Weierstraß-Institut für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint

ISSN 2198-5855

Data-driven methods for quantitative imaging

Guozhi Dong¹, Moritz Flaschel², Michael Hintermüller²,

Kostas Papafitsoros³, Clemens Sirotenko², Karsten Tabelow²

submitted: April 26, 2024

School of Mathematics and Statistics
HNP-LAMA
Central South University
Lushan South Road 932
410083 Changsha
China
E-Mail: guozhi.dong@csu.edu.cn

 ² Weierstrass Institute Mohrenstrasse 39 10117 Berlin Germany
E-Mail: moritz.flaschel@wias-berlin.de michael.hintermueller@wias-berlin.de clemens.sirotenko@wias-berlin.de karsten.tabelow@wias-berlin.de

³ School of Mathematical Sciences Queen Mary University of London Mile End Road E1 4NS UK E-Mail: k.papafitsoros@qmul.ac.uk

> No. 3105 Berlin 2024



2020 Mathematics Subject Classification. 35R30, 49J52, 49K20, 49M41, 62G05, 62G99, 62P10, 68T07, 90C46.

Key words and phrases. Quantitative MRI, quantitative image reconstruction, regularization, variational methods, machine learning, neural networks, learning-informed physics.

Edited by Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS) Leibniz-Institut im Forschungsverbund Berlin e. V. Mohrenstraße 39 10117 Berlin Germany

Fax:+493020372-303E-Mail:preprint@wias-berlin.deWorld Wide Web:http://www.wias-berlin.de/

Data-driven methods for quantitative imaging

Guozhi Dong, Moritz Flaschel, Michael Hintermüller, Kostas Papafitsoros, Clemens Sirotenko, Karsten Tabelow

ABSTRACT. In the field of quantitative imaging, the image information at a pixel or voxel in an underlying domain entails crucial information about the imaged matter. This is particularly important in medical imaging applications, such as quantitative Magnetic Resonance Imaging (qMRI), where quantitative maps of biophysical parameters can characterize the imaged tissue and thus lead to more accurate diagnoses. Such quantitative values can also be useful in subsequent, automatized classification tasks in order to discriminate normal from abnormal tissue, for instance. The accurate reconstruction of these quantitative maps is typically achieved by solving two coupled inverse problems which involve a (forward) measurement operator, typically ill-posed, and a physical process that links the wanted quantitative parameters to the reconstructed qualitative image, given some underlying measurement data. In this review, by considering qMRI as a prototypical application, we provide a mathematically-oriented overview on how data-driven approaches can be employed in these inverse problems eventually improving the reconstruction of the associated quantitative maps.

1. INTRODUCTION

The combined and sophisticated use of large imaging data and deep neural networks towards improving image processing and reconstruction algorithms is nowadays ubiquitous. Essentially most of the current state-of-the-art imaging tasks incorporate some (machine) learning component that provides a priori information or structure to the final image reconstruction. This a priori information is necessary as imaging problems are often ill-posed and available data are typically degraded and incomplete. A typical example is Magnetic Resonance Image (MRI) reconstruction in which the data are formed from noisy subsampled Fourier coefficients of the contrasting tissue image [193]. In the classical setting of variational, i.e. minimization based image reconstruction, prior information is included via a regularization functional. Often the latter represents a model for imposing sparsity of certain image related features, while suppressing adverse noise. An example for such a functional is the total variation (TV) regularization [165, 37, 131] which promotes sparsity of discontinuities (often referred to as "edges") of reconstructed image intensities while featuring piecewise constant reconstructions. The latter is often an unwanted side-effect of the TV-regularization leading to the development of generalized versions of TV; see, e.g., [27, 114]. Rather than defining regularizers "manually" on a case-by-case basis, in modern data-driven regimes the regularization functional is merely *learned* automatically from reference data. Inspired by an underlying class of image features (contained also in the training data), this typically allows to impose a more sophisticated structure to the final image. For such a technique, several approaches are conceivable. Here we mention only two. Indeed, for once one may "train" the regularizer by using only the given datum to adapt the reconstruction to its paramount features in the spirit of unsupervised learning or one uses large datasets of degraded ground truth data (paired or unpaired) in order to learn the desired image distribution and impose this as prior information [9, 101, 112, 140, 168]. These techniques have been intensively developed concerning qualitative imaging. Regarding the latter, the reconstruction values at pixels do not themselves entail any crucial information (e.g. any diagnostic value), rather the quality of a reconstruction is judged according to its overall image structure and visual quality. This coins the term structure-informed learning in imaging for this category of reconstruction methods.

Prior information to image reconstruction can also be imposed by taking advantage of the physical processes governing the data acquisition. Mathematically, the acquisition physics can often be expressed in terms of (partial (P) or ordinary (O)) differential equations (DE) which then govern the reconstruction process. Considering such DE constraints is particularly relevant in *quantitative imaging* where one is interested in the numerical values of biophysical parameters (in physical units) associated with the imaged object as these values provide objective characteristics of the imaged matter or tissue. Among others, quantitative MRI (qMRI) is a prominent example in this imaging category. Indeed, in qMRI the relevant biophysical parameters are the (longitudinal and transversal) relaxation times (T_1 and T_2) and the proton density (ρ); see Subsection 2.2 below for an explanation of these parameters. Specific values of these parameters are related to specific tissue types. Modelling the time evolution of tissue magnetization, given the aforementioned biophysical parameters, leads to the Bloch equations, a system of ODEs associated with every voxel of an imaged domain [21, 193]. Often, however, such a differential equation model is an idealized representation or a simplification of a far more complex underlying physical process. Sometimes the latter may be even unknown in detail or too complicated to be modelled precisely (from first principles). Under such circumstances, experimental data can provide insights into the underlying physics and may help to (approximately) learn a mathematical model of the involved physical processes. For the latter, nowadays one benefits from the remarkable versatility and approximation properties of deep neural networks (DNNs) [79]. Once models of the underlying physical laws have been learned, then these data-driven models are embedded into the image reconstruction process providing further information [59]. For this overall category of methods, we use the term *learning-informed physics in imaging* throughout this review.

Notably as far as imaging applications are concerned, learning-informed physics approaches are significantly less studied than structure-informed learning techniques. In general, the aim of this review paper is to provide a unified presentation of both categories, with a special focus on recent developments of the former one. For the sake of illustration we use (q)MRI as a prototype application.

The remainder of this paper is organized as follows: In Section 2, we describe model-based methods with an emphasis on qualitative and quantitative MRI reconstruction. We recall the classical variational regularization methods for inverse problems for the qualitative MRI problem and physics-integrated methods (e.g. Magnetic Resonance Fingerprinting and related extensions), but still model-based, for the quantitative setting. We proceed with data-driven approaches in Section 3 for both qualitative and quantitative MRI, starting from dictionary approaches to methods incorporating deep neural networks. We also highlight certain statistics-based post-processing techniques for qualitative MR images towards a subsequent improved reconstruction of the quantitative maps. In the same section we also collect recent developments in learning-inform physics in quantitative imaging, with a focus on learning the Bloch solution map for qMRI. The review ends by a section on FAIR data management in mathematical image processing in Section 4.

2. MODEL-BASED METHODS

In order to set the stage for our main focus on quantitative reconstruction, we start this review by discussing classical model-based methods for qualitative imaging. In particular, we collect here variational methods with a special focus on popular regularization strategies and solution algorithms. In this way, we believe that the similarities and (yet) differences between qualitative and quantitative approaches become apparent. Regarding quantitative methods, in Subsection 2.2 we shall see that typical quantitative imaging problems admit a variational formulation with an objective inspired by a qualitative approach but subject to certain physics-based constraints which involve (as unknowns) the quantitative parameters of interest. Let us mention that qualitative methods are by now rather classical and there already exists a plethora of reviews, see, e.g., [18, 26, 30, 39, 35, 101, 173]. Hence, we can be brief here.

When considering the quantities of interest (such as image intensities or physical parameters) as functions over an image domain, rather than vector, matrix or tensor quantities with finitely many entries, respectively, relating to, e.g, a pixel grid which represents the (discrete and finite dimensional) image domain, many of the problems presented here can be studied in an infinite dimensional function space setting. Such a framework raises numerous interesting and challenging mathematical, algorithmic, and numerical questions. However, for the sake of exposition and unification of presentation we formulate problems in a discrete, i.e., finite dimensional, setting (e.g., on a grid of finitely many pixels/voxels) and make specific remarks regarding their function space analogues.

2.1. Variational methods for imaging and qualitative MRI. The classical problem in imaging is to compute an approximation u of an underlying ground truth image u_{true} , where $u, u_{true} \in X$ with X denoting a set of functions mapping from some Euclidean space to \mathbb{R}^m . When m = 1, often u(x) models the image intensity at a point x. In the discrete setting for u this corresponds to the intensity at a specific pixel (in a 2D image domain). In order to make the distinction from quantitative imaging, we already stress here that in the qualitative regime, the precise values of u, whose range is often dictated by some a priorly chosen scale such as, e.g., $u(x) \in [0, 1]$, are not of paramount importance. Rather one is merely interested in salient features of the image, like structure, geometry, contrast, or in any other information that can be used, for instance in diagnostic procedures upon direct visual inspection, e.g., for detecting tumours. Typically, u_{true} is not directly accessible, but rather through available (measured) data y. A common model for y is

$$(2.1) y = Au_{true} + \eta$$

Here $A : X \to Y$ is the forward operator which models some degradation (e.g. by convolution, incompleteness, etc.) or it reflects an image transformation from the image space X into a data space Y. For instance in MRI, it denotes a subsampling P of the Fourier coefficients of the original image u_{true} [131], i.e., $A := P\mathcal{F}$ with \mathcal{F} the Fourier transform. In contrast, in computerized tomography (CT) it denotes the Radon transform [142], which collects (typically a limited number of) integrals along lines reflecting X-ray attenuations by different tissues in the body. The variable η in (2.1) represents a *noise* component and it is modelled as a highly oscillatory function in Y (with zero mean). Noise is often the result of measurement errors, random degradation processes or transmission losses. Note also that while we consider an additive noise term in (2.1), noise may also occur in the measurement process in a nonlinear way. This may give rise to, e.g., multiplicative noise (see [80]) which then requires an according adaptation of (2.1). Unfortunately, the direct inversion of A and, thus, the exact recovery of u_{true} from y is hampered by the presence of noise, but also by the fact that A is often singular or ill-conditioned. These facts render the reconstruction problem an ill-posed inverse problem. A popular way to overcome the latter is by imposing regularization on the reconstruction process via utilizing prior information on u. Following the variational approach to inverse problems, this leads to solving the minimization problem

(2.2)
$$\min_{u \in X} \mathcal{J}(u) := \mathcal{D}(Au, y) + \mathcal{R}(\alpha; u).$$

Above, \mathcal{D} represents a data discrepancy or fidelity term. It is designed to ensure that the reconstruction, i.e., the minimizer of the overall energy \mathcal{J} , is close to the data y in a suitable sense. In fact, its proper choice is often mainly dictated by the statistics of the noise, with $\mathcal{D}(Au, y) = \frac{1}{2} ||Au - y||_2^2$ being the appropriate form for Gaussian-type noise. In contrast, when salt-and-pepper or random-valued impulse noise is present, respectively, then the use of the ℓ^1 norm is preferable, i.e., $\mathcal{D}(Au, y) = ||Au - y||_1$, and the Kullback-Leibler divergence [118] is the appropriate choice for Poisson noise. The latter is characteristic for tomographic problems, such as CT or Positron Emission Tomography (PET). For the sake of exposition and also motivated by our example modality of MRI where the noise can be considered of Gaussian type, in what follows we will use an ℓ^2 fidelity term. The quantity \mathcal{R} denotes the regularization term or prior with respect to the quantity of interest u, and $\alpha > 0$ represents a scalar or distributed regularization weight. The main purpose of \mathcal{R} is to counteract ill-conditioning (due to properties of A) and to filter noise (due to the presence of η in (2.1)).



FIGURE 1. Qualitative MRI reconstruction example for small (left) and large (right) scalar regularization parameter α . Here the data come from brain phantom data (middle) from the Brain Web Simulated Brain Database Data [24, 48, 119].

In general, the literature on regularization functionals is rather extensive. It went through a significant development especially after the introduction of the Total Variation (TV) regularization in the early 1990s [165]. Its discrete version corresponds to the ℓ_1 -norm of the gradient of the image intensity, balanced by a positive regularization parameter $\alpha > 0$ against data fidelity:

(2.3)
$$\mathrm{TV}(u) = \alpha \|\nabla u\|_1.$$

TV-regularization based reconstructions typically exhibit a piecewise constant structure, also referred to as *staircasing effect*. The latter is often undesirable. Consequently, higher-order extensions of the total variation have been considered as a remedy. The most prominent choice of such higher order regularizers, is perhaps the Total Generalized Variation (TGV) of order 2 [27], whose discrete version reads

(2.4)
$$\operatorname{TGV}(u) = \min_{w} \alpha \|\nabla u - w\|_{1} + \beta \|\mathcal{E}w\|_{1},$$

where $\alpha, \beta > 0$ are regularization parameters and \mathcal{E} denotes the symmetrized gradient $(1/2)(\nabla w + \nabla w^{\top})$. It promotes piecewise affine reconstructions, which are often visually more appealing than piecewise constant ones.

Remark 2.1 (Function space versions of classical regularizers). In the function space setting, TV is defined as the total variation of the finite Radon measure Du that represents the distributional derivative of a function $u \in L^1(\Omega)$. When indeed the derivative of u admits such a representation, then we say that $u \in BV(\Omega)$, the space of functions of bounded variation [73, 6]. Similar are the corresponding definitions for TGV. The study of TV, TGV and other derivative-based regularizers in their respective function space formulation has also received considerable attention in imaging. This is due to their useful mathematical properties that can provide information about the structure imposed on images, such as the preservation of edges (discontinuities in u) and the promotion of piecewise constant/affine structures [138, 34, 93, 173, 162, 149].

Very large values of regularization parameters typically result in over-regularization leading to oversmoothed reconstructions. While this might benefit the reconstruction of large homogeneous areas in images (where $\nabla u \approx 0$) it can also result in cartoon-like reconstructions with a significant loss of details. On the other hand, very small values of regularization parameters generally preserve details but may also yield reconstructions with noticeable noise and artifacts due to the underlying ill-posedness; compare Figure 1 for a representative example in (qualitative) MRI reconstruction. This suggests a local adaptation of the regularization strength which can be achieved by a *spatially varying regularization parameter / weight*. Here the parameter α (and also β in TGV) is spatially varying (pixel-dependent) rather than a scalar only. Thus, one has $\alpha = \alpha(x)$, and similarly for β . Then the corresponding regularization functionals, i.e., *weighted* TV and TGV, are defined in their discrete version respectively as follows:

(2.5)
$$\mathrm{TV}(u;\alpha) = \|\alpha \nabla u\|_1,$$

(2.6)
$$\operatorname{TGV}(u;\alpha,\beta) = \min_{w} \|\alpha(\nabla u - w)\|_{1} + \|\beta \mathcal{E}w\|_{1}.$$

Clearly, the automatic determination of the spatially varying parameters α and β along with u represents a formidable challenge: (i) One has to identify criteria for the quality of a choice of (α, β) . Such criteria will depend on u and the noise statistics. (ii) A mathematical model for jointly finding u and (α, β) has to be developed. The formulation needs to accomodate the identification of several quantities which enter the problem in different ways. (iii) An efficient algorithmic treatment has to be developed to keep the overall computational burden acceptable. Indeed, one has to cope with a non-convex and also non-smooth optimization task.

