青年评述

图像反问题中的数学与深度学习方法[*1)]

董 彬
(北京大学北京国际数学研究中心, 北京, 100871)

摘 要

我们生活在数字的时代, 数据已经成为了我们生活中不可或缺的一部分, 而图像无疑是最重要的数据类型之一. 图像反问题, 包括图像降噪, 去模糊, 修复, 生物医学成像等, 是图像科学中的重要领域. 计算机技术的飞速发展使得我们可以用精细的数学和机器学习工具来为图像反问题设计有效的解决方案. 本文主要回顾图像反问题中的三大类方法, 即以小波 (框架) 为代表的计算调和分析法、偏微分方程 (PDE) 方法和深度学习方法. 我们将回顾这些方法的建模思想和一些具体数学形式, 探讨它们之间的联系与区别, 优点与缺点, 探讨将这些方法有机融合的可行性与优势.

关键词: 图像反问题; 图像重建; 医疗影像; 图像识别; 变分模型; 偏微分方程; 小波变换; 卷积神经网络; 深度学习

MR (2010) 主题分类: 35A15, 41A25, 42C40, 45Q05, 65K10, 68U10, 68T05, 68T45

1. 绪 论

我们生活在数字的时代, 数据的产生、传播、处理、整合、解释已经成为我们社会生活、工业发展和科学研究中的重要组成部分. 数字图像无疑是最重要和使用最广泛的数据类型之一. 图像是以简单和直观的方式来展现物理世界. 通过对图像的研究, 有助于还原真实客观世界, 从而发现和预测客观规律. 近年来, 计算机技术的进步, 尤其是并行和分布式计算, 使得人们能够将数学和机器学习中的一些更加精细的工具应用到图像科学的模型构架、算法设计和实现中. 现在, 图像处理和分析技术现已广泛应用于自然科学、工程和多媒体等领域, 并已正式进入每个人的生活当中. 图像反问题, 包括图像降噪, 去模糊, 修复, 生物医学成像等, 是图像科学中最重要的领域之一. 其主要目的是从观测数据重建高质量的图像, 使我们能够在图像中观测到科学问题中所关心的细节特征 (为此, 图像反问题有时也被称为图像重建).

[*] 2019 年 9 月 29 日收到.

[1)] 作者简介: 董彬, 北京大学北京国际数学研究中心长聘副教授、主任助理, 北京大数据研究院深度学习实验室研究员、生物医学影像分析实验室副主任. 2003 年本科毕业于北京大学数学科学学院、2005 年在新加坡国立大学数学系获得硕士学位、2009 年在美国加州大学洛杉矶分校数学系获得博士学位. 博士毕业后曾在美国加州大学圣迭戈分校数学系任访问助理教授、2011–2014 年在美国亚利桑那大学数学系任助理教授, 2014 年底入职北京大学. 主要研究领域为应用调和分析、反问题计算、深度学习及其在图像和数据科学中的应用. 在理论上, 与合作者一起将图像领域独立发展近 30 年的两个数学分支 (PDE/变分方法和小波方法) 建立深刻的联系, 改变了领域内对这两类方法的一些既定认识. 应用上, 以数学理论为指导思想, 为来源于医学影像、计算机视觉、深度学习等领域中的重要问题提供行之有效的解决方案. 在国际重要学术期刊和会议上发表论文 60 余篇; 现任期刊《Inverse Problems and Imaging》编委. 于 2014 年获得香港求是基金会颁发的求是杰出青年学者奖.

图像重建问题一般可以表示为以下的线性反问题:

$$\boldsymbol{f}=\boldsymbol{A}\boldsymbol{u}+\boldsymbol{\eta} \tag{1.1}$$

其中矩阵 $\boldsymbol{A}$ 是一些线性算子 (通常不可逆), $\boldsymbol{\eta}$ 表示由观察图像中加性噪声引起的扰动, 比如高斯白噪声、拉普拉斯噪声、泊松噪声等. 不同的图像重建问题对应于不同类型的 $\boldsymbol{A}$. 例如, 图像降噪问题对应的 $\boldsymbol{A}$ 是恒同算子; 图像填充问题对应的 $\boldsymbol{A}$ 是一个限制算子; 图像去模糊问题对应的 $\boldsymbol{A}$ 是卷积算子, 当卷积核未知时, 该问题被称为盲去模糊 [1]; 断层扫描 (Computed Tomography, CT) 和核磁共振成像 (Magnetic Resonance Imaging, MRI) 问题对应的 $\boldsymbol{A}$ 分别是 (下采样后的)Radon 变换和傅里叶变换 [2,3]; Quantitative Susceptibility Mapping(QSM) 对应的 $\boldsymbol{A}$ 是个偶极核 (dipole kernel) [4–7]. 图像反问题 (1.1) 中的线性映射 $\boldsymbol{A}$ 通常是不可逆或者病态的, 因此它的求解是非平凡的. 若简单的对 $\boldsymbol{A}$ 求逆, 比如伪逆或者 Tikhonov 正则化 [8,9], 通常会导致重建的图像中噪音被放大, 或者边缘被模糊掉. 一个好的图像重建方法应该能够在保持图像平滑性的同时保留重要的图像特征 (如图像边缘). 然而, 平滑化和特征保持往往是相互矛盾的, 因此高质量的图像重建是一项非常有挑战性的任务.

现在很多流行的图像重建模型和算法都是基于变换的, 该变换可以是线性的 (如傅里叶变换、小波变换), 也可以是非线性的 (非线性偏微分方程、深层卷积神经网络). 傅里叶变换是最早基于变换的方法之一. 傅里叶变换对于平滑和正弦曲线信号很有效, 然而傅里叶变换不能在空间域进行定位, 所以只能很好的表示图像的全局特征. 为此, 人们引入了窗口傅里叶变换 [10] 来克服傅里叶变换的空间定位的不足. 但是, 由于窗口傅里叶变换具有固定时频分辨率, 对于图像来说, 其变换域中高频系数的稀疏性并不理想. 然而, 小波和小波框架具有不同的时间 - 频率分辨率, 这使得它们能够为局部图像特征提供更好的稀疏逼近. 这就是在图像重建中, 小波和小波框架比傅立叶变换或窗口傅立叶变换更有效的原因, 也是正交或双正交小波在图像压缩中取得成功的原因 [11,12].

(小波) 框架等冗余系统已经成功的应用到一些经典和更具挑战性的图像反问题, 并延申图像科学其它问题. 小波框架在经典图像重建问题中的应用包括图像修复 [13], 超分辨率 [14], 去模糊 [15–18], 彩色图像去马赛克 [19] 和彩色图像色彩增强 [20] 等; 小波框架可以应用于更复杂的图像重建问题, 包括图像盲去模糊 [21,22], 图像盲填充 (blind inpainting) [23], 以及未知噪声类型的降噪 [24] 等. 近年来, 基于小波框架相关的算法可以用来进一步提高生物医学成像质量. 例如, X 射线 CT 图像重建 [25–27] 、四维 CT 图像重建 [28,29]、压缩感知 MRI [30–33]、冷冻电子显微成像中的蛋白质分子三维重建 [34].

除了基于变换和稀疏表达的图像重建方法, 基于偏微分方程 (PDE) 的方法也被学术界和工业界广泛的采纳 [35–37]. PDE 方法和诸如小波框架方法有着十分不同的发展轨迹. 两类方法使用的数学工具也有所不同: 小波框架方法基于调和分析; PDE 方法则基于变分和(非线性)PDE. 变分模型的基本思想是将图像理解为某个函数空间中的函数 (如有界变差 (B-V) 函数空间 [38]) 并根据函数空间性质来设计能量泛函, 通过极小化该能量泛函来复原图像. 代表性的变分方法包括全变差模型 (total variation) [38]、广义全变差模型 (total generalized variation) [39]、卷积下确界 (infimal convolution) 模型 [40] 和 Mumford-Shah 模型 [41] 等. 典型的 PDE 模型包括各向异性扩散方程、Perona-Malik 方程 [42]、冲激滤波器 [43]、和基于 Navier-Stokes 方程的图像处理模型 [44] 等.

小波框架方法和 PDE 方法从不同的角度来刻画图像: 小波框架方法大多通过稀疏性来刻画图像, 而 PDE 方法往往是用函数空间和几何来刻画图像. 在图像科学中, 这两大类方法独

立发展了近 30 年. 虽然一些工作表明, 这两类方法在特殊的情况下存在一定的联系 [45, 46], 但是它们之间是否存在一般性的联系一直缺乏系统性研究. 直到最近, 在文献 [47–49] 中, 基于小波框架的图像重建模型和变分模型的一般性联系被提出并给予了严格证明, 这包括全变差模型和小波框架分析模型的联系 [47]、分片光滑小波框架模型与 Mumford-Shah 模型的联系 [48]、更为一般的小波框架模型和更多变分模型之间的联系 [49]. 此外, [50] 建立了迭代小波框架收缩算法与一般非线性演化 PDE 之间的深层联系, 这些非线性演化 PDE 包含了 Perona-Malik 方程和冲激滤波器等在图像重建问题中被普遍采纳的 PDE 模型. 该系列工作将小波框架方法和 PDE 方法在图像问题中统一起来, 一方面赋予了小波框架几何直观, 另一方面也赋予了 PDE 模型稀疏逼近的性质, 使得我们可以设计出融合两类方法的新模型和算法. 在理论上, 该系列工作通过渐进分析严格证明了离散化后的优化问题和迭代算法与对应的变分模型和 PDE 模型的一致性, 完善了这方面的理论工作, 而此前只有个别的变分和 PDE 模型有严格的渐近分析.

最近几年, 深度学习已经在计算机视觉、图像识别、目标检测、语音识别、自然语言处理、文本生成等应用领域取得显著进展, 成为了人工智能研究领域重要的基础性方法. 通过巧妙设计的多层神经网络, 结合 GPU 强大的计算能力, 神经网络可以从大量的样本数据中学习到数据中隐藏的高级语义特征. 基于神经网络, 大量人工智能的方法, 比如语音识别、人脸识别、机器翻译等方法, 逐渐投入了商业应用, 逐渐改变了人们的生活.

深度学习在图像反问题中也取得了令人瞩目的进展. 与传统方法相比, 深度学习方法有区别, 但也有大量相似之处. 如上文所述, 图像重建问题的关键是找到待重建图像的一个很好的稀疏表达, 传统方法大多假设该表达是线性的, 通过人为构造或者基于数据学习, 可以得到图像好的线性稀疏表达; 然而, 深度学习方法假设该表达是由深层神经网络构成, 因此是非线性的稀疏表达, 该表达的模型参数必须从数据中去学习. 代表性工作包括基于单图学习的深层图像先验模型 (Deep Image Prior, DIP) [51]、通过预训练的基于生成对抗网络 (Generative Adversarial Network, GAN) 的图像重建模型 [52–55]、基于展开动力系统 (Unrolling Dynamics, UD) 的方法 [56–58] 等等.

纵观图像反问题过去 30 年的发展, 我们可以把基于数学和机器学习方法的建模分为三个阶段: 1) 模型驱动、模型 + 数据驱动、数据驱动. 传统图像反问题建模基本都是模型驱动, 基于我们对待重建图像的先验知识, 将其转换为数学刻画, 并设计相应的变分、优化或者 PDE 模型. 模型与数据驱动混合建模始于 1999 年的最优方向法 (Method of Optimal Directions, MOD) [59], 这类方法还是以模型驱动为主, 数据驱动为辅, 代表工作包括 K-SVD [60]、数据驱动 (紧) 框架等等 [61, 62]. 不同于前面两类方法, 深度学习方法是一类以数据驱动为主导的方法, 其模型除了一些基本单元之外, 大量参数需要通过数据集训练来确定, 通过精心设计深层网络构架和训练策略, 深度学习方法在很多图像反问题任务中都不同程度上获得了目前最佳图像重建效果.

深度学习模型和传统图像反问题模型 (包括模型驱动和模型与数据混合驱动方法) 在解决实际问题中各有优劣. 大多传统方法对应的模型都有坚实的理论基础, 有良好的可解释和透明度, 但缺点是模型复杂度有限, 不能有效的利用大规模数据集. 深度学习方法设计灵活且有较高的模型复杂度, 伴随随机优化策略, 可以从大规模数据集中有效的提取数据特征, 但其缺点是模型和算法的理论性质难以刻画、工作机理不明确、存在不稳定性等诸多问题. 因此, 在最近几年, 图像反问题领域有越来越多的工作致力于寻找传统建模与深度学习建模的有机融合, 结合两类方法的优点, 在提升模型性能的同时保持模型的可解释性.

本文将介绍部分传统图像重建模型与算法, 讨论这些模型和算法间的联系与区别, 最后介绍深度学习与传统建模方法融合的建模思想与研究现状.

2. 小波框架方法

本节我们将介绍图像反问题中的一些小波框架模型, 对小波框架的理论或应用感兴趣的读者可参考文献 [11, 63–65] (框架和小波框架理论)、 [66](框架的理论和应用的简洁的描述) 和 [67](更详细的描述).

