} = \sum_{j=1}^{n} \alpha_j K_{x_j} \,, \tag{2.42}$$

where $\alpha_j \in \mathbb{R}$*, for all* $j \in \{1, \ldots, n\}$ *are appropriate coefficients.*

Proof. Let $\lambda > 0$ and let ψ be a continuous convex and strictly increasing real-valued function. Fixed $n \in \mathbb{N}$, the sequence $(R_{\mathcal{Z}_n} + \lambda\psi(\|\cdot\|_{\mathcal{H}_K}))_\lambda$ satisfies the hypotheses of Proposition 3 and it Γ-converges to $R_{\mathcal{Z}_n}$ as $\lambda \to 0$. This proves the convergence of minimizers, i.e. $\hat{g}_{R,\lambda}^{(n)} \to \hat{g}_R^{(n)}$ as $\lambda \to 0$, uniformly in $\|\cdot\|_{\mathcal{H}_K}$ for all $n \in \mathbb{N}$, where $\hat{g}_{R,\lambda}^{(n)}$ is defined in equation (2.9). Moreover, $\hat{g}_{R,\lambda}^{(n)}$ admits the following representation

$$\hat{g}_{R,\lambda}^{(n)} = \sum_{j=1}^{n} \beta_j^\lambda K_{x_j} \,, \tag{2.43}$$

where $\beta_j^\lambda \in \mathbb{R}$ for all $j \in \{1, \ldots, n\}$. Therefore, $\sum_{j=1}^n \beta_j^\lambda \, K_{x_j}$ pointwise converges to $\hat{g}_R^{(n)} \in \mathcal{H}_K$ as $\lambda \to 0$ and each β_j^λ has to converge to some value β_j^0. The limit can be written as $\sum_{j=1}^n \beta_j^0 \, K_{x_j}$ and this shows that $\hat{g}_R^{(n)} \in \mathcal{H}_K^{(n)}$. □

Corollary 4. *Under assumptions of Proposition 6, consider* $(\hat{f}_R^{(n)})^\dagger$ *defined in equation*

(2.7). *Then* $(\hat{f}_R^{(n)})^\dagger$ *admits the following representation*

$$(\hat{f}_R^{(n)})^\dagger = \sum_{j=1}^{n} \alpha_j \phi_{x_j} \,, \tag{2.44}$$

where $\alpha_j \in \mathbb{R}$*, for all* $j \in \{1, \ldots, n\}$ *are the same coefficients in equation* (2.42).

The proof is a straightforward consequence of Proposition 6 and Corollary 1.

Chapter 3

Convergence rates comparison

Convergence rates are studied in both learning and inverse problems theory. In inverse problems theory convergence rates have been analyzed in the infinite dimensional setting with decreasing noise level [43]. In this case the noise is considered as a deterministic quantity and it is natural to study the worst–case error. If the noise is modeled as a random quantity, the convergence of estimators should be studied in statistical terms, e.g. computing the expected mean square error [16]. In the discrete setting, they are studied under both deterministic and statistical hypotheses, with increasing number of samples. In the deterministic setting the proof techniques usually consist in decomposing the error in two terms, i.e. the approximation term and the noise amplification term, and in taking the worst-case error by computing an upper bound. Analogously, in the statistical setting the proof techniques are based on the bias-variance decomposition and one usually bounds the mean square error (see [16] and references therein). On the other hand, more recent convergence rates for learning algorithms have been established: they are studied with increasing the number of samples and the proof techniques are based on the use of concentration inequalities in order to establish upper, lower and minmax bounds on the mean square error. However, the decomposition of the error is slightly different from the usual one in the statistical setting. In this Chapter, we focus on convergence rates for spectral regularization applied

to infinite dimensional inverse problems and learning theory, with respect to the noise level and the number of samples, respectively. A common thread is the fact that convergence rates have been studied in both fields under the same source conditions (the hypotheses on the sought solution) i.e. namely the Holder-type source conditions, and on the model, i.e. the polynomial decay of the linear operator [3]. Therefore, convergence rates have been provided with respect to the smoothness parameter of the ideal solution (a priori assumption) and to the parameter of the eigenvalue decay of the operator. The question naturally arises whether the above rates are comparable and, if it is the case, which relation occurs between δ and n for quantifying the difference between optimal rates in the two contexts.

We answer to this question by making use of a statistical estimator with the following two properties [87; 139]. First, it has the same upper rates of the spectral regularization considered in statistical learning: our analysis of the convergence rates of this estimator is based on the results in [16] where a comprehensive study on the convergence rates with infinite dimensional deterministic and stochastic noise is given. Second, the rates of this estimator are related to the ones of the classical spectral regularization for deterministic ill-posed inverse problems. Indeed, we prove that the expected error of this estimator given n samples is an upper bound of the error of the spectral regularization given the noise level δ, provided that a suitable relation between n and δ holds true. This allows us to convert upper rates with respect to the number of samples n to upper rates with respect to the noise level δ and, conversely, lower rates depending on δ to lower rates depending on n. Then, we compare optimal convergence rates obtained in the two contexts for the class of spectral regularized algorithms and we quantify their difference showing that they exactly match when the rank of the linear operator is finite. However, we prove that, in general, they do not correspond to each other.

The proofs of results of the current Chapter are provided in section 3.7.

3.1 Preliminaries

Both ill-posed inverse problems and statistical learning deal with a bounded linear operator A: in inverse problems A is the operator to be (approximately) inverted; in statistical learning A relates to the feature map (see Chapter 1) and it is usually known as inverse learning [17] if A is not the canonical inclusion. We consider the assumption 1 on the operator A, which we recall in the following. Let $\mathcal{X}$ be a standard Borel space endowed with a measure ν. Let $\mathcal{H}_1$ be a separable Hilbert space, $\mathcal{H}_2 := L^2(\mathcal{X}, \nu)$ and A be a bounded linear operator $A : \mathcal{H}_1 \to \mathcal{H}_2$. We assume that A is uniformly bounded, i.e. there exists a constant $c > 0$ such that

$$|Af(x)| \leq c\|f\|_{\mathcal{H}_1}, \tag{3.1}$$

for all $x \in \mathcal{X}$ and for all $f \in \mathcal{H}_1$. This assumption leads to the implication that for all x there exists an element $\phi_x \in \mathcal{H}_1$ such that

$$(Af)(x) = \langle f, \phi_x \rangle_{\mathcal{H}_1}. \tag{3.2}$$

Moreover, the range of A is a subset of $L^2(\mathcal{X}, \nu)$ and it is well known that it is a RKHS with kernel $K(x, x') = \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}_1}$ (for details see Chapter 1). Therefore, the adjoint operator $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ of A takes the form

$$A^* g = \int_{\mathcal{X}} g(x)\phi_x \, d\nu(x) \tag{3.3}$$

and the operator A^*A is given by

$$A^*A = \int_{\mathcal{X}} \langle \cdot, \phi_x \rangle_{\mathcal{H}_1} \phi_x \, d\nu(x). \tag{3.4}$$

As the operator A^*A is self-adjoint and compact, there exists an orthonormal basis consisting of eigenfunctions of A with real eigenvalues. Furthermore, the operator A^*A is of trace class. The proofs of such properties can be found in [37].

Finally, for any set $\{X_1, \ldots, X_n\} \subset \mathcal{X}$ we consider the operator $A_n : \mathcal{H}_1 \to \mathbb{R}^n$ as follows

$$(A_n f)_i := \langle f, \phi_{X_i} \rangle_{\mathcal{H}_1} \tag{3.5}$$

for $i = 1, \ldots, n$ and $f \in \mathcal{H}_1$. Its adjoint operator $A_n^* : \mathbb{R}^n \to \mathcal{H}_1$ is given by

$$A_n^* z = \frac{1}{n} \sum_{i=1}^{n} z_i \phi_{X_i} \quad \text{for} \quad z \in \mathbb{R}^n \tag{3.6}$$

and the operator $A_n^* A_n : \mathcal{H}_1 \to \mathcal{H}_1$ is given by

$$A_n^* A_n = \frac{1}{n} \sum_{i=1}^{n} \langle \cdot, \phi_{X_i} \rangle_{\mathcal{H}_1} \phi_{X_i}. \tag{3.7}$$

The assumption $\mathcal{H}_2 := L^2(\mathcal{X}, \nu)$ allows us to consider convergence rate results in both inverse problems and statistical learning theory.

3.1.1 Spectral regularization

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces and $\mathcal{B}$ be the space of bounded linear operators $\mathcal{H}_1 \to \mathcal{H}_2$. A spectral regularization is a map $\mathfrak{R} : \mathcal{B} \times \mathcal{H}_2 \times \mathbb{R}_+ \to \mathcal{H}_1$ defined by

$$\mathfrak{R}(A, y, \lambda) := s_\lambda(A^* A) A^* y, \tag{3.8}$$

where $A \in \mathcal{B}$, $y \in \mathcal{H}_2$, $\lambda \in \mathbb{R}_+$ and s_λ denotes the *regularization function* defined as follows.

Definition 13. *The regularization (or filtering) function s_λ for $\lambda > 0$ is defined on the spectrum of $A^* A$, denoted by $\tau(A^* A)$, and satisfies the following properties:*

1. *there exists a constant $D > 0$ such that*

$$\sup_{t \in \tau(A^* A)} |t s_\lambda(t)| \leq D \quad \textit{uniformly in} \quad \lambda > 0\,, \tag{3.9}$$

2. *there exists a constant* $E > 0$ *such that*

$$\sup_{\lambda>0} \sup_{t\in\tau(A^*A)} |\lambda s_\lambda(t)| \leq E \,, \tag{3.10}$$

3. *there exists* $q > 0$ *called qualification of the method and constants* $C_a > 0$ *such that*

$$\sup_{t\in\tau(A^*A)} |t^a(1 - ts_\lambda(t))| \leq C_a\lambda^a \quad \forall\ \lambda > 0 \ \textit{ and } \ 0 \leq a \leq q. \tag{3.11}$$

The idea of spectral regularization is to provide approximated solutions of a linear operator equation with noisy data. Two typical examples are the following.

- Tikhonov regularization: in this case the regularization function is given by $s_\lambda(t) = (\lambda + t)^{-1}$ and the qualification is $q = 1$.
- Truncated singular value decomposition (or spectral cut-off): in this case the regularization function is given by

$$s_\lambda(t) = \begin{cases} \frac{1}{t} \,, & \text{if } t \geq \lambda \\ 0 \,, & \text{if } t < \lambda \end{cases} \tag{3.12}$$

  and q is arbitrary.

Now we introduce the two main assumptions on the noise: the first is usually considered in the study of ill-posed inverse problems whereas the second has been considered in both the case of inverse problems and of statistical (inverse) learning.

3.1.2 Deterministic noise

Spectral regularization has been introduced in ill-posed inverse problems theory to approximately solve

$$Af = y \tag{3.13}$$

when $y \in \mathcal{H}_2$ is not known and only a noisy version y^δ of the data is available. The spectral regularized solution takes the form

$$f_\delta^\lambda := \mathfrak{R}(A, y^\delta, \lambda) = s_\lambda(A^*A)A^*y^\delta \ . \tag{3.14}$$

In this context the noisy data is infinite dimensional and the relation with the exact data y is $\|y^\delta - y\|_{\mathcal{H}_2} \leq \delta$ for some $\delta > 0$ representing the noise level. As f_δ^λ continuously depends on the data, it converges to the generalized solution $f^\dagger$. The convergence rates are studied with respect to $\delta \to 0$, i.e.

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} \in O(\delta^d) \tag{3.15}$$

where $\lambda = \lambda(\delta)$ is such that $d > 0$.

3.1.3 Stochastic noise

In the context of supervised (inverse) learning the noise is formalized differently, as we have seen in Chapter 2. In this case, instead of knowing an infinite dimensional noisy version of the data y, we assume to know a set of noisy samples $\{(X_i, Y_i)\}_{i=1}^n$. In particular, one supposes that each (X_i, Y_i) is independently drawn from a given (but unknown) probability distribution ρ on $\mathcal{X} \times \mathcal{Y}$ where the input space $\mathcal{X} \subseteq \mathbb{R}^p$ and the output space $\mathcal{Y} \subseteq \mathbb{R}$. We assume that ρ satisfies the following factorization property $\rho(X, Y) = \rho(Y|X)\nu(X)$ where ν is the marginal distribution on $\mathcal{X}$ and $\rho(\cdot|X = x)$ is the conditional distribution on $\mathcal{Y}$ for almost all $x \in \mathcal{X}$. We assume that the conditional expectation with

respect to $\rho(\cdot|\cdot)$ of Y given X is equal to

$$\mathbb{E}(Y|X = x) = Af^{\dagger}(x) = y(x) \tag{3.16}$$

and that the variance of the conditional probability is

$$Var(Y|X = x) = \sigma^2 \tag{3.17}$$

for ν-almost $x \in \mathcal{X}$, where σ is a constant. We notice that this is the case where the loss function in Chapter 2 is chosen such that the characteristic to estimate of the unknown distribution is the mean (e.g. the loss function is the square loss). Therefore, y is the regression function and it is assumed to be modeled through the linear operator A (equation (3.16)). In this way the sought solution to estimate is $f^{\dagger}$. In this setting, spectral regularization takes the form

$$\hat{f}^{\lambda}_{n,\text{learn}} = \mathfrak{R}(A_n, \mathbf{y}, \lambda) = s_{\lambda}(A_n^* A_n) A_n^* \mathbf{y}, \tag{3.18}$$

where $\mathbf{y} = (Y_1, \ldots, Y_n)^T$. The convergence rates of the estimator $\hat{f}^{\lambda}_{n,\text{learn}}$ are studied with respect to $n \to \infty$ in expectation (or in probability), i.e

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}^{\lambda}_{n,\text{learn}} - f^{\dagger}\|_{\mathcal{H}_1}) \in O\left(\frac{1}{n^{d'}}\right), \tag{3.19}$$

where $\lambda = \lambda(n)$ is such that $d' > 0$ and $\rho^{\otimes n}$ indicates the distribution tensor product.

3.2 Assumptions

In the following we give a brief review about the main results on convergence rates for spectral regularized estimators in statistical learning theory and in deterministic infinite dimensional inverse problems setting under the same Holder-type source condition. In the statistical learning setting (Holder-type) source conditions are expressed in terms of restrictions of the probability $\rho(\cdot|\cdot)$

whereas in inverse problems they are formalized as direct requirements on the form of the solution. The standard Holder-type source condition is

$$f^{\dagger} \in \omega(r, R) := \{f \in \mathcal{H}_1 \ : \ f = B^r w, \ \|w\|_{\mathcal{H}_1} \leq R\}, \tag{3.20}$$

where $B := A^*A$, $r > 0$ and $R > 0$. In the statistical framework the equivalent requirement is that $\rho(\cdot|\cdot)$ has to be such that equation (3.16) holds and the solution belongs to the set $\omega(r, R)$.

This assumption is common in both statistical learning and infinite dimensional deterministic inverse problems theory and it is interpreted as an assumption on the smoothness of the sought solution. Furthermore, especially in the statistical learning setting another assumption about the eigenvalue decay of the operator B is considered in order to improve the convergence rates. We assume that

$$\frac{c}{j^b} \leq \tau_j \leq \frac{d}{j^b} \tag{3.21}$$

where τ_j are the eigenvalues of B for each $j \in \mathbb{N}$, $j \geq 1$, $d, c > 0$ and $b > 1$. In the inverse problems literature, such an eigenvalue decay assumption is related to the so-called degree of ill-posedness of the inverse problem. In the statistical framework this assumption is given as a requirement on the probability ν which B depends on. Together with the first assumption, they can be expressed in statistical learning as a single restriction on the probability space by requiring that ρ belongs to a suitable subspace $\mathcal{M}(r, R, b)$ representing the class of models (for details see [17]). We will see in section 3.3 that this last assumption on the eigenvalue decay can not improve the convergence rates given in the deterministic infinite dimensional inverse problems setting, which, instead, are independent from the eigenvalue decay assumptions [85].

3.3 Existing convergence rates

In Table 3.1 we report a summary about the convergence rates given in inverse learning and in deterministic ill-posed inverse problems for spectral

regularization methods according to different assumptions on the operator A. Whereas convergence rates for ill-posed inverse problems are independent of the assumption on the operator A, these hypotheses are crucial to improve the convergence rates in the case of statistical learning. First, assuming a polynomial decay of the eigenvalues of the operator $B = A^*A$ with exponent $b > 1$, convergence rates improve and they become faster and faster as b increases. When b goes to 1, the rate corresponds to the one obtained without assuming any further condition to the eigenvalue decay. Second, assuming that just a finite set of eigenvalues are nonzero, i.e. A has finite rank, convergence reaches its faster rate. This rate is the limit rate achieved when b goes to infinity. Indeed, this last case can be seen as the limit case of the request of the fastest eigenvalue decay, i.e. the eigenvalues are "definitively zeros" ($\tau_j = 0$ for $j > Q$, where Q is the rank of A).

Table 3.1: Comparison of upper rates under the Holder-type source condition in equation (3.20) in statistical learning and ill-posed inverse problems with increasing n and decreasing δ, respectively.

Assumption on A	$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}^{\lambda}_{n,\text{learn}} - f^{\dagger}\|_{\mathcal{H}_1})$	$\|f^{\lambda}_{\delta} - f^{\dagger}\|_{\mathcal{H}_1}$
none	$\left(\frac{1}{n}\right)^{\frac{r}{2r+2}}$	$\delta^{\frac{2r}{2r+1}}$
eigenvalue decay (3.21)	$\left(\frac{1}{n}\right)^{\frac{r}{2r+1+\frac{1}{b}}}$	$\delta^{\frac{2r}{2r+1}}$
finite rank	$\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}$	$\delta^{\frac{2r}{2r+1}}$

3.4 A link between the number of samples n and the noise level δ

The main difference between the study of the convergence rates in statistical learning and ill-posed inverse problems with deterministic noise lies in the independent variable which the error depends on. Whereas for learning problems the independent variable is the number of examples n, for inverse

problems it is the noise level δ of infinite dimensional noisy data. The relation between the rates provided in these two settings under the same source condition is not straightforward. It is evident that there is no direct transformation between n and δ. To establish such a relation we need to introduce the following estimator

$$\hat{f}_n^\lambda := s_\lambda(A^*A)A_n^*\mathbf{y}. \tag{3.22}$$

We refer to estimator in equation (3.22) as the *hybrid* estimator as it is halfway between the spectral regularization for ill-posed problems and for statistical learning: indeed, it is composed by the infinite dimensional term $s_\lambda(A^*A)$ (which is in the definition of f_δ^λ in equation (3.14)) and the term $A_n^*\mathbf{y}$ (which is in the definition of $\hat{f}_{n,\text{learn}}^\lambda$ in equation (3.18)). This estimator has been introduced in [87; 139]. We are interested in this estimator as it has the following two properties:

(i) the error given by this estimator is always larger than the error given by the standard spectral regularized solution provided that a suitable relation between n and δ holds true;

(ii) it has the same upper rates of the spectral regularized estimator $\hat{f}_{n,\text{learn}}^\lambda$.

The first property allows us to convert upper convergence rates depending on n to upper convergence rates depending on δ and viceversa, lower convergence rates depending on δ to lower convergence rates depending on n (as we will show in Theorems 4 and 5). The second property assures that this estimator in terms of upper rate is the same as the spectral estimator in statistical learning (as we will show in section 3.5)

The first property of the hybrid estimator $\hat{f}_n^\lambda$ is summarized in the following:

Proposition 7. *Consider the spectral regularized solution f_δ^λ defined in equation* (3.14) *and the hybrid estimator $\hat{f}_n^\lambda$ defined in equation* (3.22)*. Let $\mathcal{H}_1$ be embedded in the space of square integrable functions. Let us consider n samples identically and*

independently drawn according to a distribution ρ as in section 3.1.3. Let

$$\varepsilon(\lambda) := \frac{\|f^\lambda - f^\dagger\|_{\mathcal{H}_1}}{\|s_\lambda(A^*A)A^*\|_{HS}} \tag{3.23}$$

where $f^\lambda := R(A, y, \lambda) = s_\lambda(A^*A)A^*y$ and $\|\cdot\|_{HS}$ denotes the Hilbert Schmidt norm. For each $n \in \mathbb{N}$ there exists a function

$$\Delta(n, \lambda) = \frac{1}{\sqrt{\frac{\sigma^2}{n} + \varepsilon(\lambda)^2} + \varepsilon(\lambda)} \frac{\sigma^2}{n}, \tag{3.24}$$

such that for each $0 < \delta \leq \Delta(n, \lambda)$ and infinite dimensional noisy data y^δ such that $\|y^\delta - y\|_{\mathcal{H}_2} \leq \delta$, the following inequality holds

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1}^2 \leq \mathbb{E}_{\rho^{\otimes n}}\left(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2\right). \tag{3.25}$$

Conversely, for each $\delta > 0$ there exists a function

$$N(\delta, \lambda) = \frac{\sigma^2}{\delta^2 + 2\delta\varepsilon(\lambda)} \tag{3.26}$$

such that for each $n \in \mathbb{N}$ such that $n \leq N(\delta, \lambda)$ inequality (3.25) applies.

Thanks to the result in Proposition 7 we can relate a given upper convergence rate computed with respect to n (for the hybrid estimator $\hat{f}_n^\lambda$) to the one computed with respect to δ (for the spectral regularized solution f_δ^λ). From now on, in order to express asymptotic behaviors we make use of the Landau symbols O, Ω and Θ.

Theorem 4. *Let the upper rate of the hybrid estimator $\hat{f}_n^\lambda$ defined in equation (3.22) be equal to $n^{-\alpha}$, as $n \to \infty$, for a given $\alpha > 0$, i.e.*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2) \in O\left(\frac{1}{n}\right)^\alpha, \tag{3.27}$$

for a given $\lambda = \lambda_n = \Theta(n^{-p})$, with $p > 0$ and $\varepsilon(\lambda) = \Theta(\lambda^\gamma)$, with $\gamma > 0$. Then the upper rate of the estimator f_δ^λ defined in equation (3.14) with respect to the noise

level $\delta \to 0$ *is given by*

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1}^2 \in O\left(\delta^{\min\left(2\alpha, \frac{\alpha}{1-p\gamma}\right)}\right), \tag{3.28}$$

where y^δ *is such that* $\|y^\delta - y\|_{\mathcal{H}_2} \leq \delta$, *and* $\lambda = \lambda_\delta$ *(is defined in Lemma 7 in the section 3.7 of Proofs and it) has the following rate*

$$\lambda_\delta \in \Theta\left(\delta^{\min\left(2p, \frac{p}{1-p\gamma}\right)}\right). \tag{3.29}$$

Now we give the converse result on lower rates.

Theorem 5. *Let the lower rate of the spectral regularized solution* f_δ^λ, *defined in equation* (3.14), *be equal to* δ^α, *as* $\delta \to 0$, *for a given* $\alpha > 0$, *i.e.*

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1}^2 \in \Omega(\delta^\alpha), \tag{3.30}$$

where y^δ *is such that* $\|y^\delta - y\|_{\mathcal{H}_2} \leq \delta$, $\lambda = \lambda_\delta = \Theta(\delta^{p^*})$, *with* $p^* > 0$ *and* $\varepsilon(\lambda) = \Theta(\lambda^\gamma)$, *with* $\gamma > 0$. *Then the lower rate of the hybrid estimator* $\hat{f}_n^\lambda$ *defined in equation* (3.22) *with respect to the number of samples* $n \to \infty$ *is given by*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2) \in \Omega\left(n^{-\max\left(\frac{\alpha}{2}, \frac{\alpha}{1+p^*\gamma}\right)}\right) \tag{3.31}$$

where $\lambda = \lambda_n$ *(is defined in Lemma 8 in section 3.7 of Proofs and) it has the following rate*

$$\lambda_n \in \Theta\left(n^{-\max\left(\frac{p^*}{2}, \frac{p^*}{1+p^*\gamma}\right)}\right). \tag{3.32}$$

3.5 Upper rates of the hybrid estimator

We now present the result on the upper rates of the hybrid estimator $\hat{f}_n^\lambda$ under the classical source condition in equation (3.20) and according to the assumption on the operator A. If we do not make assumption on the singular values of A, we have the following result.

Lemma 2. *Let $\hat{f}_n^\lambda$ be defined in equation* (3.22) *and let the model be described by equations in* (3.16) *and* (3.17)*. Under the source condition in equation* (3.20) *we have*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2) \in O\left(\left(\frac{1}{n}\right)^{\frac{2r}{2r+2}}\right), \tag{3.33}$$

with $\lambda \in \Theta\left((\frac{1}{n})^{\frac{1}{2r+2}}\right)$.

Under the hypothesis of the polynomial eigenvalue decay of A^*A we have the following result.

Lemma 3. *Let $\hat{f}_n^\lambda$ be defined in equation* (3.22) *and let the model be described by equations in* (3.16) *and* (3.17)*. Under the source conditions in equations* (3.20) *and* (3.21) *we have*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2) \in O\left(\left(\frac{1}{n}\right)^{\frac{2r}{2r+1+\frac{1}{b}}}\right), \tag{3.34}$$

with $\lambda \in \Theta\left((\frac{1}{n})^{\frac{1}{2r+1+\frac{1}{b}}}\right)$.

Finally, in the case the operator A has finite rank we have the following:

Lemma 4. *Let $\hat{f}_n^\lambda$ be defined in equation* (3.22) *and let the model be described by equations in* (3.16) *and* (3.17)*. Under the source condition in equation* (3.20) *and under the hypothesis that the operator A^*A has finite rank ($\tau_j > 0$ for all $0 < j \leq Q$) we have*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}^2) \in O\left(\left(\frac{1}{n}\right)^{\frac{2r}{2r+1}}\right), \tag{3.35}$$

with $\lambda \in \Theta\left((\frac{1}{n})^{\frac{1}{2r+1}}\right)$.

We remark that the upper rates given in equations (3.34), (3.33) and (3.35) are the same ones of the classical spectral estimator $\hat{f}_{n,\text{learn}}^\lambda$ defined in equation (3.18) (see Table 3.1).

3.6 Conversion of convergence rates

In section 3.5 we have shown that the estimator $\hat{f}_n^\lambda$ defined in equation (3.22) has the same upper rates of the standard statistical learning estimator $\hat{f}_{n,\text{learn}}^\lambda$ defined in equation (3.18) for the same choice of the sequence λ_n. This allows us to use Theorem 4 to transform the upper rates depending on n in Table 3.1 to upper rates for the classical spectral regularization depending on δ. Let f_δ^λ be defined in equation (3.14) and γ be defined as in Theorem 4.

3.6.1 Upper rates

We now focus on the hybrid estimator and on its upper rates in the three cases considered in section 3.5. We have the following results.

Corollary 5. *Consider the hybrid estimator $\hat{f}_n^\lambda$ and its upper rates (Lemma 2, Lemma 3 and Lemma 4). Then, thanks to Theorem 4 for the spectral regularization solution f_δ^λ we have the following cases.*

1. Under assumption in equation (3.20)

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} \in \begin{cases} O\left(\delta^{\frac{2r}{2r+2}}\right) & , \quad \gamma \geq r+1 \quad \text{and} \quad \lambda \in \Theta\left(\delta^{\frac{2}{2r+2}}\right) \\ O\left(\delta^{\frac{r}{2r+2-\gamma}}\right) & , \quad \gamma < r+1 \quad \text{and} \quad \lambda \in \Theta\left(\delta^{\frac{1}{2r+2-\gamma}}\right) \end{cases} . \tag{3.36}$$

2. Under assumptions in equations (3.20) *and* (3.21)

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} \in \begin{cases} O\left(\delta^{\frac{2r}{2r+1+\frac{1}{b}}}\right), & \gamma \geq r+\frac{1}{2}+\frac{1}{2b} \text{ and } \lambda \in \Theta\left(\delta^{\frac{2}{2r+1+\frac{1}{b}}}\right) \\ O\left(\delta^{\frac{r}{2r+1+\frac{1}{b}-\gamma}}\right), & \gamma < r+\frac{1}{2}+\frac{1}{2b} \text{ and } \lambda \in \Theta\left(\delta^{\frac{1}{2r+1+\frac{1}{b}-\gamma}}\right) \end{cases} . \tag{3.37}$$

3. Under assumption in equation (3.20) *and assuming the rank of A^*A is finite*

$$\|f_\delta^\lambda - f^\dagger\|_{\mathcal{H}_1} \in \begin{cases} O\left(\delta^{\frac{2r}{2r+1}}\right) & , \quad \gamma \geq r+\frac{1}{2} \quad \text{and} \quad \lambda \in \Theta\left(\delta^{\frac{2}{2r+1}}\right) \\ O\left(\delta^{\frac{r}{2r+1-\gamma}}\right) & , \quad \gamma < r+\frac{1}{2} \quad \text{and} \quad \lambda \in \Theta\left(\delta^{\frac{1}{2r+1-\gamma}}\right) \end{cases} . \tag{3.38}$$

In all these three cases, if γ is sufficiently large, the upper rates are independent of γ. Otherwise, they are bounded from below by the γ-independent upper rates. We remark that in the first two cases in equations (3.36) and (3.37) the upper rates are always slower than the classical optimal one, i.e. $O\left(\delta^{\frac{2r}{2r+1}}\right)$ [43]. This rate can be achieved only in the case A^*A has finite rank (see equation (3.38)). The proof of Corollary 5 is omitted since it is a straightforward application of Theorem 4, using results in Lemmas 2, 3 and 4.

3.6.2 Lower rates

Now we exploit Theorem 5 to convert the lower rate of the spectral regularized solution f_δ^λ in a lower rate depending on n for the hybrid estimator $\hat{f}_n^\lambda$. As shown in Table 3.1, the lower rate of f_δ^λ depends only on the source condition in equation (3.20) and it is independent of the eigenvalue decay of A^*A. Therefore, under assumption in equation (3.20) we have the following:

Corollary 6. *Consider the spectral regularized solution f_δ^λ and its lower rate $\Omega\left(\delta^{\frac{2r}{2r+1}}\right)$, achieved with $\lambda = \Theta\left(\delta^{\frac{2}{2r+1}}\right)$. Then, thanks to Theorem 5 for the hybrid estimator $\hat{f}_n^\lambda$ we have*

$$\mathbb{E}_{\rho^{\otimes n}}(\|\hat{f}_n^\lambda - f^\dagger\|_{\mathcal{H}_1}) \in \begin{cases} \Omega\left(\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}\right), & \gamma \geq r+\frac{1}{2} \text{ and } \lambda \in \Theta\left(\left(\frac{1}{n}\right)^{\frac{1}{2r+1}}\right) \\ \Omega\left(\left(\frac{1}{n}\right)^{\frac{r}{r+\gamma+\frac{1}{2}}}\right), & \gamma < r+\frac{1}{2} \text{ and } \lambda \in \Theta\left(\left(\frac{1}{n}\right)^{\frac{1}{r+\gamma+\frac{1}{2}}}\right) \end{cases}. \tag{3.39}$$

We remark that rates obtained in Corollary 6 are lower bounds of the classical lower rates of the spectral regularized estimator $\hat{f}_{n,\text{learn}}^\lambda$. The proof of Corollary 6 is omitted since it is straightforward from Theorem 5.

In the case of Tikhonov regularization and truncated singular value decomposition it is readily to prove that $\gamma \geq r + \frac{1}{2}$. We show in Table 3.2 a summary of the conversion of upper rates and lower rates in this case.

Table 3.2: Conversion of convergence rates in the case $\gamma = r + \frac{1}{2}$ under the Holder-type source condition in equation (3.20).

	Upper rates (from n to δ)			
Assumption on A	Hybrid	λ_n	λ_δ	Spectral reg.
none	$\left(\frac{1}{n}\right)^{\frac{r}{2r+2}}$	$\left(\frac{1}{n}\right)^{\frac{1}{2r+2}}$	$\delta^{\frac{1}{r+\frac{3}{2}}}$	$\delta^{\frac{r}{r+\frac{3}{2}}}$
eigenvalue decay (3.21)	$\left(\frac{1}{n}\right)^{\frac{r}{2r+1+\frac{1}{b}}}$	$\left(\frac{1}{n}\right)^{\frac{1}{2r+1+\frac{1}{b}}}$	$\delta^{\frac{1}{r+\frac{1}{2}}+\frac{1}{b}}$	$\delta^{\frac{r}{r+\frac{1}{2}+\frac{1}{b}}}$
finite rank	$\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}$	$\left(\frac{1}{n}\right)^{\frac{1}{2r+1}}$	$\delta^{\frac{2}{2r+1}}$	$\delta^{\frac{2r}{2r+1}}$
	Lower rates (from δ to n)			
Assumption on A	Spectral reg.	λ_δ	λ_n	Hybrid
none	$\delta^{\frac{2r}{2r+1}}$	$\delta^{\frac{2}{2r+1}}$	$\left(\frac{1}{n}\right)^{\frac{1}{2r+1}}$	$\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}$

We remark that, in the finite rank hypothesis, the rates of the hybrid estimator and the spectral regularization match each other and, in this case, the number of samples n turns out to be inversely proportional to the noise level δ^2. However, this is not true if the rank of A^*A is not finite: in such a case, convergence rates given in statistical learning are weaker than the ones given for ill-posed deterministic inverse problems. Indeed, the conversion of statistical learning rates yields slower rates than the classical $\delta^{\frac{2r}{2r+1}}$ resulting from the inverse problems theory [43]. The fact that learning rates are generally slower should not be surprising: as the noise level δ goes to zero the assumption $\|y - y^\delta\|_{\mathcal{H}_2} \leq \delta$ implies that there exists a subsequence of noisy data that converges to the exact data y on the set $\mathcal{X}$ almost everywhere; by contrast, taking the set of samples $\{(X_i, Y_i)\}_{i=1}^n$ as n goes to infinity is an assumption on the set $\{X_i\}_{i=1}^n \subset \mathcal{X}$, which is at most countable.

Remark 4. *It is worth noticing that the techniques to bound the errors in learning and inverse problem settings in order to prove convergence rates are different: the*

errors are split in different ways and bounded with different techniques (in learning theory concentration inequalities are exploited [9; 17; 115]). We showed that the learning and inverse problems convergence rates coincide when the operator A has finite rank. We give an alternative proof of the optimal learning upper rate under the assumption that the input space is a discrete set, i.e. $\mathcal{X}$ is a set of a finite number Q of elements. This case applies in learning problems with categorical variables (see e.g. [51]). This case is equivalent to consider the case of finite rank: it is treated in [30] and it corresponds to the case in which $\mathcal{H}_K$ is finite dimensional, therefore the effective dimension $\mathcal{N}(\lambda) \leq Q$ and the rate is retrieved by taking $b = +\infty$. In our alternative proof, which we report for the sake of completeness in section 3.7 (proof of Proposition 8) we split the convergence error of learning in a more similar way to the one used in the convergence error analysis in the inverse problem setting. The idea is to re-organize the discrete data in order to write the spectral regularized estimator depending on the sample average of the responses Y_i: this procedure is possible under the discrete input space hypothesis, equivalently under the hypothesis that A has finite rank. In this way we highlight that the error between the spectral regularized estimator and the spectral regularization applied on the true samples $A_n f^\dagger$ can be bounded in expectation by a term which is of the order $\sqrt{\frac{\sigma^2}{\lambda n}}$, i.e.

$$\mathbb{E}(\|\hat{f}^{\lambda}_{n,\text{learn}} - \mathfrak{R}(A_n, A_n f^\dagger, \lambda)\|_{\mathcal{H}_1}) \in O\left(\sqrt{\frac{\sigma^2}{\lambda n}}\right). \tag{3.40}$$

The corresponding error in the inverse problem setting between the spectral regularized solution and the spectral regularization applied on the noise-free data y is bounded by a term of the order $\frac{\delta}{\sqrt{\lambda}}$, i.e.

$$\|f^{\lambda}_{\delta} - \mathfrak{R}(A, A f^\dagger, \lambda)\|_{\mathcal{H}_1} \in O\left(\frac{\delta}{\sqrt{\lambda}}\right). \tag{3.41}$$

This confirms that δ is proportional to $\sqrt{\frac{\sigma^2}{n}}$, where $\frac{\sigma^2}{n}$ represents the variance of the sample average.

3.7 Proofs

We prove Proposition 7, Theorems 4 and 5 and the upper convergence rates given in Lemmas 2, 3 and 4.

The first property of the hybrid estimator depends on the fact that it can be seen as an empirical version of the standard spectral regularization. To see this, we now introduce a linear regularization operator family L^λ as follows. We consider two positive and finite measures ν and μ. We suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces of square integrable functions on $\mathcal{T}$ with respect to the measure μ, $L^2(\mathcal{T}, \mu)$, and on $\mathcal{X}$ with respect to the measure ν, $L^2(\mathcal{X}, \nu)$, respectively. Let the linear regularization operator family $L^\lambda : \mathcal{H}_2 \to \mathcal{H}_1$, with $\lambda > 0$ be of the form

$$L^\lambda y = \int_{\mathcal{X}} \ell_x^\lambda \, y(x) \, d\nu(x) \, , \tag{3.42}$$

where $\ell_x^\lambda \in \mathcal{H}_1$, $\ell_x^\lambda(t) := \ell^\lambda(x, t)$ and $\ell^\lambda(\cdot, t) \in \mathcal{H}_2$ for each $x \in \mathcal{X}$ and for each $t \in \mathcal{T}$. Thanks to this last assumption the integral in equation (3.42) is finite. Moreover, we assume $\sup_{t \in \mathcal{T}} \|\ell^\lambda(\cdot, t)\|_{\mathcal{H}_2} < +\infty$. Such an assumption implies that L^λ is uniformly bounded and then for each $y \in \mathcal{H}_2$, $L^\lambda y$ is bounded in supremum norm which assures that $L^\lambda y \in \mathcal{H}_1$. We denote with F^λ the regularized solution given by the linear regularization operator L^λ applied to the noise free data y, i.e.

$$F^\lambda = L^\lambda y \, , \tag{3.43}$$

and with F_δ^λ the regularized solution given by the noisy data y^δ, i.e.

$$F_\delta^\lambda = L^\lambda y^\delta \, , \tag{3.44}$$

when $\|y - y^\delta\|_{\mathcal{H}_2} \leq \delta$. We introduce the following estimator computed from a set of discrete data as follows

$$\hat{F}_n^\lambda = L_{\mathbf{x}}^\lambda \mathbf{y} = \frac{1}{n} \sum_{i=1}^{n} \ell_{X_i}^\lambda \, Y_i \tag{3.45}$$

where $\mathbf{x} = (X_1, \ldots, X_n)^T$ and $\mathbf{y} = (Y_1, \ldots, Y_n)^T$ denote the samples. We

consider the model assumptions in equations (3.16) and (3.17). The spectral regularization can be seen as a special case of the linear regularization L^λ defined in equation (3.44) by setting

$$\ell_x^\lambda = s_\lambda(A^*A)\phi_x, \tag{3.46}$$

with $x \in \mathcal{X}$. Indeed, hypotheses on ℓ_x^λ are satisfied since $\sup_{t\in\mathcal{T}} \|\phi(\cdot,t)\|_{\mathcal{H}_2} < +\infty$. In this case we have

$$f_\delta^\lambda = F_\delta^\lambda \quad \text{and} \quad \hat{f}_n^\lambda = \hat{F}_n^\lambda. \tag{3.47}$$

We start by proving an inequality which will be used in the proof of the result in Proposition 7. In what follows, to make it easier the writing, we do not write the subscript of the norms and we denote with $\mathbb{E}$ the mean computed with respect to the measure $\rho^{\otimes n}$.

Lemma 5. *Let $\hat{F}_n^\lambda$ be defined in equation* (3.45)*. Under assumptions in equations* (3.16) *and* (3.17) *we have*

$$\mathbb{E}(\|\hat{F}_n^\lambda - f^\dagger\|^2) \geq \frac{\sigma^2}{n}\|L^\lambda\|_{HS}^2 + \|F^\lambda - f^\dagger\|^2, \tag{3.48}$$

where $\|\cdot\|_{HS}$ denotes the Hilbert Schmidt norm.

Proof. Denote with ϵ_n the difference between the estimate $\hat{F}_n^\lambda$ obtained with n samples and the sought solution $f^\dagger$. For any $t \in \mathcal{T}$ we have

$$\begin{aligned}
\epsilon_n^2(t) &= \left(\frac{1}{n}\sum_{i=1}^n \ell_{X_i}^\lambda(t)Y_i - f^\dagger(t)\right)^2 \\
&= \frac{1}{n^2}\sum_{i,j=1}^n \ell_{X_i}^\lambda(t)Y_i\ell_{X_j}^\lambda(t)Y_j - \frac{2}{n}f^\dagger(t)\sum_{i=1}^n \ell_{X_i}(t)Y_i + (f^\dagger(t))^2. \qquad (3.49)
\end{aligned}$$

By integrating over $\mathcal{Y}^n$, we get

$$\begin{aligned}\int_{\mathcal{Y}^n} \epsilon_n^2(t)\; d\rho(\cdot|\cdot)^{\otimes n} &= \frac{1}{n^2}\sum_{i=1}^{n}(\ell_{X_i}^{\lambda}(t))^2\sigma^2 + \frac{1}{n^2}\sum_{i,j=1}^{n}\ell_{X_i}^{\lambda}(t)\ell_{X_j}^{\lambda}(t)\; y(X_i)y(X_j) \\ &\quad -\frac{2}{n}\; f^{\dagger}(t)\sum_{i=1}^{n}\ell_{X_i}^{\lambda}(t)y(X_i) + (f^{\dagger}(t))^2, \end{aligned} \tag{3.50}$$

where $d\rho(\cdot|\cdot)^{\otimes n} = d\rho(Y_1|X_1)\cdots d\rho(Y_n|X_n)$ and by using that $\rho(\cdot|\cdot)$ is a probability measure on $\mathcal{Y}$. Then, by integrating over $\mathcal{X}^n$ we obtain

$$\begin{aligned}\int_{\mathcal{X}^n}\int_{\mathcal{Y}^n} \epsilon_n^2(t)\; d\rho(\cdot|\cdot)^{\otimes n}d\nu^{\otimes n} &= \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{X}}(\ell_{X_i}^{\lambda}(t))^2 d\nu(X_i) \\ &\quad +\frac{1}{n^2}\sum_{i=1}^{n^2-n}\left(\int_{\mathcal{X}}\ell_{X_i}^{\lambda}(t)y(X_i)d\nu(X_i)\right)^2 \\ &\quad +\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{X}}\left(\ell_{X_i}^{\lambda}(t)y(X_i)\right)^2 d\nu(X_i) \\ &\quad -\frac{2}{n}f^{\dagger}(t)\sum_{i=1}^{n}\int_{\mathcal{X}}\ell_{X_i}^{\lambda}(t)y(X_i)d\nu(X_i) + (f^{\dagger}(t))^2 \\ &\geq \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{X}}(\ell_{X_i}^{\lambda}(t))^2 d\nu(X_i) \\ &\quad +\frac{1}{n^2}\sum_{i=1}^{n^2}\left(\int_{\mathcal{X}}\ell_{X_i}^{\lambda}(t)y(X_i)d\nu(X_i)\right)^2 \\ &\quad -\frac{2}{n}f^{\dagger}(t)\sum_{i=1}^{n}\int_{\mathcal{X}}\ell_{X_i}^{\lambda}(t)y(X_i)d\nu(X_i) + (f^{\dagger}(t))^2 \\ &= \frac{\sigma^2}{n}\int_{\mathcal{X}}(\ell_X^{\lambda}(t))^2 d\nu(X) \\ &\quad +\left(F^{\lambda}(t)\right)^2 - 2f^{\dagger}(t)F^{\lambda}(t) + (f^{\dagger}(t))^2\,, \end{aligned} \tag{3.51}$$

where we used that ν is a probability measure on $\mathcal{X}$. Therefore, we have

$$\begin{aligned}\mathbb{E}\left(\|\hat{F}_n^{\lambda} - f^{\dagger}\|^2\right) &\geq \int_{\mathcal{T}}\frac{\sigma^2}{n}\int_{\mathcal{X}}(\ell_X^{\lambda}(t))^2\; d\nu(X) + \left(F^{\lambda}(t) - f^{\dagger}(t)\right)^2 d\mu(t) \\ &= \frac{\sigma^2}{n}\|L^{\lambda}\|_{HS}^2 + \|F^{\lambda} - f^{\dagger}\|^2\,, \end{aligned} \tag{3.52}$$

as required. □

In the following we prove the result in Proposition 7.

Proof of Proposition 7. We start from the result of Lemma 5. Easy manipulation of formula in equation (3.48) leads to

$$\sqrt{\mathbb{E}\left(\|\hat{F}_n^\lambda - f^\dagger\|^2\right)} \geq \Delta(n,\lambda)\|L^\lambda\|_{HS} + \|F^\lambda - f^\dagger\| \tag{3.53}$$

where $\Delta(n,\lambda)$ is defined as in equation (3.24). For each $\delta > 0$, let y^δ such that $\|y^\delta - y\| \leq \delta$, then a simple calculation gives

$$\|F_\delta^\lambda - f^\dagger\| \leq \delta\|L^\lambda\| + \|F^\lambda - f^\dagger\|. \tag{3.54}$$

Further, for each $\delta \leq \Delta(n,\lambda)$ we have

$$\sqrt{\mathbb{E}\left(\|\hat{F}_n^\lambda - f^\dagger\|^2\right)} \geq \delta\|L^\lambda\| + \|F^\lambda - f^\dagger\| \tag{3.55}$$

as $\|\cdot\|_{HS} \geq \|\cdot\|$. From equations (3.54) and (3.55) we obtain $\forall\ \delta \leq \Delta(n,\lambda)$

$$\|F_\delta^\lambda - f^\dagger\|^2 \leq \mathbb{E}\left(\|\hat{F}_n^\lambda - f^\dagger\|^2\right) \tag{3.56}$$

for each y^δ for which $\|y^\delta - y\| \leq \delta$.

Conversely, let $\delta > 0$. For each $n \leq N(\delta,\lambda)$, with $N(\lambda,\delta)$ defined by equation (3.26) we have

$$\delta \leq \Delta(n,\lambda) \tag{3.57}$$

and so the thesis is proved. □

Functions $\Delta(n,\lambda)$ and $N(\delta,\lambda)$ express the dependency between the noisy level δ and the number of samples n. To make explicit this dependency we need to specify the rate of convergence of $\lambda \to 0$ both considered as a function of δ and n. For the sake of convenience, we introduce the following

Notation 1. *For any given λ_n we define*

$$\tilde{\delta}(n) := \Delta(n, \lambda_n) \ . \tag{3.58}$$

Conversely, for any given λ_δ we define

$$\tilde{n}(\delta) := \lfloor N(\delta, \lambda_\delta) \rfloor \ , \tag{3.59}$$

where the symbol $\lfloor \cdot \rfloor$ denotes the integer part.

