

数据驱动的稀疏编码综述及展望

周箴¹

1. 福州大学 物理与信息工程学院, 福建 福州 350108

摘要: 稀疏编码是一种高效的表达信息的手段, 它在图像信号处理、计算机视觉以及模式识别等领域具有广泛应用。然而, 当前许多大规模的实际应用依然受限于稀疏求解器的速度。数据驱动的稀疏编码是一种将传统的基于迭代的稀疏编码算法展开为深度神经网络从而实现快速逼近稀疏解的高效的编码手段。近年来, 从理论创新到应用的扩展, 这种新颖的稀疏编码策略取得了巨大的成功并得到了越来越多的关注。文章首先简要概述传统的稀疏编码算法。其次, 文章介绍经典算法 LISTA (Learned Iterative Shrinkage Threshold Algorithm), 着重从类 LISTA 算法、理论和应用研究三个方面对前沿工作的进展综述, 并对主要算法进行结构性的归纳与总结。最后, 文章讨论现有工作的不足以及建议未来可能的研究方向。

关键词: 稀疏编码; 深度学习; LISTA 算法; 数据驱动

文献标志码: A 中图分类号: TP391

Review and Prospects of Data-driven Sparse Coding

ZHOU Zheng¹, YANG Wen², SAKAI Tomoya^{3*}

1. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China

2. Network and Information Center, Chengdu University, Chengdu 610106, China

2. Graduate School of Engineering, Nagasaki University, Nagasaki 852-8521, Japan

Abstract: Sparse coding is an efficient method to represent information, it has achieved widespread applications in image signal processing, computer vision and pattern recognition. However, many large-scale practical tasks are still limited by the speed of sparse solvers. Data-driven-based sparse coding is an efficient coding strategy that unfolds traditional iterative algorithms into DNNs to speed-up the approximation of sparse solutions. Extending from theoretical innovations to application expansions, in more recent years, this new sparse coding strategy has achieved great success and received more and more attention. This paper briefly summarizes traditional sparse coding algorithms. Next, the paper introduces the classic algorithm LISTA (Learned Iterative Shrinkage Threshold Algorithm), gives an outline of the progress of the frontier works out of three aspects: LISTA-based algorithms, theoretical and practical researches, and classifies and sums up the main achievements in a structural way. At last, this article discusses the limitations of current works and suggests possible future research directions.

Key words: Sparse coding; Deep learning; LISTA; Data-driven

从已知的存在噪声的信号近似未知信号是一个 寻找欠定线性系统的稀疏解的问题, 对这个问题进行

求解可以解决很多具体的实际问题, 包括图像信号恢复、重建^[1-2]、合成^[3]以及去噪^[4]、计算机视觉中的识别任务^[5-7]等。近年来, 采用稀疏编码^[8]方法近似未知信号经成为当前最主流手段之一。传统的估计稀疏解的手段主要包括, 使用正交匹配算法(Orthogonal Matching Pursuit, OMP)^[9]及其扩展算法^[10-11]或从“凸松弛”的角度采用 L1 范数替换 L0 范数^[12-14], 并应用数学优化方法^[15-17]等。这些方案的思想是通过反复迭代逼近“真实”的稀疏解来实现, 通常都需要较多次的迭代才能得到理想的结果, 同时会产生大量资源消耗且不能适应实时工作。而一种使用编码器进行稀疏编码预测的基本思想^[18-19], 以及将这种编码器转变成深度学习模型的特殊形式和进行特殊参数化的策略^[20]仅需少数几次迭代就能更加快速的逼近稀疏解。此外, 这种将迭代算法展开为深度神经网络的思想结合了深度学习能够快速学习的特点, 还让深度学习这种“黑匣子”式的训练方式在稀疏编码器中得到有效的解释。

近十年来, 这种数据驱动的稀疏编码方法在算法改进以及应用创新上被国内外的学者持续不断的突破, 得到了越来越多的重视与发展。本文主要针对这种新颖的数据驱动的稀疏编码手段进行综述以及对未来可能的研究方向进行展望。本文的主要贡献如下:

(1) 文章首先概述传统的稀疏编码算法, 进而重点介绍什么是数据驱动的稀疏编码算法, 并对主要算法分类总结, 介绍其核心思想并对比分析其局限性;

(2) 文章分别从理论与应用研究两方面对前沿研究进行论述, 从基准算法、目标问题、训练参数数量、应用场景等几个方面详细归纳与总结数据驱动的稀疏编码算法, 通过仿真实验对主要算法进行性能评估并给出评价;

(3) 文章分两步从理论研究和应用研究方面梳理数据驱动的稀疏编码需要解决的关键性问题并对未来工作进行展望。

本文第一章介绍基本的稀疏编码迭代算法; 第二章对数据驱动的稀疏编码算法进行综述与总结, 第三章对未来研究工作方向进行展望。最后进行总结。

1 迭代算法

考虑一个有约束的最小化问题: 其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是光滑的函数 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一个可能不可微的凸函数:

$$\arg \min_x f(x) + g(x). \quad (1)$$

如果 $f(x) = \frac{1}{2} \|b - Ax\|_2^2$ 并且 $g(x) = \lambda \|x\|_1$, 其中 $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $m \ll n$, λ 是一个正标量, 可以调节稀疏强度, 则等式(1)中的最小化问题称为 LASSO 回归^[14]。针对此类 LASSO 最小化问题, 许多学者^[15-17,21]提出可以通过迭代收缩阈值(Iterative Shrinkage Thresholding, IST)思想对其进行求解。本节针对这种问题的求解简单介绍几种基本的迭代收缩阈值算法。

1.1 迭代软阈值算法

迭代软阈值算法 (Iterative Shrinkage Thresholding Algorithm, ISTA)^[15]是一种经典的求解 LASSO 回归的算法:

$$x^{(k+1)} = SF(x^{(k)} + \alpha A^T(b - Ax^{(k)}), \alpha\lambda). \quad (2)$$

其中, SF 是一个收缩算子, 定义为 $SF(x, \theta) = \text{sign}(x) \max(|x| - \theta, 0)$, k 表示迭代次数, $x^{(0)} = 0$ 。正常数 α 是步长, 步长的最佳值通常被分析得到是 $\frac{1}{L}$, L 是 ∇f 的最小的利普希茨常数, 其值为 $A^T A$ 的最大特征值。

1.2 快速迭代软阈值算法

与 ISTA 相比, 快速迭代软阈值算法 (Fast Iterative Shrinkage Thresholding Algorithm, FISTA)^[21]通过对 ISTA 添加“Nesterov”加速度, 即用一个特殊的线性组合乘以软阈值收缩运算的最后两个输出之差。此时, 收缩运算针对 $z^{(k)}$ 起作用, FISTA 可以写成:

$$z^{(1)} = x^{(0)}, t_{(1)} = 1, \quad (3)$$

$$x^{(k)} = SF(z^{(k)} + \alpha A^T(b - Ax^{(k)}), \alpha\lambda), \quad (4)$$

$$t_{(k+1)} = \frac{1}{2} \left(1 + \sqrt{1 + 4t_{(k)}^2} \right), \quad (5)$$

$$z^{(k+1)} = x^{(k)} + \frac{t_{(k)} - 1}{t_{(k+1)}} (x^{(k)} - x^{(k-1)}). \quad (6)$$