2.1. 小波框架变换

基于 MRA 的小波框架的构造, 尤其是紧小波框架, 始于对平移不变系统的研究 [63, 64, 68–70]. 其中, Gabor 框架的对偶原则 [69]、小波框架的酉扩展原则 (Unitary Extension Principle, UEP) 和混合扩展原则 (Mixed Extension Principle, MEP) [63, 64] 均源于这项研究. 这些理论不仅为小波框架的构造提供了系统的指导, 同时也自然的诱导出了小波框架变换的快速算法. 这里, 我们简短的回顾一下小波框架的基本概念以及 MEP 和 UEP.

定义 1(框架和紧框架). 集合 $X = \{g_j : j \in \mathbb{Z}\} \subset L_2(\mathbb{R}^d)$, $d \in \mathbb{N}$ 被称作一个框架 (frame), 如果

$$A\|f\|^2_{L_2(\mathbb{R}^d)} \le \sum_{j\in\mathbb{Z}} |\langle f, g_j\rangle|^2 \le B\|f\|^2_{L_2(\mathbb{R}^d)}, \quad \forall f \in L_2(\mathbb{R}^d),$$

其中 $\langle\cdot,\cdot\rangle$ 是空间 $L_2(\mathbb{R}^d)$ 中的内积. 进一步的, 如果上式中 $A = B = 1$, 则称 X 是一个紧框架 (tight frame), 此时有

$$f = \sum_{j\in\mathbb{Z}} \langle f, g_j\rangle g_j \quad \forall f \in L_2(\mathbb{R}^d).$$

定义 2(对偶框架和框架对). 对任意给定的框架 X, 若存在另一个框架 $\widetilde{X} = \{\widetilde{g}_j : j \in \mathbb{Z}\}$ 使得

$$f = \sum_{j\in\mathbb{Z}} \langle f, g_j\rangle \widetilde{g}_j \quad \forall f \in L_2(\mathbb{R}^d),$$

称 $\widetilde{X}$ 是 X 的一个对偶框架 (dual frame), 称 $(X, \widetilde{X})$ 为一个框架对 (bi-frame).

给定 $\Psi = \{\psi_1, \dots, \psi_r\} \subset L_2(\mathbb{R}^d)$, 记拟仿射系统 $X(\Psi)$ 是由 Ψ 的伸缩和平移构成的集合 [63], 即

$$X(\Psi) = \{\psi_{\ell,n,\boldsymbol{k}} :\ 1 \le \ell \le r;\ n \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^d\}, \tag{2.1}$$

其中

$$\psi_{\ell,n,\boldsymbol{k}} = \begin{cases} 2^{\frac{nd}{2}}\psi_\ell(2^n \cdot - \boldsymbol{k}), & n \ge 0; \\ 2^{nd}\psi_\ell(2^n \cdot - 2^n\boldsymbol{k}), & n < 0. \end{cases} \tag{2.2}$$

若 $X(\Psi)$ 是一个 (紧) 框架, 称 ψ_ℓ, $\ell = 1, \dots, r$ 为 (紧) 小波框架函数, 称系统 $X(\Psi)$ 为 (紧) 小波框架系统. 在很多文献中, 抽样的小波 (框架) 变换的仿射系统比较常用. 但在本文中, 我们仅讨论拟仿射系统 (2.2), 因为它在图像重建中的效果更好, 且与变分模型和 PDE 之间的联系比仿射系统更加自然 [47, 48, 50].

函数 Ψ 及其对偶 $\widetilde{\Psi}$ 的构造通常是基于多分辨分析 (Multiresolution Analysis, MRA). 主 MRA 由某些可细分函数 (refinable function)ϕ 和可细分掩码 (refinement mask)$\boldsymbol{p}$ 生成, 对偶 MRA 则由对偶可细分函数 $\widetilde{\phi}$ 和可细分掩码 $\widetilde{\boldsymbol{p}}$ 生成:

$$\phi = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{p}[\boldsymbol{k}]\phi(2\cdot-\boldsymbol{k}) \quad \text{and} \quad \widetilde{\phi} = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \widetilde{\boldsymbol{p}}[\boldsymbol{k}]\widetilde{\phi}(2\cdot-\boldsymbol{k}).$$

基于 MRA 的框架函数对 $\Psi=\{\psi_1,\dots,\psi_r\}$ 和 $\widetilde{\Psi}=\{\widetilde{\psi}_1,\dots,\widetilde{\psi}_r\}$ 的构建方法为: 寻找掩码 $\boldsymbol{q}^{(\ell)}$ 和 $\tilde{\boldsymbol{q}}^{(\ell)}$ 使得对任意的 $\ell=1,2,\dots,r$ 有

$$\psi_\ell = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{q}^{(\ell)}[\boldsymbol{k}]\widetilde{\phi}(2\cdot-\boldsymbol{k}) \quad 和 \quad \widetilde{\psi}_\ell = 2^d \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \widetilde{\boldsymbol{q}}^{(\ell)}[\boldsymbol{k}]\phi(2\cdot-\boldsymbol{k}). \tag{2.3}$$

然而, 掩码 $\boldsymbol{p}$、$\widetilde{\boldsymbol{p}}$、$\boldsymbol{q}^{(\ell)}$ 和 $\tilde{\boldsymbol{q}}^{(\ell)}$ 需要满足一定条件才能使 $(X(\Psi),X(\widetilde{\Psi}))$ 构成一个小波框架对. 文献 [64] 提出的 MEP 提供了构建小波双框架的一般理论.

给定数列 $\{\boldsymbol{p}[\boldsymbol{k}]\}_{\boldsymbol{k}}$, 记 $\widehat{\boldsymbol{p}}(\boldsymbol{\omega})$ 为对应的傅里叶级数, 即 $\widehat{\boldsymbol{p}}(\boldsymbol{\omega})=\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\boldsymbol{p}[\boldsymbol{k}]e^{-i\boldsymbol{k}\cdot\boldsymbol{\omega}}$. 给定两组有限支撑的掩码

$$\{\boldsymbol{p},\boldsymbol{q}^{(1)},\dots,\boldsymbol{q}^{(r)}\},\ \{\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}}^{(1)},\dots,\tilde{\boldsymbol{q}}^{(r)}\},$$

MEP 表明, 只要对所有的 $\boldsymbol{\nu}\in\{0,\pi\}^d\setminus\{\boldsymbol{0}\}$ 和 $\boldsymbol{\xi}\in[-\pi,\pi]^d$ 有

$$\widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\xi})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\widetilde{\boldsymbol{q}}}^{(\ell)}(\boldsymbol{\xi})} = 1 \quad 和 \quad \widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\widetilde{\boldsymbol{p}}}(\boldsymbol{\xi}+\boldsymbol{\nu})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\widetilde{\boldsymbol{q}}}^{(\ell)}(\boldsymbol{\xi}+\boldsymbol{\nu})} = 0, \tag{2.4}$$

那么由 (2.3) 生成地 Ψ, $\widetilde{\Psi}$ 和拟仿射系统 $X(\Psi)$, $X(\widetilde{\Psi})$ 就是 $L_2(\mathbb{R}^d)$ 中的一个框架对. 特别的, 若 $\boldsymbol{p}=\widetilde{\boldsymbol{p}}$ 且 $\boldsymbol{q}^{(\ell)}=\widetilde{\boldsymbol{q}}^{(\ell)}$ 对任意 $\ell=1,\dots,r$, MEP(2.4) 就退化为下面的 UEP [63]: 若掩码 $\boldsymbol{p}$、$\boldsymbol{q}^\ell$ 满足

$$|\widehat{\boldsymbol{p}}(\boldsymbol{\xi})|^2 + \sum_{\ell=1}^r |\widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})|^2 = 1 \quad 和 \quad \widehat{\boldsymbol{p}}(\boldsymbol{\xi})\overline{\widehat{\boldsymbol{p}}(\boldsymbol{\xi}+\boldsymbol{\nu})} + \sum_{\ell=1}^r \widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi})\overline{\widehat{\boldsymbol{q}}^{(\ell)}(\boldsymbol{\xi}+\boldsymbol{\nu})} = 0, \tag{2.5}$$

则系统 $X(\Psi)$ 是 $L_2(\mathbb{R}^d)$ 中的一个紧框架. 在这里 $\boldsymbol{p}$ 和 $\widetilde{\boldsymbol{p}}$ 是低通滤波器, $\boldsymbol{q}^{(\ell)},\widetilde{\boldsymbol{q}}^{(\ell)}$ 是高通滤波器. 这些滤波器生成了 $\ell_2(\mathbb{Z}^d)$ 空间中的离散的框架对 (若 UEP 条件成立, 它就对应了函数空间上的紧框架).

下面是两个简单但很常用的紧小波框架的例子.

例 1. 令 $\boldsymbol{p}=\frac{1}{2}[1,1]$ 是分段常数 B 样条 $B_1(x)=1$, $x\in[0,1]$(其它点的值为 0) 的细分掩码. 定义 $\boldsymbol{q}^{(1)}=\frac{1}{2}[1,-1]$. 那么 $\boldsymbol{p}$ 和 $\boldsymbol{q}^{(1)}$ 满足 (2.5) 中的两个等式. 因此由 (2.1) 生成的系统 $X(\psi_1)$ 是 $L_2(\mathbb{R})$ 中的一个紧框架.

例 2. [63] 令 $\boldsymbol{p}=\frac{1}{4}[1,2,1]$ 是分段线性 B 样条 $B_2(x)=\max(1-|x|,0)$ 的细分掩码. 定义 $\boldsymbol{q}^{(1)}=\frac{\sqrt{2}}{4}[1,0,-1]$, $\boldsymbol{q}_2=\frac{1}{4}[-1,2,-1]$. 那么 $\boldsymbol{p}$, $\boldsymbol{q}^{(1)}$ 和 $\boldsymbol{q}^{(2)}$ 满足 (2.5) 中的两个等式. 因此系统 $X(\Psi)$(其中 $\Psi=\{\psi_1,\psi_2\}$ 的定义见式 (2.1)) 是 $L_2(\mathbb{R})$ 中的一个紧框架.

在离散情形下, 记图像 $\boldsymbol{f}$ 是一个 d- 维数组, 而 $\mathcal{I}_d=\mathbb{R}^{N_1\times N_2\times\cdots\times N_d}$ 表示所有 d- 维图像的集合. 记滤波器为 $\{\boldsymbol{q}^{(0)}=\boldsymbol{p},\boldsymbol{q}^{(1)},\cdots,\boldsymbol{q}^{(r)}\}$ 的 d- 维快速 $(L+1)$- 层小波框架分解变换为

$$\boldsymbol{W}\boldsymbol{u}=\{\boldsymbol{W}_{\ell,l}\boldsymbol{u}:(\ell,l)\in\mathbb{B}\},\quad \boldsymbol{u}\in\mathcal{I}_d, \tag{2.6}$$

其中

$$\mathbb{B}=\{(\ell,l):\ 1\le\ell\le r,0\le l\le L\}\cup\{(0,L)\}.$$

那么图像 $\boldsymbol{u}$ 的小波框架系数为 $\boldsymbol{W}_{\ell,l}\boldsymbol{u}=\boldsymbol{q}_{\ell,l}[-\cdot]\circledast\boldsymbol{u}$, 其中 $\circledast$ 是某种边界条件 (如循环边界条件) 下的卷积算子且

$$\boldsymbol{q}_{\ell,l}=\check{\boldsymbol{q}}_{\ell,l}\circledast\check{\boldsymbol{q}}_{l-1,0}\circledast\ldots\circledast\check{\boldsymbol{q}}_{0,0}\quad\text{with}\quad\check{\boldsymbol{q}}_{\ell,l}[\boldsymbol{k}]=\begin{cases}\boldsymbol{q}^{(\ell)}[2^{-l}\boldsymbol{k}], & \boldsymbol{k}\in 2^l\mathbb{Z}^d;\\ 0, & \boldsymbol{k}\notin 2^l\mathbb{Z}^d.\end{cases}\tag{2.7}$$

对于对偶滤波器 $\{\tilde{\boldsymbol{p}},\tilde{\boldsymbol{q}}^{(1)},\ldots,\tilde{\boldsymbol{q}}^{(r)}\}$, 可类似的定义 $\widetilde{\boldsymbol{W}}\boldsymbol{u}$ 和 $\widetilde{\boldsymbol{W}}_{\ell,l}\boldsymbol{u}$. 逆小波框架变换 (或者小波框架分解)$\widetilde{\boldsymbol{W}}^\top$ 是 $\widetilde{\boldsymbol{W}}$ 的伴随算子. 根据 MEP, 可得到下述完美的重建公式

$$\boldsymbol{u}=\widetilde{\boldsymbol{W}}^\top\boldsymbol{W}\boldsymbol{u},\quad\forall\boldsymbol{u}\in\mathcal{I}_d.$$

特别的, 当 $\boldsymbol{W}$ 是紧框架变换时, 由 UEP 可得

$$\boldsymbol{u}=\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{u},\quad\forall\boldsymbol{u}\in\mathcal{I}_d.\tag{2.8}$$

为了简便, 本文主要关注 $d=2$(即二维图图像) 的情形.

