Author for correspondence:

Simon R. Arridge

e-mail: s.arridge@ucl.ac.uk

(An Overview of) Synergistic Reconstruction for Multimodality/Multichannel Imaging Methods

Simon R. Arridge [iD] [1],

Matthias J. Ehrhardt [iD] [2,3],

and Kris Thielemans [iD] [4]

[1] Department of Computer Science, University College London, UK
[2] Department of Mathematical Sciences, University of Bath, UK
[3] Institute for Mathematical Innovation, University of Bath, UK
[4] Institute of Nuclear Medicine, University College London, UK

Imaging is omnipresent in modern society with imaging devices based on a zoo of physical principles, probing a specimen across different wavelengths, energies and time. Recent years have seen a change in the imaging landscape with more and more imaging devices combining that which previously was used separately. Motivated by these hardware developments, an ever increasing set of mathematical ideas is appearing regarding how data from different imaging modalities or channels can be synergistically combined in the image reconstruction process, exploiting structural and/or functional correlations between the multiple images. Here we review these developments, give pointers to important challenges and provide an outlook as to how the field may develop in the forthcoming years.

THE ROYAL SOCIETY
PUBLISHING

1. Introduction

Images have been an ever-present component of human civilisation for thousands of years because of their unique ability to record and represent complex information in a form directly interpretable by human intelligence. Whereas *directly* recorded images are ubiquitous in static and moving formats, the enormous advances in engineering, physics, mathematics and computer science in the last half-century has led to the increasing availability of *indirect* imaging methods. Specifically, we refer to technologies where data g, belonging to some space Y, is measured and the desired image f, in a different space X, is recovered by solving an *inverse problem* of the form

$$g = A(f) + e \tag{1.1}$$

where A is the (possibly non-linear) mapping that models the generation of data from a given estimate of f and e represents noise, arising from one or several sources of error.

Such image reconstruction problems arise in many areas of science, including geophysics, non-destructive testing, atmospheric physics and notably in medical imaging which is the principal focus of this article. Many different physical phenomena can be measured (representing the space Y) such as electro-magnetic, acoustic, and optical waves, as well as particle counting processes, and many different physical parameters can be reconstructed (representing the space X), including density, sound speed, attenuation coefficients, molecular relaxation rates, tracer concentration, amongst many others. Modalities based on X-Rays (computed tomography) and magnetic resonance (both of which led to the award of Nobel prizes for their development) are well known examples present in every hospital.

With the success of such technologies has come the quest to push the boundaries of achievable imaging in regard to speed, resolution and additional physical parameters. Although most imaging techniques were originally envisioned as purely 2D methods, the relentless increase in computing power has made 3D imaging routine. Nevertheless, the challenge to address so called "4D" or even "5D" imaging (adding the dimension of time and/or spectral variations, or potentially both) still presents difficulties, both in terms of computation time, and, more fundamentally, in terms of adequate data acquisition within constraints such as patient tolerance and safety. We will refer to these extensions as *Multichannel Imaging (MCI)*.

A separate, but related, extension to conventional image reconstruction modalities is the development of *Multimodality Imaging (MMI)*. The key difference here is that the measurements usually are of different physical phenomena, and/or the recovered images represent two or more different physical parameters. The increasing interest in MMI is accelerating with the ever increasing advances in systems and reconstruction techniques [1–3].

This article is a brief overview of some recent developments in these topics with a focus on image reconstruction methods. The emphasis of this article is on the various different incarnations of *synergistic imaging* wherein several images are reconstructed simultaneously from data acquired simultaneously. We also briefly cover the joint reconstruction of multiple images from data that was sequentially acquired, e.g. such as in dynamic imaging, follow-up studies or many multi-modality cases.

The article is organised as follows. We provide brief definitions and terminology in section 2 as well as a taxonomy of applications. Section 3 on methods for "guided reconstruction", where a single image is reconstructed with regularisation based on one or more other images, will serve as a gentle introduction to the main body of this review, section 4, where we provide an overview of the dominant notions for synergistic reconstruction. We conclude this review in section 5 with a discussion and point out important challenges and an outlook for the future of the field. For completeness we summarise useful concepts from inverse problems and image reconstruction in an appendix section A.

2. A Taxonomy of Problems

(a) Basic Definitions

In this review we concentrate on scalar-valued images as most of the synergistic reconstruction literature has been developed in this context. Some concepts could be generalised to (geometrical) vector or tensor-based images (for instance for velocity or diffusion). We use the word "image" both for the continuous function $f(x)$ and the discretised version where the function is obtained as a sum over basis functions

$$f(x) = \sum_i f_i b_i(x) \tag{2.1}$$

where x is a coordinate in space (most often three dimensional: $x \in \mathbb{R}^3$). Most authors use cuboid basis functions ("voxels"). We will introduce multi-channel images below.

A commonly used generic setting for solving problem (1.1) is to solve an associated variational problem

$$f_* = \underset{f}{\operatorname{argmin}} \left[\mathcal{D}(g, A(f)) + \alpha \Psi(f) \right] \tag{2.2}$$

where $\mathcal{D}$ is the *data fit functional* measuring a suitable distance between the observed data g and the output of the model $A(f)$, and Ψ is a regularisation functional. The approach (2.2) is often referred to as a *variational regularization* of the inverse problem (1.1). Under the Bayesian interpretation, (2.2) is the negative logarithm of the product of the likelihood and the prior, i.e.

$$\mathcal{D}(g, A(f)) \equiv -\log(P(g|A(f))), \quad \alpha \Psi(f) \equiv -\log(P(f)).$$

In MCI we consider a vector of measurements $\mathbf{g} \in Y := \prod_{m=1}^{M} Y_m$, a vector of images $\mathbf{f} \in X := \prod_{p=1}^{P} X_p$ and the forward operator $\mathbf{A} : X \to Y$ mapping between these product spaces. Within this description we distinguish

- Multi-Channel Single-Image (MCSI), which implies reconstructing a single image from multiple channels, i.e. $M > 1, P = 1$.
- Single-Channel Multi-Image (SCMI), which implies reconstructing multiple images from a single channel, i.e. $M = 1, P > 1$.
- Multi-Channel Multi-Image (MCMI) which implies reconstructing multiple images from multiple channels, i.e. $M > 1, P > 1$.

In the case of MMI the various domain and range spaces are composed of different modalities and quantitative image representations, and we may write all terms in stacked form:

$$\mathbf{g} \equiv \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \quad \mathbf{A} \equiv \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \quad \mathbf{f} \equiv \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \tag{2.3}$$

Each modality may be independently linear or nonlinear, well-posed, weakly ill-posed or strongly ill-posed. Unless requiring reference to particular case details, we will use a single notation for all the above:

$$\mathbf{g} = \mathbf{A}(\mathbf{f}). \tag{2.4}$$

(b) Guided Reconstruction

Closely related to the synergistic reconstruction problem in MCI or MMI is the possibility of using one acquired modality or channel, with a robustly reconstructed image, as a prior for subsequent, usually less well-posed, image reconstruction problems. This constitutes a sequential process where two different acquisitions are made, image 1 is reconstructed and then image 2 guided by image 1. Most often, the key idea is that image 1 has high resolution structural information whereas image 2 is low resolution and/or a functional image.

Although (within medical imaging) the above concept is sometimes called *anatomically-guided image reconstruction* or *structure-driven image regularisation*, we will refer to this approach as *guided image reconstruction* in order to generalise to other applications. We describe the main techniques in §3. A more detailed review of applications and methods can be found in [4], see also [5] for Positron Emission Tomography (PET)/Magnetic Resonance Imaging (MRI) applied to neurology.

(c) Synergistic Reconstruction

In terms of the definitions given in §2(a) we can list a number of applications, grouped according to their similarity in regard to image reconstruction strategies:

- SCMI. Examples are acoustic speed and attenuation from Ultrasound (US) data, absorption and scattering from unscattered photons only, or from photon intensity only in Diffuse Optical Tomography (DOT), and attenuation and activity estimation in PET and Single Photon Emission Computed Tomography (SPECT). However, SCMI is usually a very ill-conditioned problem with unsatisfactory results, often exhibiting non-uniqueness [6]. An exception is PET with Time of Flight (TOF) where the extra information on the approximate location of the activity along a line of response is sufficient to provide uniqueness up-to some constant [7].
- MCMI : Multispectral Imaging. Modalities that can be classified as *multispectral imaging* include multispectral Computed Tomography (CT) [8–14], multispectral Electrical Impedance Tomography (EIT) [15,16], multispectral DOT [17], multispectral Photoacoustic Tomography (PAT) [18] and Quantitative Photoacoustic Tomography (QPAT) [19–21].

  A common feature of multispectral imaging is the expression of $f(\lambda)$ as a mixture of component images z_m

  $$f(\lambda) = \sum_m E_m(\lambda) z_m \tag{2.5}$$

  where λ is the energy/wavelength and the spectral signatures $E_m(\lambda)$ of the components may be fully or partially known. Therefore these problems are often posed in two steps: a channel by channel reconstruction for each λ followed by a *spectral unmixing problem* solving (2.5) for $\mathbf{z}$. If the prior is defined in terms of $\mathbf{f}$ it may introduce extra bias into the recovery of $\mathbf{z}$. Alternatively, the prior may be directly imposed on $\mathbf{z}$; see [16,22,23] for examples. A benefit of a one-step reconstruction procedure is that there is no propagation of errors from the channel-wise tomographic inverse problem to the spectral unmixing one. A drawback is that the full inverse problem may become nonlinear, which potentially leads to a longer computation time.
- MCMI : Multi-Energy imaging. An example is the reconstruction of both attenuation and tracer concentration from detection of both unscattered and scattered photons, the latter having reduced energy. This has been demonstrated in both SPECT [24–26] and PET [27–29].
- MCMI : Multi-Time Imaging : in dynamic/kinetic imaging the aim is to explicitly separate different temporal variations as separate images. The time-series of images can be reconstructed with *e.g.* a total nuclear norm constraint [30]. The assumption is that the number of separate temporal components is small and could be concisely expressed using PCA, ICA or NNMF for example. This can be extended to allow outlier representations via the *Low-Rank plus sparse* approach [31]. Alternatively, the temporal behaviour could be expressed as an approximately known function with some random components modelled by a Kalman process [32], or it can be constrained to follow a (potentially non-linear) kinetic model

  $$f(t) = \Phi(k, t) \tag{2.6}$$

  with k parametric images, similar to (2.5). However, we are not aware of any literature yet where there is a prior that couples the parametric images.