Lemma 6. *Let $\varepsilon(\lambda) \in \Theta(\lambda^\gamma)$, with $\gamma \geq 0$. If $\lambda_n \in \Theta(n^{-p})$, with $p > 0$, then*

$$\tilde{\delta}(n) \in \Theta\left(n^{-\max(\frac{1}{2}, 1-p\gamma)}\right) \ . \tag{3.60}$$

If $\lambda_\delta \in \Theta(\delta^{p^*})$, with $p^* > 0$, then

$$\tilde{n}(\delta) \in \Theta\left(\delta^{-\min(2, p^*\gamma+1)}\right) \ . \tag{3.61}$$

Proof. The equation (3.60) follows from the definition of $\tilde{\delta}$ and from hypotheses $\lambda_n \in \Theta(n^{-p})$ and $\varepsilon(\lambda) \in \Theta(\lambda^\gamma)$. In the same way the equation (3.61) follows from the definition of $\tilde{n}$ and from hypotheses $\lambda_\delta \in \Theta(\delta^{p^*})$ and $\varepsilon(\lambda) \in \Theta(\lambda^\gamma)$. □

Lemma 7. *Given λ_n there exists a unique λ_δ such that*

$$\tilde{\delta} \circ \tilde{n} = id_{\Im(\tilde{\delta})}, \tag{3.62}$$

where $id_{\Im(\tilde{\delta})}$ indicates the identity on the set $\Im(\tilde{\delta}) = \{\delta > 0 \mid \frac{\sigma^2}{\delta^2 + 2\delta\varepsilon(\lambda_\delta)} \in \mathbb{N}\}$ and

$$\Lambda^n = \Lambda^\delta \circ \tilde{\delta}, \tag{3.63}$$

where $\Lambda^n : \mathbb{N} \to \mathbb{R}$ and $\Lambda^\delta : \mathbb{R} \to \mathbb{R}$ are such that $\lambda_n = \Lambda^n(n)$ and $\lambda_\delta = \Lambda^\delta(\delta)$. Furthermore,

$$\tilde{n} \circ \tilde{\delta} = id_{\mathbb{N}}. \tag{3.64}$$

Proof. The existence and uniqueness of λ_δ such that equations (3.62) and (3.63) are verified follow by defining $\lambda_\delta := \Lambda^n(\tilde{n}(\delta))$. With straightforward calculus it can be verified that equation (3.63) implies equation (3.64). □

Similarly, we give the converse result.

Lemma 8. *Given λ_δ, there exists a unique λ_n such that*

$$\tilde{n} \circ \tilde{\delta} = id_{\mathbb{N}}$$

and

$$\Lambda^\delta = \Lambda^n \circ \tilde{n} \tag{3.65}$$

where we have used the same notation of Lemma 7. Furthermore,

$$\tilde{\delta} \circ \tilde{n} = id_{\Im(\tilde{\delta})}. \tag{3.66}$$

The proof is analogous to the one of Lemma 7 by defining $\lambda_n = \Lambda^\delta(\tilde{\delta}(n))$.

Now we prove the main theorems. In the following we prove the result in Theorem 4.

Proof of Theorem 4. Given $\lambda_n = \Lambda^n(n)$, we define $\lambda_\delta = \Lambda^\delta(\delta)$ according to Lemma 7, so that equations (3.62) and (3.63) hold. The rate of λ_δ given in equation (3.29) can be found by using the hypothesis $\lambda_n = \Theta(n^{-p})$ and Lemma 6. Now we prove equation (3.28). Thanks to Proposition 7 and Lemma 6, for each $\lambda > 0$ and $\delta > 0$ there exists $\tilde{n}(\delta)$ such that for all $n \leq \tilde{n}(\delta)$

$$\|F_\delta^\lambda - f^\dagger\|^2 \leq \mathbb{E}(\|\hat{F}_n^\lambda - f^\dagger\|^2). \tag{3.67}$$

Let $n = \tilde{n}(\delta)$, then

$$\|F_\delta^\lambda - f^\dagger\|^2 \leq \mathbb{E}(\|\hat{F}_{\tilde{n}(\delta)}^\lambda - f^\dagger\|^2). \tag{3.68}$$

Let $\lambda = \lambda_\delta$. Then there exist $n_0 \in \mathbb{N}$ and $M > 0$ such that

$$\|F_\delta^{\lambda_\delta} - f^\dagger\|^2 \leq \mathbb{E}(\|\hat{F}_{\tilde{n}(\delta)}^{\Lambda^\delta(\delta)} - f^\dagger\|^2) = \mathbb{E}(\|\hat{F}_{\tilde{n}(\delta)}^{\Lambda^n(\tilde{n}(\delta))} - f^\dagger\|^2) \leq M \left(\frac{1}{\tilde{n}(\delta)}\right)^\alpha, \tag{3.69}$$

for all $\tilde{n}(\delta) > n_0$. From equations (3.29) and (3.69), and by using Lemma 6 we obtain

- if $p\gamma \geq \frac{1}{2}$ then $\lambda_\delta = \Theta(\delta^{2p})$, therefore from Proposition 7 we have $\tilde{n}(\delta) \in \Theta(\delta^2)$ and from equation (3.69) we obtain $\|F_\delta^\lambda - f^\dagger\|^2 \in O(\delta^{2\alpha})$
- if $p\gamma < \frac{1}{2}$ then $\lambda_\delta \in \Theta\left(\delta^{\frac{p}{1-p\gamma}}\right)$, therefore from Proposition 7 we have $\tilde{n}(\delta) \in \Theta\left(\delta^{\frac{1}{1-p\gamma}}\right)$ and from equation (3.69) we obtain $\|F_\delta^\lambda - f^\dagger\|^2 \in O\left(\delta^{\frac{\alpha}{1-p\gamma}}\right)$.

This completes the proof. □

We give the proof of the result in Theorem 5.

Proof of Theorem 5. The proof exploits a similar argument to the one used for Theorem 4. Given $\lambda_\delta = \Lambda^\delta(\delta)$, by defining $\lambda_n = \Lambda^n(n)$ according to Lemma 8, it can be proved that the rate of λ_n is given by equation (3.32). To prove equation (3.31) one has to reverse the role of n and δ in the proof of Theorem 4 and use Proposition 7 and hypothesis in equation (3.30). In such a way one obtains that for each $n \in \mathbb{N}$, there exist $\delta_0 > 0$ and $M' > 0$ such that

$$\mathbb{E}(\|\hat{F}_n^{\lambda_n} - f^\dagger\|^2) \geq \|F_{\tilde{\delta}(n)}^{\Lambda^\delta(\tilde{\delta}(n))} - f^\dagger\|^2 \geq M'(\tilde{\delta}(n))^\alpha , \tag{3.70}$$

for all $\tilde{\delta}(n) < \delta_0$. The thesis follows from equations (3.32), (3.70) and Lemma 6. □

Now we provide the proofs of upper rates of the hybrid estimator. We remark that $\mathbb{E}(\|\hat{f}_n^\lambda - f^\dagger\|^2)$ satisfies the bias-variance decomposition as follows

$$\mathbb{E}(\|\hat{f}_n^\lambda - f^\dagger\|^2) = B(\hat{f}_\lambda)^2 + \mathbb{E}(\|\hat{f}_\lambda - \mathbb{E}(\hat{f}_\lambda)\|^2), \tag{3.71}$$

where $B(\hat{f}_\lambda) := \|\mathbb{E}(\hat{f}_n^\lambda) - f^\dagger\|$ is the bias term and $\mathbb{E}(\hat{f}_n^\lambda) = f^\lambda$. Under the source condition in equation (3.20) the bias term can be bounded by

$$B(\hat{f}_\lambda) \leq C_r \lambda^r R, \tag{3.72}$$

where C_r is the constant of the property in equation (3.11) of the regularization function s_λ. Hereafter, we consider $r \leq q$, where q is the qualification of the method. The estimation of the variance term needs more manipulations. In detail, to bound the variance term we follow the argument given in [16] where a more general mixed type noise model is considered and the stochastic part of the noise is modeled as a Hilbert-space process. In particular we follow the argument in the section 4.3 in [16]. We consider a Hilbert-space noise process $\tilde{\epsilon}$ such that $A^*\tilde{\epsilon} = A_n^*\mathbf{y} - A^*Af^\dagger$. The noise $\epsilon = \tilde{\sigma}\tilde{\epsilon}$, where $\tilde{\sigma} = \frac{\sqrt{C}}{\sqrt{n}}$ with C a constant depending on the variance σ^2, satisfies the assumption of the Theorem 3 in [16]. Then, we have

$$\mathbb{E}(\|\hat{f}_n^\lambda - \mathbb{E}(\hat{f}_n^\lambda)\|^2) = \mathbb{E}(\|s_\lambda(A^*A)A^*\tilde{\sigma}\epsilon\|^2). \tag{3.73}$$

In the following we provide proofs of results in Lemmas 2, 3 and 4. The proofs mainly consist of bounding the term in equation (3.73) in different ways in according to the hypothesis on the eigenvalue decay. We start to prove Lemma 2.

Proof of Lemma 2. From equation (3.73) we have

$$\begin{aligned}\mathbb{E}(\|\hat{f}_n^\lambda - \mathbb{E}(\hat{f}_n^\lambda)\|^2) \leq \frac{C}{n}\sum_{j:\tau_j\geq\lambda} s_\lambda^2(\tau_j)\tau_j &\leq \frac{C}{n}\left(\sup_{\tau_j\in\tau(A^*A)} s_\lambda^2(\tau_j)\right)\sum_{j:\tau_j\geq\lambda}\tau_j \\ &\leq \frac{C}{n}\frac{E^2}{\lambda^2}C',\end{aligned} \tag{3.74}$$

where we have used the property in equation (3.9) of the regularization function s_λ and the fact that the operator A^*A is of trace class where C' represents a constant which bounds the trace norm of A^*A. Therefore, under assumption

in equation (3.20) we obtain

$$\mathbb{E}(\|\hat{f}_n^\lambda - f^\dagger\|^2) \leq C_r^2 \lambda^{2r} R^2 + \frac{CE^2C'}{n\lambda^2}. \tag{3.75}$$

By balancing terms in the r.h.s. of equation (3.75) we have the thesis.

□

Now we prove the result in Lemma 3.

Proof of Lemma 3. Under assumption in equation (3.21) we have the following bound

$$\mathbb{E}(\|\hat{f}_n^\lambda - \mathbb{E}(\hat{f}_n^\lambda)\|^2) = \mathbb{E}(\|s_\lambda(A^*A)A^*\tilde{\sigma}\epsilon\|^2) \leq \frac{C}{n} L \frac{1}{\lambda^2} \int_0^\lambda \beta^{-\frac{1}{b}} d\beta = \frac{C}{n} L \frac{1}{\lambda^{1+\frac{1}{b}}}, \tag{3.76}$$

where L is a constant which depends on D and E (see properties in equations (3.9) and (3.10)) and constants in the assumption in equation (3.21). Therefore, under assumption in equation (3.20) we obtain

$$\mathbb{E}(\|\hat{f}_n^\lambda - f^\dagger\|^2) \leq C_r^2 \lambda^{2r} R^2 + \frac{C}{n} L \frac{1}{\lambda^{1+\frac{1}{b}}}. \tag{3.77}$$

By balancing terms in the r.h.s. of equation (3.77) we have the thesis. □

Finally, we give the proof of Lemma 4.

Proof of Lemma 4. The variance term can be bounded as follows

$$\begin{aligned}
\mathbb{E}(\|\hat{f}_n^\lambda - \mathbb{E}(\hat{f}_n^\lambda)\|^2) &\leq \frac{C}{n} \sum_{j:\tau_j \geq \lambda} s_\lambda^2(\tau_j)\tau_j && (3.78)\\
&\leq \frac{C}{n} \left(\sup_{\tau_j \in \tau(A^*A)} s_\lambda(\tau_j) \right) \sum_{j:\tau_j \geq \lambda} s_\lambda(\tau_j)\tau_j && (3.79)\\
&\leq \frac{C}{n} Q \frac{E}{\lambda} \left(\sup_{\tau_j \in \tau(A^*A)} s_\lambda(\tau_j)\tau_j \right) \leq \frac{C}{n} \frac{E}{\lambda} QD, && (3.80)
\end{aligned}$$

for the properties in equations (3.9) and (3.10) of the regularization function. Therefore,

$$\mathbb{E}(\|\hat{f}_n^\lambda - f^\dagger\|^2) \leq C_r^2 \lambda^{2r} R^2 + \frac{CEQD}{n\lambda}, \tag{3.81}$$

by balancing terms in the r.h.s. of equation (3.81) we obtain the thesis. □

For the sake of completeness we give an alternative proof of the known result that the upper rate of the spectral regularized estimator in the finite dimensional space case is $O\left(n^{-\frac{r}{2r+1}}\right)$ under the Holder-type source condition in equation (3.20).

Proposition 8. *Let $\mathcal{X}$ be the input space consisting of a finite set of elements (Q different elements), i.e. $\mathcal{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_Q\}$ with $\boldsymbol{x}_j \in \mathbb{R}^p$ and $\boldsymbol{x}_j \neq \boldsymbol{x}_{j'}$, for $j \neq j'$. Let $\hat{f}^\lambda_{n,\text{learn}}$ be defined in equation* (3.18) *and let the model be described by equations in* (3.16) *and* (3.17)*. Under the source condition in equation* (3.20) *we have*

$$\mathbb{E}(\|\hat{f}^\lambda_{n,\text{learn}} - f^\dagger\|) \in O\left(\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}\right), \tag{3.82}$$

with $\lambda \in \Theta\left(\left(\frac{1}{n}\right)^{\frac{1}{2r+1}}\right)$.

Proof. At first we notice that, in the case $n > Q$ the samples X_i can be repeated. We denote with $h^{(l)}$ the number of times that $\boldsymbol{x}_l$ is repeated between the samples $(X_i)_{i=1}^n$ and we denote with $\mathrm{Y}_j^{(l)}$ the j-th response associated to the feature $\boldsymbol{x}_l$, for $j = 1, \dots, h^{(l)}$. We can define $\bar{\mathbf{Y}}$ the Q-dimensional vector as follows

$$\bar{\mathbf{Y}} = \begin{bmatrix} \bar{\mathrm{Y}}_1 := \frac{1}{h^{(1)}} \sum_{j=1}^{h^{(1)}} \mathrm{Y}_j^{(1)} \\ \vdots \\ \bar{\mathrm{Y}}_Q := \frac{1}{h^{(Q)}} \sum_{j=1}^{h^{(Q)}} \mathrm{Y}_j^{(Q)} \end{bmatrix}. \tag{3.83}$$

Empirically, we can write the problem by considering the set of Q samples $(\mathrm{X}^{(l)}, \frac{1}{h^{(l)}} \sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)})$ i.i.d., where $\mathrm{X}^{(l)}$ is a deterministic variable ($\mathrm{X}^{(l)} = \boldsymbol{x}_l$ with probability equal to 1) and each $\mathrm{Y}_j^{(l)}$ has distribution $\rho(\mathrm{Y}|\boldsymbol{x}_l) = \rho(\mathrm{Y}|\mathrm{X}^{(l)})$ (therefore $\mathrm{Y}_j^{(l)}$ are i.i.d. for $j = 1, \cdots, h^{(l)}$ and for $l = 1, \dots, Q$). An empirical

distribution $\hat{v}$ can be associated to the set of samples in the following way: we define $\mathrm{f}^{(l)} := \frac{h^{(l)}}{n}$ the frequency of the event $X_i = \boldsymbol{x}_l$. Therefore we can construct the empirical measure $\hat{v} = \sum_{l=1}^{Q} \mathrm{f}^{(l)} \delta_{\boldsymbol{x}_l}$, which is a probability measure since it satisfies $\sum_{l=1}^{Q} \mathrm{f}^{(l)} = \sum_{l=1}^{Q} \frac{h^{(l)}}{n} = 1$.

We define the following operators which are associated to the set $\mathcal{X}$. We define $A_{\bar{\mathbf{x}}} : \mathcal{H}_1 \to \mathbb{R}^Q$,

$$(A_{\bar{\mathbf{x}}} f)_l = \langle f, \phi_{x_l} \rangle_{\mathcal{H}_1}, \tag{3.84}$$

for $l = 1, \ldots, Q$, for $f \in \mathcal{H}_1$ and we define $A_{\bar{\mathbf{x}}}^* : \mathbb{R}^Q \to \mathcal{H}_1$,

$$A_{\bar{\mathbf{x}}}^* w = \sum_{l=1}^{Q} w_l \phi_{x_l}, \tag{3.85}$$

for $w \in \mathbb{R}^Q$. Furthermore we define $\mathbf{H}$ as the $Q \times Q$ diagonal matrix which is defined as

$$\mathbf{H} = diag(\mathrm{f}^{(1)}, \cdots, \mathrm{f}^{(Q)}). \tag{3.86}$$

With simple computations we have

$$A_n^* A_n = A_{\bar{\mathbf{x}}}^* \mathbf{H} A_{\bar{\mathbf{x}}} \quad \text{and} \quad A_n^* \mathbf{y} = A_{\bar{\mathbf{x}}}^* (\mathbf{H} \bar{\mathbf{Y}}). \tag{3.87}$$

By defining

$$A_{\mathbf{x}} := \sqrt{\mathbf{H}} A_{\bar{\mathbf{x}}} \quad \text{and} \quad \mathbf{Y} := \sqrt{\mathbf{H}} \bar{\mathbf{Y}}, \tag{3.88}$$

where $\sqrt{\mathbf{H}}$ is the diagonal matrix with entries the square root of the matrix $\mathbf{H}$, we obtain

$$A_n^* A_n = A_{\mathbf{x}}^* A_{\mathbf{x}} \quad \text{and} \quad A_n^* \mathbf{y} = A_{\mathbf{x}}^* \mathbf{Y}. \tag{3.89}$$

We remark that the operators $A_{\mathbf{x}}$ and $A_{\mathbf{x}}^*$ are defined as follows

$$(A_{\mathbf{x}} f)_l = \langle f, \sqrt{\mathrm{f}^{(l)}} \phi_{x_l} \rangle_{\mathcal{H}_1} \quad \text{and} \quad A_{\mathbf{x}}^* w = \sum_{l=1}^{Q} w_l \sqrt{\mathrm{f}^{(l)}} \phi_{x_l}, \tag{3.90}$$

for each $l = 1, \ldots, Q$. Let $\hat{f}_{n,\text{learn}}^{\lambda}$ be the spectral regularized estimator defined

in equation (3.18). From equation (3.87) we have the following equality

$$\hat{f}^{\lambda}_{n,\text{learn}} = \mathfrak{R}(A_{\mathbf{x}}, \mathbf{Y}, \lambda) = s_{\lambda}\left(A^{*}_{\mathbf{x}}A_{\mathbf{x}}\right) A^{*}_{\mathbf{x}}\mathbf{Y}. \tag{3.91}$$

We focus on giving an upper rate for the estimator $\hat{f}^{\lambda}_{n,\text{learn}}$. Now we bound the following error as follows

$$\|\hat{f}^{\lambda}_{n,\text{learn}} - f^{\dagger}\|_{\mathcal{H}_1} \leq \|\hat{f}^{\lambda}_{n,\text{learn}} - s_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})A^{*}_{\mathbf{x}}A_{\mathbf{x}}f^{\dagger}\| + \|s_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})A^{*}_{\mathbf{x}}A_{\mathbf{x}}f^{\dagger} - f^{\dagger}\|. \tag{3.92}$$

In the inverse problem terminology the first term in the r.h.s. of inequality (3.92) is the approximation error and the second one is the propagation of the noise in the regularization. For easy of writing we define $\hat{f}^{\dagger,\lambda}_{n} := \mathfrak{R}(A_{\mathbf{x}}, A_{\mathbf{x}}f^{\dagger}, \lambda) = s_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})A^{*}_{\mathbf{x}}A_{\mathbf{x}}f^{\dagger}$. We start by bounding the second term of the r.h.s. of equation (3.92):

$$\|\hat{f}^{\dagger,\lambda}_{n} - f^{\dagger}\| = \|\left(s_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})A^{*}_{\mathbf{x}}A_{\mathbf{x}} - I\right) f^{\dagger}\| \leq R\|\mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})B^{r}\|, \tag{3.93}$$

using the Holder-type source condition in equation (3.20) and where $\mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}}) = s_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})A^{*}_{\mathbf{x}}A_{\mathbf{x}} - I$. With simple computations we have

$$\mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})B^{r} = \mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})(A^{*}_{\mathbf{x}}A_{\mathbf{x}})^{r} + \mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})\left(B^{r} - (A^{*}_{\mathbf{x}}A_{\mathbf{x}})^{r}\right). \tag{3.94}$$

For the property in equation (3.11) of the regularization function the first term of the r.h.s. of equation (3.94) is bounded as follows

$$\|\mathrm{r}_{\lambda}(A^{*}_{\mathbf{x}}A_{\mathbf{x}})(A^{*}_{\mathbf{x}}A_{\mathbf{x}})^{r}\| \leq C_r\lambda^{r}. \tag{3.95}$$

Now we focus on the second term of the r.h.s. of equation (3.94). We denote $\hat{B} := A^{*}_{\mathbf{x}}A_{\mathbf{x}}$. Then

$$\|\mathrm{r}_{\lambda}(\hat{B})(\hat{B}^{r} - B^{r})\| \leq \|\mathrm{r}_{\lambda}(\hat{B})\|\|\hat{B}^{r} - B^{r}\| \leq C_0C'_r\|\hat{B} - B\|, \tag{3.96}$$

where we have used the property in equation (3.11) of the regularization

function and the result in Proposition 5.6 in [17] with C_r' a suitable constant. By remarking that $\hat{B} = A_n^* A_n$ and by using the result in Proposition 5.5 in [17] we obtain that

$$\mathbb{E}(\|\hat{B} - B\|) \leq \frac{12c^2}{\sqrt{n}}, \tag{3.97}$$

where c is the constant in equation (3.1). Therefore, the term in equation (3.93) can be bounded as follows

$$\mathbb{E}(\|\mathrm{r}_\lambda(\hat{B}) f^\dagger\|) \leq RC_r\lambda^r + RC_0C_r'\frac{12c^2}{\sqrt{n}}. \tag{3.98}$$

Now we bound the first term of the r.h.s. of equation (3.92) as follows

$$\begin{aligned} \|\hat{f}^\lambda_{n,\text{learn}} - \hat{f}^{\dagger,\lambda}_n\|^2 &= \langle \hat{f}^\lambda_{n,\text{learn}} - \hat{f}^{\dagger,\lambda}_n, A^*_{\mathbf{x}} s_\lambda(A^*_{\mathbf{x}} A_{\mathbf{x}})(\mathbf{Y} - A_{\mathbf{x}} f^\dagger)\rangle \\ &= \langle A_{\mathbf{x}}(\hat{f}^\lambda_{n,\text{learn}} - \hat{f}^{\dagger,\lambda}_n), s_\lambda(A^*_{\mathbf{x}} A_{\mathbf{x}})(\mathbf{Y} - A_{\mathbf{x}} f^\dagger)\rangle \\ &\leq \|A_{\mathbf{x}}\hat{f}^\lambda_{n,\text{learn}} - A_{\mathbf{x}}\hat{f}^{\dagger,\lambda}_n\| \|s_\lambda(A_{\mathbf{x}} A^*_{\mathbf{x}})\| \|\mathbf{Y} - A_{\mathbf{x}} f^\dagger\|. \end{aligned} \tag{3.99}$$

Now we bound the following term

$$\begin{aligned} \|A_{\mathbf{x}}\hat{f}_\lambda - A_{\mathbf{x}}\hat{f}^{\dagger,\lambda}_n\| &= \|A_{\mathbf{x}} s_\lambda(A^*_{\mathbf{x}} A_{\mathbf{x}}) A^*_{\mathbf{x}}(\mathbf{Y} - A_{\mathbf{x}} f^\dagger)\| \\ &= \|A_{\mathbf{x}} A^*_{\mathbf{x}} s_\lambda(A_{\mathbf{x}} A^*_{\mathbf{x}})(\mathbf{Y} - A_{\mathbf{x}} f^\dagger)\| \\ &\leq \|A_{\mathbf{x}} A^*_{\mathbf{x}} s_\lambda(A_{\mathbf{x}} A^*_{\mathbf{x}})\| \|\mathbf{Y} - A_{\mathbf{x}} f^\dagger\| \\ &\leq D\|\mathbf{Y} - A_{\mathbf{x}} f^\dagger\|, \end{aligned} \tag{3.100}$$

where we have used the property in equation (3.9) of the regularization function. By using the bound in equation (3.100) and the property in equation (3.10) of the regularization function we obtain

$$\|\hat{f}^\lambda_{n,\text{learn}} - \hat{f}^{\dagger,\lambda}_n\|^2 \leq D\frac{E}{\lambda}\|\mathbf{Y} - A_{\mathbf{x}} f^\dagger\|^2. \tag{3.101}$$

We compute the following expectation

$$\begin{aligned}\mathbb{E}(\|\mathbf{Y} - A_{\mathbf{x}} f^{\dagger}\|^2) &= \sum_{l=1}^{Q} \mathbb{E}\left(\left(\sqrt{\mathrm{f}^{(l)}}\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)} - \langle f^{\dagger}, \sqrt{\mathrm{f}^{(l)}}\phi_{x_l}\rangle\right)^2\right) \\ &= \sum_{l=1}^{Q} \mathrm{f}^{(l)}\mathbb{E}\left(\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)} - \langle f^{\dagger}, \sqrt{\mathrm{f}^{(l)}}\phi_{x_l}\rangle\right)^2\right).\end{aligned}$$

By remarking that

$$\mathbb{E}_{\rho(\cdot|\cdot)}(\mathrm{Y}_j^{(l)}|\boldsymbol{x}_l) = \langle f^{\dagger}, \phi_{\boldsymbol{x}_l}\rangle = Af^{\dagger}(\boldsymbol{x}_l), \tag{3.102}$$

$$\mathbb{E}_{\rho(\cdot|\cdot)}\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)}|\boldsymbol{x}_l\right) = \langle f^{\dagger}, \phi_{\boldsymbol{x}_l}\rangle = Af^{\dagger}(\boldsymbol{x}_l), \tag{3.103}$$

$$Var\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)}|\boldsymbol{x}_l\right) = \frac{\sigma^2}{h^{(l)}} \tag{3.104}$$

and

$$\begin{aligned}\mathbb{E}\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathrm{Y}_j^{(l)}\right) &= \frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \mathbb{E}(\mathrm{Y}_j^{(l)}) \\ &= \frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \int_{\mathcal{X}}\int_{\mathcal{Y}} \mathrm{Y}_j^{(l)} d\rho(\mathrm{Y}_j^{(l)}|\boldsymbol{x}_l) d\nu(\boldsymbol{x}_l) \\ &= \frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}} \int_{\mathcal{X}} \langle f^{\dagger}, \phi_{\boldsymbol{x}_l}\rangle d\nu(\boldsymbol{x}_l) \\ &= \mathbb{E}_{x\sim\nu}(Af^{\dagger}(x)) = \mathbb{E}_{x\sim\nu}(\langle f^{\dagger}, \phi_x\rangle)\end{aligned}$$

then

$$\begin{aligned}
&\mathbb{E}\left(\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}}\Upsilon_j^{(l)} - \langle f^\dagger, \phi_{\boldsymbol{x}_l}\rangle\right)^2\right) \\
&= \int_{\mathcal{X}}\int_{\mathcal{Y}}\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}}\Upsilon_j^{(l)} - \langle f^\dagger, \phi_{\boldsymbol{x}_l}\rangle\right)^2 d\rho(\Upsilon_j^{(l)}|\boldsymbol{x}_l)d\nu(\boldsymbol{x}_l) \\
&= \int_{\mathcal{X}} Var\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}}\Upsilon_j^{(l)}|\boldsymbol{x}_l\right)d\nu(\boldsymbol{x}_l) \\
&= \frac{\sigma^2}{h^{(l)}} \qquad (3.105)
\end{aligned}$$

and

$$\sum_{l=1}^{Q}\mathrm{f}^{(l)}\mathbb{E}\left(\left(\frac{1}{h^{(l)}}\sum_{j=1}^{h^{(l)}}\Upsilon_j^{(l)} - \langle f^\dagger, \phi_{\boldsymbol{x}_l}\rangle\right)^2\right) = \sum_{l=1}^{Q}\mathrm{f}^{(l)}\frac{\sigma^2}{h^{(l)}} = Q\frac{\sigma^2}{n}. \qquad (3.106)$$

Therefore,

$$\mathbb{E}(\|\hat{f}_{n,\text{learn}}^{\lambda} - \hat{f}_n^{\dagger,\lambda}\|^2) \leq D\frac{E}{\lambda}Q\frac{\sigma^2}{n} \qquad (3.107)$$

and we obtain the following bound in expectation

$$\mathbb{E}(\|\hat{f}_{n,\text{learn}}^{\lambda} - f^\dagger\|) \leq RC_r\lambda^r + 12c^2RC_0C_r'\frac{1}{\sqrt{n}} + \sqrt{DQ}\sqrt{\frac{E}{\lambda}}\sqrt{\frac{\sigma^2}{n}}. \qquad (3.108)$$

By balancing terms in the r.h.s. of equation (3.108) we obtain the following optimal upper rate

$$\mathbb{E}(\|\hat{f}_{n,\text{learn}}^{\lambda} - f^\dagger\|) \in O\left(\left(\frac{1}{n}\right)^{\frac{r}{2r+1}}\right), \qquad (3.109)$$

with $\lambda \in \Theta\left(\left(\frac{1}{n}\right)^{\frac{1}{2r+1}}\right)$.

□

Chapter 4

A fast and consistent sparsity-enhancing method for Poisson data

In this Chapter we focus on Poisson data. This restriction can be read as a particular choice of the loss function introduced in the previous Chapters. By taking

$$V(Y, Af(X)) = Y \log \frac{Y}{Af(X)} + Af(X) - Y \tag{4.1}$$

we implicitly assume, thanks to the Maximum Likelihood formulation, that the noise on the data is distributed according to a Poisson law. Poisson noise is quite common in inverse problems and particularly in imaging applications, due to the quantum nature of the recorded light radiation. On the other hand, it is relevant also in learning applications when the response variables are counts. More in general, the objective of statistical learning is two-fold: (1) ensuring a good estimation and (2) selecting the relevant variables. The first objective means that the learning algorithm shall provide accurate predictions, (in this regard the correct noise hypothesis could be crucial) and the second one means that the algorithm shall identify the most relevant features, i.e. those variables which play an important role for the prediction. Selecting the most relevant variables is an issue also in inverse problems, e.g. in sparse

signal recovery. In this case, the goal is to find the smallest number of elements of a suitable basis to represent a signal: sparsity strategies apply to imaging applications, e.g. in astronomy and medical imaging.

In sparse signal recovery with Poisson data a lot of attention has been paid on fast and efficient optimization methods especially when the number of data is high and therefore a large scale inverse problem has to be solved. The penalized Maximum Likelihood approach in the context of Poisson noise leads to the minimization of a penalized functional where the discrepancy term is the well-known non-quadratic functional Kullback-Leibler divergence (see section 2.2). Recent improvements have been focused on acceleration of the usual proximal gradient methods requiring sophisticated optimization techniques and first order approximations of the objective function [21; 46; 59; 60; 61; 67; 129].

On the other hand, in statistical learning a special effort has been provided in promoting consistent variable selection and estimation. To this aim, one of the most used strategy is to use the ℓ_1-penalty, i.e. the Lasso method [134] which performs sign consistent selection under the so-called Irrepresentable Condition [155]. A major step forward in this direction was the introduction of the Adaptive Lasso, which guarantees variable selection consistency in the case of Generalized Linear Models (GLMs) under less restrictive statistical assumptions [156].

In this Chapter we propose a data-dependent global quadratic approximation of the Kullback-Leibler divergence enabling us to formulate simplified Lasso and Adaptive Lasso estimators suitable for sparse Poisson regression. We call them as Poisson Reweighted Lasso (PRiL) and Adaptive Poisson Reweighted Lasso (APRiL). These estimators can be computed by taking advantages of the fastest available algorithms, i.e. those developed for ℓ_1-penalized least squares regression [10; 50; 52]. We prove that the adaptive estimator satisfies the property of consistent variable selection. Finally we show the performances of the proposed estimators both on a statistical learning application (with synthetic data) and on a sparse signal recovery one (with

synthetic images).

The proofs of results of the current Chapter are provided in section 4.5.

4.1 Sparsity: a tool for learning and inverse problems

Sparsity has become a key concept in both statistical learning and inverse problems. Roughly speaking, a sparse statistical model is one in which only a relatively small number of parameters, called also predictors or features, play an important role. In image reconstruction problems the idea is that the information content of images is small compared to the number of pixels we use to represent them. As a consequence images can be compressed on a proper basis, in which few coefficients are non zero, i.e. they are sparse.

Promoting the sparsity of a solution would be ideally obtained by minimizing the ℓ_0 norm of the solution, which limits explicitly the number of non-zero elements, represented on a suitable basis. However, such a regularization term is non convex yielding combinatorial complexity. Since the relaxation to ℓ_p norm with $0 < p < 1$ leads again to a non-convex minimization problem, the most common approximation is the ℓ_1 norm, which represents a good trade-off between sparsity promotion and computational tractability.

In the following we introduce the ℓ_1-penalized method, known as Least Absolute Shrinkage and Selection Operator (Lasso) [134] method and we show one of its variation, the Adaptive Lasso [156]. In the last decade this approach has been widely investigated, also thanks to the development of new efficient algorithms for convex optimization [10].

4.1.1 Lasso and Adaptive Lasso: a reminder

Lasso is a regularization technique for simultaneous estimation and variable selection. It was introduced by [134] as a technique for linear regression and it has become a very attractive method [74; 155; 156], since its entire

regularization path can be computed efficiently [42; 52; 105; 147]. Lasso is known as basis pursuit [31] in the context of signal processing. In the last decade many Lasso-type methods have been proposed, including extensions or variations of the classical Lasso (Fused Lasso [135], Group Lasso [151], Multi-task Lasso [103], Trace Lasso [57] to mention a few). In the current section we deal with a particular variation of Lasso which is the Adaptive Lasso [157]. First we introduce the classical Lasso method in the usual setting of linear regression models, according to the notations introduced in the previous Chapters. In Chapter 2 we show that the conditional expectation $\mathbb{E}(Y|X)$ is the ideal solution of the optimization problem consisting in minimizing the expected risk characterized by a particular choice of the loss function (e.g. the square loss). Therefore, the problem of regression is often stated as finding an estimator $\hat{g} : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}^p$ is the input space, which approximates the function $g^\dagger(x) = \mathbb{E}(Y|X = x)$ from noisy samples. Therefore, given the samples $\{(X_i, Y_i)\}_{i=1}^n$, we can assume that observations are modeled as follows

$$Y_i = g^\dagger(X_i) + \epsilon_i, \tag{4.2}$$

where ϵ_i are independent centered noise variables for $i = 1, \dots, n$. Recalling the feature maps introduced in Chapter 1, it can be assumed that $g^\dagger$ admits the following representation $g^\dagger(x) = \langle f^\dagger, \phi_x \rangle_{\mathcal{H}_1}$, with ϕ_x the feature map and $f^\dagger \in \mathcal{H}_1$ the sought solution in a Hilbert space $\mathcal{H}_1$. In the particular case in which $\mathcal{H}_1 := \mathbb{R}^p$ then g can be parameterized by a coefficient vector β^*, i.e.

$$g^\dagger(x) = \langle \beta^*, \phi_x \rangle_{\mathbb{R}^p} = \phi_x^T \beta^*. \tag{4.3}$$

This case can be considered as an extension of the the usual linear regression case, which is found by taking $\phi_x = x$: in this case

$$g^\dagger(x) = \langle \beta^*, x \rangle_{\mathbb{R}^p} = x^T \beta^*. \tag{4.4}$$

Therefore, the linear regression model can be described as

$$Y_i = X_i^T \beta^* + \epsilon_i, \tag{4.5}$$

where ϵ_i are the noise components and they are usually assumed to be i.i.d. with mean 0 and a fixed constant variance σ^2 (e.g. standard Gaussian) and β^* is a suitable vector of parameters. Without loss of generality, we assume that the data are centered, in this way the intercept is not included in the regression function, otherwise a constant intercept β_0^* is considered in the equation (4.5) (Chapter 2 in [68]). The Lasso estimator is given by

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1, \tag{4.6}$$

where $\mathbf{Y}$ is the vector containing the samples Y_i and $\mathbf{X}$ is the matrix which has X_i^T as rows. The functional to minimize is the sum of two contributions: the first is the residual term, which is in this case the least square functional, and the second is the penalized term which is represented by the ℓ_1-penalty. The ℓ_1-penalty is crucial to enhance sparsity in the solution and the regularization parameter $\lambda > 0$ has the role to create a trade off between the two terms. However, the variable selection provided by the Lasso method has been shown to be consistent under certain conditions. In particular, it is sign-consistent under the Irrepresentable condition [155], i.e. it is sign-consistent if and only if the correlation between the relevant and irrelevant variables is low. Therefore, in order to assure the variable selection consistency under less restrictive assumptions the Adaptive Lasso has been introduced [156]. On the contrary of the Lasso procedure, which forces coefficients to be equally penalized in the ℓ_1-penalty, the idea of the adaptive approach is to introduce weights in the ℓ_1-penalty which allow penalizing the coefficients in different ways. The Adaptive Lasso is given by

$$\hat{\beta}_{\mathbf{w}} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|, \tag{4.7}$$

where w_j are the positive weights. The choice of the weights is an issue. They have to be chosen such that consistency properties hold. Furthermore, the Adaptive Lasso, with a suitable choice of weights, guarantees variable selection consistency in the case of Generalized Linear Models (GLMs) [156].

4.2 Sparsity and Poisson data

Let us consider a Poisson random vector $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ made of independently distributed components with mean $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$, i.e.

$$Y_i \sim \text{Poisson}(\mu_i^*), \tag{4.8}$$

$\forall\, i \in \{1, \dots, n\}$. Suppose that the parameter μ_i^* can be expressed as

$$\mu_i^* = h^{-1}((\mathbf{X}\beta^*)_i), \tag{4.9}$$

$\forall\, i \in \{1, \dots, n\}$, where $h : \mathbb{R} \to \mathbb{R}$ is an invertible function, $\mathbf{X}$ is the $n \times p$ matrix wich has X_i^T as rows for $i \in \{1, \dots, n\}$ and $\beta^* = (\beta_1^*, \dots, \beta_p^*)^T$ is a suitable vector of parameters. We denote with x_{ij} the (i, j)-entry of the matrix $\mathbf{X}$ and we denote with $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})^T$ the j-th column of $\mathbf{X}$ for $j \in \{1, \dots, p\}$. In statistical estimation $\mathbf{Y}$ is called the response vector, $\mathbf{X}$ is the predictor (or feature, or design) matrix, h is called the link function and equation (4.9) describes the GLMs [93]. On the other hand, in signal recovery $\mathbf{Y}$ represents the vector of noisy measurements of a given random signal and $\mathbf{X}$ describes a linear signal formation process depending on the parameters β^*. We assume that the true unknown vector β^* is sparse. More formally, let us denote by

$$\mathcal{A}^* := \{j \in \{1, \dots, p\} : \beta_j^* \neq 0\}$$

the set of indexes corresponding to relevant variables of the model, namely the active set, and with $\#\mathcal{A}^*$ its cardinality. We suppose that

$$q := \#\mathcal{A}^* < p\,.$$

In applications we consider q as being a substantially smaller fraction of p. Such an assumption leads to the variable selection and estimation problem, i.e. to compute a model with a small number of relevant variables with good prediction capabilities [51]. The standard choice for h in the statistical learning framework is the so-called canonical link function of the GLM theory, which is the logarithm function $h(z) := \ln(z)$ in the case of Poisson data (see Chapter 3 in [135]). In this way, Poisson means are equal to the exponents of linear predictors, i.e. $\mu^* = \exp(\mathbf{X}\beta^*)$, taking positive values only.

In the case of Poisson regression with canonical link, an usual variable selection method comes from extending the Adaptive Lasso to the GLMs, suggesting the following estimator

$$\hat{\beta}^{(n)}(\text{log link}) := \arg\min_{\beta} \quad \sum_{i=1}^{n} \exp(\mathbf{X}\beta)_i - Y_i(\mathbf{X}\beta)_i \\ +\lambda \sum_{j=1}^{p} w_j|\beta_j|, \tag{4.10}$$

where $\beta \in \mathbb{R}^p$, λ is the positive regularization parameter, $\mathbf{w} = (w_j)_{j=1,\dots,p}$ is the weights vector, which has the role of weighting the contribution of the coefficients β_j. This estimator can be derived from the adaptive ℓ_1-penalized Maximum Likelihood approach (see section 2.2) applied to the Poisson GLM with canonical log-link (equations (4.8) and (4.9)).

Another possible choice for h is the identity link, i.e. $h(z) := \mathrm{id}(z)$ under the non-negativity constraint on z. This choice is natural in a large variety of applications, e.g. in emission tomography and in astronomical image reconstruction and deblurring, since the matrix $\mathbf{X}$ is able to describe a linear transformation which approximates the physical signal formation process [112; 130]. In the unconventional case of Poisson GLM with identity-link the adaptive ℓ_1-penalized Maximum Likelihood approach leads to minimize the following functional

$$\hat{\beta}^{(n)}(\text{id link}) := \arg\min_{\beta \in \mathcal{C}} \sum_{i=1}^{n} D((\mathbf{X}\beta)_i, Y_i) + \lambda \sum_{j=1}^{p} w_j|\beta_j|, \tag{4.11}$$

where

$$\mathcal{C} = \{\beta \in \mathbb{R}^p : (\mathbf{X}\beta)_i > 0 \; \forall \; i \in \{1, \ldots, n\}\} \tag{4.12}$$

is the subset of feasible β solutions and D is the Kullback-Leibler divergence [15] which is defined as

$$D(z, y) := y \log \frac{y}{z} + z - y, \tag{4.13}$$

with $z, y > 0$ and $D(z, 0) := z$. The presence of such additional constraint $(\mathbf{X}\beta)_i > 0$, $\forall \; i \in \{1, \ldots, n\}$ can be a disadvantage of using the identity link. Indeed, this can result in the need for much more computationally expensive optimization methods. However, in applications the vector β^* often contains an offset parameter associated with a constant value predictor, which usually makes the quantity $\mathbf{X}\beta$ substantially larger than zero. As a consequence the solution of the problem is an interior point of the feasible solution set (4.12). This offset is called “the intercept" in the statistical language and “the background" in signal recovery. The choice of the adaptive weights is put forward against different ways [20; 28]. For Poisson GLMs the use of data-driven adaptive weights has been recently proposed: in [72] authors adapted Lasso to work with Poisson data by means of a particular choice of the adaptive weights, while in [71] authors proposed a choice based on concentration inequalities for solving an adaptive problem arising from the Poisson GLM with the canonical log link.

In [156] it has been proven that, by choosing the weights in an appropriate manner, the estimator $\hat{\beta}^{(n)}(\text{log link})$ performs consistent variable selection and estimation, under some mild regularity conditions where both $\mathbf{X}$ and $\mathbf{Y}$ are thought of as random variables. Now we introduce an approximation of the functional (4.11) which allows us to define an adaptive penalized reweighted least squares method with the property to identify the exact relevant explanatory variables when the number of observations diverges in a deterministic matrix design framework. At the same time, such an approximation overcomes the need for expensive optimization methods such as the Iteratively

Reweighted Least Squares (IRLS) commonly applied in the case of GLMs [35; 39; 55].

4.3 Adaptive Poisson Reweighted Lasso

In this section we present two new ℓ_1-penalized methods for sparse Poisson regression: Poisson Reweighted Lasso (PRiL) and Adaptive Poisson Reweighted Lasso (APRiL). They are based on a globally quadratic approximation of the Kullback-Leibler (KL) divergence and they enhance sparsity with a classical ℓ_1-penalty and a weighted ℓ_1-penalty, respectively. We prove the theoretical properties of such proposed estimators and after we show a numerically efficient approach to compute them.

4.3.1 Approximation of the Kullback-Leibler divergence

We now show a global quadratic approximation of the KL divergence and we prove that such an approximation is an asymptotically unbiased estimator of the KL divergence. Formally, we have the following

Theorem 6. *Let y be a Poisson random variable with mean θ. For any $z > 0$ such that $|z - \theta| \leq c\sqrt{\theta}$, where $c > 0$ is constant (or even bounded from above when $\theta \to \infty$) such that $\theta - c\sqrt{\theta} > 0$, we have*

$$\mathbb{E}\left(D(z,y) - \frac{1}{2}\frac{(y-z)^2}{y+1}\right) = O\left(\frac{1}{\sqrt{\theta}}\right), \tag{4.14}$$

as $\theta \to \infty$.

The proof of Theorem 6 is given in the section 4.5 devoted to proofs of the novel results in the current Chapter. Theorem 6 implies that for all $i \in \{1, \ldots, n\}$, in a neighborhood of the exact values $(\mathbf{X}\beta^*)_i$, such an approximation is more and more accurate with $(\mathbf{X}\beta^*)_i \to \infty$. This approximation calls up to the Anscombe transform [4]. Nonetheless, the substantial difference is that

the proposed approximation (4.14) is globally quadratic making its numerical treatment extremely easier.