It has turned out that *primal-dual equilibrium formulations* are particularly suitable for these tasks. In [61] for TV-regularization and Gaussian noise an augmented Lagrangian type update scheme for a fidelity (*not* regularization) weight was proposed, while simultaneously reconstructing u. It utilizes Gumbel statistics, i.e., statistics of the extremes, to develop a local quality measure for the associated fidelity weight. Indeed, the spatially varying weight can be interpreted as a Lagrangian multiplier for a localized version of the image residual. Then the multiplier gets updated as long as the localized image residual does not satisfy a local variance bound; otherwise the fidelity weight remains fixed. The choice of the variance bound is delicate and utilizes the statistics of the extremes. This approach was later extended to ℓ^1 -TV reconstruction [98], the reconstruction of color images [62], and to TGV [25]. Another very suitable approach to simultaneously identifying u and (α, β) relies on *bilevel optimization* which we briefly describe next, but we also refer the interested reader to the reviews [101, 31, 54]. The bilevel formulation for computing optimal spatially varying regularization parameters reads

(2.7)
$$\begin{cases} \min_{u,\alpha} \mathcal{L}(u,\alpha) \\ \text{subject to} \quad u \in \operatorname*{arg\,min}_{u \in X} \|Au - y\|_2^2 + \mathcal{R}(\alpha;u), \end{cases}$$

where α denotes all relevant regularization parameters, such as for instance α in (2.5) or (α, β) in (2.6). Furthermore, \mathcal{L} denotes an upper level objective which is to be minimized over both the image u and α under the constraint that u is the solution of the variational problem with regularization parameter α . Borrowing terminology from mathematical game theory, one may consider α the "leader" and $u = u(\alpha)$ the "follower", that accepts any choice of α issued by the leader and reacts optimally by minimizing the objective in the constraints of (2.7). Consequently, as in the above primal-dual equilibrium setting \mathcal{L} has to implement criteria for measuring the quality of $u(\alpha)$. For instance, \mathcal{L} may equal the distance of $u(\alpha)$ to the ground truth u_{true} , i.e. $||u(\alpha) - u_{true}||_2^2$. In practice, this approach typically involves a training set of true data in the spirit of supervised learning [31, 54] which, however, makes the approach susceptible to overfitting [55]. This can be remedied by developing ground-truthfree upper level objectives like the one introduced in [96, 97] for TV, which is inspired by a dual version of the localized residual estimators of [61]. Later this approach was also used for TGV and general convex regularizers [102, 147]. As in [61], there are statistics-based objectives aiming to enforce a localized version of the residuals Au - y that allows the computation of spatially varying regularization parameters without using any ground truths. In Figure 2 we depict a TGV-based qualitative MRI reconstruction where the regularization parameter α of TGV is spatially varying and computed by the ground-truth-free bilevel approach in [102, 177].



FIGURE 2. Qualitative MRI reconstruction (left) using TGV with a spatial varying regularization parameter α (right), computed via a bilevel optimization scheme. Observe how the weight attains small values in the bottom right part of the image, where most of the fine scale details are present, leading to their preservation.

Remark 2.2 (Function space versions of bilevel problems in imaging). Bilevel problems in imaging of the type (2.7) have also been extensively studied in infinite dimensions; see all of the aforementioned references for bilevel approaches. Here additional challenges arise. For instance, one must ensure that the pairing of the (now function) α with the gradient of u (now measure Du) e.g. $\int \alpha(x) d|Du|$ is well defined. The latter is true for instance when α is continuous [96, 147], see [53] for an investigation concerning even weaker requirements. Additional complications arise in the case of vanishing weights [93, 15], which is typically avoided by constraining the corresponding minimization variable in (2.7) to be bounded away from zero. Further challenges exist when devising numerical solution algorithms for (2.7). In general, these problems fall into the realm of mathematical programs with equilibrium constraints (MPECs); see, e.g., [130, 146] for rather general accounts in finite dimensions and [95] for a comprehensive discussion of MPEC stationarity for a specific problem class in infinite dimensions. Typically, MPECs suffer from constraint degeneracy. As a consequence, for the derivation of stationarity conditions it requires advanced non-smooth analysis techniques other than the classical Karush-Kuhn-Tucker (KKT) theory; see [55, 95]. More specifically, one is usually confronted with a variety of stationarity conditions; some providing strong conditions avoiding spurious solutions, but being hard to compute, whereas others are somewhat easier to compute, but too weak a filter for (local) optimality. Also, due to the structure of the feasible set, the design of solution algorithms is challenging for this problem class.

We close this section by mentioning that there is a broad literature on solvers for the TV- resp. TGVregularized reconstruction problem. However, as in this review we are primarily interested in different reconstruction models, here we do not dwell on discussing solvers in detail. Rather, we only refer to the first-order primal-dual algorithm for non-smooth convex minimization problems in [38], the secondorder-type semismooth Newton method for TV-regularization in [92, 99], the duality-based multigrid scheme in [40], and the divide-and-conquer algorithm using graph-cuts in [50]. This selection indicates the versatility in solver developments for variational regularization approaches.

2.2. Model-based methods for quantitative MRI. In addition to the single equation (2.1) in qualitative imaging, the quantitative imaging problem also involves a (typically nonlinear) operator e that links the image u_{true} to a (bio-)physical quantity of interest q_{true} via

$$(2.8) e(u_{true}, q_{true}) = 0$$

Often, e(u,q) = 0 is given by a system of ordinary or partial differential equations modelling an underlying physical process. Consequently, the aim of quantitative imaging is to identify the value q for

every pixel (or voxel in the case of MRI) given some measurements y that correspond to the qualitative parameter u_{true} . Hence, the resulting general (and idealized) formulation of the reconstruction problem can be stated as follows: Given y, estimate q (close to q_{true}) and an associated u (close to u_{true}) such that

(2.9)
$$\begin{cases} e(u,q) = 0, \\ y = Au + \eta. \end{cases}$$

In practice, however, due to ill-posedness of this reconstruction task, one can only expect approximate satisfaction of the above equalities.

In what follows, we focus on quantitative MRI as the underlying imaging modality. But let us also mention that other modalities fit the framework (2.9), too. Two specific examples are Photoacoustic Imaging [56, 64] and Magnetic Resonance Elastography [103], respectively. In the case of MRI we recall that the measured data y represent an incomplete (i.e. subsampled via P) set of the Fourier coefficients of an emitted signal. The latter encodes information about the total magnetization m of the nuclear spins of hydrogen atoms with (proton) density ρ of the tissue of interest aligned to a strong external magnetic field B. The underlying physical dynamics of m are governed by the Bloch equations, which are a system of ODEs:

(2.10)
$$\frac{\partial m(t)}{\partial t} = m(t) \times \gamma B(t) - (m_x(t)/T_2, m_y(t)/T_2, (m_z(t) - m_{eq})/T_1),$$

given some initial (equilibrium) condition. The relaxation terms in (2.10) occur due to the excitation of m from equilibrium state with an electromagnetic radio-frequency wave at resonance frequency of the nuclear spins with the gyromagnetic ratio γ . The excitation usually distorts m(t) from its equilibrium direction by a flip angle FA to some initial value m(0), where the relaxation starts. The quantity m_{eq} is the equilibrium magnetization, and T_1 and T_2 are the relaxation times of the longitudinal and transverse component of m (relative to B) with tissue dependent values, respectively. In case of magnetic field inhomogeneities the transverse relaxation time is slightly different and denoted by T_2^* ; in order to access T_2 a special re-focussing pulse is needed with the scope of so-called spin-echo sequences. The relaxation process releases again energy in the form of electromagnetic waves which can then be recorded by receiver coils as so-called echos after echo times TE. Thus, the equation e(u, q) = 0 of the general setting (2.9) takes now the form

$$(2.11) u = \rho \mathcal{B}(q)$$

where $u = \rho m$, $q = (\rho, T_1, T_2)$ and $\mathcal{B} : q \mapsto m$ is the solution map of the Bloch equations (2.10). For convenience of notation, by $\Pi : q \mapsto u$ we denote the map defined in (2.11), and sometimes in a slight misuse of notation we also refer to this map as the Bloch solution map.

MRI generates its power from the possibility to employ a variety of different sequences, i.e., spin excitation schemes, each focussing on different tissue properties and resulting in a different image contrast. Often, a second spin excitation is performed after repetition time TR even if full recovery of the equilibrium state of the magnetization m(t) has not yet been reached. This is typically reflected in the respective signal equations derived as solutions of (2.10). Specific choices of the flip angle FA, the repetition time TR, and the echo time TE enhance the tissue contrast with respect to T_1 , T_2 (or T_2^*), or ρ . This leads to T_1 -weighted, T_2 -weighted, proton density-weighted images, respectively.

Bloch solution maps for specific excitation sequences. For some specific magnetization-excitationsequences the Bloch solution map Π , or a satisfactory approximation thereof, can be stated explicitly. This is the case, for example, for a *multi-echo Fast Low-Angle Shot (FLASH)* sequence [85] for small flip angles *a*. In this case, the Bloch map is given by the Ernst equation [65]. By denoting the echo time and repetition time of this specific sequence by TE and TR, respectively, then we have (see [91]):

(2.12)
$$u = C \cdot \sin a \cdot \frac{1 - e^{-R_1 \cdot TR}}{1 - \cos a \cdot e^{-R_1 \cdot TR}} \cdot e^{-R_2^{\star} \cdot TE},$$

where $R_1 = 1/T_1$ and $R_2^{\star} = 1/T_2^{\star}$ is the apparent transverse relaxation time. Here, C is proportional to the equilibrium magnetization m and thus also to the proton density $C = c \cdot \rho$, with c a spatially varying factor related to the detection sensitivity [190]. The variation of TE within the multi-echo sequence and the flip angle allows for the estimation of R_2^{\star} , ρ and R_1 [154]. In practice, the approach uses two flip angles, one for proton weighted images and one for T_1 -weighted images, with six to eight echos at different TE each. Then, (2.12) can be re-parametrized with the so-called ESTATICS model [189] for estimating the apparent transverse relaxation time R_2^{\star} jointly from the two weightings:

(2.13)
$$u = u_{T_1} \cdot e^{-R_2^{\star} \cdot TE}$$
, for the T_1 -weighted echos,

(2.14)
$$u = u_{PD} \cdot e^{-R_2^{\star} \cdot TE}$$
, for the proton density-weighted echos.

It makes use of the fact that the decay constant R_2^{\star} of the exponential signal decay with TE, cf. Eq. (2.12), is identical for all weightings. Thus, we can simply write

(2.15)
$$u = \Pi(u_{PD}, u_{T_1}, R_2^{\star}),$$

where Π can be written as in (2.13) and (2.14).

Another popular flip angle sequence pattern is the *Inversion Recovery balanced Steady State Free Precession (IR-bSSFP)* [171]. Through this choice, the solution of the Bloch equations can be simulated by solving a discrete linear dynamical system; see for instance [52, 57].

Returning to (2.9) and recalling the notation for the Bloch solution map, one needs to estimate q where

(2.16)
$$\begin{cases} u = \Pi(q), \\ y = Au + \eta \end{cases}$$

An alternative and perhaps more practical formulation is

(2.17)
$$\begin{cases} u = \Pi(q), \\ u = \operatorname{Recon}(y) \end{cases}$$

where "Recon" denotes an arbitrary reconstruction scheme for the qualitative MR image u given the k-space data y. For instance, in view of our previous discussion on classical variational methods, one may consider

(2.18)
$$\begin{cases} u = \Pi(q), \\ u = \arg\min_{\tilde{u}} \frac{1}{2} \|A\tilde{u} - y\|_{2}^{2} + \mathcal{R}(\alpha; \tilde{u}). \end{cases}$$

Note that (2.18) can be interpreted as a coupled equilibrium system (for finding (u, q)) when replacing the convex reconstruction problem by its (necessary and sufficient) first-order optimality or Euler-Lagrange condition. Below we summarize two families of approaches for tackling (2.17), namely the *two-step approaches* and the *integrated-physics* method.

Two-step approaches for qMRI. As the name suggests, in this family of approaches the qualitative MRI reconstruction problem, i.e. the second problem in (2.17), is solved first to obtain u and then one uses the first equation to obtain the quantitative parameters q. However, due to the presence of noise, artifacts, the fact that u cannot be expected to equal u_{true} , as well as due to challenges that arise from

inverting Π , one relaxes the physical law to arrive at a nonlinear regression problem of the type

(2.19)
$$\begin{cases} \min_{q} \frac{1}{2} \|\Pi(q) - u\|_{2}^{2}, \\ u = \operatorname*{arg\,min}_{\tilde{u}} \frac{1}{2} \|A\tilde{u} - y\|_{2}^{2} + \mathcal{R}(\alpha; \tilde{u}). \end{cases}$$

We note that different quantitative reconstruction schemes will result from considering (2.19) and solving the qualitative problem with different choices of \mathcal{R} , including the spatially adapted TV/TGV from Section 2.1. Two particular instances of this technique are highlighted in the next section on data-driven methods.

The original magnetic resonance fingerprinting (MRF) [132] is a popular method also closely related to (2.19). In MRF, a series of L qualitative images $(u^{(1)}, \ldots, u^{(L)})$, corresponding to a sufficiently rich excitation sequence with L read-out times, are reconstructed with no regularisation, i.e., with $\mathcal{R} = 0$. This sequence is chosen, such that the Bloch solution map can be explicitly computed (in closed form, e.g. when using (IR-bSSFP)). In an offline phase, a series of Bloch responses $\mathcal{B}(q)$ pertinent to a range of realistic values for q are simulated (fingerprints) and stored in form of a dictionary denoted by Dic. Finally, the time series that results from considering the values in time of the qualitative images at a specific voxel is mapped to its closest fingerprint. Then this voxel is assigned the value of q that corresponds to that fingerprint. We outline this procedure in more detail in Algorithm 1 and note that the associated $\mathcal{B}(q)$ refers to the x, y-components of the magnetizations.