其研究表明, ISTA 的复杂度为 $O\left(\frac{1}{k}\right)$, 而 FISTA 的复杂度为 $O\left(\frac{1}{k^2}\right)$, 其性能得到了有效提升。

1.3 分裂增广拉格朗日收缩算法

分裂增广拉格朗日收缩算法 (Split augmented Lagrangian shrinkage algorithm, SALSA)^[16-17]是一种引申于基于乘数的交替方向方法(Alternating Direction

Method of Multipliers, ADMM)的算法, 其数学表达为:

$$\mathbf{v} = SF\left(\mathbf{x}^{(k)} + \mathbf{y}^{(k)}, \frac{\lambda}{\rho}\right) - \mathbf{y}^{(k)}, \quad (7)$$

$$\mathbf{x}^{(k+1)} = (\mathbf{A}^T \mathbf{A} + \rho \mathbf{I})^{-1} (\mathbf{A}^T \mathbf{b} + \rho \mathbf{v}) \quad (8)$$

$$\mathbf{y}^{(k+1)} = \mathbf{x}^{(k+1)} - \mathbf{v}. \quad (9)$$

$\rho \in \mathbb{R}$ 是一个正标量。这种方法在解决 L1 范数问题或 TV 范数问题的磁共振图像问题上较 FISTA 更快。

1.4 近似消息传递算法

近似消息传递算法(Approximate Message Passing, AMP)^[22-23]是一种基于概率图思想通过状态演化迭代产生稀疏解的方法:

$$\mathbf{q}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)} + \mathbf{s}^{(k)}\mathbf{q}^{(k-1)}, \quad (10)$$

$$\mathbf{x}^{(k+1)} = SF(\mathbf{x}^{(k)} + \mathbf{A}^T \mathbf{q}^{(k)}, \lambda^{(k)}), \quad (11)$$

其中 $\mathbf{q}^{(-1)} = \mathbf{0}$, $\mathbf{s}^{(k)} = \frac{1}{m} \|\mathbf{x}^{(k)}\|_0$, $\lambda^{(k)} = \frac{\mu}{\sqrt{m}} \|\mathbf{q}^{(k)}\|_2$, $\mu \in \mathbb{R}$ 是一个和 λ 有关的可以进行调整的参数。这种算法要求矩阵 \mathbf{A} 必须是一个高斯或亚高斯随机矩阵。

1.5 迭代硬阈值算法

和以上几种算法不同, 迭代硬阈值算法 (Iterative Hard Thresholding, IHT)^[24]是一种通常用于求解 L0 范数最小化问题的方法, 此时, $g(x) = \lambda \|\mathbf{x}\|_0$, IHT 的数学表达如下:

$$\mathbf{x}^{(k+1)} = \text{hard}_{\lambda^{0.5}}(\mathbf{x}^{(k)} + \mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)})). \quad (12)$$

其中 $\text{hard}_{\lambda^{0.5}}(u)$ 是一个逐元素的硬阈值算子:

$$\text{hard}_{\lambda^{0.5}}(u_i) = \begin{cases} 0, & \text{if } |u_i| < \lambda^{0.5} \\ u_i, & \text{if } |u_i| \geq \lambda^{0.5} \end{cases} \quad (13)$$

2 数据驱动的迭代算法

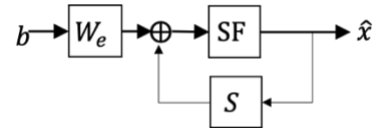
近十年中, 将迭代算法展开为深度神经网络来加速求解稀疏解的做法已经成为图像信号处理新的主要内容。LISTA(Learned Iterative Shrinkage Thresholding Algorithm)^[20]是一种基于 ISTA 的改进算法, 这种算法只需几次迭代就能实现较传统稀疏编码方法更优的性能。他们重写 ISTA 为:

$$\mathbf{x}^{(k+1)} = SF(\mathbf{W}_e \mathbf{b} + \mathbf{S}\mathbf{x}^{(k)}, \theta), \quad (14)$$

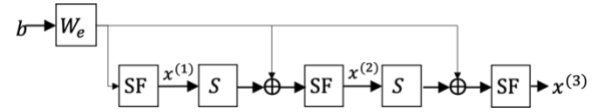
其训练参数为 $\theta = \{(\mathbf{W}_e, \mathbf{S}, \theta)\}$ 。各参数初始化为 $\mathbf{W}_e = \frac{1}{L} \mathbf{A}^T$, $\mathbf{S} = \mathbf{I} - \frac{1}{L} \mathbf{A}^T \mathbf{A}$, $\theta = \frac{\lambda}{L}$ 。该方法将 ISTA 看作是一个循环神经网络, 其结构框图如图 1(a)所示, 通过将 ISTA 迭代层展开作为深度神经网络层, 他们进而展示了 LISTA 的共享权重的结构框图(如图 1(b)所示)。其中, 共享权重指的是, 对于 K 层模型, 每个权重只有

一个可训练参数, 并且该参数在每一层(或每次迭代)都得到训练, 即权重设计与索引无关。三次迭代的 ISTA 框图可以转化为图 1(b)所示的三层 LISTA 的深度框架, LISTA 通过模型重构加快了逼近“真实”的稀疏解的速度。同时他们对坐标下降算法(Coordinate Descent algorithm, CoD)^[25]也采取了同样的策略, 并在 MNIST 手写字^[26]识别实验上较好的验证了这种策略。

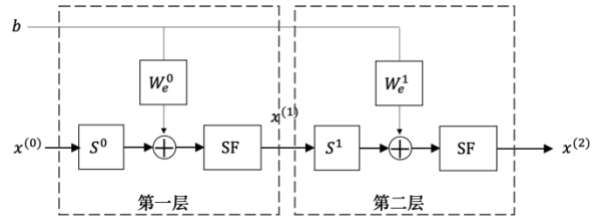
这种数据驱动的稀疏编码方法的新颖性主要在于将一般的求解 LASSO 回归的迭代算法展开形成计算图, 权重结构的设计确保可以用深度学习的方法来训练它们, 并确保可以通过它们对梯度进行反向传播。



(a) ISTA 框图



(b) 共享权重的 LISTA 框图



(c) 非共享权重的 LISTA 框图

图 1 ISTA 和 LISTA 算法的结构框图

Fig.1 Block diagram of ISTA and LISTA.