2.2. 小波框架图像重建模型与算法

小波框架的图像重建模型源于对图像填充 (image inpainting) 问题的算法设计. 图像填充问题的的数学模型可以叙述如下 (可参考 [71, 72]). 原始图像 $\boldsymbol{u}\in\mathbb{R}^{m\times n}$ 定义在 $\boldsymbol{\Omega}=\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}$ 上且非空集 $\boldsymbol{\Lambda}\subsetneq\boldsymbol{\Omega}$ 是观测区域. 那么观测到的不完整图像 $\boldsymbol{f}$ 为

$$\boldsymbol{f}[\boldsymbol{k}]=\begin{cases}\boldsymbol{u}[\boldsymbol{k}]+\boldsymbol{\eta}[\boldsymbol{k}], & \boldsymbol{k}\in\boldsymbol{\Lambda},\\ \text{任意}, & \boldsymbol{k}\in\boldsymbol{\Omega}\setminus\boldsymbol{\Lambda}.\end{cases}\tag{2.9}$$

我们的目标是从带噪声数据 $\boldsymbol{f}$ 重建 $\boldsymbol{u}$. 对于这类问题, 线性系统 $\boldsymbol{A}$ 是一个限制算子, 将 $\boldsymbol{u}\in\mathbb{R}^{m\times n}$ 限制在索引集 $\boldsymbol{\Lambda}\subsetneq\boldsymbol{\Omega}$ 上. 记该问题的线性算子为 $\boldsymbol{A}_{\boldsymbol{\Lambda}}$.

在早期小波框架方法中 [14], 求解图像填充问题的思想是, 对于给定的数据, 先通过插值得到一个近似重建图像. 该估计图像的边缘可能是模糊的且噪声仍然存在. 文献 [14] 中的提供了一个简单策略, 将小的小波框架系数设为 0, 从而锐化图像并且消除噪声. 使用上述修正后的小波系数重建图像时, 不再进行插值, 而是用观测区域数据进行简单的修正. 重复这个过程直到收敛. 具体的算法如下:

基于小波框架的图像填充算法: 初始化 $\boldsymbol{u}^0=\boldsymbol{0}$. 对 $k=1,2,\ldots$ 做下述迭代直至收敛:

$$\begin{cases}\boldsymbol{u}^k=(\boldsymbol{I}-\mu\boldsymbol{A}_{\boldsymbol{\Lambda}}^\top\boldsymbol{A}_{\boldsymbol{\Lambda}})\boldsymbol{u}^{k-1}+\mu\boldsymbol{A}_{\boldsymbol{\Lambda}}^\top\boldsymbol{f},\\ \boldsymbol{u}^{k+1}=\boldsymbol{W}^\top\mathcal{T}_{\boldsymbol{\lambda}}^s(\boldsymbol{W}\boldsymbol{u}^k).\end{cases}\tag{2.10}$$

其中 $\boldsymbol{W}$ 紧小波框架系统, 即 $\boldsymbol{W}$ 满足 $\boldsymbol{W}^\top\boldsymbol{W}=\boldsymbol{I}$, $\mathcal{T}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha})$ 是 (各向异性) 软阈值算子

$$\mathcal{T}_{\boldsymbol{\lambda}}(\boldsymbol{\alpha})=\left\{\mathcal{T}^s_{\lambda_{\ell,l,\boldsymbol{k}}}(\alpha_{\ell,l,\boldsymbol{k}})=\frac{\alpha_{\ell,l,\boldsymbol{k}}}{|\alpha_{\ell,l,\boldsymbol{k}}|}\max\{|\alpha_{\ell,l,\boldsymbol{k}}|-\lambda_{\ell,l,\boldsymbol{k}},0\}:\ \boldsymbol{k}\in\boldsymbol{\Omega}\right\}.\tag{2.11}$$

相应的数值实验结果可以参见 [14]. 填充算法 (2.10) 也可用于一般的图像重建问题 (1.1), 只需要将 $\boldsymbol{A}_{\boldsymbol{\Lambda}}$ 替换为相应的线性系统.

文献 [73] 分析了算法 (2.10) 的收敛性. 如果令 $\boldsymbol{\alpha}^k = \mathcal{T}_{\boldsymbol{\lambda}}(\boldsymbol{W}\boldsymbol{u}^k)$, 那么由上述算法得到的序列 $\boldsymbol{\alpha}^k$ 和使用邻近向前向后算法 [74–79] 求解下面的均衡模型 [73, 80–82],

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha} - \boldsymbol{f}\|_2^2 + \frac{\kappa}{2}\|(\boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^\top)\boldsymbol{\alpha}\|_2^2 + \|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1 \tag{2.12}$$

所得到的迭代序列是等价的. 均衡模型是一个较为一般性的模型, 合成模型 [83–87] 和分析模型 [15, 88, 89] 都是均衡模型的特殊情形. 当 $\kappa = 0$, 我们可以得到合成模型:

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{W}^\top\boldsymbol{\alpha} - \boldsymbol{f}\|_2^2 + \|\boldsymbol{\lambda}\cdot\boldsymbol{\alpha}\|_1. \tag{2.13}$$

当 $\kappa = \infty$ 时, 如果记 $\boldsymbol{u} = \boldsymbol{W}^\top\boldsymbol{\alpha}$, 可以得到分析模型

$$\min_{\boldsymbol{u}} \ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \|\boldsymbol{\lambda}\cdot\boldsymbol{W}\boldsymbol{u}\|_1. \tag{2.14}$$

上述三类模型是基于小波框架变换图像反问题模型中最常用的模型.

均衡模型和合成模型都可以由近似前向 - 后向拆分 (proximal forward-backward splitting, PFBS) 来求解 [90, 91], 也可以使用 Nesterov 加速 [92, 93]. 分析模型可以用交替方向乘子法 (Alternating Direction Method of Multipliers, ADMM) [15, 94–97] 和原对偶混合梯度法 (Primal Dual Hybrid Gradient, PDHG) [98–100] 来快速求解.

这里我们简短回顾一下 ADMM 和 PDHG 算法 (后面也会用到). 针对形如 (2.14) 的分析模型, 定义该优化问题对应的增广拉格朗日函数 [101, Chapter 17]

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{d}; \boldsymbol{b}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \lambda\|\boldsymbol{d}\|_1 + \langle \boldsymbol{W}\boldsymbol{u} - \boldsymbol{d}, \boldsymbol{b}\rangle + \frac{\mu}{2}\|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{d}\|_2^2,$$

其中 $\boldsymbol{b}$ 为拉格朗日乘子. 这样, ADMM 算法形式如下:

$$\begin{aligned}
\boldsymbol{u}^{k+1} &= \left(\boldsymbol{A}^\top\boldsymbol{A} + \mu\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\left[\boldsymbol{A}^\top\boldsymbol{f} + \mu\boldsymbol{W}^\top(\boldsymbol{d}^k - \boldsymbol{\nu}^k)\right],\\
\boldsymbol{d}^{k+1} &= \mathcal{T}_{\lambda/\mu}\left(\boldsymbol{W}\boldsymbol{u}^{k+1} + \boldsymbol{\nu}^k\right),\\
\boldsymbol{\nu}^{k+1} &= \boldsymbol{\nu}^k + (\boldsymbol{W}\boldsymbol{u}^{k+1} - \boldsymbol{d}^{k+1}).
\end{aligned} \tag{2.15}$$

PDHG 算法可以求解比分析模型更一般形式的优化问题:

$$\min_{\boldsymbol{u}} F(\boldsymbol{u}) + \Phi(\boldsymbol{W}\boldsymbol{u}), \tag{2.16}$$

其中 $F(\boldsymbol{u})$ 是数据拟合项 (如 $F(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{u} - \boldsymbol{f}\|_2^2$)、$\Phi(\boldsymbol{W}\boldsymbol{u})$ 是正则项 (如 $\Phi(\cdot) = \|\cdot\|_1$). 假设 $F: \mathbb{R}^n \to (-\infty, +\infty]$ 和 $\Phi: \mathbb{R}^m \to (-\infty, +\infty]$ 是两个闭的、恰当的凸函数, 这样问题 (2.16) 具有如下的等价形式

$$\min_{\boldsymbol{u}} \max_{\boldsymbol{w}} F(\boldsymbol{u}) + \langle \boldsymbol{W}\boldsymbol{u}, \boldsymbol{w}\rangle - \Phi^*(\boldsymbol{w}).$$

PDHG 算法具有如下形式

$$\begin{aligned}
\boldsymbol{w}^{k+1} &= (I + \partial\Phi^*)^{-1}\left(\boldsymbol{w}^k + \alpha_k\boldsymbol{W}\boldsymbol{u}^k\right),\\
\boldsymbol{u}^{k+1} &= (I + \partial F)^{-1}\left(\boldsymbol{u}^k - \beta_k\boldsymbol{W}^\top\boldsymbol{w}^{k+1}\right),
\end{aligned} \tag{2.17}$$

这里 α_k 和 β_k 是预设超参数. 文献 [100] 在 PDHG 算法 (2.17) 的基础上引入了一个额外的修正项:

$$\bar{\boldsymbol{u}}^{k+1}=\boldsymbol{u}^{k+1}+\theta(\boldsymbol{u}^{k+1}-\boldsymbol{u}^{k}), \tag{2.18}$$

且在 $\boldsymbol{w}^{k+1}$ 这一步更新中将 $\boldsymbol{u}^k$ 替换为 $\bar{\boldsymbol{u}}^k$.

3. 偏微分方程 (PDE) 方法

图像科学领域发展了许多偏微分方程来描述图像的演化和刻画图像函数的性质, 本节讨论的 PDE 方法泛指变分模型和 PDE 模型. 在图像反问题中, PDE 模型主要有两个来源: 一是求解图像反问题中诱导出的变分模型, 转换为求解 Euler-Lagrange 方程或者梯度流所对应的方程; 另一个来源是图像重建中直接设计出的 PDE 模型, 这类 PDE 方程不保证有对应的能量泛函. 常用的 PDE 方法包括全变差 (Total Variation, TV) 模型、广义 (高阶) 全变差 (Total Generalized Variation) 模型、Mumford-Shah 模型、Shock-Filter 和 Perona-Malik(PM) 模型等. 更多的内容可以参考 [102, 103]. 本节我们主要介绍 TV 模型和 PM 模型.

3.1. 全变差 (Total Variation, TV) 模型

1963 年由 Tikhonov [8] 提出求解不适定问题的正则化方法后, 正则化方法成为图像反问题中 PDE 方法的一个重要工具. 在大量实际应用场景中, 图像的边缘反映了物体轮廓, 从人类视觉上讲, 图像边缘是重要的图像特征, 也是我们对图像进行进一步分析的基础. 从数学上看, 图像边缘反映的是图像的不连续性, 因此经典的 Sobolev 空间不能很好的刻画图像. 为了应对这一挑战, Rudin, Osher, Fatemi 提出了著名的全变差 (TV) 降噪模型 [38]. 该模型用有界变差 (Bounded Variation, BV) 函数空间来刻画图像. BV 空间是一个足够大的函数空间, 包含了带有间断的函数, 但同时也把噪声排除在空间之外. 因此, TV 降噪模型具有在降噪的同时保持边缘的良好性质. 接下来, 我们针对一般图像反问题简单介绍一下 TV 图像重建模型.

定义 3. 令 $\Omega\subset\mathbb{R}^2$ 为一个开集, 且 $u\in L_1(\Omega)$. u 的全变差定义为

$$\mathrm{TV}(u):=\sup\left\{\int_\Omega u\,\mathrm{div}\,v\,\mathrm{d}x\colon v\in C_c^1(\Omega,\mathbb{R}^2),\ \|v\|_{L_\infty(\Omega)}\le 1\right\}, \tag{3.1}$$

其中 $C_c^1(\Omega,\mathbb{R}^2)$ 是 Ω 上所有具有紧支集的连续可微向量函数构成的集合, $\|\cdot\|_{L_\infty(\Omega)}$ 是本征上确界范数.

函数 u 的全变差的另一个比较方便的记号为: $\mathrm{TV}(u)=\int_\Omega|Du(x)|\,\mathrm{d}x$. 其中 Du 是 u 的广义导数. 直观上看, 全变差就是函数值的变化的总和. 如果 $u(x)$ 可微, 则 $\mathrm{TV}(u)=\int_\Omega|\nabla u(x)|\,\mathrm{d}x$. 我们将有界全变差 (BV) 函数空间记为

$$\mathrm{BV}(\Omega)=\{u\in L_1(\Omega)\colon\ \mathrm{TV}(u)<+\infty\}. \tag{3.2}$$

给定观测图像 f, 考虑下列 (函数型) 线性图像反问题

$$f=Au+\eta,$$

其中 A 是对应于某个图像重建问题的线性算子, η 是个随机过程 (噪声). TV 图像重建模型通过最小化下面的能量泛函来获得重建的图像:

$$E_{\mathrm{TV}}(u)=\mathrm{TV}(u)+\frac{\lambda}{2}\int_\Omega(Au(x)-f(x))^2\mathrm{d}x, \tag{3.3}$$

这里 $\lambda > 0$ 是模型预设超参数. 求解 TV 模型 (3.3), 经典做法包括求解其对应的 Euler-Lagrange 方程、求解梯度流 (TV 流方程) 或用更加精致的优化算法去求解 (如上文描述的 ADMM 或 PDHG).