- MCMI : MRI imaging with multiple sequences also generate multiple images, with many different applications and sometimes overlap with the aforementioned categories, e.g. multimodal dynamical MRI [33], multiparametric MR [34], multi-contrast MRI [35–37], MR Fingerprinting which aims to estimate multiple images corresponding to different tissue properties [38,39], see also [40].
- Multi-Modality Multi-Image (MMMI) : The distinguishing aspect is that we have multiple modalities which may have very different physical measurements. Notable examples : PET and MRI (PET/MRI) [41–43], PAT and US [44], geophysics applications with multiple data (e.g. electromagnetic waves, seismic waves, radar, DC resistivity, groundwater flow) [45–52].
- MMMI : Coupled Physics Imaging (CPI). These methods are so-named because the measurement involves the cross-generation of one type of wave from another [53]. Examples include PAT and Optical Coherence Tomography (OCT) [54], quasi-static elasticity imaging [55] and Acousto-Electric Tomography [56].

3. Guided Reconstruction

In this section we briefly describe the main methods that have been developed for guided reconstruction, as many of these ideas can and have been extended to the synergistic case.

(a) The Continuous Setting

An obvious way to include information from an auxiliary image is to take a regularisation strategy for one modality, f_1 say, and introduce a local dependence on f_2. For example, the form expressed in eq. (A 6) can be extended to the form

$$\Psi(f_1) \quad := \quad \int_\Omega w(f_2(x))\psi(|\nabla f_1(x)|)\mathrm{d}x\,, \tag{3.1}$$

$$\rightarrow \quad \Psi'(f_1) \quad = \quad -\nabla \cdot \left(w(f_2(x))\frac{\psi'(|\nabla f_1(x)|)}{|\nabla f_1(x)|}\right)\nabla f_1(x) = -\nabla \cdot w(f_2(x))\kappa(f_1(x))\nabla f_1(x) \tag{3.2}$$

The choice for the weighting $w(f_2)$ could be quite general, and need not be strictly local. A recurring concept is to make $w \in [0,1]$ an *edge-indicator* such as (A 9) computed on f_2, i.e.

$$w(f_2) = \exp\left(-|\nabla f_2|^2/\epsilon_2^2\right) \tag{3.3}$$

This form of prior favours a reconstruction of f_1 where its edges are colocated with those of f_2; see figure 1(a). A more powerful approach may be also to encourage the direction of image gradients to be aligned; see figure 1(b). This is similar to the Edge Enhancement Diffusion concept outlined in (A 10)-(A 12). Again, the new concept is to control the flow of f_1 based on the gradient directions in f_2 rather than only on those of f_1 itself. For image reconstruction the methodology was introduced by Kaipio *et al.* [57] by defining a tensor field, $B(f_2)$, that incorporates directional structure from f_2:

$$\Psi(f_1) = \int_\Omega |\nabla f_1|^2_{B(f_2)}\mathrm{d}x = \int_\Omega |U^\mathrm{T}\nabla f_1|^2\mathrm{d}x \tag{3.4}$$

Specifically, for two-dimensional images, the choice

$$B = I - (1-\beta)\hat{\nu}\hat{\nu}^\mathrm{T} = \hat{\tau}\hat{\tau}^\mathrm{T} + \beta\hat{\nu}\hat{\nu}^\mathrm{T} = R\mathsf{K}R^\mathrm{T} = UU^\mathrm{T} \tag{3.5}$$

where $R = [\hat{\nu}\ \hat{\tau}] = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is a rotation matrix, $\mathsf{K} = \begin{pmatrix} \beta & 0 \\ 0 & 1 \end{pmatrix}$ is an anisotropy matrix, $U = R\mathsf{K}^{1/2}$ and $\beta \in [0,1]$ is again an edge-indicator, leads to the diffusion flow

$$\frac{\partial f_1}{\partial t} = \nabla \cdot R\mathsf{K}R^\mathrm{T}\nabla f_1\,. \tag{3.6}$$

Here $\tilde{\nabla} = R^{\mathrm{T}}\nabla = \begin{pmatrix} \frac{\partial}{\partial\tilde{\nu}} \\ \frac{\partial}{\partial\tilde{\tau}} \end{pmatrix}$ is the gradient in local "gauge" coordinates. This approach can be generalised by using other local functions ψ resulting in the form

$$\Psi(f_1) \quad := \quad \int_{\Omega} \psi(|U^{\mathrm{T}}\nabla f_1|)\mathrm{d}x\,, \tag{3.7}$$

$$\rightarrow \quad \Psi'(f_1) \quad = \quad -\nabla \cdot U \left(\frac{\psi'(|U^{\mathrm{T}}\nabla f_1(x)|)}{|U^{\mathrm{T}}\nabla f_1(x)|} \right) U^{\mathrm{T}}\nabla f_1(x) \tag{3.8}$$

$$= \quad -\tilde{\nabla} \cdot \underbrace{\mathsf{K}^{1/2} \left(\frac{\psi'(|\nabla f_1(x)|_B)}{|\nabla f_1(x)|_B} \right) \mathsf{K}^{1/2}}_{\tilde{\mathsf{K}}} \tilde{\nabla} f_1(x)\,. \tag{3.9}$$

Of course here we can only give a glimpse into continuous regularisation for guided reconstruction. Similar ideas exist for regularisers that do not fit (3.1) such as the total variation [58] and the total generalised variation [59], see for instance [4,60,61].

(b) The Discrete Setting

The above description is given in the continuous setting, but can be readily discretised. Here we briefly describe some methods that are specific to the discrete case.

(i) Markov Random Field Priors

One prominent example is based on MRFs (A 14), where weights are made dependent on f_2 :

$$\Psi(f_1) = \frac{1}{p} \sum_i \sum_{j \in \mathcal{N}(i)} w_{ij}(f_2)\|f_{1i} - f_{1j}\|^p\,. \tag{3.10}$$

The simplest choice is to set the weights to zero across known edges (derived from f_2), ideally with some blurring to accommodate imperfect edge information [62]. Leahy & Yan estimated both image values and edge-indicators by incorporating known edge information (obtained from MRI) into an MRF prior that reduced the weights across edges, while encouraging continuous edges [63].

To avoid having to determine the edges, the most popular choice nowadays to choose the weights in (3.10) is called the *Bowsher prior* [64]. Here for every voxel i, only the n weights are kept non-zero which correspond to the n smallest differences $\|f_{2i} - f_{2j}\|$.

Another choice for the weights, inspired by the kernel-method described in A(e), is to use a similarity function $\mathcal{K}$ between "features" computed on f_2, potentially together with a dependence on the distance between the i and j voxels to enforce locality, *e.g.*:

$$w_{ij}(f_2) = \mathcal{K}(\mathcal{T}_i(f_2), \mathcal{T}_j(f_2)) \exp\left(\frac{-\|r_i - r_j\|^2}{2\sigma_s^2} \right) \tag{3.11}$$

with $\mathcal{T}_i(f_2)$ a feature-vector computed at voxel i, r_i the spatial coordinate of voxel i, and σ_s a tunable parameter. This has been used with the Radial Basis Function (RBF) similarity function (A 19) for multi-tracer PET by Ellis *et al.* [65], who sparsified the weights by keeping only the $n_{\max}$ largest weights for each voxel, similar to the Bowsher prior[1]. Bland *et al.* extended this idea by further adapting the weights over iterations by including Gaussian differences of the PET of the previous update [66].

(ii) Kernel Methods

Wang & Qi have used the kernel method (see section A(e)) to represent the image as a linear combination of transformed "features" computed on the guidance [67]. The image that needs to

---

[1]Note the relation between this approach and the discretised version of (3.3)

be estimated can then be written in terms of a coefficient image ζ as

$$f_{1,i} = \sum_j \mathcal{K}(\mathcal{T}_i(f_2), \mathcal{T}_j(f_2))\zeta_j \tag{3.12}$$

with $\mathcal{K}$ a similarity function ("kernel function") between features $\mathcal{T}_i(f_2)$. The kernel matrix $\mathcal{K}(\mathcal{T}_i(f_2), \mathcal{T}_j(f_2))$ was sparsified to improve computational performance and stability. The reconstruction then becomes an optimisation problem in terms of ζ. This approach has been used in MMI – see [67–69] (PET with MRI), [70] (SPECT with PET), [71] (fluorescence molecular tomography with CT or MRI), [72] (DOT with CT) – but also for MCMI – PET dynamic imaging using static images as guide [67] or temporal features derived from the raw data [73].

One potential problem with kernel-based methods is that the kernel matrix can be too restrictive such that "unique" features in the image that is reconstructed are suppressed. This can be mitigated by limiting the spatial extent of the kernel function [74], or by adapting the kernel matrix by using features computed on both the guidance and current estimate of the image. The latter approach can also be called the "hybrid" kernel method [43].

(iii) Basis Function Selection

The basis functions (2.1) can be chosen based on the guidance, for instance increasing spatial extent i areas where f_2 is smooth.

An approach originating in the machine learning community is "dictionary learning", where images are written in terms of a dictionary, obtained from some training data. However, most of the literature does not fit in the guided reconstruction category as it uses data from the same modality for learning the dictionary, and/or adapts the dictionary from a current estimate of the image. Tang *et al.* reconstruct PET images using a quadratic penalty encouraging similarity with the previous iterate denoised using a pre-defined dictionary. The dictionary was trained on various images, including MRI images from the same subject [75]. Tahaei *et al.* instead reconstruct the PET image directly as a sparse combination of the dictionary [76]. The dictionary was learned from an MRI image of the same subject, and then changed to allow different contrast and impose non-negativity. Both papers show promising results, although somewhat surprisingly Tang *et al.* obtained good results with a dictionary trained on a simple hollow sphere as well. Sudarshan *et al.* extended the method of [75], by using a coupled dictionary encoding both PET and MRI images, with the latter obtained from the same subject, adapting the dictionary at each iteration [77].