4.3.2 PRiL/APRiL estimators and properties

In view of the KL approximation given in section 4.3.1, we can introduce a novel estimator on the basis of a positive weight vector $\mathbf{w} = \{w_j\}_{j\in\{1,\dots,p\}}$, as follows

$$\hat{\beta}^{(n)}_{(\mathbf{w},\lambda)} := \arg\min_{\beta\in\mathcal{C}} \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i - (\mathbf{X}\beta)_i)^2}{Y_i+1} + \lambda\sum_{j=1}^{p} w_j|\beta_j|, \tag{4.15}$$

where λ is the regularization parameter. Therefore, functional in the r.h.s. of equation (4.15) is an asymptotically unbiased estimator of the functional in the r.h.s. of equation (4.11). We point out that the fit term in the r.h.s. of equation (4.15) is a re-weighted least square functional and it can be written in the following form: let Λ be the following $n \times n$ diagonal matrix

$$\Lambda = diag\left(\frac{1}{\sqrt{Y_1+1}},\dots,\frac{1}{\sqrt{Y_n+1}}\right), \tag{4.16}$$

then

$$\frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i - (\mathbf{X}\beta)_i)^2}{Y_i+1} = \frac{1}{2}\left\|\frac{\mathbf{Y}-\mathbf{X}\beta}{\sqrt{\mathbf{Y}+1}}\right\|_2^2 = \frac{1}{2}\left\|\Lambda(\mathbf{Y}-\mathbf{X}\beta)\right\|_2^2, \tag{4.17}$$

where the division and the square root of the vector in the second term have to be intended as element-wise. In the case weights are all equal to 1, i.e. $w_j = 1$ for any j, the estimator is the minimizer of a functional that we call "Poisson Reweighted Lasso" (PRiL), which is defined as follows

$$\hat{\beta}^{(n)}(\text{PRiL}) := \arg\min_{\beta\in\mathcal{C}} \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i - (\mathbf{X}\beta)_i)^2}{Y_i+1} + \lambda_1\|\beta_j\|_1, \tag{4.18}$$

where λ_1 denotes its regularization parameter. We prove in the following that $\hat{\beta}^{(n)}(\text{PRiL})$ is a $\sqrt{n}$-consistent estimator, provided an appropriate asymptotics of the regularization parameter λ_1 is given. As we mentioned before, data-

dependent choices of the weights w_j in the case of Poisson problems have been recently proposed in [66; 72] and are based on Poisson concentration inequalities. In all cases the idea is to choose such weights in order to provide the method with the asymptotic model selection consistency property. Inspired by the choice in [157] for the adaptive elastic net, we introduce the following weights

$$\hat{w}_j = \frac{1}{\left(|\hat{\beta}^{(n)}(\text{PRiL})_j| + \left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta}\right)^{\gamma}}, \tag{4.19}$$

where γ and δ are strictly positive constants. The estimator in equation (4.15) when provided with such weights is called "APRiL" for Adaptive Poisson Reweighted Lasso and we denote it by $\hat{\beta}^{(n)}(\text{APRiL})$, defined as follows

$$\hat{\beta}^{(n)}(\text{APRiL}) := \arg\min_{\beta \in \mathcal{C}} \frac{1}{2} \sum_{i=1}^{n} \frac{(Y_i - (\mathbf{X}\beta)_i)^2}{Y_i + 1} + \lambda \sum_{j=1}^{p} \hat{w}_j |\beta_j|, \tag{4.20}$$

where $\hat{w}_j$ are defined in (4.19). Now, the main goal is to prove that the $\hat{\beta}^{(n)}(\text{APRiL})$ estimator has the model selection consistency property in the case of Poisson data and under some assumptions on the matrix $\mathbf{X}$. We assume that:

(H1) the matrix $\mathbf{X}^T \Lambda^2 \mathbf{X}$ is positive definite, and

$$\mathbb{E}\left(\left(\frac{1}{\tau_{\min}(\mathbf{X}^T \Lambda^2 \mathbf{X})}\right)^4\right) \leq \frac{1}{(bn)^4} \text{ and}$$
$$\tau_{\max}(\mathbf{X}^T \mathbf{X}) \leq Bn$$

where $\tau_{\min}(A)$ and $\tau_{\max}(A)$ are the minimum and maximum eigenvalues of the matrix A respectively, b and B are two strictly positive constants

(H2) $\lim_{n \to +\infty} \frac{\lambda_1}{\sqrt{n}} = 0$

(H3) a) $\lim_{n \to +\infty} \lambda n^{\frac{\gamma}{2}-1} = \infty$, b) $\lim_{n \to +\infty} \lambda n^{\delta\gamma} = \infty$,
c) $\lim_{n \to +\infty} \lambda n^{\delta\gamma - \frac{1}{2}} = 0$

(H4) there exists an $L > 0$ such that

$$\max_{j\in\{1,\dots,p\}} \|\mathbf{x}_j\|_2 \leq L.$$

Assumptions in (H1) involve the matrix $\mathbf{X}$ and the random variable $\mathbf{Y}$. The hypothesis concerning $\tau_{\min}$ implies that

$$\mathbb{E}\left(\tau_{\min}\left(\frac{\mathbf{X}^T\Lambda^2\mathbf{X}}{n}\right)\right) \geq b \tag{4.21}$$

which calls up to the assumption used by [157]. Assumption (H2) involves the convergence rate of the regularization parameter λ_1 whereas assumptions described in (H3) involve the convergence rate of regularization parameter λ. Assumption (H4) is necessary for consistent model selection and it is automatically verified after the feature standardization/normalization procedure. In the following theorem we give a general bound of the expected error for the estimator (4.15).

Theorem 7. *Assuming hypothesis (H1), then it exists a constant $G < +\infty$, such that*

$$\mathbb{E}(\|\hat{\beta}^{(n)}_{(\mathbf{w},\lambda)} - \beta^*\|_2^2) \leq \frac{4\lambda^2\sqrt{\mathbb{E}\left(\left(\sum_{j=1}^p w_j^2\right)^2\right)} + pGBn}{(bn)^2}. \tag{4.22}$$

The proof of Theorem 7 is given in section 4.5. Such a bound takes into account that weights can be random variables. In the case weights are constants all equal to 1, the previous result boils down to the following

Corollary 7. *Assuming hypothesis (H1) then*

$$\mathbb{E}(\|\hat{\beta}^{(n)}(\text{PRiL}) - \beta^*\|_2) \leq \frac{2\lambda_1\sqrt{p} + \sqrt{pGBn}}{bn}. \tag{4.23}$$

It is worth observing that under assumption (H2) Corollary 7 implies that $\hat{\beta}^{(n)}(\text{PRiL})$ is a $\sqrt{n}$-consistent estimator. We now consider the weights given by equation (4.19). Although it is possible to have the consistent estimation

property by letting λ goes to 0 fast enough, assumptions (H1)-(H4) do not permit to conclude the consistency of the APRiL estimator. However, we now prove that such assumptions make APRiL a variable selection consistent estimator. We introduce the estimated active set

$$\hat{\mathcal{A}}^{(n)} = \{j \in \{1, \dots, p\} : \hat{\beta}^{(n)}(\text{APRiL})_j \neq 0\} \tag{4.24}$$

of the estimator $\hat{\beta}^{(n)}(\text{APRiL})$. The model selection consistency property reads

$$\lim_{n \to +\infty} \mathbb{P}(\hat{\mathcal{A}}^{(n)} = \mathcal{A}^*) = 1 \, . \tag{4.25}$$

Theorem 8. *Under assumptions (H1), (H2), (H3), (H4) the APRiL estimator has the model selection consistency property.*

The proof of the Theorem (given in section 4.5) exploits the $\sqrt{n}$-consistency property of the estimator $\hat{\beta}^{(n)}(\text{PRiL})$. This property underpins the choice of the weights defined in equation (4.19).

It is worth noticing that the consistency property has different implications depending on the application: for signal recovery problems, consistency is computed against the increasing number of bins/pixels in which the signal is measured, whereas for statistical learning it is evaluated against the increasing number of available examples. For a detailed discussion on this topic see, for example [38].

For the sake of completeness, we notice that a similar result can be obtained in the case $p = p(n) \to \infty$ provided that we replace hypotheses (H2) and (H3) with suitable conditions on the convergence rate of the regularization parameters λ_1 and λ and we assume an upper bound on the asymptotic behavior of p. We give the following

Proposition 9. *Consider assumptions (H1), (H4) and the following*

(H5) $p = O(n^c)$, *with* $0 \leq c < \frac{1}{6}(7 - \sqrt{37})$

(H6) $\lim_{n \to +\infty} \frac{\sqrt{p}}{\sqrt{n}\eta} = 0$

(H7) *a)* $\lim_{n\to+\infty} \frac{\lambda_1\sqrt{p}}{n\eta} = 0$, *b)* $\lim_{n\to+\infty} \frac{\lambda_1 n^{\delta+\frac{1}{\gamma}-1}}{\sqrt{p}} = 0$

(H8) *a)* $\lim_{n\to+\infty} \lambda n^{\gamma\delta-\frac{1}{2}} p\sqrt{p} = 0$, *b)* $\lim_{n\to+\infty} \frac{\lambda n^{\delta\gamma}}{p^{\gamma}} = \infty$,
c) $\lim_{n\to+\infty} \frac{\lambda p}{n\eta^{\gamma}} = 0$,

where $\eta := \min_{j\in\mathcal{A}^*} |\beta_j^*| + \left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta}$ *and* $\delta + \frac{1}{\gamma} < \frac{c+1}{2}$. *Then the APRiL estimator has the model selection consistency property.*

The condition on c in hypothesis (H5) ensures the possibility to choose γ and δ so that assumptions of the Proposition 9 are consistent. We remark that such assumptions allow the non zero coefficients β_j^* with $j \in \mathcal{A}^*$ to vanish. The proof of Proposition 9 is given in section 4.5.

4.3.3 Algorithm

The computation of the APRiL estimator can be performed by means of the same numerically efficient algorithms developed for the solution of the Lasso problem. We propose a numerical strategy which consists of two steps. First we reweight the columns of the matrix $\mathbf{X}$ and the vector $\mathbf{Y}$ by left-multiplying by Λ defined in equation (4.16). Second, following the approach proposed by [156], we reweight the predictor matrix $\mathbf{X}$ for computing the adaptive solution. These two steps need the computation of the solution of two Lasso problems. In Algorithm 1 we outline the scheme of the procedure.

In many applications the presence of an offset - be it a regression intercept or a constant background signal - makes the vector $\mathbf{X}\hat{\beta}_\lambda$ an interior point of the feasible set $\mathcal{C}$, i.e all its components are positive. Moreover, we notice that, unlike the functional in equation (4.11) which is based on the KL divergence, the proposed functional in equation (4.15) is meaningful for each β, even when $\mathbf{X}\beta$ has negative components. In such cases, the constraint $\mathcal{C}$ can be neglected during the optimization process and standard algorithms can be used in place of sophisticated constrained techniques. Therefore, steps 3 and 6 of the Algorithm 1 can be performed by solving the unconstrained Lasso problem.

Algorithm 1 APRiL estimator computation

1: Input: $\mathbf{X}$, $\mathbf{Y}$.
2: **Data driven reweighting**. Define

$$\tilde{\mathbf{X}} := \Lambda \mathbf{X} \qquad \tilde{\mathbf{Y}} := \Lambda \mathbf{Y},$$

where Λ is defined in equation (4.16).
3: Compute the regularization path

$$\hat{\beta}_{\lambda_1} = \arg\min_{\beta \in \mathcal{C}} \frac{1}{2} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\beta\|_2^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j|$$

and select $\hat{\beta}_{\lambda_1}(\text{PRiL})$ with λ_1 according to a cross validation process.
4: Compute the adaptive weights $\hat{\mathbf{w}}$ as in formula (4.19).
5: **Adaptive reweighting**. Define $\tilde{\tilde{\mathbf{X}}}$ so that

$$\tilde{\tilde{\mathbf{x}}}_j = \tilde{\mathbf{x}}_j / \hat{w}_j, \ \forall\, j \in \{1, \dots, p\}.$$

6: Compute the regularization path

$$\tilde{\tilde{\beta}}_\lambda = \arg\min_{\beta \in \mathcal{C}} \frac{1}{2} \|\tilde{\mathbf{Y}} - \tilde{\tilde{\mathbf{X}}}\beta\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

and select $\tilde{\tilde{\beta}}_\lambda$ with λ according to a cross validation process.
7: Output: $\hat{\beta}_\lambda(\text{APRiL})$ is such that

$$\hat{\beta}_\lambda(\text{APRiL})_j = (\tilde{\tilde{\beta}}_\lambda)_j / \hat{w}_j \quad \forall j \in \{1, \dots, p\}.$$

In this way, APRiL method can take advantage of numerically efficient solvers and of the piece-wise linear form of the regularization path [42].

4.4 Simulations: learning and sparse signal recovery

In this section we show two applications of the proposed methods. In the first one, we apply them to some statistical learning test problems and in the second one, we show that they can be successfully applied to wavelet-based Poisson denoising and deblurring. One of the main difference between these applications is that in the first case the model (or the link function) is not known whereas in the second case it is a linear operator representing the signal formation process. This leads us to make a performance comparison between our methods and the Lasso techniques for GLMs with Poisson data in the statistical learning application, and to check the performance of the proposed method in the sparse signal recovery one.

4.4.1 Statistical learning application

We present a synthetic variable selection problem in order to compare the proposed methods (PRiL and APRiL) with Lasso and Adaptive Lasso for GLMs for Poisson data [156]: we refer to these last methods as GLM and AGLM, respectively. The main goal of this synthetic experiment is to assess the variable selection performance of the proposed methods as the number of samples increases and its computational advantages when the number of samples reaches the order of million. It is worth observing that in statistical learning regression methods are based on a given data model (equation (4.9)), i.e. on a particular choice of the link function. The standard method based on GLM theory uses the log-link function (see equation (4.10)), which is the canonical choice for Poisson data, whereas the proposed methods are based on the identity-link. Therefore, in order to perform a comprehensive comparison of the methods, we consider two sets of data generated according to the log-link and the identity-link function based model, respectively. We are interested in evaluating the performance of the APRiL method and the AGLM method by applying them to both datasets. Furthermore, we test PRiL and

GLM methods on the same datasets in order to compare performances also of the non-adaptive methods. In particular, these two datasets are generated according to the following assumptions. We fix $p = 15$ and $q = \#\mathcal{A}^* = 5$. We construct the $n \times p$ predictor matrix $\mathbf{X}$ for $n = 125, 250, 500$, so that each of its columns is extracted by a p-dimensional normal multivariate distribution with zero mean and covariance Σ with $\Sigma_{jr} = \rho^{|j-r|}$, for $j, r \in \{1, \dots, p\}$. We assume $\rho = 0.5$ and $\rho = 0.75$. We consider the following two cases:

1. Log-link dataset. We generate the data $\mathbf{Y}$ by using log-link function as follows
$$Y_i = \text{Poisson}(\beta_0^* \exp((\mathbf{X}\beta^*)_i)), \ \forall\, i \in \{1, \dots, n\}, \tag{4.26}$$
where $\beta^* = (0.7, -0.5, 0.3, -0.4, 0.6, \mathbf{0}_{p-5})^T$ is the true coefficient vector ($\mathbf{0}_{p-5}$ denotes the zero vector of dimension $p - 5$) and β_0^* is a suitable constant intercept.

2. Identity-link dataset. We generate $\mathbf{Y}$ by using the identity-link function as follows
$$Y_i = \text{Poisson}((\mathbf{X}\beta^{**})_i + \beta_0^{**}), \ \forall\, i \in \{1, \dots, n\}, \tag{4.27}$$
where $\beta^{**} = (e^{0.7}, e^{-0.5}, e^{0.3}, e^{-0.4}, e^{0.6}, \mathbf{0}_{p-5})^T$ is the true coefficient vector and β_0^{**} is a suitable constant intercept.

In the second case we select the intercept in order to make each component of the vector $(\mathbf{X}\beta^{**})_i + \beta_0^{**}$ positive. In the first case we tune the intercept value so that data generated in the first case has about the same signal to noise ratio of the data generated in the second case. Moreover, for each problem, we generate 100 realizations of Poisson data and therefore we obtained 600 estimation problems ($\#n = 3$ and $\#\rho = 2$). For each one of these problems we perform regression by means of PRiL, APRiL, GLM and AGLM methods.

The APRiL weights are parametrized according with the assumptions in Theorem 8. In particular we use $\hat{w}_j$ as defined in equation (4.19), and we fix constants $\gamma = 3$ and $\delta = \frac{1}{8}$. For what concerns AGLM defined in equation

Table 4.1: Mean Square Error values obtained by averaging over 100 replicates the results provided by GLM, AGLM, PRiL and APRiL methods for each problem.

	log-link dataset			**identity-link dataset**		
n	125	250	500	125	250	500
	$\rho = 0.5$					
GLM	$4.1_{(\pm 2)}\ 10^{-4}$	$2.1_{(\pm 1.1)}\ 10^{-4}$	$7.2_{(\pm 4.7)}\ 10^{-5}$	$6.2_{(\pm 0.1)}\ 10^{-1}$	$6.2_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.03)}\ 10^{-1}$
AGLM-I	$4.1_{(\pm 2.1)}\ 10^{-4}$	$2.1_{(\pm 1.2)}\ 10^{-4}$	$8_{(\pm 5)}\ 10^{-5}$	$6.2_{(\pm 0.1)}\ 10^{-1}$	$6.2_{(\pm 0.1)}\ 10^{-1}$	$6.2_{(\pm 0.1)}\ 10^{-1}$
AGLM-II	$3_{(\pm 0.7)}\ 10^{-3}$	$4.1_{(\pm 0.7)}\ 10^{-3}$	$2.2_{(\pm 0.3)}\ 10^{-3}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.1)}\ 10^{-1}$
PRiL	$1.5_{(\pm 0.6)}$	$1.7_{(\pm 0.6)}$	$4.4_{(\pm 0.5)}$	$2.4_{(\pm 0.4)}\ 10^{-1}$	$2.1_{(\pm 0.3)}\ 10^{-1}$	$1.9_{(\pm 0.2)}\ 10^{-1}$
APRiL	$1.5_{(\pm 0.6)}$	$1.8_{(\pm 0.6)}$	$4.4_{(\pm 0.5)}$	$2.4_{(\pm 0.5)}\ 10^{-1}$	$2.1_{(\pm 0.4)}\ 10^{-1}$	$1.9_{(\pm 0.2)}\ 10^{-1}$
	$\rho = 0.75$					
GLM	$7_{(\pm 4)}\ 10^{-4}$	$4_{(\pm 3.1)}\ 10^{-4}$	$1.5_{(\pm 0.9)}\ 10^{-4}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.03)}\ 10^{-1}$
AGLM-I	$7.3_{(\pm 4.5)}\ 10^{-4}$	$4.1_{(\pm 2.2)}\ 10^{-4}$	$1.4_{(\pm 0.7)}\ 10^{-4}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.03)}\ 10^{-1}$
AGLM-II	$1.1_{(\pm 0.2)}\ 10^{-2}$	$1.5_{(\pm 0.2)}\ 10^{-2}$	$8.0_{(\pm 0.7)}\ 10^{-3}$	$6.4_{(\pm 0.1)}\ 10^{-1}$	$6.3_{(\pm 0.05)}\ 10^{-1}$	$6.3_{(\pm 0.05)}\ 10^{-1}$
PRiL	$2.8_{(\pm 1.2)}$	$5.7_{(\pm 1)}$	$8.6_{(\pm 0.9)}$	$2.7_{(\pm 0.7)}\ 10^{-1}$	$2.3_{(\pm 0.5)}\ 10^{-1}$	$2.1_{(\pm 0.4)}\ 10^{-1}$
APRiL	$2.7_{(\pm 1.2)}$	$5.6_{(\pm 1)}$	$8.3_{(\pm 1.2)}$	$2.7_{(\pm 0.7)}\ 10^{-1}$	$2.2_{(\pm 0.5)}\ 10^{-1}$	$2.1_{(\pm 0.4)}\ 10^{-1}$

(4.10) we fix the weights according to the following two strategies.

1. First strategy [156]:

$$\hat{w}_j = \frac{1}{|\hat{\beta}(\mathrm{MLE})_j|^{\bar{\gamma}}} \ \forall\, j \in \{1, \ldots, p\}\ , \tag{4.28}$$

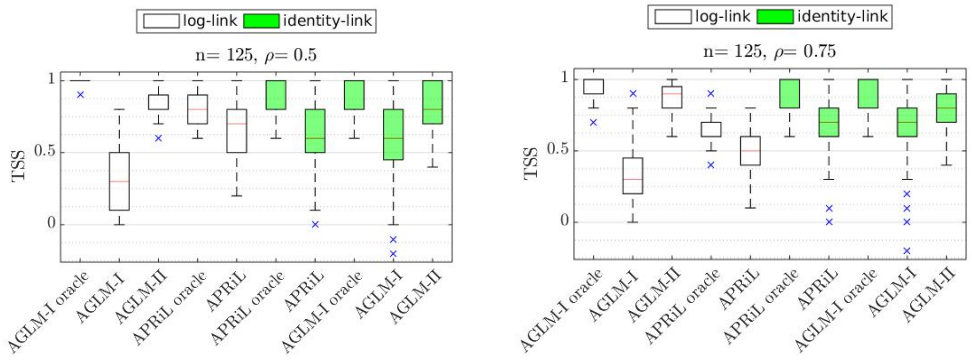
   where $\hat{\beta}(\mathrm{MLE})$ is the Maximum Likelihood estimate in Poisson log-linear regression model and $\bar{\gamma}$ is a positive constant. We denote the resulting algorithm by AGLM-I.

2. Second strategy [71]:

$$\hat{w}_j = \sqrt{2\tilde{\gamma} \log p \tilde{V}_j} + \frac{\tilde{\gamma} \log p}{3} \max_i |x_{ij}|, \tag{4.29}$$

   where $\tilde{V}_j = \hat{V}_j + \sqrt{2\tilde{\gamma} \log p \hat{V}_j \max_i x_{ij}^2} + 3\tilde{\gamma} \log p \max_i x_{ij}^2$, $\hat{V}_j = \sum_{i=1}^n x_{ij}^2 Y_i$ and $\tilde{\gamma}$ is a positive constant. In such a case the regularization parameter has to be fixed equal to 1. We denote the resulting algorithm by AGLM-II.

For computing the solution of these optimization problems we use the *glmnet*



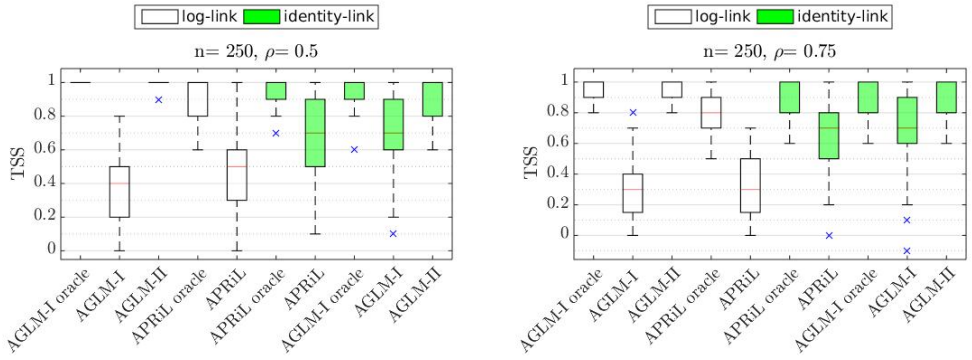
(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

Figure 4.1: Comparing distributions of TSS, fixing number of samples equal to $n = 125$.

MATLAB package [52]. Moreover, in the case of PRiL, APRiL, GLM and AGLM-I, we select the regularization parameter by means of the 10-fold Cross Validation (CV) [49] implemented in the same package. We use the mean squared error (MSE) for measuring the estimation accuracy of each solution. In Table 6.4 we show MSE values for the algorithms. It is evident that the algorithm based on the same model by which data have been generated achieves a lower MSE. In other words, the GLM and AGLM methods perform better when applied to the log-link dataset and the PRiL and APRiL methods when applied to the identity-link dataset. Finally, the MSE provided by the AGLM-II method is always smaller than the one obtained with AGLM-I.

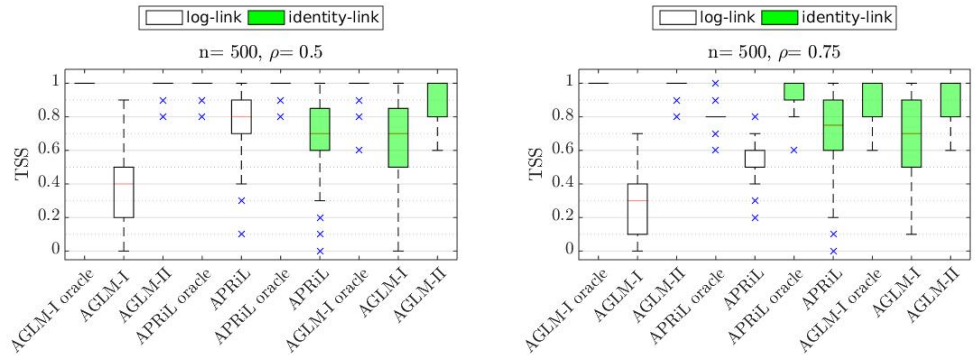
Moreover, we compare the variable selection performance of the GLM, AGLM, PRiL and APRiL methods by computing the confusion matrix which represents matches and mismatches between predicted active variables and exact ones. On the basis of the components of the confusion matrix, i.e. false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN), we compute the True Skill Score (TSS) which is defined as the balance between the true positive rate (or probability of detection) and the false alarm rate, i.e.

$$\mathrm{TSS} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} - \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}\,, \tag{4.30}$$



(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

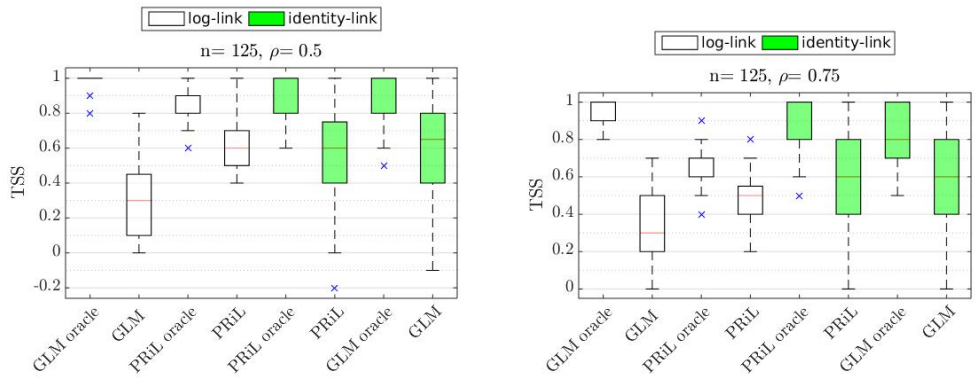
Figure 4.2: Comparing distributions of TSS, fixing number of samples equal to $n = 250$.



(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

Figure 4.3: Comparing distributions of TSS, fixing number of samples equal to $n = 500$.

and ranges from -1 to 1. The optimal variable selection is obtained when the TSS is 1 and a direct consequence of Theorem 8 is that the TSS value provided by the APRiL estimator converges to one in probability as n goes to infinity. For having a broader picture, for each method, in addition to the 10-fold cross-validated solution, we compute the solution which maximizes the TSS value along the regularization path, and we refer to it as the oracle solution. Oracle solutions allow us to make a performance assessment of the algorithms



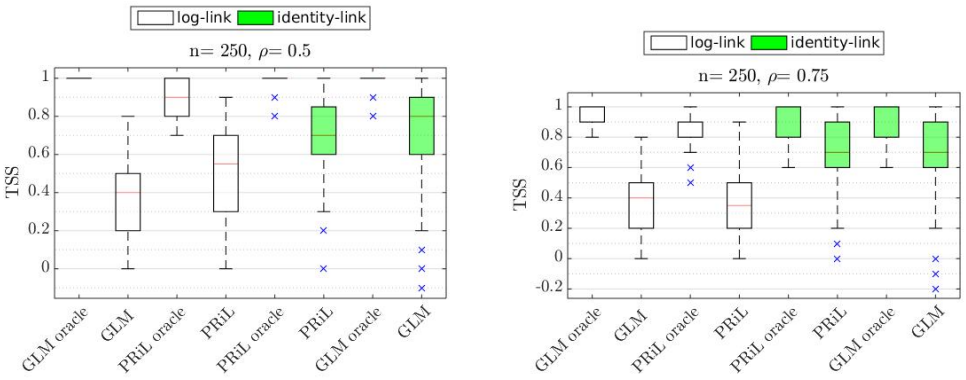
(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

Figure 4.4: Comparing distributions of TSS, fixing number of samples equal to $n = 125$.

independently of the choice of the regularization parameter. Each box-plot in Figures 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 shows the TSS distribution obtained by applying the algorithm written in the x-axis label to one hundred replicates of $\mathbf{Y}$. The first three figures contain the comparison between the adaptive methods whereas the last three figures contain the comparison between the non-adaptive methods. In Figures 4.1, 4.2 and 4.3 the first five box-plots refer to the AGLM and APRiL algorithms applied to the log-link dataset (equation (4.26)), whereas the second five box-plots refer to the algorithms applied to the identity-link dataset (equation (4.27)). In Figures 4.4, 4.5 and 4.6 the first four box-plots refer to the GLM and PRiL algorithms applied to the log-link dataset (equation (4.26)), whereas the second four box-plots refer to the algorithms applied to the identity-link dataset (equation (4.27)).

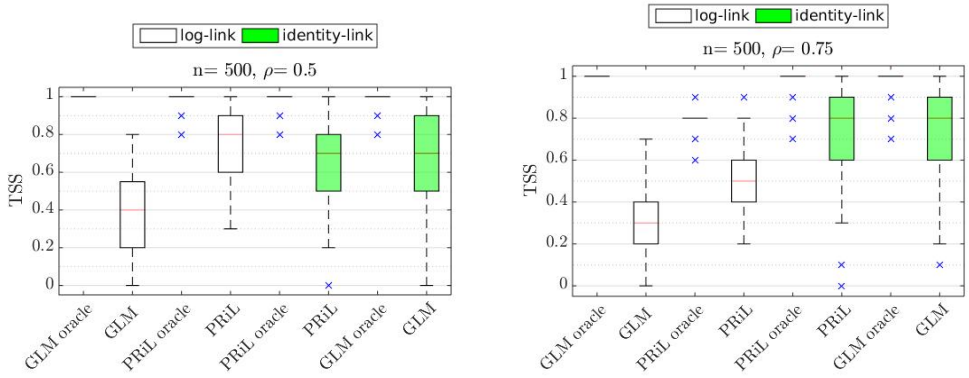
Some comments about variable selection results.

1. The TSS provided by oracle AGLM-I solutions is larger than the one provided by the oracle APRiL solutions in all the experiments we performed. This can be explained by the fact that AGLM-I method is based on the maximization of the Poisson likelihood, which is the actual distribution used for generating data. Oracle solutions provided by the APRiL method, which is based on an approximation of the Poisson



(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

Figure 4.5: Comparing distributions of TSS, fixing number of samples equal to $n = 250$.



(a) case $\rho = 0.5$ (b) case $\rho = 0.75$

Figure 4.6: Comparing distributions of TSS, fixing number of samples equal to $n = 500$.

log-likelihood, do not achieve the same performance.

2. The use of CV procedure for finding the regularization parameter reduces the performance of the variable selection so that it does not seem to be an efficient method in the case of small and moderately sized samples. However, for large scale problems the regularization path is more stable and the CV selects a solution closer to the oracle one [89]. In general, TSS distributions corresponding to cross validated solutions are over-

Table 4.2: Computation mean time in seconds.

Link	**Method**	time in s				
		$n = 10^4$	$n = 5\ 10^4$	$n = 10^5$	$n = 5\ 10^5$	$n = 10^6$
log	AGLM	$1.6\ 10^{-2}$	$7.5\ 10^{-2}$	$1.5\ 10^{-1}$	$8.3\ 10^{-1}$	2.0
	APRiL	$9.6\ 10^{-4}$	$5.0\ 10^{-3}$	$1.0\ 10^{-2}$	$6.1\ 10^{-2}$	$1.3\ 10^{-1}$
identity	AGLM	$1.4\ 10^{-2}$	$6.5\ 10^{-2}$	$1.3\ 10^{-1}$	$7.3\ 10^{-1}$	1.8
	APRiL	$9.7\ 10^{-4}$	$5.1\ 10^{-3}$	$1.1\ 10^{-2}$	$6.2\ 10^{-2}$	$1.3\ 10^{-1}$

dispersed and for each problem among the 100 replicates we can find a variable selection with a very low TSS value. Moreover CV behaves differently across algorithms. The striking fact is that the cross validated APRiL solution tends to produce a better variable selection than the cross validated AGLM-I one, overall in the case of smaller sized samples and log-link dataset.

3. In general, the TSS provided by AGLM-II solutions is larger than the one provided by cross-validated solutions. However, it is smaller than the one provided by oracle AGLM-I solutions for both datasets, and smaller than the one provided by oracle APRiL solutions in the case of the identity-link dataset.

4. In such simulations the performances of PRiL method are similar to the ones provided by APRiL and also the performances of GLM are similar to the ones provided by AGLM-I. This is mainly due to two factors: one is the fact that the correlation between predictors in the design matrix is not strong enough to compromise the variable selection property of non adaptivity methods and the other one is the fact that the number of samples is not so large. However, in some cases the oracle solutions provided by the adaptive methods are slightly better than the ones provided by the non-adaptive methods.

We now compare performances of the methods in terms of MSE and TSS

values. We note that, when the penalized criterion to be minimized is not adapted to the dataset, MSE values, evaluating the estimation capabilities, do not improve for AGLM, and deteriorate for APRiL with increasing n (see Table 6.4). At the same time, TSS values, evaluating the quality of variable selection, tend to increase with increasing n, or, at least, they do not deteriorate (see Figures 4.1, 4.2 and 4.3). Here, it is fundamental to note that the proposed datasets differ in values, but they have the same support $\mathcal{A}^*$, or, in other words, relevant variables are common to both datasets despite they assume different values. We can conclude that the use of the wrong model primarily affects the quality of estimation (MSE) and it is of minor importance with regard to the quality of variable selection (TSS). This is also confirmed by the fact that oracle solutions (including AGLM-II) provide almost optimal TSS values in spite of poor MSE values. Furthermore, we notice that the MSE values provided by the adaptive methods are similar to the ones provided by the corresponding non-adaptive methods: this is expected since the non adaptivity influences only the variable selection property but not the estimation property. Obtained results have been proven to be robust by varying the number of folds in the cross validation analysis and the definition of the adaptive weights. In this regard, we replicated the experiments introduced above by using the 5-fold cross validation and by choosing the adaptive weights of the AGLM-I method in a way analogous to the one described in equation (4.19) obtaining similar outcomes.

Finally, we check the numerical efficiency of the AGLM and APRiL algorithms. Following the above described setup, for each method we estimate the required CPU time for computing a solution of the problem having fixed the regularization parameter λ, for $\rho = 0.5$ and $n = 10^4$, $5\ 10^4$, 10^5, $5\ 10^5$, 10^6. In Table 4.2 we show the computational time by reporting the mean time in seconds to compute a solution of the regularization path. From Table 4.2 the benefit in terms of computational efficiency provided by the use of the APRiL method with respect to the AGLM method is evident. Indeed, in each case the computational cost is shrunk by a factor of about 15. In addition, another

advantage of the proposed method is that it does not suffer of convergence issues which are instead well-known in the case of the Poisson regression [88; 125].

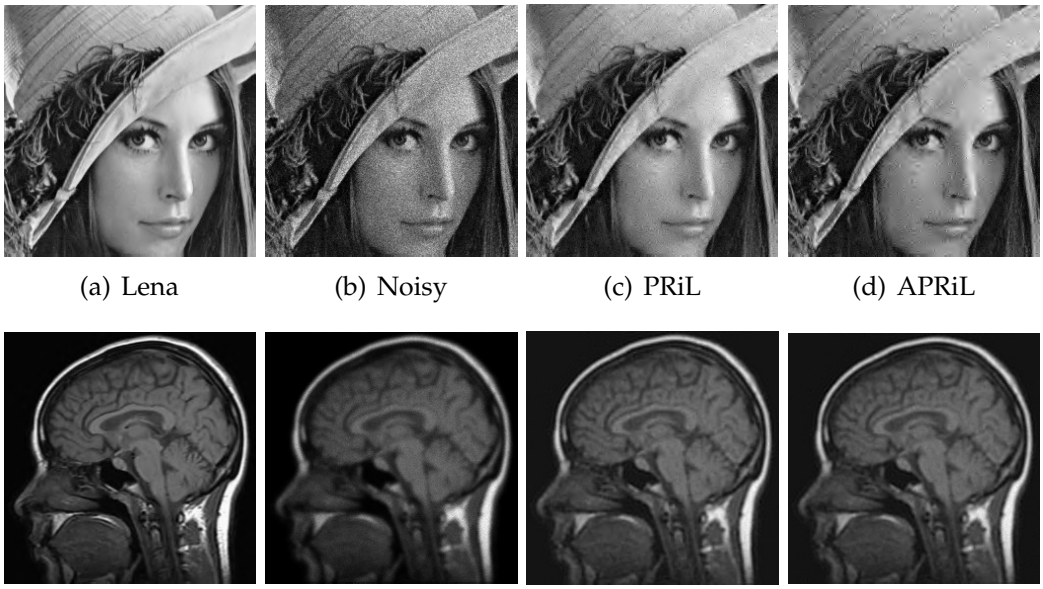
4.4.2 Sparse signal recovery application

We present two simulated experiments in sparse signal recovery: the first is an example of image denoising and the second is an example of image deblurring. Formally, these problems are described by equation (4.27) with $\mathbf{X} := \Omega\Psi$ where Ω represents the convolution with a given point-spread-function and Ψ is the standard synthesis operator which decomposes a given image f on an orthogonal wavelet basis $\{\psi_j\}_{j\in\{1,\dots,p\}}$. The image to recover is characterized by coefficients denoted by $(\beta_j^*)_{j\in\{1,\dots,p\}}$, i.e.

$$f^* = \sum_{j=1}^{p} \beta_j^* \psi_j. \tag{4.31}$$

In both cases we consider 256×256 images leading to large scale inverse problems with size $n = 65536$. For the denoising application we generate a compressed version of the “lena" image by thresholding its coefficients in the wavelet basis and we use the resulting image as the “true" image to recover. The true image is then represented by 17368 non-zero coefficients in the wavelet basis (about 74% of sparsity) with a Relative Square Error (RSE) of about 0.001% with respect to the original image. In this case the operator Ω is the identity. For the deblurring application we used a medical image and we performed the above described procedure for obtaining a “true" image represented by 10005 non-zero coefficients (about 85% of sparsity) corresponding to a RSE value of about 0.003% with respect to its original version. The convolution kernel of the operator Ω is a Gaussian function with $\sigma = 1.5$. We apply PRiL and APRiL methods to both problems. Thanks to their particular form, we can solve optimization problems by using an iterative forward-backward splitting algorithm: we perform a gradient step with step-size $\tau = 1.5$ and then we

apply the soft thresholding operator in the wavelet domain. Iteration stops when convergence is reached. The numerical optimization has been performed by using the MATLAB Numerical Tours [109].

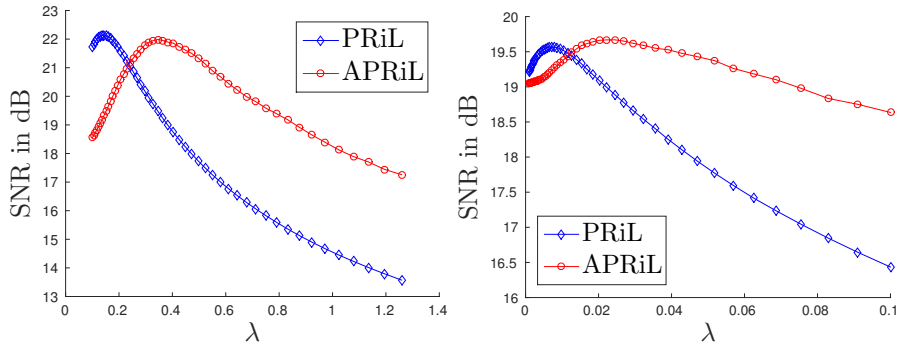


(a) Lena (b) Noisy (c) PRiL (d) APRiL

(e) MRI (f) Blurred and noisy (g) PRiL (h) APRiL

Figure 4.7: First row. Image denoising application: (a) true object, (b) noisy image, (c) recovered image with PRiL method, (d) recovered image with APRiL method. Second row. Image deblurring application: (e) true object, (f) blurred and noisy image, (g) recovered image with PRiL method, (h) recovered image with APRiL method.

In both examples we select the regularization parameter in order to maximize the Signal-to-Noise Ratio (SNR). We recall that SNR is one of the measure used for evaluating the reconstruction image quality (Chapter 3 in [13]) and it measures the ratio between the signal level and the noise. Figure 4.7 shows the results in the case of denoising (first row) and deblurring (second row) problems: for each example we show the true image, the noisy image, the best recovered image with PRiL and APRiL method, respectively. In the deblurring application the reconstructions provided by the two methods are very similar to each other whereas in the denoising application APRiL introduces some artifacts near the Lena's left eye (fourth panel top row in Figure 4.7). Figure



(a) Denoising application (b) Deblurring application

Figure 4.8: Comparison of SNRs as functions of λ between PRiL and APRiL methods. Left panel: SNR in the image denoising application. Right panel: SNR in the image deblurring application.

Table 4.3: Recovery performance results for image denoising and deblurring applications, respectively.

	denoising				
Method	**RSE**	**SNR in dB**	**PSNR in dB**	**confusion matrix**	
PRiL	0.09 %	22.14	27.85	TP = 11471 FP = 17002	FN = 5897 TN = 31166
APRiL	0.05 %	21.99	28.52	TP = 5668 FP = 2257	FN = 11700 TN = 45911
	deblurring				
Method	**RSE**	**SNR in dB**	**PSNR in dB**	**confusion matrix**	
PRiL	0.08 %	19.57	31.57	TP = 8052 FP = 17219	FN = 1953 TN = 38312
APRiL	0.08 %	19.67	31.66	TP = 5420 FP = 2684	FN = 4585 TN = 52847

4.8 shows the SNR of the recovered images as a function of the regularization parameter: in detail, we compare the SNR functions provided by the PRiL and APRiL method in each of the two applications. As we expect, the regularization parameter which maximizes the SNR function is different for each method: in both applications the optimal regularization parameter of PRiL method is smaller than the one of APRiL method. Therefore, if we choose the optimal

regularization parameter for PRiL method and we use this choice to compute the adaptive solution, the result provided by APRiL has a worse SNR. In Table 4.3 we show the following performance values: the RSE, the SNR and the Peak SNR (PSNR). PSNR is a commonly used image quality measure: it expresses the ratio between the power of the signal (the maximum possible value) and the power of noise that affects the quality of the signal. In addition, in Table 4.3 we provide, for each problem, the confusion matrix showing how many wavelet coefficients have been correctly recovered.

In both imaging applications, APRiL provides a smaller (or equal) RSE value and a higher PSNR value than the ones provided by PRiL. The SNR value provided by APRiL is higher in the deblurring application and it is smaller in the denoising one than the SNR value provided by PRiL. Confusion matrices show that APRiL provides a higher number of TN and a smaller number of FP, but also a higher number of FN. However, most of such incorrectly estimated coefficients have very small absolute value: indeed, they do not significantly contribute to the signal formation. Whereas the PRiL method provides higher number of TP, the sum of FN and FP is higher than the one provided by APRiL method in both applications. As we expected, we notice that the adaptive method tends to find more sparse solutions than the ones provided by the PRiL method.