3.2. Perona-Malik(PM) 方程

PDE 模型在处理图像重建问题中的一个优势是其较为明确的几何直观. Perona-Malik(PM) 方程是图像重建问题中比较经典的 PDE 模型 [42]. 除了 PM 方程以外, 图像反问题中经典的 PDE 模型还包括 Weickert's 各向异性扩散方程 [104–106]、冲激滤波器 [43] 以及基于 Navier-Stokes 方程的图像处理模型 [44] 等等. 此外, 反向尺度空间理论 (Inverse Scale Space, ISS) 告诉我们用 PDE 去建模图像反问题是自然的选择, 因为如果把图像处理运算用一族非线性算子来刻画 $\{\boldsymbol{T}_t\}_{t\geq 0}$, 则如果算子族满足一些图像处理运算必须满足的公理, 该算子族一定对应于某个非线性 PDE [107]. 这里, 我们简短回顾 PM 方程及其建模思想.

给定观测图像 $u_0(x)$(比如含噪声的图像), PM 方程形式如下

$$\begin{cases} u_t = \operatorname{div}\big(g(|\nabla u|^2)\nabla u\big), \\ u(0,x) = u_0(x), \end{cases}$$

这里函数 g 满足下列条件

$$\begin{cases} g: [0,\infty) \mapsto (0,\infty) \text{ 单调下降}; \\ g(0) = 1;\ g(x) \to 0 \text{ as } x \to \infty; \\ g(x) + 2xg'(x) > 0 \text{ for } x \leq K; \quad g(x) + 2xg'(x) < 0 \text{ for } x > K. \end{cases} \tag{3.4}$$

在 PM 模型中, 扩散系数 $g(\cdot)$ 的作用是在不同空间位置和方向上定义不同的扩散力度. 所以 PM 方程是一个各向异性的扩散方程. 由 g 的性质 (3.4) 可以看出, 在图像 u 相对于光滑的区域 (即 $|\nabla u|$ 取值较小的区域), $g(|\nabla u|^2)$ 的值较大. 因此, PM 方程在图像光滑区域的平滑效果相对显著一些. 然而在图像边缘附近 (即 $|\nabla u|$ 取值较大的区域), $g(|\nabla u|^2)$ 的值较小. 因此, PM 方程在图像边缘附近做的平滑力度较小.

此外, 如果我们将 PM 方程按照 u 水平集的法向和切向展开, 则 PM 方程可以改写为:

$$u_t = g(|\nabla u|^2)u_{TT} + \widetilde{g}(|\nabla u|^2)u_{NN},$$

这里

$$\widetilde{g}(x) = g(x) + 2xg'(x), \quad N = \frac{\nabla u}{|\nabla u|} \quad \text{and} \quad T = N^{\perp},\ |T| = 1.$$

其中, u_{TT} 和 u_{NN} 分别是 u 在其水平集的切向 T 和法向 N 的二阶偏导. 从 (3.4) 的最后一个条件可以看出, PM 方程在与图像边缘正交的方向做的是反向扩散, 从而达到边缘锐化的效果. PM 方程常用的扩散系数 g 包括

$$g(s) = e^{-\frac{s}{2\sigma^2}}, \tag{3.5}$$

或

$$g(s) = \frac{1}{1 + s^p/\lambda^2}, \qquad p > \frac{1}{2},\ \lambda > 0. \tag{3.6}$$

如果观测图像噪声过大, 扩散系数 $g(|\nabla u|^2)$ 也会有较大噪声, 由于 PM 方程是个病态的 PDE(含有反向扩散), 噪声会使方程的解有较大的震荡, 降噪效果变差. 为了解决这个问题, 文献 [108] 提出了 PM 方程更加稳健的版本, 并且证明了该方程的适定性. 修正后的 PM 方程具有如下形式:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \operatorname{div}\left(g(|\nabla G_\sigma * u|^2)\nabla u\right), \\ u(0,x) = u_0(x), \end{cases} \tag{3.7}$$

其中 G_σ 是一个均值为 0, 方差为 σ^2 的标准高斯函数. 更多细节可以参考 [108].

4. 小波方法和 PDE 方法的联系与融合

小波框架 (其中正交、双正交小波是小波框架的特例) 和 PDE 方法在图像科学中有非常广泛的应用. 正如上文所述, 小波框架方法是基于线性变换的模型, 关键思想是分片光滑函数 (如图像) 在小波框架变换域是稀疏的. 小波框架方法包括基于 ℓ_p $(0 \le p \le 1)$ 等稀疏化范数的优化模型和基于稀疏化非线性算子 (如软阈值、硬阈值函数) 的迭代算法. 其中优化模型的代表包括分析模型 (analysis based model) [15, 88, 89]、合成模型 (synthesis based model) [83, 84, 86] 和均衡模型 (balanced model) [73, 80]; 迭代算法的代表是软阈值算法 [14, 109, 110]. PDE 方法包括变分模型和非线性演化 PDE 模型. 其中变分模型的代表是著名的全变差 (TV) 模型 [38]、Mumford-Shah 模型 [41]、卷积下确界 (Inf-Convolution) 模型 [40] 和广义全变差 (TGV) 模型 [39]; PDE 模型的代表是 Perona-Malik 方程 [42]、图像填充扩散方程 [71] 和 Osher-Rudin's 冲激滤波器 (shock filters) [43]. 小波框架方法和 PDE 方法是从两个不同的角度来刻画图像. 在近期的一系列工作中 [47–50], 小波框架方法和 PDE 方法之间深层的联系被建立起来. 本文主要对小波框架分析模型与 TV 模型的联系 [47]、小波框架迭代阈值算法与非线性 PDE 的联系 [50] 进行简短介绍, 具体细节和其它扩展研究请参考前文提及的文献.

4.1. 小波框架高通滤波与微分算子

小波框架变换是由一系列卷积组成, 卷积核是小波框架函数所对应的滤波器. 我们系列理论的一个重要的观察就是小波框架高通滤波器的消失矩和微分算子的阶具有对应关系, 同消失矩的不同滤波器对微分算子的逼近阶也会有所不同. 给定一个有限冲激响应 (finite impulse response, FIR) 的高通滤波器 $\boldsymbol{q}$, 令 $\widehat{\boldsymbol{q}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}]e^{-i\boldsymbol{k}\boldsymbol{\omega}}$. 我们再引入多维指标记号 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{Z}_+^2$ 和 $\boldsymbol{\omega} \in \mathbb{R}^2$, 且记

$$\boldsymbol{\alpha}! = \alpha_1!\alpha_2!,\ |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2,\ \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\alpha}}} = \frac{\partial^{\alpha_1+\alpha_2}}{\partial \omega_2^{\alpha_2}\partial\omega_1^{\alpha_1}}.$$

我们称 $\boldsymbol{q}$ (或 $\widehat{\boldsymbol{q}}(\boldsymbol{\omega})$) 的消失矩为 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$, 当

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\beta}} \boldsymbol{q}[\boldsymbol{k}] = i^{|\boldsymbol{\beta}|} \frac{\partial^{\boldsymbol{\beta}}}{\partial\boldsymbol{\omega}^{\boldsymbol{\beta}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega})\Big|_{\boldsymbol{\omega}=\underline{0}} = 0.$$

对所有满足下述条件的 $\boldsymbol{\beta} \in \mathbb{Z}_+^2$ 都成立: $|\boldsymbol{\beta}| < |\boldsymbol{\alpha}|$ 且 $|\boldsymbol{\beta}| = |\boldsymbol{\alpha}|$ 但 $\boldsymbol{\beta} \ne \boldsymbol{\alpha}$. 当 $\sum_{\boldsymbol{k}} \boldsymbol{q}[\boldsymbol{k}] \ne 0$ 时, 我们称 $\boldsymbol{q}$ 有 $(0,0)$ 阶消失矩. 我们称 $\boldsymbol{q}$ 有 $K \in \mathbb{Z}_+$ 阶全消失矩, 如果

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\beta}} \boldsymbol{q}[\boldsymbol{k}] = i^{|\boldsymbol{\beta}|} \frac{\partial^{\boldsymbol{\beta}}}{\partial\boldsymbol{\omega}^{\boldsymbol{\beta}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega})\Big|_{\boldsymbol{\omega}=\underline{0}} = 0, \quad \text{对任意}\ \ \boldsymbol{\beta} \in \mathbb{Z}_+^2,\ \text{且}\ |\boldsymbol{\beta}| < K. \tag{4.1}$$

如果 (4.1) 对所有 $|\boldsymbol{\beta}| < K$ 成立, 且除了某个 $\boldsymbol{\beta}_0 \in \mathbb{Z}_+^2$ 使得 $|\boldsymbol{\beta}_0| = J < K$, 我们称 $\boldsymbol{q}$ 有 $K\backslash\{J+1\}$ 阶消失矩. FIR 滤波器与微分算子的关系由下面的命题刻画, 该命题是保证小波框架变换和微分算子局部一致性的关键. 该命题也激发了我们近期将数值 PDE 与深层神经网络融合的工作 PDE-Net [111, 112].

命题 1. [50] 令 $\boldsymbol{q}$ 为一个高通 FIR 滤波器且有 $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$ 阶消失矩. 给定任意一个 $\mathbb{R}^2$ 上的光滑函数 $F(\boldsymbol{x})$, 我们有

$$\frac{1}{\varepsilon^{|\boldsymbol{\alpha}|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x}+\varepsilon\boldsymbol{k}) = C_{\boldsymbol{\alpha}} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} F(\boldsymbol{x}) + O(\varepsilon), \text{ 当 } \varepsilon \to 0, \tag{4.2}$$

其中常数 $C_{\boldsymbol{\alpha}}$ 为

$$C_{\boldsymbol{\alpha}} = \frac{1}{\boldsymbol{\alpha}!} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{k}^{\boldsymbol{\alpha}} \boldsymbol{q}[\boldsymbol{k}] = \frac{i^{|\boldsymbol{\alpha}|}}{\boldsymbol{\alpha}!} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{\omega}^{\boldsymbol{\alpha}}} \widehat{\boldsymbol{q}}(\boldsymbol{\omega}) \Big|_{\boldsymbol{\omega}=\boldsymbol{0}}. \tag{4.3}$$

如果 $\boldsymbol{q}$ 有 $\boldsymbol{\alpha} \in \mathbb{Z}_+^2$ 阶消失矩, 且有 $K\backslash\{|\boldsymbol{\alpha}|+1\}$ 阶全消失矩, 则

$$\frac{1}{\varepsilon^{|\boldsymbol{\alpha}|}} \sum_{\boldsymbol{k}\in\mathbb{Z}^2} \boldsymbol{q}[\boldsymbol{k}] F(\boldsymbol{x}+\varepsilon\boldsymbol{k}) = C_{\boldsymbol{\alpha}} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}}} F(\boldsymbol{x}) + O(\varepsilon^{K-|\boldsymbol{\alpha}|}), \text{ 当 } \varepsilon \to 0. \tag{4.4}$$

4.2. 小波框架分析模型与 TV 模型的联系

小波框架模型 (如基于 ℓ_1 范数的优化模型) 和变分模型 (如 TV 模型、Inf-Convolution 模型、TGV 模型、Mumford-Shah 模型) 在图像乃至数据科学中的诸多问题中有着广泛的应用. 这两类模型虽然在形式和数值表现方面有很多类似的地方, 但两者之间严格而系统的联系由 Cai, Dong, Osher, Shen 在 2012 年给出 [47]. 这篇文章给出了基于 ℓ_1 范数和小波框架变换的分析模型与基于微分算子的分析模型之间的联系, 在能量泛函极小化的构架下架起小波框架变换与微分算子之间的桥梁.

文 [47] 中研究的小波框架模型如下:

$$\inf_{u\in W_1^s(\Omega)} E_n(u) := \nu \|\boldsymbol{\lambda}_n \cdot \boldsymbol{W}_n \boldsymbol{T}_n u\|_1 + \frac{1}{2} \|\boldsymbol{A}_n \boldsymbol{T}_n u - \boldsymbol{T}_n f\|_2^2. \tag{4.5}$$

变分模型如下:

$$\inf_{u\in W_1^s(\Omega)} E(u) := \nu \|\boldsymbol{D}u\|_1 + \frac{1}{2} \|Au - f\|_{L_2(\Omega)}^2. \tag{4.6}$$

这里 $\boldsymbol{W}_n$ 为尺度 n 下的小波框架变换, $\boldsymbol{T}_n$ 为对应 $\boldsymbol{W}_n$ 的采样算子, $\boldsymbol{A}_n$ 为算子 A 的离散化, $\boldsymbol{D}$ 为某个最高阶为 s 的微分算子 (比如 TV 模型: $\boldsymbol{D} = \nabla$, $s = 1$). 这里 $W_p^r(\Omega)$ 是一个 Sobolev 空间, 即由 r- 阶弱导数属于 $L_p(\Omega)$ 的函数组成的集合, 并且配备了范数: $\|f\|_{W_p^r(\Omega)} := \sum_{|\boldsymbol{k}|\le r} \|D_{\boldsymbol{k}} f\|_p$, 其中 $D_{\boldsymbol{i}}(f(x,y)) = \frac{\partial^{|\boldsymbol{i}|} f}{\partial x^{i_1} \partial y^{i_2}}$.