2.1 类 LISTA 的迭代算法

近年来, 国内外的研究者通过对 LISTA^[20]算法的思想进行学习与扩展, 提出了许多类 LISTA 的变体算法^[27-32]。本小节对当前主要的算法进行梳理与总结并进行对比分析。

2.1.1 LISTA 的改进算法

Moreau 等人^[33]发现他们的数学分析和研究^[34]的实验结果都表明 LISTA 算法的共享权重设计较非共享权重设计效率更低。其中非共享权重表明在 K 层 LISTA 模型中, 每个权重都有 K 个可训练参数, 并且这 K 个参数在相应的层中独立训练, 即权重设计与索

引有关。研究^[35]将提出 LISTA-tied, 其将非共享的 LISTA 的权重结构改写为 $\mathbf{W}_e^{(0)}, \mathbf{S}^{(0)}$ 和 $\theta^{(k)}$, 进而减少了训练参数并相比于非共享权重下的 LISTA 实现了更优的结果。

Chen 等人^[36]介绍了 LISTA 的非共享的权重图(如图 1(c)所示)并提出了一种部分权重耦合算法(LISTA with partial weight coupling, LISTA-CP):

表 1 LISTA 的改进算法总结

Table 1 Summary of LISTA-based improved algorithms

算法	训练参数	应用场景	优点	局限性
LISTA ^[20] (2010)	$O(mn + n^2 + 1)$	分类任务	较 ISTA 算法得到了可观的性能提升, 较少迭代次数就能较快的逼近“真实”稀疏解	研究没有提供理论证明
非共享的 LISTA ^[33-34] (2016,2017)	$O(Kmn + Kn^2 + K)$	--	性能相比 LISTA 提升, 收敛速度更快, 得到理论验证分析	训练参数较 LISTA 更多, 计算、时间等资源消耗更多
LISTA-tied ^[35] (2016)	$O(mn + n^2 + K)$	--	随着迭代次数增加, 较 LISTA 和非共享 LISTA, 其性能更优	仅存在仿真结果, 研究没有提供应用场景
LISTA-CP ^[36] (2018)	$O(Kmn + K)$	压缩感知	提出权重耦合的观点, 给出了理论证明与实验分析, 重构了 LISTA 模型	训练参数较 LISTA 和非共享的 LISTA 较少, 但由于是非共享权重结构, 矩阵权重参数依旧非常多
TiLISTA/ALISTA ^[37] (2019)	$O(mn + 2K)/O(2K)$	图像去噪	训练参数大幅度减少、存储计算以及时间消耗大幅降低	ALISTA 算法在无监督任务上失效
SLISTA ^[38] (2019)	$O(K)$	无监督学习任务	训练参数最少, 无监督任务中当稀疏性提高性能较 LISTA 更优	研究没有对监督场景下的算法进行验证, 稀疏性需手动调节

LISTA-CP 可以进一步简化为 TiLISTA (Tied LISTA)^[37], Liu 等人继用 LISTA-CP 的模型框架, 在此基础上结合研究^[34]所提的 tied 权重结构^[34]的思想将权重矩阵 $\mathbf{W}_e^{(k)}$ 改写为 $\mathbf{W}_e^{(k)} = \gamma^{(k)} \mathbf{W}_e$, 进而只需要训练一个与迭代索引无关的矩阵权重 \mathbf{W}_e 和非共享权重参数 $\gamma^{(k)}$ 和 $\theta^{(k)}$, 更进一步的大幅度减少了训练参数, 其训练参数变化为 $\boldsymbol{\theta} = \{\mathbf{W}_e, \{\gamma^{(k)}, \theta^{(k)}\}_{k=0}^{K-1}\}$, 其模型为:

$$\mathbf{x}^{(k+1)} = SF(\mathbf{x}^{(k)} - \gamma^{(k)} \mathbf{W}_e (\mathbf{b} - \mathbf{A} \mathbf{x}^{(k)}), \theta^{(k)}), (16)$$

此外, 他们提出一种基于提前分析计算的算法 ALISTA (Analytic LISTA), 通过分析给定的字典 \mathbf{A} 解决一个凸二次优化问题, 预计算得到一个“好的”权重矩阵 $\tilde{\mathbf{W}}$, 从而只需要训练非共享的权重系数 $\theta = \{\gamma^{(k)}, \theta^{(k)}\}_{k=0}^{K-1}$ 。此方法较 LISTA 算法^[20]及其不同权重结构下的改进算法^[33-35]性能更优同时避免了字典学习这种大规模矩阵计算从而实现了计算以及内存的减负。

然而, 研究^[38]提出 ALISTA 不适用于非监督学

$$\mathbf{x}^{(k+1)} = SF(\mathbf{x}^{(k)} - \mathbf{W}_e^k (\mathbf{b} - \mathbf{A} \mathbf{x}^{(k)}), \theta^k), (15)$$

其训练参数为 $\boldsymbol{\theta}^{(k)} = \{(\mathbf{W}_e^{(k)}, \theta^{(k)})\}_{k=0}^{K-1}$ 。他们进而验证了这种结构在一定条件下, 随着迭代次数 k 的增大阈值 θ^k 会趋向于 0, 耦合矩阵之间存在一定的关系, 即 $\mathbf{S}_e^{(k)} - (\mathbf{I} - \mathbf{W}_e^{(k)} \mathbf{A}) \rightarrow 0$ 。LISTA-CP 算法在一定程度上简化了 LISTA 模型并减少了训练参数。

习场景, 当应用于非监督学习的场景时, 他们的研究表明一种只学习步长 α 的策略 SLISTA (Step-LISTA) 在稀疏解足够稀疏时可以有效提高算法收敛速度。这种策略的训练参数仅为 $\theta = \{\alpha^{(k)}\}_{k=0}^{K-1}$, 他们也采用了 LISTA-CP 的权重耦合框架:

$$\mathbf{x}^{(k+1)} = SF(\mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{A}^T (\mathbf{b} - \mathbf{A} \mathbf{x}^{(k)}), \alpha^{(k)} \lambda). (17)$$

相比 ALISTA, 这种方案更进一步的减少了训练参数。

以上几种算法都主要针对 LISTA 的权重设计进行改进, 分别提出采用非共享、绑定、耦合、预计算以及去除矩阵权重计算等思想来减少训练参数并提升编码器的速度。相比于 LISTA 算法, 这些算法在不同程度上减少了计算以及资源消耗。针对这一类的算法, 本小节给出具体的训练参数、应用场景、优点以及局限性等方面的总结 (如表 1 所示)。其中当前性能综合最优的 ALISTA 算法得到了计算以及效率的提升却被证明不适用于非监督学习场景, 训练参数最少的 SLISTA 算法仅仅针对非监督学习得到了理想的结果,