4.5 Proofs

This section is devoted to prove the main results of the current Chapter. We prove Theorems 6, 7, 8, Corollary 7 and Proposition 9. In order to prove Theorem 6, we start by proving the following

Lemma 9. *Let y be a Poisson random variable with mean θ. Let $z > 0$ be such that $|z - \theta| \leq c\sqrt{\theta}$, where c is a positive constant smaller than $\sqrt{\theta}$. Then*

$$\mathbb{E}\left(D(z,y) - \frac{1}{2}\frac{(y-z)^2}{z}\right) = O\left(\frac{1}{\theta}\right), \text{ as } \theta \to \infty. \tag{4.32}$$

Proof. Following [153] we obtain

$$D(z,k) = \frac{1}{2}\frac{(k-z)^2}{z} - \frac{1}{6}\frac{(k-z)^3}{z^2} + \frac{1}{3}\frac{(k-z)^4}{z^3} + kR_3\left(\frac{k-z}{z}\right),$$

where $k \in \mathbb{N}$ and R_3 is defined as follows

$$R_3(\xi) = \int_0^{\xi} \frac{(t-\xi)^3}{(1+t)^4} dt,$$

where $\xi \geq -1$. By computing the moments of the Poisson random variable y centered in z we obtain

$$\begin{aligned} \mathbb{E}\left(D(z,y) - \frac{1}{2}\frac{(y-z)^2}{z}\right) &= -\frac{1}{6}\frac{\theta - 3\theta r - r^3}{z^2} \\ &+ \frac{1}{3}\frac{3\theta^2 + 6\theta r^2 - 4\theta r + \theta + r^4}{z^3} + \mathcal{E}(\theta), \end{aligned} \tag{4.33}$$

where $r := z - \theta$ and

$$\mathcal{E}(\theta) = \mathbb{E}\left(yR_3\left(\frac{y-z}{z}\right)\right) = \sum_{k=1}^{\infty} \frac{e^{-\theta}\theta^k}{k!} kR_3\left(\frac{k-z}{z}\right).$$

To conclude we now prove that $\mathcal{E}(\theta) = O(\frac{1}{\theta})$. Following the idea of the proof given in [153], we split the series into two parts: in the first ranging k between 0 and $\lfloor \frac{z}{2} \rfloor$ and in the second one $k \geq \lfloor \frac{z}{2} \rfloor + 1$, where $\lfloor \chi \rfloor$ denotes the integer part of χ. We observe that for k from 1 to $\lfloor \frac{z}{2} \rfloor$, or equivalently $\xi \in (-1, -\frac{1}{2}]$, then

$$(1+\xi)|R_3(\xi)| \leq \frac{1}{e}. \tag{4.34}$$

Since $\frac{\theta^s}{s!} = \frac{\theta^s}{\Gamma(s+1)}$ is monotonically increasing for $0 \leq s \leq \lfloor \frac{\theta + c\sqrt{\theta}}{2} \rfloor$, using

equation (4.34) and the Stirling formula we obtain

$$\begin{aligned}\left|\sum_{k=1}^{\lfloor \frac{z}{2} \rfloor} \frac{e^{-\theta}\theta^k}{k!} k R_3\left(\frac{k-z}{z}\right)\right| &\leq \frac{1}{e}\sum_{k=1}^{\lfloor \frac{z}{2} \rfloor} z e^{-\theta}\frac{\theta^k}{k!} \\ &\leq \frac{1}{e}\sum_{k=1}^{\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor} z e^{-\theta}\frac{\theta^k}{k!} \leq \frac{1}{e}\left\lfloor \frac{\theta+c\sqrt{\theta}}{2}\right\rfloor e^{-\theta}\frac{\theta^{\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor}}{\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor !} z \\ &\leq \frac{e^{-\theta-1+\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor}}{\sqrt{2\pi}}\left(\frac{\theta}{\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor}\right)^{\lfloor \frac{\theta+c\sqrt{\theta}}{2} \rfloor}\left\lfloor \frac{\theta+c\sqrt{\theta}}{2}\right\rfloor^{\frac{1}{2}} z\,.\end{aligned} \tag{4.35}$$

Since $\frac{\theta+c\sqrt{\theta}}{2} - 1 \leq \left\lfloor \frac{\theta+c\sqrt{\theta}}{2}\right\rfloor \leq \frac{\theta+c\sqrt{\theta}}{2}$ and $\frac{\theta}{\left\lfloor \frac{\theta+c\sqrt{\theta}}{2}\right\rfloor} > 1$ for θ large enough, the upper bound in inequality (4.35) can be bounded by

$$M(\theta) := \frac{e^{-\frac{1}{2}\theta\left(1+2\theta^{-1}-c\theta^{-\frac{1}{2}}-v\log\left(\frac{2}{v-2\theta^{-1}}\right)\right)}(\theta v)^{\frac{3}{2}}}{2\sqrt{\pi}} \tag{4.36}$$

where $v := 1 + c\theta^{-\frac{1}{2}}$. Then $M(\theta) \to 0$ exponentially as $\theta \to \infty$. Now we consider $k \geq \lfloor \frac{z}{2} \rfloor + 1$, or equivalently $\xi > -\frac{1}{2}$. Since

$$|R_3(\xi)| \leq 4\xi^4, \tag{4.37}$$

we obtain

$$\begin{aligned}&\left|\sum_{k=\lfloor \frac{z}{2} \rfloor+1}^{\infty} \frac{e^{-\theta}\theta^k}{k!} k R_3\left(\frac{k-z}{z}\right)\right| \leq 4\sum_{k=0}^{\infty}\frac{e^{-\theta}\theta^k}{k!}k\left(\frac{k-z}{z}\right)^4 \\ =\;& 4\mathbb{E}\left(\frac{(y-z)^5}{z^4}+\frac{(y-z)^4}{z^3}\right) \\ =\;& 4\frac{3\theta^3+6\theta^2r^2-16\theta^2r+11\theta^2+\theta(r-1)^4}{z^4} \\ \leq\;& 4\frac{(3+6c^2)\theta^3+16c\theta^2\sqrt{\theta}+11\theta^2+\theta(c\sqrt{\theta}+1)^4}{(\theta-c\sqrt{\theta})^4} = O\left(\frac{1}{\theta}\right),\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}((y-z)^5) &= \theta^5 - 5\theta^4(z-2) + 5\theta^3(2z^2-6z+5)\\ &- 5\theta^2(2z^3-6z^2+7z-3)\\ &+ \theta(5z^4-10z^3+10z^2-5z+1) - z^5\end{aligned}$$

$$\begin{aligned}\mathbb{E}((y-z)^4) &= \theta^4 + \theta^3(6-4z) + \theta^2(6z^2-12z+7)\\ &+ \theta(-4z^3+6z^2-4z+1) + z^4,\end{aligned}$$

are the 5-th and 4-th moments of the Poisson random variable y centered in z. □

Proof of Theorem 6. By the triangular inequality we have

$$\begin{aligned}\left|\mathbb{E}\left(D(z,y) - \frac{1}{2}\frac{(y-z)^2}{y+1}\right)\right| &\leq \left|\mathbb{E}\left(D(z,y) - \frac{1}{2}\frac{(y-z)^2}{z}\right)\right| \\ &+ \left|\mathbb{E}\left(\frac{1}{2}\frac{(y-z)^2}{z} - \frac{1}{2}\frac{(y-z)^2}{y+1}\right)\right|.\end{aligned} \tag{4.38}$$

Then, to get the thesis, thanks to Lemma 9, it is sufficient to prove that

$$\mathbb{E}\left(\frac{(y-z)^2}{z} - \frac{(y-z)^2}{y+1}\right) = O\left(\frac{1}{\sqrt{\theta}}\right), \text{ as } \theta \to \infty. \tag{4.39}$$

By writing the left hand side of the equation (4.39) as the difference between the second moments of a Poisson variable centered in z and in $z+1$, we obtain

$$\begin{aligned}&\frac{1}{z}\mathbb{E}\left((y-z)^2\right) - \frac{1}{\theta}\mathbb{E}\left((y-z-1)^2\right) + \frac{1}{\theta}e^{-\theta}(z+1)^2\\ =\ &\frac{\theta^2-2\theta z+\theta+z^2}{z} - \frac{\theta^2-\theta(2z+1)+(z+1)^2-(z+1)^2e^{-\theta}}{\theta}.\end{aligned}$$

By some manipulations and by using that $|z-\theta| \leq c\sqrt{\theta}$ we get

$$\begin{aligned}\left|\mathbb{E}\left(\frac{(y-z)^2}{z}-\frac{(y-z)^2}{y+1}\right)\right| &\leq e^{-\theta}\left|\frac{(z+1)^2}{\theta}\right|+\left|\frac{(\theta-z)^3}{z\theta}\right| \\ &+ \left|\frac{(\theta-z)^2}{z\theta}\right|+\left|\frac{3(\theta-z)}{\theta}-\frac{1}{\theta}\right| \\ &\leq e^{-\theta}\frac{(\theta+c\sqrt{\theta}+1)^2}{\theta}+\frac{c^3}{\sqrt{\theta}-c} \\ &+ \frac{c^2}{\theta-c\sqrt{\theta}}+\frac{3c}{\sqrt{\theta}}+\frac{1}{\theta}=O\left(\frac{1}{\sqrt{\theta}}\right).\end{aligned}$$

□

To prove Theorem 7 we need some preliminary results. We start by defining

$$\epsilon := \mathbf{Y}-\mathbf{X}\beta^*. \tag{4.40}$$

We observe that the components ϵ_i are independent random variables with zero mean and $Var(\epsilon_i) = (\mathbf{X}\beta^*)_i$, for all $i \in \{1,\dots,n\}$. Hereafter, for easy of notation we suppress the superscript (n) from the estimators.

Lemma 10. *There exists a constant $G < +\infty$ such that*

$$\mathbb{E}\left(\|\mathbf{X}^T\Lambda^2\epsilon\|_2^4\right) \leq p^2G^2(\tau_{\max}(\mathbf{X}^T\mathbf{X}))^2. \tag{4.41}$$

Proof of Lemma 10. We compute the term in the l.h.s. in equation (4.41). We have

$$\mathbb{E}\left(\|\mathbf{X}^T\Lambda^2\epsilon\|_2^4\right) = \mathbb{E}((\mathbf{D}^T\mathbf{X}\mathbf{X}^T\mathbf{D})^2) \tag{4.42}$$

with $\mathbf{D} := \Lambda^2\epsilon$. For the Singular Value Decomposition we can write

$$\mathbf{X}\mathbf{X}^T = \mathbf{U}^T\Sigma\mathbf{U}\,,$$

where $\mathbf{U}$ is an orthogonal matrix and Σ a diagonal matrix containing the eigenvalues of $\mathbf{X}^T\mathbf{X}$. We define $\mathbf{H} := \mathbf{U}\mathbf{D} = \mathbf{U}\Lambda^2\epsilon$. The i-th component of $\mathbf{H}$ is

given by

$$H_i = \sum_{l=1}^{n} u_{il} \frac{\epsilon_l}{Y_l + 1} \,,$$

where u_{il} represents the (i,l)-entry of the matrix $\mathbf{U}$. Since $\frac{\epsilon_l}{Y_l+1}$ takes values between $-(\mathbf{X}\beta^*)_l$ and 1, we have that each component H_i takes values in a compact subset $[R,S]$. Therefore, as $\mathbf{H} \in [R,S]^n$, the quadratic form $(\mathbf{H}^T\Sigma\mathbf{H})^2$ admits a maximum, i.e. there exists an $\mathbf{H}^*$ such that $(\mathbf{H}^T\Sigma\mathbf{H})^2 \leq ((\mathbf{H}^*)^T\Sigma\mathbf{H}^*)^2$. Then

$$\begin{aligned} \mathbb{E}\left(\|\mathbf{X}^T\Lambda^2\epsilon\|_2^4\right) = \mathbb{E}((\mathbf{H}^T\Sigma\mathbf{H})^2) &\leq ((\mathbf{H}^*)^T\Sigma\mathbf{H}^*)^2 \\ &\leq p^2G^2(\tau_{\max}(\mathbf{X}^T\mathbf{X}))^2, \end{aligned} \tag{4.43}$$

with

$$G := \max_{\substack{i\in\{1,\dots,n\}:\\ \Sigma_{ii}\neq 0}} (H_i^*)^2. \tag{4.44}$$

□

Corollary 8. *Under assumption (H1) there exists a constant $G < +\infty$ such that we have the following bound*

$$\mathbb{E}(\|\hat{\beta}(\mathrm{PRLS}) - \beta^*\|_2^2) \leq \frac{pGBn}{(bn)^2}, \tag{4.45}$$

where $\hat{\beta}(\mathrm{PRLS})$ is the reweighted least square estimator defined as follows

$$\hat{\beta}(\mathrm{PRLS}) = \arg\min_{\beta} \frac{1}{2}\|\Lambda(\mathbf{Y} - \mathbf{X}\beta)\|_2^2 \,. \tag{4.46}$$

Proof of Corollary 8. By using optimality conditions of problem in equation (4.46) and the definition of ϵ we have

$$\hat{\beta}(\mathrm{PRLS}) - \beta^* = (\mathbf{X}^T\Lambda^2\mathbf{X})^{-1}(\mathbf{X}^T\Lambda^2\epsilon) \,. \tag{4.47}$$

Then, by using the Cauchy-Schwartz inequality we obtain

$$\mathbb{E}(\|\hat{\beta}(\text{PRLS}) - \beta^*\|_2^2) \leq \sqrt{\mathbb{E}(\|(\mathbf{X}^T\Lambda^2\mathbf{X})^{-1}\|_2^4)\mathbb{E}(\|\mathbf{X}^T\Lambda^2\epsilon\|_2^4)}, \tag{4.48}$$

where

$$\|(\mathbf{X}^T\Lambda^2\mathbf{X})^{-1}\|_2^4 = \frac{1}{(\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X}))^4}\,. \tag{4.49}$$

By assumption (H1) and Lemma 10 we have the thesis. □

Proof of Theorem 7. We want to prove the bound in equation (4.22). From Corollary 8, since

$$\begin{aligned}\mathbb{E}(\|\hat{\beta}_{(\mathbf{w},\lambda)} - \beta^*\|_2^2) \leq \mathbb{E}(\|\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS})\|_2^2) \\ + \mathbb{E}(\|\hat{\beta}(\text{PRLS}) - \beta^*\|_2^2),\end{aligned} \tag{4.50}$$

we have to establish a bound for the first term of the r.h.s. of (4.50). In order to do so, we follow similar arguments as in the proof of Theorem 3.1 in [157]. By definition of $\hat{\beta}_{(\mathbf{w},\lambda)}$ (equation (4.15)), the following inequality applies

$$\begin{aligned}&\frac{1}{2}\|\Lambda(\mathbf{Y} - \mathbf{X}\hat{\beta}_{(\mathbf{w},\lambda)})\|_2^2 - \frac{1}{2}\|\Lambda(\mathbf{Y} - \mathbf{X}\hat{\beta}(\text{PRLS}))\|_2^2 \\ &\leq \lambda\sum_{j=1}^{p} w_j(|\hat{\beta}(\text{PRLS})_j| - |(\hat{\beta}_{(\mathbf{w},\lambda)})_j|).\end{aligned} \tag{4.51}$$

From the optimality conditions of the optimization problem in equation (4.46), we have

$$\begin{aligned}&\frac{1}{2}\|\Lambda(\mathbf{Y} - \mathbf{X}\hat{\beta}_{(\mathbf{w},\lambda)})\|_2^2 - \frac{1}{2}\|\Lambda(\mathbf{Y} - \mathbf{X}\hat{\beta}(\text{PRLS}))\|_2^2 \\ &= \frac{1}{2}(\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS}))^T\mathbf{X}^T\Lambda^2\mathbf{X}(\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS})),\end{aligned} \tag{4.52}$$

and we notice that

$$\begin{aligned}&\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})\|\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS})\|_2^2 \\ &\leq (\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS}))^T\mathbf{X}^T\Lambda^2\mathbf{X}(\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS}))\end{aligned} \tag{4.53}$$

and

$$\sum_{j=1}^{p} w_j(|\hat{\beta}(\text{PRLS})_j| - |(\hat{\beta}_{(\mathbf{w},\lambda)})_j|) \leq \sqrt{\sum_{j=1}^{p} w_j^2}\|\hat{\beta}(\text{PRLS}) - \hat{\beta}_{(\mathbf{w},\lambda)}\|_2. \tag{4.54}$$

Using (4.51), (4.52), (4.53) and (4.54) we obtain

$$\|\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS})\|_2 \leq \frac{2\lambda\sqrt{\sum_{j=1}^{p} w_j^2}}{\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})}, \tag{4.55}$$

and finally, the Cauchy Schwartz inequality and assumption (H1) lead to

$$\mathbb{E}(\|\hat{\beta}_{(\mathbf{w},\lambda)} - \hat{\beta}(\text{PRLS})\|_2^2) \leq \frac{4\lambda^2\sqrt{\mathbb{E}\left(\left(\sum_{j=1}^{p} w_j^2\right)^2\right)}}{(bn)^2}. \tag{4.56}$$

The thesis follows from equations (4.50), (4.56) and Corollary 8. □

Proof of Theorem 8. For brevity we denote the APRiL estimator by $\hat{\beta}$. To prove the model selection consistency we prove that for $n \to +\infty$

$$\mathbb{P}(\forall j \in (\mathcal{A}^*)^C, \hat{\beta}_j = 0) \longrightarrow 1 \tag{4.57}$$

and

$$\mathbb{P}(\forall j \in \mathcal{A}^*, |\hat{\beta}_j| > 0) \longrightarrow 1\ . \tag{4.58}$$

We now prove equation (4.57). The functional defined in equation (4.15) is convex and not differentiable and $\mathcal{C}$ is a convex set. Then the solution $\hat{\beta}$ is characterized by the Karush-Kuhn-Tucker (KKT) optimality conditions [23]:

- $(\mathbf{X}\hat{\beta})_i \geq 0\ \forall\ i \in \{1, \ldots, n\}$;

- $\nu_i \geq 0 \; \forall \; i \in \{1, \dots, n\}$;

- $\nu_i(\mathbf{X}\hat{\beta})_i = 0 \; \forall \; i \in \{1, \dots, n\}$;

- if $\hat{\beta}_j \neq 0$
$$-\mathbf{x}_j^T \Lambda^2(\mathbf{Y} - \mathbf{X}\hat{\beta}) + \lambda \hat{w}_j sgn(\hat{\beta}_j) - \mathbf{x}_j^T \nu = 0, \tag{4.59}$$
where sgn is the sign function;

- if $\hat{\beta}_j = 0$
$$-\mathbf{x}_j^T \Lambda^2(\mathbf{Y} - \mathbf{X}\hat{\beta}) + \lambda \hat{w}_j s_j - \mathbf{x}_j^T \nu = 0, \tag{4.60}$$
with $s_j \in [-1, 1]$.

ν is the n-dimensional vector whose components are the Lagrangian multipliers. Thanks to KKT conditions, the event $\{\forall j \in (\mathcal{A}^*)^C, \hat{\beta}_j = 0\}$ can be written as

$$\{\mathbf{x}_j^T \Lambda^2(\mathbf{Y} - \mathbf{X}_{\mathcal{A}^*}\hat{\beta}_{\mathcal{A}^*}) + \mathbf{x}_j^T \nu = \lambda \hat{w}_j s_j, \forall j \in (\mathcal{A}^*)^C\}, \tag{4.61}$$

where $|s_j| \leq 1$ (see equation (4.60)), $\mathbf{X}_{\mathcal{A}^*}$ is the matrix constituted by the columns $\mathbf{x}_j$ and $\hat{\beta}_{\mathcal{A}^*}$ is the vector constituted by the components $\hat{\beta}_j$ with $j \in \mathcal{A}^*$. By taking the absolute value of each equation in (4.61) the event takes the form

$$\{|\mathbf{x}_j^T(\Lambda^2(\mathbf{Y} - \mathbf{X}_{\mathcal{A}^*}\hat{\beta}_{\mathcal{A}^*}) + \nu)| \leq \lambda \hat{w}_j, \forall j \in (\mathcal{A}^*)^C\}. \tag{4.62}$$

This implies that equation (4.57) is equivalent to

$$\mathbb{P}\left(\exists j \in (\mathcal{A}^*)^C, \left|\mathbf{x}_j^T(\Lambda^2(\mathbf{Y} - \mathbf{X}_{\mathcal{A}^*}\hat{\beta}_{\mathcal{A}^*}) + \nu)\right| > \lambda \hat{w}_j\right) \to 0$$

as $n \to +\infty$. We set

$$\hat{S}_j := |\hat{\beta}(\text{PRiL})_j| + \left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta},$$

$$\hat{\eta} := \min_{j \in \mathcal{A}^*} \hat{S}_j,$$

$$\eta := \min_{j \in \mathcal{A}^*} |\beta_j^*| + \left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta},$$

and

$$\hat{E}_j := \left| \mathbf{x}_j^T (\Lambda^2 (\mathbf{Y} - \mathbf{X}_{\mathcal{A}^*} \hat{\beta}_{\mathcal{A}^*}) + \nu) \right|.$$

Then

$$\begin{aligned} \mathbb{P}\left(\exists j \in (\mathcal{A}^*)^C \hat{E}_j > \lambda w \hat{o}_j\right) \quad &\leq \textstyle\sum_{j \in (\mathcal{A}^*)^C} \mathbb{P}\left(\hat{E}_j > \lambda w \hat{o}_j, \hat{\eta} > \frac{\eta}{2}, \hat{S}_j \leq \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) \\ &\quad + \textstyle\sum_{j \in (\mathcal{A}^*)^C} \mathbb{P}\left(\hat{S}_j > \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) + \mathbb{P}\left(\hat{\eta} \leq \frac{\eta}{2}\right). \quad (4.63) \end{aligned}$$

The idea is to determine three bounds M_1, M_2 and M_3 depending on n, such that

$$\mathbb{P}\left(\hat{\eta} \leq \frac{\eta}{2}\right) \leq M_1, \quad (4.64)$$

$$\sum_{j \in (\mathcal{A}^*)^C} \mathbb{P}\left(\hat{S}_j > \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) \leq M_2, \quad (4.65)$$

$$\sum_{j \in (\mathcal{A}^*)^C} \mathbb{P}\left(\hat{E}_j > \lambda w \hat{o}_j, \hat{\eta} > \frac{\eta}{2}, \hat{S}_j \leq \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) \leq M_3 \quad (4.66)$$

and M_1, M_2 and M_3 go to 0 as $n \to +\infty$. Let us start with the determination of bound M_1. Using Corollary 7, it follows that

$$\begin{aligned} \mathbb{P}\left(\hat{\eta} \leq \frac{\eta}{2}\right) &\leq \mathbb{P}\left(\|\hat{\beta}(\text{PRiL}) - \beta^*\|_2 \geq \frac{\eta}{2}\right) \\ &\leq \frac{2}{\eta}\left(\frac{2\lambda_1\sqrt{p} + \sqrt{pGBn}}{bn}\right) =: M_1. \quad (4.67) \end{aligned}$$

For the determination of bound M_2, we use again Corollary 7. We have that

$$\begin{aligned}\sum_{j\in(\mathcal{A}^*)^C}\mathbb{P}\left(\hat{S}_j > \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) &\leq \frac{\mathbb{E}\left(\|\hat{\beta}(\mathrm{PRiL}) - \beta^*\|_2\right) + p\left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta}}{\left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}} \\ &\leq \frac{1}{(\frac{\lambda}{n})^{\frac{1}{\gamma}}}\left(\frac{2\lambda_1\sqrt{p}+\sqrt{pGBn}}{bn} + \frac{p}{n^{\frac{1}{\gamma}+\delta}}\right) \\ &=: M_2. \end{aligned} \tag{4.68}$$

Finally, for the determination of bound M_3, we write

$$\sum_{j\in(\mathcal{A}^*)^C}\mathbb{P}\left(\hat{E}_j > \lambda\hat{w}_j, \hat{\eta} > \frac{\eta}{2}, \hat{S}_j \leq \left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right) \leq 2\frac{\mathbb{E}\left(\sum_{j\in(\mathcal{A}^*)^C}\hat{E}_j\mathbf{1}_{\left\{\hat{\eta} > \frac{\eta}{2}\right\}}\right)}{n},$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. By definition of ϵ (equation (4.40)), we have

$$\begin{aligned}\sum_{j\in(\mathcal{A}^*)^C}\hat{E}_j &= \sum_{j\in(\mathcal{A}^*)^C}\left|\mathbf{x}_j^T\Lambda^2\mathbf{X}_{\mathcal{A}^*}(\beta^*_{\mathcal{A}^*} - \hat{\beta}_{\mathcal{A}^*}) + \mathbf{x}_j^T\Lambda^2\epsilon + \mathbf{x}_j^T\nu\right| \\ &\leq \sum_{j\in(\mathcal{A}^*)^C}\|\mathbf{x}_j^T\Lambda\|_2\sqrt{\tau_{\max}(\mathbf{X}^T\Lambda^2\mathbf{X})}\|\beta^*_{\mathcal{A}^*} - \hat{\beta}_{\mathcal{A}^*}\|_2 \\ &+ \sum_{j\in(\mathcal{A}^*)^C}\left|\mathbf{x}_j^T\nu\right| + \sum_{j\in(\mathcal{A}^*)^C}\|\mathbf{x}_j^T\Lambda\|_2\|\Lambda\epsilon\|_2. \end{aligned} \tag{4.69}$$

By using assumption (H4), we get

$$\sum_{j\in(\mathcal{A}^*)^C}\|\mathbf{x}_j^T\Lambda\|_2 \leq p\max_{j=1,\ldots,p}\|\mathbf{x}_j\|_2 \leq pL, \tag{4.70}$$

and

$$\mathbb{E}(\|\Lambda\epsilon\|_2) = \mathbb{E}\left(\sqrt{\sum_{i=1}^n\left(\frac{\epsilon_i}{\sqrt{Y_i+1}}\right)^2}\right) \leq \sqrt{2n} \tag{4.71}$$

where we used that $\mathbb{E}\left(\frac{\epsilon_i^2}{Y_i+1}\right) \leq 2$ for all $i \in \{1,\dots,n\}$. Following the idea of the proof of Lemma 7, in particular the calculus which leads to equation (4.45), we have

$$\|\hat{\beta}_{\mathcal{A}^*} - \hat{\beta}(\mathrm{PRLS})_{\mathcal{A}^*}\|_2 \leq \frac{2\lambda\sqrt{p}\frac{1}{\hat{\eta}^\gamma}}{\tau_{\min}(\mathbf{X}_{\mathcal{A}^*}^T \Lambda^2 \mathbf{X}_{\mathcal{A}^*})}. \tag{4.72}$$

Thanks to the Cauchy-Schwartz inequality, equations (4.69), (4.70), (4.71), (4.72) and hypothesis (H1) we obtain the following bound

$$\begin{aligned}
& \mathbb{E}\left(\sum_{j\in(\mathcal{A}^*)^C} \hat{E}_j \mathbf{1}_{\left\{\hat{\eta} > \frac{\eta}{2}\right\}}\right) \\
\leq\ & pL\sqrt{\mathbb{E}\left(\tau_{\max}(\mathbf{X}^T\Lambda^2\mathbf{X})\right)\mathbb{E}\left(\|\beta^*_{\mathcal{A}^*} - \hat{\beta}_{\mathcal{A}^*}\|_2^2 \mathbf{1}_{\left\{\hat{\eta} > \frac{\eta}{2}\right\}}\right)} \\
+\ & \mathbb{E}\left(\sum_{j\in(\mathcal{A}^*)^C} \left|\mathbf{x}_j^T \nu\right| \mathbf{1}_{\left\{\hat{\eta} > \frac{\eta}{2}\right\}}\right) + pL\mathbb{E}(\|\Lambda\epsilon\|_2) \\
\leq\ & pL\sqrt{Bn}\left(\frac{2\lambda\sqrt{p}\left(\frac{\eta}{2}\right)^{-\gamma} + \sqrt{pGBn}}{bn}\right) \\
+\ & \mathbb{E}\left(\sum_{j\in(\mathcal{A}^*)^C} \left|\mathbf{x}_j^T \nu\right| \mathbf{1}_{\left\{\hat{\eta} > \frac{\eta}{2}\right\}}\right) + pL\sqrt{2n}.
\end{aligned} \tag{4.73}$$

From optimality conditions in equations (4.59) and (4.60) it follows that

$$|\mathbf{x}_j^T \nu| \leq |\mathbf{x}_j^T \Lambda^2 (\mathbf{Y} - \mathbf{X}\hat{\beta})| + \lambda \hat{w}_j,$$

so we have

$$\begin{aligned}
&\mathbb{E}\left(\sum_{j\in(\mathcal{A}^*)^C}\left|\mathbf{x}_j^T\nu\right|\mathbf{1}_{\left\{\hat{\eta}>\frac{\eta}{2}\right\}}\right)\\
\leq\ &\mathbb{E}\left(\sum_{j\in\mathcal{A}^*}|\mathbf{x}_j^T\Lambda^2\epsilon|\right)+\mathbb{E}\left(\sum_{j\in\mathcal{A}^*}|\mathbf{x}_j^T\Lambda^2\mathbf{X}(\beta^*-\hat{\beta})|\mathbf{1}_{\left\{\hat{\eta}>\frac{\eta}{2}\right\}}\right)\\
+\ &\lambda\mathbb{E}\left(\sum_{j\in\mathcal{A}^*}\hat{w}_j\mathbf{1}_{\left\{\hat{\eta}>\frac{\eta}{2}\right\}}\right)\\
\leq\ &pL\sqrt{2n}+pL\sqrt{Bn}\left(\frac{2\lambda\sqrt{p}n^{1+\gamma\delta}+\sqrt{pGBn}}{bn}\right)+\lambda p\left(\frac{2}{\eta}\right)^{\gamma},
\end{aligned}$$

where we have used the following bound

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{j=1}^{p}\hat{w}_j^2\right)^2\right) &= \mathbb{E}\left(\left(\sum_{j=1}^{p}\frac{1}{\left(|\hat{\beta}(\mathrm{PRiL})_j|+\left(\frac{1}{n}\right)^{\frac{1}{\gamma}+\delta}\right)^{2\gamma}}\right)^2\right)\\
&\leq p^2n^{4(1+\delta\gamma)}. \qquad (4.74)
\end{aligned}$$

Then, we obtain

$$\begin{aligned}
&\sum_{j\in(\mathcal{A}^*)^C}\mathbb{P}\left(\hat{E}_j>\lambda\hat{w}_j,\hat{\eta}>\frac{\eta}{2},\hat{S}_j\leq\left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}\right)\\
\leq\ &\frac{4pL}{n}\left(\sqrt{2n}+\frac{B\sqrt{pG}}{b}+\sqrt{Bnp}\frac{\lambda\left(\frac{2}{\eta}\right)^{\gamma}+\lambda n^{1+\delta\gamma}}{bn}+\lambda\frac{2^{\gamma-1}}{\eta^{\gamma}L}\right)\\
=:\ &M_3. \qquad (4.75)
\end{aligned}$$

Now we prove that M_1, M_2 and M_3 go to 0 as $n\to+\infty$.

$$M_3\to 0$$

since $\frac{\sqrt{n}\lambda\left(\frac{2}{\eta}\right)^{\gamma}}{n^2} \longrightarrow 0$, $\frac{\sqrt{n}\lambda n^{1+\delta\gamma}}{n^2} \longrightarrow 0$ and $\frac{\lambda}{n}\left(\frac{2}{\eta}\right)^{\gamma} \longrightarrow 0$ as $n \to +\infty$, for the assumption (H3 c) and for the positivity of constants γ and δ;

$$M_2 = \frac{1}{(\frac{\lambda}{n})^{\frac{1}{\gamma}}}\left(\frac{2\lambda_1\sqrt{p}+\sqrt{pGBn}}{bn} + \frac{p}{n^{\frac{1}{\gamma}+\delta}}\right) \to 0$$

since $\frac{\lambda_1}{\left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}n} \longrightarrow 0$ as $n \to +\infty$ for assumptions (H2) and (H3 a), $\frac{\sqrt{n}}{n\left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}} = \frac{1}{\left(\lambda n^{\frac{\gamma}{2}-1}\right)^{\frac{1}{\gamma}}} \longrightarrow 0$ as $n \to +\infty$ for the assumption (H3 a), and $\frac{1}{\left(\frac{\lambda}{n}\right)^{\frac{1}{\gamma}}n^{\frac{1}{\gamma}+\delta}} \longrightarrow 0$ as $n \to +\infty$ for the assumption (H3 b);

$$M_1 = \frac{2}{\eta}\left(\frac{2\lambda_1\sqrt{p}+\sqrt{pGBn}}{bn}\right) \to 0$$

since $\frac{\lambda_1}{n\eta} \longrightarrow 0$ as $n \to +\infty$, for the assumption (H2) and the definition of η, and $\frac{\sqrt{n}}{n\eta} = O(\frac{1}{\sqrt{n}})$ as $n \to +\infty$.

Now we prove equation (4.58). It is sufficient to show that

$$\mathbb{P}\left(\min_{j\in\mathcal{A}^*}|\hat{\beta}_j| > 0\right) \longrightarrow 1, \quad n \to +\infty.$$

By equation (4.72) we have

$$\min_{j\in\mathcal{A}^*}|\hat{\beta}_j| > \min_{j\in\mathcal{A}^*}|\hat{\beta}(\mathrm{PRLS})_j| - \frac{2\lambda\sqrt{p}\hat{\eta}^{-\gamma}}{\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})}\,, \tag{4.76}$$

where

$$\min_{j\in\mathcal{A}^*}|\hat{\beta}(\mathrm{PRLS})_j| \geq \min_{j\in\mathcal{A}^*}|\beta^*_j| - \|\beta^*_{\mathcal{A}^*} - \hat{\beta}(\mathrm{PRLS})_{\mathcal{A}^*}\|_2\,. \tag{4.77}$$

Since $\min_{j\in\mathcal{A}^*}|\beta^*_j| > 0$, to conclude we show that $\|\beta^*_{\mathcal{A}^*} - \hat{\beta}(\mathrm{PRLS})_{\mathcal{A}^*}\|_2$ and $\frac{2\lambda\sqrt{p}\hat{\eta}^{-\gamma}}{\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})}$ go to 0 in probability. Equation (4.45) implies that the second term in the r.h.s of equation (4.77) goes to zero. Moreover, for the second term in

equation (4.76) we have that, given $M > 0$

$$\begin{aligned}
&\mathbb{P}\left(\frac{2\lambda\sqrt{p}}{\hat{\eta}^{\gamma}\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})} > M\right) \\
\leq\ &\mathbb{P}\left(\frac{2\lambda\sqrt{p}}{\hat{\eta}^{\gamma}\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})} > M, \{\hat{\eta} > \frac{\eta}{2}\}\right) + \mathbb{P}\left(\hat{\eta} \leq \frac{\eta}{2}\right) \\
\leq\ &\frac{2\lambda\sqrt{p}}{M}\sqrt{\mathbb{E}\left(\left(\frac{1}{\tau_{\min}(\mathbf{X}^T\Lambda^2\mathbf{X})}\right)^2\right)\mathbb{E}\left(\frac{1}{\hat{\eta}^{2\gamma}}\mathbf{1}_{\{\hat{\eta}>\frac{\eta}{2}\}}\right)} + M_1 \\
\leq\ &\frac{2\lambda\sqrt{p}}{bnM}\left(\frac{2}{\eta}\right)^{\gamma} + M_1 \longrightarrow 0 \ \text{ as } \ n \to +\infty
\end{aligned} \tag{4.78}$$

since $M_1 \to 0$ as $n \to +\infty$ and $\frac{\lambda}{n}\left(\frac{2}{\eta}\right)^{\gamma} \to 0$ as $n \to +\infty$ thanks to assumption (H3 c). This proves equation (4.58) and concludes the proof. □

Proof of Proposition 9. The proof is analogous to the one of Theorem 8. We start by proving that equation (4.57) holds. It is sufficient to show that the bounds M_1, M_2 and M_3 defined in equations (4.67), (4.68) and (4.75), respectively go to 0 as $n \to \infty$ under assumptions (H5), (H6), (H7) and (H8). We have that

$$M_1 = \frac{2}{\eta}\left(\frac{2\lambda_1\sqrt{p} + \sqrt{pGBn}}{bn}\right) \to 0$$

since $\frac{\lambda_1\sqrt{p}}{\eta n} \to 0$ for the assumption (H7 a) and $\frac{\sqrt{p}}{\sqrt{n}\eta} \to 0$ for the assumption (H6). Moreover,

$$M_2 = \frac{1}{(\frac{\lambda}{n})^{\frac{1}{\gamma}}}\left(\frac{2\lambda_1\sqrt{p} + \sqrt{pGBn}}{bn} + \frac{p}{n^{\frac{1}{\gamma}+\delta}}\right) \to 0$$

since $\left(\frac{\lambda}{n}\right)^{-\frac{1}{\gamma}}\frac{\lambda_1\sqrt{p}}{n} = \left(\frac{\lambda n^{\delta\gamma}}{p^{\gamma}}\right)^{-\frac{1}{\gamma}}\frac{\lambda_1 n^{\delta+\frac{1}{\gamma}-1}}{\sqrt{p}} \to 0$ for assumptions (H8 b) and (H7 b), $\left(\frac{\lambda}{n}\right)^{-\frac{1}{\gamma}}\frac{\sqrt{p}}{\sqrt{n}} = \left(\frac{\lambda n^{\delta\gamma}}{p^{\gamma}}\right)^{-\frac{1}{\gamma}}\frac{n^{-\frac{1}{2}+\delta+\frac{1}{\gamma}}}{\sqrt{p}} \to 0$ for the assumption (H8 b) and for the hypothesis $\frac{1}{\gamma} + \delta < \frac{c+1}{2}$, and $p\left(\frac{\lambda}{n}\right)^{-\frac{1}{\gamma}} n^{-\frac{1}{\gamma}-\delta} \to 0$ for the assumption (H8 b).

Finally,

$$M_3 = \frac{4pL}{n}\left(\sqrt{2n} + \frac{B\sqrt{pG}}{b} + \sqrt{Bnp}\frac{\lambda\left(\frac{2}{\eta}\right)^{\gamma} + \lambda n^{1+\delta\gamma}}{bn} + \lambda\frac{2^{\gamma-1}}{\eta^{\gamma}L}\right) \to 0$$

since the first two terms in the definition of M_3 go to 0 for the assumption (H5), $\frac{\lambda p}{n\eta^{\gamma}}\frac{\sqrt{p}}{\sqrt{n}} \to 0$ for assumptions (H5) and (H8 c), $\frac{p}{n}\frac{\lambda n^{1+\delta\gamma}}{\sqrt{n}}\sqrt{p} = \lambda n^{\delta\gamma-\frac{1}{2}}p\sqrt{p} \to 0$ for the assumption (H8 a) and $\frac{p}{n}\frac{\lambda}{\eta^{\gamma}} \to 0$ for the assumption (H8 c).

Now we have to prove that equation (4.58) holds. It is sufficient to observe that, as $n \to \infty$, the bound in equation (4.45) goes to 0 for the assumption (H5) and the bound in equation (4.78) goes to 0 for the assumption (H8 c). □

Chapter 5

Solar flares prediction as a learning problem

In the present Chapter we present a learning problem in solar physics for the solution of which we make use of sparsity-enhancing methods. This learning problem concerns the prediction of solar flares and the selection of the most predictive features. Solar flares are the most energetic explosive events on the solar surface: they are characterized by an intense electromagnetic emission and are often followed by particle emissions, namely as Coronal Mass Ejections (CMEs) during which the solar material (as electrons and solar plasma) is ejected throughout the solar corona into the interplanetary space. They are one of the primary drivers of space weather and they may cause damage to space-based technological systems, communication links on our planet, radio blackouts etc. The NASA satellite SDO has been launched in February 2010 [108] with the scientific goal of a more complete understanding of the solar magnetic field dynamics related to emissions in Ultraviolet (UV) and Extreme UV (EUV). It comprises three instruments: *Extreme Ultraviolet Variability Experiment* (EVE), *Atmospheric Imaging Assembly* (AIA) and *Helioseismic and Magnetic Imager* (HMI). The instrument HMI [117] is devoted to provide observations of the solar magnetic field activity. Many observational studies and previous machine learning studies confirmed the important role of the magnetic field

for prediction of solar flares. We attempt to forecast solar flares and to analyze the most predictive features of the magnetic field by using sparsity-enhancing methods, i.e. Lasso-type methods as PRiL and APRiL (see Chapter 4). We aim to select and rank the most relevant features by training algorithms on dataset where the labeling is not only the occurrence of solar flares but other tasks, as instance the number of the originated flares and the peak intensity of the strongest originated flare, which are reasonably assumed to be affected by Poisson noise.

5.1 Introduction to the problem

Solar flares are flashes of brightness on the surface of the Sun and they are the most energetic events in the heliosphere [122]. They may extend to over 10^4 km while releasing more than 10^{32} erg in less than 100 seconds, accelerating billions of tons of material to more than 10^6 km/h, emitting electromagnetic radiation at all wavelengths and, in this way, triggering the whole space weather connection. These events, although occurring far from the Earth, could be a threat for our planet, affecting satellite operations, aviation and communication technologies (in Figure 5.1 a summary of the most common effects occur due to space weather, produced due to solar storms, is shown). They are classified according to their peak soft X-ray flux/emission in the $1-8$ Angstrom channel measured by the *Geostationary Operational Environmental Satellites* (GOES). The flare classes are A, B, *C*, M and X with decimal sub-classes. Usually only flares of *C* class and above (we denote it as $C+$ class) are potentially dangerous from a point of view of the space weather effects, e.g. M and X class flares can cause radio blackouts. The full comprehension of solar (and stellar) flare physics is still an open issue, to such an extent that we can talk about a sort of flare paradox: simple computation based on their physical and geometrical properties and on magnetohydrodynamic (MHD) equations would lead to predict a light-up time for flares longer than 10^5 years, while the observed flash phase for these mysterious events is of the order of

some minutes. The numerical modeling of solar flare physics may rely on two different perspectives. On the one hand, finite and boundary element methods applied against MHD partial differential equations allow the simulation of the electromagnetic fields and plasma properties in time and space; on the other hand, artificial intelligence allows pattern identification in the data mess and both source reconstruction with inverse methods and flare prediction with machine learning.

For this reason, the space weather community looks for methods for forecasting solar flares and this was the aim of the Horizon 2020 (H2020) project FLARECAST (Flare Likelihood And Region Eruption foreCASTing). The main purpose was the creation of an advanced technological infrastructure for the solar flare prediction from the data provided by the instrument HMI on the solar satellite SDO. The Vector Magnetic Field data product from HMI gives a quantitative measurement of the free magnetic field energy, magnetic field stress and the helicity. Thanks to the automated Active Region (AR) tracking system, of which HMI is provided, the active regions, which could origin solar flares in correspondence of the sunspots, are identified. However, the active regions do not always origin solar flares and their occurrence is related to the size and complexity of the magnetic patterns characterizing these regions. Therefore, once active regions are localized, the aim is to predict if such active regions will or will not give rise to solar flares (in Figure 5.2 we show an example of HMI magnetogram). Efficient prediction relies on parameters which quantify the eruptive capability of solar active regions. A working package of the FLARECAST project was devoted to the extraction of features from the HMI data characterizing active regions and another working package was focused on developing flare prediction algorithms. Machine learning techniques used for prediction can be exploited also to identify the most predictive features (i.e. to do feature selection). In this part we show a first (preliminary) analysis of the most relevant properties among all the features extracted by the FLARECAST project, using sparsity-enhancing methods.

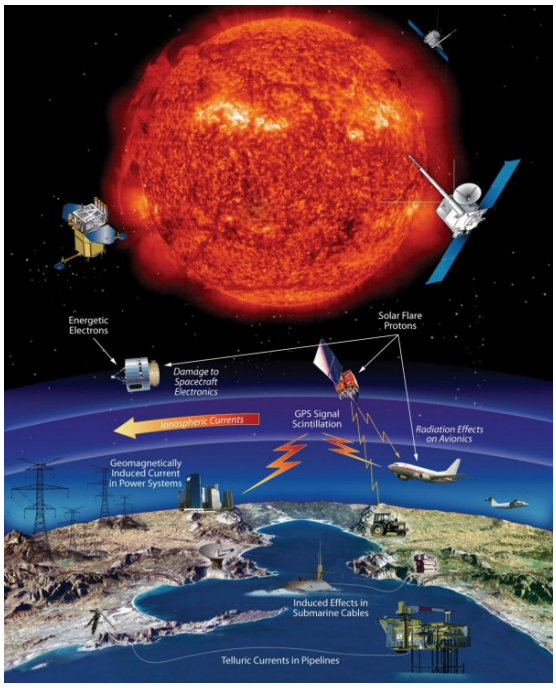


Figure 5.1: A summary of the most common effects due to solar storms.

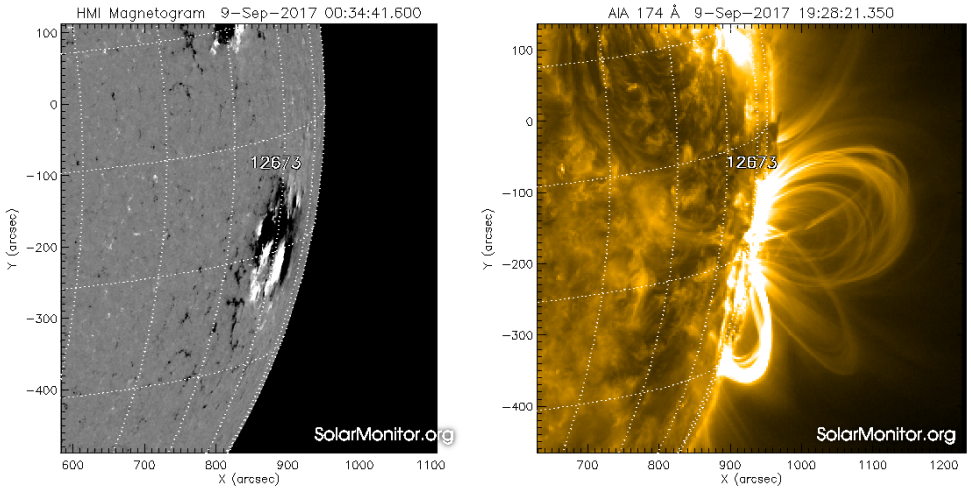


Figure 5.2: An example of active region which will give rise to solar flares. Left panel: HMI magnetogram and identification of an active region (on 9th September 2017 at 00:34:41 UT). At right: AIA image at the bandwidth 171 Å, which shows a solar flare (on 9th September 2017 at 19:28:21 UT) originated by the active region shown in the left panel. These two images are provided by www.SolarMonitor.Org.

5.2 Data description

Data preparation consists of two steps: the feature extraction, which is devoted to compute properties from HMI data in order to create feature vectors (feature vectors constitute the input space, denoted as $\mathcal{X}$ in previous Chapters) and the flare association, which consists mainly in labeling the feature vectors from GOES data (labels constitute the output space, denoted as $\mathcal{Y}$ in previous Chapters). We see in the following paragraphs these two procedures.

5.2.1 Feature extraction

The data we consider are provided by the Helioseismic and Magnetic Imager in the payload of the Solar Dynamics Observatory (SDO/HMI). This telescope provides full disk vector magnetograms with a temporal cadence of 12 minutes, starting from February 2010 [70; 117]. Relying on the Near-Realtime (NRT) Space Weather HMI Archive Patch (SHARP) data product of the HMI database and also using property extraction algorithms developed within FLARECAST, the input data at disposal for the machine learning analysis are feature vectors of dimension up to 171 characterizing properties of the active regions (ARs) present in the SDO/HMI maps. Features extracted by FLARECAST algorithms, which often duplicate the property calculation step on B_{los} (the line-of-sight component of the magnetic field vector) and B_{radial} (the radial component of the magnetic field vector) input data, include the following:

- Schrijver's R value [119]: 1 property yielding a total of 2 features.
- Multifractal structure function spectrum on a 2D image: 2 properties yielding a total of 4 features.
- Falconer's total free magnetic energy proxy WL_{SG} [44]: 1 property yielding a total of 2 features.
- Distance between the leading and following sunspot subgroups and the S_{l-f} [76] separation parameter: 1 property yielding a total of 2 features.

- Spectral power indices extracted by means of the Fourier transform and of a continuous wavelet transform: 2 properties yielding a total of 4 features.
- Magnetic polarity inversion line (MPIL) characteristics: 3 properties yielding a total of 6 features.
- Effective connected magnetic field strength (B_{eff}): 1 property yielding a total of 2 features.
- Vertical decay index of potential field: 4 properties yielding a total of 8 features.
- Non-neutralized electric currents: 1 property yielding 1 feature.
- Ising energy (E): 1 property yielding a total of 4 features.
- Fractal dimension (D): 1 property yielding a total of 2 features.
- Flow field characteristics: 8 properties yielding a total of 16 features.
- Magnetic helicity and energy injection rate: 14 properties yielding a total of 14 features.
- SHARP keywords calculated from their corresponding vector and line-of-sight magnetograms: 16 properties yielding a total of 96 features (including the maximum, total, median, mean, standard deviation, skewness and kurtosis over the SHARP field-of-view).

Eventually, this analysis provides 167 properties extracted from the HMI images. Four further features come from the NOAA/SRS (Solar Region Summary) database: the mean heliographic longitude and latitude of each AR, a binary label encoding the presence of a flare in the past 24 hours and the flare index of events occurring within the past 24 hours. The list of the overall features with a brief description is reported in Tables 5.9, 5.10 and 5.11.

5.2.2 Flare association

Once the features are extracted from each active region we have to associate the information of the occurrence (or not) of solar flares: this procedure is called "flare association". From GOES data we have the information if the active region gave rise to a fixed class energetic solar flare. In this work we consider only GOES class $C1$ and above ($C1+$). In order to do the association between the SHARP ARs and the occurrence of $C1+$ class solar flares another FLARECAST algorithm is then applied. The algorithm first verifies whether the SHARP data contain NOAA-numbered regions (i.e., sunspot groups) by comparison with NOAA's daily SRS file immediately before the SHARP observations. Then, if any NOAA number is assigned to the SHARP data, the algorithm searches for GOES flares occurring in the same source region during the entire disk passage. Once the flare association is realized, each active region is characterized by the 171-dimensional vector of features and a binary label (1 if the active region originated a solar flare 0 otherwise). Therefore, given n active regions we can construct a training set $\{(X_i, Y_i)\}_{i=1}^{n}$ where X_i is the feature vector of the i-th active region and Y_i is the corresponding label. We use this training to train machine learning techniques in order to be able to predict the occurrence of solar flares when the feature vector of a new active region is given. However, from GOES data many different types of information on solar flares are available, such as the number of flares originated from an active region and their intensity. In the case where many flares were originated from an active region we refer to the "maximum flare" as the flare with the most intensity. Therefore, we extract from GOES data the following information.