Algorithm 1 Magnetic Resonance Fingerprinting (MRF) for $q = (\rho, T_1, T_2)$ [132]

(i) Qualitative reconstructions:

Reconstruct the series of L qualitative images $u=(u^{(1)},\ldots,u^{(L)})$ by solving

$$u^{(\ell)} \in \operatorname*{arg\,min}_{u} \|P^{(\ell)}\mathcal{F}u - y^{(\ell)}\|_{2}^{2}$$
 using $u^{(\ell)} = \mathcal{F}^{-1}(P^{(\ell)})^{\top}y^{(\ell)}, \quad \ell = 1, \dots, L.$

(ii) Voxel-wise matching to the pre-computed fingerprints:

For every voxel i = 1, ..., N, extract the closest fingerprint to the time-series $u(i) = (u^{\ell}(i))_{\ell=1}^{L}$ according to

$$\mathcal{B}(q_i) = \underset{\mathcal{B}(q)\in\text{Dic}}{\arg\min} \frac{1}{2} \left\| \frac{\mathcal{B}(q)}{\|\mathcal{B}(q)\|_2} - u(i) \right\|_2^2.$$

Then assign $(T_1(i), T_2(i))$ to the *i*-th voxel and compute its proton density value ρ_i as

$$\rho_i = \frac{\|u(i)\|_2}{\|\mathcal{B}(q_i)\|_2}.$$

We note that the qualitative reconstructions in the first step of MRF suffer from severe artifacts due to the typical subsampling and the absence of regularization. As a result, L must be quite large to produce robust results; see [57] for an analysis of the interplay of these two steps. Another drawback of the method is the computational burden of dictionary matching, since the latter is typically very large. Improved versions of MRF aiming to tackle some of these challenges have subsequently appeared. For instance, in [52] a projected Landweber iteration, called BLIP, is proposed. In this approach the qualitative images are projected onto the dictionary in every iteration; see also related extensions in [77]. In [136], low-rank regularization on the qualitative reconstructions u was introduced. In [137] the dictionary containing the pre-computed magnetization trajectories was compressed using a low rank approximation in order to reduce the complexity of the second projection step in (2.19). Building on this idea of simplifying the expensive projection step several more sophisticated methods were developed. A brief summary over existing data-driven methods is provided in Section 3 below and for a very comprehensive and systematic review we refer to the recent overview article [176].

Integrated-physics approaches for qMRI. The previous approaches are highly dependent on the fineness of the dictionary. The latter is related to a sufficiently fine discretisation of the quantitative parameter space in order to yield accurate results. This, however, also results in an increased computational effort. In order to avoid the use of a precomputed dictionary, one can revisit (2.17) and impose the physical equation $u = \Pi(q)$ as a constraint instead of projecting onto the dictionary. In that case the regularization \mathcal{R} is applied directly to the quantitive map q:

(2.20)
$$\begin{cases} \min_{u,q\in\mathcal{C}_{ad}} \frac{1}{2} \|Au - y\|_2^2 + \mathcal{R}(\alpha;q), \\ \text{subject to} \quad u = \Pi(q), \end{cases}$$

which leads to the following reduced (integrated physics) formulation

(2.21)
$$\min_{q \in \mathcal{C}_{ad}} \frac{1}{2} \|A \circ \Pi(q) - y\|_2^2 + \mathcal{R}(\alpha; q).$$

Notice that the minimization for q must be performed over a range of *feasible* (i.e. realistic) physical values, denoted here by a convex set C_{ad} . The latter encodes a specific range of values depending on the inspected tissue types. The above optimization without the explicit regularization term $\mathcal R$ was proposed [57]. In that work, a projected Levenberg-Marquardt (L-M) iteration was adopted to solve (2.21), which is a regularized version of a projected Gauss-Newton scheme. We outline this technique in Algorithm 2.

Algorithm 2 Levenberg-Marguardt iteration for physics-integrated gMRI (e.g. Eq. (2.20) with no explicit regularization) [57]

Given $q_0 \in C_{ad}$ and a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ of positive real numbers with $\lambda_n \downarrow 0$, iteratively solve the following problems for $n = 0, 1, 2, \ldots$:

 $\tilde{y}_n = y - A \circ \Pi(q_n),$

(2.22)

$$h_n = \underset{h}{\arg\min} \|\Pi'(q_n)h - A^{\dagger}\tilde{y}_n\|_2^2 + \lambda_n \|h\|_2^2,$$
$$q_{n+1} = P_{\mathcal{C}_{nd}}(q_n + h_n),$$

Here, $P_{\mathcal{C}_{ad}}$ is the projection operator onto the feasible set \mathcal{C}_{ad} . The operator Π' is the Frechét derivative of the parameter-to-solution map of the Bloch equations, and A^{\dagger} is the generalized inverse of the subsampled Fourier transform, i.e. $A^{\dagger} = \mathcal{F}^{-1}P^{\dagger}$, where P^{\dagger} is the zero filling operator, and \mathcal{F}^{-1} is the inverse Fourier transform. For comparing a variety of algorithms for qMRI, we present a set of examples in the following. Our tests are based on synthetic data from an anatomical brain phantom, publicly available from the Brain Web Simulated Brain Database [24, 48, 119]. More details on how to generate this data can be found in [57] or [52].

The results for each of the above mentioned algorithms can be found in Figures 3 and 4. We use the MRF reconstruction as initialization for the L-M algorithm, and compare that with the result of the BLIP algorithm when using a relatively refined dictionary. The advantage of integrated-physics approaches is evident by checking the reconstructions in Figure 3 as well as quantitatively by looking at the error maps in Figure 4. In this example, we have used a time series of the 1/8 Cartesian-subsampled kspace data (Fourier coefficients of magnetizations) of length 40. In the original MRF algorithm (typically requiring a large time series of k-space data), this test setting is far from yielding a reasonable result, while BLIP has improved a little by enforcing the projection to the Bloch manifold. But still one observes deficiencies. The method with integrated physics, however, appears to be efficient and it returns the best results among the three methods.



FIGURE 3. From left to right (columns): The estimated parameters T_1 , T_2 and the proton density. From top to bottom (rows): The ground truth, MRF, BLIP, and the integrated model with the L-M algorithm.

The results in [57] show that the proposed integrated physics model can work well for settings where an explicit solution formula for the Bloch equations is available. However, in most cases (excitation sequences) the Bloch equations have no explicit solution formula. In addition, from a practitioners perspective the Bloch equations appear to be only a simplified mathematical model under some assumptions. As in reality these assumptions are often not realistic, one indeed requires a more complex model. In this vein, deep learning turns out to be a useful data-driven technique for learning or approximating such physical models. In Subsection 3.3 we will therefore discuss extracting physical laws from data.



FIGURE 4. Error maps of each estimated parameters for the three algorithms. From left to right: the errors for T_1 , T_2 and the proton density. From top to bottom: MRF, BLIP, and the integrated model with the L-M algorithm. Note the different scales in all the three methods reflecting the largest errors in the estimation.

3. DATA-DRIVEN METHODS

From variational to data-driven regularization. Traditional regularization techniques as introduced in Section 2.1 are based on strong theoretical foundations and often provide good results in many practical applications. However, in many situations important structure in the data may escape "handcrafted" regularizers. An example substantiating this claim is depicted in Figure 7 where the wavelet decomposition of a clean image is shown on the left. Natural clean images are assumed to have a sparse wavelet decomposition which is exploited to reconstruct images by solving the problem

(3.1)
$$\min_{u \in X} \frac{1}{2} \|Au - y\|_2^2 + \lambda \|Wu\|_1.$$

Here W denotes the discrete wavelet transform. Such strategies have been analyzed for instance in [51, 81, 82, 32, 63, 67]. Despite good practical results the underlying assumption in (3.1) is simply that the wavelet-coefficients of the true image Wu_{true} are sparse in the sense that $||Wu_{true}||_0$ is small. However, looking at the first image in Figure 7, we clearly observe that image information is "hidden" in the coefficients. Indeed, they are not just sparse, but accumulated at edges and highly correlated across different scales. This correlation between wavelet-coefficients is not captured by the simple model in (3.1). To address this issue, one natural strategy is to not only design regularization strategies by taking into account obvious a priori assumptions like sparsity, but to learn parts of the regularization from data. In this vein, during the past decade various methods have emerged; see [9]

for a comprehensive overview. A more recent article along the same lines is [86], while data-driven reconstruction strategies with a focus on medical imaging are presented in [167, 161].

One of the first approaches in this direction which directly builds on the ideas of wavelet-regularization is *dictionary learning* where the goal is to find a sparse linear representation of a large amount of given image patches which can then be used as a data-driven regularizer. Dictionary learning usually involves the solution of the non-convex and non-differentiable problem

(3.2)
$$\min_{D \in \mathcal{D}, C \in \mathcal{C}} \|DC - X\|_F^2 + \lambda \|C\|_0,$$

where \mathcal{D} is a set of admissible dictionaries, \mathcal{C} is the set of sparse coefficients, X is a collection of clean image patches and again $\lambda > 0$ denotes some regularization parameter that controls the sparsity of the coefficients. Here and in the forthcoming part of the article $||M||_F$ denotes the classical Frobenius norm of a matrix M. In a second step the dictionary can be used as a regularizer in the sense that the following problem is solved

(3.3)
$$\min_{u,C} \frac{\mu}{2} \|Au - y\|_2^2 + \frac{1}{2} \|Ru - DC\|_F^2 + \lambda \|C\|_0.$$

to obtain an estimate for the ground truth. Here R is an operator that extracts patches from the image, cf. Section 3.2 for details. It has been recognized later that a decent regularization effect can also be obtained by training the dictionary simultaneously to reconstructing the image u. The corresponding method is called *blind compressed sensing* and is reviewed below in Section 3.2, where also its applicability to qMRI is demonstrated. Dictionary learning methods in general have been proven to be very successful in medical imaging during the past decade [158, 121, 125] and are still being used frequently in practice due to their high interpretability, cf. e.g. [148, 161]. A potential drawback is the limited expressivity as the final reconstruction method is still of the same type as (3.1) and it is not yet clear whether the approach is suited to capture all complex imaging structures needed to represent natural images.

Towards interpretable neural networks for linear inverse problems. For more flexibility and to better model these complex structures, methods based on various types of neural networks have become increasingly popular over the recent years and represent the state of the art in many imaging applications today [129, 188, 124]. However, these methods often lead to uninterpretable outcomes and unstable reconstruction algorithms, in the sense that small changes in the setup can lead to significant changes in the image reconstruction [7, 144]. Nevertheless, the practical results are very impressive, which is why there is a natural demand to combine deep learning strategies with interpretable and robust reconstruction methods for inverse problems. One of the earliest attempts in this direction is called *algorithm-unrolling*. The motivation is to start from an iterative scheme which is known to converge against the solution of (2.2) and to interpret this scheme as a neural network. In [83] the ISTA algorithm (see [17] for details) is under consideration, which is defined subsequently via

(3.4)
$$u_{k+1} = \operatorname{prox}_{s_k \mathcal{R}(\alpha, \cdot)}(u_k - s_k A^T (Au_k - y)) = \operatorname{prox}_{s_k \mathcal{R}(\alpha, \cdot)}(W_k^1 u_k + W_k^2 y) \quad u_0 \in X,$$

when applied to the problem (2.2) with $\mathcal{D}(Au, y) = \frac{1}{2} ||Au - y||_2^2$. Here we set $W_k^1 = I - s_k A^T A$ and $W_k^2 = s_k A^T$ and define $S^k(\alpha, y, W, u_0)$ to be the k-th iterate of the algorithm when initialized at u_0 , with data y, regularization parameter α , step sizes $s_k > 0$ and matrices $W = (W_0^1, \ldots, W_k^1, W_0^2, \ldots, W_k^2)$. Here $\operatorname{prox}_{s_k \mathcal{R}} : X \to X$ denotes the classical proximal operator from convex analysis, see [163]. Note that for $\mathcal{R}(u) = ||u||_1$ this proximal operator is a piecewise differentiable and nonlinear function which acts component-wise. Hence in this case $S^k(\alpha, y, W, u_0)$ can be understood as a k layer neural network with input u_0 , weights W and activation function $\operatorname{prox}_{s_k \mathcal{R}}$. In [83] the W-matrices are replaced by general matrices that can be learned from data by solving the following problem

(3.5)
$$\min_{W} \frac{1}{2M} \sum_{i=1}^{M} \|S^{k}(\alpha, y_{i}, W, u_{0,i}) - u_{true,i}\|_{2}^{2}$$

using automatic differentiation such as implemented in nowadays popular software libraries [1, 151] again using training data $(u_{true,i}, y_i)_{i=1,...,M}$. For an overview on automatic differentiation and how it is applied to learning problems, we refer to [16]. It could be shown in experiments that this approach leads to very good outcomes, already when only a few iterations are unrolled. However by replacing all matrices W_k^1, W_k^2 by neural networks, a massive loss of interpretability has to be claimed such that many (open) questions naturally arise: Does the method still converge as the number of unrolled iterates tends to infinity? What is the connection to the optimization problem (2.2), and which regularization properties are inherited from the corresponding variational problem? Moreover due to the slow convergence of proximal gradient type algorithms many iterations need to be unrolled, which might lead to memory issues [107]. Some research has been conducted in order to overcome these limitations, and also the convergence issue has been addressed, see [43, 127] and [41] for an overview over existing results. A natural idea to retain more interpretability compared to pure unrolling is to reduce the number of parameters that are learned in the variational model. For instance, in [2] only the step-sizes are learned. Despite the facts that algorithm unrolling delivers nice practical results in general and also addresses the issues of interpretability to a certain extent there are also shortcomings, such as, e.g., the large amount of memory consumption. In addition it is noted in [72] that the method does not converge in all scenarios when $k \to \infty$, and the reconstruction quality might heavily deteriorate when the number of unrolled parameters is increased during test-time. This is why in [13, 72] the authors provide a strategy, called (deep) equilibrium models, to learn parameters directly in the limit-point of the unrolled method. For this purpose note that any minimizer u of (2.2) has to satisfy the fixed point equation

(3.6)
$$u = \operatorname{prox}_{s\mathcal{R}(\alpha,\cdot)}(u - sA^T(Au - y)) \approx D_\theta(W^1u + W^2y), \quad s > 0.$$

where again $W^1 = I - sA^TA$, $W^2 = sA^T$ and D_θ is a parametrized approximation of $\operatorname{prox}_{s\mathcal{R}}$, e.g., by a neural network with weights θ . The idea is now to learn $W = (W^1, W^2)$ and θ from training data by differentiating through the implicitly given fixed-point in (3.6). The strategy is applied to linear inverse problems in [72] and is further analyzed in [192, 23, 68]. A related method which addresses the convergence problem and parametrizes the prox-operator is the highly flexible Plug-and-Play approach [186]. Here also primal-dual algorithms like ADMM [186] or variants of ISTA [70] are considered which make use of proximal operators. But in contrast to deep equilibrium models where the parameters W, θ are learned in an end-to-end fashion, the authors in [186] interpret the prox-operator as a denoiser and replace it by a more powerful denoiser $D_\theta : X \to X$, often parametrized by a neural network with weights θ . In a second step after the denoiser is chosen the method is used iteratively as

(3.7)
$$u_{k+1} = D_{\theta}(u_k - s_k A^T (A u_k - y)),$$

where again $s_k > 0$ denotes the step size in iteration k. After the first attempts of the method using off-the-shelf denoisers like BM3D [49] have been proven to be successful, practitioners started to use state-of-the-art methods based on neural networks as denoisers. Due to the possibly complex structure of these denoisers it is a priori unclear whether the plug-and-play algorithm converges or not. This issue has been addressed in [164, 169, 70, 110] using quite restrictive classes of denoisers or neural networks, see also the recent overview article [90] where additionally regularization properties of the plug-and-play approach are discussed.