没有对监督学习进行扩展同时还需要手动调节稀疏性。

2.1.2 LISTA 的变体算法

由于 FISTA 可以有效加快逼近“真实”稀疏解的速度,因此许多研究者通过结合 FISTA 和 LISTA 的思想来进一步构建新的算法获取稀疏解。Moreau 等人^[33]将 FISTA 中的特殊线性组合到训练参数中,对新的权值参数采用手动初始化的方式展开 FISTA,得到了一种学习的 FISTA 算法(Learned FISTA, LFISTA):

$$\mathbf{x}^{(k+1)} = SF(\mathbf{W}_g^{(k)} \mathbf{x}^{(k)} + \mathbf{W}_m^{(k)} \mathbf{x}^{(k-1)} + \mathbf{W}_e^{(k)} \mathbf{b}, \theta^{(k)}), \quad (18)$$

其中 $\mathbf{W}_g^{(k)} = \left(1 + \frac{t_{(k-1)} - 1}{t_{(k)}}\right) \mathbf{S}$, $\mathbf{W}_m^{(k)} = \left(\frac{1 - t_{(k-1)}}{t_{(k)}}\right) \mathbf{S}$, 训练参数为 $\theta = \{(\mathbf{W}_g^{(k)}, \mathbf{W}_m^{(k)}, \mathbf{W}_e^{(k)}, \theta^{(k)})\}_{k=0}^{K-1}$ 。LFISTA 的实验结果体现了其算法的有效性,但这种模型的训练参数非常多,采用非共享的方式使得每一次迭代都包含了三个矩阵权重需要训练,这使得 LFISTA 难以像 TiLISTA 一样在训练参数上得到进一步的优化。此外,训练参数 $\mathbf{W}_g^{(k)}$ 和 $\mathbf{W}_m^{(k)}$ 的每一层初始化值都需要手动计算,算法相对复杂并增加了算法的人工成本。

FISTA 中添加的“动能”项是一种对历史信息计算(即计算最后两次迭代的输出之差)得到有效动能

补偿来实现快速逼近稀疏解的逐元素计算的方法。在类 LISTA 的迭代算法中,一些算法^[39-41]也采用这种思路来实现,他们通过对历史信息回顾来刺激或补偿稀疏解在估计中比预期更低的成分从而实现稀疏编码器的加速。研究^[39]作用于阈值算子;Zhou 等人^[40]介绍了一种自适应的“动能”策略来保证参数的更新以及对历史信息的封装,他们是第一个将稀疏编码算法和长短期记忆模型(Long Short Term Memory, LSTM)^[42]结合到一起形成一个新的循环神经网络模型的研究,使用 LSTM 单元而不是简单的 RNN 单元,因此更加有助于捕获历史信息并能加快收敛速度。Wu 等人^[41]通过假设 LISTA 算法中的稀疏解的成分增益不足需要补偿,提出使用“门”策略,一种“增益门”的策略单独作用于软阈值收缩算子运算前的稀疏解,而“过冲门”的策略作用于阈值收缩算子运算之前和之后的稀疏解。这两种“门”策略都需要添加相关的新的计算步骤,也会增加相应的训练参数。其中在阈值算子之后添加的“门”策略是根据时间和输入变化的,但是由于额外的方程以及计算的加入,这种策略下的稀疏编码会变得更加复杂同时增加了资源消耗。

表 2 LISTA 的变体算法总结

Table 2 Summary of LISTA-based variant algorithms

算法	训练参数	应用场景	优点	局限性
FISTA ^[21] (2009)	--	图像去噪	较 ISTA 算法在收敛速度上有较大提升	传统的方法,不是学习的方法。较 LISTA 在实验中的收敛速度更慢
LFISTA ^[33] (2017)	$O(Kmn + Kn^2 + Kn^2 + K)$	无监督学习任务	提出新的变体模型,性能提升,具备理论证明	仅在无监督任务上得到验证,且训练参数非常多,模型更加复杂
SC2Net ^[40] (2018)	$O(mn + n^2 + K + K\phi)$, 训练参数由改进的 LSTM 单元 ϕ 决定	图像去噪	第一个将经典稀疏优化方法和 LSTM 结合到一起的算法,在监督和无监督学习中得到有效验证	添加 LSTM 单元后的模型相对更加复杂
GLISTA ^[41] (2020)	$O(Kmn + Kn^2 + K + K\phi)$ 根据算法添加的功能方程不同,需要增加不同训练参数 ϕ	光度立体视觉任务	采用类似于 gated recurrent unit(GRU) ^[43] 设计,提出了多种“门”功能函数,提供了新的视野并提高了性能	训练参数相比于 LISTA 在某些条件下设置下参数增加较多,同时功能函数的计算使得资源消耗更多

FISTA 算法是一种改进的 ISTA 算法,能够实现更加快速的收敛速度。从表 2 可知,这些算法借鉴了 FISTA 添加 Nesterov 加速度以及 LISTA 将迭代层次展开为深度网络的思想,通过对单元模块的整合与添加等方式构建了不同 LISTA 变体模型。他们从不同的角度考虑如何加速稀疏编码器,为领域内的其他学者提供了新的思路,但相比于前文中基于 LISTA 的改进算

法,他们都在不同程度上使得模型变得更加复杂并增加了一定的训练参数。

2.1.3 其他变体算法

研究^[44]将 SALSALSA 算法展开成深度神经网络并将其应用于图像重建以及分类任务中。LISTA 的想法启发了研究^[45]对 AMP 算法及其的贝叶斯扩展。

Borgerding 和 Schniter^[35]通过将 AMP 中 Onsager 修正项 $s^{(k)} q^{(k-1)}$ (等式) 应用到基于深度学习思想的稀疏编码算法中, 提出了一种学习的 AMP 算法(Learned AMP, LAMP)。研究^[46]提出类似的基于学习的模型, 该模型也是受到 AMP 算法的启发。

Wang 等人^[47]对迭代硬阈值 IHT 算法进行展开从而得到了它的深度模型。研究^[48]将迭代算法展开为类似于神经网络的分层结构。研究^[49]针对一种 LASSO 的变体问题提出了一种构造前馈网络的方法, 以加快稀疏编码器的编码速度, 该方法的基准算法为 ADMM 算法。研究^[50]构造了可学习的追踪过程架构, 该架构是从近端下降算法的迭代中得出。当涉及到 L_{∞} 范数最小化问题, 研究^[51]基于 ADMM 算法将原始凸最小化问题表示为一个前馈神经网络。

2.2 理论与应用研究

2.2.1 理论研究

数据驱动的稀疏编码方法在仿真数据以及真实数据的实验中都较好的展现出了其更高效、更精准的结果, 其理论研究^[33-34,36,41,52-53]也在逐步完善中。研究^[33]用一种特殊的因式分解的数学分析方法阐述了 LISTA 算法能快速逼近稀疏解的原因, 并重新参数化 LISTA 为新的因式分解后的结构, 得到了与 LISTA 近似的实验结果。Xin 等人^[34]调查剖析了将传统稀疏编码算法展开迭代构建深度网络层的好处, 精确阐述了各种不同的体系结构选择如何影响和改善限制性等距特性 (RIP) 常数的能力。他们使用 LISTA 结构的思想来解决 L_0 范数最小化问题, 被认为是一种纯粹基于优化讨论的深度学习的方法。Chen 等人^[36]证明了存在一组理想的参数可以使得 LISTA 实现线性收敛速度, 同时提出一种“支持选择”策略, 这种策略通过假设作用于阈值算子选择学习足够大的偏置可以保证阈值算子后的结果中不存在“假阳性”的成分来实现, Wu 等人的研究^[41]对这种假设进行了进一步的研究并作出更合理的假设从而得到了更严格的界限。Mardani 等人^[52]通过使用一种特殊的估计器来评估迭代算法展开神经网络的预测误差并给出了相应的理论分析。Takabe 和 Wadayama^[53]通过理论分析证实了合