"Super-voxels" (or super-pixels) are another closely related concept for selecting basis functions. They were originally developed for segmentation where voxels are grouped based on similarity and spatial closeness. Multiple "layers" of different super-voxel realisations were later proposed as an overcomplete basis for image registration [78]. Jiao *et al.* proposed creating super-voxels from both a pre-computed MRI image and the gradient of the PET log-likelihood to avoid bias in PET-only features [79].

4. Synergistic Reconstruction

This section reviews important concepts for synergistic image reconstruction. Some methods are related to the concepts for guided reconstruction (see section 3) whereas some are directly formulated for the synergistic setting.

(a) Joint MAP Estimation

In a Bayesian framework, a central concept for many synergistic image reconstruction approaches is to formulate the multi-modality inverse problem (2.4) as a joint maximum *a posteriori* (JMAP)

estimate given by

$$(f_{1*}, f_{2*}) = \operatorname*{argmin}_{f_1, f_2} \left[-\log P(f_1, f_2 | g_1, g_2) = -\log P(g_1, g_2 | f_1, f_2) + \alpha \Psi(f_1, f_2) + \text{const} \right] , \quad (4.1)$$

where $P(g_1, g_2 | f_1, f_2)$ is the multi-modality likelihood. This is in contrast to the Bayesian framework for guided reconstruction which would suggest a conditional *a posteriori* (CMAP) estimate

$$f_{1*} = \operatorname*{argmin}_{f_1} \left[-\log P(f_1 | g_1, f_2) = -\log P(g_1 | f_1) + \alpha \Psi(f_1 | f_2) \right] . \quad (4.2)$$

While (4.1) is conceptually simple, it is in general difficult to specify a good multi-modality likelihood. The situation significantly simplifies when certain conditional independence are assumed (see [80] for more details), since then the likelihood factors as $P(g_1, g_2 | f_1, f_2) = P(g_1 | f_1) P(g_2 | f_2)$ and the JMAP (4.1) becomes

$$(f_{1*}, f_{2*}) = \operatorname*{argmin}_{f_1, f_2} \left[\mathcal{D}(g_1, A_1(f_1)) + \mathcal{D}(g_2, A_2(f_2)) + \alpha \Psi(f_1, f_2) \right] . \quad (4.3)$$

In the special case of white Gaussian distributed noise in each modality, the JMAP then reads

$$(f_{1*}, f_{2*}) = \operatorname*{argmin}_{f_1, f_2} \left[\frac{1}{2\sigma_1} \|g_1 - A_1 f_1\|^2 + \frac{1}{2\sigma_2} \|g_2 - A_1 f_2\|^2 + \alpha \Psi(f_1, f_2) \right] , \quad (4.4)$$

which defines a natural statistical scaling between the two least squares terms.

Such conditional independence assumptions are used in almost all contributions based on the JMAP although often not explicitly mentioned.

We mention also that similar questions arise w.r.t. the different image channels, including scaling between terms and quantitative difference in images scales, as many (joint) priors depend on image scale. In the optimisation literature pre-scaling between different dimensioned variables is known as *sphereing* (referring to the ellipticity of the posterior covariance), but this is rarely made explicit in the synergistic literature.

(b) Joint Sparsity

The variational synergistic reconstruction problem (4.3) needs a regularizer Ψ which encodes the desired properties between the images f_1 and f_2. Choosing such a regularizer is a highly nontrivial task and a good choice will generally depend on each individual application depending on what properties f_1 and f_2 are expected to share.

In many applications it is desirable for f_1 and f_2 to have many common edges, i.e. edges are more likely to occur at the same locations in f_1 and f_2 than it is for edges to occur at different locations. If, in addition, both images are likely to have a small jump set, then this *a priori* information can be encoded via the *Joint Total Variation* (JTV) [41,46,81]

$$\text{JTV}(f) = \sum_i \sqrt{\|\nabla_i f_1\|^2 + \|\nabla_i f_2\|^2} \quad (4.5)$$

where $\nabla_i f$ denotes the spatial gradient of f at location (e.g. voxel) i. This regularizer can be readily extended to an arbitrary number of images by summing over all gradient norms. An alternative but completely equivalent viewpoint is to define the joint total variation as the sum over the 2-norm of the Jacobian of the vector-valued image f. In the context of colour image processing, this regularizer is also called *Vectorial Total Variation* (VTV) [82,83].

An alternative to sparsity of the Jacobian is to consider joint sparsity of the wavelet coefficients [84]. Since wavelets are highly localized one would expect this prior to be useful in similar situations as described before for the joint total variation. Joint sparsity can also be used in other over-complete basis settings such as "super-voxels".

The notion of joint sparsity is not limited to explicitly given transforms such as the gradient or the wavelet transform. For instance, when considering a dynamic sequence $f_1, \ldots, f_n$ it is often
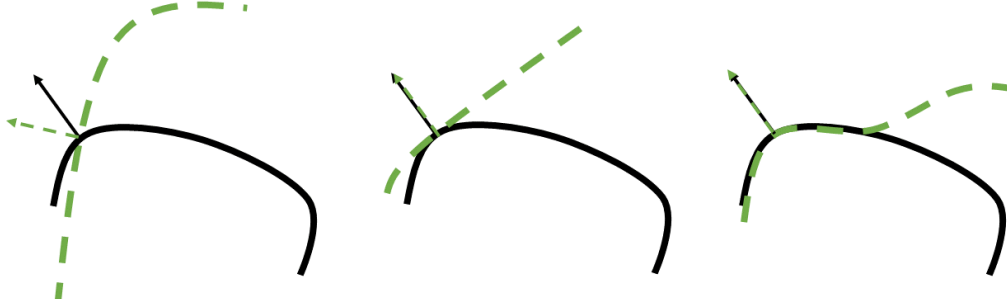
Figure 1: Level Sets with coincident edges : left) intersecting at a point with different orientations; centre) intersecting with parallel orientations but different curvatures; right) intersecting with parallel orientations and the same curvature.

desirable to promote a low rank of the Casoratti matrix

$$[f_1(:), \ldots, f_n(:)] \tag{4.6}$$

where we abused MATLAB notation for simplicity. Such low rank can be for instance promoted with the nuclear norm [30,31].

(c) Joint Geometric Regularisation

The idea in joint geometric regularisation is to define a prior that enforces consistent geometric structure between channels or modalities. Although similar to the motivation in §3(a) the key point is that is optimised with respect to each image.

The *Parallel Level Sets* (PLS) prior [80,85] is defined as

$$\Psi^{\mathrm{PLS}}(f_1, f_2) = \int_\Omega \varphi\bigg(\psi\left(\|\nabla f_1(x)\|\|\nabla f_2(x)\|\right) - \psi\left(|\langle\nabla f_1(x), \nabla f_2(x)\rangle|\right)\bigg)\mathrm{d}x \tag{4.7}$$

for strictly increasing functions φ, ψ. One can see that this prior achieves its smallest value if the two gradients are co-linear (or parallel) at each point, i.e. for almost all $x \in \Omega$ there exists a scaling factor $\beta \in \mathbb{R}$ such that $\nabla f_1(x) = \beta\nabla f_2(x)$ or $\nabla f_2(x) = \beta\nabla f_1(x)$. See also [86] for a comprehensive overview of this concept.

This generalized framework encompasses some previously used regularizers. For instance, if $\phi(t) = t$ and $\psi(t) = t^2$, then it measures the squared norm of the cross-product of ∇f_1 and ∇f_2, which has been successfully used in the geophysics literature [46,47]. The gradient of Ψ^{PLS} w.r.t f_1 produces an anisotropic diffusive flow where the diffusivity depends on f_2, and the gradient w.r.t f_2 produces an anisotropic diffusive flow where the diffusivity depends on f_1 so as to overall favour the alignment of both image gradients [80].

There are also other priors which promote parallel level sets or parallel gradients. For instance, the *Total Nuclear Variation* (TNV) [8,14,87,88] promotes sparsity in the singular values of the matrix of gradients (the Jacobian) $\mathsf{V}(x) := [\nabla f_1(x), \nabla f_2(x), \ldots, \nabla f_M(x)]$ at almost every $x \in \Omega$. One can see that the Jacobian has rank 1 if and only if all images $f_1, \ldots, f_M$ have pairwise parallel level sets. An advantage of TNV compared to PLS is that the former is convex in the joint argument $(f_1, \ldots, f_M)$ whereas the latter is in general nonconvex. See also [4,86] for a deeper discussion of the connections between TNV and PLS. The same idea has also been exploited for higher-order regularization such as the total generalized variation, see [89] for more details.

In principle, the concepts of joint structure could be extended to higher order geometrical features, for example using the "jet" of derivatives $\mathcal{M}^J : f \mapsto \left\{ \frac{\partial f}{\partial x_i}, \frac{\partial^2 f}{\partial x_{i_1}\partial x_{i_2}}, \cdots \frac{\partial^J f}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_J}} \right\}$, which is the basis of several approaches to analysing shape in images, including at multiple scales [90,91]. For example, one could consider the definition of curvature $\gamma = \hat{\nu}^{\mathrm{T}} H \hat{\nu}$, where

$\hat{\nu} = \frac{\nabla f}{|\nabla f|}$ is the local level set normal direction and $H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}$ is the image Hessian. Images with the same local curvature may have stronger corresponding structural similarity then those with only local normal coincidences; see figure 1 (c).

(d) Regularisation exploiting Joint Statistics

Joint entropy is a measure of randomness characterizing the joint probability density function $P(f_1, f_2)$ of two random variables f_1 and f_2. The joint Shannon Entropy is given by:

$$S(f_1, f_2) = -\int_\Omega P(f_1(x), f_2(x)) \log(P(f_1(x), f_2(x))) dx. \tag{4.8}$$

See figure 2. Joint Entropy and Mutual Information (i.e. the difference between joint and marginal entropies) are routinely used in image registration [92].

Setting $\Psi(f_1, f_2) = S(f_1, f_2)$ specifies a regularisation scheme that minimises the joint entropy between f_1 and f_2. Qualitatively, entropy measures the "peakiness" of a probability distribution; i.e. the more concentrated the distribution around cluster points the lower the entropy, and minimising it will lead to images that have less uncertainty [93–98]. One drawback of the definition in (4.8) is that it depends only on pixel intensity value and does not admit any spatial correlation. This motivated Tang and Rahmim to extend the Joint Entropy concept to a multi-resolution description based on wavelets [99].