Data-driven methods for qMRI. As already pointed out in Section 2.2 the existing methods for quantitative MRI can be roughly divided into two different approaches. The first one, explained in (2.17), reconstructs a high number of MRI-images solving a linear inverse problem to subsequently use these images in a second step to find quantitative parameters that match the reconstructed signal from the first stage. The second class of methods, also called physics-integrated approaches directly computes the physical parameters using the corresponding physical model based on the Bloch equation in (2.10). In principle every data-driven method for linear inverse problems from the introduction of this section can be used to enhance the two-step-procedure. First, the method reconstructs the magnetization trajectories for each pixel in the domain of interest and then the nonlinear projection step is carried out. In [108, 33, 88] sparsity and low rank regularization techniques are used for the signal reconstruction while in [20, 195] deep learning based methods are proposed. A very recent example which also falls into this category and combines total variation based regularization with an unrolling methodology already discussed above, the lower level problem (2.7). In the spirit of the unrolling methodology already discussed above, the lower level problem is replaced by $k \in \mathbb{N}$ steps of an iterative solver (in this case the primal-dual-hybrid-gradient method proposed in [38]) that is known to converge against a solution of the lower level problem. Mathematically speaking,

(3.8)
$$S^k(\alpha, y, u_0) \to \underset{v \in X}{\operatorname{arg\,min}} \frac{1}{2} \|Au - y\|_2^2 + \|\alpha \nabla u\|_1, \quad \text{as } k \to \infty,$$

where $S^k(\alpha, y, u_0)$ denotes the k-th iterate of the unrolled algorithm for data y, with regularization parameter α and initialization u_0 . The spatially dependent regularization parameter α is then replaced by a rather small neural network $y \rightarrow \alpha_{\theta}(A^{\dagger}y)$ with weights θ and the input being the zero filling solution $A^{\dagger}y$. The overall learning problem eventually reads

(3.9)
$$\min_{\theta} \ \frac{1}{2M} \sum_{i=1}^{M} \|S^k(\alpha_{\theta}(A^{\dagger}y_i), y_i, A^{\dagger}y_i) - u_{true,i}\|^2.$$

Here we use again a training set of clean and distorted images $(y_i, u_{true,i})_{i=1,\dots,M}$ of size M. Given new test-data that is generated according to $y = A \circ \Pi(q_{true}) + \eta$ with noise η as in (2.16), the method reconstructs first

(3.10)
$$u^* \in \underset{u \in X}{\operatorname{arg\,min}} \ \frac{1}{2} \|Au - y\|_2^2 + \|\alpha_\theta(A^{\dagger}y)\nabla u\|_1,$$

using $A^{\dagger}y$ here as the input of the neural network since it already contains significant image structure. Subsequently the nonlinear projection problem in (2.19) is solved to obtain an estimate of q_{true} . Besides data-driven two-step techniques for qMRI where the data-driven aspect mostly refers to the first linear reconstruction step there are also methods which aim at modifying the nonlinear projection step in (2.19) to infer the tissue parameters. In [137] it was empirically shown that the magnetization for different MR sequences actually features a low-rank structure, which was then used to reduce the complexity needed for the tissue parameter inference given a reconstructed magnetization signal. Building on this idea more complex models where used in order to modify or even fully replace the projection step, cf. [10, 47, 44]. Most recent techniques combine data-driven regularization techniques with the first and the second reconstruction step and end up using an end-to-end learning based strategy. We mention here [75, 179] and refer to the comprehensive overview article [176] for a general account including recent references.

In the subsequent part of this article we will present three approaches to incorporate data-driven techniques into the reconstruction process for qMRI. The first one, outlined in Section 3.1, is a two step-approach which uses a data adaptive technique for the linear reconstruction step in (2.19). In particular, it addresses the problem that traditional variational reconstruction methods such as TV reconstruction [165] often lead to a bias in contrast, which might result in a systematic error in the tissue parameters. The second approach, described in Section 3.2, uses integrated physics and a regularization strategy based on dictionary learning, which has been proven successful in particular for unbiased linear estimation of highly under sampled k-space data when no or only a small amount of training data is available. The method which is summarized in Section 3.3 falls also into the category of

integrated physics but uses training data to modify the solution map of the Bloch equation. In contrast to the neural network based methods described above, the goal here is not only to allow for faster inference of the parameters, but also to account for uncertainties in the physical model.

3.1. Data-driven bias-free post-processing methods suitable for qMRI. Statistical methods for imaging focus on properties of the error term η that is usually considered to be a random variable. Its distribution usually depends on the image acquisition method, often, but not always, additive Gaussian noise in equation (2.1) is appropriate. Specifically, let us assume data y_i at locations i to be distributed as $y_i \sim P_{\theta_i}$ with some density $p(\cdot, \theta_i)$ depending on some local parameter θ_i (typically from \mathbb{R}^p) with the probability distribution belonging to some parametric (typically exponential) family.

For imaging data usually a structural assumption can be formulated, i.e., there exists a disjoint partitioning of the space of the positions i, such that θ_i is constant within each partition. Based on this assumption, an iterative adaptive reconstruction method for the true parameter θ_i called Propagation-Separation approach was defined in [153]. There, a locally adaptive weighting scheme $W_i^{(k)} = \{w_{ii}^{(k)}\}$ at a voxel i at iteration step k is employed for a weighted maximum likelihood estimator

(3.11)
$$\hat{\theta}_i^{(k)} = \underset{\theta}{\arg\max} \sum_{j} w_{ij}^{(k)} \log(p(y_i, \theta_i)),$$

where the sum is over all neighboring voxels j with non-vanishing weights $w_{ij}^{(k)}$. More specifically, the adaptive weights at step k are defined as

(3.12)
$$w_{ij}^{(k)} = K_{loc}(l_{ij}^{(k)}) \cdot K_{st}(s_{ij}^{(k)}),$$

i.e., the product of two kernel functions K_{loc} and K_{st} . The first factor in the kernel definition then refers to the Euclidean distance

(3.13)
$$l_{ij}^{(k)} = ||i - j||_2 / h^{(k)}$$

in the design space using a bandwidth $h^{(k)}$, the second factor refers to a distance in feature space

(3.14)
$$s_{ij}^{(k)} = N_i^{(k-1)} \cdot \mathcal{KL}(\hat{\theta}_i^{(k-1)}, \hat{\theta}_j^{(k-1)}) / \lambda$$

where \mathcal{KL} refers to the Kullback-Leibler distance of the two respective probability distributions, λ is the adaptation parameter of the procedure and $N_i^{(k-1)} = \sum_j w_{ij}^{(k)}$ approximates the variance reduction from the k - 1-th step. The resulting procedure, adaptive weights smoothing, with some increasing sequence of bandwidths $h^{(0)}, \ldots, h^{(k^*)}$ performing k^* steps results in noise reduced parameter estimates $\hat{\theta}_i^{(k^*)}$ that do not suffer from the blurring observed in non-adaptive methods but does edge-preserving smoothing.

In [152], a variant of the adaptive weights smoothing employing a statistical penalty $s_{ij}^{(k)}$ as the maximum over a local (rectangular) patch of voxels was defined. This patchwise adaptive smoothing combines the iterative procedure outlined above with the idea of patchwise comparisons used by non-local means (NLM) [29]. The resulting procedure combines the edge preserving property of the original method while avoiding the cartoon-like appearance of the reconstructed image that is the result of the structural assumption above. In [152] the method was extensively compared with TV and TGV based reconstruction methods, see the Figure 6 and 7 within this paper. In Figure 5, we provide an example reconstruction of a noisy MR image (from fully sampled k-space), where T_1 weighted example data has been taken from the IXI Dataset [111].

The adaptive weights smoothing procedure can be applied to estimate quantitative MRI parameters using a physical model. We will rely on model (2.12) and more specifically the ESTATICS reparametrization (2.13) and (2.14) to estimate quantitative parameters like the proton density ρ or



FIGURE 5. From left to right: MR image with artificial Gaussian noise and fully sampled k-space. Reconstruction with patchwise adaptive smoothing. Original noise-free image. The adaptive smoothing is capable of removing the noise and preserving the details of the image.

relaxation times T_1 or T_2^{\star} . First, we solve the optimization problem:

(3.15)
$$\hat{\theta}_{LS} = \underset{\theta}{\arg\min} \sum_{i=1}^{n} (u_i - \Pi(u_{PD}, u_{T_1}, R_2^{\star}))^2$$

to infer on $\theta = (u_{PD}, u_{T_1}, R_2^{\star})$. The sum is calculated over all *n* echos of the two sequences from the multi-echo MPM measurement. Comparing (2.12) and (2.13), (2.14) it is straightforward to determine the remaining quantitative parameters R_1 , and *A* (if TR_{T_1} and TR_{PD} are equal) by

(3.16a)
$$\hat{R}_1 = -\ln\left(\frac{\hat{u}_{T_1} - \hat{u}_{PD} \cdot \frac{\sin a_{T_1}}{\sin a_{PD}}}{\hat{u}_{T_1} \cdot \cos a_{T_1} - \hat{u}_{PD} \cdot \frac{\sin a_{T_1}}{\sin a_{PD}} \cdot \cos a_{PD}}\right) / TR_2$$

(3.16b)
$$\hat{A} = \frac{(1 - \cos a_{T_1} \cdot e^{-R_1 \cdot TR})}{\sin a_{T_1} \cdot (1 - e^{-\hat{R}_1 \cdot TR})} \cdot \hat{u}_{T_1}.$$

The actual measurements within an MPM sequence are corrupted by noise which will be propagated to the model parameter maps. The adaptive weights smoothing procedure outlined above can be used to reduce the noise in the quantitative maps while preserving the fine structures of the brain tissue that can be seen in the images. First we make use of the fact that the ESTATICS re-parametrization (2.13) and (2.14) of the signal model (2.12) has a low parameter-induced nonlinearity and leads to approximate Gaussianity of the estimates

$$\vec{u}^{(0)} = \left(\hat{u}_{T_1}^{(0)}, \hat{u}_{PD}^{(0)}, \hat{R}_2^{\star(0)}\right)^{\top}.$$

A voxelwise estimate $\hat{\Sigma}$ for the covariance of the parameter estimates can be obtained from the least squares estimation procedure [154]. These three-dimensional parameter maps are then iteratively smoothed by an increasing sequence of bandwidths h_k for $k = 0, \ldots, k^*$ and the definition of locally adaptive as outlined above. Specifically, the the statistical penalty $s_{ij}^{(k)}$ defined as

(3.17)
$$s_{ij}^{(k)} = N_i^{(k-1)} \cdot \left(\vec{u}_i^{(k-1)} - \vec{u}_j^{(k-1)}\right)^\top \hat{\Sigma}_i^{-1} \left(\vec{u}_i^{(k-1)} - \vec{u}_j^{(k-1)}\right)$$

based on the estimates $\vec{u}_i^{(k-1)}$ and $\vec{u}_j^{(k-1)}$ and the sum of weights $N_i^{(k-1)} = \sum_j w_{ij}^{(k-1)}$ from the previous step. The new estimates in step k are obtained by

(3.18)
$$\vec{u}_i^{(k)} = \sum_j w_{ij}^{(k)} \vec{u}^{(0)} / \sum_j w_{ij}^{(k)}.$$



FIGURE 6. Reconstruction of quantitative MR parameters, i.e., R_1 , R_2^{\star} , and ρ , i.e., proton density PD, after applying adaptive smoothing with comparison to the result obtained from threefold repeated data (NEX=3) and with non-adaptive smoothing, i.e., when a classical Gaussian filtering is applied. While the non-adaptive Gaussian filtering does remove the noise at the cost of blurring, the adaptive method is able to preserve the fine structural details. The comparison between the single data with the threefold repetition shows, that adaptive smoothing is capable of saving image acquisition time by using less data with a comparable resulting map.

At the last iteration step $k = k^{\star}$ the final smoothed maps

$$\vec{u}^{(k^{\star})} = \left(\hat{u}_{T_1}^{(k^{\star})}, \hat{u}_{PD}^{(k^{\star})}, \hat{R}_2^{\star(k^{\star})}\right)^{\top}$$

can be used within (3.16a) and (3.16b) to get implicitly smoothed quantitative maps. The application of this procedure to real MPM data from [139] is shown in Figure 6.

3.2. Dictionary learning approaches for qMRI. Instead of learning the, possibly spatially dependent regularization parameter as proposed in Section 2.1, the goal of dictionary-learning is to learn a

sparse representation of a class of clean images. This is motivated by the success of Wavelet based methods and the observation that most natural images admit a sparse representation in some orthonormal basis, cf. [133] for an overview on wavelets. In Figure 7 a clean image is shown together with its wavelet-decomposition. In the second and third image a large number of wavelet coefficients are gradually set to zero. Still, the image retains most of its structural features which indicates that the essential information is stored in only a small number of coefficients. Since learning a basis or dictionary for the entire image has too many degrees of freedom, the classical approach is to learn sparse representations for small image patches [4]. During this subsection, we will work with pre-discretized complex images $u \in \mathbb{C}^{n \times m} =: X$, where $m \times n$ is the number of pixels. For such an image let $R_i : \mathbb{C}^{n \times m} \to \mathbb{C}^P$ denote the linear operator that cuts out a quadratic image patch of size $p \times p$ and puts it into a row-vector of size $P = p^2$. Here the index set I is an enumeration of a selection of the possible $n \times m$ patches that can be cut out of the image u, see [159] for details. For simplicity we will work with I indexing the set of all possible $m \cdot n$ patches of an image. The underlying assumption is that every patch $R_i u_{true}, i \in I$, which is cut out of the ground-truth image can be represented as a sparse linear combination of $K \in \mathbb{N}$ image atoms $\varphi_1, \ldots, \varphi_K \in \mathbb{C}^P$ in the sense that

(3.19)
$$R_i u_{true} = \sum_{l=1}^{K} c_{li} \varphi_l \qquad \|c_i\|_0 \le s_i \ll \min\{K, P\}, \ i \in I,$$

where the sparsity $s_i \in \mathbb{N}$ of every patch $R_i u_{true}$ is a priori unknown but significantly smaller then the number P of pixels in each patch and the number K of dictionary elements. Note that (3.19) can also be written in a matrix vector form Ru = DC where the operator $R : \mathbb{C}^{m \times n} \to \mathbb{C}^{P \times mn}$ applies R_i to every patch indexed by $i \in I$, $D = (\varphi_1, \ldots, \varphi_K) \in \mathbb{C}^{P \times K}$ denotes the unknown dictionary and $C = (c_1, \ldots, c_{mn}) \in \mathbb{C}^{K \times mn}$ is the collection of sparse coefficients. If the number of atoms K is strictly larger then P the dictionary is called overcomplete. The approach in [4] suggests to start with a training set $X = (x_1, \ldots x_M) \in \mathbb{C}^{P \times M}$ of M training patches of clean images that represent a large variety of image structures. These patches are collected in a large matrix $X \in \mathbb{C}^{P \times M}$ for which the following problem is solved

(3.20)
$$\min_{D\in\mathcal{D},C\in\mathbb{C}^{M\times K}}\frac{1}{2}\|X-DC\|_F^2+\lambda\|C\|_s.$$

Here $\mathcal{D} \subset \mathbb{R}^{P \times K}$ is the set of admissible dictionaries, often chosen to be the set of column normalized dictionaries, i.e. $\|\varphi_i\|_2 = 1$ for $i = 1, \ldots, K$ and $s \in [0, 1]$ defines the sparsity promoting regularizer. Note that the problem (3.20) is non-convex and non-differentiable and finding a global solution of this problem is known to be NP-hard, cf. [185]. Therefore possible solution algorithms focus on computing



FIGURE 7. The image on the left shows the Haar-Wavelet coefficients of a clean image, the second image shows the clean image when the only the largest 10% of the coefficients are kept. The rest is set to zero. On the right image less than 5% of the coefficients are kept.

stationary points under usage of techniques from variational analysis [159, 178] or on greedy type algorithms, see [4]. Once the dictionary $D \in \mathcal{D}$ is found, the following problem is solved

(3.21)
$$\min_{u \in X, C \in \mathbb{C}^{P \times mn}} \frac{\mu}{2} \|Au - y\|_2^2 + \frac{1}{2} \|Ru - DC\|_F^2 + \lambda \|C\|_s$$

to find the estimation of u_{true} . For s = 1 the problem is convex and can be efficiently solved using classical (accelerated) splitting algorithms like FISTA [17]. For s < 1 again the problem is non-convex (even locally non-Lipschitz for C = 0 which challenges solution algorithms and stationarity characterizations [100]).