小波框架模型 (4.5) 和变分模型 (4.6) 从形式上看是非常类似的. 命题 1 告诉我们当变量 u 有足够的光滑性的话, 小波框架变换的确是微分算子的一个离散逼近. 但是, 为了降低我们对变量 u 的光滑性假设 (因为图像并不光滑), 我们需要在较弱的设定下探讨两者之间的联系. 令 $\psi_{n,\boldsymbol{k}}$ 为处在尺度 n 和位置 $\boldsymbol{k}$ 的小波框架函数, 其对应的可细分函数为 $\phi_{n,\boldsymbol{k}}$. 我们把图像 $\boldsymbol{u}$ 视作函数 u 的采样: $\boldsymbol{u}[\boldsymbol{k}] = \alpha_n \langle u, \phi_{n,\boldsymbol{k}}\rangle$, 其中 α_n 是一个依赖于 n 的常数. 当小波框架变换作用在 $\boldsymbol{u}$ 上时, 对应于 $\psi_{n,\boldsymbol{k}}$ 的小波框架系数可以写成 $\beta_n \langle u, \psi_{n-1,\boldsymbol{k}}\rangle$. 我们可以

证明, 若 ψ 是一个具有紧支集的小波框架函数, 总存在另一个具有紧支集且积分不为零的函数 φ 使得 $\langle u, \psi_{n-1,\boldsymbol{k}}\rangle = \gamma_n \langle Du, \varphi_{n-1,\boldsymbol{k}}\rangle$, 其中 D 是由 ψ 唯一确定的一个微分算子 (严格证明见 [49,113]), 即小波框架系数是微分算子在某个尺度和位置的采样.

为了严格刻画能量泛函 E_n 与 E 的关系, 文献 [47] 采用了 Γ- 收敛 (Γ-convergence) 这个重要的数学分析工具. 我们先来回顾一下 Γ- 收敛的定义, 关于 Γ- 收敛的详细介绍可以参考 [114]. 给定某拓扑空间 X 上的泛函 E 和泛函序列 $\{E_n\}$, 如果下面两个条件满足, 我们称 E_n Γ- 收敛到 E:

(1) X 中的任何一个收敛序列 $u_n \to u$ 都满足 $E(u) \le \liminf_{n\to\infty} E_n(u_n)$;

(2) 对任意的 $u \in X$, 均存在一个收敛序列 $u_n \to u$, 使得 $E(u) \ge \limsup_{n\to\infty} E_n(u_n)$.

文献 [47] 的主要结论如下:

定理 1. 给定任何一个形如 (4.6) 的变分模型, 存在系数 $\boldsymbol{\lambda}_n$, 使得 (4.5) 中的泛函 E_n 在 W_1^s 中 Γ- 收敛到 (4.6) 中的泛函 E. 令 $u_n^\star$ 为 $\inf_u E_n(u)$ 的 ε- 近似解, 即 $E_n(u_n^\star) \le \inf_u E_n(u) + \epsilon$ $(\epsilon > 0)$, 则

$$\limsup_{n\to\infty} E_n(u_n^\star) \le \inf_u E(u) + \epsilon,$$

且 $\{u_n^\star\}$ 的任何一个聚点均为 $\inf_u E(u)$ 的 ε- 近似解.

定理 1 在能量泛函优化的框架下建立起了小波框架变换和微分算子之间的联系. 同时, 该联系赋予了小波框架的几何直观, 然而在此之前, 小波框架变换的一个较大的诟病就是缺乏几何直观 [115]. 此外, 该联系也将微分算子的几何直观和小波框架的稀疏逼近统一起来, 为微分算子的数值逼近提供了一个新的视角, 也为变分模型提供了一套新的离散化方法. 该离散化方法同时具有多尺度和稀疏逼近的良好性质, 并在图像分割、图像增强、医学成像等重要图像问题中明显优越于传统的有限差分离散化方法 (部分数值比对结果可参见 [49, 113, 116, 117]). 通过架起小波框架变换与微分算子之间的桥梁, 该工作进一步拓宽了小波框架的应用范畴, 为解决更多图像科学中的重要问题打下了坚实的理论基础(如图像分割 [116, 118]、曲面重建 [119]、半监督学习问题 [120]).

4.3. 小波框架迭代阈值算法和 PDE 模型的联系

Weickert 在 2000 年初研究了 Haar 小波和 Perona-Malik 方程的关系 [45, 121, 122], Jiang 在一维研究了小波框架迭代阈值算法和非线性热方程之间的联系 [46]. 但是, 小波方法和 PDE 方法之间的一般性的联系并没有被深入的研究. 现有工作仅在特殊情况下提出了小波框架迭代阈值算法与非线性热方程的对应关系, 然而在图像问题中除了非线性热方程, 还有许多被广泛使用的方程, 比如冲激滤波器 [43], 以及基于 Navier-Stokes 方程的图像处理微分方程模型 [44] 等等.

小波方法与 PDE 方法之间是否存在既一般又深刻的联系? 研究它们之间的联系能为我们带来什么样的新思想、能为图像问题的数学建模带来怎样的新思路? 这些问题在 Dong, Jiang 和 Shen 的工作 [50] 中得到了深入的研究和探讨. 文献 [50] 从一个新的角度, 把一般形式的小波框架迭代阈值算法和非线性演化偏微分方程 (其中 Perona-Malik 方程、冲激滤波器和基于 Navier-Stokes 方程的图像处理微分方程模型都是该方程的特例) 联系在一起, 在理论上严格证明了小波框架迭代阈值算法是偏微分方程的离散逼近, 也从数值上证明了用小波框

架迭代来逼近微分方程在很多图像问题中比传统的有限差分方法更加优越. 同时, 该工作发现了小波框架变换在对微分算子的采样形式上与有限元、有限差分及小波伽辽金 (wavelet-Galerkin) 方法都有本质的区别, 该区别也是小波框架迭代阈值算法在图像问题中成功的关键之一. 同时, 从该项研究拓展出了众多全新的小波框架算法和全新偏微分方程结合了两类方法的优点, 在数值上比传统的小波框架算法和偏微分方程能更加有效地解决图像科学中的很多问题. 本节我们将总结 [50] 中的一些重要结果, 详细内容请参看原文.

4.3.1. 小波框架迭代阈值算法和非线性演化偏微分方程的联系

令 $\boldsymbol{d} := \boldsymbol{W}\boldsymbol{u}$ 为 $\boldsymbol{u}$ 的小波框架变换, $\widetilde{\boldsymbol{W}}^\top \boldsymbol{d}$ 为小波框架逆变换, 其中 $\widetilde{\boldsymbol{W}}$ 和 $\boldsymbol{W}$ 对应的小波框架互为对偶, 即 $\widetilde{\boldsymbol{W}}^\top \boldsymbol{W} = \boldsymbol{I}$. 简便起见, 我们仅考虑单层小波框架变换. 给定小波框架系数 $\boldsymbol{d} = \{d_{\ell,\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{Z}^2, 0 \le \ell \le L\}$ 和阈值 $\boldsymbol{\alpha}(\boldsymbol{d}) = \{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d}) : \boldsymbol{n} \in \mathbb{Z}^2, 0 \le \ell \le L\}$, 收缩算子 (shrinkage operator) $\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d})$ 的定义如下

$$\boldsymbol{S}_{\boldsymbol{\alpha}}(\boldsymbol{d}) = \{S_{\alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}}) = d_{\ell,\boldsymbol{n}}(1 - \alpha_{\ell,\boldsymbol{n}}(\boldsymbol{d})) : \boldsymbol{n} \in \mathbb{Z}^2, 0 \le \ell \le L\}. \tag{4.7}$$

定义 (4.7) 中的收缩算子的两个特殊情形是各向同性和各项异性软阈值算子 [47, 50]. 给定收缩算子 $\boldsymbol{S}_{\boldsymbol{\alpha}}$, 我们将一般形式的小波框架迭代阈值算法记为

$$\boldsymbol{u}^k = \widetilde{\boldsymbol{W}}^\top \boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1}), \quad k = 1, 2, \cdots. \tag{4.8}$$

现在我们考虑具有如下形式的非线性演化偏微分方程

$$u_t = \sum_{\ell=1}^{L} \frac{\partial^{\boldsymbol{\alpha}_\ell}}{\partial x^{\boldsymbol{\alpha}_\ell}} \Phi_\ell(\boldsymbol{D}u, u), \quad \boldsymbol{D} = (\frac{\partial^{\boldsymbol{\beta}_1}}{\partial x^{\boldsymbol{\beta}_1}}, \dots, \frac{\partial^{\boldsymbol{\beta}_L}}{\partial x^{\boldsymbol{\beta}_L}}), \tag{4.9}$$

其中 (4.9) 是定义在 $\mathbb{R}^2$ 上的微分方程, 且 $|\boldsymbol{\alpha}_\ell|, |\boldsymbol{\beta}_\ell| \ge 0$, $1 \le \ell \le L$. 微分方程 (4.9) 涵盖了很多在图像问题中被提出的非线性热方程和非线性双曲方程.

文献 [50] 中的一个核心结论是: 给定任何一个形如 (4.9) 的 PDE, 可以构造小波框架变换 $\boldsymbol{W}$ 和 $\widetilde{\boldsymbol{W}}$ 以及算子 $\boldsymbol{S}_{\boldsymbol{\alpha}}$ 使得小波框架迭代 (4.8) 是该 PDE 的一个局部离散化; 当方程 (4.9) 是某种适定的各向异性热方程时, (4.8) 所对应的离散解收敛到 PDE 的解. 该结论的关键是命题 1 中对 $\boldsymbol{W}$($\widetilde{\boldsymbol{W}}$ 同理) 和微分算子关系的刻画. 我们举一个例子来帮助读者理解这一结论.

考虑以下的 PDE

$$u_t = \frac{\partial \Phi_1}{\partial x_1}\left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, u\right) + \frac{\partial \Phi_2}{\partial x_2}\left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, u\right).$$

令 $\boldsymbol{W} = \widetilde{\boldsymbol{W}}$ 为 Haar 小波框架, 由命题 1 不难看出上述 PDE 有如下的离散格式

$$\begin{aligned}\widetilde{\boldsymbol{u}}^k = \widetilde{\boldsymbol{u}}^{k-1} &- \tau\widetilde{\lambda}_1 \boldsymbol{W}_1^\top \Phi_1(\lambda_1 \boldsymbol{W}_1 \widetilde{\boldsymbol{u}}^{k-1}, \lambda_2 \boldsymbol{W}_2 \widetilde{\boldsymbol{u}}^{k-1}, \widetilde{\boldsymbol{u}}^{k-1}) \\ &- \tau\widetilde{\lambda}_2 \boldsymbol{W}_2^\top \Phi_2(\lambda_1 \boldsymbol{W}_1 \widetilde{\boldsymbol{u}}^{k-1}, \lambda_2 \boldsymbol{W}_2 \widetilde{\boldsymbol{u}}^{k-1}, \widetilde{\boldsymbol{u}}^{k-1}).\end{aligned}$$

同时, 迭代算法 (4.8) 具有如下的形式

$$\begin{aligned}\boldsymbol{u}^k = \boldsymbol{u}^{k-1} &- \boldsymbol{W}_1^\top \left[\boldsymbol{W}_1 \boldsymbol{u}^{k-1} - \mathcal{S}_1(\boldsymbol{W}_1 \boldsymbol{u}^{k-1}, \boldsymbol{W}_2 \boldsymbol{u}^{k-1}, \boldsymbol{u}^{k-1})\right] \\ &- \boldsymbol{W}_2^\top \left[\boldsymbol{W}_2 \boldsymbol{u}^{k-1} - \mathcal{S}_2(\boldsymbol{W}_1 \boldsymbol{u}^{k-1}, \boldsymbol{W}_2 \boldsymbol{u}^{k-1}, \boldsymbol{u}^{k-1})\right].\end{aligned}$$

通过比较上面的两个式子, 不难看出, 我们只需将算子 $\mathcal{S}_\ell$, $\ell=1,2$, 定义为

$$\mathcal{S}_\ell(\xi_1,\xi_2,\zeta):=\xi_\ell-\tau\widetilde{\lambda}_\ell\Phi_\ell(\xi_1,\xi_2,\zeta)=\xi_\ell\left(1-\tau\widetilde{\lambda}_\ell\Phi_\ell(\xi_1,\xi_2,\zeta)/\xi_\ell\right),\quad \xi_\ell,\zeta\in\mathbb{R}$$

(当 $\Phi_\ell(\xi_1,\xi_2,\zeta)/\xi_\ell$ 有良定义时), 则小波框架迭代阈值算法和上述 PDE 就有了严格的对应关系. 特别的, 当 $\Phi_\ell\left(\frac{\partial u}{\partial x_1},\frac{\partial u}{\partial x_2},u\right)=g_\ell(|\nabla u|^2,u)\frac{\partial u}{\partial x_\ell}$ 时 (非线性热方程), 我们可以令

$$\mathcal{S}_\ell(\xi_1,\xi_2,\zeta)=\xi_\ell\left(1-\tau\widetilde{\lambda}_\ell g_\ell(\xi_1^2+\xi_2^2,\zeta)\right).$$

这里, 我们选取参数 λ_ℓ 和 $\widetilde{\lambda}_\ell$ 使得 $\lambda_\ell\boldsymbol{W}_\ell\approx\frac{\partial}{\partial x_\ell}$ 且 $\widetilde{\lambda}_\ell\boldsymbol{W}_\ell^\top\approx\frac{\partial}{\partial x_\ell}$.