适的步长能够提高基于深度展开的梯度下降方法的收敛速度。

2.2.2 应用研究

当前数据驱动的稀疏编码方法主要应用于图像信号处理与压缩感知方面, 包括图像去噪、修复、重建等方面。越来越多的研究将稀疏编码模型扩展到卷积稀疏编码模型 (Convolutional Sparse Code, CSC) 领域^[54-58]。研究^[54]提出了一种受约束的递归稀疏自动编码器体系结构, 该体系结构可学习真正的底层卷积滤波器。采用卷积字典替换常规字典能够更好的保留图像特征, 研究^[55]重写了 ISTA 作为 CSC 问题的解决方案, 并且在此框架中学习了稀疏编码和字典。他们的创新工作表明, 训练过的 CSC 可以很好地处理整个图像, 相比于 KSVD 的去噪技术^[59]与其他一些 CSC 模型的图像修复技术^[60-61], 他们的算法表现出更加出色的性能。ALISTA 同样将这种 CSC 思想融入进他们的算法中, 进而提出一种鲁棒性更强的卷积 ALISTA。Zhang 和 Ghanem^[62]提出了一种新颖的 ISTA-Net, 他们通过使用非线性变换来解决与稀疏诱导正则化器相关的近端映射, 从而将 ISTA 转换为深度神经网络形式, 他们建议学习所有的参数, 阈值, 步长和非线性变换。这是使用基于数据驱动的 ISTA 算法解决大规模的自然图片的压缩感知重建问题的一种重要扩展方式, 但这种策略的可训练参数会相对较多。研究^[24,36-37,63]将可训练的稀疏编码分别应用于自然图片的压缩感知, 自然图片的去噪以及图像修补等应用场景中, 得到了较好的结果。此外, 一些研究^[64-66]也将迭代算法展开为深度神经网络的思想运用于医学图像处理当中。传统算法 ISTA 是一种非常适合解决大规模的线性逆问题的方法, 现有研究表明^[67], 基于数据驱动的稀疏编码方法在处理大规模的应用问题时同样具有较大的优势。

2.3 算法归纳与总结

在前面的小节中, 文章分类对部分算法的进行总结了归纳, 主要包括算法 K 次迭代的训练参数、应用场景、优点以及局限性等方面。为了更进一步的全面总结与对比, 本小节首先按照时间顺序梳理了当

前主流的数据驱动的稀疏编码算法，列出其基准算法、目标问题、K 次迭代的训练参数的数量、是否给出理论分析以及他们的应用场景。其次，本小节对主流的方法通过实验进行量化对比分析。

表 3 基于数据驱动的稀疏编码算法总结

Table 3 Summary of algorithms of data-driven-based sparse coding

算法	基准算法	目标问题	训练参数数量	理论分析	应用场景
LISTA/LCoD ^[20] (2010,ICML)	ISTA/CoD	L1 范数最小化问题	$O(mn + n^2 + 1)$	×	分类任务
Analysis-ADMM ^[49] (2013,NeurIPS)	ADMM	类 LASSO 回归问题	训练参数采用滤波器的形式表示	√	图像重构以及分类任务
Wang 等人 ^[51] (2016,IJCAI)	ADMM	L_∞ 范数最小化问题	$O(mn + n^2 + Kn + K)$	×	图像哈希应用
Deep l_0 -Regularized ^[47] (2016,AAAI)	IHT	L0 范数最小化问题	$O(mn + n^2 + K)$	√	分类任务以及聚类任务
LAMP/LAMP-tied ^[35] (2016,GlobalSIP)	AMP	L1 范数最小化问题	$O(Kmn + K)/O(mn + K)$	×	--
LFISTA ^[33] (2017,ICLR)	FISTA	L1 范数最小化问题	$O(Kmn + Kn^2 + Kn^2 + K)$	√	无监督学习任务
ACSC ^[56] (2018,ICASSP)	ISTA	L1 范数最小化问题	训练参数根据卷积滤波器的大小变化	×	图像去噪以及修复
LISTA-CP/LISTA-CPss ^[36] (2018,NeurIPS)	ISTA	L1 范数最小化问题	$O(Kmn + K)$	√	压缩感知
ISTA-Net ^[62] (2018,CVPR)	ISTA	L1 范数最小化问题的变体	$O(K + \psi)$ 训练参数由非线性变换 ψ 决定	√	压缩感知与核磁共振成像
SC2Net ^[40] (2018,AAAI)	ISTA	L1 范数最小化问题	$O(mn + n^2 + K + K\phi)$	√	图像重建
TISTA ^[32] (2018, IEEE Transactions on Signal Processing)	ISTA	L1 范数最小化问题	$O(K)$	√	图像重建
LSALSA ^[44] (2018,Machine Learning)	SALSA	L1 范数最小化问题	$O(mn + n^2 + K)$	√	图像重建以及分裂任务
ALISTA ^[37] (2019,ICLR)	ISTA	L1 范数最小化问题	$O(K)$	√	图像去噪
SLISTA ^[38] (2019,NeurIPS)	ISTA	L1 范数最小化问题	$O(K)$	√	无监督学习任务
GLISTA ^[41] (2020,ICLR)	ISTA	L1 范数最小化问题	$O(Kmn + Kn^2 + K + K\phi)$	√	光度立体视觉任务
LISTA-AT ^[39] (2020,IEEE Access)	ISTA	L1 范数最小化问题	$O(Kmn + K)$	√	图像重建
Ada-LISTA ^[30] (2020)	ISTA	L1 范数最小化问题	$O(mn + n^2 + K)$	√	图像修复

从表 3 可知，L1 目标问题是当前学术界研究最多的一种问题，以 ISTA 为基准算法的研究较其他算法更多。权重结构的设计决定了训练参数的多少是当前研究者比较关注的内容。大部分的研究都包含了理论分析，但理论分析部分还是相对不够充分，需要进行更多的研究。应用场景部分主要研究重点在于图像信号处理任务、计算机视觉的分类任务以及基于压缩感知的感知任务和医学图像处理等问题。