Blind compressed sensing. A related method is *blind compressed sensing* or *blind transform learning*, cf. [158, 159, 178]. While the dictionary learning approach presented above requires the solution of the non-convex problem (3.20) in a first offline-training-phase requiring lots of training data, the authors in [74] noticed that very good results can already be obtained by assuming a relatively small (often quadratic) dictionary size and by learning the dictionary simultaneously to the reconstruction process. This idea leads to the optimization problem

(3.22)
$$\min_{u \in X, D \in \mathcal{D}, C \in \mathbb{C}^{P \times mn}} \frac{\mu}{2} \|Au - y\|_2^2 + \|Ru - DC\|_F^2 + \lambda \|C\|_s$$

Let us briefly sketch the details of [159] where an efficient update strategy is combined with a convergence analysis. We will focus on learning an orthogonal transform here, while the approach presented in [159] also allows for more general dictionaries. This corresponds to the choice

$$\mathcal{D} = \{ D \in \mathbb{C}^{K \times K} \mid D^H D = I \}.$$

The general idea of the solution algorithm is to update one variable while keeping the others fixed and then iterate cyclically. We start with the update of the coefficients, i.e. we want to optimize over $C \in \mathbb{C}^{K \times K}$ while keeping D and u fixed. The resulting optimization problem in the k-th step of the algorithm reads for given $D_k \in \mathcal{D}$ and $u_k \in X$:

(3.24)
$$C_{k+1} \in \underset{C \in \mathbb{C}^{P \times mn}}{\operatorname{arg\,min}} \ \frac{1}{2} \|Ru_k - D_k C_k\|_F^2 + \frac{\lambda_{C,k}}{2} \|C - C_k\|_F^2 + \lambda \|C\|_s,$$

whose solution has a closed form in terms of the well known proximal operator, see [163]. Here we took advantage of the orthogonality of the transform $D^{-1} = D^H$. Let us now consider the update for the transform D for fixed u_k , C_{k+1} and previous iterate D_k . We want to calculate in this step

(3.25)
$$D_{k+1} \in \underset{D \in \mathcal{D}}{\operatorname{arg\,min}} \ \frac{1}{2} \|DC_{k+1} - Ru_k\|_F^2 + \frac{\lambda_{D,k}}{2} \|D - D_k\|_F^2$$

It is remarkable that the non-convex problem in (3.25) which involves the solution of a quadratic objective on the matrix manifold $\mathcal{D} \subset \mathbb{C}^{K \times K}$ has a closed form solution, that can be computed as fast as the singular-value-decomposition of $Ru_kC^H + \lambda_{D,k}D_k$ can be computed. In general the minimizer is not unique, cf. [159]. Once we have updated the dictionary D the last step requires the update for the image u. Therefore the following problem needs to be solved

(3.26)
$$u_{k+1} \in \underset{u}{\operatorname{arg\,min}} \ \frac{\mu}{2} \|Au - y\|_2^2 + \frac{1}{2} \|Ru - D_{k+1}C_{k+1}\|_F^2 + \frac{\lambda_{u,k}}{2} \|u - u_k\|^2.$$

The equivalent first-order system of this quadratic optimization problem reads

(3.27)
$$(\mu A^H A + R^T R + \lambda_{u,k} \mathrm{Id}) u = A^H y + R^T D_{k+1} C_{k+1} + \lambda_{u,k} D_{k+1}$$

The system in (3.27) is in general very large and can only be solved approximately for a general forward operator A, e.g. using methods of conjugate gradient type [78]. However in the case of the MRI operator and a special form of R this system has a closed form solution which can be computed easily [158, 159] for details. The schematic overall algorithm for the linear MRI problem is depicted in Algorithm 3. The procedure can be shown to converge against a stationary point in the

Algorithm 3 Blind Compressed Sensing [158]

Given $u_0 \in X, D_0 \in \mathcal{D}, C_0 \in \mathbb{C}^{P \times nm}$ and a sequence $\{\lambda_n^u, \lambda_n^D, \lambda_n^C\}_{n \in \mathbb{N}}$ of positive real numbers, iteratively solve the following problems for n = 0, 1, 2, ...:

- (1) Compute C_{k+1} given u_k, D_k by formula (3.24).
- (2) Compute D_{k+1} formula (3.25).
- (3) Compute u_{k+1} using formula (3.27).
- If a stopping criterion is met, return u_k .

limiting sense [163] under the assumption that the overall objective function satisfies the so called Kurdyka-Lojasiewicz-inequality (KL-inequality) [12, 11] and there exists a constant c > 0 such that $\min\{\lambda_{k,u}, \lambda_{k,D}, \lambda_{k,C}\} \ge c$ for every k. In [159] a refined convergence result is presented which does neither require the KL-inequality nor the step-size bounds but only yields convergence up to a subsequence. For a proof of this result and details see again [159]. A prototypical numerical result of the discussed approach is presented in Figure 8.



FIGURE 8. From left to right: Sampling mask, zero filling solution, dictionary learning solution as described in Algorithm 3.

Blind compressed sensing for qMRI. Despite the relatively high computational effort the methodology of blind dictionary learning has recently been proven to be successful also for qMRI, [116, 178]. We recall the optimization problem for qMRI which was considered also in (2.21):

$$\min_{q \in \mathcal{C}_{ad}} \frac{\mu}{2} \|A \circ \Pi(q) - y\|_2^2 + \frac{1}{2} \|q\|_2^2$$

with the minor modification that we put the penalty parameter in front of the data discrepancy term. However instead of using a Tikhonov regularisation approach with a smooth quadratic penalty as proposed in [57], a regularization strategy via blind dictionary learning is proposed in [116] and some ongoing work [178]. Analogously to the corresponding method for qualitative MRI which is described at the beginning of this subsection we aim at solving the optimization problem

(3.28)
$$\min_{q \in \mathcal{C}_{ad}} \frac{\mu}{2} \|A \circ \Pi(q) - y\|_2^2 + \mathcal{R}(q),$$

where the regulariser $\mathcal{R}(q)$ is given by another minimization problem

(3.29)
$$\mathcal{R}(q) = \inf_{D \in \mathcal{D}, C \in \mathcal{C}} \frac{1}{2} \|Rq - DC\|_F^2 + \alpha \|C\|_s.$$

Here again $s \in [0,1]$ specifies the sparsity promoting regularizer which was also proposed in the linear case and the set C is the set of sparse coefficients. R denotes the linear patch-extraction operator and $\mu, \alpha > 0$ balance overall the regularization effect introduced by learning the dictionary.

The eventual optimization problem reads

(3.30)
$$\min_{(q,D,C)\in\mathcal{C}_{ad}\times\mathcal{D}\times\mathcal{C}}\frac{\mu}{2}\|A\circ\Pi(q)-y\|_{2}^{2}+\frac{1}{2}\|Rq-DC\|_{F}^{2}+\alpha\|C\|_{s}$$

In [178] a model based optimization algorithm is proposed in order to find limiting-stationary points of (3.30). While the update steps for the sparse coefficients C and the dictionary D look similar to the steps in (3.24) and (3.25) respectively, the update step of the tissue parameter q does not require the solution of a linear equation as in (3.27) but the solution of a non convex constrained optimization problem. For this purpose a Levenberg-Marquardt approximation is employed and the eventual update step reads (3.31)

$$q_{k+1} = \underset{q \in \mathcal{C}_{ad}}{\operatorname{arg\,min}} \frac{\mu}{2} \|A \circ \Pi'(q_k)[q - q_k] + A \circ \Pi(q_k) - y\|_2^2 + \|Rq - D_{k+1}C_{k+1}\|_F^2 + \frac{\lambda_{q,k}}{2} \|q - q_k\|_2^2.$$

In [178] a line-search strategy is proposed to find the step size-parameter $\lambda_{q,k} > 0$ that has to be adjusted in dependence of the unknown Lipschitz constant of Π' in order to guarantee a sufficient descent of the objective in each step of the algorithm. The overall method alternates between the three optimization steps and is shown to converge against a limiting-stationary point. Prototypical numerical results from [178] are depicted in Figure 9, where the dictionary learning approach is compared against the pure Levenberg-Marquardt method from Section (2.2) and the BLIP method in [52]. Here the number of image frames was chosen to be L = 200 and the under-sampling factor is 16. Moreover small complex noise was added to the under sampled Fourier data. The details can be found in [178] where also extensions of the methods are discussed.

Despite the practical success of the dictionary learning methods described above, there are still many open theoretical questions. First of all even for qualitative MRI the corresponding optimization problem (3.22) is a non-convex and non-differentiable program which depends on the initialization, the optimization algorithm, the parameter choice, and also the choice of the sampling pattern. Moreover it is a priori unclear if a solution of (3.22) or even more difficult a solution of (3.30) has anything to do with ground truth data-generating dictionary. This question of the so called identifiability has been discussed in [180, 71, 175, 84] but is in general still an open problem. Further theoretical progress has concentrated mostly on different facets of the classical dictionary learning problem in (3.20). In [181, 182] the structure of minimizers and stationary points is studied. More sophisticated minimization algorithms are for instance considered in e.g. [113, 120]. The question of whether alternating minimization schemes are able to recover the data generating dictionary is investigated in [8, 3, 160, 123] under some specific probabilistic model.

Remark 3.1 (Function space versions of dictionary learning). Most of the work on patch-based dictionary learning was done in the finite dimensional set-up, i.e. a pre-discretized image with a fixed number of pixels was assumed. Generalisations of patch-based dictionary learning are inherently difficult as the generalisation of the patch extraction operator is not clear. The straightforward way would first apply a discretization to an image in the function space X and cut the patches out of the discretization. But in this way R has a nullspace where no regularization acts at all and also the solution again depends on a pre selected discretization. Recently a variant of dictionary learning called convolutional dictionary learning was proposed which does not rely on patches but on kernels which are convolved against sparse coefficients [69, 150] to reconstruct the image. Convolutional dictionary learning admits a natural extension to infinite dimensions which is studied in [36]. In addition to these difficulties on the modeling side, also the extension of the convergence theory, in particular of the KL-inequality, towards infinite dimensions seems delicate. While there are some generalizations towards Hilbert spaces [22] and also to general metric spaces [89] available which are used to study the convergence of gradient flow problems, the application to inverse problems is rarely studied in the literature. This is due to the fact that the verification of the KL-inequality is more involved since its connection to real-semialgebraic geometry, as it is usually used for finite dimensional problems [12, 11], is to the best of our knowledge not available.



DOI 10.20347/WIAS.PREPRINT.3105

FIGURE 9. From left to right: T_1 , T_2 , ρ reconstruction. From top to bottom: Ground truth, BLIP, Levenberg-Marquardt, Dictionary learning approach.

Berlin 2024

3.3. Learning-informed physics in quantitative imaging.

3.3.1. *General set-up.* The quantitative methods for imaging discussed in Section 2.2 rely on the assumption that the equations that govern the physical constraints of the inverse problem are known a priori. This is a valid assumption if the governing physical model is derived from well-established physical arguments that are validated experimentally. Often, however, the physical equations are of phenomenological nature only and rooted in empiricism, such that it might be advisable to substitute the physical model or parts of it by learning-informed components, which are trained based on available physical data. *Learning-informed models* have the advantage that they are versatile in describing correlations and structures in given physical data. Consequently, utilizing learning-informed models reduces the risk of introducing modeling errors caused by false physical assumptions. Further, learninginformed models can reduce the computational cost associated with the evaluation of the models, e.g., when learning an efficient-to-evaluate control-to-state map for control problems like those described in Section 2.2. In the following, we discuss some recent advances of learning-informed physical constraints for general inverse problems and then in particular with respect to the qMRI problem.

We begin by considering a physics-constrained inverse problem, which is a general version of the integrated-physics qMRI approach (2.20) discussed in Section 2.2

(3.32)
$$\begin{cases} \min_{u,q\in\mathcal{C}_{ad}} \frac{1}{2} \|Au-y\|_2^2 + \mathcal{R}(\alpha;q), \\ \text{subject to} \quad e(u,q) = 0. \end{cases}$$



FIGURE 10. From left to right: Error-maps for the T_1 , T_2 and ρ reconstruction. Top to bottom: BLIP, Levenberg-Marquardt, Dictionary learning approach.



FIGURE 11. Learning a (partially) unknown physical process Π and embedding the learned map $\Pi_{\mathcal{N}}$ to the integrated-physics optimization problem.