4.3.2. 小波框架迭代阈值算法诱导出的新 PDE

上一小节我们讨论了小波框架迭代阈值算法和非线性演化偏微分方程的联系. 基于该联系, 本小节我们列举一个由小波框架迭代阈值算法诱导出的新 PDE, 更多的例子参见 [50].

我们考虑基于 Nesterov 加速的小波框架迭代阈值算法 [34]

$$\boldsymbol{u}^k=(I-\mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{S}_{\boldsymbol{\alpha}^{k-1}}\left((1+\gamma^{k-1})\boldsymbol{W}\boldsymbol{u}^{k-1}-\gamma^{k-1}\boldsymbol{W}\boldsymbol{u}^{k-2}\right)+\mu\boldsymbol{A}^\top\boldsymbol{f},\quad k=1,2,\cdots. \tag{4.10}$$

该算法具有最优的优化复杂度 $O(1/k^2)$ [123, 124]. 在参数和小波框架的适当选取下, 迭代算法 (4.10) 对应的 PDE 如下:

$$u_{tt}+Cu_t=\sum_{\ell=1}^{L}(-1)^{1+|\boldsymbol{\beta}_\ell|}\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_\ell}}\left[g_\ell\left(u,\frac{\partial^{\boldsymbol{\beta}_1}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_1}},\cdots,\frac{\partial^{\boldsymbol{\beta}_L}u}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_L}}\right)\frac{\partial^{\boldsymbol{\beta}_\ell}}{\partial\boldsymbol{x}^{\boldsymbol{\beta}_\ell}}u\right]-\kappa A^\top(Au-f). \tag{4.11}$$

该方程可被视作一个抛物与双曲混合偏微分方程, 方程左边的 u_{tt} 项的作用是通过波传导来加速扩散速度使其能更快的达到稳定解, 这个发现与文献 [125] 中提出的加速思想有类似之处. 另外, 在文献 [126] 中也有类似的发现. 这一发现也激发了我们近期用数值微分方程去设计深层神经网络构架的工作 [127].

4.3.3. PDE 诱导的新小波框架迭代阈值算法

小波框架迭代阈值算法往往被认为是一类缺乏几何直观的方法, 但是基于前文对小波框架迭代阈值算法和微分方程的联系, 我们可以将两者有机的融合在一起, 设计出带有明确几何直观的小波框架迭代算法. 比如, 我们可以用 Perona-Malik 方程扩散系数来设计一个自适应的软阈值算子, 算法的具体形式如下

$$\boldsymbol{u}^k=(I-\mu\boldsymbol{A}^\top\boldsymbol{A})\boldsymbol{W}^\top\boldsymbol{T}_{\boldsymbol{\theta}^{k-1}}(\boldsymbol{W}\boldsymbol{u}^{k-1})+\mu\boldsymbol{A}^\top\boldsymbol{f},$$

这里 $\boldsymbol{u}^0=\boldsymbol{f}$ 是初始观测数据. 算子 $\boldsymbol{T}_{\boldsymbol{\theta}}$ 是各向同性软阈值算子:

$$\boldsymbol{T}_{\boldsymbol{\theta}}(\boldsymbol{d})=\left\{\mathcal{T}_{\theta_{\ell,\boldsymbol{n}}(\boldsymbol{d})}(d_{\ell,\boldsymbol{n}})=\frac{d_{\ell,\boldsymbol{n}}}{R_{\ell,\boldsymbol{n}}}\max\left\{R_{\ell,\boldsymbol{n}}-\theta_{\ell,\boldsymbol{n}}(\boldsymbol{d}),0\right\}:\boldsymbol{n}\in\mathbb{Z}^2,0\le\ell\le L\right\},$$

这里 $R_{\ell,\boldsymbol{n}}=\left(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|}|d_{\ell',\boldsymbol{n}}|^2\right)^{\frac{1}{2}}$, $\boldsymbol{\beta}_\ell$ 为小波框架变换中第 ℓ 个滤波器的消失矩. 阈值 $\boldsymbol{\theta}^k$ 的取法如下

$$\boldsymbol{\theta}^{k-1}=\left\{\theta_{l,\ell}(\boldsymbol{W}_{l,1}\boldsymbol{u}^{k-1},\boldsymbol{W}_{l,2}\boldsymbol{u}^{k-1},\ldots,\boldsymbol{W}_{l,L}\boldsymbol{u}^{k-1}):\ 0\le l\le \mathrm{Lev}-1,1\le\ell\le L\right\},$$

这里

$$\theta_{l,\ell}(\xi_1, \xi_2, \ldots, \xi_L) = C_\ell g\Big(\sum_{|\boldsymbol{\beta}_{\ell'}|=|\boldsymbol{\beta}_\ell|} \frac{\xi_{\ell'}^2}{C^{(\ell')}_{\boldsymbol{\beta}_{\ell'}} h^{\boldsymbol{\beta}_{\ell'}}} \Big).$$

函数 g 可以选取 Perona-Malik 方程中的热扩散系数 (Perona-Malik 方程的介绍见 3.2 节).

5. 图像反问题中的深度学习建模

从前文不难看出, 图像反问题的关键是对 (高质量) 图像的刻画, 小波框架方法和 PDE 方法都假设图像在某个精心设计的线性变换下是稀疏的, 这也是这些方法在图像重建任务中成功的关键. 近些年, 深度学习在图像反问题中也有着令人瞩目的进展. 与小波框架和 PDE 方法类似, 深度学习模型的核心思想是找到图像更好的表达. 然而不同之处是, 深度学习模型假设该表达是由深层神经网络构成的, 因此是更复杂的非线性稀疏表达, 该表达的模型参数必须从数据中去学习. 和传统方法相比, 深度学习模型更加灵活、适用任务更加多样、在特定场景下性能更优; 但与此同时, 深度学习模型可解释性、稳健性较差, 对数据量要求过大, 很多实际场景不能满足复杂深度学习模型对训练样本量的需求. 深度学习的劣势恰好是传统方法的优势所在, 因此, 在图像反问题领域, 人们开始探索传统模型与深度学习模型的有机融合, 以求设计出综合两类方法优点的新模型. 在这些融合建模方法中比较成功的一类方法被称作展开动力系统 (Unrolling Dynamics, UD) 方法, 我们在本节将着重介绍这类方法.

图像反问题的求解可以用机器学习模型来刻画,

$$\min_{\mathcal{F}} \mathbb{E}_{(\boldsymbol{u},\boldsymbol{f})\sim\mathcal{P}}\ \ell(\mathcal{F}(\boldsymbol{f}), \boldsymbol{u}), \tag{5.1}$$

即我们想找到一个 (参数化) 映射 $\mathcal{F}$ 使得模型重建的图像 $\mathcal{F}(\boldsymbol{f})$ 与真实图像 $\boldsymbol{u}$ 足够接近. 这里 $\ell(\cdot,\cdot)$ 是某个损失函数, $\mathcal{P}$ 是数据对 $(\boldsymbol{u},\boldsymbol{f})$ 所服从的概率分布.

传统小波框架或 PDE 模型所对应的 $\mathcal{F}$ 没有可以训练的参数, 因此我们无需求解优化问题 (5.1), 因为 $\mathcal{F}$ 是一个确定性的完全通过手动设计的映射 (如用 ADMM 或 PDHG 求解 TV 模型、小波框架迭代阈值算法、PM 方程). 但这也反映了传统算法的缺点, 即不能够充分利用现有的大规模数据集 $\{(\boldsymbol{u},\boldsymbol{f})\}$ 从而得到比手动设计更优的映射 $\mathcal{F}$. 在数据相对丰富的情况下, 人们尝试直接用神经网络来参数化映射 $\mathcal{F}$, 通过合理的网络构架设计并用随机梯度法求解问题 (5.1), 我们可以得到有较好泛化能力的图像重建模型. 在正问题 $\boldsymbol{A}$ 机理不清楚, 或者与真实物理正过程有较大误差时, 这种直接搭建从 $\boldsymbol{f}$ 到 $\boldsymbol{u}$ 的神经网络方法具有较大优势, 这也是大量计算机视觉和医疗影像重建方法采用的思路. 然而, 虽然 $\boldsymbol{A}$ 只是对真正物理正问题的一个逼近, 但也包含了问题的大量信息, 完全忽略 $\boldsymbol{A}$ 的内部机理并不是最佳方案. 因此, 如何将机理和数据有机的融合在一起, 是当前图像反问题的一个研究热点. 本小节我们介绍一类有效的将机理与数据融合的建模思想, 即展开动力系统 (UD) 方法. 这里我们只介绍网络构架的设计, 具体损失函数设计、优化算法选取等细节请参考原文.

UD 方法不同于其它方法之处在于 UD 将 $\mathcal{F}$ 视为为一个被展开的动力系统. UD 的基本思路是先从一个求解反问题的迭代算法出发, 将迭代算法中原先比较难以预设的参数 (比如正则化参数、稀疏变换等) 设置为可训练参数, 再将原先难以设计或者求解的算子 (如 proximal 算子、矩阵求逆变换等) 用神经网络来逼近, 这样就得到一个前馈神经网络, 即 $\mathcal{F}$, 最后通过设计合适的损失函数来进行端到端的训练, 即求解问题 (5.1). 可以被展开成前馈神经网络的

动力系统可以是某个优化算法, 如迭代阈值算法 [56,128,129]、ADMM [130,131]、PDHG [132]、加速梯度或 Heavy Ball 算法 [133], 也可以是某个离散后的 PDE [57,134]. 和直接用深层神经网络建模 $\mathcal{F}$ 相比, 通过 UD 对 $\mathcal{F}$ 建模具有如下三个优点: 1) 由 UD 建模的 $\mathcal{F}$ 有更好的可解释性, 因为起始模型往往是一个已经被很好的理解了的动力系统; 2) 由 UD 建模的 $\mathcal{F}$ 往往有更少的可训练参数, 因为由 $\boldsymbol{f}$ 到 $\boldsymbol{u}$ 的复杂映射已经有相当一部分由起始的动力系统模型很好的刻画, 因此我们仅需要对整个模型进行微调就可以得到很好的重建结果; 3) UD 给我们提供了一个机理与数据有机结合的建模思路, 因此 UD 并不仅仅局限于图像反问题, 类似想法也其它任务中取得了成功, 这包括图像识别 [127,135–140]、对抗训练算法加速 [141]、自然语言处理 [142]、PDE 反问题 [111,112] 等. 下面, 我们简短介绍一下 UD 的思想是如何运用在 ADMM 和 PDHG 上的.

ADMM-Net [130]

ADMM 迭代算法由 (2.15) 给出, 其中 μ, λ 和 β_k 为预设参数, 需要提前指定, 算子 $\boldsymbol{W}$ 是一个事先设计好的线性变换 (如小波框架变换或者某个微分算子的离散). 但是对于给定数据集, 这些预先指定的参数和变换很难保证是最优的, 我们需要对它们针对给定数据进行调整. [130] 的作者从 ADMM 算法出发, 将迭代展开成一个前馈神经网络 (称作 ADMM-Net), 这样就可以把前面提到的预设参数和变换 $\boldsymbol{W}$ 都作为网络的可训练参数, 参数值从数据中去学习. 此外, ADMM-Net 还用一个分片线性函数去替换 ADMM 算法中的阈值函数 $\mathcal{T}_\lambda(\cdot)$, 分片线性函数的参数也是从数据学习得到. 在 ADMM-Net 中, 变量 $\boldsymbol{d}^{k+1}$ 的更新变为

$$\boldsymbol{d}^{k+1} = \mathcal{T}_{\boldsymbol{\Theta}_1}\left(\boldsymbol{\mathcal{W}}_{\boldsymbol{\Theta}_2}(\boldsymbol{u}^{k+1}) + \boldsymbol{b}^k\right),$$

其中 $\mathbb{T}_{\boldsymbol{\Theta}_1}(\cdot)$ 是上述参数化的分片线性函数, $\boldsymbol{\Theta}_1$ 为参数, 且 $\boldsymbol{\mathcal{W}}_{\boldsymbol{\Theta}_2}$ 是由多个卷积组成的线性变换, 卷积核也是可训练参数并由 $\boldsymbol{\Theta}_2$ 表示. 之后, ADMM-Net 又被进一步的扩展 [143], 逐渐形成模型驱动的深度学习的建模思想 [144], 这类方法可以兼顾模型的可解释性, 同时也能得到明显优于传统压缩感知算法的图像重建效果.