Differentiation of $S(f_1, f_2)$ can be made computationally efficient using Parzen kernel density estimators to develop a continuous function based on the sample pixels in f_1, f_2 [95,97]. However (4.8) is extremely non-convex, and minimisation of joint entropy requires careful attention such as applying multiple re-starts from different initialisations [96] which has prevented its wider uptake as a regularisation scheme. As an alternative to Shannon entropy the Burg entropy has been proposed [98] which is computationally more tractable whilst still providing comparable results.

An approach combining joint feature-space clustering with image diffusion was developed in [100]. Here a distance measure in $P(f_1, f_2)$ was defined to define a diffusivity that favoured intra-class smoothing above inter-class smoothing. The method was applied to enhance multichannel MRI images, but could in principle be used to regularise image reconstruction problems as well.

(e) Recycling of Guided Reconstruction for Synergistic Reconstruction

An alternative to joint reconstruction via joint regularization (4.3) is to alternate between guided reconstructions. In its most generality let $\mathcal{A}(A, g, f, v)$ be an algorithm that takes data f, current estimate f of the solution to the inverse problem $Af = g$ and guide image v, then one can always create an algorithm to perform joint reconstruction by iterating

$$f_1^{k+1} = \mathcal{A}(A_1, g_1, f_1^k, f_2^k) \tag{4.9}$$

$$f_2^{k+1} = \mathcal{A}(A_2, g_2, f_2^k, f_1^{k+1}) \tag{4.10}$$

or choose a guide image based on previous iterates, $v = \mathcal{B}(f^k)$ and update in parallel

$$f_1^{k+1} = \mathcal{A}(A_1, g_1, f_1^k, v) \tag{4.11}$$

$$f_2^{k+1} = \mathcal{A}(A_2, g_2, f_2^k, v). \tag{4.12}$$

The algorithm $\mathcal{A}$ may be derived from an algorithm to solve a guided variational reconstruction problem but does not have to be. It is also possible to use different algorithms $\mathcal{A}_1, \mathcal{A}_2$.

This *ad hoc* approach has a number of advantages. First of all, this approach is highly modular, meaning that any guided reconstruction algorithm can be recycled into a joint reconstruction algorithm. Second, most guided reconstruction algorithms are better understood and have
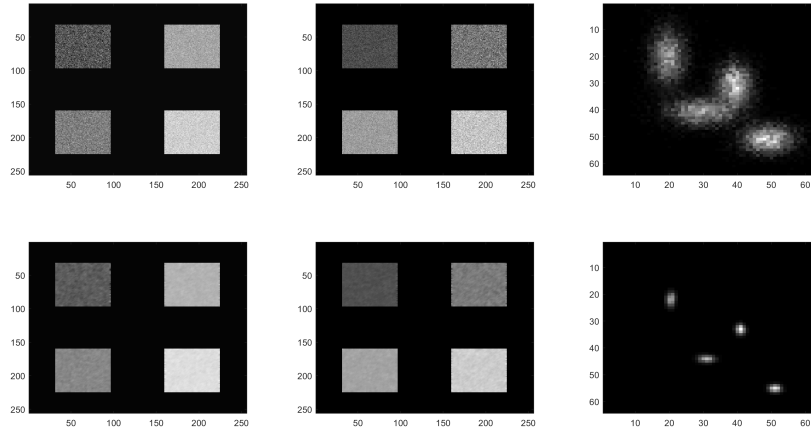
Figure 2: Joint Probability measures of multiple images. Top row : two images f_1, f_2 and their joint histogram. Bottom row : the result of applying edge-preserving diffusion to f_1 and f_2 and their new joint histogram. The joint entropy of the bottom right is lower than the top right, and the classes are more clearly separated.

favourable properties when compared with joint reconstruction algorithms. For example, many guided variational reconstruction problems are convex and are independent of the scaling of the guide image; two properties which joint variational approaches often lack.

That being said, it has also a number of disadvantages. Most importantly, there is no guarantee that the sequence f^k will converge and if it does, how the limit can be characterised. Second, the sequence f^k will depend on the actual implementation of the algorithm $\mathcal{A}$ like number of subiterations, step size etc even if the algorithm $\mathcal{A}$ itself is well-characterized as converging to the optimal solution of a guided variational regularization problem.

This approach has been used for spectral CT [10] where the algorithm $\mathcal{A}$ was solving a directional total variation regularized least squares problem and the guide image v is either chosen to be a weighted average over the previous iterate or a randomly chosen image from the previous iterate. See also the next section.

(f) Optimisation-inspired Synergistic Reconstruction

As we will highlight in section 5, many joint reconstruction variational problems and algorithms have unfavourable properties. For example, when solving (4.3) with joint total variation (4.5) regularization and grouped coordinate descent, the iterations read

$$
\begin{aligned}
f_1^{k+1} &= \arg\min_f \mathcal{D}(g_1, A_1(f)) + \alpha \operatorname{JTV}(f, f_2^k) \\
f_2^{k+1} &= \arg\min_f \mathcal{D}(g_2, A_2(f)) + \alpha \operatorname{JTV}(f_1^{k+1}, f),
\end{aligned}
\tag{4.13}
$$

where $\nabla_i f$ denotes the spatial gradient of f at the ith pixel/voxel. This algorithm (and the underlying variational problem) has two potential drawbacks. First, the same regularization parameter α is being used for both modalities. Second, since the regularizer compares the magnitude of the gradients of the two modalities, the reconstruction will favour the modality with the larger scale even though often they capture two very different physical phenomena which should not be compared.

A way out of these problems is to change the iterations (4.13) and introduce a weighting $\mu_j^k > 0$ and one regularization parameter for each modality α_j and iterate

$$f_1^{k+1} = \arg\min_f \mathcal{D}(g_1, A_1(f)) + \alpha_1 \, \mathrm{JTV}(f, \mu_2^k f_2^k)$$

$$f_2^{k+1} = \arg\min_f \mathcal{D}(g_2, A_2(f)) + \alpha_2 \, \mathrm{JTV}(\mu_1^k f_1^{k+1}, f) \, .$$

This approach has been studied for another algorithm (ADMM [101]) in the context of PET/MRI [102] and with a weighted quadratic prior (similar to (3.11)) instead of joint total variation for the same application in [103]. While this *ad hoc* modification potentially overcomes the aforementioned problems, it has the drawback that it is unclear how to choose the parameters μ_j^k, α_j and if the iterates f^k converge.

A related but different approach are the infimal-convolution Bregman total variation iterations [42]. Here the starting point are Bregman iterations based on the total variation regularizer [104] which given an iterate f^k and a subgradient $p^k \in \partial \mathrm{TV}(f^k)$ read

$$f^{k+1} = \arg\min_f \frac{1}{2}\|Af - g\|^2 + \alpha D_{\mathrm{TV}}^{p^k}(f, f^k) \tag{4.14}$$

$$p^{k+1} = \alpha^{-1}\left(p^k - A^*(Af^{k+1} - g)\right) . \tag{4.15}$$

Bregman iterations are proven to converge to a total variation minimizing solution of $Af = g$ (1.1), so early stopping is required in order to provide regularization, see for instance [105, chapter 6].

Bregman iterations can be extended to multiple modalities by replacing the Bregman distance $D_{\mathrm{TV}}^{p^k}(w, f^k)$ with a weighted sum of pairwise channel correlations

$$\sum_{i,j=1}^{M} w_{ij} D_{\mathrm{TV}}^{p_j^k}(f_i, f_j^k) . \tag{4.16}$$

The resulting algorithm is coined "Color Bregman iterations" [106]. Its convergence is guaranteed in some special cases, see [106] for more details. A problem with (4.16) is that it only promotes positive correlations between the edges in channels and negative correlations are suppressed. In order to circumvent this problem the Bregman distance of the total variation was replaced by the infimal convolution of Bregman distances with opposite sign subgradients. While no proof of convergence for resulting iterations is known, these were shown in [42] to be competitive with the state-of-the-art for joint PET-MR reconstruction.

(g) Joint-dictionary learning methods

Methods from §3.(b)iii can be extended to the joint problem.

Sudarshan *et al.* recently extended their work on PET reconstruction guided by a joint-dictionary using an MRI image from the same subject [77] to the joint problem for PET and (undersampled) MRI data [107]. The proposed method uses a pre-trained joint dictionary for PET and MRI magnitude images. The MRI phase image was not included in the dictionary nor penalised as it is usually sensitive to noise and phase wrap-around. The prior penalises the square of the differences between both reconstructed images and their denoised images (obtained by sparse coding). The method alternates between PET, MRI phase and magnitude reconstruction and sparse coding. The method performed well on simulated data. However, a practical difficulty would be how to obtain high quality data for training the dictionary.

Song *et al.* proposed an algorithm for multi-contrast MRI that alternates three stages: coupled dictionary learning (from random patches of current images), coupled sparse denoising (of all patches) and k-space consistency (as a gradient-step of normal image reconstruction with a quadratic penalty that encourages similarity between the denoised and reconstructed image [108]. An interesting point is that multiple dictionaries are used: a coupled dictionary for all contrasts, and independent dictionaries for each contrast. The latter are used to encode residual image

features that do not fit the coupled dictionary. As any adaptive dictionary learning problem, this problem is highly non-convex and the authors acknowledge that the suggested approach will at most converge to a local minimum, with no convergence proofs available.

5. Discussion and Outlook

In this article we have given a (somewhat personal) overview of the history and current state of synergistic reconstruction, with an emphasis on medical imaging. We have attempted to frame a common setting for commonly used methods and present some of the key concepts that arise in quite different applications. Due to space constraints as well as consistency of readability we have inevitably had to omit certain topics which we discuss here. At the same time we indicate the current challenges which will motivate the outlook for research in the next several years.

We have implicitly assumed that images have been acquired and reconstructed in a consistent coordinate system. This may not always be the case. In the case of multiple modalities, images might have different natural sampling, or more generally basis functions. It is possible to insert the necessary transformations into the joint priors, although in practice most authors choose the highest sampling for all images. If data-sets are not acquired at the same time, misalignment due to motion can create difficulties with some methods more sensitive to misalignment than others [109]. This leads to the topic of joint misalignment estimation and reconstruction, with recent contributions in the Guided context jointly estimating the image and the misalignment modelled either via a convolution kernel [110] or a spatial transformation [111,112]. If there is motion during the acquisition, it is possible to jointly estimate the image and the spatial transformation between motion different states [113,114]. There is considerable scope to combine this with methods from the image registration literature, such as joint penalties on the image and motion fields [115] and synergistic image registration for multi-modality data with the same underlying motion which benefits from different contrast and structures in the different modalities [116,117].