Recall here that the (possibly nonlinear) physical constraint e(u,q) = 0, which may take the form of an ordinary or partial differential equation, relates the state variable u and the control variable q. We assume that the physical constraint e(u,q) = 0 is well-posed such that for a given control q, the state u that satisfies the physical constraint is unique, i.e., we can define (e.g., via the implicit function theorem) the control-to-state (or parameter-to-solution) map

$$(3.33) u = \Pi(q)$$

In Section 2.2, it was assumed that the physical constraint e(u,q) = 0 and the control-to-state operator Π are a priori known, and in particular it could be expressed by the Bloch equations. As aforementioned, in this section, we seek to substitute the physical model or parts of it by learninginformed components. Here we distinguish between two conceptually different strategies. The first strategy is to learn the physical model in its implicit form e(u,q) = 0 from data. If e(u,q) = 0 is an ordinary or partial differential equation, this could mean to learn the unknown differential operator or an unknown nonlinear term in the differential equation. This strategy has been followed by [174, 28, 166, 155, 156, 45, 59, 58, 60], to name only a few. When the target of learning is the implicit form of the physical model, computing the state for a given control is still needed after the learning has taken place and it usually requires solving a differential equation. The alternative, conceptually different strategy is to directly learn the explicit control-to-state operator Π from available data such that the computation of the state for a given control becomes computationally more efficient and closer to the true physical process. This approach was introduced in [59] and the overall idea is depicted in the diagram in Figure 11, The focus of the following discussions lies on this latter strategy because it is well-aligned with the integrated-physics approached discussed in Section 2.2. The strategy for substituting the physical model by a learning-informed ansatz is to parameterize the unknown controlto-state operator, i.e.,

(3.34)
$$u = \Pi(q) \approx \Pi_{\mathcal{N}}(\theta; q),$$

where $\theta \in \mathbb{R}^{\ell}$ is a set of ℓ tunable parameters, for example the weights of a neural network. If labeled data pairs (q_i, u_i) with $i \in \{1, \ldots, N\}$ are available, the tunable parameters can be calibrated by minimizing the mismatch between model prediction and data in a supervised learning fashion

(3.35)
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \|\Pi(\theta; q_i) - u_i\|_2^2 + r(\theta).$$

where r denotes some potential regularization for θ reflecting prior information which one wants to impose on θ . Also, in some cases additional constraints on θ might be relevant. Solving (3.35) is often a very challenging tasks, giving rise to many current efforts in solver design. Also, non-smooth variants of (3.35) are relevant in practice.

3.3.2. *Learning the Bloch solution map for qMRI.* We turn back to the physics-constrained inverse problem in (3.32) and consider it in the context of qMRI, i.e., we consider the Bloch equations as the physical constraint (see Section 2.2) where, following the terminology from optimal control, the state variable is the magnetization and the control variables are the physical tissue parameters. The literature reports different attempts to tackle the Bloch equations with machine learning. These attempts can be classified into approaches that seek to learn the Bloch solution map or its inverse, or both through encoder-decoder type models, as discussed in the following.

Learning the inverse of the Bloch solution map can, for example, be used to efficiently identify the physical tissue parameters given the magnetization response, by leveraging deep neural networks as demonstrated by [104, 187, 47, 143, 76, 105]. For instance, the work in [47] introduced the framework DRONE (Deep RecOnstruction NEtwork), in which the inverse Bloch solution map is approximated by a neural network that is trained on a simulated MRF look-up-table. The trained network requires less memory and allows for faster predictions than the MRF look-up-table. In the other direction, neural networks can for example be used to efficiently generate the Bloch look-up-table for MRF as shown by [194] using generative adversarial neural networks (GANs), by [87] using feed-forward ANNs, and by [126] using recurrent neural networks (RNNs).

Another approach in this context, aligned with Figure 11, is to learn the Bloch solution map and substitute it in the reduced (integrated physics) formulation given in (2.21). This idea was realized in [59] where feed-forward ANNs have been utilized for learning the control-to-state operator $\mathcal{N}(q) \approx \Pi(q)$. This resulted in the following learning-informed optimization problem:

(3.36)
$$\min_{q \in \mathcal{C}_{ad}} \frac{1}{2} \|A \circ \mathcal{N}(q) - y\|_2^2 + \frac{\alpha}{2} \|\nabla q\|_2^2,$$

where for simplicity a standard Tikhonov regularisation was employed for q albeit with small α in order to stabilize the solution process. In Algorithm 4, we summarize the above procedure. We note that in this approach the neural network \mathcal{N} is trained in an *offline phase*, and then embedded into the minimization problem (3.36). However, integrated learning techniques, i.e., training \mathcal{N} while solving (3.36), are conceivable.

Algorithm 4 A sketch of the learning-informed models for quantitative imaging

- (1) Collect a set of training data which are resulting from the targeting physical models with (experimental or numerical) measurements.
- (2) Build a neural network architecture to learn this physical model using the training data from Step (1).
- (3) Solve the optimization problem (3.36) with learning informed operators or constraints with learning informed differential equations.

Remark 3.2. The description of Algorithm 4 looks quite abstract. Here we take an example to explain the ideas with more details. For our qMRI example, the training data are collected from the dictionary which is introduced in the MRF paper [132], see also [57]. The data consist of time series which mimic the discretization of the Bloch equations, where the input is the tissue parameter, and the output is the magnetization. Using this training data, we can learn the map from tissue parameters to magnetization (it is a type of Nemytskii operator in the case of [132]). However, the learning approach can be more general in the case where the Bloch equations are considered to be a simplification of a more complicated physical process, and more data are available for the training. The architecture of the neural networks are popular, see a description in [59], which is also the one we have used in the numerical examples of this review. In terms of the third step, there are typically many ways to solve the optimization problem. In the considered example, we have used a semismooth Newton type algorithm

[94] to take care of the box constraints for the tissue parameters. For more details one can also refer to [59].

We present some numerical examples in Figure 12, where we compute solutions of the optimization problem (3.36). In this example, the neural network simulating the Bloch map is trained via a training set which was obtained through discrete dynamics of simulated Bloch data. In particular, a set of parameter pairs $\{q_i\}_{i=1}^N$ in the feasible domain of T_1 and T_2 is taken as inputs of the network, and then the simulated Bloch dynamics resulting from all these pairs are regarded as the outputs. The simulated k-space data in this example are generated using exactly the same setting as the one used for Figures 3 and 4. For such a discrete time series, the Bloch map is a Nemytskii type operator, and thus can be approximated sufficiently well in a uniformly bounded feasible set.



FIGURE 12. The estimated parameters T_1 , T_2 and ρ using the optimization framework with learning-informed models. Top row are the estimated parameters, and bottom row are the relative errors. By comparing the result with the ones in Figures 3 and 4, we see that using learning-informed operator, the obtained results are comparable, and the computational time is less in solving the optimization problem, since the learning is done before hand, and heavy evaluation of the equation is avoided in every iteration.

Remark 3.3 (Learning-informed qMRI in function space). The learning-informed optimization problem for qMRI (3.36) can be also posed and analyzed in a function space setting [59]. The key observation here is that the Bloch solution map Π and its approximation \mathcal{N} can be regarded as Nemytskii operators, with a voxel-wise action, i.e., $\hat{\Pi}, \hat{\mathcal{N}} : \mathcal{C}_{ad} \to [(L^{\infty}(\Omega))^3]^L$:

- (3.37) $\hat{\Pi}(\hat{\rho}, \hat{T}_1, \hat{T}_2)(x) = \Pi(\rho(x), T_1(x), T_2(x))$
- (3.38) $\hat{\mathcal{N}}(\hat{\rho}, \hat{T}_1, \hat{T}_2)(x) = \mathcal{N}(\rho(x), T_1(x), T_2(x))$

for almost every $x \in \Omega$. Here C_{ad} is a subset of $[L^{\infty}(\Omega)^+]^3$, i.e. L^{∞} -functions that are bounded away from zero, denoting feasible values for $\hat{\rho}, \hat{T}_1, \hat{T}_2 \in L^{\infty}(\Omega)$. Note that $\hat{\mathcal{N}}$ can be in principle learned via the network $\mathcal{N} \approx \Pi$ with inputs in \mathbb{R}^3 . Under this setting, the analysis of (3.36) in the function space setting can be performed. For instance in [59] convergence of the solution of the learning-informed problem to the solution of the problem given in (2.21) was proved upon increasing the approximation capability of the low dimensional network \mathcal{N} and corresponding error estimates were also provided.

Towards learning infinite dimensional physics operators for gMRI. Conventionally, the universal approximation theorem [106] considers neural network functions between finite dimensional spaces. The idea of studying neural networks as approximators of functionals on infinite dimensional spaces has been initiated by [42], where two architectures were introduced; one for approximating functionals (see also the related earlier work [170]) and one for approximating operators. The number of parameters in these models can be substantially larger than in a standard feed-forward network. Thus, training such a model requires more computational resources; and [42] did not provide a practical training example. In recent years, as computers and optimizers for machine learning models have become more and more powerful, a blossoming of the idea of learning functionals or operators is observed [128, 19, 122, 117, 157]. Prominent ANN architectures for learning operators between infinite-dimensional spaces are DeepONets [128], which are inspired by the architecture presented in [42], and (Fourier) Neural Operators. We see the opportunities of operator learning within the problems setting of the current paper, where the qMRI physical model can be realized as an operator as well. Once a mature operator learning scheme has been developed, it will not only improve the modeling process but also contribute efficient solvers. Note that more realistic qMRI dynamics may involving partial differential equations. In this context, operator learning is an even more recent topic and one may refer to the papers [128, 19, 122, 117, 157] and the references therein.

4. TOWARDS FAIR DATA

Similar to all mathematical research, the imaging field is intrinsically tied to research data. This encompasses the image data y used for applying the mathematical methods, its parameter sets, and even the algorithms themselves along with intricate workflows employed for image processing. The digitalization trend has spurred substantial data growth. Particularly evident in data-driven methodologies, this data's availability has become fundamental for research. However, overlooking effective research data management could result in "dark data" [172], posing challenges to research reproducibility and contributing to the broader issue known as the reproducibility crisis in science [14].

In response to these challenges and the overarching call for open science and open data, Tim Berners-Lee's 5-star principles [66] and the more recent FAIR (Findable, Accessible, Interoperable, Reusable) principles [191] have emerged. These principles provide guidance and categorization of compliance levels for research output and data. Similarly, the complementary FAIR4RS [46] principles have been formulated to address the specific requirements of research software. Despite their distinct characteristics, open data and FAIR principles share a unifying objective: enhancing the accessibility of research data, much like how journal publications have been provided to the scientific community, facilitating the practice of building upon existing knowledge. Open data and opensource software should be accessible to all under fitting open licenses. In contrast, the FAIR principles delineate the methods and conditions for human and machine data access. Moreover, they introduce stringent prerequisites for metadata descriptions and formats, enhancing findability, interoperability, and potential reusability.

Specifically, findability encompasses assigning a distinct identifier, furnishing comprehensive metadata descriptions, and indexing the data within a resource that's both searchable by humans and machines, streamlining researcher access. Accessibility ensures that researchers with interest can approach the data using standardized, open, and cost-free communication protocols. The metadata remains perpetually accessible, even if the data itself isn't or becomes inaccessible over time. Interoperability empowers data use across different contexts from their original purpose. Reusability necessitates meticulous documentation aligned with community standards, facilitating reproducibility and enabling other researchers to utilize a dataset for their own applications. This reusability also mandates releasing the data with a fitting license.

In an attempt of a community-driven approach to make research data FAIR and provide appropriate infrastructures for storage, collection of the scientific knowledge, and innovative services to enable new scientific results, the German government decided to fund the building of a National Research Data Infrastructure [141]. In particular, the community approach is realized by discipline-specific consortia that address the needs of the respective area in science. The consortium for mathematics, the Mathematical Research Data Initiative (MaRDI)[184, 135, 134], has started its work in 2021 and is gradually releasing new services to the mathematical community such as specific data repositories, services for creating and maintaining a data management plan and a MaRDI knowledge graph to connect the mathematical knowledge, software, publications, and data.

Mathematical research projects in the area of imaging may benefit from an existing data infrastructure while also contributing to its contents. This already starts at the proposal phase, where an appropriate research data management plan has to be developed in accordance with the requirements of the funding agency [183]. Besides the description of the data used and created within the project such a plan usually contains the elaboration on its documentation, e.g., with rich metadata, thoughts on storage in appropriate data repositories, assignment of suitable licenses, descriptions of availability to third-parties and responsibilities within the project, and more general measures to increase the FAIRness of the project data and software. Such a research data management plan is usually a living document adjusted to the needs of the project over time.

Projects for mathematical imaging general should consider research data such as: images used for algorithmic evaluation, collections of images used for training and testing, software scripts for processing examples, implementations of algorithms in some more general-use software, knowledge about the algorithm and related publications and software as contribution to the MaRDI knowledge graph. Naturally, some of these data are generated within the project, others are often re-used from other available resources, such as OpenNeuro [145], the Human Connectome Project [109], the Alzheimer's Disease Neuroimaging Initiative [5], or [24] for simulated brain data in the field of neuroimaging. It should be considered natural for research project to publish its generated data and software in a FAIR manner as described above in the same way as it is common for the scientific articles describing and discussing the scientific results. This will increase visibility of the research, enhance its reproducibility, prevent re-invention of the wheel by duplicated efforts for data, software, and knowledge creation, and lower the burden to start new research. **Acknowledgments.** GD acknowledges the support from NSF of Hunan Province (project No. 2024JJ5413). This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - MATH+: The Berlin Mathematics Research Center [EXC-2046/1 - project ID: 390685689].

Author contributions. The various authors have made the following contributions according to the CRediT-taxonomy.

Guozhi Dong: Writing-original draft, Writing-review & editing, Software, Conceptualization, **Moritz Flaschel**: Writing-original draft, **Michael Hintermüller**: Project administration, Supervision, Writing-review & editing, Conceptualization, Funding acquisition, **Kostas Papafitsoros**: Project administration, Supervision, Writing-original draft, Writing-review & editing, Conceptualization **Clemens Sirotenko**: Writingoriginal draft, Writing-review & editing, Software, Conceptualization, **Karsten Tabelow**: Writing-original draft, Writing-review & editing, Software.

Financial disclosure. None reported.

Conflict of interest. The authors declare no potential conflict of interests.

REFERENCES

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.
- [2] Pierre Ablin, Thomas Moreau, Mathurin Massias, and Alexandre Gramfort. Learning step sizes for unfolded sparse coding. *Advances in Neural Information Processing Systems*, 32, 2019.
- [3] Alekh Agarwal, Animashree Anandkumar, Prateek Jain, and Praneeth Netrapalli. Learning sparsely used overcomplete dictionaries via alternating minimization. *SIAM Journal on Optimization*, 26(4):2775–2799, 2016.
- [4] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.
- [5] Alzheimer's Disease Neuroimaging Initiative. https://adni.loni.usc.edu.
- [6] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford University Press, USA, 2000.
- [7] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of Al. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.
- [8] Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra. Simple, efficient, and neural algorithms for sparse coding. In *Conference on learning theory*, pages 113–149. PMLR, 2015.
- [9] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using datadriven models. *Acta Numerica*, 28:1–174, May 2019.
- [10] Jakob Assländer, Martijn A Cloos, Florian Knoll, Daniel K Sodickson, Jürgen Hennig, and Riccardo Lattanzi. Low rank alternating direction method of multipliers reconstruction for MR fingerprinting. *Magnetic resonance in medicine*, 79(1):83–96, 2018.
- [11] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of* operations research, 35(2):438–457, 2010.
- [12] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.
- [13] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. *Advances in Neural Information Processing Systems*, 32, 2019.
- [14] Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533:452–454, 2016.
- [15] Annalisa Baldi. Weighted bv functions. Houston Journal of Mathematics, 27(3):683-705, 2001.
- [16] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of Marchine Learning Research*, 18:1–43, 2018.