原对偶网络 (Primal-Dual Net, PD-Net) [132]

文献 [132] 考虑的是 CT 图像重建问题, 文章从 PDHG 的迭代 (2.17) 和 (2.18) 出发, 将其展开成前馈网络, 并用卷积神经网络来逼近 PDHG 算法中的近邻算子 [145], 从而使得网络可以从数据中学习最佳的 Φ 和 F. 这个前馈网络被称作 PD-Net, 具体形式如下:

$$\begin{aligned} \boldsymbol{w}^{k+1} &= \boldsymbol{\mathcal{N}}_{\boldsymbol{w}}\left([\boldsymbol{w}^k, \boldsymbol{W}\boldsymbol{u}^k]; \boldsymbol{\Theta}_{\boldsymbol{w}}^k\right), \\ \boldsymbol{u}^{k+1} &= \boldsymbol{\mathcal{N}}_{\boldsymbol{u}}\left([\boldsymbol{u}^k, \boldsymbol{W}^\top\boldsymbol{w}^{k+1}, \boldsymbol{A}^\top\boldsymbol{f}]; \boldsymbol{\Theta}_{w}^k\right), \end{aligned} \tag{5.2}$$

这里的 $\boldsymbol{f}$ 和 $\boldsymbol{A}$ 是线性反问题 (1.1) 中相应的变量, $\boldsymbol{\mathcal{N}}_{\boldsymbol{w}}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{w}}^k)$ 和 $\boldsymbol{\mathcal{N}}_{\boldsymbol{u}}(\cdot; \boldsymbol{\Theta}_{\boldsymbol{u}}^k)$ 是神经网络其对应参数分别记为 $\boldsymbol{\Theta}_{\boldsymbol{w}}^k$ 和 $\boldsymbol{\Theta}_{\boldsymbol{u}}^k$. 记号 $[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m]$ 表示将向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ 叠在一起形成一个高维的张量. 线性算子 $\boldsymbol{W}$ 可以是预先设计好的也可以通过数据去学习. 在 CT 图像重建问题中, PD-Net 和行业标准算法 FBP [146] 以及一些传统模型相比有非常明显的性能提升, 即在固定剂量下, PD-Net 能够得到明显优质于 FBP 的 CT 图像 [132,147].

6. 结 论

图像反问题中的建模在过去 30 年间走过了模型驱动、模型 + 数据驱动和数据驱动这三个阶段, 期间的变化主要源于数据量的大幅增加和计算能力的大幅提升, 使得我们可以从数据中训练出比以往基于手动设计更优的模型. 然而, 过度依赖数据, 把对模型和机理的依赖降到最低也给这些模型 (特别是深度学习模型) 在可解释性、稳定性、泛化性等方面带来诸多问题. 解决这些问题的一个可行方案就是机理与数据融合的建模思路, 继承传统模型对问题刻画较为准确的环节或计算框架, 将设计任意、原理不清晰、维度高的环节用机器学习方法从数据中去学习. 因此, 基于调和分析和 PDE 的图像反问题方法与机器学习方法的有机融合是一个今后值得去探索的研究方向.

参 考 文 献

[1] Pavlovic G, Tekalp A. Maximum likelihood parametric blur identification based on acontinuous spatial domain model[J]. IEEE Transactions on image processing, 1992, 1(4): 496–504.

[2] Buzug T M. Computed tomography: from photon statistics to modern cone-beam CT. Springer Science & Business Media, 2008.

[3] Brown R W, Haacke E M, Cheng Y C N, Thompson M R, Venkatesan R. Magnetic resonance imaging: physical principles and sequence design. John Wiley & Sons, 2014.

[4] Choi J K, Park H S, Wang S, Wang Y, Seo J K. Inverse problem in quantitative susceptibility mapping[J]. SIAM J. Imaging Sci., 2014, 7(3): 1669–1689.

[5] Natterer F. Image reconstruction in quantitative susceptibility mapping[J]. SIAM J. Imaging Sci., 2016, 9(3): 1127–1131.

[6] de Rochefort L, Liu T, Kressler B, Liu J, Spincemaille P, Lebon V, Wu J, Wang Y. Quantitative susceptibility map reconstruction from mr phase data using bayesian regularization: validation and application to brain imaging[J]. Magn. Reson. Med., 2010, 63(1): 194–206.

[7] Wang Y, Liu T. Quantitative susceptibility mapping (qsm): decoding mri data for a tissue magnetic biomarker[J]. Magn. Reson. Med., 2015, 73(1): 82–101.

[8] Tikhonov A, Arsenin V, John F. Solutions of ill-posed problems, VH Winston Washington, DC, 1977.

[9] Bell J B, Tikhonov A N, Arsenin V Y. Solutions of ill-posed problems[J]. Mathematics of Computation, 1977, 32(144): 1320.

[10] Gröchenig K. Foundations of time-frequency analysis, Birkhauser, 2001.

[11] Daubechies I. Ten lectures on wavelets, Vol. CBMS-NSF Lecture Notes, SIAM, nr. 61, Society for Industrial and Applied Mathematics, 1992.

[12] Mallat S. A wavelet tour of signal processing: the sparse way, Academic press, 2008.

[13] Cai J, Osher S, Shen Z. Convergence of the linearized Bregman iteration for ℓ_1-norm minimization[J]. Mathematics of Computation, 2009, 78: 2127–2136.

[14] Chan R, Chan T, Shen L, Shen Z. Wavelet algorithms for high-resolution image reconstruction[J]. SIAM Journal on Scientific Computing, 2003, 24(4): 1408–1432.

[15] Cai J, Osher S, Shen Z. Split Bregman methods and frame based image restoration[J]. Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 2009, 8(2): 337–369.

[16] Cai J, Osher S, Shen Z. Linearized Bregman iterations for frame-based image deblurring[J]. SIAM J. Imaging Sci, 2009, 2(1): 226–252.

[17] Zhang Y, Dong B, Lu Z. ℓ_0 minimization of wavelet frame based image restoration[J]. Mathematics of Computation, 2013, 82: 995–1015.
[18] Dong B, Zhang Y. An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration[J]. Journal of Scientific Computing, 2013, 54(2-3): 350–368.
[19] Liang J, Li J, Shen Z, Zhang X. Wavelet frame based color image demosaicing[J]. Inverse Problems and Imaging, 2013, 7(3): 777–794.
[20] Hou L, Ji H, Shen Z. Recovering over-/underexposed regions in photographs[J]. SIAM J. Imaging Sciences, 2013, 6(4): 2213–2235.
[21] Cai J, Ji H, Liu C, Shen Z. Blind motion deblurring using multiple images[J]. Journal of Computational Physics, 2009, 228(14): 5057–5071.
[22] Cai J, Ji H, Liu C, Shen Z. Blind motion deblurring from a single image using sparse approximation, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, 104–111.
[23] Dong B, Ji H, Li J, Shen Z, Xu Y. Wavelet frame based blind image inpainting[J]. Applied and Computational Harmonic Analysis, 2012, 32(2): 268–279.
[24] Gong Z, Shen Z, Toh K C. Image restoration with mixed or unknown noises[J]. Multiscale Modeling & Simulation, 2014, 12(2): 458–487.
[25] Jia X, Dong B, Lou Y, Jiang S B. GPU-based iterative cone-beam CT reconstruction using tight frame regularization[J]. Physics In Medicine And Biology, 2011, 56(13): 3787–3807.
[26] Dong B, Li J, Shen Z, X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting[J]. Journal of Scientific Computing, 2013, 54(2-3): 333–349.
[27] Zhang H, Dong B, Liu B. A reweighted joint spatial-radon domain ct image reconstruction model for metal artifact reduction[J]. SIAM J. Imaging Sci. 2018, 11(1): 707–733.
[28] Gao H, Cai J F, Shen Z, Zhao H. Robust principal component analysis-based four-dimensional computed tomography[J]. Physics in medicine and biology, 2011, 56(11): 3181.
[29] Cai J, Jia X, Gao H, Jiang S, Shen Z, Zhao H. Cine cone beam ct reconstruction using low-rank matrix factorization: Algorithm and a proof-of-principle study[J]. IEEE transactions on medical imaging, 2014, 33(8): 1581–1591.
[30] Lustig M, Donoho D, Pauly J, Sparse MRI: The application of compressed sensing for rapid MR imaging[J]. Magnetic Resonance in Medicine, 2007, 58(6): 1182–1195.
[31] Lustig M, Donoho D L, Santos J M, Pauly J M. Compressed sensing mri[J]. IEEE signal processing magazine, 2008, 25(2): 72.
[32] Liu Y, Cai J F, Zhan Z, Guo D, Ye J, Chen Z, Qu X. Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging[J]. PloS one, 2015, 10(4): e0119584.
[33] Liu Y, Zhan Z, Cai J F, Guo D, Chen Z, Qu X. Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging[J]. IEEE transactions on medical imaging, 2016, 35(9): 2130–2140.
[34] Li M, Fan Z, Ji H, Shen Z. Wavelet frame based algorithm for 3d reconstruction in electron microscopy[J]. SIAM Journal on Scientific Computing, 2014, 36(1): B45–B69.
[35] Sapiro G. Geometric partial differential equations and image analysis. Cambridge University Press, 2001.
[36] Osher S, Fedkiw R. Level set methods and dynamic implicit surfaces. Vol. 153, Springer Science & Business Media, 2006.
[37] Chan T F, Shen J. Image processing and analysis: variational, PDE, wavelet, and stochastic

methods, SIAM, 2005.

[38] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: Nonlinear Phenomena, 1992, 60(1-4): 259–268.

[39] Bredies K, Kunisch K, Pock T. Total Generalized Variation[J]. SIAM Journal on Imaging Sciences, 2010, 3: 492.

[40] Chambolle A, Lions P. Image recovery via total variation minimization and related problems[J]. Numerische Mathematik, 1997, 76(2): 167–188.

[41] Mumford D, Shah J. Optimal approximations by piecewise smooth functions and associated variational problems[J]. Communications on pure and applied mathematics, 1989, 42(5): 577–685.

[42] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629–639.

[43] Osher S, Rudin L. Feature-oriented image enhancement using shock filters[J]. SIAM Journal on Numerical Analysis, 1990, 27(4): 919–940.
http://www.jstor.org/stable/2157689

[44] Bertalmio M, Bertozzi A L, Sapiro G. Navier-stokes, fluid dynamics, and image and video inpainting, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, Vol. 1, IEEE, 2001, I–I.

[45] Steidl G, Weickert J, Brox T, Mrázek P, Welk M. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides[J]. SIAM Journal on Numerical Analysis, 2004, 42(2): 686–713.

[46] Jiang Q. Correspondence between frame shrinkage and high-order nonlinear diffusion[J]. Applied Numerical Mathematics, 2012, 62(1): 51–66.

[47] Cai J, Dong B, Osher S, Shen Z. Image restorations: total variation, wavelet frames and beyond[J]. Journal of American Mathematical Society, 2012, 25(4): 1033–1089.

[48] Cai J, Dong B, Shen Z. Image restorations: a wavelet frame based model for piecewise smooth functions and beyond[J]. Applied and Computational Harmonic Analysis, 2016, 41(1): 94–138.

[49] Dong B, Shen Z, Xie P. Image restoration: a general wavelet frame based model and its asymptotic analysis[J]. SIAM Journal on Mathematical Analysis, 2017, 49(1): 421–445.

[50] Dong B, Jiang Q, Shen Z. Image restoration: wavelet frame shrinkage, nonlinear evolution PDEs, and beyond, Multiscale Modeling & Simulation: A SIAM Interdisciplinary Journal, 2017, 15(1): 606–660.

[51] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 9446–9454.

[52] Hand P, Voroninski V. Global guarantees for enforcing deep generative priors by empirical risk, arXiv preprint arXiv:1705.07576.

[53] Bora A, Jalal A, Price E, Dimakis A G. Compressed sensing using generative models, in: Proceedings of the 34th International Conference on Machine Learning-Volume 70, JMLR. org, 2017, 537–546.

[54] Shah V, Hegde C. Solving linear inverse problems using gan priors: An algorithm with provable guarantees, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, 4609–4613.

[55] Kabkab M, Samangouei P, Chellappa R. Task-aware compressed sensing with generative adversarial networks, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[56] Gregor K, LeCun Y. Learning fast approximations of sparse coding, 2010, 399–406.
[57] Chen Y, Yu W, Pock T. On learning optimized reaction diffusion processes for effective image restoration, in: CVPR, 2015, 5261–5269.
[58] Sun J, Li H, Xu Z, et al., Deep admm-net for compressive sensing mri, in: NeurIPS, 2016, 10–18.
[59] Engan K, Aase S O, Husoy J H. Method of optimal directions for frame design, in: Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, Vol. 5, IEEE, 1999, 2443–2446.
[60] Aharon M, Elad M, Bruckstein A, et al., K-svd: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Trans. Signal Process., 2006, 54(11): 4311.
[61] Cai J F, Ji H, Shen Z, Ye G B. Data-driven tight frame construction and image denoising[J]. Applied and Computational Harmonic Analysis, 2014, 37(1): 89–105.
[62] Tai C, Weinan E. Multiscale adaptive representation of signals: I. the basic framework[J]. J. Mach. Learn. Res., 2016, 17(1): 4875–4912.
[63] Ron A, Shen Z. Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator[J]. Journal of Functional Analysis, 1997, 148(2): 408–447.
[64] Ron A, Shen Z. Affine systems in $L_2(\mathbb{R}^d)$ II: dual systems[J]. Journal of Fourier Analysis and Applications, 1997, 3(5): 617–638.
[65] Daubechies I, Han B, Ron A, Shen Z. Framelets: MRA-based constructions of wavelet frames[J]. Applied and Computational Harmonic Analysis, 2003, 14(1): 1–46. doi:10.1016/S1063-5203(02)00511-0.
[66] Shen Z. Wavelet frames and image restorations[J]. in: Proceedings of the International Congress of Mathematicians, 2010, (4): 2834–2863.
[67] Dong B, Shen Z. MRA-Based Wavelet Frames and Applications, IAS Lecture Notes Series, Summer Program on "The Mathematics of Image Processing", Park City Mathematics Institute.
[68] Ron A, Shen Z. Generalized shift-invariant systems[J]. Constructive Approximation, 2005, 22(1): 1–45.
[69] Ron A, Shen Z. Weyl-Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$[J]. Duke Mathematical Journal, 1997, 89(2): 237–282.
[70] Ron A, Shen Z. Frames and Stable Bases for Shift-Invariant Subspaces of $L_2(\mathbb{R}^d)$[J]. Canadian Journal of Mathematics, 1995, 47(5): 1051–1094.
[71] Bertalmio M, Sapiro G, Caselles V, Ballester C. Image inpainting, in: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., 2000, 417–424.
[72] Chan T, Shen J. Variational image inpainting[J]. Commun. Pure Appl. Math, 2005, 58: 579–619.
[73] Cai J, Chan R, Shen Z. A framelet-based image inpainting algorithm[J]. Applied and Computational Harmonic Analysis, 2008, 24(2): 131–149.
[74] Tseng P. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities[J]. SIAM Journal on Control and Optimization, 1991, 29(1): 119–138.
[75] Chen G H, Rockafellar R. Convergence rates in forward–backward splitting[J]. SIAM Journal on Optimization, 1997, 7(2): 421–444.
[76] Combettes* P L. Solving monotone inclusions via compositions of nonexpansive averaged operators[J]. Optimization, 2004, 53(5-6): 475–504.
[77] Combettes P, Wajs V. Signal recovery by proximal forward-backward splitting[J]. Multiscale Modeling and Simulation, 2006, 4(4): 1168–1200.