There are also several open questions in regard to algorithm choices and their implementation. Should (4.3) be solved with grouped coordinate descent or should all coordinates be updated in parallel? What are efficient algorithms to solve (4.3)? Since the computational cost of separate reconstruction of all channels/modalities scales linearly with respect to their number, one might aim to have a similar computational cost for synergistic reconstruction. It is currently unknown how existing methods theoretically scale in this regard. Another challenge is the non-convexity of many JMAP approaches, with in some cases presence of local minima. In such a case, heuristic update mechanisms are often used, for instance by using small steps in an alternating update scheme such that artefacts do not arise.

There is an ever increasing trend towards learning based methods in imaging, which is also seen in almost all branches of science and data analysis in general. See [118–120] for some reviews on learned image reconstruction methods. There is however currently very little research on applying these techniques to the multi-channel/modality problem, possibly due to the relative scarcity of data and the difficulties in obtaining ground truth in the situations that synergistic image reconstruction tries to tackle. Future research in this area is likely to emphasise the unification of model-driven and data-driven methods.

A practical difficulty with synergistic methods, especially in multi-modality imaging, is the need for software that can handle large amounts of data, is capable to accurately compute the system models (2.3), and ideally allows easy experimentation with novel algorithms. It is therefore often necessary to combine several software packages, ideally via an overarching framework [121] or by writing interfaces to other packages such that they can be used in an optimisation library such as [122–124].

Finally, although synergistic methods hold considerable promise to expand imaging into application areas where ill-conditioning of the single channel/modality otherwise impedes sufficient image quality, it also comes with its dangers. Like any regularisation, prior information can generate bias. This can lead to cross-talk, creating structure in one image when it should only

be present in the other. Addressing this will need both prior design [23] and extensive studies to validate these methods for each application.

Acknowledgements

References

1. Wang G, Zhang J, Gao H, Weir V, Yu H, Cong W, Xu X, Shen H, Bennett J, Furth M, Wang Y, Vannier M. 2012 Towards Omni-Tomography—Grand Fusion of Multiple Modalities for Simultaneous Interior Tomography. *PLoS ONE* **7**, e39700.
2. Cherry SR. 2006 Multimodality in Vivo Imaging Systems: Twice the Power or Double the Trouble?. *Annual Review of Biomedical Engineering* **8**, 35–62.
3. Vogler N, Heuke S, Bocklitz TW, Schmitt M, Popp J. 2015 Multimodal Imaging Spectroscopy of Tissue. *Annu. Rev. Anal. Chem.* **8**, 359–387.
4. Ehrhardt MJ. 2020 Multi-modality imaging with structure-promoting regularisers. accepted for Springer Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging, `arxiv.org/abs/2007.11689`.
5. Zhu Y, Zhu X. 2019 MRI-Driven PET Image Optimization for Neurological Applications. *Frontiers in Neuroscience* **13**.
6. Arridge SR, Lionheart WRB. 1998 Non-Uniqueness in Diffusion-Based Optical Tomography. *Optics Letters* **23**, 882–884.
7. Defrise M, Rezaei A, Nuyts J. 2012 Time-of-flight PET data determine the attenuation sinogram up to a constant. *Physics in Medicine and Biology* **57**, 885+.
8. Rigie DS, Rivière PJL. 2015 Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. *Physics in Medicine and Biology* **60**, 1741–1762.
9. Niu S, Yu G, Ma J, Wang J. 2018 Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction. *Inverse Problems* **34**.
10. Kazantsev D, Jørgensen JS, Andersen MS, Lionheart WR, Lee PD, Withers PJ. 2018 Joint image reconstruction method with correlative multi-channel prior for X-ray spectral computed tomography. *Inverse Problems* **34**.
11. Wu W, Zhang Y, Wang Q, Liu F, Luo F, Yu H. 2018 Spatial-spectral cube matching frame for spectral CT reconstruction. *Inverse Problems* **34**.
12. Hu Y, Nagy JG, Zhang J, Andersen MS. 2019 Nonlinear optimization for mixed attenuation polyenergetic image reconstruction. *Inverse Problems* **35**.
13. Abascal JFPJ, Ducros N, Peyrin F. 2018 Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance. *Inverse Problems* **34**.
14. Kai C, Min J, Yu J, Yi S. 2019 Multi-energy CT reconstruction using nonlocal total nuclear generalized variation. .
15. Malone E, dos Santos GS, Holder D, Arridge S. 2014 Multifrequency electrical impedance tomography using spectral constraints. *IEEE Trans Med Imaging* **33**, 340–350.
16. Malone E, dos Santos G, Holder D, Arridge S. 2015 A reconstruction-classification method for multifrequency electrical impedance tomography. *IEEE transactions on medical imaging* **34**, 1486–1497.
17. Corlu A, Choe R, Durduran T, Lee K, Schweiger M, Arridge S, Hillman E, Yodh A. 2005 Diffuse optical tomography with spectral constraints and wavelength optimization. *Applied Optics* **44**, 2082–2089.
18. Razansky D, Distel M, Vinegoni C, Ma R, Perrimon N, Köster, W. R, Ntziachristos V. 2009 Multispectral opto-acoustic tomography of deep-seated fluorescent proteins in vivo. *Nature Photonics* **3**, 412–417.

19. Cox B, Arridge S, Beard P. 2009 Estimating chromophore distributions from multiwavelength photoacoustic images. *Journal Optical Society America A* **26**, 443–455.

20. Bal G, Ren K. 2012 On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inv. Prob.* **28**, 025010.

21. An L, Cox B. 2018 Estimating relative chromophore concentrations from multiwavelength photoacoustic images using independent component analysis. *Journal of biomedical optics* **23**, 076007.

22. Blumensath T, Boardman R. 2015 Non-convexly constrained image reconstruction from nonlinear tomographic X-ray measurements. *Philosophical Transactions of the Royal Society A* **373**, 20140393.

23. Wang W, Tilley S, Tivnan M, Stayman JW. 2019 Local response prediction in model-based CT material decomposition. *Proceedings of SPIE–the International Society for Optical Engineering* **11072**.

24. Bousse A et al.. 2016 Joint Activity/Attenuation Reconstruction in SPECT Using Photopeak and Scatter Sinograms. *IEEE NSS/MIC Conf. Rec*.

25. Courdurier M, Monard F, Osses A, Romero F. 2015 Simultaneous source and attenuation reconstruction in SPECT using ballistic and single scattering data. *Inverse Problems* **31**, 095002 (30pp).

26. Rahman MA, Zhu Y, Clarkson E, Kupinski MA, Frey EC, Jha AK. 2020 Fisher information analysis of list-mode SPECT emission data for joint estimation of activity and attenuation distribution. *Inverse Problems* **36**, 084002.

27. Berker Y, Schulz V, Karp JS. 2019 Algorithms for joint activity–attenuation estimation from positron emission tomography scatter. *EJNMMI Physics* **6**.

28. Berker Y, Karp JS, Schulz V. 2017 Joint Reconstruction of PET Attenuation and Activity from Scattered and Unscattered Data. *IEEE NSS/MIC Conference Record*.

29. Brusaferri L, Bousse A, Emond EC, Brown R, Tsai YJ, Atkinson D, Ourselin S, Watson CC, Hutton BF, Arridge S, Thielemans K. 2020 Joint Activity and Attenuation Reconstruction From Multiple Energy Window Data With Photopeak Scatter Re-Estimation in Non-TOF 3-D PET. *IEEE Transactions on Radiation and Plasma Medical Sciences* **4**, 410–421.

30. Trémoulhéac B, Dikaios N, Atkinson D, Arridge S. 2014 Dynamic MR image reconstruction—separation from undersampled (k-t)-space via low-rank plus sparse prior. *IEEE Trans. Med. Imaging* **33**, 1689—1701.

31. Otazo R, Candès E, Sodickson DK. 2015 Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magn. Reson. Med.* **73**, 1125–1136.

32. Prince S, Kolehmainen V, Kaipio J, Franceschini M, Boas D, Arridge S. 2003 Time-series estimation of biological factors in optical diffusion tomography. *Physics in medicine & biology* **48**, 1491.

33. Rasch J, Kolehmainen V, Nivajarvi R, Kettunen M, Gröhn O, Burger M, Brinkmann EM. 2018 Dynamic MRI reconstruction from undersampled data with an anatomical prescan. *Inverse Problems* **34**.

34. Cruz G, Jaubert O, Botnar R, Prieto C. 2019 Cardiac magnetic resonance fingerprinting: Technical developments and initial clinical validation. *Current cardiology reports* **21**, 91.

35. Bilgic B, Goyal VK, Adalsteinsson E. 2011 Multi-Contrast Reconstruction with Bayesian Compressed Sensing. *Magnetic Resonance in Medicine* **66**, 1601–15.

36. Huang J, Chen C, Axel L. 2014 Fast Multi-contrast MRI Reconstruction. *Magnetic Resonance Imaging* **32**, 1344–52.

37. Xiang L, Chen Y, Chang W, Zhan Y, Lin W, Wang Q, Shen D. 2019 Deep-Learning-Based Multi-Modal Fusion for Fast MR Reconstruction. *IEEE Transactions on Biomedical Engineering* **66**, 2105–2114.

38. Ma D, Gulani V, Seiberlich N, Liu K, Sunshine JL, Duerk JL, Griswold MA. 2013 Magnetic Resonance Fingerprinting. *Nature* **495**, 187–92.

39. Tang S, Fernandez-Granda C, Lannuzel S, Bernstein B, Lattanzi R, Cloos M, Knoll F, Asslander J. 2018 Multicompartment magnetic resonance fingerprinting. *Inverse Problems* **34**.

40. Sodickson DK, Feng L, Knoll F, Cloos M, Ben-Eliezer N, Axel L, Chandarana H, Block KT, Otazo R. 2015 The Rapid Imaging Renaissance: Sparser Samples, Denser Dimensions, and Glimmerings of a Grand Unified Tomography. In *Proceedings of SPIE* vol. 9417 pp. 94170G1–14.