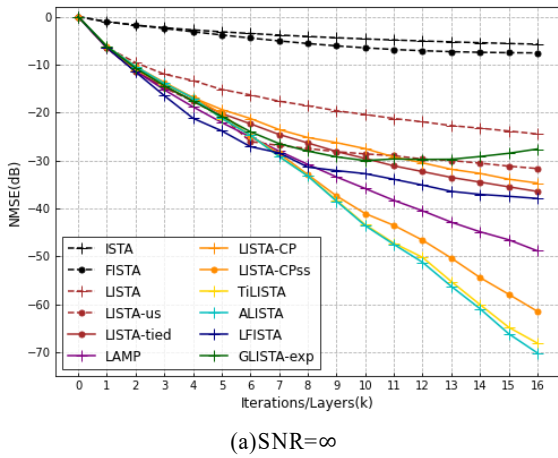
本文通过实验对当前主要的算法进行性能评估，包括传统算法和数据驱动的稀疏编码算法，首先由于大多算法研究采用图的形式体现其性能，本文同样通过采取这种方式，其次由于实验条件以及各类算法针对的问题不同，本文仅对主要的算法进行实验及性能

比较。具体的实验设置为：感知矩阵 A 形状为 $m = 250$, $n = 500$ 。 A 是从标准高斯分布 $A_{ij} \sim N(0, 1/m)$ 中采样得到的。对于基准稀疏向量 x^* ，非零项的采样个数为 n 的 10%，每个非零项都从标准高斯分布中采样，并且该样本假定所有项上的分布均匀， $\lambda = 0.4$ 。训练批尺度大小为 64，验证批尺度大小为 1000，测试数据为 1000，优化器为 Adam。应对不同的噪音设置，文章模拟产生相应的测试数据，所有的算法在相应的测试数据上进行测试，确保每个算法的实验环境一致。对于训练和验证数据，本文基于一种理想情况产生，每个批尺度的产生都遵循同样的分布并实时产生，训练过程遵循研究^[36-37]中以不同的学习速率训练数据的方法。

法¹。实验框架环境为 TensorFlow-GPU 2.2.0, GPU 是 GeForce RTX 2080。文章使用 NMSE^[36]来评估算法:

$$NMSE(\hat{x}, x^*) = 10 \log_{10} \left(\frac{\| \hat{x} - x^* \|_2^2}{\| x^* \|_2^2} \right). \quad (19)$$

本文通过两组仿真实验来进行算法评估, 分别为无噪声和有噪声情况下的对照实验。无噪声时 SNR (Signal-to-Noise Ratio) 为 ∞ , 有噪声时 SNR 为 20。为



了便于比较, 文章采用 LISTA-us 表示非共享权重的 LISTA 算法。LISTA-CPss 表示添加了“支持选择”策略^[35]后的算法, TiLISTA, ALISTA 以及 GLISTA 中都添加了该策略。Wu 等人^[40]研究了多种“门”策略, 由于在他们所展示的实验结果中“exp”策略与他们所提的其他策略性能相似, 所以本文仅对“exp”策略进行实验, 本文用“GLISTA-exp”对其进行表示。

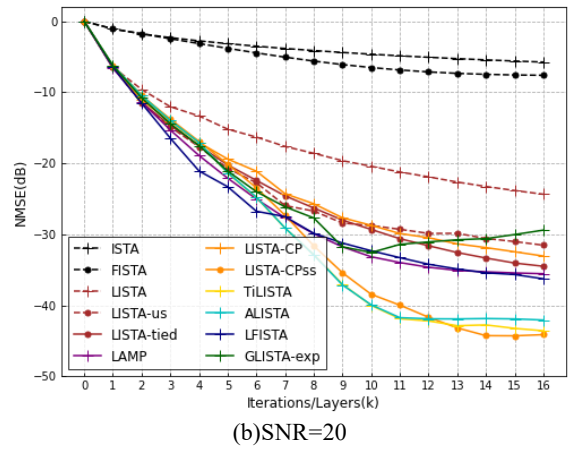


图 2 算法的仿真实验结果

Fig.2 Simulation results of algorithms

如图 2 所示, 文章对 12 种算法进行了仿真实验, 包括传统稀疏编码算法, LISTA 的改进算法, LISTA 的变体算法以及基于 AMP 的数据驱动稀疏编码算法。从实验结果可知:

(1) 相比于传统算法 ISTA 和 FISTA, 在不同信噪比的条件下, 数据驱动的稀疏编码算法在性能上都有明显提升。

(2) LISTA 的权重设计不同, 性能也明显不同, 其中 tied 权重设计最优。

(3) 多种 LISTA 的改进算法中, ALISTA 的训练参数最少, 同时性能相对较优。

(4) 多种 LISTA 的变体算法中, 在不同的信噪比的条件下, 他们的性能相比 LISTA 更优, 一定程度上可以为数据驱动的稀疏编码算法提供新的研究方向。

(5) 其他的变体算法中可以看到 LAMP 算法性能也具有一定可比性, 这能为其他变体算法的研究提供新的思路。

3 展望

数据驱动的稀疏编码方法在近十年中得到了迅速发展, 但其理论研究以及实际应用方面仍然存在不足, 本节主要从这方面来阐述未来可能的研究方向。

3.1 理论展望

本小节将从三个方面来讨论理论研究未来可能的工作方向。当前, 许多科研工作者都关注于 LISTA 算法的变体算法的研究, 变体算法不仅仅需要实现优越的收敛性能, 同时还需要从数学分析中得到解释和证明。一方面, 实现更加快速的收敛性能可以从研究^[36,41]中所提出的补偿稀疏编码估计能力方面考虑, Wu 等人的研究^[40]提供了多种“门”策略来实现补偿, 但是他们都不不同程度的使算法变得更加复杂, 并且为了更快收敛, 牺牲了时间以及其他资源的消耗, 在大规模的实际应用中不太实用, 如何均衡有效补偿与资源耗损将是一个具有现实意义的议题。另一方面, 现存的稀疏编码算法还没得到完全的研究和解释, 比

¹ <https://github.com/VITA-Group/LISTA-CPSS>

如 FISTA, SALSA, ADMM 等算法还没得到真正的延伸与拓展, 究竟哪一种稀疏编码算法更适用于什么样的领域, 性能能否更加优越, 与一般的深度学习的算法相比可训练的稀疏编码更加适用于什么样的领域, 还需要更多的研究者进行学习探讨。第三, 基于数据驱动的稀疏编码算法的权重结构的设计直接关系到训练参数的多少, 而训练参数对深度神经网络起着关键的作用。迭代型算法由于其迭代的特性和一般深度学习网络有着不一样的权重设计, 深度学习网络可以对不同网络层设置不同的权重结构, 而数据驱动的稀疏编码算法的权重设计较为固定, 主要根据其基准算法来确定, 其展开为深度神经网络后, 不同迭代之间的权重几乎一致, 但是他们可以以共享或不共享的方式存在, 如何充分利用这一特性提出更加轻便的网络模型也可能是一个重要的议题。其次, 此类权重结构的设计也需要得到数学上的证明与剖析。