[78] Hale E, Yin W, Zhang Y. A fixed-point continuation method for ℓ_1-regularization with application to compressed sensing, CAAM Technical Report TR, Rice University, Houston, TX (2007) 07–07CAAM Technical Report TR07-07, Rice University, Houston, TX.

[79] Bredies K. A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in banach space[J]. Inverse Problems, 2009, 25(1): 015005.

[80] Cai J, Chan R, Shen L, Shen Z. Convergence analysis of tight framelet approach for missing data recovery[J]. Advances in Computational Mathematics, 2009, 31(1): 87–113.

[81] Cai J, Chan R, Shen Z. Simultaneous cartoon and texture inpainting[J]. Inverse Problems and Imaging (IPI), 2010, 4(3), 379–395.

[82] Cai J, Shen Z. Framelet based deconvolution[J]. J. Comp. Math., 2010, 28(3): 289–308.

[83] Daubechies I, Teschke G, Vese L. Iteratively solving linear inverse problems under general convex constraints[J]. Inverse Problems and Imaging, 2007, 1(1): 29.

[84] Fadili M, Starck J. Sparse representations and Bayesian image inpainting. Proc. SPARS 5.

[85] Fadili M, Starck J, Murtagh F. Inpainting and zooming using sparse representations[J]. The Computer Journal, 2009, 52(1): 64.

[86] Figueiredo M, Nowak R, An EM algorithm for wavelet-based image restoration[J]. IEEE Transactions on Image Processing, 2003, 12(8): 906–916.

[87] Figueiredo M, Nowak R. A bound optimization approach to wavelet-based image deconvolution, in: Image Processing, 2005. ICIP 2005. IEEE International Conference on, Vol. 2, IEEE, 2005, pp. II–782.

[88] Elad M, Starck J, Querre P, Donoho D. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)[J]. Applied and Computational Harmonic Analysis, 2005, 19(3): 340–358.

[89] Starck J, Elad M, Donoho D. Image decomposition via the combination of sparse representations and a variational approach[J]. IEEE transactions on image processing, 2005, 14(10): 1570–1582.

[90] Bruck Jr R E. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in hilbert space[J]. J. Math. Anal. Appl., 1977, 61(1): 159–164.

[91] Passty G B. Ergodic convergence to a zero of the sum of monotone operators in hilbert space[J]. J. Math. Anal. Appl., 1979, 72: 383–290.

[92] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183–202.

[93] Shen Z, Toh K C, Yun S. An accelerated proximal gradient algorithm for frame-based image restoration via the balanced approach[J]. SIAM Journal on Imaging Sciences, 2011, 4(2): 573–596.

[94] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends® in Machine Learning, 2011, 3(1): 1–122.

[95] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Comput. Math. Appl., 1976, 2(1): 17–40.

[96] Glowinski R, Marroco A. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, 1975, 9(R2), 41–76.

[97] Goldstein T, Osher S. The split bregman method for l_1-regularized problems[J]. SIAM J. Imaging

Sci., 2009, 2(2): 323–343.

[98] Zhu M, Chan T. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report 34.

[99] Esser E, Zhang X, Chan T F. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science[J]. SIAM J. Imaging Sci., 2010, 3(4): 1015–1046.

[100] Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging[J]. J. Math. Imaging Vis., 2011, 40(1): 120–145.

[101] J. Nocedal, S. J. Wright, Numerical Optimization, 2nd Edition, Springer, 2006.

[102] Aubert G, Kornprobst P. Mathematical problems in image processing: partial differential equations and the calculus of variations. Springer, 2006.

[103] Chan T, Shen J. Image processing and analysis: variational, PDE, wavelet, and stochastic methods, Society for Industrial Mathematics, 2005.

[104] Weickert J. Coherence-enhancing diffusion filtering[J]. International Journal of Computer Vision, 1999, 31(2): 111–127.

[105] Weickert J. Anisotropic diffusion in image processing, Vol. 1, Teubner Stuttgart, 1998.

[106] Weickert J. Theoretical foundations of anisotropic diffusion in image processing[J]. Computing-Wien-Supplements, 1996, 11: 221–236.

[107] Alvarez L, Guichard F, Lions P, Morel J. Axioms and fundamental equations of image processing[J]. Archive for Rational Mechanics and Analysis, 1993, 123(3): 199–257.

[108] Catté F, Lions P L, Morel J M, Coll T. Image selective smoothing and edge detection by nonlinear diffusion[J]. SIAM Journal on Numerical analysis, 1992, 29(1): 182–193.

[109] Coifman R, Donoho D. Translation-invariant de-noising[J]. Wavelets and statistics, 1995, 103: 125.

[110] Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on pure and applied mathematics, 2004, 57(11): 1413–1457.

[111] Long Z, Lu Y, Ma X, Dong B. Pde-net: Learning PDEs from data, in: International Conference on Machine Learning, 2018, 3214–3222.

[112] Long Z, Lu Y, Dong B. Pde-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network, Journal of Computational Physics, 2019, 108925.

[113] Choi J K, Dong B, Zhang X. An edge driven wavelet frame model for image restoration, arXiv preprint arXiv:1701.07158.

[114] Braides A. Gamma-Convergence for Beginners. Vol. 22, Oxford University Press, 2002.

[115] Chan T, Shen J, Zhou H. Total variation wavelet inpainting[J]. Journal of Mathematical Imaging and Vision, 2006, 25(1): 107–125.

[116] Dong B, Chien A, Shen Z. Frame based segmentation for medical images[J]. Communications in Mathematical Sciences, 2010, 9(2): 551–559.

[117] Jia X, Dong B, Lou Y, Jiang S. GPU-based iterative cone-beam CT reconstruction using tight frame regularization[J]. Physics in Medicine and Biology, 2011, 56: 3787–3807.

[118] Tai C, Zhang X, Shen Z. Wavelet frame based multiphase image segmentation[J]. SIAM Journal on Imaging Sciences, 2013, 6(4): 2521–2546.

[119] Dong B, Shen Z. Frame based surface reconstruction from unorganized points[J]. Journal of Computational Physics, 2011, 230: 8247–8255.

[120] Dong B. Sparse representation on graphs by tight wavelet frames and applications[J]. Applied

and Computational Harmonic Analysis, 2017, 42(3): 452–479.

[121] Mrázek P, Weickert J. From two-dimensional nonlinear diffusion to coupled haar wavelet shrinkage[J]. Journal of Visual Communication and Image Representation, 2007, 18(2): 162–175.

[122] Mrázek R, Weickert J, Steidl G. Correspondences between wavelet shrinkage and nonlinear diffusion. in: Scale Space Methods in Computer Vision, Springer, 2003, 101–116.

[123] Nesterov Y. On an approach to the construction of optimal methods for minimizing smooth convex functions[J]. Ehkon. Mat. Metody, 1988, 24(3): 509–517.

[124] Nesterov Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$[J]. in: Soviet Mathematics Doklady, 1983, 27(2): 372–376.

[125] Mao Y, Osher S, Dong B. A nonlinear PDE-based method for sparse deconvolution. Multiscale Modeling and Simulation, 8(3).

[126] Su W, Boyd S, Candes E J. A differential equation for modeling nesterov' s accelerated gradient method: theory and insights[J]. Journal of Machine Learning Research, 2016, 17(153): 1–43.

[127] Lu Y, Zhong A, Li Q, Dong B. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. in: International Conference on Machine Learning, 2018, 3276–3285.

[128] Chen X, Liu J, Wang Z, Yin W. Theoretical linear convergence of unfolded ista and its practical weights and thresholds. in: NeurIPS, 2018, 9079–9089.

[129] Liu J, Chen X, Wang Z, Yin W. Alista: Analytic weights are as good as learned weights in lista. in: ICLR, 2019.

[130] yang Y, Sun J, Li H, Xu Z. Deep admm-net for compressive sensing mri, in: D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems 29, Curran Associates, Inc., 2016, 10–18.

[131] Zhang H, Dong B, Liu B. Jsr-net: A deep network for joint spatial-radon domain ct reconstruction from incomplete data, in: IEEE International Conference on Acoustics. Speech and Signal Processing (ICASSP)-2019, 2019, 3657–3661. doi:10.1109/ICASSP.2019.8682178.

[132] Adler J, Öktem O. Learned primal-dual reconstruction[J]. IEEE Trans. Med. Imaging, 2018, 37(6): 1322–1332.

[133] Li H, Yang Y, Chen D, Lin Z. Optimization algorithm inspired deep neural network structure design, in: J. Zhu, I. Takeuchi (Eds.), Proceedings of The 10th Asian Conference on Machine Learning, Vol. 95 of Proceedings of Machine Learning Research, PMLR, 2018, 614–629.

[134] Zhang X, Lu Y, Liu J, Dong B. Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration, in: ICLR, 2019.

[135] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. in: CVPR, 2016, 770–778.

[136] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. in: ECCV, 2016, 630–645.

[137] Weinan E. A proposal on machine learning via dynamical systems[J]. Communications in Mathematics and Statistics, 2017, 5(1): 1–11.

[138] Haber E, Ruthotto L. Stable architectures for deep neural networks[J]. Inverse Problems, 2017, 34(1): 014004.

[139] Chen T Q, Rubanova Y, Bettencourt J, Duvenaud D K. Neural ordinary differential equations. in: Advances in neural information processing systems, 2018, 6571–6583.

[140] Wang B, Yuan B, Shi Z, Osher S J. Enresnet: Resnet ensemble via the feynman-kac formalism,

arXiv preprint arXiv:1811.10745.

[141] Zhang D, Zhang T, Lu Y, Zhu Z, Dong B. You only propagate once: Painless adversarial training using maximal principle, in: NeurIPS, 2019.

[142] Lu Y, Li Z, He D, Sun Z, Dong B, Qin T, Wang L, Liu T Y. Understanding and improving transformer from a multi-particle dynamic system point of view, arXiv preprint arXiv:1906.02762.

[143] Yang Y, Sun J, Li H, Xu Z. Admm-csnet: A deep learning approach for image compressive sensing. IEEE transactions on pattern analysis and machine intelligence.

[144] Xu Z, Sun J. Model-driven deep-learning[J]. National Science Review, 2017, 5(1): 22–24.

[145] Parikh N, Boyd S, et al., Proximal algorithms[J]. Foundations and Trends® in Optimization, 2014, 1(3): 127–239.

[146] Katsevich A. Theoretically exact filtered backprojection-type inversion algorithm for spiral ct[J]. SIAM Journal on Applied Mathematics, 2002, 62(6): 2012–2026.

[147] Adler J, Öktem O, Solving ill-posed inverse problems using iterative deep neural networks. Inverse Probl. 33 (2017) 124007 (24pp).

MATHEMATICAL AND DEEP LEARNING METHODS IN IMAGE INVERSE PROBLEMS

Dong Bing

(*Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China*)

Abstract

We live in the digital age, and data has become an essential part of our lives. Images are undoubtedly one of the most important types of data. Image inverse problems, including image denoising, deblurring, restoration, biomedical imaging, etc., are important areas in imaging science. The rapid development of computer technology has enabled us to use sophisticated mathematics and machine learning tools to design effective algorithms for image inverse problems. This paper mainly reviews three types of methods in image inverse problem, namely, applied and computational harmonic analysis method (represented by wavelets and wavelet frames), partial differential equation (PDE) method and deep learning method. We will review the modeling philosophies of these methods, explore the connections and differences among them, their advantages and disadvantages, and further discuss the feasibility and benefit of the integration of these methods.

Keywords: Image inverse problem; image reconstruction; medical imaging; image recognition; variational model; partial differential equations; wavelet transform; convolutional neural network; deep learning

2010 Mathematics Subject Classification: 35A15, 41A25, 42C40, 45Q05, 65K10, 68U10, 68T05, 68T45