- [17] Amir Beck. First-order methods in optimization. SIAM, 2017.
- [18] Martin Benning and Martin Burger. Modern regularization methods for inverse problems. Acta Numerica, 27:1–111, 2018.
- [19] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model Reduction And Neural Networks For Parametric PDEs. *The SMAI journal of computational mathematics*, 7:121–157, July 2021.
- [20] Berkin Bilgic, Itthi Chatnuntawech, Mary Kate Manhard, Qiyuan Tian, Congyu Liao, Siddharth S Iyer, Stephen F Cauley, Susie Y Huang, Jonathan R Polimeni, Lawrence L Wald, et al. Highly accelerated multishot echo planar imaging through synergistic machine learning and joint reconstruction. *Magnetic resonance in medicine*, 82(4):1343– 1358, 2019.
- [21] Felix Bloch. Nuclear induction. Physical review, 70(7-8):460, 1946.
- [22] Jérôme Bolte, Aris Daniilidis, Olivier Ley, and Laurent Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.
- [23] Jérôme Bolte, Tam Le, Edouard Pauwels, and Tony Silveti-Falls. Nonsmooth implicit differentiation for machinelearning and optimization. Advances in neural information processing systems, 34:13537–13549, 2021.
- [24] Brainweb: Simulated brain database. http://www.bic.mni.mcgill.ca/brainweb/.
- [25] Kristian Bredies, Yiqiu Dong, and Michael Hintermüller. Spatially dependent regularization parameter selection in total generalized variation models for image restoration. *International Journal of Computer Mathematics*, 90(1):109– 123, 2013.
- [26] Kristian Bredies and Martin Holler. Higher-order total variation approaches and generalisations. *Inverse Problems*, 36(12):123001, dec 2020.
- [27] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. SIAM Journal on Imaging Sciences, 3(3):492–526, 2010.
- [28] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932– 3937, April 2016.
- [29] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4:490–530, 2005.
- [30] Martin Burger and Stanley Osher. A guide to the TV zoo. In Martin Burger and Stanley Osher, editors, Level Set and PDE Based Reconstruction Methods in Imaging, pages 1–70. Springer International Publishing, 2013.
- [31] Luca Calatroni, Chung Cao, Juan Carlos De Los Reyes, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Bilevel approaches for learning of variational imaging models. *Variational Methods: In Imaging and Geometric Control*, 18(252):2, 2017.
- [32] Emmanuel Candes, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.
- [33] Xiaozhi Cao, Congyu Liao, Siddharth Srinivasan Iyer, Zhixing Wang, Zihan Zhou, Erpeng Dai, Gilad Liberman, Zijing Dong, Ting Gong, Hongjian He, et al. Optimized multi-axis spiral projection mr fingerprinting with subspace reconstruction for rapid whole-brain high-isotropic-resolution quantitative imaging. *Magnetic Resonance in Medicine*, 88(1):133–150, 2022.
- [34] Vicent Caselles, Antonin Chambolle, and Matteo Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Modeling & Simulation*, 6(3):879–894, 2007.
- [35] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In Massimo Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, pages 263–340. De Gruyter, Berlin, New York, 2010.
- [36] Antonin Chambolle, Martin Holler, and Thomas Pock. A convex variational model for learning convolutional image atoms from incomplete data. *Journal of Mathematical Imaging and Vision*, 62(3):417–444, 2020.
- [37] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
- [38] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40:120–145, 2011.
- [39] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.
- [40] Ke Chen, Yiqiu Dong, and Michael Hintermüller. A nonlinear multigrid solver with line Gauss-Seidel-semismooth-Newton smoother for the Fenchel pre-dual in total variation based image restoration. *Inverse Problems and Imaging*, 5(2):323–339, 2011.
- [41] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Zhangyang Wang, Howard Heaton, Jialin Liu, and Wotao Yin. Learning to optimize: A primer and a benchmark. *The Journal of Machine Learning Research*, 23(1):8562–8620, 2022.

- [42] Tianping Chen and Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, July 1995.
- [43] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds. *Advances in Neural Information Processing Systems*, 31, 2018.
- [44] Yong Chen, Zhenghan Fang, Sheng-Che Hung, Wei-Tang Chang, Dinggang Shen, and Weili Lin. High-resolution 3D MR fingerprinting using parallel imaging and deep learning. *Neuroimage*, 206:116329, 2020.
- [45] Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nature Communications*, 12(1):6136, December 2021.
- [46] Neil P Chue Hong, Daniel S Katz, Michelle Barker, Anna-Lena Lamprecht, Carlos Martinez, Fotis E Psomopoulos, Jen Harrow, Leyla J Castro, Morane Gruenpeter, Paula A Martinez, Tom Honeyman, et al. FAIR principles for research software (FAIR4RS principles). Research Data Alliance, 2021.
- [47] Ouri Cohen, Bo Zhu, and Matthew S Rosen. MR fingerprinting Deep RecOnstruction NEtwork (DRONE). Magnetic Resonance in Medicine, 80(3):885–894, 2018.
- [48] D L Collins, A P Zijdenbos, V Kollokian, J G Sled, N J Kabani, C J Holmes, and A C Evans. Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging*, 17(3):463–468, Jun 1998.
- [49] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.
- [50] Jérôme Darbon and Marc Sigelle. Image restoration with discrete constrained total variation. I. Fast and exact optimization. Journal of Mathematical Imaging and Vision, 26(3):261–276, 2006.
- [51] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.
- [52] Mike Davies, Gilles Puy, Pierre Vandergheynst, and Yves Wiaux. A compressed sensing framework for magnetic resonance fingerprinting. SIAM Journal on Imaging Sciences, 7(4):2623–2656, 2014.
- [53] Elisa Davoli, Rita Ferreira, Irene Fonseca, and José A Iglesias. Dyadic partition-based training schemes for tv/tgv denoising. arXiv preprint arXiv:2305.07150, 2023.
- [54] Juan Carlos De los Reyes and David Villacís. Bilevel optimization methods in imaging. In Ke Chen, Carola-Bibiane Schönlieb, Xue-Cheng Tai, and Laurent Younces, editors, *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1–34. Springer International Publishing, Cham, 2021.
- [55] Juan Carlos De los Reyes and David Villacís. Optimality conditions for bilevel imaging learning problems with total variation regularization. *SIAM Journal on Imaging Sciences*, 15(4):1646–1689, 2022.
- [56] Tian Ding, Kui Ren, and Sarah Vallelian. A one-step reconstruction algorithm for quantitative photoacoustic imaging. *Inverse Problems*, 31:095005, 2015.
- [57] Guozhi Dong, Michael Hintermüller, and Kostas Papafitsoros. Quantitative magnetic resonance imaging: From fingerprinting to integrated physics-based models. SIAM Journal on Imaging Sciences, 12(2), 2019.
- [58] Guozhi Dong, Michael Hintermüller, and Kostas Papafitsoros. A descent algorithm for the optimal control of ReLU neural network informed PDEs based on approximate directional derivatives. arXiv preprint arXiv:2210.07900, 2022.
- [59] Guozhi Dong, Michael Hintermüller, and Kostas Papafitsoros. Optimization with learning-informed differential equation constraints and its applications. ESAIM: COCV, 28:3, 2022.
- [60] Guozhi Dong, Michael Hintermüller, Kostas Papafitsoros, and Kathrin Völkner. First-order conditions for the optimal control of learning-informed nonsmooth PDEs. *arXiv:2206.00297*, 2022.
- [61] Yiqiu Dong, Michael Hintermüller, and M Monserrat Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40(1):82– 104, 2011.
- [62] Yiqiu Dong, Michael Hintermüller, and M Monserrat Rincon-Camacho. A multi-scale vectorial L^τ-TV framework for color image restoration. International Journal of Computer Vision, 92(3):296–307, 2011.
- [63] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [64] Peter Elbau, Leonidas Mindrinos, and Otmar Scherzer. Quantitative reconstructions in multi-modal photoacoustic and optical coherence tomography imaging. *Inverse Problems*, 34(1):014006, 2018.
- [65] R R Ernst and W A Anderson. Application of Fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments*, 37:93–102, 1966.
- [66] Five stars of open data. https://opendatahandbook.org/glossary/en/terms/ five-stars-of-open-data/.
- [67] Simon Foucart, Holger Rauhut, Simon Foucart, and Holger Rauhut. *An invitation to compressive sensing*. Springer, 2013.

- [68] Samy Wu Fung, Howard Heaton, Qiuwei Li, Daniel McKenzie, Stanley Osher, and Wotao Yin. JFB: Jacobian-free backpropagation for implicit networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(6):6648– 6656, 2022.
- [69] Cristina Garcia-Cardona and Brendt Wohlberg. Convolutional dictionary learning: A comparative review and new algorithms. *IEEE Transactions on Computational Imaging*, 4(3):366–381, 2018.
- [70] Ruturaj G Gavaskar and Kunal N Chaudhury. Plug-and-play ISTA converges with kernel denoisers. IEEE Signal Processing Letters, 27:610–614, 2020.
- [71] Quan Geng and John Wright. On the local correctness of ℓ_1 -minimization for dictionary learning. In 2014 IEEE International Symposium on Information Theory, pages 3180–3184. IEEE, 2014.
- [72] Davis Gilton, Gregory Ongie, and Rebecca Willett. Deep equilibrium architectures for inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 7:1123–1133, 2021.
- [73] Enrico Giusti. *Minimal surfaces and functions of bounded variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1984.
- [74] Sivan Gleichman and Yonina C Eldar. Blind compressed sensing. *IEEE Transactions on Information Theory*, 57(10):6958–6975, 2011.
- [75] Mohammad Golbabaee, Guido Buonincontri, Carolin M Pirkl, Marion I Menzel, Bjoern H Menze, Mike Davies, and Pedro A Gómez. Compressive MRI quantification using convex spatiotemporal priors and deep encoder-decoder networks. *Medical Image Analysis*, 69:101945, 2021.
- [76] Mohammad Golbabaee, Dongdong Chen, Pedro A Gomez, Marion I Menzel, and Mike E Davies. Geometry of Deep Learning for Magnetic Resonance Fingerprinting. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 7825–7829, Brighton, United Kingdom, 2019. IEEE.
- [77] Mohammad Golbabaee, Zhouye Chen, Yves Wiaux, and Mike Davies. CoverBLIP: accelerated and scalable iterative matched-filtering for magnetic resonance fingerprint reconstruction. *Inverse Problems*, 36(1):015003, 2019.
- [78] Gene H Golub and Charles F Van Loan. Matrix computations. JHU press, 2013.
- [79] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
- [80] J W Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11):1145, 1976.
- [81] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. Sparse regularization with l^q penalty term. Inverse Problems, 24(5):055020, 2008.
- [82] Markus Grasmair, Otmar Scherzer, and Markus Haltmeier. Necessary and sufficient conditions for linear convergence of l^1 -regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011.
- [83] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010.
- [84] Rémi Gribonval, Rodolphe Jenatton, and Francis Bach. Sparse and spurious: dictionary learning with noise and outliers. IEEE Transactions on Information Theory, 61(11):6298–6319, 2015.
- [85] A Haase, J Frahm, D Matthaei, W Hanicke, and K-D Merboldt. FLASH imaging. rapid NMR imaging using low flipangle pulses. *Journal of Magnetic Resonance*, 67(2):258 – 266, 1986.
- [86] Andreas Habring and Martin Holler. Neural-network-based regularization methods for inverse problems in imaging. arXiv preprint arXiv:2312.14849, 2023.
- [87] Jesse I Hamilton and Nicole Seiberlich. Machine Learning for Rapid Magnetic Resonance Fingerprinting Tissue Property Quantification. *Proceedings of the IEEE*, 108(1):69–85, January 2020.
- [88] Paul Kyu Han, Thibault Marin, Yanis Djebra, Vanessa Landes, Yue Zhuo, Georges El Fakhri, and Chao Ma. Freebreathing 3D cardiac T1 mapping with transmit B1 correction at 3T. *Magnetic Resonance in Medicine*, 87(4):1832– 1845, 2022.
- [89] Daniel Hauer and José Mazón. Kurdyka–Łojasiewicz–simon inequality for gradient flows in metric spaces. Transactions of the American Mathematical Society, 372(7):4917–4976, 2019.
- [90] Andreas Hauptmann, Subhadip Mukherjee, Carola-Bibiane Schönlieb, and Ferdia Sherry. Convergent regularization in inverse problems and linear plug-and-play denoisers. *arXiv preprint arXiv:2307.09441*, 2023.
- [91] Gunther Helms, Henning Dathe, and Peter Dechent. Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magnetic Resonance in Medicine*, 59:667–672, 2008.
- [92] Walter Hinterberger, Michael Hintermüller, Karl Kunisch, Markus von Oehsen, and Otmar Scherzer. Tube methods for BV regularization. *Journal of Mathematical Imaging and Vision*, 19(3):219–235, 2003.
- [93] Michael Hintermüller, Martin Holler, and Kostas Papafitsoros. A function space framework for structural total variation regularization with applications in inverse problems. *Inverse Problems*, 34(6):064002, 2018.
- [94] Michael Hintermüller, Kazufumi Ito, and Karl Kunisch. The primal-dual active set strategy as a semismooth Newton method. SIAM Journal on Optimization, 13(3):865–888, 2003.
- [95] Michael Hintermüller and Ian Kopacka. Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm. SIAM Journal on Optimization, 20(2):868–902, 2009.