41. Ehrhardt MJ, Thielemans K, Pizarro L, Atkinson S, Ourselin S, Hutton BF, Arridge SR. 2015 Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems* **31**.

42. Rasch J, Brinkmann EM, Burger M. 2018 Joint reconstruction via coupled Bregman iterations with applications to PET-MR imaging. *Inverse Problems* **34**.
43. Deidda D, Karakatsanis NA, Robson PM, Tsai YJ, Efthimiou N, Thielemans K, Fayad ZA, Aykroyd RG, Tsoumpas C. 2019 Hybrid PET–MR list-mode kernelized expectation maximization reconstruction. *Inverse Problems* **35**.
44. Matthews TP, Anastasio MA. 2017 Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements. *Inverse Problems* **33**.
45. Haber E, Oldenburg D. 1997 Joint inversion: a structural approach. *Inverse Problems* **13**, 63.
46. Haber E, Holtzman Gazit M. 2013 Model Fusion and Joint Inversion. *Surv. Geophys.* **34**, 675–695.
47. Gallardo LA, Meju MA. 2003 Characterization of heterogeneous near-surface materials by joint 2D inversion of DC resistivity and seismic data. *Geophys. Res. Lett.* **30**, 1658.
48. Gallardo LA, Meju MA. 2004 Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *J. Geophys. Res.: Solid Earth* **109**, B03311.
49. Gallardo LA. 2007 Multiple cross-gradient joint inversion for geospectral imaging. *Geophys. Res. Lett.* **34**, L19301.
50. Gallardo LA, Meju MA. 2011 Structure-coupled multiphysics imaging in geophysical sciences. *Rev. Geophys.* **49**, RG1003.
51. Rovetta D, Colombo D. 2018 Analysis of inter-domain coupling constraints for multi-physics joint inversion. *Inverse Problems* **34**.
52. Crestel B, Stadler G, Ghattas O. 2019 A comparative study of structural similarity and regularization for joint inverse problems governed by PDEs. *Inverse Problems* **35**.
53. Arridge SR, Scherzer O. 2012 Imaging from coupled physics. *Inverse Problems* **28**, 080201.
54. Elbau P, Mindrinos L, Scherzer O. 2018 Quantitative reconstructions in multi-modal photoacoustic and optical coherence tomography imaging. *Inverse Problems* **34**.
55. Smyl D, Antin KN, Liu D, Bossuyt S. 2018 Coupled digital image correlation and quasi-static elasticity imaging of inhomogeneous orthotropic composite structures. *Inverse Problems* **34**.
56. Adesokan BJ, Knudsen K, Krishnan VP, Roy S. 2018 A fully non-linear optimization approach to acousto-electric tomography. *Inverse Problems* **34**.
57. Kaipio JP, Kolehmainen V, Vauhkonen M, Somersalo E. 1999 Inverse problems with structural prior information. *Inverse Problems* **15**, 713.
58. Burger M, Osher S. 2013 A Guide to the TV Zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging* vol. 2090*Lecture Notes in Mathematics* pp. 1–70. Cham: Springer International Publishing.
59. Bredies K, Kunisch K, Pock T. 2010 Total Generalized Variation. *SIAM Journal on Imaging Sciences* **3**, 492–526.
60. Ehrhardt MJ, Betcke MM. 2016 Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences* **9**, 1084–1106.
61. Bredies K, Holler M. 2020 Higher-order total variation approaches and generalisations. *Inverse Problems* **36**, 123001.
62. Fessler J, Clinthorne N, Rogers W. 1992 Regularized emission image reconstruction using imperfect side information. *IEEE Transactions on Nuclear Science* **39**, 1464–1471.
63. Leahy R, Yan X. 1991 Incorporation of anatomical MR data for improved functional imaging with PET. In *12th Int. Conf. Information Processing in Medical Imaging, (Wye, UK, 7–12 July 1991)* p. 105–120. Springer (Berlin).
64. Bowsher JE, Yuan H, Hedlund L, Turkington T, Akabani G, Badea A, Kurylo W, Wheeler C, Cofer G, Dewhirst M, Johnson GA. 2004 Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors. *IEEE Nuclear Science Symp. and Medical Imaging Conf.* p. 2488–92.
65. Ellis S, Mallia A, McGinnity CJ, Cook GJR, Reader AJ. 2018 Multitracer Guided PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2**, 499–509.
66. Bland J, Mehranian A, Belzunce MA, Ellis S, Costa-Luis Cd, McGinnity CJ, Hammers A, Reader AJ. 2019 Intercomparison of MR-informed PET image reconstruction methods. *Medical Physics* **46**, 5055–5074.
67. Wang G, Qi J. 2015 PET Image Reconstruction Using Kernel Method. *IEEE Transactions on Medical Imaging* **34**, 61–71.
68. Hutchcroft W, Wang G, Chen KT, Catana C, Qi J. 2016 Anatomically-aided PET reconstruction using the kernel method. *Physics in Medicine and Biology* **61**, 6668–6683.

69. Novosad P, Reader AJ. 2016 MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions. *Physics in Medicine and Biology* **61**, 4624–4644.

70. Marquis H, Deidda D, Gillman A, Willowson K, Gholami Y, Hioki T, Eslick E, Thielemans K, Bailey D. 2020 Theranostic SPECT Reconstruction for Improved Resolution: Application to Radionuclide Therapy Dosimetry. *European Journal of Nuclear Medicine and Molecular Imaging*. submitted.

71. Baikejiang R, Zhao Y, Fite BZ, Ferrara KW, Li C. 2017a Anatomical image-guided fluorescence molecular tomography reconstruction using kernel method. *Journal of Biomedical Optics* **22**, 055001.

72. Baikejiang R, Zhang W, Zhu D, Hernandez AM, Shakeri SA, Wang G, Qi J, Boone JM, Li C. 2017b Kernel-based anatomically-aided diffuse optical tomography reconstruction. *Biomedical Physics & Engineering Express* **3**, 055002.

73. Wang G. 2019 High Temporal-Resolution Dynamic PET Image Reconstruction Using a New Spatiotemporal Kernel Method. *IEEE Transactions on Medical Imaging* **38**, 664–674.

74. Bland J, Belzunce MA, Ellis S, McGinnity CJ, Hammers A, Reader AJ. 2018 Spatially Compact MR-Guided Kernel EM for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* **2**, 470–482.

75. Tang J, Yang B, Wang Y, Ying L. 2016 Sparsity-constrained PET image reconstruction with learned dictionaries. *Physics in Medicine and Biology* **61**, 6347–6368.

76. Tahaei MS, Reader AJ. 2016 Patch-based image reconstruction for PET using prior-image derived dictionaries. *Physics in Medicine and Biology* **61**, 6833–6855.

77. Sudarshan VP, Chen Z, Awate SP. 2018 Joint PET+MRI Patch-Based Dictionary for Bayesian Random Field PET Reconstruction. In Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* Lecture Notes in Computer Science pp. 338–346 Cham. Springer International Publishing.

78. Heinrich MP, Simpson IJ, Papież BW, Brady SM, Schnabel JA. 2016 Deformable image registration by combining uncertainty estimates from supervoxel belief propagation. *Medical Image Analysis* **27**, 57–71.

79. Jiao J, Markiewicz P, Burgos N, Atkinson D, Hutton B, Arridge S, Ourselin S. 2015 Detail-Preserving PET Reconstruction with Sparse Image Representation and Anatomical Priors. In Ourselin S, Alexander DC, Westin CF, Cardoso MJ, editors, *Information Processing in Medical Imaging* Lecture Notes in Computer Science pp. 540–551 Cham. Springer International Publishing.

80. Ehrhardt MJ, Thielemans K, Pizarro L, Atkinson D, Ourselin S, Hutton BF, Arridge SR. 2015 Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems* **31**, 015001.

81. Chen C, Li Y, Huang J. 2013 Calibrationless Parallel MRI with Joint Total Variation Regularization. In *Medical Image Computing and Computer-Assisted Intervention* pp. 106–114.

82. Bresson X, Chan TF. 2008 Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problem & Imaging* **2**, 455.

83. Moreno JC, Prasath VBS, Neves JC. 2016 Color image processing by vectorial total variation with gradient channels coupling. *Inverse Problems & Imaging* **10**, 461.

84. Knoll F, Koesters T, Otazo R, Block T, Feng L, Vunckx K, Faul D, Nuyts J, Boada F, Sodickson DK. 2014 Joint reconstruction of simultaneously acquired MR-PET data with multi sensor compressed sensing based on a joint sparsity constraint. In *PSMR 2013* vol. 1 pp. A26+.

85. Ehrhardt MJ, Arridge SR. 2014 Vector-Valued Image Processing by Parallel Level Sets. *IEEE Trans. Med. Processing* **23**, 9–18.

86. Ehrhardt MJ. 2015 *Joint Reconstruction for Multi-Modality Imaging with Common Structure*. Phd thesis University College London.

87. Holt KM. 2014 Total Nuclear Variation and Jacobian Extensions of Total Variation for Vector Fields. *IEEE Transactions on Image Processing* **23**, 3975–3989.

88. Wu Z, Wang Q, Jin J, Shen Y. 2017 Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising. *Signal Processing* **131**, 202–219.

89. Knoll F, Holler M, Koesters T, Otazo R, Bredies K, Sodickson DK. 2017 Joint MR-PET reconstruction using a multi-channel image regularizer. *IEEE Trans. Med. Imaging* **36**, 1–16.

90. Koenderink JJ. 1990 *Solid Shape*. MIT Press.

91. Romeny BMH. 1994 *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic.

92. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. 1997 Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* **16**, 187–198.

93. Nuyts J. 2007 The use of mutual information and joint entropy for anatomical priors in emission tomography. In *Nuclear Science Symp. Conf. Record* p. 4149–4154.

94. Somayajula S, Asma E, Leahy RM. 2005 PET image reconstruction using anatomical information through mutual information based priors. *IEEE Nuclear Science Symposium Conference Record* **5**, 2722–2726.

95. Panagiotou C, Somayajula S, Gibson A, Schweiger M, Leahy RM, Arridge S. 2009 Information Theoretic Regularization in Diffuse Optical Tomography. *J. Opt. Soc. Am. A* **26**, 1277–1290.