3.2 实际应用的展望

纵观目前可训练稀疏编码的研究成果可知, 实际应用的场景还远远不够。现今的应用场景主要在一般的图像处理、信号处理等场景中, 包括, 图像去噪、重建、修复、聚类以及压缩感知等方面。研究者主要关注点还在理论与算法改进突破方面, 提出算法的改进创新再将其应用于一般的信号图像处理场景加以证实其实际应用的可行性以及优势还是当前的主要的研究趋势。然而, 传统的稀疏编码算法的利用场景众多, 比如, 推荐系统、医学图像处理^[68-69]以及识别等领域都有涉及。如何在快速高效获得稀疏解的情况下, 将迭代算法结合深度神经网络模型进行学习, 进而推动这种可训练稀疏编码算法的应用有望成为下一个新的研究方向。本文主要提出三个方面的可能的研究趋势, 第一, 替代传统稀疏编码算法更高效的实现准确推荐; 第二, 继续推进图像信号处理场景下的应用发展, 包括医学图像处理, 识别应用等, 其中识别应用可以包括计算机视觉中的人脸识别应用以及网络信息安全中恶意病毒、软件等方面的识别系统的应用研究等。

4 结束语

本文对数据驱动的稀疏编码算法进行了回顾以

及展望。文章从传统的稀疏编码迭代算法出发, 重点论述了将传统迭代算法展开为深度学习网络的思想及其扩展研究; 从理论创新到实际应用分别综述了数据驱动的稀疏编码的前沿研究; 结构性的归纳和总结了当前主要的算法研究并通过仿真实验给出性能评估。最后, 文章分两步从多个方面系统的阐述了未来可能的研究方向。

参考文献:

- [1] BLOUCAS-DIAS J M, FIGUEIREDO M A T. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration[J]. IEEE Transactions on Image processing, 2007, 16(12): 2992-3004.
- [2] MAIRAL J, ELAD M, SAPIRO G. Sparse representation for color image restoration[J]. IEEE Transactions on image processing, 2007, 17(1): 53-69.
- [3] 张华, 曹林. 结合 pHash 和稀疏编码的素描人脸合成方法[J]. 计算机工程与应用, 2019, 22.
ZHANG H, CAO L. Face Sketch Synthesis Method Combining pHash and Sparse Coding[J]. Computer Engineering and Applications, 2019, 22.
- [4] ELAD M, AHARON M. Image denoising via sparse and redundant representations over learned dictionaries[J]. IEEE Transactions on Image processing, 2006, 15(12): 3736-3745.
- [5] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 31(2): 210-227.
- [6] 潘婷婷. 稀疏编码算法改进及其在人脸识别中的应用[D]. 山东大学, 2014.
PAN T T. Improvement of sparse coding and its application in face recognition. Jinan: Shandong University, 2014.
- [7] 蔡家柱. 基于稀疏表达的人脸识别算法研究与实现[D]. 成都: 电子科技大学, 2015.
CAI J Z. The research and implementation of algorithm of face recognition based on sparse representation[D]. Chengdu: University of electronic science and technology of China, 2015.
- [8] OLSHAUSEN B A, FIELD D J. Sparse coding with an overcomplete basis set: A strategy employed by V1?[J]. Vision research, 1997, 37(23): 3311-3325.
- [9] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit[J]. IEEE Transactions on information theory, 2007, 53(12): 4655-4666.
- [10] NEEDELL D, VERSHYNIN R. Uniform uncertainty

- principle and signal recovery via regularized orthogonal matching pursuit[J]. *Foundations of computational mathematics*, 2009, 9(3): 317-334.
- [11] WANG J, KWON S, SHIM B. Generalized orthogonal matching pursuit[J]. *IEEE Transactions on signal processing*, 2012, 60(12): 6202-6216.
 - [12] CHEN S, DONOHO D. Basis pursuit[C]//*Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*. IEEE, 1994, 1: 41-44.
 - [13] GILL P R, WANG A, MOLNAR A. The in-crowd algorithm for fast basis pursuit denoising[J]. *IEEE Transactions on Signal Processing*, 2011, 59(10): 4595-4605.
 - [14] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996, 58(1): 267-288.
 - [15] 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: A Journal Issued by the Courant Institute of Mathematical Sciences*, 2004, 57(11):1413-1457.
 - [16] AFONSO M V, BIOUCAS-DIAS J M, FIGUEIREDO M A T. Fast image recovery using variable splitting and constrained optimization[J]. *IEEE transactions on image processing*, 2010, 19(9): 2345-2356.
 - [17] AFONSO M V, BIOUCAS-DIAS J M, FIGUEIREDO M A T. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems[J]. *IEEE Transactions on Image Processing*, 2010, 20(3): 681-695.
 - [18] KAVUKCUOGLU K, RANZATO M A, LECUN Y. Fast inference in sparse coding algorithms with applications to object recognition[J]. *arXiv preprint arXiv:1010.3467*, 2010.
 - [19] JARRETT K, KAVUKCUOGLU K, RANZATO M A, et al. What is the best multi-stage architecture for object recognition?[C]//*2009 IEEE 12th international conference on computer vision*. IEEE, 2009: 2146-2153.
 - [20] GREGOR K, LECUN Y. Learning fast approximations of sparse coding[C]//*Proceedings of the 27th international conference on international conference on machine learning*. 2010: 399-406.
 - [21] BECK A, TEOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. *SIAM journal on imaging sciences*, 2009, 2(1): 183-202.
 - [22] DONOHO D L, MALEKI A, MONTANARI A. Message passing algorithms for compressed sensing: I. motivation and construction[C]//*2010 IEEE information theory workshop on information theory (ITW 2010, Cairo)*. IEEE, 2010: 1-5.
 - [23] MONTANARI A, ELDAR Y C, KUTYNIOK G. Graphical models concepts in compressed sensing[J]. 2012.
 - [24] BLUMENSATH T, DAVIES M E. Iterative thresholding for sparse approximations[J]. *Journal of Fourier analysis and Applications*, 2008, 14(5-6): 629-654.
 - [25] LI Y, OSHER S. Coordinate descent optimization for l_1 minimization with application to compressed sensing; a greedy algorithm[J]. *Inverse Problems & Imaging*, 2009, 3(3): 487.
 - [26] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
 - [27] BORGERDING M, SCHNITER P, RANGAN S. AMP-inspired deep networks for sparse linear inverse problems[J]. *IEEE Transactions on Signal Processing*, 2017, 65(16): 4293-4308.
 - [28] YANG M, LIU W, LUO W, et al. Analysis-synthesis dictionary learning for universality-particularity representation based classification[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2016, 30(1).
 - [29] RUBINSTEIN R, ELAD M. Dictionary learning for analysis-synthesis thresholding[J]. *IEEE Transactions on Signal Processing*, 2014, 62(22): 5962-5972.
 - [30] ABERDAM A, GOLTS A, ELAD M. Ada-lista: Learned solvers adaptive to varying models[J]. *arXiv preprint arXiv:2001.08456*, 2020.
 - [31] KAMILOV U S, MANSOUR H. Learning optimal nonlinearities for iterative thresholding algorithms[J]. *IEEE Signal Processing Letters*, 2016, 23(5): 747-751.
 - [32] ITO D, TAKABE S, WADAYAMA T. Trainable ISTA for sparse signal recovery[J]. *IEEE Transactions on Signal Processing*, 2019, 67(12): 3113-3125.
 - [33] MOREAU T, BRUNA J. Understanding trainable sparse coding via matrix factorization[J]. *arXiv preprint arXiv:1609.00285*, 2016.
 - [34] XIN B, WANG Y, GAO W, et al. Maximal sparsity with deep networks?[J]. *arXiv preprint arXiv:1605.01636*, 2016.
 - [35] BORGERDING M, SCHNITER P. Onsager-corrected deep learning for sparse linear inverse problems[C]//*2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2016: 227-231.
 - [36] CHEN X, LIU J, WANG Z, et al. Theoretical linear convergence of unfolded ISTA and its practical weights and thresholds[J]. *arXiv preprint arXiv:1808.10038*, 2018.
 - [37] LIU J, CHEN X. ALISTA: Analytic weights are as good as learned weights in LISTA[C]//*International Conference on Learning Representations (ICLR)*. 2019.
 - [38] ABLIN P, MOREAU T, MASSIAS M, et al. Learning step sizes for unfolded sparse coding[C]//*Advances in Neural Information Processing Systems*. 2019: 13100-13110.
 - [39] KIM D, PARK D. Element-Wise Adaptive Thresholds for Learned Iterative Shrinkage Thresholding Algorithms[J]. *IEEE Access*, 2020, 8: 45874-45886.
 - [40] ZHOU J T, DI K, DU J, et al. Sc2net: Sparse lstms for sparse coding[C]//*Proceedings of the AAAI Conference on*