1. The number of flares originated by an AR.

2. The intensity of the maximum flare. This number is obtained by converting the GOES flare class (e.g. $C3.6$ flare class is converted in the quantity 3.6, M1.8 is converted in the quantity 18, etc. in synthesis the decimal number is multiplied by a power of 10 in according to the corresponding letter (from 10^{-2} for A class to 10^{2} for X class)).

3. The imminence of the maximum flare computed as 1 divided by the peak time of the maximum flare (computed in hours).

We refer to such information on solar flares as "tasks" to predict. The task which is typically used is the occurrence of solar flares (1 if a solar flare is occurred and 0 if not) and we refer to it with the name 'flaring'. Theoretically, a better prediction can be done since a more complete information on solar flares (not only on the occurrence of flares) is available.

5.2.3 Training and test sets

Fixed a task, machine learning methods can be applied in order to predict the occurrence of solar flares. If we denote n the number of active regions and p the number of features associated to each active region, we can construct a training set $\{(X_i, Y_i^{(t)})\}_{i=1}^n$ where X_i is the feature vector and $Y_i^{(t)}$ is the label of the t-th task for the i-th active region. Therefore the training set is given by

$$(\mathbf{X}(\text{training}), \mathbf{Y}^{(t)}(\text{training})), \tag{5.1}$$

where $\mathbf{X}(\text{training})$ is the feature matrix whose rows are the p-dimensional feature vectors X_i^T, and $\mathbf{Y}^{(t)}(\text{training})$ is the label vector which has $Y_i^{(t)}$ as elements $i = 1, \dots n$. Therefore, $\mathbf{X}(\text{training})$ has dimension $n \times p$ and $\mathbf{Y}^{(t)}(\text{training})$ has dimension n. Once, the machine learning method is trained we evaluate the performances in prediction by defining a test set

$$(\mathbf{X}(\text{test}), \mathbf{Y}^{(t)}(\text{test})), \tag{5.2}$$

where $\mathbf{X}(\text{test})$ is a $m \times p$ test matrix which has as rows the feature vectors $X_i^T(\text{test})$ for $i = 1, \dots, m$ and $\mathbf{Y}^{(t)}(\text{test})$ is the m-dimensional test label vector made of the observations $Y_i^{(t)}(\text{test})$ for $i = 1, \dots, m$.

5.3 Solar flare prediction and feature selection

The ingredients of a supervised approach for the prediction of solar flares of a given intensity class are

- a historical data set of feature vectors extracted from SDO/HMI data to create the feature matrix $\mathbf{X}(\text{training})$ for the training set;
- a set of labels, each one associated with an active region and encoding the outcome information to create the label vector $\mathbf{Y}(\text{training})$ for the training set;
- a computational method trained on the historical (training) set and the corresponding set of labels. When a new magnetogram arrives, the pattern recognition method extracts the features from it and the trained machine learning method both predicts the outcome corresponding to the new feature set and assesses the impact of each feature against the prediction effectiveness.

To this purpose, within the FLARECAST project, different machine learning techniques are used as Support Vector Machine (SVM), Random Forest, multi-layer perceptrons, k-nearest neighbors and so on [19; 48; 82; 101]. In particular in [27; 110] a kind of Lasso method, called Hybrid Lasso [12], which combines the Lasso method with an unsupervised fuzzy clustering technique, is used.

In this Chapter we apply some Lasso-type methods: the classical Lasso, the Adaptive Lasso [156] and the two Lasso-type methods introduced in Chapter 4 PRiL and APRiL. As we discussed in Chapter 4, this kind of machine learning methods allows us to simultaneously predict and select the most predictive features. Lasso-type methods are supervised regularization methods for regression, therefore, in order to have a binary prediction (YES or NO solar flares) we apply a simple technique to partition the regression outcome through the optimization of a specific skill score on the training set. Usually the feature selection is analyzed using the 'flaring' task as labeling [19; 27; 48; 82; 110]. In the current Chapter we analyze the feature selection using also different tasks

(the ones listed in section 5.2.2). In particular we are interested in analyzing if (and how) the prediction improves using different tasks and using different Lasso-type methods taking into account the nature of noise on each task. We expect that it is more reasonable to consider the tasks 'number of flares' and 'maximum intensity' affected by Poisson noise rather than Gaussian, whereas both choices, Poisson or Gaussian noise, should be not deemed appropriate for the 'imminence' task. We apply both Lasso (which is usually applied to Gaussian data) and PRiL (which has to be applied to Poisson data) and their adaptive versions (Adaptive Lasso and APRiL) on each of the available tasks and we compare their performances.

5.3.1 Algorithm scheme

Fixed a task t, we assume that the observed data $(X_i, Y_i^{(t)})$ are i.i.d. and satisfy

$$\mathbb{E}(Y_i^{(t)}|X_i) = X_i^T(\beta^*)^{(t)}, \tag{5.3}$$

for all $i = 1, \ldots, n$, where $(\beta^*)^{(t)}$ is a suitable vector of parameters. Under this formulation we can consider the following cases: $Y_i^{(t)}$ is affected by Gaussian noise; by Poisson noise; by an unknown noise, as concerns machine learning setting. We apply a Lasso-type method in order to estimate the sought parameter vector $(\beta^*)^{(t)}$. An estimator $\hat{\beta}^{(t)}$ can be used to both select the relevant features and to predict the outcomes. In order to evaluate the selection of the relevant features we consider the active set: we recall that the active set $\hat{\mathcal{A}}$ is defined as follows

$$\hat{\mathcal{A}} = \{j \in \{1, \ldots, p\} : \hat{\beta}_j^{(t)} \neq 0\}. \tag{5.4}$$

$\hat{\mathcal{A}}$ provides the set of relevant features.

Flare prediction with regression algorithms is typically obtained by accounting for numerical skill scores for the assessment of flare prediction performances [18], and thresholding the regression outcome in such a way that one of the skill score is optimized. We follow a similar procedure in order

to convert the quantitative information of the prediction (e.g. the number of occurred flares, the maximum intensity of the maximum flare, and so on) into a binary information of the occurrence or not of solar flares. To do so we choose a threshold which allows us to cluster the predictions in YES or NO so that the TSS (defined in equation (4.30)) is maximized. This procedure is described in details in Algorithm 2. In this way we can compare the skill scores on the binary prediction 'YES/NO solar flares' using different tasks.

We describe in the following the algorithm scheme which we apply on each task.

- Train the machine learning method on the training set $(\mathbf{X}(\text{training}), \mathbf{Y}^{(t)}(\text{training}))$. The result is an estimator $\hat{\beta}^{(t)}$ of the sought parameter vector $(\beta^*)^{(t)}$.
- Compute the set of relevant features (i.e., the active set $\hat{\mathcal{A}}$).
- Select a threshold $L_{opt}^{(t)}$ in the training phase in order to convert the information of the task to be predicted into a binary outcome (this thresholding procedure is summarized in Algorithm 2).
- Given a new feature vector $X_i(\text{new})$, compute

$$\hat{Y}_i(\text{new}) = X_i^T(\text{new})\hat{\beta}^{(t)}, \tag{5.5}$$

  and convert the information of the prediction $\hat{Y}_i(\text{new})$ in a binary outcome as follows

  - if $\hat{Y}_i(\text{new}) > L_{opt}^{(t)}$ then YES solar flare (i.e. in the next 24 hours at least a $C1+$ class solar flare will occur)
  - if $\hat{Y}_i(\text{new}) \leq L_{opt}^{(t)}$ then NO solar flare (i.e. in the next 24 hours no $C1+$ class solar flares will occur).

In order to assess the effectiveness of the prediction we compute some skill scores on the test set. Computationally we consider the following scheme.

- Compute
$$\hat{Y}_i(\text{test}) = X_i^T(\text{test})\hat{\beta}^{(t)}, \quad \text{for } i = 1, \ldots, m. \tag{5.6}$$

- Compute the classes
$$\mathcal{C}_1^{(t)} = \{i \in \{1, \ldots, m\} : Y_i^{(t)}(\text{test}) > 0\} \tag{5.7}$$
$$\mathcal{C}_2^{(t)} = \{i \in \{1, \ldots, m\} : Y_i^{(t)}(\text{test}) = 0\}, \tag{5.8}$$
where $Y_i^{(t)}(\text{test})$ are the true labels of the test set and compute
$$\hat{\mathcal{C}}_1^{(t)} = \{i \in \{1, \ldots, m\} : \hat{Y}_i^{(t)}(\text{test}) > L_{opt}^{(t)}\} \tag{5.9}$$
$$\hat{\mathcal{C}}_2^{(t)} = \{i \in \{1, \ldots, m\} : \hat{Y}_i^{(t)}(\text{test}) \leq L_{opt}^{(t)}\}. \tag{5.10}$$

- Compute the confusion matrix, i.e. evaluate the TP, TN, FP and FN as follows
$$\text{TP} = \#(\mathcal{C}_1^{(t)} \cap \hat{\mathcal{C}}_1^{(t)}) \tag{5.11}$$
$$\text{TN} = \#(\mathcal{C}_2^{(t)} \cap \hat{\mathcal{C}}_2^{(t)}) \tag{5.12}$$
$$\text{FP} = \#(\mathcal{C}_2^{(t)} \cap \hat{\mathcal{C}}_1^{(t)}) \tag{5.13}$$
$$\text{FN} = \#(\mathcal{C}_1^{(t)} \cap \hat{\mathcal{C}}_2^{(t)}), \tag{5.14}$$
where $\#(\cdot)$ indicates the cardinality of the set in the argument.

- From these four quantities compute some skill scores in order to evaluate the performance in prediction.

The skill scores as TSS, Heidke Skill Score (HSS), accuracy (ACC), Probability of Detection (POD) are metrics computed from the confusion matrix, or contingency table, which is represented in Table 5.1. In this context TP are the ARs which originated solar flares correctly predicted as YES flares, TN are the ARs which did not originate solar flares correctly predicted as NO flares, FP are the ARs which did not originate solar flares incorrectly predicted as YES flares and FN are the ARs which originated solar flares incorrectly predicted as NO

Table 5.1: Confusion matrix definition.

		predicted	
		YES	NO
observed	YES	TP	FN
	NO	FP	TN

flares. The TSS and HSS are the most popular metrics used in the context of Space Weather, as forecasting solar flares [48]. ACC is the most popular classification metric, but it is less meaningful in rare events: it can be very high even if the prediction of the positive events is not so accurate since the number of negative events correctly estimated is very high. The TSS is defined in equation (4.30), it covers the range between -1 and 1 and it is optimal when it is equal to 1. A negative value means that forecasting behaves in a wrong way i.e. it mixes the role of the positive events with the one of the negative events. The HSS is defined as follows

$$\mathrm{HSS} = \frac{2(\mathrm{TP}\cdot\mathrm{TN} - \mathrm{FN}\cdot\mathrm{FP})}{((\mathrm{TP}+\mathrm{FN})\cdot(\mathrm{FN}+\mathrm{TN}) + (\mathrm{TP}+\mathrm{FP})\cdot(\mathrm{FP}+\mathrm{TN}))}, \tag{5.15}$$

it measures the improvement of the forecast over the random forecast. HSS values are in the range between $-\infty$ and 1. The optimal value is equal to 1, a negative value means that the forecast is worse than the random forecast and the 0 value means that the forecast has the same skill of the random forecast. The ACC, defined as follows

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{TN}+\mathrm{FP}}, \tag{5.16}$$

is the ratio between the number of correct predictions over the total number of predictions. It ranges between 0 and 1 and the optimal value is achieved in 1. Finally, POD, defined as follows

$$\mathrm{POD} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{5.17}$$

measures the ability to find the positive examples. It is the first addend in the definition of TSS (see equation (4.30)).

Algorithm 2 Thresholding procedure to classify predictions.

1: Input: $\mathbf{X}$(training), $\mathbf{Y}^{(t)}$(training), $\hat{\beta}^{(t)}$ (where $\hat{\beta}^{(t)}$ denotes an estimator computed in according to the chosen method).

2: Cluster the values of the label vector $\mathbf{Y}^{(t)}(\text{training}) = (Y_1^{(t)}, \dots, Y_n^{(t)})^T$ in two classes, i.e.

$$\mathcal{C}_1^{(t)} = \{i \in \{1, \dots, n\} : Y_i^{(t)} > 0\} \tag{5.18}$$

$$\mathcal{C}_2^{(t)} = \{i \in \{1, \dots, n\} : Y_i^{(t)} = 0\}. \tag{5.19}$$

$\mathcal{C}_1^{(t)}$ represents the set of the active regions which produce at least one flare (YES flare) as a positive label correlates with the flare occurrence. $\mathcal{C}_2^{(t)}$ represents the set of the active regions which do not produce any flares (NO flare).

3: Compute

$$\hat{\mathbf{Y}}^{(t)}(\text{training}) = \mathbf{X}(\text{training})\hat{\beta}^{(t)}. \tag{5.20}$$

4: Cluster the values of the t-th predicted task $\hat{\mathbf{Y}}^{(t)}(\text{training}) = (\hat{Y}_1^{(t)}(\text{training}), \dots, \hat{Y}_n^{(t)}(\text{training}))^T$ in two classes according to YES or NO flares by choosing a threshold $L_{\text{opt}}^{(t)}$ which optimizes the TSS on the training set, as follows: given a set of values $\{\ell_q^{(t)}\}_q^Q$,

5: **for** $L^{(t)} \in \{\ell_q^{(t)}\}_q^Q$ **do**

6:

$$\hat{\mathcal{C}}_1^{(t),L^{(t)}} = \{i \in \{1, \dots, n\} : \hat{Y}_i^{(t)}(\text{training}) > L^{(t)}\} \tag{5.21}$$

$$\hat{\mathcal{C}}_2^{(t),L^{(t)}} = \{i \in \{1, \dots, n\} : \hat{Y}_i^{(t)}(\text{training}) \leq L^{(t)}\}. \tag{5.22}$$

7: Compute the TSS between the predicted classes $\hat{\mathcal{C}}_1^{(t),L^{(t)}}$, $\hat{\mathcal{C}}_2^{(t),L^{(t)}}$ and the true classes $\mathcal{C}_1^{(t)}$, $\mathcal{C}_2^{(t)}$, we denote such a value as $\text{TSS}^{(t),L^{(t)}}$.

8: **end for**

9: Choose the threshold such that

$$L_{opt}^{(t)} = \arg \max_{L^{(t)} \in \{\ell_q^{(t)}\}_{q=1}^Q} \text{TSS}^{(t),L^{(t)}}. \tag{5.23}$$

10: Return $L_{opt}^{(t)}$ in order to classify the outcomes when a new feature vector is given.

5.4 Experiments

In our numerical results we consider point-in-time SDO/HMI images in the time range between 09/14/2012 and 04/30/2016, with a time cadence of 24 hours corresponding to a specific forecast issuing time, which is 00:00 expressed Universal Time (UT). After the application of the pattern recognition step (or feature extraction) we had at disposal 4061 point-in-time 171-dimension feature vectors. In the following sections we report some results about the feature selection using different tasks by applying Lasso, Adaptive Lasso, PRiL and APRiL as machine learning methods.

5.4.1 Data

Within the FLARECAST project, the usual way to rank the importance of features consists in applying some machine learning methods, suitable for feature selection, on data where the label is the 'flaring' information. In this Chapter we analyze the feature selection when the label is not the 'flaring' but it is one of the tasks listed in section 5.2. We train the machine learning methods following the procedure used in [27]. The training set is built by randomly extracting around 2/3 active regions from the set of all ARs and labeling the 171-dimension feature vectors associated to each AR with 1 if a GOES $C1+$ flare occurred in the next 24 hours 0 otherwise. The set of feature vectors associated to the remaining 1/3 ARs was provided as test set for experiments to supervised learning algorithms trained on the training set. Training and test sets are built such that they do not overlap in any way, neither in time nor in terms of ARs examined. Finally the random complete separation of ARs into training and test sets was replicated 100 times to enable statistical robustness of the results. We follow a similar procedure for each task, with the difference that the feature vectors are labeled with the information contained in the chosen task (e.g. if we use the task 'number of flares' the feature vector is labeled by annotating the number of $C1+$ solar flares occurred in the next 24 hours and 0 means no occurrence of $C1+$ solar flares). In our experiments

we consider separately the four tasks: 'flaring', 'number of flares', 'maximum intensity' and 'imminence' (see section 5.2).

5.4.2 Results

We follow the algorithm scheme described in section 5.3 and we use Lasso, Adaptive Lasso (AdaLasso) (see equations (4.6) and (4.7), respectively), PRiL and APRiL methods to compute the estimator $\hat{\beta}^{(t)}$. For the adaptive strategy AdaLasso we define weights as in [156], i.e. weights are defined as in equation (4.28) where, in this case, the Maximum Likelihood estimate coincides with the least square estimate. We apply each of these four methods on each dataset fixing one task t of the four above listed tasks. We first focus on the feature selection. We evaluate the importance of a feature according to its presence in the 100 active sets computed for each method. In Table 5.2 we report, for each method and for each task, the number of features which belong to at least 1 active set (occurrence ≥ 1) and the number of features which belong to at least 10 active sets (occurrence > 10). Some comments on results in Table 5.2.

- The number of features selected by the Lasso method is always higher than the one provided by the other methods.
- The number of inactive features (i.e. the features never selected by any method in any active sets) is higher using the task 'maximum intensity' than the one using the other tasks (this is clearly visible in results provided by PRiL method, since only 20 features occur in at least one active set).
- For any method, few features, with respect to the total number, occur in more than 10 active sets (for Lasso and AdaLasso methods the maximum number is 58 and 26, respectively, obtained using the task 'number of flares' and for PRiL and APRiL methods the maximum number is 35 and 14, respectively obtained using the task 'imminence' against the total number of features equal to 171).

- As we expected the adaptive methods return solutions more sparse than the ones computed by the corresponding non-adaptive methods.

Furthermore, we remark that the set of relevant features selected by APRiL with occurrence > 10 is contained in the set of relevant features selected by PRiL with occurrence > 10, whereas the set of relevant features selected by AdaLasso is usually different with respect to the one returned by Lasso: for the task 'flaring' only 5 features (over 18) are in common between Lasso and AdaLasso, for the task 'number of flares' only 6 features (over 26), for the task 'maximum intensity' only 8 features (over 23) and for the task 'imminence' only 3 features (over 23).

In order to rank features we order them in according to their occurrence in the active sets and we report histograms of the top-10 features according to this principle. In Figures 5.3, 5.4, 5.5 and 5.6 we report the top-10 rankings provided by each method for each task: in detail, in Figure 5.3 we report the histograms counting the number of times each feature is selected in the 100 active sets using the task 'flaring', in Figure 5.4 using the task 'number of flares', in Figure 5.5 the task 'maximum intensity' and in Figure 5.6 the task 'imminence'. In the following we provide some comments about the top-10 rankings by comparing them with the ones obtained in [27]. In [27] the top-10 rankings of features are provided by following a different principle based on a Recursive Feature Elimination and differentiate the top-10 ranking obtained by forecasting $C1+$ class flares and $M1+$ class flares whereas in our analysis we consider only $C1+$ class flares. Furthermore, in [27] two machine learning methods are compared: the Hybrid Lasso (which exploits an unsupervised fuzzy clustering technique to classify the regression outcome provided by the classical Lasso method) and the Random Forest, which provides good results in flare prediction as shown in [48]. However, we notice that some of the top-10 features are the same found in [27].

We report some comments about the selected features.

- The features *wlsg_blos/value_int* and *sharp_kw/snetjzpp/total* belong to all 4 top-10 rankings of PRiL and APRiL and also to 3 and 4 top-10 rankings

Table 5.2: Number of features with occurrence in at least 1 active set (occurrence ≥ 1) and in more than 10 active sets (occurrence > 10). For each method, 100 active sets are computed.

task	**flaring**		**number of flares**		**maximum intensity**		**imminence**	
	Occurrence of features in 100 **active sets**							
	≥ 1	> 10	≥ 1	> 10	≥ 1	> 10	≥ 1	> 10
Lasso	99	41	135	58	83	27	95	34
AdaLasso	34	18	30	26	27	23	30	23
PRiL	72	22	62	20	20	13	93	35
APRiL	18	9	17	10	13	7	38	14

Table 5.3: Number of times each feature is selected in the top-10 rankings of each method (the maximum possible number of times is equal to 4, which is the number of tasks considered in the analysis).

	Lasso	**AdaLasso**	**PRiL**	**APRiL**
feature	Number of times in the top-10 rankings			
wlsg_blos/value_int	3	-	4	4
sharp_kw/snetjzpp/total	4	-	4	4
sharp_kw/twistp/kurtosis	4	-	4	4
wlsg_br/value_int	1	-	3	3
flare_past	2	-	2	1
flare_index_past	-	-	2	3
sharp_kw/sflux/max	2	4	2	2

Table 5.4: Presence of each feature (yes or no) in the top-10 ranking provided with the task 'flaring' for forecasting C1+ class flares, considering Lasso, AdaLasso, PRiL, APRiL, HL and RF (results of HL and RF are provided in [27]).

	Lasso	**AdaLasso**	**PRiL**	**APRiL**	**HL**	**RF**
feature	Presence (yes or no) in the top-10 ranking using the task 'flaring'					
wlsg_blos/value_int	yes	no	yes	yes	yes	no
sharp_kw/snetjzpp/total	yes	no	yes	yes	no	yes
sharp_kw/twistp/kurtosis	yes	no	yes	yes	no	no
wlsg_br/value_int	no	no	yes	yes	yes	no
flare_past	yes	no	yes	no	yes	no
flare_index_past	no	no	no	yes	yes	yes
sharp_kw/sflux/max	yes	yes	yes	yes	no	no

of Lasso, respectively. The features *wlsg_blos/value_int* has almost always occurrence equal to 100, which means that it is present in all 100 active sets. Such a feature is in the top-10 ranking provided by Hybrid Lasso for forecasting $C1+$ class flares and in the one provided by Random Forest for forecasting M1+ class flares [27]. The feature *sharp_kw/snetjzpp/total* is in the top-10 ranking provided by Random Forest for forecasting $C1+$ class flares and M1+ class flares and in the top-10 ranking of Hybrid Lasso method for forecasting M1+ class flare [27].

- The feature *sharp_kw/twistp/kurtosis* belongs to all 4 top-10 rankings of Lasso, PRiL and APRiL, but it does not belong to any top-10 rankings of Hybrid Lasso and Random Forest.

- The feature *wlsg_br/value_int* belongs to 3 of the top-10 rankings of PRiL and APRiL and 1 of the top-10 rankings of Lasso. Such a feature is in the top-10 ranking provided by Hybrid Lasso for forecasting $C1+$ class flares [27].

- The feature *flare_index_past*, which is present in both top-10 rankings of Hybrid Lasso and Random Forest for forecasting $C1+$ class flares [27], belongs to 2 of the top-10 rankings of PRiL (for tasks 'number of flares' and 'maximum intensity') and 3 of the top-10 rankings of APRiL (for tasks 'flaring', 'maximum intensity' and 'imminence'). However, in 2 top-10 rankings of both PRiL and Lasso (for tasks 'flaring' and 'imminence') *flare_past* has very high occurrence (occurrence equal to 100 for 3 of these rankings). These two features are quite correlated in meaning: *flare_index_past* is a binary flag for the occurrence of at least one flare in the previous 24 hours and *flare_past* measures the flare peak magnitudes of the previous 24 hours. The fact that the magnitude of the flares in the past 24 hours is a relevant variable for the prediction of the imminence is a coherent result.

- AdaLasso generally selects different features than the ones selected by the other methods. The feature which is present in all 4 top-10 rankings of

AdaLasso is *sharp_kw/sflux/max*, which belongs to 2 of the top-10 rankings of Lasso, PRiL and APRiL, but it is not in any top-10 rankings provided by Hybrid Lasso and Random Forest.

- Most of the features presented in the top-10 rankings are of the family *sharp_kw*. They are mainly specific properties of the magnetic field (see Tables 5.10 and 5.11).

We summarize some of these comments in Tables 5.3 and 5.4. In Table 5.3 we report the number of times each feature discussed above is selected in the top-10 rankings by Lasso, AdaLasso, PRiL and APRiL: for each method there are 4 top-10 rankings, therefore the maximum number of times is equal to 4. In Table 5.4 we report if each feature is present (yes) or not (no) in the top-10 ranking provided with the task 'flaring' using Lasso, AdaLasso, PRiL, APRiL, Hybrid Lasso (HL) and Random Forest (RF) (we refer to the top-10 rankings of HL and RF provided in [27] for forecasting $C1+$ class flares).

Now we focus on the prediction performances. In Tables 5.5, 5.6, 5.7 and 5.8 we report the TSS, HSS, ACC and POD values obtained by averaging over the 100 replicates the results provided by each method for each task on both the training and test set. We notice that Adaptive Lasso provides bad results with respect to the other methods. PRiL usually provides higher values in TSS, HSS and ACC than the other methods except for the TSS and HSS values in the test phase for the task 'imminence', which are a little smaller than the ones provided by Lasso. In detail, the TSS and HSS values provided in training by PRiL and APRiL are much higher than the ones provided by Lasso and AdaLasso (the highest TSS mean value (approximatively equal to 0.82) and the highest HSS mean value (approximatively equal to 0.75) are provided by PRiL). The TSS values in the test set provided by PRiL, APRiL and Lasso are approximatively equal or better with respect to the ones shown in [27] (using the task 'flaring' Lasso and PRiL achieve approximatively the value equal to 0.57 against the value 0.53 provided by RF (see [27])). The HSS values in the test set provided by PRiL and APRiL are approximatively close to the ones shown in [27]: using the task 'flaring' Lasso, PRiL and APRiL achieve

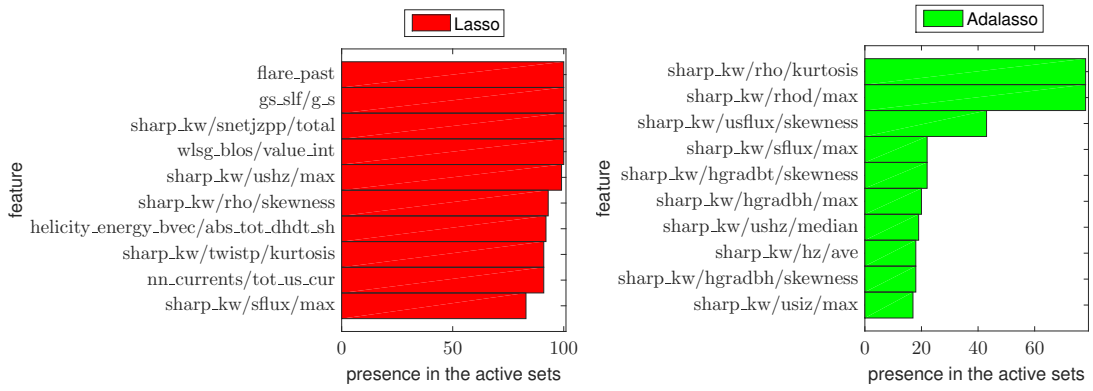
Table 5.5: TSS values obtained by averaging over 100 replicates the results provided by Lasso, Adaptive Lasso, PRiL and APRiL methods for each task.

task		TSS			
		Lasso	**AdaLasso**	**PRiL**	**APRiL**
flaring	Train	$0.59_{(\pm 0.01)}$	$0.43_{(\pm 0.04)}$	$0.72_{(\pm 0.01)}$	$0.71_{(\pm 0.01)}$
	Test	$0.57_{(\pm 0.03)}$	$0.43_{(\pm 0.06)}$	$0.57_{(\pm 0.03)}$	$0.55_{(\pm 0.03)}$
number of flares	Train	$0.56_{(\pm 0.01)}$	$0.51_{(\pm 0.02)}$	$0.74_{(\pm 0.01)}$	$0.73_{(\pm 0.01)}$
	Test	$0.55_{(\pm 0.03)}$	$0.5_{(\pm 0.03)}$	$0.56_{(\pm 0.03)}$	$0.56_{(\pm 0.03)}$
maximum intensity	Train	$0.53_{(\pm 0.02)}$	$0.5_{(\pm 0.03)}$	$0.82_{(\pm 0.009)}$	$0.79_{(\pm 0.01)}$
	Test	$0.53_{(\pm 0.02)}$	$0.5_{(\pm 0.03)}$	$0.55_{(\pm 0.03)}$	$0.54_{(\pm 0.03)}$
imminence	Train	$0.53_{(\pm 0.02)}$	$0.5_{(\pm 0.02)}$	$0.59_{(\pm 0.01)}$	$0.59_{(\pm 0.02)}$
	Test	$0.58_{(\pm 0.02)}$	$0.5_{(\pm 0.02)}$	$0.57_{(\pm 0.03)}$	$0.57_{(\pm 0.03)}$

approximatively the value equal to 0.51 against the value 0.52 provided by RF (see [27]). However, we notice that using other tasks the HSS in the test set achieve higher values (using the task 'maximum intensity' PRiL provides an HSS approximatively equal to 0.54). We notice that the POD values provided by all methods are very high (see Table 5.8): this leads to worse results in the False Alarm Ratio (FAR), since the TSS is obtained by balancing POD and FAR.

In Figure 5.7 we report the distributions of TSS and HSS over the 100 replicates: the TSS distributions on training and test sets are provided in the top row and the HSS distributions are provided in the bottom row. We notice that the TSS and HSS provided by PRiL and APRiL in the training phase are much better with respect to the other methods, in particular for the task 'maximum intensity'. HSS in the test phase is usually better for PRiL and APRiL. We notice that only for the task 'imminence' Lasso has usually a better performance than PRiL and APRiL. This result is coherent with the nature of noise on the different tasks: we can assume that the tasks 'number of flares' and 'maximum intensity' follow a Poisson statistic whereas it is more reliable that the task 'imminence' is affected by Gaussian noise instead of Poisson noise.

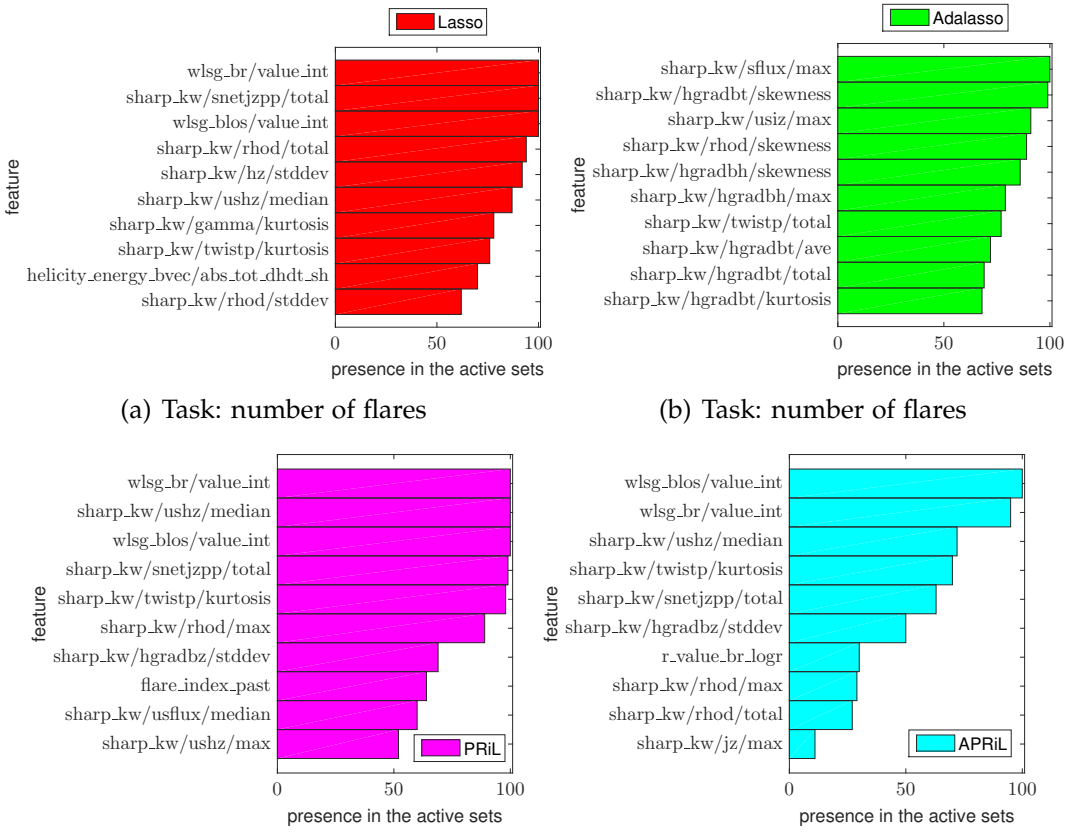
We report some comments about the obtained results and possible improvements.



(a) Task: flaring (b) Task: flaring

(c) Task: flaring (d) Task: flaring

Figure 5.3: Top-10 rankings using the task named 'flaring': the histograms count the number of times each feature is selected in the 100 active sets by Lasso and AdaLasso methods (in the top row) and by PRiL and APRiL methods (in the bottom row).



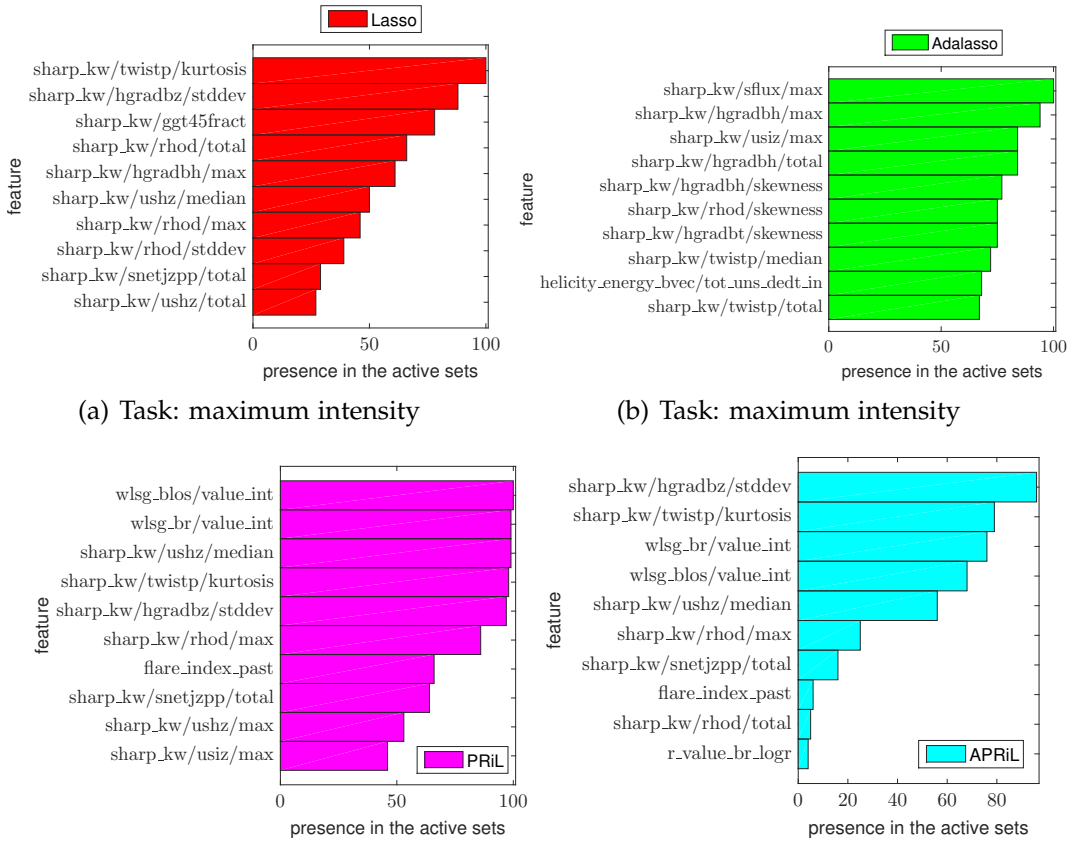
(a) Task: number of flares

(b) Task: number of flares

(c) Task: number of flares

(d) Task: number of flares

Figure 5.4: Top-10 rankings using the task named 'number of flares': the histograms count the number of times each feature is selected in the 100 active sets by Lasso and AdaLasso methods (in the top row) and by PRiL and APRiL methods (in the bottom row).



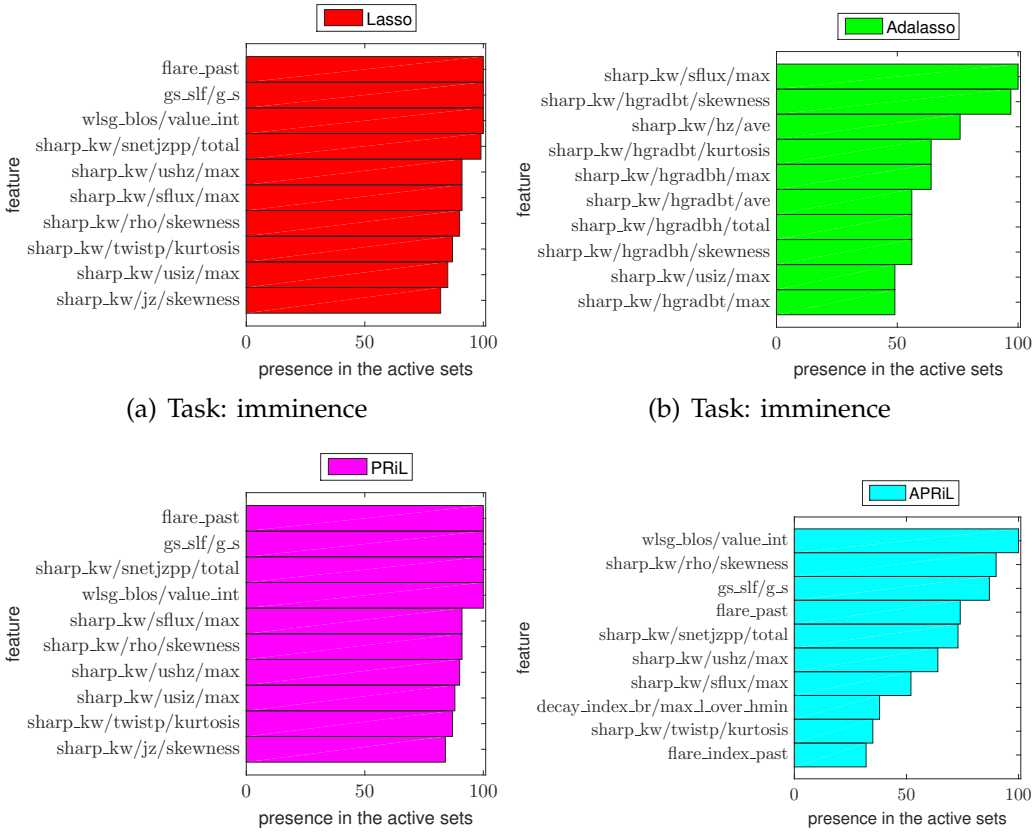
(a) Task: maximum intensity

(b) Task: maximum intensity

(c) Task: maximum intensity

(d) Task: maximum intensity

Figure 5.5: Top-10 rankings using the task named 'maximum intensity': the histograms count the number of times each feature is selected in the 100 active sets by Lasso and AdaLasso methods (in the top row) and by PRiL and APRiL methods (in the bottom row).



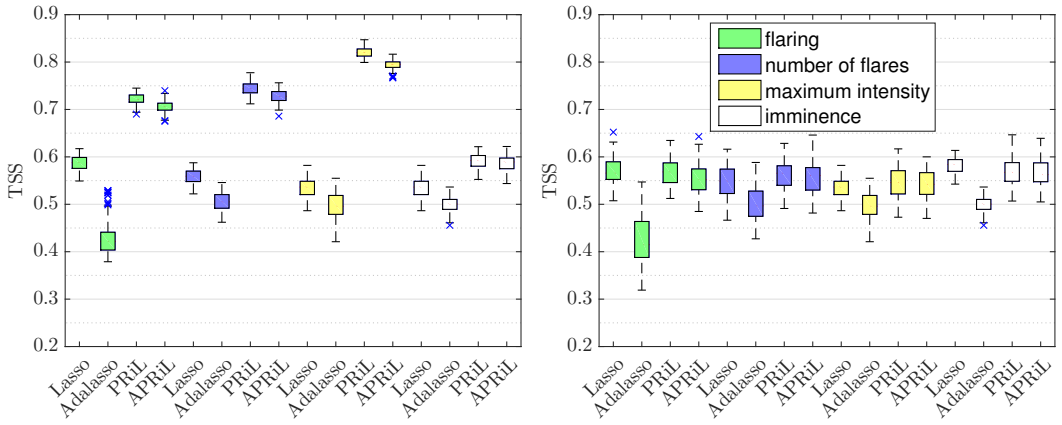
(a) Task: imminence

(b) Task: imminence

(c) Task: imminence

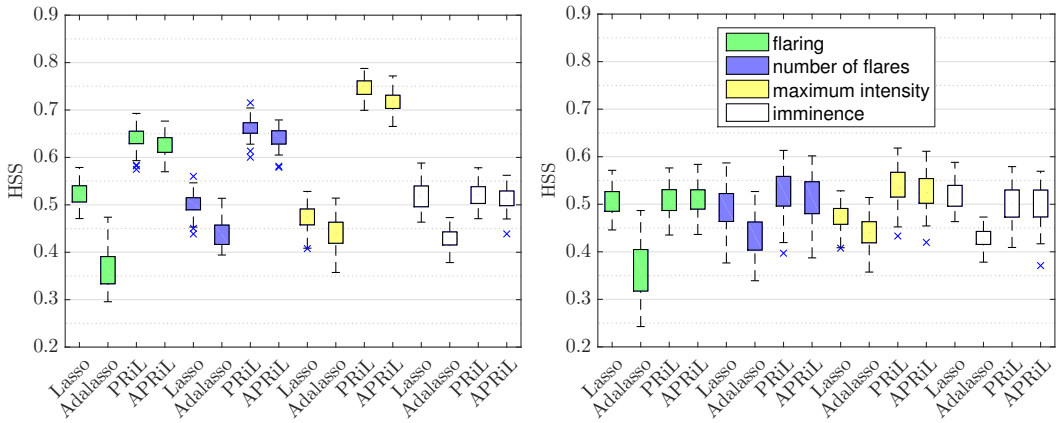
(d) Task: imminence

Figure 5.6: Top-10 rankings using the task named 'imminence': the histograms count the number of times each feature is selected in the 100 active sets by Lasso and AdaLasso methods (in the top row) and by PRiL and APRiL methods (in the bottom row).



(a) TSS on training set

(b) TSS on test set



(c) HSS on training set

(d) HSS on test set

Figure 5.7: Distributions of TSS (top row) and HSS (bottom row) over 100 replicates using the method indicated in the x-axis and using the tasks indicated in the legend ('flaring' green boxplots, 'number of flares' blue boxplots, 'maximum intensity' yellow boxplots and 'imminence' white boxplots). Left column: TSS and HSS distributions computed on 100 replicates of the training set. Right column: TSS and HSS distributions computed on 100 replicates of the test set.

Table 5.6: HSS values obtained by averaging over 100 replicates the results provided by Lasso, Adaptive Lasso, PRiL and APRiL methods for each task.

task		HSS			
		Lasso	**AdaLasso**	**PRiL**	**APRiL**
flaring	Train	$0.52_{(\pm 0.02)}$	$0.37_{(\pm 0.04)}$	$0.64_{(\pm 0.02)}$	$0.62_{(\pm 0.02)}$
	Test	$0.51_{(\pm 0.03)}$	$0.36_{(\pm 0.06)}$	$0.51_{(\pm 0.03)}$	$0.51_{(\pm 0.03)}$
number of flares	Train	$0.5_{(\pm 0.02)}$	$0.44_{(\pm 0.03)}$	$0.66_{(\pm 0.02)}$	$0.64_{(\pm 0.02)}$
	Test	$0.49_{(\pm 0.04)}$	$0.43_{(\pm 0.04)}$	$0.53_{(\pm 0.04)}$	$0.51_{(\pm 0.04)}$
maximum intensity	Train	$0.47_{(\pm 0.03)}$	$0.44_{(\pm 0.03)}$	$0.75_{(\pm 0.02)}$	$0.72_{(\pm 0.02)}$
	Test	$0.47_{(\pm 0.03)}$	$0.44_{(\pm 0.03)}$	$0.54_{(\pm 0.04)}$	$0.53_{(\pm 0.04)}$
imminence	Train	$0.52_{(\pm 0.03)}$	$0.43_{(\pm 0.02)}$	$0.52_{(\pm 0.03)}$	$0.51_{(\pm 0.02)}$
	Test	$0.52_{(\pm 0.03)}$	$0.43_{(\pm 0.02)}$	$0.5_{(\pm 0.04)}$	$0.5_{(\pm 0.04)}$

Table 5.7: ACC values obtained by averaging over 100 replicates the results provided by Lasso, Adaptive Lasso, PRiL and APRiL methods for each task.

task		ACC			
		Lasso	**AdaLasso**	**PRiL**	**APRiL**
flaring	Train	$0.8_{(\pm 0.02)}$	$0.71_{(\pm 0.03)}$	$0.85_{(\pm 0.01)}$	$0.84_{(\pm 0.01)}$
	Test	$0.79_{(\pm 0.02)}$	$0.71_{(\pm 0.04)}$	$0.79_{(\pm 0.02)}$	$0.8_{(\pm 0.02)}$
number of flares	Train	$0.79_{(\pm 0.01)}$	$0.76_{(\pm 0.02)}$	$0.86_{(\pm 0.01)}$	$0.85_{(\pm 0.01)}$
	Test	$0.78_{(\pm 0.02)}$	$0.75_{(\pm 0.02)}$	$0.81_{(\pm 0.02)}$	$0.8_{(\pm 0.02)}$
maximum intensity	Train	$0.77_{(\pm 0.02)}$	$0.76_{(\pm 0.02)}$	$0.89_{(\pm 0.009)}$	$0.88_{(\pm 0.01)}$
	Test	$0.77_{(\pm 0.02)}$	$0.76_{(\pm 0.02)}$	$0.82_{(\pm 0.02)}$	$0.82_{(\pm 0.02)}$
imminence	Train	$0.79_{(\pm 0.02)}$	$0.75_{(\pm 0.01)}$	$0.79_{(\pm 0.02)}$	$0.79_{(\pm 0.01)}$
	Test	$0.79_{(\pm 0.02)}$	$0.75_{(\pm 0.01)}$	$0.79_{(\pm 0.02)}$	$0.78_{(\pm 0.02)}$

- As we expected, taking into account the nature of noise of the tasks gives better results (PRiL works better on tasks 'number of flares' and 'maximum intensity' than Lasso).

