1 稳健的矩阵回归模型与方法*

杨 健 罗 雷

南京理工大学模式计算与应用实验室，南京 210094

1 引言

随着网络和信息技术的不断发展，全球数据量呈现爆炸式增长。特别地，人类在最近两年产生的数据量相当于之前产生的全部数据量。这些海量、复杂的数据已经对社会经济、政治、文化以及生活等方面产生了深远的影响。如何有效地分析它们使其更好地为人类服务是一个亟待解决的问题。

近年来，机器学习方法已经渗透到了数据分析的各个方面，成为了其中的重要组成部分。它利用已知数据来学习和推理其中未知的、潜在的概率分布等重要信息，解释数据样本中变量(或特征)之间的关系。也就是说，它能从庞大的数据中揭示出数据的内在规律或本质结构。这种本质结构可方便人类对数据的理解，提高人类对数据的驾驭能力。为实现这一任务，我们首先要熟知数据分析中可能面临的问题。一方面，在实际的信号和图像采集与处理中，数据的维数越高，给数据的采集和处理带来越多的限制。大规模的数据不仅存在着较多的冗余、无关的属性，还会带来维数的灾难。另一方面，实际中数据往往结构复杂，例如采集的数据本身不完整或者存在大量的噪声，这导致了一些传统的数据处理技巧也许失效。

基于线性回归(linear regression, LR)分析的方法一直是机器学习领域的研究热点。它的目的是估计测试数据与训练数据之间的关系。值得注意的是，这种关系可理解成数据的一种内在结构。为避免过拟合，一个正则项需要强加到 LR 模型中。基于 L_1 范数的正则项和基于 L_2 范数的正则项是目前使用最广泛的两种正则项。L_2 范数正则化的线性回归一般称为脊回归(ridge regression)。该方法使用 L_2 范数的平方来约束表示系数，并使用同样的方式来刻画表示残差。最近，I. Naseem 等人[1]将此方法应用于人脸识别并由

* 本文得到国家自然科学基金(91420201 和 61472187)资助。

此提出了线性回归分类器(LRC)。在这个方法中,他们先通过脊回归模型获得测试样本在每类训练样本中的表示系数,然后将测试样本分在离其最近的那一类。事实上,一些以往的工作,诸如最近邻特征线[2]、最近邻特征面以及最近邻特征空间[3]都可看作 LRC 的不同变体。

L_1范数正则化的线性回归称为 Lasso,它已被广泛地应用于稀疏表示中。例如,J. Wright 等人[4]引入了稀疏表示分类器 (SRC)。SRC 使用所有训练样本作为一个字典来表示一个测试样本, 并且假设表示系数是稀疏的,即这些稀疏的非零表示系数应该集中在与测试样本具有相同类型标签的训练样本上。为了得到一个稳健的模型,他们进一步假设噪声也是稀疏的,由此建立了推广的 SRC 模型。这个模型对随机像素损坏与块状遮挡的噪声具有一定的稳健性。A. Wagner 等人[5]进一步推广了 SRC 模型,并将人脸对齐和识别统一到一个框架中。另一方面,一些近期的工作主要集中在探究稀疏在图像分类中的角色[6~9]上。J. Yang 等人[8]对 SRC 的合理性提供了一些理论性的解释并得出一个结论:SRC 的有效性是由 L_1约束带来的,而不是 L_0约束 (原本的稀疏限制) 的作用。L. Zhang 等人[9]分析了 SRC 的运行原理且发现:合作表示策略比基于 L_1范数的系数约束扮演了更为重要的角色。于是,基于脊回归的思想,他们提出了合作表示分类器(collaborative representation classifier, CRC)。可是,CRC 虽然具有速度上的优势,但并不能提供一种噪声移除的机制, 所以它不是图像分类的一种稳健的回归方法。

在 LRC、CRC 以及 SRC 模型中,表示余差使用误差向量的 L_2范数或者 L_1范数来约束。从统计学来讲,L_2范数为服从高斯分布的数据提供了一种最优的刻画。但是,在现实的人脸识别中,表示误差的分布是很复杂的,高斯分布无法完美地拟合它们[10,11]。所以,以上方法对现实中的噪声并不稳健。为解决这类问题,稳健回归分析[12]应运而生。该方法是对传统的最小均方法在异常点存在情况下的一种改进,它提供给我们这样一个信息:什么是一个有效的观测,并且哪一种观测应被剔除。稳健回归的最初目的是用来拟合一个能够表示出大多数数据的模型。

M-估计是稳健回归中一种常用的技巧,在 19 世纪 60 年代由 P. J. Huber 等人[13]提出。选择权重函数对稳健回归的性能具有重要影响,由此衍生了许多稳健回归模型,例如,推广的稀疏表示模型[4]、稳健稀疏编码模型(RSC)[10]、基于相关熵的稀疏表示模型(CESR)[14]以及稳健的线性回归分类器 (RLRC)[15]。特别地,RLRC 本质上是一种 Huber 估计方法。此外,R. He 等人[16]统一了两种稀疏稳健回归模型,即在误差纠正中以 SRC 为代表的相加模型和在误差探测中以 CESR 为代表的相乘模型。R. He 等人[16]还通过定义不同的半平方函数建立了一种半平方的人脸识别框架。该框架可同时执行误差纠正与探测任务。M-估计的稳健性体现在: 异常像素值会被赋予一个小的权值来减小它们在编码过程中的影响,而对其他元素则赋予较大的权值。所以,与传统的最小均方法相比,稳健回归更适合处理实际中的噪声。但是,以上所提到的稳健回归方法都使用了