- Artificial Intelligence. 2018, 32(1).
- [41] WU K, GUO Y, LI Z, et al. Sparse coding with gated learned ista[C]//International Conference on Learning Representations. 2019.
 - [42] GERS F A, SCHRAUDOLPH N N, SCHMIDHUBER J. Learning precise timing with LSTM recurrent networks[J]. *Journal of machine learning research*, 2002, 3(Aug): 115-143.
 - [43] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. *arXiv preprint arXiv:1406.1078*, 2014.
 - [44] COWEN B, SARIDENA A N, CHOROMANSKA A. LSALSA: accelerated source separation via learned sparse coding[J]. *Machine Learning*, 2019, 108(8): 1307-1327.
 - [45] DONOHO D L, MALEKI A, MONTANARI A. Message-passing algorithms for compressed sensing[J]. *Proceedings of the National Academy of Sciences*, 2009, 106(45): 18914-18919.
 - [46] METZLER C A, MOUSAVIA, BARANIUK R G. Learned D-AMP: Principled neural network based compressive image recovery[J]. *arXiv preprint arXiv:1704.06625*, 2017.
 - [47] WANG Z, LING Q, HUANG T. Learning deep ℓ_0 encoders[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2016, 30(1).
 - [48] HERSHEY J R, ROUX J L, WENINGER F. Deep unfolding: Model-based inspiration of novel deep architectures[J]. *arXiv preprint arXiv:1409.2574*, 2014.
 - [49] SPRECHMANN P, LITMAN R, BEN YAKAR T, et al. Supervised sparse analysis and synthesis operators[J]. *Advances in Neural Information Processing Systems*, 2013, 26: 908-916.
 - [50] SPRECHMANN P, BRONSTEIN A M, SAPIRO G. Learning efficient sparse and low rank models[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(9): 1821-1833.
 - [51] WANG Z, YANG Y, CHANG S, et al. Learning A Deep ℓ_∞ Encoder for Hashing[J]. *arXiv preprint arXiv:1604.01475*, 2016.
 - [52] GIRYES R, ELDAR Y C, BRONSTEIN A M, et al. Tradeoffs between convergence speed and reconstruction accuracy in inverse problems[J]. *IEEE Transactions on Signal Processing*, 2018, 66(7): 1676-1690.
 - [53] MARDANI M, SUN Q, PAPYAN V, et al. Degrees of freedom analysis of unrolled neural networks[J]. *arXiv preprint arXiv:1906.03742*, 2019.
 - [54] TAKABE S, WADAYAMA T. Theoretical interpretation of learned step size in deep-unfolded gradient descent[J]. *arXiv preprint arXiv:2001.05142*, 2020.
 - [55] TOLOOSHAMS B, DEY S, BA D. Scalable convolutional dictionary learning with constrained recurrent sparse auto-encoders[C]//2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2018: 1-6.
 - [56] SRETER H, GIRYES R. Learned convolutional sparse coding[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 2191-2195.
 - [57] PAPYAN V, ROMANO Y, ELAD M. Convolutional neural networks analyzed via convolutional sparse coding[J]. *The Journal of Machine Learning Research*, 2017, 18(1): 2887-2938.
 - [58] WOHLBERG B. Efficient convolutional sparse coding[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014: 7173-7177.
 - [59] ELAD M, AHARON M. Image denoising via sparse and redundant representations over learned dictionaries[J]. *IEEE Transactions on Image processing*, 2006, 15(12): 3736-3745.
 - [60] HEIDE F, HEIDRICH W, WETZSTEIN G. Fast and flexible convolutional sparse coding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5135-5143.
 - [61] BRISTOW H, ERIKSSON A, LUCEY S. Fast convolutional sparse coding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 391-398.
 - [62] ZHANG J, GHANEM B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1828-1837.
 - [63] WU S, DIMAKIS A, SANGHAVI S, et al. Learning a compressed sensing measurement matrix via gradient unrolling[C]//International Conference on Machine Learning. PMLR, 2019: 6828-6839.
 - [64] YANG Y, SUN J, LI H, et al. ADMM-CSNet: A deep learning approach for image compressive sensing[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2018, 42(3): 521-538.
 - [65] HOSSEINI S A H, YAMAN B, MOELLER S, et al. Dense recurrent neural networks for inverse problems: History-cognizant unrolling of optimization algorithms[J]. *Unknown Journal*, 2019.
 - [66] WU D, KIM K, DONG B, et al. End-to-end lung nodule detection in computed tomography[C]//International workshop on machine learning in medical imaging. Springer, Cham, 2018: 37-45.
 - [67] MONGA V, LI Y, ELDAR Y C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing[J]. *arXiv preprint arXiv:1912.10557*, 2019.
 - [68] LUSTIG M, DONOHO D, PAULY J M. Sparse MRI: The application of compressed sensing for rapid MR imaging[J]. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 2007, 58(6): 1182-1195.

- [69] SAKAI T, SATOMOTO H, KIYASU S, et al. Sparse representation-based extraction of pulmonary sound components from low-quality auscultation signals[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2012: 509-512.