- The fact that the thresholding process to classify predictions (see Algorithm 2) is based on optimizing a specific skill score does not seem to affect so much the other scores: although the optimization is based on the TSS value, we retrieve good values in HSS (especially for PRiL method) and in accuracy.

Table 5.8: POD values obtained by averaging over 100 replicates the results provided by Lasso, Adaptive Lasso, PRiL and APRiL methods for each task.

task		POD			
		Lasso	**AdaLasso**	**PRiL**	**APRiL**
flaring	Train	$0.79_{(\pm 0.03)}$	$0.72_{(\pm 0.04)}$	$0.89_{(\pm 0.02)}$	$0.88_{(\pm 0.02)}$
	Test	$0.78_{(\pm 0.05)}$	$0.72_{(\pm 0.06)}$	$0.77_{(\pm 0.04)}$	$0.73_{(\pm 0.05)}$
number of flares	Train	$0.76_{(\pm 0.03)}$	$0.75_{(\pm 0.03)}$	$0.9_{(\pm 0.02)}$	$0.9_{(\pm 0.02)}$
	Test	$0.75_{(\pm 0.03)}$	$0.75_{(\pm 0.04)}$	$0.73_{(\pm 0.03)}$	$0.73_{(\pm 0.03)}$
maximum intensity	Train	$0.75_{(\pm 0.03)}$	$0.72_{(\pm 0.04)}$	$0.94_{(\pm 0.02)}$	$0.93_{(\pm 0.02)}$
	Test	$0.75_{(\pm 0.03)}$	$0.72_{(\pm 0.04)}$	$0.67_{(\pm 0.04)}$	$0.68_{(\pm 0.04)}$
imminence	Train	$0.79_{(\pm 0.04)}$	$0.76_{(\pm 0.03)}$	$0.8_{(\pm 0.04)}$	$0.8_{(\pm 0.03)}$
	Test	$0.79_{(\pm 0.04)}$	$0.76_{(\pm 0.03)}$	$0.78_{(\pm 0.04)}$	$0.79_{(\pm 0.04)}$

- The regularization parameter for all Lasso-type methods is chosen by the 3-fold cross validation. From results in Chapter 4 we have already noticed that Lasso and Adaptive Lasso do not provide optimal results with this regularization parameter choice procedure. This could be a reason which explains the low performances in TSS and HSS for the Adaptive Lasso in the current analysis. Indeed, a better regularization parameter choice strategy could be based on the optimization of a skill score as the TSS value.

- We noticed that some features seem to be robust with respect to the use of different tasks. In this analysis we applied methods on each task taking them separately. However, we can hold more tasks together and apply Multi-task learning methods. A preliminary analysis is in progress: we applied the Multi-task Lasso method [103] on more tasks simultaneously and we made the analysis fixing a feature matrix and varying the number of tasks in the label matrix. This procedure confirmed that the features *wlsg_ blos/value_int* and *sharp_kw/snetjzpp/total* occur in all active sets obtained by applying the Multi-task Lasso method using all possible combinations of tasks.

5.5 Discussion

The mechanism of solar flares is an open issue in solar physics and it has remained unsolved for more than one century. They are triggered at the Sun's surface and they propagate from the solar atmosphere toward Earth. Although they are far from our planet the solar flare radiation may be damaging to infrastructures, instruments and astronauts in space, therefore flare forecasting is an integral part of contemporary space-weather forecasting. Solar flares originate from magnetically active regions but not all solar active regions give rise to solar flares. Therefore, the challenge of solar flare prediction is nowadays based on an intelligent computational analysis of physics-based properties extracted from active region observables, most commonly line-of-sight or vector magnetograms of the active-region photosphere. To deal with the recent large amount of solar observation data new approaches have been developed using machine learning algorithms. In this Chapter we use Lasso-type methods (as PRiL and APRiL) to flare forecasting and identify which features are most effective for predicting flares. The analysis has been done using different labeling (not only the binary information of 'flaring' but other tasks as the number of the originated flares, the maximum flare class intensity between the originated ones and the imminence of the most intensive flare between the originated ones). Furthermore, the nature of noise affecting labels has been taken into account improving the results in terms of skill scores (PRiL and APRiL methods, applied on tasks reasonably affected by Poisson noise, gave better results). However, in our analysis we focused on a linear model. In literature, different strategies have been applied also using non linear kernels (as in [48]) or deep neural networks [100] providing comparable results.

Possible improvements concern the creation of different feature spaces: instead of considering the linear model as in equation (5.3), where we estimate the "best" weights (i.e. the estimator $\hat{\beta}$) in such a way the linear combination of features returns a good prediction, we can consider non linear combinations of features with simple operations, as division or multiplication of powers of features, driven by their physical meaning and physics models, to return a

better prediction.

Table 5.9: Feature list with short descriptions.

Feature	Description
alpha_exp_fft_blos/alpha, alpha_exp_fft_br/alpha	Fourier power spectral index
alpha_exp_cwt_blos/alpha, alpha_exp_cwt_br/alpha	Continuous wavelet transform power spectral index
beff_blos/beff, beff_br/beff	Effective connected magnetic field strength B_{eff}
decay_index_blos/max_l_over_hmin	Max. ratio of MPIL length to min. height of critical decay index
decay_index_br/max_l_over_hmin	$l/h(n_{\mathrm{cr}})_{\mathrm{min}}$
decay_index_blos/tot_l_over_hmin, decay_index_br/tot_l_over_hmin	Total of all separate MPIL ratios of $l/h(n_{\mathrm{cr}})_{\mathrm{min}}$
decay_index_blos/l_over_minhmin, decay_index_br/l_over_minhmin	Ratio of MPIL $l/h(n_{\mathrm{cr}})_{\mathrm{min}}$ (for MPIL having lowest $h(n_{\mathrm{cr}})_{\mathrm{min}}$)
decay_index_blos/maxl_over_hmin, decay_index_br/maxl_over_hmin	Ratio of MPIL $l/h(n_{\mathrm{cr}})_{\mathrm{min}}$ (for longest MPIL)
flare_past	Binary flag for occurrence of $\geqslant$1 flare in previous 24 hr
flare_index_past	Accumulated GOES flare peak magnitudes in previous 24 hr
frdim_blos/frdim, frdim_br/frdim	Fractal dimension
gs_slf/g_s	Sum of the horizontal magnetic gradient
gs_slf/slf	Separation distance lead. and follow. polarity subgroups
ising_energy_blos/ising_energy, ising_energy_br/ising_energy	Ising energy (calculated pixel-by-pixel)
ising_energy_part_blos/ising_energy_part, ising_energy_part_br/ising_energy_part	Ising energy (calculated using B_{eff} flux partitions)
lat_hg	Heliographic latitude of SHARP centroid
lon_hg	Heliographic longitude of SHARP centroid
mf_spectrum_blos/dq, mf_spectrum_br/dq	Multi-fractal generalized correlation dimension spectrum
mpil_blos/max_length, mpil_br/max_length	Maximum length of a single MPIL
mpil_blos/tot_length, mpil_br/tot_length	Total length of all MPILs
mpil_blos/tot_usflux, mpil_br/tot_usflux	Total unsigned flux around all MPILs
r_value_blos_logr, r_value_br_logr	Schrijver's R ($\log_{10}$ form)
sfunction_blos/zq, sfunction_br/zq	Multi-fractal structure function inertial range index
wlsg_blos/value_int, wlsg_br/value_int	Falconer's ${}^{\mathrm{L}}\mathrm{WL}_{\mathrm{SG}}$
flow_field_bvec/diver, flow_field_bvec/diver_max, flow_field_bvec/diver_mean	Flow field divergence (total, maximum, mean)
flow_field_bvec/shear, flow_field_bvec/shear_max, flow_field_bvec/shear_mean	Flow field shear (total, maximum, mean)
flow_field_bvec/v_ mean, flow_field_bvec/v_median	Flow field total velocity magnitude (mean, median)
flow_field_bvec/vz_max, flow_field_bvec/vz_mean	Flow field vertical velocity magnitude (mean, median)
flow_field_bvec/w_diver, flow_field_bvec/w_diver_max	Flux-weighted flow field divergence (total, maximum)
flow_field_bvec/w_diver_mean	Flux-weighted flow field divergence (mean)
flow_field_bvec/w_shear, flow_field_bvec/w_shear_max	Flux-weighted flow field shear (total, maximum)
flow_field_bvec/w_shear_mean	Flux-weighted flow field shear (mean)

Table 5.10: Feature list with short descriptions.

Feature	Description
helicity_energy_bvec/abs_tot_dedt	Abs. val. net vertical Poynting flux
helicity_energy_bvec/abs_tot_dedt_in	Abs. val. net vertical Poynting flux (emerg. comp.)
helicity_energy_bvec/abs_tot_dedt_sh	Abs. val. net vertical Poynting flux (shear. comp.)
helicity_energy_bvec/abs_tot_dedt_in_plus_sh	Emerg. + shear. abs. values net vertical Poynting flux
helicity_energy_bvec/abs_tot_dhdt	Abs. val. net vertical helicity flux
helicity_energy_bvec/abs_tot_dhdt_in	Abs. val. net vertical helicity flux (emerg. comp.)
helicity_energy_bvec/abs_tot_dhdt_sh	Abs. val. net vertical helicity flux (emerg. comp.)
helicity_energy_bvec/abs_tot_dhdt_in_plus_sh	Emerg. + shear. abs. values net vertical helicity flux
helicity_energy_bvec/tot_uns_dedt	Total unsigned vertical Poynting flux
helicity_energy_bvec/tot_uns_dedt_in	Tot. unsign. vertical Poynting flux (emerg. comp.)
helicity_energy_bvec/tot_uns_dedt_sh	Tot. unsign. vertical Poynting flux (shear. comp.)
helicity_energy_bvec/tot_uns_dhdt	Tot. unsign. vertical helicity flux
helicity_energy_bvec/tot_uns_dhdt_in	Tot. unsign. vertical helicity flux (emerg. comp.)
helicity_energy_bvec/tot_uns_dhdt_sh	Tot. unsign. vertical helicity flux (shear. comp.)
nn_currents/tot_us_cur	Total unsigned non-neutralized currents
sharp_kw/gamma/ave, sharp_kw/gamma/stddev	Field inclin. ang. (mean, st. dev.)
sharp_kw/gamma/skewness, sharp_kw/gamma/kurtosis	Field inclin. ang. (skewn., kurt.)
sharp_kw/gamma/total, sharp_kw/gamma/max, sharp_kw/gamma/median	Field inclin. ang. (tot., max., med.)
sharp_kw/ggt45fract	% tot. area with shear angle $> 45^{\circ}$
sharp_kw/hgradbh/ave, sharp_kw/hgradbh/stddev	Horiz. grad. B_{hor} (mean, st. dev.)
sharp_kw/hgradbh/skewness, sharp_kw/hgradbh/kurtosis	Horiz. grad. B_{hor} (skewn., kurt.)
sharp_kw/hgradbh/total, sharp_kw/hgradbh/max, sharp_kw/hgradbh/median	Horiz. grad. B_{hor} (tot, max, med)
sharp_kw/hgradbt/ave, sharp_kw/hgradbt/stddev	Horiz. grad. B_{tot} (mean, st. dev.)
sharp_kw/hgradbt/skewness, sharp_kw/hgradbt/kurtosis	Horiz. grad. B_{tot} (skewn., kurt.)
sharp_kw/hgradbt/total, sharp_kw/hgradbt/max, sharp_kw/hgradbt/median	Horiz. grad. B_{tot} (tot., max., med.)
sharp_kw/hgradbz/ave, sharp_kw/hgradbz/stddev	Horiz. grad. B_r (mean, st. dev.)
sharp_kw/hgradbz/skewness, sharp_kw/hgradbz/kurtosis	Horiz. grad. B_r (skewn., kurt.)
sharp_kw/hgradbz/total, sharp_kw/hgradbz/max, sharp_kw/hgradbz/median	Horiz. grad. B_r (tot., max., med.)

Table 5.11: Feature list with short descriptions.

Feature	Description
sharp_kw/hz/ave, sharp_kw/hz/stddev	Vert. curr. hel. (mean, st. dev.)
sharp_kw/hz/skewness, sharp_kw/hz/kurtosis	Vert. curr. hel. (skewn., kurt.)
sharp_kw/hz/total, sharp_kw/hz/max, sharp_kw/hz/median	Vert. curr. hel. (tot., max., med.)
sharp_kw/jz/ave, sharp_kw/jz/stddev	Vert. curr. (mean, st. dev.)
sharp_kw/jz/skewness, sharp_kw/jz/kurtosis	Vert. curr. (skewn., kurt.)
sharp_kw/jz/total, sharp_kw/jz/max, sharp_kw/jz/median	Vert. curr. (tot., max., med.)
sharp_kw/rho/ave, sharp_kw/rho/stddev	Photosph. excess magn. en. (mean, st. dev.)
sharp_kw/rho/skewness, sharp_kw/rho/kurtosis	Photosph. excess magn. en. (skewn., kurt.)
sharp_kw/rho/total, sharp_kw/rho/max, sharp_kw/rho/median	Photosph. excess magn. en. (tot., max., med.)
sharp_kw/rhod/ave, sharp_kw/rhod/stddev	Photosph. excess magn. en. dens. (mean, st. dev.)
sharp_kw/rhod/skewness, sharp_kw/rhod/kurtosis	Photosph. excess magn. en. dens. (skewn., kurt.)
sharp_kw/rhod/total, sharp_kw/rhod/max, sharp_kw/rhod/median	Photosph. excess magn. en. dens (tot., max., med.)
sharp_kw/sflux/ave, sharp_kw/sflux/stddev	Signed flux (mean, st. dev.)
sharp_kw/sflux/skewness, sharp_kw/sflux/kurtosis	Signed flux (skewn., kurt.)
sharp_kw/sflux/total, sharp_kw/sflux/max, sharp_kw/sflux/median	Signed flux (tot., max., med.)
sharp_kw/sheargamma/ave, sharp_kw/sheargamma/stddev	B_{tot} shear angle (mean, st. dev.)
sharp_kw/sheargamma/skewness, sharp_kw/sheargamma/kurtosis	B_{tot} shear angle (skewn., kurt.)
sharp_kw/sheargamma/total, sharp_kw/sheargamma/max, sharp_kw/sheargamma/median	B_{tot} shear angle (tot., max., med.)
sharp_kw/snetjzpp/total	Sum abs. val. net currents per polarity
sharp_kw/twistp/ave, sharp_kw/twistp/stddev,	Twist parameter (mean, st. dev.,)
sharp_kw/twistp/skewness, sharp_kw/twistp/kurtosis	Twist parameter (skewn., kurt.)
sharp_kw/twistp/total, sharp_kw/twistp/max, sharp_kw/twistp/median	Twist parameter (tot., max., med.)
sharp_kw/usflux/ave, sharp_kw/usflux/stddev,	Uns. flux (mean, st. dev.)
sharp_kw/usflux/skewness, sharp_kw/usflux/kurtosis	Uns. flux (skewn., kurt.)
sharp_kw/usflux/total, sharp_kw/usflux/max, sharp_kw/usflux/median	Uns. flux (tot., max., med.)
sharp_kw/ushz/ave, sharp_kw/ushz/stddev,	Uns. vert. curr. hel. (mean, st. dev.)
sharp_kw/ushz/skewness, sharp_kw/ushz/kurtosis	Uns. vert. curr. hel. (skewn., kurt.)
sharp_kw/ushz/total, sharp_kw/ushz/max, sharp_kw/ushz/median	Uns. vert. curr. hel. (tot., max., med.)
sharp_kw/usiz/ave, sharp_kw/usiz/stddev,	Uns. vert. curr. (mean, st. dev.)
sharp_kw/usiz/skewness, sharp_kw/usiz/kurtosis	Uns. vert. curr. (skewn., kurt.)
sharp_kw/usiz/total, sharp_kw/usiz/max, sharp_kw/usiz/median	Uns. vert. curr. (tot., max., med.)

Chapter 6

Solar image desaturation as an inverse problem

Image saturation is an issue for several instruments in solar astronomy, mainly at Extreme Ultraviolet (EUV) wavelengths: an example is the AIA instrument on board SDO which realizes an unprecedented EUV view of solar corona and its dynamics. EUV imaging is crucial for providing a clear-cut picture of the dynamical structure of the solar corona at many different time and spatial scales [7; 69; 141]. Observations at these wavelengths are probably the only data that can provide direct clear visualizations of magnetic reconnection as the trigger of magnetic energy release [47; 113; 149; 154], reveal in detail the thermal structure of the solar atmosphere [65] and therefore explain basic plasma physics processes like coronal heating [8] and irradiance [146], and unveil still unresolved diagnostic issues concerned with coronal waves and oscillations [83]. From a space weather perspective, the ability of EUV imaging to point out, both spatially and dynamically, the connection between solar flares and coronal mass ejections (CMEs) paves the way to understand how Sun's variability impacts the escape of energetic particles into the heliosphere [91]. As typically happens in EUV imaging, SDO/AIA observations of solar flares may be significantly limited by the presence of two kinds of image artifacts, diffraction and saturation. A tool for desaturating such images, called

DESAT, has been developed in [120; 121; 137]. An inverse diffraction problem is formulated in order to restore the primary saturated region. However DESAT has some limitations: the main one is the fact that it can not be used when a reliable estimate of the background is not available. This is the case for instance of the super solar storm on September 10, 2017.

In this Chapter we develop a novel computational approach for the analysis of SDO/AIA saturated images able to recover the signal in the primary saturation region without any a priori estimate of the background. Such a method, called Sparsity-Enhancing DESAT (SE-DESAT) is based on alternating the PRiL method (see Chapter 4) with Expectation Maximization (EM) algorithm for Poisson data. In order to introduce the method we first formalize the process of saturation which comprises two phenomena: primary saturation and blooming. We formulate an equation model for the signal acquisition process which takes into account both the diffraction model and the blooming process. Finally, we test the performance on both simulated and real data comparing them with the DESAT ones. Finally we show the effectiveness of the new method also on the solar storm occurred on September 10, 2017 on which DESAT can not be used.

6.1 Introduction to the problem

The classical model of astronomical image reconstructions describes the observed image as the result of a convolution between an unknown object and the Point Spread Function (PSF) of the observation instrument. The PSF models the impulse response of the whole optical system to a point source placed far from the optical system, thus encoding the image degradation due to the optical system of the telescopes. Most PSFs are made of just a core peak that induces diffusion effects but there are also PSFs which have a non local component: indeed, in addition to the core one, a PSF can have more complex structures due to wave scattering against the filters support which replicates the central peak according to regular diffraction patterns of varying intensity. In addition to the

image blur, telescopes based on the Charged Coupled Device (CCD) imaging technology can present artifacts due to the saturation effect. Saturation [90] happens when the incident photons exceed the sensor capacity and includes two phenomena: the primary saturation, which refers to the condition where CCD pixels lose their ability to accommodate additional charge and therefore for intense incoming photon flux a set of pixel cells reaching its Full Well Capacity, stores the maximum number possible of photon-induced electrons; the blooming, named also secondary saturation, which refers to the fact that the additional charge spreads into neighboring pixels, causing bright artifacts in a circular region around a single pixel or along the horizontal or vertical axis in the image. The blooming effect takes place with the exceeding of the limit to how much charge each pixel can store, and its consequence is that the electrons excited by the incoming photons spill out onto adjacent pixels. Despite the efforts undertaken to build more and more efficient devices, the fact remains that making an instrument with higher spatial resolution requires smaller pixels, which are more likely affected by saturation and blooming effects with increasing incoming photon flux.

In the solar images provided by the Atmospheric Imaging Assembly (AIA) telescopes [80], these three kinds of degradation, i.e. non-local diffraction, primary saturation and blooming, become clearly visible when the incoming light is enough (see Figure 6.1). AIA is mounted on the Solar Dynamics Observatory (SDO) NASA satellite has been launched in February 2010 and provides an unprecedented EUV view of the solar corona and its dynamics, allowing to obtain several significant scientific results. The four telescopes of AIA capture images of the Sun's atmosphere in ten separate wave bands, seven of which centered at EUV wavelengths (94 Å, 131 Å, 171 Å, 193 Å, 211 Å, 304 Å, 335 Å), providing full-disk 4096×4096 pixel images with a time cadence of 12 s and with pixel width in the range $0.6 - 1.5$ arcsec. The image reconstruction problem in AIA/SDO is an important scientific issue as the brightest images of the Sun, showing highest energetic events such as big solar flares, are degraded to such an extent that they cannot be useful to the solar

scientists. It is also a big data issue since saturation effects involve, around 10^5 images per year. Diffraction and primary saturation play a competing role in the image processing effort for AIA. Indeed, just part of the incoming signal accumulates in the CCD pixels up to saturation, while the other part is coherently and linearly scattered to produce diffraction pixels unaffected by saturation. As shown in [121; 137], this fact has a crucial implication for image restoration: all information lost due to primary saturation is actually present, as regular ghosts, in the diffraction fringes and therefore such an information can be recovered by means of an inverse diffraction procedure. The method presented in these papers, called DESAT, is able to estimate the saturated region with some limitations. The core idea of the DESAT method is that the non local effect of diffraction brings information on the pixel intensity of the saturated region from which it has been generated. Therefore, to recover the photometry of the saturated region, the DESAT method (1) estimates the support of the saturated region mainly producing the diffraction effect, (2) performs an inversion process which restores the pixel content of the saturated region from the diffraction values and (3) makes an interpolation step to estimate the pixel intensity of the image in the remaining bloomed region. As mentioned before, in order to work properly, the DESAT method needs an estimation of the solar activity without the diffraction effect, i.e. an estimation of the background. This is the actual drawback of DESAT: the background estimation consists in exploiting the fact that a typical AIA observation along a time range of some minutes, is characterized by some unsaturated frames since such a telescope is equipped with a feedback system which automatically reduces the exposure time in correspondence of intense emission. The unsaturated frames are used for providing an a priori estimate of the background. In detail, the estimation consists in interpolating the pixel values of the two unsaturated images recorded just before and just after the image which has to be de-saturated. However, some of the most interesting events are correlated to acquired images where strong saturation effects occur for a whole time series, as for example the solar storm on September 2017: in

particular on September 10 2017, at the wavelength 171 Å all images suffer of significant primary saturation and blooming effects for more than an hour, involving about 300 consecutive EUV maps completely deteriorated [63]. In these cases DESAT can not be used since an a priori estimate of the background is not available.

We propose a new method, called Sparsity-Enhancing DESAT (SE-DESAT), which can be used in these dramatic cases. The idea is that diffraction effects in the original image come from a subset of pixels of the saturated region. Such pixels can be identified using a sparsity-enhancing (in detail, a Lasso-type) method which selects those pixels whose diffraction PSF most correlates the original signal. In the Lasso terminology such pixels belong to the *active set*, whereas the pixels that do not correlate much belong to the *inactive set*. In our case sparsity is not used in the standard way concerning astronomical image reconstruction problems [41; 45; 123] in which the signal is usually compressed in some suitable basis (e.g. wavelet basis), but rather in a way more similar to the one considered in learning applications: the sparsity in the pixel space is used to select the variables most explaining the diffraction model, or the features most predictive in machine learning terminology (see Chapter 5). Furthermore, our approach takes into account that in the original signal the diffraction effect is superimposed to the normal solar activity and therefore it considers an unknown background to be estimated. As a final step of the method we proposed to use an inpainting procedure to fill the pixels of the inactive set, being, in practice, the information on these bloomed pixels irremediably lost. However, when the bloomed pixels are considerably much more than the pixels whose diffraction fringes are clearly visible in the image, the above strategy does not perform correctly. This is due to the fact that enlarging the saturated region, the probability that the background solar activity spuriously correlates with the diffraction effects increases, and this degrades the restoration of the saturated region. This issue led us to consider, together with the standard diffraction model, an additional model for the signal in the saturated part of the image which relies on a peculiar feature of

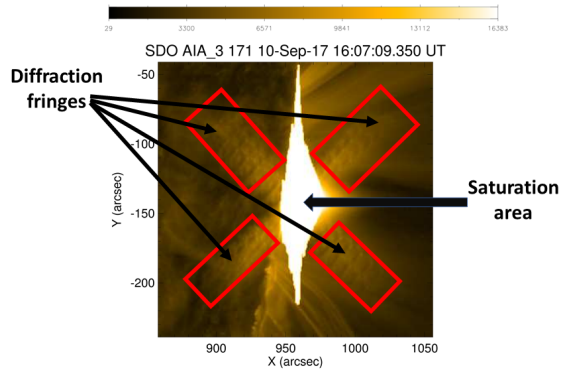


Figure 6.1: An example of saturated AIA image at 171 Å wavelength with highlighted the overall (primary + blooming) saturation region and the diffraction fringes (the event occurred on September 10, 2017 at the acquisition time 16:07:09 UT).

the non linear behavior of the blooming effect. Indeed, the blooming effect appears along the columns of the image and we can consider that the total amount of charge excited is maintained along columns despite the presence of blooming. This approximation allows us to consider that the integral of the restored images along the columns of the saturated region should be kept approximately equal to the same integral applied to the original image. We will show that using this additional data in the inverse restoration problem permits to obtain more reliable results even when the blooming region is large.

6.2 Signal formation process

The signal formation process of an optical system can be described by a model equation as follows

$$h = K * f \tag{6.1}$$

where $K \in \mathcal{L}^2(\mathcal{X} \times \mathcal{Y})$ is the point spread function of the instrument, where $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$ are two intervals, $h \in \mathcal{L}^2(\mathcal{X} \times \mathcal{Y})$ is the ideal data, $f \in \mathcal{L}^2(\mathcal{X} \times \mathcal{Y})$ is the incoming photon flux and $*$ indicates the convolution operation. Therefore,

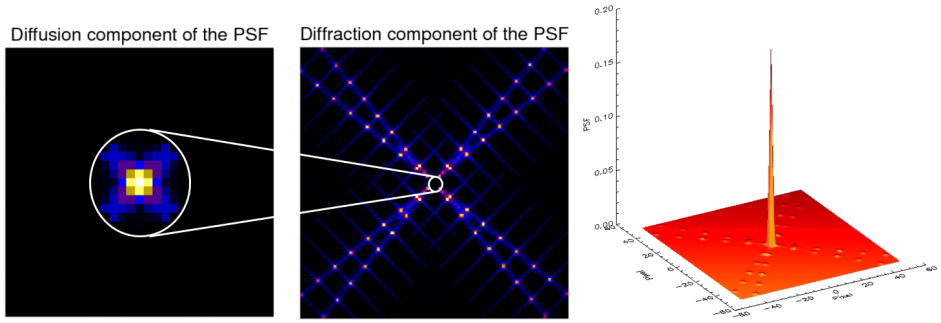


Figure 6.2: The AIA PSF of the bandwidth 171 Å. The diffusion (or core) and diffraction components of the AIA PSF are in the left and central panels, respectively. In the left panel a zoom of the core component is reported. The core component is located at the center of the image in the central panel, in which the diffraction component is reported. In the right panel we report a 3D view of the AIA PSF.

we have the following integral equation

$$h(x,y) = \int_{\mathcal{X}\times\mathcal{Y}} K(x-x',y-y')f(x',y')d(x',y'), \tag{6.2}$$

for each $(x,y) \in \mathcal{X} \times \mathcal{Y}$. The measured data h belongs to the data space or image space $\mathcal{I}$, whereas the incoming photon flux f belongs to the object space $\mathcal{O}$ (see Figure 6.3).

The PSF of the AIA instrument can be modeled as the sum of two contributions [111]: the core PSF which takes into account the local diffusion of the light and the diffraction PSF which describes the non-local component given by the diffraction pattern corresponding to a point source (see Figure 6.2). Therefore, the equation model is characterized by a linear integral operator whose kernel is the sum of the core and the diffraction PSF, i.e.

$$K = K_c + K_d \tag{6.3}$$

is the sum of the local and non-local components. These two components can be thought of as compact supported as the diffraction patterns are elongated along two axis at 40 and 50 degrees with respect the x-axis in Cartesian

coordinate [106]. The blur (and the diffraction) is a degradation due to the signal formation process. However other kinds of degradation are introduced during the recorded process, or signal acquisition process, which we see in the next section.

6.3 Signal acquisition process

Beside blur and artifacts associated to the PSF, images are affected by noise, which is a degradation introduced during the signal acquisition process. The recording hardware is placed in the image space and in the case of AIA the acquisition is according to a standard CCD-based imaging technique. The noise which affects counting processes is the so-called Poisson noise. AIA data can be thought approximately Poisson, since the returned data are Data Number (DN), which are obtained by dividing the recorded charge by the average charge per photon. However, during the acquisition not only noise is added but AIA CCD pixels are affected by saturation effects. We refer to saturation as the whole of two phenomena: the primary saturation which refers to the fact that a pixel reaches the maximum possible value ($M = 2^{14}$ DN), and the blooming which refers to the fact that, once a pixel is saturated, it can affect the value of its neighboring pixels. In the following we first describe the primary saturation phenomenon and then we propose a formalization for the entire saturation process which includes also the blooming effect, by means of a nonlinear operator between Hilbert spaces.

6.3.1 Primary saturation

Let us consider a function h representing the result of the signal formation process (see equation (6.1)). We define the primary saturated region as

$$S = \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid h(x, y) \geq M\} \, . \tag{6.4}$$

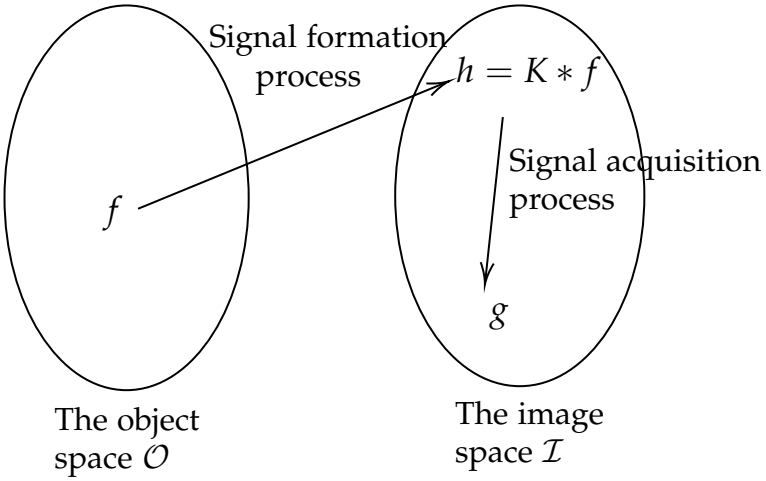


Figure 6.3: Geometric representation of the signal formation and acquisition processes.

The primary saturation is a threshold operator

$$\mathcal{S}(h)(x,y) = \begin{cases} h(x,y) & (x,y) \notin S \\ M & (x,y) \in S \end{cases} . \tag{6.5}$$

Therefore, the primary saturated data g is given by

$$g = \mathcal{S}(h). \tag{6.6}$$

We consider the sub-region of the image defined by

$$F = \{(x,y) \in \mathcal{X} \times \mathcal{Y} \mid (K_d * 1_S)(x,y) > 0 \ , \ (K_c * 1_S)(x,y) = 0\} \ , \tag{6.7}$$

where 1_Z indicates the characteristic function on the set Z. F contains the replicates of the primary saturated region S due to the diffraction effect and does not contain core effects.

By splitting the signal f in the two regions S and S^C, where S^C denotes the complementary of the set S, from equations (6.1) and (6.3) we obtain

$$h = K_c * f + K_d * f_{|S} + K_d * f_{|S^C} \ , \tag{6.8}$$

where the convolution of K_d with f restricted to a set T (with $T = S$ or $T = S^C$) is defined as follows

$$(K_d * f_{|T})(x,y) = \int_T K_d(x-x', y-y') f(x',y') d(x',y'), \tag{6.9}$$

where $(x,y) \in \mathcal{X} \times \mathcal{Y}$. Since in AIA images the visible diffraction effects come only from the primary saturated parts of the images, we can assume that $K_d * f_{|S^C}$ is negligible with respect to the other terms. Then, by restricting the signal formation model (6.8) to F and by using equation (6.6) we get

$$g_{|F} = (K_d * f_{|S})_{|F} + b \tag{6.10}$$

where $b := (K_c * f)_{|F}$ can be thought of as a background restricted to the set F. In this context, the background is the image deprived by diffraction effects. Equation (6.10) states that if we know the primary saturation region S and the background b, we can retrieve the photon flux intensity in S by solving an inverse diffraction problem.

6.3.2 Saturation (primary saturation and blooming)

In this section, we include the blooming effect in the saturation model. As the saturation process of the CCD of the AIA telescope takes place along the vertical axes of the images, the primary saturated region can be written in the following form

$$S := \bigcup_{x \in \mathcal{X}} S_x \quad , \quad S_x = \{y \in \mathcal{Y} \mid h(x,y) > M\} \, . \tag{6.11}$$

Provided that S_x is connected for each $x \in \mathcal{X}$, we start modeling the saturation process by considering the two-dimensional nonlinear operator $\mathcal{S} : \mathcal{L}^2(\mathcal{X} \times \mathcal{Y}) \to \mathcal{L}^2(\mathcal{X} \times \mathcal{Y})$ which is defined by

$$g(x,y) := \mathcal{S}(h)(x,y) = \begin{cases} h(x,y) & y \notin \tilde{S}_x \\ M & y \in \tilde{S}_x \end{cases} \tag{6.12}$$

where, for each $x \in \mathcal{X}$,

$$\tilde{S}_x = \{y \in \mathcal{Y} \mid |y - y_0(x)| \leq t(x)\} , \tag{6.13}$$

and where $y_0(x) = \frac{1}{2}(\max(S_x) + \min(S_x))$ and $t(x)$ satisfies the condition

$$\int_{y_0(x)-t(x)}^{y_0(x)+t(x)} (M - \min(h(x,y), M))dy = \int_{S_x} (h(x,y) - M)dy . \tag{6.14}$$

The saturation operator $\mathcal{S}$ describes the process of vertically and symmetrically charge spilling out of the primary saturated region S from the medium point $y_0(x)$ for each $x \in \mathcal{X}$. If $S = \emptyset$ the r.h.s. of equation (6.14) is identically zero and this implies that $t(x) = 0$ for each $x \in \mathcal{X}$, i.e. no charges are spilled out as no saturation effect takes place. The saturation process makes the observed image flat in the total saturated region

$$\tilde{S} := \bigcup_{x \in \mathcal{X}} \tilde{S}_x . \tag{6.15}$$

The diffraction fringes of the saturated region can be modeled as for the primary saturation. Therefore, we define the diffraction fringes corresponding to the saturated region as

$$\tilde{F} = \{(x,y) \in \mathcal{X} \times \mathcal{Y} \mid (K_d * 1_{\tilde{S}})(x,y) > 0 \ , \ (K_c * 1_{\tilde{S}})(x,y) = 0\} , \tag{6.16}$$

and again from equations (6.1) and (6.12), by splitting f in the two regions $\tilde{S}$ and $\tilde{S}^C$ and then by restricting the resulting signal formation model to $\tilde{F}$, with K given by equation (6.3), we get

$$g_{|\tilde{F}} = (K_d * f_{|\tilde{S}})_{|\tilde{F}} + b \tag{6.17}$$

where $b := (K_c * f)_{|\tilde{F}}$ is the background restricted to the fringes $\tilde{F}$. Then, as we consider that the diffraction effects generated by the pure blooming region $\tilde{S} - S$ are negligible with respect to the ones produced by the primary

saturation region, solving an inverse diffraction problem to retrieve the photon flux intensity in S needs an estimate of the primary saturation region. This is the reason why the DESAT method [120] performs an estimate of the primary saturation region as a starting point. However, in section 6.5, we propose a sparsity-enhancing method which automatically performs the segmentation of the saturation region in primary and blooming.

6.3.3 Integrated core model

In section 6.3.2, we have considered the restriction of the signal g to the diffraction fringe support. However, we can provide an image acquisition model also for the signal in the saturated region. Indeed, having supposed that the signal integrated along the vertical axes of the image is maintained (see equation (6.14)), we can consider an integral equation

$$\bar{g}(x) := \int_{\tilde{S}_x} g(x,y)dy = \int_{\tilde{S}_x} (K_c * f_{|\tilde{S}})(x,y)dy + w, \tag{6.18}$$

where $w = \int_{\tilde{S}_x} (K_c * f_{|\tilde{S}^C})(x,y)dy$ includes the local effects coming from the non saturated region $\tilde{S}^C$. We refer to this model as the *integrated core model*. This additional model equation can be considered together with the diffraction model (6.17). We will show with some numerical experiments that the addition of the integrated core model can be very useful to improve the reconstruction of the unknown photon flux of the saturated region.

6.4 Discretization

Since the AIA data are images of 4096×4096 pixels, we have to consider the discretization of equations (6.17) and (6.18). By taking $n = 4096$ equispaced points for each axis, $\{(x_p, y_q)\}_{p,q=1,\dots,n}$ we can write the discretization of equation (6.1) as

$$\mathbf{h} = \mathbf{A}\beta^*, \tag{6.19}$$

where $\mathbf{h} = \{h(x_p, y_q)\}_{p,q=1,\dots,n} \in \mathbb{R}^N$, $\beta^* = \{f(x_p, y_q)\}_{p,q=1,\dots,n} \in \mathbb{R}^N$ and $\mathbf{A} = \{K(x_p - x_{p'}, y_q - y_{q'})\}_{p,q,p',q'=1,\dots,n} \in \mathbb{R}^N \times \mathbb{R}^N$ with $N = n^2$ having used the lexicographic order for the image pixel rearrangement and cyclic boundary condition for the convolution operator. Moreover, we consider the discretization of the saturated data g in equation (6.12) denoting it with $\mathbf{I}$, therefore,

$$\mathbf{I} = \{g(x_p, y_q)\}_{p,q=1,\dots,n} \in \mathbb{R}^N. \tag{6.20}$$

Given an index i of the vector $\mathbf{h}$, we define with r the index transformation $r : \{1, \dots, N\} \to \{1, \dots, n\} \times \{1, \dots, n\}$ which returns the row and column indexes $(q, p) \in \{1, \dots, n\} \times \{1, \dots, n\}$ associated with the index i before the rearrangement. As in the infinite dimensional case, the matrix $\mathbf{A}$ can be split into two parts

$$\mathbf{A} = \mathbf{A}_D + \mathbf{A}_C, \tag{6.21}$$

with $\mathbf{A}_D$ and $\mathbf{A}_C$ the circulant matrices associated with the diffraction component of the PSF and with the diffusion component, respectively (see equation (6.3)). With a slight abuse of notation, we keep using the same symbols for the saturated region, the fringes region and the saturation operator. The saturated region is given by

$$\tilde{S} = \{i \in \{1, \dots, N\} : \mathbf{I}_i \geq M\}, \tag{6.22}$$

and the saturation operator $\mathcal{S} : \mathbb{R}^N \to \mathbb{R}^N$ is defined as

$$\mathcal{S}(\mathbf{h}) = \begin{cases} \mathbf{h}_i & i \notin \tilde{S} \\ M & i \in \tilde{S} \end{cases}. \tag{6.23}$$

The saturated data $\mathbf{I} \in \mathbb{R}^N$ is given by

$$\mathbf{I} = \mathcal{S}(\mathbf{h}). \tag{6.24}$$

The set of the diffraction fringes $\tilde{F}$ defined in equation (6.16) is computed in the following way

$$\tilde{F} = \{i \in \{1, \dots, N\} : (\mathbf{A}_D \mathbf{1}_{\tilde{S}})_i > 0, (\mathbf{A}_C \mathbf{1}_{\tilde{S}})_i = 0\}, \tag{6.25}$$

where $\mathbf{1}_{\tilde{S}}$ indicates the index rearrangement of the mask of the region $\tilde{S}$, i.e. $(\mathbf{1}_{\tilde{S}})_i = 1$ when $i \in \tilde{S}$ and 0 otherwise. Therefore, the diffraction model equation (6.17) can be written as follows

$$\mathbf{I}_{\tilde{F}} = \mathbf{A}_D^{\tilde{S}} \beta^*_{\tilde{S}} + \mathbf{b}, \tag{6.26}$$

where $\mathbf{A}_D^{\tilde{S}} : \mathbb{R}^{\#\tilde{S}} \to \mathbb{R}^{\#\tilde{F}}$ is the sub-matrix of $\mathbf{A}_D$ given by $(\mathbf{A}_D^{\tilde{S}})_{i,j} = (\mathbf{A}_D)_{i,j}$ when $i \in \tilde{S}$ and $j \in \tilde{F}$, mapping the photon flux emitted in the region $\tilde{S}$ into the flux recorded into the diffraction fringes $\tilde{F}$; $\beta^*_{\tilde{S}}$ denotes the restiction of β^* in $\tilde{S}$ and $\mathbf{b}$ is the discretized background (see equation (6.17)), i.e. $\mathbf{b} := (\mathbf{A}_C \beta^*)_{\tilde{F}}$.

The integrated core model can be discretized as follows. We denote with $\tilde{L}_{\tilde{S}}$ the discretized saturation region

$$\tilde{L}_{\tilde{S}} := \bigcup_{p \in \{1,\dots,n\}} \tilde{L}_p, \quad \tilde{L}_p = \{q \in \{1, \dots, n\} \mid g(x_p, y_q) \geq M\}. \tag{6.27}$$

We consider the column indexes for which at least one pixel is saturated $J_{\tilde{S}} = \{p \in \{1, \dots, n\} \mid \tilde{L}_p \neq \emptyset\}$. The discrete version of the integrated core model equation reads as

$$\mathbf{y} = \mathbf{C} \beta^*_{\tilde{S}} + \mathbf{w}, \tag{6.28}$$

where $\mathbf{w}$ is the discretization of w in equation (6.18),

$$\mathbf{y}_j = \sum_{i \in T_j} \mathbf{I}_i, \qquad (\mathbf{C}\beta^*_{\tilde{S}})_j = \sum_{i \in T_j} (\mathbf{A}_C^{\tilde{S}} \beta^*_{\tilde{S}})_i \tag{6.29}$$

with $j \in J_{\tilde{S}}$ and

$$T_j := \{i \in \{1, \dots, N\} : (r(i))_1 \in \tilde{L}_j\} \tag{6.30}$$

where $(r(i))_1$ indicates the first entry of the rearrangement. The inverse

diffraction model (6.19) and the integrated core model (6.28) together can be written as an unique equation as follows

$$\mathbf{G} = \mathbf{T}\beta^*_{\tilde{S}} + \mathbf{W} \tag{6.31}$$

where

$$\mathbf{G} := \begin{bmatrix} \mathbf{I}_{\tilde{F}} \\ \mathbf{y} \end{bmatrix}, \mathbf{T} := \begin{bmatrix} \mathbf{A}_D^{\tilde{S}} \\ \mathbf{C} \end{bmatrix} \text{ and } \mathbf{W} := \begin{bmatrix} \mathbf{b} \\ \mathbf{w} \end{bmatrix}. \tag{6.32}$$

6.5 De-saturation with a sparsity-enhancing approach

In this section we describe in details our new approach for de-saturating AIA images and we compare it with the DESAT method [120; 121; 137], which is, to our knowledge, the only available method to desaturate AIA images. DESAT method consists mainly of three steps: the first step (segmentation) allows us to separate the primary saturation region from the blooming one and to select the diffraction fringes pixels, the second step (reconstruction) consists in estimating the photon flux in the primary saturation region by solving an inverse diffraction problem and the last step (synthesis) returns the de-saturated image by projecting in the data space the solution given in the reconstruction step. We remark that all these three steps need an a priori estimate of the background and this limitation does not allow DESAT to de-saturate many consecutive images in the AIA dataset. Here we propose an approach in which segmentation and reconstruction are performed simultaneously by means of an ℓ_1-penalized method inducing sparsity in the pixel space. This strategy is able to estimate a constant background. Such an ℓ_1-penalized method is our novel Lasso-type method PRiL (see Chapter 4), which takes into account the Poisson nature of data. In particular, we recall that PRiL method consists in minimizing a penalized functional where the fidelity term is a globally quadratic approximation of the Kullback-Leibler divergence and the penalty term is the ℓ_1 norm of the flux: this ensures to select those pixels which most correlate with the data, automatically segmenting the saturation region. The

novel approach uses PRiL first for giving an initialization of the photon flux in the saturated region and then by alternating it with an iteration of the Expectation Maximization (EM) algorithm for Poisson data [124] to estimate the background into the diffraction fringes. Finally the synthesis step is needed to project the estimated photon flux into the image space.

6.5.1 The novel method

The desaturation method is composed by three steps: first, it provides a rough estimate of the photon flux with a constant background; second, it alternates an iteration of the EM algorithm and the PRiL method in order to refine the estimation of the background in the diffraction fringes and the photon flux in the primary saturation region; third, it performs a synthesis step by projecting the estimated incoming photon flux to the image space and estimating the value of the bloomed pixels by means of an inpainting procedure. We remark that the method can be applied to both the diffraction model only (6.19) and the total model (6.31). For the sake of simplicity, we introduce a general notation which takes into account both cases. Consider

$$\mathbf{Y} = \mathbf{X}\beta^*_{\tilde{S}} + \mathbf{B}, \tag{6.33}$$

where

$$\begin{cases} \mathbf{Y} := \mathbf{I}_{\tilde{F}}, \\ \mathbf{X} := \mathbf{A}^{\tilde{S}}_{D}, \\ \mathbf{B} := \mathbf{b} \end{cases} \quad \text{or} \quad \begin{cases} \mathbf{Y} := \mathbf{G}, \\ \mathbf{X} := \mathbf{T}, \\ \mathbf{B} := \mathbf{W} \end{cases}$$

in the case of the diffraction model only (6.19) or in the case of the total model (6.31), respectively. We notice that in the first case the equation (6.33) is a vector-equation of dimension $\#\tilde{F}$, whereas in the second case the equation (6.33) is a vector-equation of dimension $\#\tilde{F} + \#J_{\tilde{S}}$.

We now describe the three steps of the proposed algorithm.