96. Somayajula S, Panagiotou C, Rangarajan A, Li Q, Arridge S, Leahy R. 2011 PET Image Reconstruction Using Information Theoretic Anatomical Priors. *IEEE Trans Med Imag* **30**, 537–549.

97. Tang J, Rahmim A. 2019 Bayesian PET image reconstruction incorporating anato-functional joint entropy. *Physics in Medicine and Biology* **54**, 7063–7075.

98. Mehranian A, Belzunce MA, Niccolini F, Politis M, Prieto C, Turkheimer F, Hammers A, Reader AJ. 2017 PET image reconstruction using multi-parametric anato-functional priors. *Physics in Medicine and Biology* **62**, 5975–6007.

99. Tang J, Rahmim A. 2015 Anatomy assisted PET image reconstruction incorporating multi-resolution joint entropy. *Physics in Medicine and Biology*.

100. Arridge S, Simmons A. 1997 Multi-Spectral Probabilistic Diffusion Using Bayesian Classification. In ter Haar Romeny B, Florak L, Koenderink J, Viergever M, editors, *Scale-Space Theory in Computer Vision : Proc. First Intl. conf., Scale-Space '97, Utrecht, The Netherlands* pp. 224–235. Springer-Verlag (Berlin).

101. Gabay D, Mercier B. 1976 A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *Computers & Mathematics with Applications* **2**, 17–40.

102. Mehranian A, Belzunce MA, Prieto C, Hammers A, Reader AJ. 2018 Synergistic PET and SENSE MR Image Reconstruction Using Joint Sparsity Regularization. *IEEE Transactions on Medical Imaging* **37**, 20–34.

103. Mehranian A, Belzunce MA, McGinnity CJ, Bustin A, Prieto C, Hammers A, Reader AJ. 2019 Multi-modal synergistic PET and MR reconstruction using mutually weighted quadratic priors. *Magnetic resonance in medicine* **81**, 2120–2134.

104. Osher S, Burger M, Goldfarb D, Xu J, Yin W. 2005 An Iterative Regularization Method for Total Variation-based Image Restoration. *Multiscale Modelling and Simulation* **4**, 460–489.

105. Benning M, Burger M. 2018 Modern regularization methods for inverse problems. *Acta Numerica* **27**, 1–111.

106. Moeller M, Brinkmann E, Burger M, Seybold T. 2014 Color Bregman TV. *SIAM J. Imaging Sci.* **7**, 2771–2806.

107. Sudarshan VP, Egan GF, Chen Z, Awate SP. 2020 Joint PET-MRI image reconstruction using a patch-based joint-dictionary prior. *Medical Image Analysis* **62**, 101669.

108. Song P, Weizman L, Mota JF, Eldar YC, Rodrigues MR. 2020 Coupled Dictionary Learning for Multi-Contrast MRI Reconstruction. *IEEE Transactions on Medical Imaging* **39**.

109. Deidda D, Karakatsanis NA, Robson PM, Efthimiou N, Fayad ZA, Aykroyd RG, Tsoumpas C. 2019 Effect of PET-MR Inconsistency in the Kernel Image Reconstruction Method. *IEEE Transactions on Radiation and Plasma Medical Sciences* **3**, 400–409.

110. Bungert L, Coomes DA, Ehrhardt MJ, Rasch J, Reisenhofer R, Schönlieb CB. 2018 Blind image fusion for hyperspectral imaging with the directional total variation. *Inverse Problems* **34**, 044003.

111. Bungert L, Ehrhardt MJ. 2020 Robust Image Reconstruction with Misaligned Structural Information. *arXiv:2004.00589 [cs, eess, math]*.

112. Tsai YJ, Bousse A, Arridge S, Stearns CW, Hutton BF, Thielemans K. 2020 Penalized PET/CT Reconstruction Algorithms with Automatic Realignment for Anatomical Priors. *IEEE Transactions on Radiation and Plasma Medical Sciences* pp. 1–1. Conference Name: IEEE Transactions on Radiation and Plasma Medical Sciences.

113. Batchelor PG, Atkinson D, Irarrazaval P, Hill DLG, Hajnal J, Larkman D. 2005 Matrix description of general motion correction applied to multishot images. *Magn. Reson. Med.* **54**, 1273–1280.

114. Bousse A, Bertolli O, Atkinson D, Arridge S, Ourselin S, Hutton BF, Thielemans K. 2016 Maximum-Likelihood Joint Image Reconstruction/Motion Estimation in Attenuation-Corrected Respiratory Gated PET/CT Using a Single Attenuation Map. *IEEE Transactions on Medical Imaging* **35**, 217–228. Citation Key Alias: bousse2015MaximumLikelihoodJointImage.

115. Ruan D, Fessler JA, Esedoglu S. 2008 *Discontinuity preserving regularization for modeling sliding in medical image registration*.

116. Fieseler M, Gigengack F, Jiang X, Schäfers KP. 2014 Motion Correction of Whole-Body PET Data with a Joint PET-MRI Registration Functional. *BioMedical Engineering OnLine* **13**, S2.

117. Kolbitsch C, Neji R, Fenchel M, Schuh A, Mallia A, Marsden P, Schaeffter T. 2018 Joint cardiac and respiratory motion estimation for motion-corrected cardiac PET-MR. *Physics in Medicine & Biology* **64**, 015007.

118. Arridge S, Maass P, Öktem O, Schönlieb CB. 2019 Solving inverse problems using data-driven models. *Acta Numerica* **28**, 1–174.

119. Hammernik K, Knoll F. 2020 Chapter 2 - Machine learning for image reconstruction. In Zhou SK, Rueckert D, Fichtinger G, editors, *Handbook of Medical Image Computing and Computer Assisted Intervention* pp. 25–64. Academic Press.

120. Reader AJ, Corda G, Mehranian A, Costa-Luis Cd, Ellis S, Schnabel JA. 2020 Deep Learning for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* pp. 1–1. Conference Name: IEEE Transactions on Radiation and Plasma Medical Sciences.

121. Ovtchinnikov E, Brown R, Pasca E, da Costa-Luis CO, Thomas B, Atkinson D, Mayer J, Gillman A, Kolbitsch C, Ehrhardt MJ, Whitehead AC, Thielemans K. 2020 Synergistic Image Reconstruction Framework. .

122. Adler J, Kohr H, Ringh A, Moosmann J, sbanert, Ehrhardt MJ, Lee GR, niinimaki, bgris, Verdier O, Karlsson J, zickert, Palenstijn WJ, Öktem O, Chen C, Loarca HA, Lohmann M. 2018 Operator Discretisation Library ODL. .

123. Jørgensen JS, Ametova E, Burca G, Fardell G, Papoutsellis E, Pasca E, Thielemans K, Turner M, Warr R, Lionheart WRB, Withers PJ. 2020 Core Imaging Library – Part I: a versatile Python framework for tomographic imaging. preprint.

124. Papoutsellis E, Ametova E, Delplancke C, Fardell G, Jørgensen JS, Pasca E, Turner M, Warr R, Lionheart WRB, Withers PJ. 2020 Core Imaging Library – Part II: Multi-channel reconstruction for dynamic and spectral tomography. preprint.

125. Scherzer O, Grasmair M, Grossauer H, Haltmeier M, Lenzen F. 2008 *Variational Methods in Imaging* vol. 167*Applied Mathematical Sciences Series*. Springer-Verlag, New York 1 edition.

126. Ito K, Jin B. 2014 *Inverse Problems - Tikhonov Theory and Algorithms*. World Scientific.

127. Perona P, Malik J. 1990 Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, 629–639.

128. Weickert J. 1998 *Anisotropic Diffusion in Image Processing*. Teubner Stuttgart.

129. Geman S, Geman D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 413–25.

130. Gilboa G, Osher S. 2007 Nonlocal linear image regularization and supervised segmentation. *Multiscale Modeling and Simulation* **6**, 595–630.

131. Smith S, Brady J. 1997 SUSAN–A New Approach to Low Level Image Processing. *International Journal of Computer Vision* **23**, 45–78.

132. Buades A, Coll B, Morel JM. 2005 A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition (CVPR'05)* vol. 2 pp. 60–65. IEEE.

133. Belkin M, Niyogi P. 2013 Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**, 1373–1396.

134. Wang Z, Nasrabadi N, Huang T. 2015 Semisupervised hyperspectral classification using task-driven dictionary learning with Laplacian regularization. *IEEE Trans. Geosci. Remote Sens.* **53**, 1161–1173.

135. Pang J, Cheung G. 2017 Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain. *arXiv*.

136. Tomasi C, Manduchi R. 1998 Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision* pp. 839–846. (IEEE Cat. No.98CH36271), Bombay, India,.

137. Candés EJ, Wakin MB. 2008 An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine* **25**, 21–30.

138. Olshausen BA, Field DJ. 1997 Sparse coding with an overcomplete basis set: A strategy employed by V1?. *Vision Research* **37**, 3311–3325.

139. Tošić I, Frossard P. 2011 Dictionary Learning. *IEEE Signal Processing Magazine* **28**, 27–38.
140. Hofmann T, Schölkopf B, Smola AJ. 2008 Kernel Methods in Machine Learning. *The Annals of Statistics* **36**, 1171–1220.
141. Meinhardt T, Moeller M, Hazirbad C, Cremers D. 2017 Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems. In *International Conference on Computer Vision* pp. 1781–1790.
142. He K, Zhang X, Ren S, Sun J. 2016 Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*.
143. Romano Y, Elad M, Milanfar P. 2017 The little engine that could: Regularization by denoising (RED). *SIAM J. Imag. Sci* **10**, 1804–1844.

A. Mathematical Tools for Image Reconstruction

In this appendix we briefly outline some main concepts and methods used in image reconstruction, which we frame in the context of ill-posed inverse problems. There is an ever growing number of strategies for the design of regularisation functionals to ameliorate ill-posedness using geometrical, statistical or learned approaches. We mention some key principles here. For a general introduction to the regularization of inverse imaging problems, we refer the reader to [105,125,126].

We note that there are some fundamental differences between the *Continuous Setting* and the *Discrete Setting* which are related through (2.1). Any approach derived in the continuous setting needs to be discretised to allow a computational implementation. However, some approaches start directly from a discretisation, where the continuous functions are written in terms of basis functions, normally as a linear combination as in (2.1). Regularisation can then proceed by choosing appropriate basis functions, or by adding a penalty $\Psi(f)$ on the coefficients, or both.