- [96] Michael Hintermüller and Carlos N Rautenberg. Optimal selection of the regularization function in a weighted total variation model. Part I: Modelling and theory. *Journal of Mathematical Imaging and Vision*, 59(3):498–514, 2017.
- [97] Michael Hintermüller, Carlos N Rautenberg, Tao Wu, and Andreas Langer. Optimal selection of the regularization function in a weighted total variation model. Part II: Algorithm, its analysis and numerical tests. *Journal of Mathematical Imaging and Vision*, 59(3):515–533, 2017.
- [98] Michael Hintermüller and M Monserrat Rincon-Camacho. Expected absolute value estimators for a spatially adapted regularization parameter choice rule in L^1 -TV-based image restoration. *Inverse Problems*, 26(8):085005, 30, 2010.
- [99] Michael Hintermüller and Georg Stadler. An infeasible primal-dual algorithm for total bounded variation-based infconvolution-type image restoration. SIAM Journal on Scientific Computing, 28(1):1–23, 2006.
- [100] Michael Hintermuüller and Tao Wu. Nonconvex TV^{*q*}-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver. *SIAM Journal on Imaging Sciences*, 6(3):1385–1415, 2013.
- [101] Michael Hintermüller and Kostas Papafitsoros. Generating structured nonsmooth priors and associated primal-dual methods. In *Handbook of Numerical Analysis*, volume 20, pages 437–502. Elsevier, 2019.
- [102] Michael Hintermüller, Kostas Papafitsoros, Carlos N Rautenberg, and Hongpeng Sun. Dualization and automatic distributed parameter selection of total generalized variation via bilevel optimization. *Numerical Functional Analysis* and Optimization, 43(8):887–932, 2022.
- [103] Sebastian Hirsch, Jürgen Braun, and Ingolf Sack. *Magnetic Resonance Elastography: Physical Background And Medical Applications*. Wiley, 2017.
- [104] Elisabeth Hoppe, Gregor Körzdörfer, Tobias Würfl, Jens Wetzl, Felix Lugauer, Josef Pfeuffer, and Andreas Maier. Deep Learning for Magnetic Resonance Fingerprinting: A New Approach for Predicting Quantitative Parameter Values from Time Series. In *Volume 243: German Medical Data Sciences: Visions and Bridges*, Studies in Health Technology and Informatics, pages 202–206. IOS Press, 2017.
- [105] Elisabeth Hoppe, Florian Thamm, Gregor Körzdörfer, Franziska Schirrmacher, Mathias Nittka, Josef Pfeuffer, Heiko Meyer, and Andreas Maier. Magnetic Resonance Fingerprinting Reconstruction Using Recurrent Neural Networks. In *Volume 267: German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine*, Studies in Health Technology and Informatics, pages 126–133. IOS Press, 2019.
- [106] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989.
- [107] Seyed Amir Hossein Hosseini, Burhaneddin Yaman, Steen Moeller, Mingyi Hong, and Mehmet Akçakaya. Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms. *IEEE Journal* of Selected Topics in Signal Processing, 14(6):1280–1291, 2020.
- [108] Chuan Huang, Christian G Graff, Eric W Clarkson, Ali Bilgin, and Maria I Altbach. T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing. *Magnetic Resonance in Medicine*, 67(5):1355–1366, 2012.
- [109] Human Connectome Project. https://www.humanconnectome.org.
- [110] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. arXiv preprint arXiv:2110.03220, 2021.
- [111] IXI dataset. https://brain-development.org/ixi-dataset/.
- [112] Ulugbek S Kamilov, Charles A Bouman, Gregery T Buzzard, and Brendt Wohlberg. Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97, 2023.
- [113] Le Thi Khanh Hien, Duy Nhat Phan, Nicolas Gillis, Masoud Ahookhosh, and Panagiotis Patrinos. Block Bregman majorization minimization with extrapolation. SIAM Journal on Mathematics of Data Science, 4(1):1–25, 2022.
- [114] Florian Knoll, Kristian Bredies, Thomas Pock, and Rudolf Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–491, 2011.
- [115] Andreas Kofler, Fabian Altekrüger, Fatima Antarou Ba, Christoph Kolbitsch, Evangelos Papoutsellis, David Schote, Clemens Sirotenko, Felix Frederik Zimmermann, and Kostas Papafitsoros. Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling. *SIAM Journal on Imaging Sciences*, 16(4):2202–2246, 2023.
- [116] Andreas Kofler, Kirsten Miriam Kerkering, Laura Göschel, Ariane Fillmer, and Christoph Kolbitsch. Quantitative MR image reconstruction using parameter-specific dictionary learning with adaptive dictionary-size and sparsity-level choice. *IEEE Transactions on Biomedical Engineering*, 2023.
- [117] Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On Universal Approximation and Error Bounds for Fourier Neural Operators. *Journal of Machine Learning Research*, (22):1–76, 2021.
- [118] S Kullback and R A Leibler. On information and sufficiency. The Annals of Mathematical Statistics, 22(1):79–86, March 1951.
- [119] R K Kwan, A C Evans, and G B Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11):1085–1097, 1999.

- [120] Jiaxiang Li, Shiqian Ma, and Tejes Srivastava. A Riemannian ADMM. arXiv preprint arXiv:2211.02163, 2022.
- [121] Shutao Li, Haitao Yin, and Leyuan Fang. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Transactions on biomedical engineering*, 59(12):3450–3459, 2012.
- [122] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.
- [123] Geyu Liang, Gavin Zhang, Salar Fattahi, and Richard Y Zhang. Simple alternating minimization provably solves complete dictionary learning. *arXiv preprint arXiv:2210.12816*, 2022.
- [124] Dana J Lin, Patricia M Johnson, Florian Knoll, and Yvonne W Lui. Artificial intelligence for MR image reconstruction: an overview for clinicians. *Journal of Magnetic Resonance Imaging*, 53(4):1015–1028, 2021.
- [125] Sajan Goud Lingala and Mathews Jacob. Blind compressive sensing dynamic MRI. IEEE transactions on medical imaging, 32(6):1132–1145, 2013.
- [126] Hongyan Liu, Oscar van der Heide, Cornelis A. T. van den Berg, and Sbrizzi Sbrizzi. Fast and accurate modeling of transient-state, gradient-spoiled sequences by recurrent neural networks. NMR in Biomedicine, 2021.
- [127] Jialin Liu and Xiaohan Chen. Alista: Analytic weights are as good as learned weights in LISTA. In *International Conference on Learning Representations (ICLR)*, 2019.
- [128] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218– 229, 2021.
- [129] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2):102–127, 2019.
- [130] Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph. *Mathematical programs with equilibrium constraints*. Cambridge University Press, Cambridge, 1996.
- [131] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing MRI. IEEE Signal Processing Magazine, 25(2):72–82, 2008.
- [132] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey Sunshine, Jeffrey L Duerk, and Mark A Griswold. Magnetic resonance fingerprinting. *Nature*, 495(187):187–193, 2013.
- [133] Stéphane Mallat. A wavelet tour of signal processing. Elsevier, 1999.
- [134] MaRDI Portal. https:/portal.mardi4nfdi.de.
- [135] Mathematical research data initiative (MaRDI). https://www.mardi4nfdi.de.
- [136] Gal Mazor, Lior Weizman, Assaf Tal, and Yonina C Eldar. Low-rank magnetic resonance fingerprinting. *Medical Physics*, 45(9):4066–4084, 2018.
- [137] Debra F McGivney, Eric Pierre, Dan Ma, Yun Jiang, Haris Saybasili, Vikas Gulani, and Mark A Griswold. SVD compression for magnetic resonance fingerprinting in the time domain. *IEEE transactions on medical imaging*, 33(12):2311–2322, 2014.
- [138] Yves Meyer. Oscillating patterns in image processing and nonlinear evolution equations, volume 22 of University Lecture Series. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.
- [139] Siawoosh Mohammadi, Chiara D'Alonzo, Lars Ruthotto, Jörg Polzehl, Isabel Ellerbrock, Martina F Callaghan, Nikolaus Weiskopf, and Karsten Tabelow. Simultaneous adaptive smoothing of relaxometry and quantitative magnetization transfer mapping. Preprint 2432, WIAS, 2017.
- [140] Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.
- [141] Nationale Forschungsdateninfrastruktur.
- [142] Frank Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, 2001.
- [143] Ilkay Oksuz, Gastao Cruz, James Clough, Aurelien Bustin, Nicolo Fuin, Rene M Botnar, Claudia Prieto, Andrew P King, and Julia A Schnabel. Magnetic Resonance Fingerprinting using Recurrent Neural Networks, 2018. arXiv:1812.08155.
- [144] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.
- [145] OpenNEURO. https://openneuro.org.
- [146] Jiři Outrata, Michal Kočvara, and Jochem Zowe. Nonsmooth approach to optimization problems with equilibrium constraints, volume 28 of Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, 1998. Theory, applications and numerical results.
- [147] Valerio Pagliari, Kostas Papafitsoros, Bogdan Raibtă, and Andreas Vikelis. Bilevel training schemes in imaging for total variation-type functionals with convex integrands. SIAM Journal on Imaging Sciences, 15(4):1690–1728, 2022.

- [148] Marie-Christine Pali, Tobias Schaeffter, Christoph Kolbitsch, and Andreas Kofler. Adaptive sparsity level and dictionary size estimation for image reconstruction in accelerated 2D radial cine MRI. *Medical Physics*, 48(1):178–192, 2021.
- [149] Kostas Papafitsoros and Kristian Bredies. A study of the one dimensional total generalised variation regularisation problem. *Inverse Problems and Imaging*, 9(2), 2015.
- [150] Vardan Papyan, Jeremias Sulam, and Michael Elad. Working locally thinking globally: Theoretical guarantees for convolutional sparse coding. *IEEE Transactions on Signal Processing*, 65(21):5687–5701, 2017.
- [151] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32, 2019.
- [152] Jörg Polzehl, Kostas Papafitsoros, and Karsten Tabelow. Patch-wise adaptive weights smoothing. Journal of Statistical Software, 95:1–27, 2020.
- [153] Jörg Polzehl and Vladimir Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135:335–362, 2006.
- [154] Jörg Polzehl and Karsten Tabelow. *Magnetic Resonance Brain Imaging: Modelling and Data Analysis Using R.* Springer International Publishing, 2023.
- [155] Maziar Raissi. Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations. Journal of Machine Learning Research, 19:1–24, 2018.
- [156] Maziar Raissi and George Em Karniadakis. Hidden Physics Models: Machine Learning of Nonlinear Partial Differential Equations. Journal of Computational Physics, 357:125–141, 2018.
- [157] Bogdan Raonić, Roberto Molinaro, Tobias Rohner, Siddhartha Mishra, and Emmanuel de Bezenac. Convolutional Neural Operators, 2023. arXiv:2302.01178.
- [158] Saiprasad Ravishankar and Yoram Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, 2010.
- [159] Saiprasad Ravishankar and Yoram Bresler. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM Journal on Imaging Sciences*, 8(4):2519–2557, 2015.
- [160] Saiprasad Ravishankar, Anna Ma, and Deanna Needell. Analysis of fast structured dictionary learning. Information and Inference: A Journal of the IMA, 9(4):785–811, 2020.
- [161] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86–109, 2019.
- [162] Wolfgang Ring. Structural properties of solutions to total variation regularization problems. ESAIM: Mathematical Modelling and Numerical Analysis, 34(4):799–810, 2000.
- [163] R Tyrrell Rockafellar and Roger J-B Wets. Variational analysis, volume 317. Springer Science & Business Media, 2009.
- [164] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.
- [165] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.
- [166] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. Science Advances, 3(4):e1602614, 2017.
- [167] Daniel Rueckert and Julia A Schnabel. Model-based and data-driven strategies in medical image computing. Proceedings of the IEEE, 108(1):110–124, 2019.
- [168] Lars Ruthotto and Eldad Haber. An introduction to deep generative modeling. *GAMM-Mitteilungen*, 44(2):e202100008, 2021.
- [169] Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pages 5546– 5557. PMLR, 2019.
- [170] Irwin W Sandberg. Approximations for nonlinear functionals. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 39(1):65–67, 1992.
- [171] Klaus Scheffler. A pictorial description of steady-states in rapid magnetic resonance imaging. *Concepts in Magnetic Resonance: An Educational Journal*, 11(5):291–304, 1999.
- [172] Björn Schembera and Juan M Durán. Dark data as the new challenge for big data science and the introduction of the scientific data officer. *Philosophy & Technology*, 33:93–115, 2020.
- [173] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen. Variational methods in *imaging*, volume 167. Springer, 2009.
- [174] Michael Schmidt and Hod Lipson. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324(5923):81–85, April 2009.

- [175] Karin Schnass. Local identification of overcomplete dictionaries. *Journal of Machine Learning Research*, 16(Jun):1211–1242, 2015.
- [176] Banafshe Shafieizargar, Riwaj Byanju, Jan Sijbers, Stefan Klein, Arnold J den Dekker, and Dirk HJ Poot. Systematic review of reconstruction techniques for accelerated quantitative MRI. *Magnetic Resonance in Medicine*, 90(3):1172– 1208, 2023.
- [177] Clemens Sirotenko. Optimal choice of spatially adaptive parameters in total generalized variation via bilevel optimization with applications to fourier inpainting. Master's thesis, Humboldt University of Berlin, 2020.
- [178] Clemens Sirotenko. *Data-driven reconstruction methods for quantitative Imaging*. PhD thesis, Humboldt University of Berlin, 2024. Ongoing.
- [179] Pingfan Song, Yonina C Eldar, Gal Mazor, and Miguel RD Rodrigues. Hydra: Hybrid deep magnetic resonance fingerprinting. *Medical physics*, 46(11):4951–4969, 2019.
- [180] Daniel A Spielman, Huan Wang, and John Wright. Exact recovery of sparsely-used dictionaries. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pages 37.1–37.18. JMLR Workshop and Conference Proceedings, 2012.
- [181] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere I: Overview and the geometric picture. *IEEE Transactions on Information Theory*, 63(2):853–884, 2016.
- [182] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere II: Recovery by riemannian trustregion method. *IEEE Transactions on Information Theory*, 63(2):885–914, 2016.
- [183] Karsten Tabelow, Peter Benner, Renita Danabalan, Dominik Göddeke, Lars Kastner, Tabea Krause, Daniel Mietchen, Marco Reidelbach, Björn Schembera, Moritz Schubotz, and Rainer Sinn. Mathematical research data management planning in mathematics, 2023.
- [184] The MaRDI consortium. MaRDI: Mathematical research data initiative proposal. https://zenodo.org/ record/6552436, 2022.
- [185] Andreas M Tillmann. On the computational intractability of exact and approximate dictionary learning. IEEE Signal Processing Letters, 22(1):45–49, 2014.
- [186] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pages 945–948. IEEE, 2013.
- [187] Patrick Virtue, Stella X Yu, and Michael Lustig. Better than real: Complex-valued neural nets for MRI fingerprinting. In 2017 IEEE International Conference on Image Processing (ICIP), pages 3953–3957, Beijing, September 2017. IEEE.
- [188] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2(12):737–748, 2020.
- [189] Nikolaus Weiskopf, Martina F. Callaghan, Oliver Josephs, Antoine Lutti, and Siawoosh Mohammadi. Estimating the apparent transverse relaxation time (R2*) from images with different contrasts (ESTATICS) reduces motion artifacts. *Frontiers in Neuroscience*, 8(278):1–10, 2014.
- [190] Nikolaus Weiskopf, John Suckling, Guy Williams, Marta M. Correia, Becky Inkster, Roger Tait, Cinly Ooi, Edward T. Bullmore, and Antoine Lutti. Quantitative multi-parameter mapping of R1, PD(*), MT, and R2(*) at 3T: a multi-center validation. *Front. Neurosci.*, 7:95, 2013.
- [191] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino Da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 'T Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene Van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan Van Der Lei, Erik Van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, March 2016.
- [192] Ezra Winston and J Zico Kolter. Monotone operator equilibrium networks. *Advances in neural information processing systems*, 33:10718–10728, 2020.
- [193] G A Wright. Magnetic resonance imaging. IEEE Signal Processing Magazine, 14(1):56–66, 1997.
- [194] Mingrui Yang, Yun Jiang, Dan Ma, Bhairav B Mehta, and Mark A Griswold. Game of Learning Bloch Equation Simulations for MR Fingerprinting, 2020. arXiv:2004.02270.
- [195] Marcelo VW Zibetti, Patricia M Johnson, Azadeh Sharafi, Kerstin Hammernik, Florian Knoll, and Ravinder R Regatte. Rapid mono and biexponential 3D-T1ρ mapping of knee cartilage using variational networks. *Scientific Reports*, 10(1):19144, 2020.