1. **Initialization.** We compute the PRiL solution for a given set of λ; i.e. we solve the following minimization problem

$$(\hat{\alpha}^{(0)}, \hat{\beta}^{(0)}) = \arg \min_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^{\#\tilde{S}}} \left|\left| \frac{\mathbf{Y} - \mathbf{X}\beta - \alpha}{\sqrt{\mathbf{Y}+1}} \right|\right|_2^2 + \lambda\|\beta\|_1, \tag{6.34}$$

where α is a constant intercept to estimate a zero-order approximation of the background. The division in the fit term has to be intended as element-wise. The regularization parameter λ is chosen such that the estimated total flux approximates the recorded one in the saturated area $\tilde{S}$. The regularization parameter choice is described in detail in section 6.5.2.

2. **Iterative alternating method**.

- Input: $\hat{\beta}^{(0)}$, $\mathbf{B}^{(0)} := \hat{\alpha}^{(0)}$, $\mathbf{X}$,$\mathbf{Y}$
- $k = 0, 1, \ldots$
  - estimate of the background by means of one iteration of the EM algorithm for Poisson data,

$$\hat{\mathbf{B}}^{(k+1)} = \mathbf{B}^{(k)} \frac{\mathbf{Y}}{\mathbf{X}\hat{\beta}^{(k)} + \mathbf{B}^{(k)}}, \tag{6.35}$$

    where the division and the first product in the r.h.s. of equation (6.35) has to be intended as element-wise;
  - estimate of the primary saturation by means of the PRiL method with $\hat{\mathbf{B}}^{(k+1)}$ provided in the previous step

$$(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}) = \arg \min_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^{\#\tilde{S}}} \left|\left| \frac{\mathbf{Y} - \mathbf{X}\beta - \hat{\mathbf{B}}^{(k+1)} - \alpha}{\sqrt{\mathbf{Y}+1}} \right|\right|_2^2 + \lambda\|\beta\|_1. \tag{6.36}$$

    The regularization parameter λ is selected according to the above mentioned procedure;

– update the background

$$\mathbf{B}^{(k+1)} = \hat{\mathbf{B}}^{(k+1)} + \hat{\alpha}^{(k+1)}; \tag{6.37}$$

- stop the iterative procedure when the C-statistic computed on diffraction fringes pixels (see equation (6.42)) is approximately close to 1. The stopping rule is described in detail in section 6.5.3;

- output: $\hat{\beta}^{(k_{opt})}$.

3. **Synthesis**. We define the de-saturated image $\hat{\mathbf{I}}$ as follows

$$\hat{\mathbf{I}} = \begin{cases} (\mathbf{A}_C^{\tilde{S}} \hat{\beta}^{(k_{opt})})_S, & \text{in } S \\ \text{inpaintig procedure}, & \text{in } B \\ \mathbf{I}_F - (\mathbf{A}_D^{\tilde{S}} \hat{\beta}^{(k_{opt})})_F, & \text{in } F \\ \mathbf{I}, & \text{elsewhere}, \end{cases} \tag{6.38}$$

where $S := \{i \in \tilde{S} : \hat{\beta}_i^{(k_{opt})} \neq 0\}$, $B := \{i \in \tilde{S} : \hat{\beta}_i^{(k_{opt})} = 0\}$ and F is the set of diffraction fringes generated by S (see equation (6.16)), i.e.

$$F := \{i \in \tilde{F} : (\mathbf{A}_D \mathbf{1}_S)_i \neq 0, (\mathbf{A}_C \mathbf{1}_S)_i = 0\}. \tag{6.39}$$

We point out that the method performs an automatic segmentation of the overall saturation region $\tilde{S}$ in the two regions S and B, the primary saturated and the blooming one by exploiting sparsity in the pixel space. This sparsity constraint is effective since the diffraction effects produced by the pixels in the blooming area are negligible with respect to the ones produced by the primary saturated region and the sparsity-enhancing method allows us to select the pixels that most correlate with the data. Therefore, according to the sparsity-enhancing method terminology, the primary saturated pixels compose the active set, whereas the blooming pixels belong to the inactive set. This means that the estimated incoming photon flux $\hat{\beta}^{(k_{opt})}$ in the blooming pixels

is 0. Therefore, in the synthesis step we use an inpainting strategy to fill these pixels with reasonable values continuously depending on the rest of the image values. To recover their value we exploit the inpainting procedure proposed in [54; 145], which is based on the Discrete Cosine Transform.

6.5.2 Regularization parameter choice

In order to select the regularization parameter we first compute the PRiL solutions $\hat{\beta}_{\lambda_j}$ for a grid of parameter values $\{\lambda_j\}_{j=0}^{T}$ by using the fast Coordinate Descent strategy proposed in [53]. We then consider the expected total flux E in the saturated region $\tilde{S}$ as a function of the regularization parameter. As we expect that E is a decreasing function of the regularization parameter, we look for the regularization parameter such that the total flux convolved with the core PSF is approximately equal to the recorded one. To do this, we exploit a simple bisection method. Operationally, for any solution $\hat{\beta}_{\lambda_j}$ we project it into the image space by means of the core component of the PSF and we fill the zero values in the pure blooming region by means of the above mentioned inpanting procedure [54; 145]. Formally, the expected total flux is defined by

$$E(\lambda_j) := \sum_{i \in \tilde{S}} (\hat{\mathbf{z}}_{\lambda_j})_i, \tag{6.40}$$

where

$$\hat{\mathbf{z}}_{\lambda_j} = \begin{cases} (\mathbf{A}_C^{\tilde{S}} \hat{\beta}_{\lambda_j})_{S_j}, & \text{in } S_j \\ \text{inpainting procedure}, & \text{in } B_j \end{cases}, \tag{6.41}$$

$S_j := \{i \in \tilde{S} : \hat{\beta}_{\lambda_j} \neq 0\}$ and $B_j := \{i \in \tilde{S} : \hat{\beta}_{\lambda_j} = 0\}$. When $E(\lambda_j)$ is equal to the overall recorded signal in the saturated region up to a given tolerance $tol > 0$, (set to 10^{-3} in our experiments), the bisection method stops. As the evaluation of the function E involves the inpainting procedure which is a quite costly operation, the bisection method allows a significant reduction of the number of evaluations. A pseudo code of this strategy is reported in Algorithm 3.

6.5.3 Stopping rule

At each iteration k of the iterative algorithm described in the second step we evaluate the C-statistic on the diffraction fringes generated by the estimated primary saturated pixels of the solution $\hat{\beta}^{(k+1)}$. Formally, we compute

$$S^{(k+1)} := \{i \in \tilde{S} : \hat{\beta}_i^{(k+1)} \neq 0\},$$

$$F^{(k+1)} := \{i \in \tilde{F} : (\mathbf{A}_D \mathbf{1}_{S^{(k+1)}})_i \neq 0, (\mathbf{A}_C \mathbf{1}_{S^{(k+1)}})_i = 0\}$$

and the C-statistic

$$C_{stat}^{(k+1)} = \frac{2}{\#F^{(k+1)}} \quad \sum_{i \in F^{(k+1)}} \quad (\mathbf{I}_{\tilde{F}})_i \log \frac{(\mathbf{I}_{\tilde{F}})_i}{(\mathbf{A}_D^{\tilde{S}} \hat{\beta}^{(k+1)})_i + \mathbf{b}_i^{(k+1)}} + (\mathbf{A}_D^{\tilde{S}} \hat{\beta}^{(k+1)})_i + \mathbf{b}_i^{(k+1)} - (\mathbf{I}_{\tilde{F}})_i, \tag{6.42}$$

where $\mathbf{b}_i^{(k+1)}$ is the estimated background at the k-th iteration (in the case of the diffraction model $\mathbf{b}^{(k+1)} = \mathbf{B}^{(k+1)}$, in the case of the total model $\mathbf{b}^{(k+1)}$ is given by the first $\#\tilde{F}$ components of $\mathbf{B}^{(k+1)}$). C-statistic measures the discrepancy between the original unsaturated image and the reconstructed de-saturated image according to the Kullback-Leibler topology. We stop the algorithm at the first iteration in which the C-statistic is smaller than 1. As the C-statistic provides a goodness of fit, a too much small value indicates data overfitting. Then, in the case the C-statistic becomes smaller than a fixed threshold, it is preferable to keep the last iteration with C-statistic higher than 1. In applications we fix the threshold equal to $tol_{C_{stat}} = 0.6$. For the sake of robustness, we fixed a maximum number of iterations $k_{\max} = 10$. In our experiments the stopping rule is always satisfied before the sixth iteration.

6.6 Relation with DESAT method

Whereas the novel method is based on the discretization of equations (6.17) and (6.18), the DESAT method relies on the discrete version of equation

(6.10) assuming an estimate of the background. The first step of DESAT is a correlation analysis to segment the primary saturation region from the blooming one. Once the primary saturation region is estimated the equation model (6.10) reads as

$$\mathbf{I}_{\hat{F}} = \mathbf{A}_D^{\hat{S}} \beta_{\hat{S}}^* + \hat{\mathbf{b}}, \tag{6.43}$$

where $\hat{S}$ is the estimated primary saturation region, $\hat{F}$ is the associated diffraction fringe set and $\hat{\mathbf{b}}$ is an a priori estimate of the background restricted to the set $\hat{F}$. The second step consists in estimating the photon flux in the primary saturation region $\hat{S}$ by means of an inverse diffraction procedure [121] applied to the equation (6.43) based on the Expectation Maximization (EM) algorithm for Poisson data regularized by early stopping [11]. Finally, the third step is the synthesis, which is analogous to the one used in our method, with the difference that the intensities in the blooming pixels are estimated with the a priori estimate of the background whereas in our method they are estimated with the inpainting strategy.

6.7 Algorithm

We provide pseudo codes for the regularization parameter choice rule, the stopping rule and the entire reconstruction procedure in Algorithms 3, 4 and 5 respectively.

Algorithm 3 Regularization parameter choice rule

Input: regularization path $\{(\hat{\alpha}_{\lambda_j}, \hat{\beta}_{\lambda_j})\}_{j=0}^{T}$, $\tilde{S}$, $\tilde{F}$, $\mathbf{A}_D^{\tilde{S}}$, $\mathbf{I}$ and tol.
Initialize $j_{\min} = 0$ and $j_{\max} = T$,
while $j_{\max} - j_{\min} \geq 2$ **do**
  Set $j = \frac{j_{\min}+j_{\max}}{2}$, $\lambda^{\text{opt}} = \lambda_j$ and take the solution $(\hat{\alpha}_{\lambda_j}, \hat{\beta}_{\lambda_j})$.
  Compute the flux $E(\lambda_j)$ as in equation (6.40).
  if $E(\lambda_j) > \sum_{i\in\tilde{S}} \mathbf{I}_i$ (overestimation) **then**
    $j_{\max} = j$
  else if $E(\lambda_j) < \sum_{i\in\tilde{S}} \mathbf{I}_i$ (underestimation) **then**
    $j_{\min} = j$
  end if
  if $\frac{|E(\lambda_j) - \sum_{i\in\tilde{S}} \mathbf{I}_i|}{\sum_{i\in\tilde{S}} \mathbf{I}_i} < tol$ (approximately equal) **then**
    stop
    return λ^{opt}
  end if
end while

Algorithm 4 Stopping rule

Input: $\hat{\beta}^{(k+1)}$, $\tilde{S}$, $\tilde{F}$, $\mathbf{A}_D^{\tilde{S}}$, $\mathbf{I}$, $\mathbf{b}^{(k+1)}$, $tol_{C_{stat}}$.
Identify the primary saturation region: $S^{(k+1)} := \{i \in \tilde{S} : \hat{\beta}_i^{(k+1)} \neq 0\}$.
Compute $F^{(k+1)} := \{i \in \tilde{F} : (\mathbf{A}_D \mathbf{1}_{S^{(k+1)}})_i \neq 0, (\mathbf{A}_C \mathbf{1}_{S^{(k+1)}})_i = 0\}$.
Compute C-statistic $C_{stat}^{(k+1)}$ as in equation (6.42).
if $C_{stat}^{(k+1)} \geq 1$ **then**
  Go to the next iteration
else if $C_{stat}^{(k+1)} < 1$ **and** $C_{stat}^{(k+1)} > tol_{C_{stat}}$ **then**
  stop and
  return $k_{opt} = k$
end if
if $C_{stat}^{(k+1)} \leq tol_{C_{stat}}$ **then**
  stop and
  return $k_{opt} = k - 1$
end if

Algorithm 5 SE-DESAT method

1: Input: $\mathbf{Y}$, $\tilde{S}$, $\mathbf{X}$
2: **Initialization**
3: Compute the PRiL solutions on a set of regularization parameters $\{\lambda_j\}_{j=1}^{T}$, i.e. for each λ_j solve

$$(\hat{\alpha}_{\lambda_j}, \hat{\beta}_{\lambda_j}) = \arg \min_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^{\#\tilde{S}}} \left|\left| \frac{\mathbf{Y} - \mathbf{X}\beta - \alpha}{\sqrt{\mathbf{Y}+1}} \right|\right|_2^2 + \lambda_j \|\beta\|_1, \tag{6.44}$$

and select λ^{opt} according to Algorithm 3.
4: Set $(\hat{\alpha}^{(0)}, \hat{\beta}^{(0)}) = (\hat{\alpha}_{\lambda^{\text{opt}}}, \hat{\beta}_{\lambda^{\text{opt}}})$.
5: **Iterative alternate method**
6: Initialize $\mathbf{B}^{(0)} := \hat{\alpha}^{(0)}$.
7: **for** $k \geq 0$ **do**
8: An iteration of EM algorithm for Poisson data to compute the background

$$\hat{\mathbf{B}}^{(k+1)} = \mathbf{B}^{(k)} \frac{\mathbf{Y}}{\mathbf{X}\hat{\beta}^{(k)} + \mathbf{B}^{(k)}}. \tag{6.45}$$

9: Compute the PRiL solutions on a set of regularization parameters $\{\lambda_j\}_{j=1}^{T}$, i.e. for each λ_j solve

$$\arg \min_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^{\#\tilde{S}}} \left|\left| \frac{\mathbf{Y} - \mathbf{X}\beta - \hat{\mathbf{B}}^{(k+1)} - \alpha}{\sqrt{\mathbf{Y}+1}} \right|\right|_2^2 + \lambda_j \|\beta\|_1 \tag{6.46}$$

and select λ^{opt} according to Algorithm 3.
10: Set $(\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}) = (\hat{\alpha}_{\lambda^{\text{opt}}}, \hat{\beta}_{\lambda^{\text{opt}}})$ and update

$$\mathbf{B}^{(k+1)} = \hat{\mathbf{B}}^{(k+1)} + \hat{\alpha}^{(k+1)}. \tag{6.47}$$

11: Stopping rule according to Algorithm 4.
12: **if** stopping rule is satisfied **then**
13: stop and
14: **return** k_{opt}.
15: **end if**
16: **end for**
17: **Segmentation**
18: Compute $S := \{i \in \tilde{S} : \hat{\beta}_i^{(k_{\text{opt}})} \neq 0\}$, $B := \{i \in \tilde{S} : \hat{\beta}_i^{(k_{\text{opt}})} = 0\}$ and F as in equation (6.39).
19: **Synthesis**
20: Output: de-saturated image $\hat{\mathbf{I}}$ defined in (6.38).

6.8 Experimental results

In this section we test the performance of the proposed method in both the case of synthetic and real data, comparing the results with the ones provided by DESAT method. We apply our method in both the case of the diffraction model in equation (6.17) and of the total model in equation (6.31). We refer to it as SE-DESAT* in the first case and as SE-DESAT in the second one. Concerning synthetic data, we show two simulations: we first recover simple gaussian sources in three different configurations and second a realistic simulated data generated by an unsaturated AIA/SDO image. In both cases we generate the saturation effect by means of a suitable algorithm based on equations (6.12), (6.13) and (6.14). The pseudo code is described in Algorithm 6. Concerning real data, we desaturate AIA/SDO images corresponding to two different events: the September 6, 2011 and the September 10, 2017. Finally in section 6.8.3 we provide an accurate analysis on the solar storm occurred on September 10, 2017. We remark that in this case the DESAT method can not be used, due to the lack of a reliable background estimate.

6.8.1 Simulation studies

In order to simulate the saturation effects we need to discretize the saturation process described in section 6.3.2. We set all pixels whose grey level is larger than the saturation level M equal to M and we artificially expand the total sum of the photon flux exceeding M along columns in a symmetric way.

We report in Algorithm 6 a pseudo code to mimic the presence of primary saturation and blooming: the algorithm approximates the process described in formulas (6.12), (6.13) and (6.14). To make easy the writing we consider that the set of saturated pixels along each column is connected and we use the following notation: $\mathbf{h}$ is the ideal image in the case the CCD does not suffer from saturation effects (see equation (6.19)) and $\mathbf{I}$ is the output image with saturation effects. We denote with $\mathbf{h}[q, p]$ the elements in the q-th row and p-th column of the image.

Algorithm 6 Primary saturation and blooming simulation

Input: $\mathbf{h}$ unsaturated image.
Initialization: $\mathbf{I} = \mathbf{h}$.
$\mathbf{I}[q, p] = M$, for (q, p) such that $\mathbf{h}[q, p] \geq M$. (discretization of the threshold operator in equation (6.5)).
Let $L_p = \{q \in \{1, \ldots, n\} : \mathbf{I}[q, p] = M\}$ and let $J_S = \{p \in \{1, \ldots, n\} : L_p \neq \emptyset\}$.
for $p \in J_S$ **do**
  $d := \sum_{q \in L_p} \mathbf{h}[q, p] - \#L_p \times M$ (discretization of the r.h.s. of equation (6.14)).
  Let $d^{(1)}_{above} := \frac{d}{2}$ and $d^{(1)}_{below} := \frac{d}{2}$. We approximate the process described in equations (6.12), (6.13) and (6.14) as follows

```
for k ≥ 1 do
  c = M − I[max(L_p) + k, p]
  if c < d^(k)_above then
    I[max(L_p) + k, p] = M
    d^(k+1)_above = d^(k)_above − c
  else if c ≥ d^(k)_above then
    stop the for loop
  end if
end for
```

```
for k ≥ 1 do
  c = M − I[min(L_p) − k, p]
  if c < d^(k)_below then
    I[min(L_p) − k, p] = M
    d^(k+1)_below = d^(k)_below − c
  else if c ≥ d^(k)_below then
    stop the for loop
  end if
end for
```

end for
return $\mathbf{I}$ saturated image.

First simulation study. We consider three ground truth objects each of which constituted by three two-dimensional Gaussian sources. The three objects differ for the parameter values of the gaussian sources as we report in Table 6.1 (the energy $\mathcal{E}$, the position of the center (x_c, y_c) and the standard deviation σ). The configuration of the first object is characterized by three well separated sources with different standard deviations and intensities; in the second one there are the same three sources but the smallest is much closer to the biggest one; in the third configuration the three sources are all close to each other in order to call up the typical loop shape of a solar flare. We convolve these three objects with the global AIA PSF of the 131 Å passband

Table 6.1: Parameters associated to the synthetic sources of the three configurations: $\mathcal{E}$ is the energy, σ is the standard deviation and (x_c, y_c) is the position of the center in arcseconds.

Configuration 1			**Configuration 2**			**Configuration 3**		
$\mathcal{E}$ (10^5)	σ	(x_c, y_c)	$\mathcal{E}$ (10^5)	σ	(x_c, y_c)	$\mathcal{E}$ (10^5)	σ	(x_c, y_c)
7	2	$(285, 123)$	7	2	$(297, 129)$	7	2	$(294, 129)$
8	4	$(273, 141)$	8	4	$(273, 141)$	10	3	$(291, 132)$
10	6	$(303, 123)$	10	6	$(303, 123)$	9	6	$(291, 126)$

wavelength and we perturb it with Poisson noise. Finally we add the primary saturation and blooming effects by applying Algorithm 6.

In Table 6.2 we compare the performance of our method with the DESAT one in terms of C-statistic, relative error (computed in norm) and the relative error computed only in the primary saturation region. Furthermore, we report in the same table the confusion matrix associated to each method for giving a quantitative measure of the goodness of the saturated pixel estimation. In this regard, true positives (TP) are the pixels correctly estimated to have values higher than the saturation level M (primary saturated pixels), true negatives (TN) are the pixels in the saturation region correctly estimated to have values smaller than M (blooming pixels), false negatives (FN) are the pixels in the primary saturation region incorrectly estimated to be blooming pixels and false positives (FP) are the pixels in the blooming region incorrectly estimated to be primary saturated pixels.

Table 6.2: C-statistic, relative error (RE), relative error in the primary saturation region (RE-P) and confusion matrix for the three configurations considered in the first simulation study provided by DESAT, SE-DESAT* and SE-DESAT.

	Configuration 1				
	C-stat	RE	RE-P	Confusion matrix	
DESAT	2.9859	0.1162	0.0841	TP = 578 FP = 87	FN = 9 TN = 790
SE-DESAT*	0.9356	0.2941	0.0853	TP = 570 FP = 11	FN = 17 TN = 866
SE-DESAT	0.8019	0.2844	0.0444	TP = 561 FP = 0	FN = 26 TN = 877
	Configuration 2				
	C-stat	RE	RE-P	Confusion matrix	
DESAT	2.9162	0.1144	0.088	TP = 586 FP = 82	FN = 8 TN = 828
SE-DESAT*	0.9749	0.2866	0.0897	TP = 582 FP = 19	FN = 12 TN = 891
SE-DESAT	0.8349	0.2749	0.0416	TP = 578 FP = 12	FN = 16 TN = 898
	Configuration 3				
	C-stat	RE	RE-P	Confusion matrix	
DESAT	1.7827	0.0606	0.0366	TP = 254 FP = 8	FN = 2 TN = 488
SE-DESAT*	0.6906	0.1464	0.02879	TP = 244 FP = 2	FN = 12 TN = 494
SE-DESAT	0.7393	0.1446	0.0218	TP = 249 FP = 1	FN = 7 TN = 495

From a morphological point of view the three methods provide reconstructions of the configuration 3 very similar to the ground truth, whereas the reconstructions of configuration 1 and 2 provided by SE-DESAT are slightly better than the ones provided by DESAT and SE-DESAT*. This is due to the fact that SE-DESAT takes advantage of the integrated core model. Moreover, we remark that DESAT needs an estimate of the background and in these simulations we used the true background, given by $\mathbf{A}_C\beta^*$, which is unknown in real experiments. Indeed, the relative error computed from the DESAT reconstruction is smaller than the one provided by both SE-DESAT methods. This is mainly due to the fact that in DESAT reconstructions the blooming pixel value is the original one as the background is exact, whereas the blooming pixels in reconstructions provided by SE-DESAT methods are estimated with

an inpainting procedure. Nonetheless, both SE-DESAT methods achieve a better relative error in the primary saturation region and a better C-statistic value which is always close to 1. From the analysis of the confusion matrices we notice that DESAT provides a higher number of TP and a smaller number of TN with respect to the ones provided by SE-DESAT methods. However, except for the third configuration, the pixels incorrectly estimated by DESAT (FN+FP) are much more than the ones incorrectly estimated by SE-DESAT methods (see Table 6.2). Finally, Figure 6.5 shows the photon flux integrated along the columns affected by the saturation effects. We compare the integral of the photon flux of the saturated image with the integral of the estimated photon flux of the reconstructed signals. We remark that for each configuration the estimated integrated profiles fit the data for all the three methods. In the case of SE-DESAT method the fit is almost exact as the integrated core model is taken into account, whereas in the case of SE-DESAT* the profiles are over- and under-estimated in such a way that the total amount of signal in the overall saturation region is maintained thanks to the regularization parameter choice rule. In the case of DESAT method the profiles are usually overestimated in agreement with the high number of FP. It is worth observing that the proposed sparsity-enhancing method is able to reconstruct these three kinds of configurations despite not having at disposal the background. Furthermore, as shown in the case of the second configuration, the method is able to reconstruct a small point source also when it is close to a broader source.

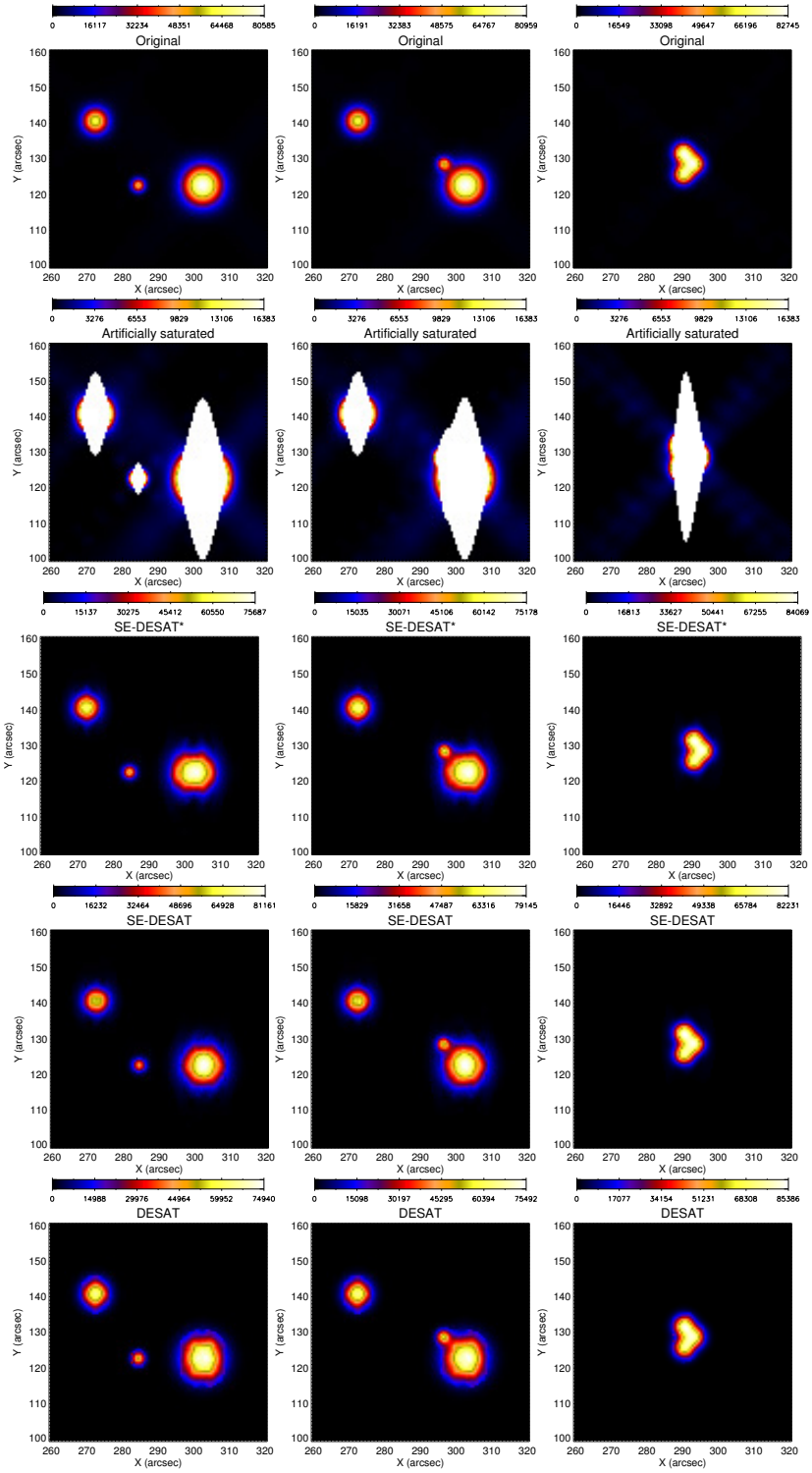
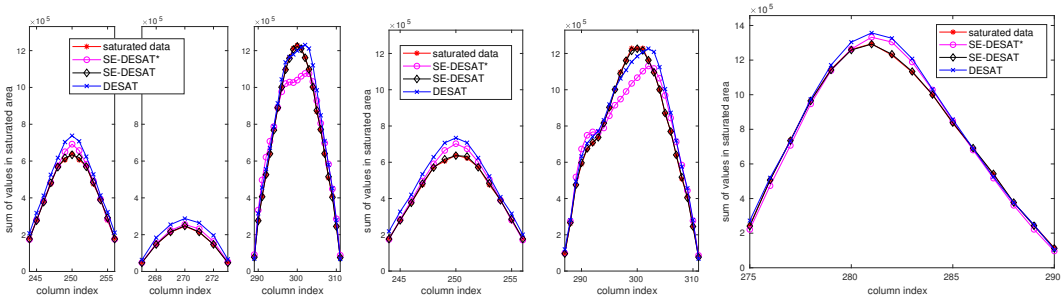


Figure 6.4: First simulation study. First row: ground-truth images. Second row: synthetic saturated images corrupted by Poisson noise. Third, fourth and fifth rows: reconstructions obtained by SE-DESAT*, SE-DESAT and DESAT methods, respectively. Left, middle and right columns refer to configurations 1, 2 and 3, respectively.



(a) Configuration 1. (b) Configuration 2. (c) Configuration 3.

Figure 6.5: First simulation study. Comparison of the reconstructed flux profiles integrated along the saturated columns obtained by SE-DESAT*, SE-DESAT and DESAT methods with the ground truth profiles. In the first (resp. the second) configuration the three (resp. two) plots correspond to the three (resp. two) connected components of the saturated region.

Second simulation study. In the second study we aim to desaturate a realistic simulated data whose saturation is artificially generated starting from the unsaturated real AIA image of September 6, 2011 at 22:07:09 UT at 131 Å wavelength (see first panel in Figure 6.6). The simulation process is implemented similarly to the one considered in [137]. It is described by the following steps:

- deconvolve the original map with the total PSF to eliminate the diffraction fringes and the blur by means of the EM algorithm for Poisson data (see second panel in Figure 6.6);
- re-scale the image in order to have a region with pixel intensity higher than the saturation level M (see third panel in Figure 6.6). The rescaled intensity is given by

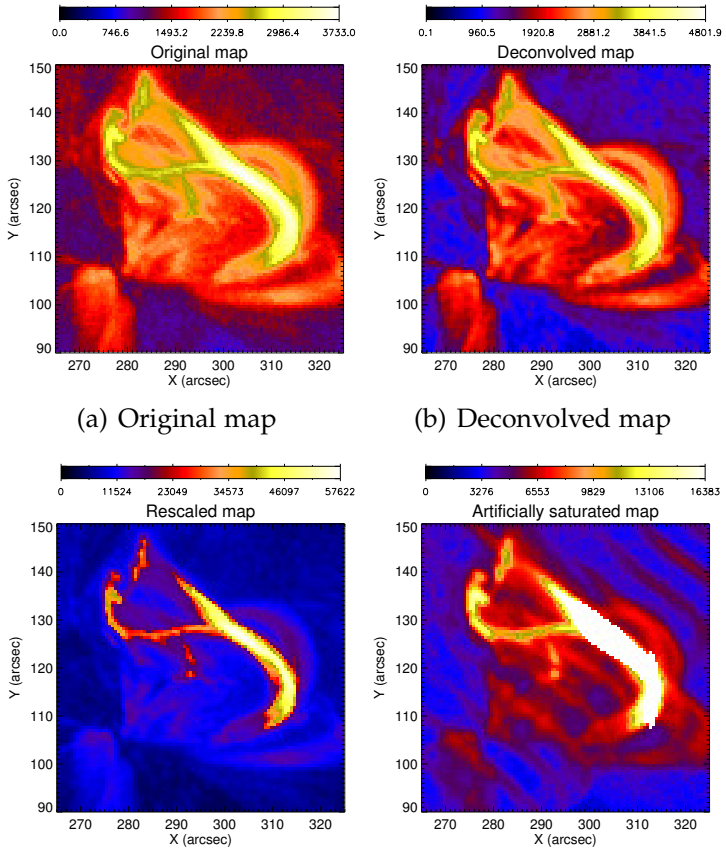
$$z_{rescaled} = \begin{cases} \frac{m\tilde{M}-T}{\tilde{M}-T} z + \frac{\tilde{M}(1-m)}{\tilde{M}-T}, & \text{if } z \geq T \\ z, & \text{if } z < T, \end{cases} \tag{6.48}$$

  where z is the pixel intensity before the rescaling, T is the threshold which defines which pixels have to be rescaled and it is set as $T = \frac{\tilde{M}}{4}$, $\tilde{M}$

is the maximum intensity in the image and $m = 12$;

- saturate the image: first convolve the rescaled map with the total PSF, second perturb data with Poisson noise and finally simulate the saturation by means of Algorithm 6 (see fourth panel in Figure 6.6).

The ground-truth which we want to recover is the result of the convolution between the rescaled image with the core of PSF (see first panel in Figure 6.7): such a map represents the ideal image that would be recorded by AIA if there were not diffraction nor saturation effects. Figure 6.7 shows the results of DESAT, SE-DESAT* and SE-DESAT methods, for which we repeat the analysis provided in the first simulation study. We report in Table 6.3 the C-statistic values, the relative errors, the relative errors in the primary saturation region and the confusion matrices. From Table 6.3 we remark that our methods give much better C-statistic than DESAT method. In this case the relative errors provided by our methods are better than the ones given by DESAT and only the relative error in the primary saturation region provided by SE-DESAT* is higher than the one given by DESAT. Exactly as in the first simulation the number of TP provided by DESAT is higher than the ones provided by our methods whereas the number of TN is smaller. From the integrated flux reconstruction analysis (see Figure 6.7) we notice that the profiles produced by SE-DESAT coherently fit the data, as expected. On the contrary, DESAT over-estimates the integrated profiles in the left part of the saturation region (see bottom panel in Figure 6.7). This over-estimation is reflected in an incorrect reconstruction of the diffraction fringe values, clearly visible in the black structures appearing at the top right part of the reconstructed image (second panel in Figure 6.7). Such a reconstruction is provided by DESAT despite the fact that the original background is used. A study of the performance of the DESAT method by randomly perturbating the original background is provided in [137] and, as we expect, the performance gets worse by increasing the standard deviation of the perturbation.



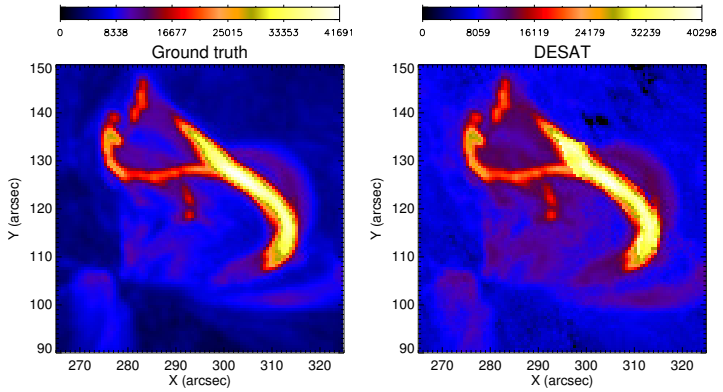
(a) Original map (b) Deconvolved map

(c) Rescaled map (d) Saturated map

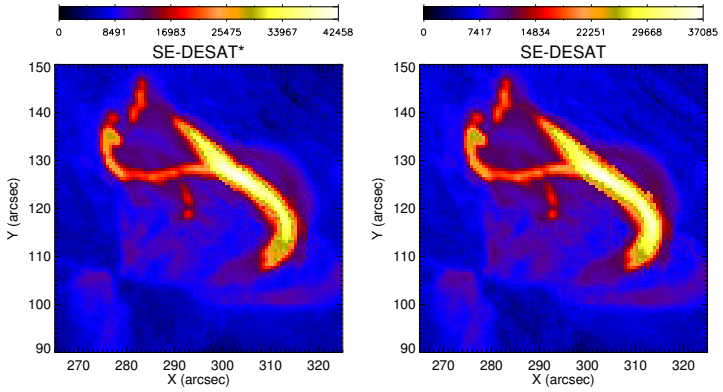
Figure 6.6: Second simulation study. From left to right: original image, deconvolved image, re-scaled image and saturated image corrupted by Poisson noise.

Table 6.3: C-statistic, relative error (RE), relative error in the primary saturation region (RE-P) and confusion matrix for the synthetic saturated image considered in the second simulation study provided by DESAT, SE-DESAT* and SE-DESAT.

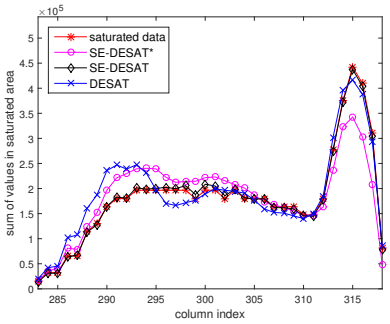
	C-stat	RE	RE-P	Confusion matrix	
DESAT	16.2052	0.2738	0.1467	TP = 217 FP = 21	FN = 12 TN = 148
SE-DESAT*	1.6892	0.1717	0.1563	TP = 197 FP = 16	FN = 32 TN = 153
SE-DESAT	1.840	0.1649	0.1399	TP = 204 FP = 11	FN = 25 TN = 158



(a) Ground truth (b) DESAT reconstruction



(c) SE-DESAT* reconstruction (d) SE-DESAT reconstruction



(e) Integrated flux profiles.

Figure 6.7: Second simulation study. First row: from left to right ground truth image and reconstruction obtained by DESAT. Second row: from left to right reconstructions obtained by SE-DESAT* and SE-DESAT. Third row: comparison of the integrated flux profiles.

6.8.2 Real data

We test the performance of our method on the event occurred on September 6, 2011. We consider two images corresponding to wavelength 131 Å, the first at 22:19:25 UT and the second at 22:19:09 UT, and a third image at wavelength 193 Å and time 22:16:43 UT. The first image represents a real data with few saturated pixels, whereas in the second image the saturation effect is really dramatic and the blooming effect dominates. The third example presents a mildly saturated image, with a quite moderate blooming. In all these cases it turns to be possible to estimate the background from some unsaturated maps before and after each acquisition time and therefore to compare with the DESAT method.

We report in Figure 6.8, the original images and the reconstructions provided by the three methods and the comparison of the reconstructed fluxes integrated along the saturated columns.

In Figure 6.8 the first column contains low saturated data and its corresponding desaturated images. From a morphological point of view the saturated region appears to be similar across the three reconstructions. However, the reconstruction in the diffraction fringes provided by DESAT presents a larger set of black pixels with 0 value: this is due to the overestimation of the pixel values in the saturated region as it is confirmed by the integrated profiles (see Figure 6.9). On the other hand, the integrated flux profiles estimated by SE-DESAT methods yield a good data fit. The second column in Figure 6.8 shows the reconstructions of a dramatically saturated image: in this case the reconstructions provided by the three methods are similar to each other solely in the brightest part of the image. Both SE-DESAT methods furnish a solution that is broader than the DESAT one. The comparison of the integrated flux profiles shows that both DESAT and SE-DESAT* do not fit correctly the integrated signal. This is not the case for SE-DESAT which implements the integrated core model. Finally, the third column in Figure 6.8 shows the reconstructions of the mildly saturated image. In this case the SE-DESAT* reconstruction seems to be affected by some artifacts (or edge effects) in the top

part of the saturation region. SE-DESAT reconstruction appears to be much more similar to the DESAT one from a morphological point of view. The main difference consists in locating the large majority of the flux in the upper part of the central structure. Also in this case DESAT method overestimates the integrated flux profiles. For all the three cases the C-statistic computed on the diffraction fringes provided by DESAT is much higher (see [137]) than the ones provided by both SE-DESAT methods which are smaller than 1 in each three cases.

Finally, we show the effectiveness of our method where DESAT can not be applied. We consider two images from the super storm occurred on September 10, 2017: the first acquired at the acquisition time 16:00:47 UT at 94 Å wavelength and the second at 16:07:09 UT at 171 Å. In these two examples DESAT method is ineffective since a reliable a priori estimate of the background can not be provided: for example in the case of the wavelength 171 Å all images are saturated for more than an hour around the considered acquisition time. In Figure 6.10 we report the reconstructions provided by SE-DESAT methods. In both cases there is an extraordinary amount of saturation with a strongly pronounced blooming effect. The reconstructions provided by SE-DESAT* are evidently corrupted by edge effects which are clearly visible on the frontier of the saturated region. These effects are dumped in the SE-DESAT reconstructions having considered here the integrated core model in addition to the diffraction one. In SE-DESAT reconstructions is more evident the loop structure of the solar flare: this is particularly emphasized in the second case (see second row third panel in Figure 6.10). From a methodological point of view the difference between SE-DESAT* and SE-DESAT is that the latter one is obliged to fit the integrated profiles (see third row in Figure 6.10): this constraint appears to be a key point to improve the quality of the reconstructions when no background estimation is available.

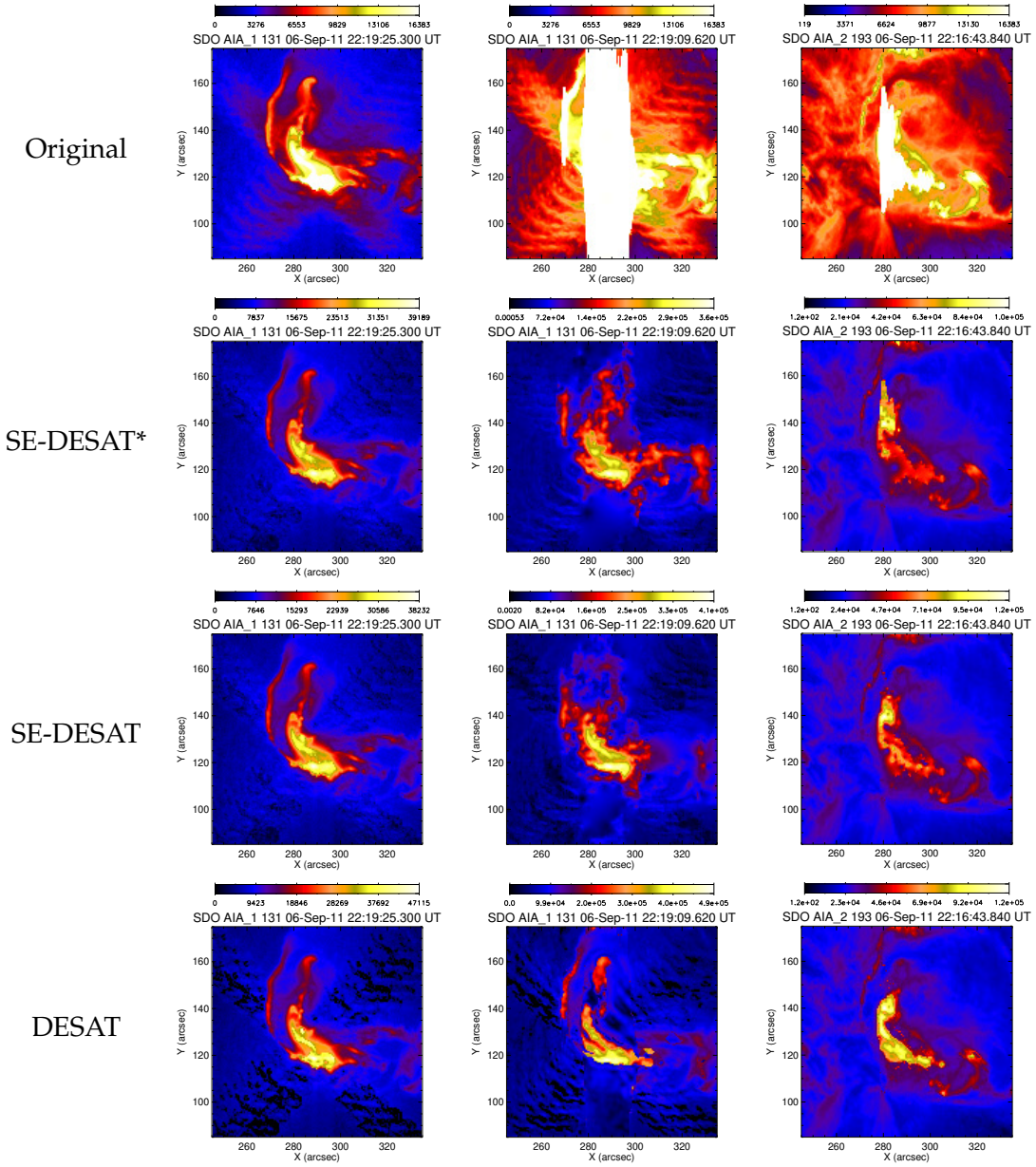
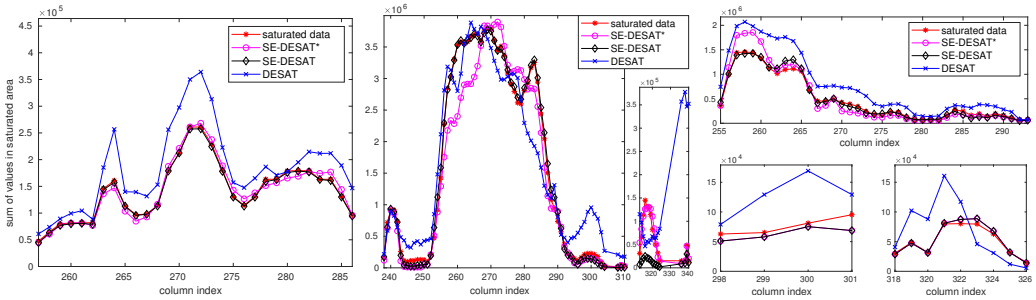
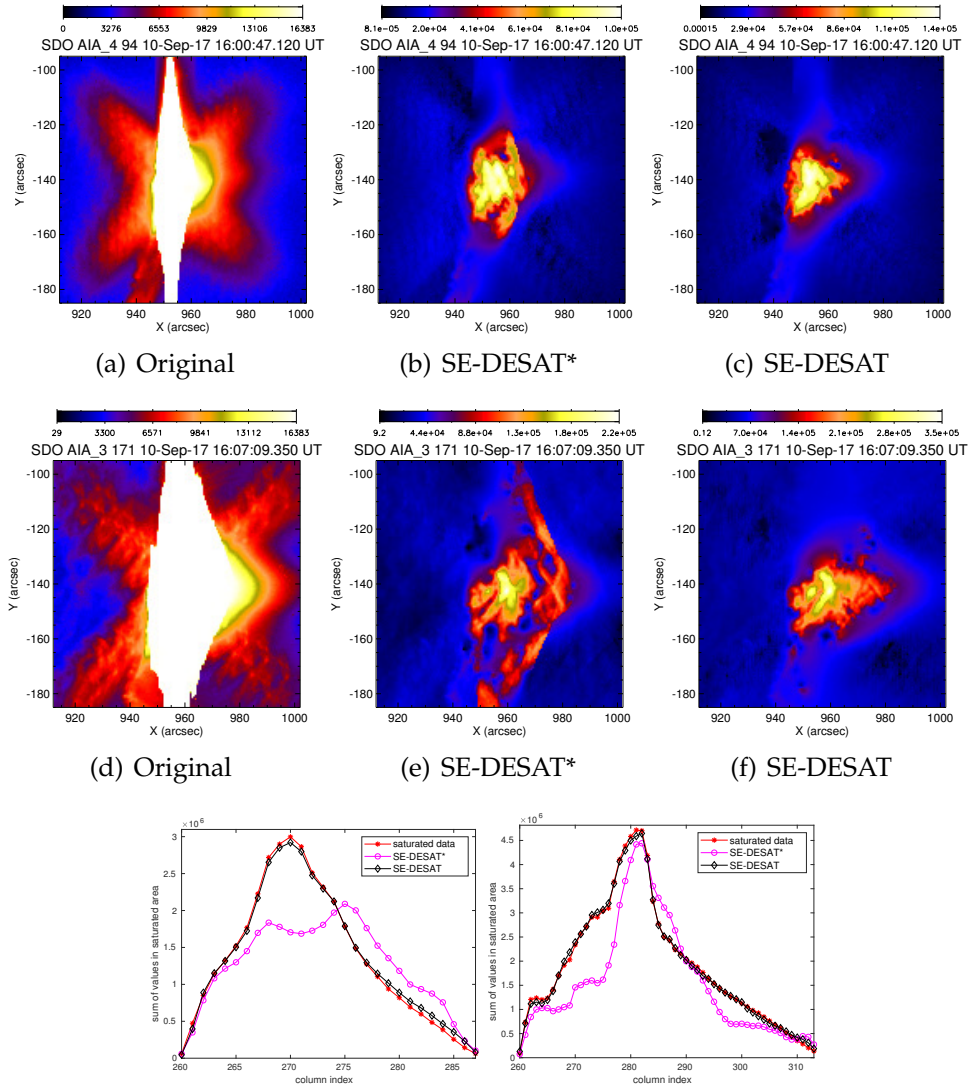


Figure 6.8: Real data: September 6, 2011 event. First row: saturated images. First column: image recorded at 22:19:25 UT at the 131 Å wavelength. Second column: image recorded at 22:19:09 UT at the 131 Å wavelength. Third column: image recorded at 22:16:43 UT at the 193 Å wavelength. Second row: reconstructions obtained by SE-DESAT* method. Third row: reconstructions obtained by SE-DESAT. Fourth row: reconstructions obtained by DESAT method.