(a) Aspects of Optimisation

The fundamental algorithm to optimise the variational representation (2.2) is gradient descent which given an initial guess $f^{(k)}$ computes the iterative updates

$$f^{(k+1)} = f^{(k)} - \tau_k \left(\mathcal{D}'(f^k) + \alpha \Psi'(f^k) \right) \tag{A 1}$$

where τ_k is a step length which can be constant or varying with the iterations.

Rather than an explicit classical descent strategy, a more general approach is the *Proximal Gradient Descent* (PGD) which take the form of two steps

$$\text{likelihood update step} \qquad f^{(k+1/2)} \;=\; f^{(k)} - \tau_k \mathcal{D}'(f^{(k)}) \tag{A 2}$$

$$\text{proximal step} \qquad f^{(k+1)} \;=\; \text{prox}_{\tau_k \alpha \Psi} \left(f^{(k+1/2)} \right) \tag{A 3}$$

where the proximal operator solves the auxiliary problem

$$\text{prox}_{\tau_k \alpha \Psi}(z) = \arg\min_f \left\{ \Phi_z(f) := \frac{1}{2} \|f - z\|^2 + \tau_k \alpha \Psi(f) \right\} \tag{A 4}$$

A possible approximate solution of the proximal step is via the evolution of a PDE induced by the form of the regularisation functional [125]

$$\frac{\partial f}{\partial t} = -\Phi'_{f^{(k+1/2)}}(f) = f - f^{(k+1/2)} + \tau_k \alpha \Psi'(f) \tag{A 5}$$

(b) Regularisation and Image Diffusion

In the continuous setting, when following the classical approach (A 1) we require the gradient $\Psi' \equiv \frac{\partial \Psi}{\partial f}$, which corresponds to the first variation (Euler-Lagrange equation) if Ψ is defined in

variational form; e.g.

$$\Psi(f) := \int_\Omega \psi(|\nabla f|)\mathrm{d}x \quad \to \quad \Psi'(f) = -\nabla \cdot \underbrace{\left(\frac{\psi'(|\nabla f|)}{|\nabla f|}\right)}_{\kappa}\nabla f \tag{A 6}$$

Furthermore, as suggested by (A 5), interpreting the iterative steps as a time evolution suggests the interpretation of the minimisation of the prior as an image flow. In the choice given in (A 6) this is of *diffusion* type, since the term on the right is a second order derivative, i.e.

$$\frac{\partial f(x)}{\partial t} = \nabla \cdot (\kappa(x)\nabla f(x)) , \tag{A 7}$$

where $\kappa(x)$ plays the role of a spatially varying diffusivity. The local function ψ in (A 6) (usually taken to be convex) admits many commonly used regularisation schemes including first order Tikhonov ($\psi(s) = \frac{1}{2}s^2$) and total variation (TV) ($\psi(s) = s$). A particular function that we will refer to in this article is the *Perona-Malik* function [127] in one of the forms

$$\text{PM1} \quad \psi(t) \quad := \quad \frac{\epsilon^2}{2}\log\left[1 + \frac{t^2}{\epsilon^2}\right] \quad \to \quad \psi'(t) = \frac{\epsilon^2 t}{\epsilon^2 + t^2} \tag{A 8}$$

$$\text{PM2} \quad \psi(t) \quad := \quad \frac{\epsilon^2}{2}\left[1 - \exp\left(-\frac{t^2}{\epsilon^2}\right)\right] \quad \to \quad \psi'(t) = t\exp\left(-\frac{t^2}{\epsilon^2}\right) \tag{A 9}$$

with ϵ a threshold indicating a level below which small gradients are considered as noise. The resultant diffusivities $\kappa = \frac{\psi'(|\nabla f|)}{|\nabla f|} \in [0, 1]$ can be interpreted as *edge-indicator* functions.

We may also define a flow without it being the variation of a functional form. Weikert [128] proposed using a tensor

$$\frac{\partial f}{\partial t} = \nabla \cdot (\mathrm{D}(\mathrm{J}_\rho(\nabla f_\sigma))\nabla f \tag{A 10}$$

where the (symmetric, positive semi-definite) structure tensor is constructed as

$$\mathrm{J}_\rho(\nabla f_\sigma)(\mathrm{x}) = G_\rho * (\nabla f_\sigma \nabla f_\sigma^T) = \left[\begin{array}{cc} G_\rho * f_x^2 & G_\rho * f_y f_x \\ G_\rho * f_x f_y & G_\rho * f_y^2 \end{array}\right], \tag{A 11}$$

with eigensystem $\{\eta_k, \mathbf{v}_k\}$ and

$$\mathrm{D} = [\mathbf{v}_1\ \mathbf{v}_2]\left[\begin{array}{cc} \zeta_1 & 0 \\ 0 & \zeta_2 \end{array}\right]\left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{array}\right] = \left[\begin{array}{cc} D_{11} & D_{12} \\ D_{12} & D_{22} \end{array}\right]. \tag{A 12}$$

Furthermore $\mathbf{v}_1 \simeq \hat{\nu}, \mathbf{v}_2 \simeq \hat{\tau}$, with equality as $\sigma, \rho \to 0$. Then the *Edge Enhancing Diffusion* (EED) approach is designed to reduce smoothing in the normal direction by an an edge-indicator derived from (A 9)

$$\zeta_1 = \kappa_\epsilon(|\nabla f|) = \exp\left(-|\nabla f|^2/\epsilon^2\right); \quad \zeta_2 = 1 \tag{A 13}$$

(c) Markov Random Fields

A fundamental concept in the discrete setting is the Markov Random Field (MRF) [129] :

$$\Psi(f) \quad = \quad \frac{1}{p}\sum_i \sum_{j\in\mathcal{N}(i)} w_{ij}|f_i - f_j|^p \tag{A 14}$$

$$\frac{\partial\Psi}{\partial f_i} \quad = \quad \sum_{j\in\mathcal{N}(i)} w_{ij}|f_i - f_j|^{p-1} \tag{A 15}$$

Sometimes the discrete MRF corresponds to the discretisation of a continuous functional; e.g for $p = 2$ we have a quadratic form

$$\Psi(f) = \langle f, Lf \rangle \quad \text{with } L_{ij} = \begin{cases} w_{ij} & j \neq i \\ -\sum_{j\neq i} w_{ij} & i = j \end{cases} \tag{A 16}$$

Taking a 4-connected neighbourhood with uniform weights $w_{ij} = 1$ corresponds to the discretisation of the Laplacian $L \simeq -\nabla^2$ which is the gradient of the first order Tikhonov prior $\psi(s) = \frac{1}{2}s^2$. However, in general, it is not always possible to explain MRFs as discretisation of a continuous model.

The MRF concept extends to a global one and the concept of *Non-local regularisation* [130]. For example the Laplacian in the local mathematical sense of a second order derivative extends to the *Graph Laplacian* [131–135]. The conjunction of an MRF with kernel-based methods (see § A(e)) is related to the so-called *bilateral filtering* technique in image processing [136].

(d) Sparsity

Sparsity has been a prominent concept in regularisation for several decades, closely connected to the principles of *Compressed Sensing* [137]. It can be formulated in both continuous and discrete settings. The assumption is that under some (possibly invertible) transform

$$\xi = \mathcal{T}(f) \quad \leftrightarrow \quad f = \mathcal{T}^{-1}(\xi) \tag{A 17}$$

the transformed parameters ξ have many/mostly zero or close to zero, which can be described by specifying that the zero-norm $L_0 := |\xi|_0$, which simply counts the non-zero components in ξ, should be minimised. However, since optimisation of L_0 norm leads to a non-convex problem, it is conveniently replaced by its *convex relaxation* given by the L_1-norm. Possibilities for the transform $\mathcal{T}$ include finite differences (gradient regularisation), wavelets, and Neural Networks (NNs) where $\mathcal{T}$ is called an encoder and $\mathcal{T}^{-1}$ a *decoder*. Methods employing sparsity also allow the possibility that the transform space is *over-complete*, i.e. the solution has a non-unique representation in the basis; an example is the use of a *dictionary* learned from example solutions [138,139].

(e) Kernel Methods

Kernel Methods compute properties such as classifiers in terms of transformed *feature vectors* obtained from one or more related images or training data [140]. The transformation can be a (generally non-linear) mapping to a high dimensional space. The features may be obtained from any abstraction model, such as patches, or geometric or statistical measures. The inner product in the transformed space defines a similarity measure on the features. In practice, the transformation is determined by specifying this similarity measure $\mathcal{K}$, called the "kernel function". Although sparsifying tranforms and feature vectors are not synonymous, For simplicity of representation, we use the same notation as (A 17)

$$\mathcal{K}(\mathcal{T}_i(f), \mathcal{T}_j(f)) \tag{A 18}$$

where $\mathcal{T}_i$ is now interpreted as the feature vector associated with pixel i and f represents the image(s) on which the features are computed. A common choice for the kernel function is the Radial Basis Function (RBF)

$$\mathcal{K}(t_1, t_2) := \exp \left\{ -\frac{||t_1 - t_2||^2}{2\sigma^2} \right\}. \tag{A 19}$$

The "kernel trick" consists of computing linear functions of the transformed variables in terms of $\mathcal{K}$.

(f) Regularisation using Neural Networks

There are many techniques for combining inverse problems with artificial intelligence in general and NNs particular; see [118] for a review.

Here we mention only one natural approach which is to replace the proximal operator (A 3) by a learned operator such that

$$f^{(k+1)} = F_\Theta \left(f^{(k+1/2)} \right) \tag{A 20}$$

where $F_\Theta : X \to X$ is a NN trained on suitable pairs of ground truth and approximately reconstructed images, and where Θ represents the weights and other parameters of the network architecture [141].

Alternatively we could consider the evolution (A 5) as an image update

$$f \mapsto f + F_\Theta \left(f^{(k+1/2)} \right) \tag{A 21}$$

where F_Θ now takes the form of a *Residual Neural Net* [142].

In neither approach is the update is explicitly derived as the variation of a function which leads to difficulties in corresponding convergence guarantees, and prevents an explicit Bayesian interpretation as the maximisation of a posterior. However, not that the *Regularisation by Denoising* (RED) framework provides a general technique to interpret denoising algorithms as variational methods under certain restrictions of their properties [143].