(a) 22:19:25 UT bandwidth 131 Å (b) 22:19:09 UT bandwidth 131 Å (c) 22:16:43 UT bandwidth 193 Å

Figure 6.9: Real data: September 6, 2011 event. Comparison of the reconstructed flux profiles integrated along the saturated columns obtained by SE-DESAT*, SE-DESAT and DESAT methods with the real profiles. In the second (resp. third) panel the two (resp. three) plots correspond to the two (resp. three) connected components of the saturated region.



(a) Original (b) SE-DESAT* (c) SE-DESAT

(d) Original (e) SE-DESAT* (f) SE-DESAT

(g) 16:00:47 UT bandwidth 94 Å (h) 16:07:09 UT bandwidth 171 Å

Figure 6.10: Real data: September 10, 2017 event. First row from left to right: image recorded at 16:00:47 UT at the 94 Å wavelength, SE-DESAT* and SE-DESAT reconstructions. Second row from left to right: image recorded at 16:07:09 UT at the 171 Å wavelength, SE-DESAT* and SE-DESAT reconstructions. Third row: comparison of the reconstructed fluxes integrated along the saturated columns for images in the first row (left panel) and for images of the second row (right panel).

6.8.3 Solar storm on September 2017

During the super storm of September 6-10 2017, 27 M flares and four X flares were emitted by the Sun, which correspondingly emitted several powerful CMEs and bursts of high-energy protons. For more than one hour observation, all AIA filters suffered saturation in the core region of their images. In particular, at 171 Å all images presented significant primary saturation and blooming effects, which correspond to a consecutive deterioration of up to 300 EUV maps: even a rather sophisticated computational method such as DESAT is ineffective in this case, since the background estimation via interpolation of unsaturated emission is completely impossible. The desaturation power of SE-DESAT in the case of solar storms is illustrated in Figure 6.12 together with, in Figure 6.11, a first example of how these recovered EUV images can be used for basic scientific purposes. These figures refer to the most saturated energy channel (171 Å) in the batch of AIA wavelengths observing the flaring storm on September 10 2017. At this wavelength, around 300 images in the time range between 15:45:09 UT and 16:45:09 UT were dramatically corrupted by wide saturation stripes so that more than one hour observation of this intriguing event could not be fully exploited for scientific investigation. The first row (in the reverse direction of the page) of Figure 6.12 shows five consecutive images in the time range 16:05:45 UT - 16:06:33 UT; the blooming effects are clearly not distinguishable from the primary saturation region, while the diffraction fringes affect around half of the remaining field-of-view. These same fringes were given as input to the algorithm that produced the restored images represented in the second row and zoomed in the third one (in the reverse direction of the page) in Figure 6.12, where the core of the area is visible during its temporal evolution. The peak intensity in these cores is larger than 10^5 DN pixel^{-1} which is well above the saturation level of 16383 DN pixel^{-1}. Figure 6.11 shows that it is now possible to determine the photon flux at 171 Å over time in the primary saturation region identified thanks to the sparsity-enhancing property of the method. The C-statistic values in Table 6.4 describe the predictive power of the desaturated signal in the primary saturation region

when reproducing the experimental diffraction patterns. These numbers are the C-statistic values averaged over the diffraction pixels and corresponding to the desaturation of 50 highly saturated images in the 171 Å band: these values go down to 1 at the third iteration of the second step of the algorithm for most images and for all examples the goodness-of-fit is completely satisfactory after just 4 iterations of the alternate iterative scheme (all desaturated images presented in the current section correspond to the last iteration with C-statistic bigger than 1). We also applied the algorithm to the processing of images of the same event but in the 94 Å wavelength, where saturation and blooming effects are typically less persistent. The particular case considered in Figure 6.13 and Figure 6.14 corresponds to the highly deteriorated frame at 16:00:23 UT, preceded and followed by two mildly saturated frames at 16:00:14 UT and 16:00:38 UT, respectively. Figure 6.13 compares the original and desaturated images whereas Figure 6.14 contains the comparison of the integrated flux profiles, the C-statistic predicted pixel-wise by the desaturated signal patterns of the image at the acquisition time 16:00:23 UT and a comparison of the location and morphology of the three desaturated images.

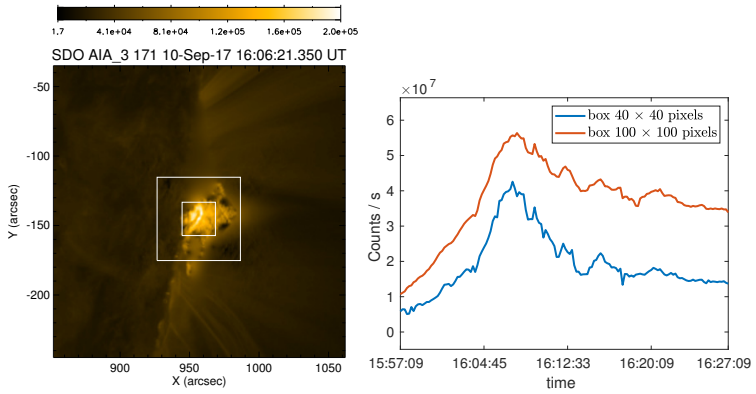


Figure 6.11: Bandwidth 171 Å for the September 10, 2017 event. Left panel: de-saturated image at 16:06:21 UT with highlighted the two boxes in which we computed the flux along the acquisition time from 15:57:09 UT to 16:27:09 UT. ‘ Right panel: reconstructed flux in the two boxes as a function of time (the inner box corresponds to the primary saturation region).

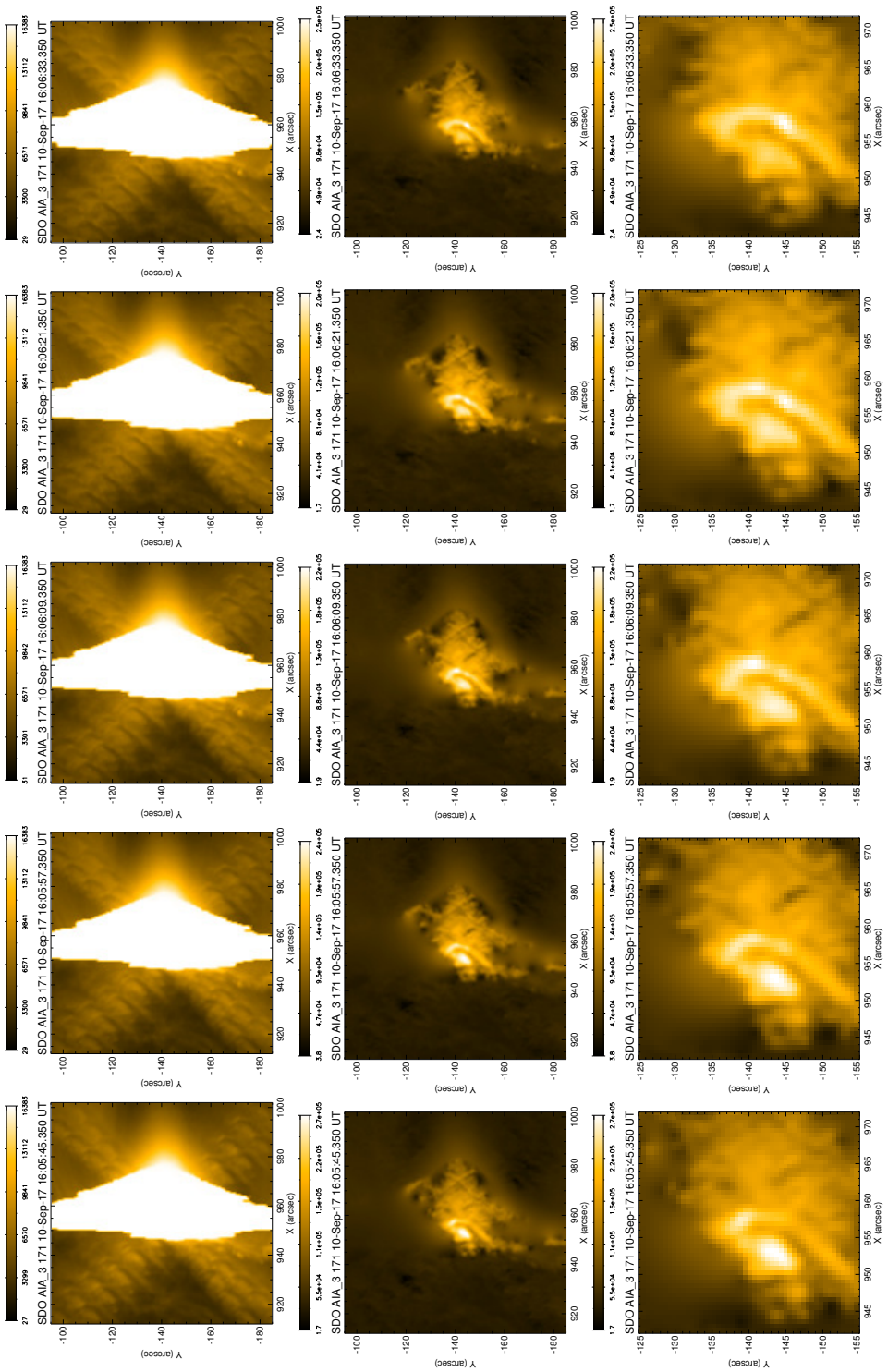
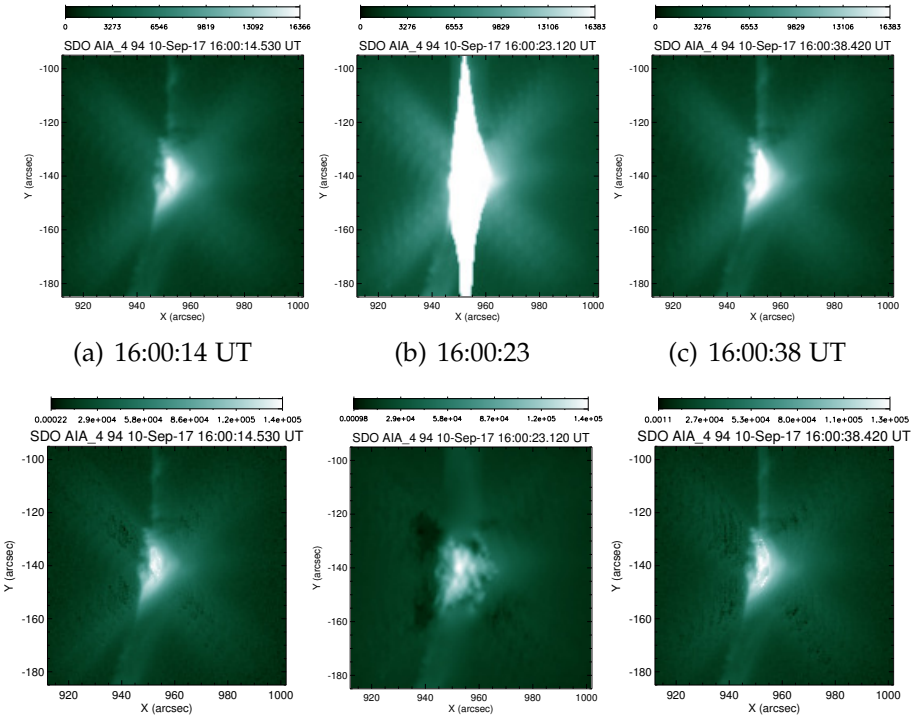


Figure 6.12: Bandwidth 171 Å for the September 10, 2017 event. First row: saturated images (from 16:05:45 to 16:06:33 UT). Second row: de-saturated images. Third row: zoom of the de-saturated images on the emission core.



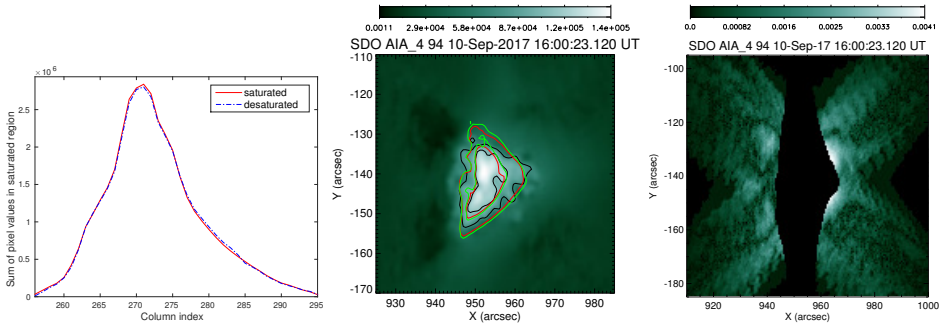
(a) 16:00:14 UT (b) 16:00:23 (c) 16:00:38 UT

(d) 16:00:14 UT (e) 16:00:23 UT (f) 16:00:38 UT

Figure 6.13: Bandwidth 94 Å for the September 10, 2017 event. First row, saturated images at 16:00:14 UT (left panel), 16:00:23 UT (middle panel), 16:00:38 UT (right panel). Second row: corresponding desaturated images provided by SE-DESAT.

Table 6.4: Performance of SE-DESAT method in C-statistic in the case of the September 10, 2017 solar storm, recorded at waveband 171 Å for different acquisition times. First and sixth column: recording time. From second to fifth column and from seventh to tenth column: C-statistic values at the first four iterations of the algorithm for the corresponding recording time.

	C-statistic					**C-statistic**			
time (UT)	iter 1	iter 2	iter 3	iter 4	time (UT)	iter 1	iter 2	iter 3	iter 4
16:04:09	2.96	0.86			16:09:33	4.82	1.45	1.02	0.72
16:04:21	2.68	1.08	0.72		16:09:45	5.19	1.08	0.97	
16:04:33	3.20	1.06	0.65		16:09:57	4.52	1.25	0.71	
16:04:45	2.89	1.39	0.68		16:10:09	4.24	1.18	0.97	
16:04:57	3.12	1.44	0.67		16:10:21	3.98	1.21	0.84	
16:05:09	3.39	1.48	0.66		16:10:33	4.38	1.19	0.76	
16:05:21	3.95	1.33	0.77		16:10:45	3.69	0.97		
16:05:33	3.69	1.49	0.80		16:10:57	3.82	0.92		
16:05:45	3.59	1.73	0.93		16:11:09	3.50	0.90		
16:05:57	3.62	1.81	0.92		16:11:21	3.88	1.00		
16:06:09	4.36	1.61	0.84		16:11:33	3.98	0.86		
16:06:21	4.72	1.78	0.90		16:11:45	3.82	0.90		
16:06:33	4.67	1.70	0.91		16:11:57	3.98	1.08	0.52	
16:06:45	4.55	1.81	0.91		16:12:09	4.22	1.01	0.57	
16:06:57	4.21	1.56	0.94		16:12:21	4.70	1.09	0.58	
16:07:09	4.11	1.72	0.94		16:12:33	4.56	1.13	0.65	
16:07:21	4.42	1.85	1.01	0.70	16:12:45	4.56	1.03	0.72	
16:07:33	4.65	1.67	0.92		16:12:57	4.38	0.98		
16:07:45	4.22	1.54	1.00	0.69	16:13:09	4.23	0.87		
16:07:57	4.42	1.77	1.03	0.71	16:13:21	4.18	0.85		
16:08:09	4.78	1.68	1.07	0.70	16:13:33	4.07	0.79		
16:08:21	4.41	1.57	1.03	0.69	16:13:45	3.73	0.71		
16:08:33	4.55	1.45	0.88		16:13:57	3.68	0.73		
16:08:45	4.30	1.29	0.87		16:14:09	4.06	0.81		
16:08:57	4.53	1.21	0.84		16:14:21	4.25	0.89		



(a) Integrated fluxes (b) Level curves (c) Pixel-wise C-statistic

Figure 6.14: Bandwidth 94 Å for the Septemper 10, 2017. Left panel: comparison of the integrated fluxes along the saturated image columns between the SE-DESAT reconstruction and the image data (at 16:00:23 UT). Middle panel: time evolution of the level curves of the three desaturated images in Figure 6.13 at 16:00:14 UT (red curves), 16:00:23 UT (black curves), and 16:00:38 UT (green curves). Right panel: pixel-wise C-statistic predicted in the diffraction fringes by the desaturated image at 16:00:23 UT in Figure 6.13.

6.9 Discussion

Why can we trust SE-DESAT results? When we use real data we have not at disposal a ground truth in order to compare results. In the case of the solar storm on September 2017, we can not compare results with DESAT method: as explained, DESAT cannot be applied since a reliable estimate of the background is not a priori available. Therefore, in this case, the reliability of SE-DESAT possibly relies on the very low values of the C-statistic in all of the almost 50 frames we investigated (see Table 6.4) and on the behavior over time of the flux in the primary saturation region (see Figure 6.11), which seems coherent with observations of previous similar events [77] and with simulation models [104]. Finally, results in Figure 6.13 and Figure 6.14 concerning the desaturation of three consecutive frames show a rather smooth evolution over time with respect to the restored emission morphology, peak location and overall photometry.

A brief comment regarding the inpainting procedure. In our approach the blooming pixel values, whose information is essentially lost, are estimated

making use of an inpainting procedure. However, such an inpainting procedure can be substituted with other approaches computationally less expensive and providing better results. A next step is the investigation of new approaches to fill blooming pixels: an idea is to use deep neural networks in order to estimate the unknown coefficients [26], i.e. the ones that can not be reconstructed by using our sparsity-enhancing approach.

What to do next is rather clear: thanks to this crucial desaturation step, all ingredients are now at disposal to design and implement an automatic pipeline for big data processing of AIA production, able to realize the whole stream of operations that from each recorded image leads to a reconstructed EUV map relieved by saturation, diffraction, and dispersion effects and therefore ready for a full exploitation within the framework of all possible physical models concerning flaring emission.

Bibliography

[1] J. Adler and O. Öktem, *Solving ill-posed inverse problems using iterative deep neural networks*, Inverse Problems, 33 (2017), p. 124007. 17

[2] V. Albani, P. Elbau, M. V. de Hoop, and O. Scherzer, *Optimal convergence rates results for linear inverse problems in Hilbert spaces*, Numerical functional analysis and optimization, 37 (2016), pp. 521–540. 6

[3] R. Andreev, P. Elbau, M. V. de Hoop, L. Qiu, and O. Scherzer, *Generalized convergence rates results for linear inverse problems in hilbert spaces*, Numerical Functional Analysis and Optimization, 36 (2015), pp. 549–566. 50

[4] F. J. Anscombe, *The Transformation of Poisson, Binomial and Negative-Binomial Data*, Biometrika, 35 (1948), pp. 246–254. 89

[5] A. Argyriou and F. Dinuzzo, *A unifying view of representer theorems*, in Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014, pp. 748–756. 47

[6] N. Aronszajn, *Theory of reproducing kernels*, Transactions of the American mathematical society, 68 (1950), pp. 337–404. 19, 20, 23

[7] M. J. Aschwanden, C. J. Schrijver, and D. Alexander, *Modeling of coronal euv loops observed with trace. i. hydrostatic solutions with nonuniform heating*, The Astrophysical Journal, 550 (2001), p. 1036. 158

[8] M. J. Aschwanden, A. Winebarger, D. Tsiklauri, and H. Peter, *The coronal heating paradox*, The Astrophysical Journal, 659 (2007), p. 1673. 158

[9] F. Bauer, S. Pereverzev, and L. Rosasco, *On regularization algorithms in learning theory*, Journal of complexity, 23 (2007), pp. 52–72. 6, 65

[10] A. Beck and M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202. 82, 83

[11] F. Benvenuto and M. Piana, *Regularization of multiplicative iterative algorithms with nonnegative constraint*, Inverse Problems, 30 (2014). 178

[12] F. Benvenuto, M. Piana, C. Campi, and A. M. Massone, *A hybrid supervised/unsupervised machine learning approach to solar flare prediction*, The Astrophysical Journal, 853 (2018), p. 90. 132

[13] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*, CRC press, 1998. 1, 12, 37, 106

[14] M. Bertero, C. De Mol, and E. R. Pike, *Linear inverse problems with discrete data. I. General formulation and singular system analysis*, Inverse problems, 1 (1985), p. 301. 1

[15] M. Bertero, H. Lanteri, and L. Zanni, *Iterative image reconstruction: a point of view*, (2008), pp. 37–63. 88

[16] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 2610–2636. 49, 50, 73

[17] G. Blanchard and N. Mücke, *Optimal rates for regularization of statistical inverse learning problems*, Foundations of Computational Mathematics, (2017), pp. 1–43. 6, 17, 18, 51, 56, 65, 78

[18] D. S. Bloomfield, P. A. Higgins, R. J. McAteer, and P. T. Gallagher, *Toward reliable benchmarking of solar flare forecasting methods*, The Astrophysical Journal Letters, 747 (2012), p. L41. 133

[19] M. G. Bobra and S. Couvidat, *Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm*, The Astrophysical Journal, 798 (2015), p. 135. 132

[20] M. Bogdan, E. van den Berg, C. Sabatti, W. Su, and E. J. Candès, *SLOPE-adaptive variable selection via convex optimization*, The annals of applied statistics, 9 (2015), p. 1103. 88

[21] S. Bonettini, F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero, *Gradient projection approaches for optimization problems in image deblurring and denoising*, in 2009 17th European Signal Processing Conference, Aug. 2009, pp. 1384–1388. 82

[22] O. Bousquet and A. Elisseeff, *Stability and generalization*, Journal of machine learning research, 2 (2002), pp. 499–526. 11

[23] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004. 115

[24] A. Braides, *A handbook of* Γ*-convergence*, in Handbook of Differential Equations: stationary partial differential equations, vol. 3, Elsevier, 2006, pp. 101–213. 39, 41, 43

[25] ——, *Local minimization, variational evolution and* Γ*-convergence*, Cours (Università di Pavia), 2013 (2012). 41

[26] T. A. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan, *Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography*, Inverse Problems, 35 (2019), p. 064002. 202

[27] C. Campi, F. Benvenuto, A. M. Massone, D. S. Bloomfield, M. K. Georgoulis, and M. Piana, *Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence*, The Astrophysical Journal, 883 (2019), p. 150. xii, 132, 139, 141, 142, 143, 144, 145

[28] E. J. Candès, M. B. Wakin, and S. P. Boyd, *Enhancing Sparsity by Reweighted $\ell 1$ Minimization*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 877–905. 88

[29] S. Canu, X. Mary, and A. Rakotomamonjy, *Functional learning through kernels*, arXiv preprint arXiv:0910.1013, (2009). 15

[30] A. Caponnetto and E. De Vito, *Optimal rates for the regularized least-squares algorithm*, Foundations of Computational Mathematics, 7 (2007), pp. 331–368. 6, 65

[31] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM review, 43 (2001), pp. 129–159. 84

[32] F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bulletin of the American mathematical society, 39 (2002), pp. 1–49. 12

[33] F. Cucker and D. X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge University Press, Mar. 2007. 17, 19

[34] G. Dal Maso, *An Introduction to* Γ*-Convergence*, Progress in Nonlinear Differential Equations and Their Applications 8, Birkhäuser Basel, 1 ed., 1993. 41

[35] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Communications on Pure and Applied Mathematics, 63 (2010), pp. 1–38. 89

[36] E. De Vito, A. Caponnetto, and L. Rosasco, *Model selection for regularized least-squares algorithm in learning theory*, Foundations of Computational Mathematics, 5 (2005), pp. 59–85. 11

[37] E. De Vito, L. Rosasco, and A. Caponnetto, *Discretization error analysis for tikhonov regularization*, Analysis and Applications, 4 (2006), pp. 81–99. 51

[38] E. De Vito, L. Rosasco, A. Caponnetto, U. De Giovannini, and F. Odone, *Learning from examples as an inverse problem*, Journal of Machine Learning Research, 6 (2005), pp. 883–904. 18, 93

[39] A. J. Dobson and A. Barnett, *An introduction to generalized linear models*, CRC press, 2008. 89

[40] R. M. Dudley, E. Giné, and J. Zinn, *Uniform and universal glivenko-cantelli classes*, Journal of Theoretical Probability, 4 (1991), pp. 485–510. 40

[41] M. A. Duval-Poo, M. Piana, and A. M. Massone, *Solar hard x-ray imaging by means of compressed sensing and finite isotropic wavelet transform*, Astronomy & Astrophysics, 615 (2018), p. A59. 162

[42] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, The Annals of statistics, 32 (2004), pp. 407–499. 84, 95

[43] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Springer, Dordrecht ; Boston, 1996 edition ed., July 1996. 1, 6, 12, 27, 36, 49, 63, 64

[44] D. Falconer, R. Moore, and G. Gary, *Magnetogram measures of total nonpotentiality for prediction of solar coronal mass ejections from active regions of any degree of magnetic complexity*, The Astrophysical Journal, 689 (2008), p. 1433. 128

[45] S. Felix, R. Bolzern, and M. Battaglia, *A compressed sensing-based image reconstruction algorithm for solar flare x-ray observations*, The Astrophysical Journal, 849 (2017), p. 10. 162

[46] M. A. T. Figueiredo and J. M. Bioucas-Dias, *Restoration of Poissonian Images Using Alternating Direction Optimization*, IEEE Transactions on Image Processing, 19 (2010), pp. 3133–3145. 82

[47] L. Fletcher and H. Hudson, *The magnetic structure and generation of euv flare ribbons*, Solar Physics, 204 (2001), pp. 69–89. 158

[48] K. Florios, I. Kontogiannis, S.-H. Park, J. A. Guerra, F. Benvenuto, D. S. Bloomfield, and M. K. Georgoulis, *Forecasting solar flares using magnetogram-based predictors and machine learning*, Solar Physics, 293 (2018), p. 28. 132, 136, 141, 153

[49] C. J. Flynn, C. M. Hurvich, and J. S. Simonoff, *Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models*, Journal of the American Statistical Association, 108 (2013), pp. 1031–1043. 99

[50] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, and others, *Pathwise coordinate optimization*, The Annals of Applied Statistics, 1 (2007), pp. 302–332. 82

[51] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1, Springer series in statistics Springer, Berlin, 2001. 12, 65, 87

[52] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent*, Journal of Statistical Software, 33 (2010). 82, 84, 99

[53] ——, *Regularization paths for generalized linear models via coordinate descent*, Journal of statistical software, 33 (2010), p. 1. 176

[54] D. Garcia, *Robust smoothing of gridded data in one and higher dimensions with missing values*, Computational statistics & data analysis, 54 (2010), pp. 1167–1178. 176

[55] J. E. Gentle, *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer Publishing Company, Incorporated, 2007. 89

[56] F. Girosi, M. Jones, and T. Poggio, *Regularization theory and neural networks architectures*, Neural computation, 7 (1995), pp. 219–269. 17

[57] E. Grave, G. R. Obozinski, and F. R. Bach, *Trace lasso: a trace norm regularization for correlated designs*, in Advances in Neural Information Processing Systems, 2011, pp. 2187–2195. 84

[58] C. W. Groetsch and C. W. Groetsch, *Inverse problems in the mathematical sciences*, vol. 52, Springer, 1993. 12

[59] R. Gu and A. Dogandžić, *Projected nesterov's proximal-gradient algorithm for sparse signal recovery*, IEEE Transactions on Signal Processing, 65, pp. 3510–3525. 82

[60] ——, *A fast proximal gradient algorithm for reconstructing nonnegative signals with sparse transform coefficients*, in Signals, Systems and Computers, 2014 48th Asilomar Conference on, IEEE, 2014, pp. 1662–1667. 82

[61] ——, *Projected nesterov's proximal-gradient signal recovery from compressive poisson measurements*, in 2015 49th Asilomar Conference on Signals, Systems and Computers, IEEE, 2015, pp. 1490–1495. 82

[62] S. Guastavino and F. Benvenuto, *A consistent and numerically efficient variable selection method for sparse poisson regression with applications to learning and signal recovery*, Statistics and Computing, (2018), pp. 1–16. 2, 17

[63] S. Guastavino, M. Piana, A. Massone, R. Schwartz, and F. Benvenuto, *Desaturating sdo/aia observations of solar flaring storms*, The Astrophysical Journal, 882 (2019), p. 12pp. 162

[64] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Dover Books on Science, Yale University Press, 1923. 10

[65] I. G. Hannah and E. P. Kontar, *Differential emission measures from the regularized inversion of hinode and sdo data*, Astronomy & Astrophysics, 539 (2012), p. A146. 158

[66] N. R. Hansen, P. Reynaud-Bouret, V. Rivoirard, and others, *Lasso and probabilistic inequalities for multivariate point processes*, Bernoulli, 21 (2015), pp. 83–143. 91

[67] Z. T. Harmany, R. F. Marcia, and R. M. Willett, *SPIRAL out of convexity: Sparsity-regularized algorithms for photon-limited imaging*, in IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2010, pp. 75330R–75330R. 82

[68] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations*, Chapman and Hall/CRC, 2015. 2, 3, 85

[69] J. Hershaw, C. Foullon, V. Nakariakov, and E. Verwichte, *Damped large amplitude transverse oscillations in an euv solar prominence, triggered by large-scale transient coronal waves*, Astronomy & Astrophysics, 531 (2011), p. A53. 158

[70] J. T. Hoeksema, Y. Liu, K. Hayashi, X. Sun, J. Schou, S. Couvidat, A. Norton, M. Bobra, R. Centeno, K. Leka, et al., *The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: overview and performance*, Solar Physics, 289 (2014), pp. 3483–3530. 128

[71] S. Ivanoff, F. Picard, and V. Rivoirard, *Adaptive Lasso and group-Lasso for functional Poisson regression*, Journal of Machine Learning Research, 17 (2016), pp. 1–46. 88, 98

[72] X. Jiang, P. Reynaud-Bouret, V. Rivoirard, L. Sansonnet, and R. Willett, *A data-dependent weighted LASSO under Poisson noise*, arXiv preprint arXiv:1509.08892, (2015). 88, 91

[73] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006. 37

[74] K. Knight, W. Fu, et al., *Asymptotics for lasso-type estimators*, The Annals of statistics, 28 (2000), pp. 1356–1378. 83

[75] M. A. Kon and L. A. Raphael, *Approximating functions in reproducing kernel hilbert spaces via statistical learning theory*, in Wavelets and Splines: Athens 2005, G. Chen and M.-J. Lai, Eds., Modern Methods in Mathematics, (2006), p. 271–286. 19

[76] M. Korsós and R. Erdélyi, *On the state of a solar active region before flares and cmes*, The Astrophysical Journal, 823 (2016), p. 153. 128

[77] E. Kraaikamp and C. Verbeeck, *Solar demon–an approach to detecting flares, dimmings, and euv waves on sdo/aia images*, Journal of Space Weather and Space Climate, 5 (2015), p. A18. 201

[78] R. Kress, *Linear integral equations*, vol. 82, Springer, 1989. 11, 18, 23

[79] V. Kůrková, *Neural Network Learning as an Inverse Problem*, Logic Journal of the IGPL, 13 (2005), pp. 551–559. 17

[80] J. R. Lemen, D. J. Akin, P. F. Boerner, C. Chou, J. F. Drake, D. W. Duncan, C. G. Edwards, F. M. Friedlaender, G. F. Heyman, N. E. Hurlburt, et al., *The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo)*, in The Solar Dynamics Observatory, Springer, 2011, pp. 17–40. 4, 160

[81] J. Lin and V. Cevher, *Optimal rates for spectral-regularized algorithms with least-squares regression over hilbert spaces*, arXiv preprint arXiv:1801.06720, (2018). 6

[82] C. Liu, N. Deng, J. T. Wang, and H. Wang, *Predicting solar flares using sdo/hmi vector magnetic data products and the random forest algorithm*, The Astrophysical Journal, 843 (2017), p. 104. 132

[83] W. Liu, N. V. Nitta, C. J. Schrijver, T. D. Tarbell, et al., *First sdo aia observations of a global coronal euv "wave": Multiple components and "ripples"*, The Astrophysical Journal Letters, 723 (2010), p. L53. 158

[84] L. Lo Gerfo, L. Rosasco, F. Odone, E. De Vito, and A. Verri, *Spectral algorithms for supervised learning*, Neural Computation, 20 (2008), pp. 1873–1897. 2, 17, 18

[85] J.-M. Loubes and V. Rivoirard, *Review of rates of convergence and regularity conditions for inverse problems*, International Journal of Tomography and Statistics, 11(S09) (2009), pp. 61–82. 56

[86] S. Loustau et al., *Inverse statistical learning*, Electronic Journal of Statistics, 7 (2013), pp. 2065–2097. 18

[87] B. A. Mair and F. H. Ruymgaart, *Statistical inverse estimation in hilbert scales*, SIAM Journal on Applied Mathematics, 56 (1996), pp. 1424–1444. 6, 50, 58

[88] I. C. Marschner and others, *glm2: fitting generalized linear models with convergence problems*, The R journal, 3 (2011), pp. 12–15. 105

[89] J. G. Martinez, R. J. Carroll, S. Müller, J. N. Sampson, and N. Chatterjee, *Empirical Performance of Cross-Validation With Oracle Methods in a Genomics Context*, The American Statistician, 65 (2011), pp. 223–228. 102

[90] P. Martinez, A. Klotz, and P. Léna, *A Practical Guide to CCD Astronomy*, Practical Astronomy Handbooks, Cambridge University Press, 1997. 160

[91] J. P. Mason, T. N. Woods, D. F. Webb, B. J. Thompson, R. C. Colaninno, and A. Vourlidas, *Relationship of euv irradiance coronal dimming slope and depth to coronal mass ejection speed and mass*, The Astrophysical Journal, 830 (2016), p. 20. 158

[92] M. T. McCann, K. H. Jin, and M. Unser, *Convolutional neural networks for inverse problems in imaging: A review*, IEEE Signal Processing Magazine, 34 (2017), pp. 85–95. 17

[93] P. McCullagh and J. A. Nelder, *Generalized Linear Models, Second Edition*, Aug. 1989. 86

[94] S. Mendelson and J. Neeman, *Regularization in kernel learning*, The Annals of Statistics, 38 (2010), pp. 526–565. 17

[95] J. Mercer, *Xvi. functions of positive and negative type, and their connection the theory of integral equations*, Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character, 209 (1909), pp. 415–446. 24

[96] C. A. Micchelli, Y. Xu, and H. Zhang, *Universal kernels*, Journal of Machine Learning Research, 7 (2006), pp. 2651–2667. 24

[97] H. Q. Minh, P. Niyogi, and Y. Yao, *Mercer's theorem, feature maps, and smoothing*, in International Conference on Computational Learning Theory, Springer, 2006, pp. 154–168. 24

[98] V. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag New York, 1984. 1

[99] S. Mukherjee, P. Niyogi, T. Poggio, and R. Rifkin, *Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization*, Advances in Computational Mathematics, 25 (2006), pp. 161–193. 40

[100] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, and M. Ishii, *Deep flare net (defn) model for solar flare prediction*, The Astrophysical Journal, 858 (2018), p. 113. 153

[101] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, S. Watari, and M. Ishii, *Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms*, The Astrophysical Journal, 835 (2017), p. 156. 132

[102] V. I. Norkin and M. A. Keyzer, *On convergence of kernel learning estimators*, SIAM Journal on Optimization, 20 (2009), pp. 1205–1223. 6, 39

[103] G. Obozinski, B. Taskar, and M. Jordan, *Multi-task feature selection*, (2006). 84, 152

[104] B. O'dwyer, G. Del Zanna, H. Mason, M. Weber, and D. Tripathi, *Sdo/aia response to coronal hole, quiet sun, active region, and flare plasma*, Astronomy & Astrophysics, 521 (2010), p. A21. 201

[105] M. R. Osborne, B. Presnell, and B. A. Turlach, *On the lasso and its dual*, Journal of Computational and Graphical statistics, 9 (2000), pp. 319–337. 84

[106] Y. S. Paolo Grigis and M. Weber, *Aia psf characterization and image deconvolution*, SDO Documentation, (2012). 165

[107] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*, vol. 152, Cambridge University Press, 2016. 19

[108] W. D. Pesnell, B. J. Thompson, and P. Chamberlin, *The solar dynamics observatory (sdo)*, in The Solar Dynamics Observatory, Springer, 2011, pp. 3–15. 124

[109] G. Peyré, *The Numerical Tours of Signal Processing*, Computing in Science & Engineering, 13 (2011), pp. 94–97. 106

[110] M. Piana, C. Campi, F. Benvenuto, S. Guastavino, and A. Massone, *Flare forecasting and feature ranking using sdo/hmi data*, IL NUOVO CIMENTO C, 42 (2019). 132

[111] B. Poduval, C. DeForest, J. Schmelz, and S. Pathak, *Point-spread functions for the extreme-ultraviolet channels of sdo/aia telescopes*, The Astrophysical Journal, 765 (2013), p. 144. 164

[112] J. L. Prince and J. M. Links, *Medical imaging signals and systems*, Pearson Prentice Hall Upper Saddle River, New Jersey, 2006. 87

[113] J. Qiu, H. Wang, C. Cheng, and D. E. Gary, *Magnetic reconnection and mass acceleration in flare-coronal mass ejection events*, The Astrophysical Journal, 604 (2004), p. 900. 158

[114] A. Rastogi and S. Sampath, *Optimal rates for the regularized learning algorithms under general source condition*, Frontiers in Applied Mathematics and Statistics, 3 (2017), p. 3. 6, 18

[115] L. Rosasco, M. Belkin, and E. D. Vito, *On learning with integral operators*, Journal of Machine Learning Research, 11 (2010), pp. 905–934. 65

[116] L. Rosasco, E. D. Vito, A. Caponnetto, M. Piana, and A. Verri, *Are loss functions all the same?*, Neural Computation, 16 (2004), pp. 1063–1076. 31

[117] P. H. Scherrer, J. Schou, R. Bush, A. Kosovichev, R. Bogart, J. Hoeksema, Y. Liu, T. Duvall, J. Zhao, C. Schrijver, et al., *The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo)*, Solar Physics, 275 (2012), pp. 207–227. 124, 128

[118] B. Schölkopf, R. Herbrich, and A. Smola, *A generalized representer theorem*, in Computational learning theory, Springer, 2001, pp. 416–426. 30, 44

[119] C. J. Schrijver, *A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting*, The Astrophysical Journal Letters, 655 (2007), p. L117. 128

[120] R. Schwartz, G. Torre, A. Massone, and M. Piana, *Desat: A solar software tool for image de-saturation in the atmospheric image assembly onboard the solar dynamics observatory*, Astronomy and Computing, 13 (2015), pp. 117–123. 159, 169, 172

[121] R. Schwartz, G. Torre, and M. Piana, *Systematic de-saturation of images from the atmospheric imaging assembly in the solar dynamics observatory*, The Astrophysical Journal Letters, 793 (2014), p. L23. 159, 161, 172, 178

[122] R. Schwenn, *Space weather: The solar perspective*, Living reviews in solar physics, 3 (2006), p. 2. 125

[123] F. Sciacchitano, S. Lugaro, and A. Sorrentino, *Sparse bayesian imaging of solar flares*, SIAM Journal on Imaging Sciences, 12 (2019), pp. 319–343. 162

[124] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE transactions on medical imaging, 1 (1982), pp. 113–122. 173

[125] J. Silva and S. Tenreyro, *Poisson: some convergence issues*, Stata journal, 11 (2011), pp. 207–212. 105

[126] S. Smale and D.-X. Zhou, *Shannon sampling and function reconstruction from point values*, Bulletin of the American Mathematical Society, 41 (2004), pp. 279–305. 17, 30

[127] ——, *Learning theory estimates via integral operators and their approximations*, Constructive approximation, 26 (2007), pp. 153–172. 6

[128] A. J. Smola, B. Schölkopf, and K.-R. Müller, *The connection between regularization operators and support vector kernels*, Neural networks, 11 (1998), pp. 637–649. 17

[129] A. Staglianò, P. Boccacci, and M. Bertero, *Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle*, Inverse Problems, 27 (2011), p. 125003. 2, 82

[130] J.-L. Starck and F. Murtagh, *Astronomical image and data analysis*, Springer Science & Business Media, 2007. 87

[131] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*, Cambridge university press, 2010. 2

[132] I. Steinwart and A. Christmann, *Support vector machines*, Springer Science & Business Media, 2008. 23

[133] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*, SIAM, 2005. 37

[134] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological), (1996), pp. 267–288. 3, 82, 83

[135] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, *Sparsity and smoothness via the fused lasso*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2005), pp. 91–108. 84, 87

[136] A. Tikhonov, A. Goncharsky, V. Stepanov, and A. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, vol. 328, Springer Science Business Media, 1995. 1

[137] G. Torre, R. A. Schwartz, F. Benvenuto, A. M. Massone, and M. Piana, *Inverse diffraction for the atmospheric imaging assembly in the solar dynamics observatory*, Inverse Problems, 31 (2015), p. 095006. 8, 159, 161, 172, 187, 188, 192

[138] A. van der Vaart and J. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Science & Business Media, 1996. 39

[139] A. C. van Rooij and F. H. Ruymgaart, *Asymptotic minimax rates for abstract linear estimators*, Journal of Statistical Planning and Inference, 53 (1996), pp. 389–402. 6, 50, 58

[140] V. Vapnik, *The nature of statistical learning theory*, Springer science & business media, 2013. 12, 36, 39

[141] A. M. Vásquez, R. A. Frazin, and F. Kamalabadi, *3d temperatures and densities of the solar corona via multi-spacecraft euv tomography: analysis of prominence cavities*, Solar Physics, 256 (2009), pp. 73–85. 158

[142] C. R. Vogel, *Computational methods for inverse problems*, SIAM, 2002. 37

[143] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM Journal on Numerical Analysis, 14 (1977), pp. 651–667. 1

[144] C. Wang and D.-X. Zhou, *Optimal learning rates for least squares regularized regression with unbounded sampling*, Journal of Complexity, 27 (2011), pp. 55–67. 6, 17

[145] G. Wang, D. Garcia, Y. Liu, R. De Jeu, and A. J. Dolman, *A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations*, Environmental Modelling & Software, 30 (2012), pp. 139–142. 176

[146] T. N. Woods, R. Hock, F. Eparvier, A. R. Jones, P. C. Chamberlin, J. A. Klimchuk, L. Didkovsky, D. Judge, J. Mariska, H. Warren, et al., *New solar extreme-ultraviolet irradiance observations during flares*, The Astrophysical Journal, 739 (2011), p. 59. 158

[147] T. T. Wu and K. Lange, *Coordinate Descent Algorithms for Lasso Penalized Regression*, The Annals of Applied Statistics, 2 (2008), pp. 224–244. 84

[148] Y. Yao, L. Rosasco, and A. Caponnetto, *On early stopping in gradient descent learning*, Constructive Approximation, 26 (2007), pp. 289–315. 2, 6, 17

[149] T. Yokoyama, K. Akita, T. Morimoto, K. Inoue, and J. Newmark, *Clear evidence of reconnection inflow of a solar flare*, The Astrophysical Journal Letters, 546 (2001), p. L69. 158

[150] Y. Yu, H. Cheng, D. Schuurmans, and C. Szepesvári, *Characterizing the representer theorem*, in Proceedings of the 30th International Conference on Machine Learning (ICML-13), 2013, pp. 570–578. 47

[151] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68 (2006), pp. 49–67. 84

[152] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, *Efficient gradient projection methods for edge-preserving removal of poisson noise*, Inverse Problems, 25 (2009), p. 045010. 3

[153] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, *Corrigendum: efficient gradient projection methods for edge-preserving removal of poisson noise*, Inverse Problems, 29 (2013), p. 119501. 109

[154] Q. Zhang, P. Chen, Y. Guo, C. Fang, and M. Ding, *Two types of magnetic reconnection in coronal bright points and the corresponding magnetic configuration*, The Astrophysical Journal, 746 (2012), p. 19. 158

[155] P. Zhao and B. Yu, *On Model Selection Consistency of Lasso*, J. Mach. Learn. Res., 7 (2006), pp. 2541–2563. 82, 83, 85

[156] H. Zou, *The adaptive lasso and its oracle properties*, Journal of the American statistical association, 101 (2006), pp. 1418–1429. 3, 82, 83, 85, 86, 88, 94, 96, 98, 132, 140

[157] H. Zou and H. H. Zhang, *On the adaptive elastic-net with a diverging number of parameters*, Annals of statistics, 37 (2009), p. 1733. 84, 91, 92, 114