一维的基于像素误差模型,即误差图像的每个像素都被独立地刻画。它们并没有考虑图像的二维结构信息。

不同于以上方法,本文将介绍几种全新的稳健回归模型。它们将图像的结构信息融入到建模中。这些方法包括:基于核范数的稳健矩阵回归,基于推广幂指数分布的稳健矩阵回归,基于核-L_1范数的联合矩阵回归,以及基于树结构核范数的稳健矩阵回归。

2 基于核范数的稳健矩阵回归

基于像素的一维模型(例如,SRC, RSC 和 CESR)存在两个问题:首先,这类模型假设误差图像的每一个像素是独立同分布的。对于随机的像素损坏,由于噪声被独立地添加在每个像素上,所以这个假设是合理的。但是,对于许多实际的人脸变化,诸如遮挡、伪装或者由光照改变产生的阴影,这个假设不再成立。例如,由黑色围巾造成的遮挡,其中的像素值是 0,那么,遮挡部分中理想的表示误差像素是相关的,因为现实图像中局部区域的像素一般是高度相关的。所以,使用基于像素的一维误差模型(诸如 SRC[4],RSC[11],RLRC[15]等)来解决带有遮挡的图像分类任务在理论上是令人质疑的。

其次,单独地刻画误差的每个像素忽视了误差图像的整个结构,因为误差图像中的所有像素误差也许包含了有意义的结构信息(例如误差图像的秩)。在基于回归分析的人脸识别方法中,可以使用训练图像来表示一张测试图像。理想情况下,误差图像应该是一个零矩阵。因此,它自然是低秩的。在一般情形下,测试图像中存在光照改变和遮挡。事实上,光照和遮挡是影响人脸识别性能的关键因素。实际的光照改变,特别是诸如阴影的部分光照改变一般会导致一个低秩(或者近似低秩)的误差图像,这与满秩的原始图像形成了鲜明的对比。诸如眼镜与围巾遮挡也会导致一个低秩的误差图像。以上所提及的回归方法可独立地处理每一个误差像素,因此无法反映这类结构信息。

为了充分利用误差图像的低秩结构信息,这里介绍我们所提出的一种基于二维图像矩阵的误差模型。相反地,以前的方法,诸如脊回归、Lasso 或者稳健回归,都是基于向量的途径。也就是说,为了处理矩阵形式的二维图像,我们在使用诸如回归的方法时,必须提前将图像转化成向量。在转化阶段,一些结构信息(例如,误差图像的秩)也许会丢失。本节的矩阵回归方法不需要矩阵到向量的转化过程。它通过最小化余差图像的秩来挖掘图像的结构信息,并以此来决定回归系数。为便于优化,秩函数最小化问题一般可转化成核范数最小化问题。根据这一想法,表示余差图像的最小核范数可作为一个准则。因此,此方法也称为基于核范数的矩阵回归(nuclear norm based matrix regression,NMR),其具体形式如下:

给定 n 个图像矩阵 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n\in R^{p\times q}$ 和一个图像矩阵 $\boldsymbol{B}\in R^{p\times q}$,则 $\boldsymbol{B}$ 可被 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n$ 线性地表示,即

$$\boldsymbol{B} = x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \cdots + x_n\boldsymbol{A}_n + \boldsymbol{E} \tag{1}$$

其中，$x_1, x_2, \cdots, x_n$ 表示稀疏的集合，$\boldsymbol{E}$ 表示余差。

定义从 R^n 到 $R^{p\times q}$ 的映射如下：

$$A(\boldsymbol{x}) = x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \cdots + x_n\boldsymbol{A}_n \tag{2}$$

那么，公式(1)变为

$$\boldsymbol{B} = A(\boldsymbol{x}) + \boldsymbol{E} \tag{3}$$

公式(3)或者公式(1)给出了线性矩阵回归模型的一般形式，它们与传统的线性向量模型形成了鲜明的对比。

在许多应用中，余差图像 $\boldsymbol{B}-A(\boldsymbol{x})$ 在最优解处是典型低秩的(或近似低秩的)。考虑到核范数是秩函数在单位球上的凸包，本节将解决以下核范数最小化问题来评估回归系数：

$$\min_{\boldsymbol{x}} \| \boldsymbol{B} - A(\boldsymbol{x}) \|_* \tag{4}$$

其中，$\| \boldsymbol{X} \|_*$ 表示矩阵 $\boldsymbol{X}$ 的核范数，即 $\boldsymbol{X}$ 的所有非零奇异值之和。此外，我们可在公式(4)上增加一个类似脊回归的正则项，由此获得了正则化的矩阵回归模型：

$$\begin{gathered}\min_{\boldsymbol{x},\boldsymbol{E}} \| \boldsymbol{E} \|_* + \frac{1}{2}\lambda \| \boldsymbol{x} \|_2^2 \\ \text{s. t.} \quad \boldsymbol{B} = A(\boldsymbol{x}) + \boldsymbol{E}\end{gathered} \tag{5}$$

其中，$\lambda>0$ 是一个平衡参数。公式(5)可使用交替方向乘子法(ADMM)来解决，其主要迭代过程被总结在算法 1 中。这里，$D_\zeta(\boldsymbol{Q})$ $(\zeta>0)$ 表示矩阵 $\boldsymbol{Q}(\in R^{p\times q})$ 的奇异值收缩算子(详细定义见文献[17])，$\mathrm{Vec}(\cdot)$ 是一个将矩阵转化为向量的操作符，$\boldsymbol{I}$ 是一个单位矩阵。此外，我们可以使用文献 [18] 中类似的途径来设置算法 1 的终止条件。

算法 1 使用 ADMM 解决 NMR

输入：一组图像矩阵 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n$ 和一个图像矩阵 $\boldsymbol{B} \in R^{p\times q}$，模型参数 λ 以及 μ。

1. 令 $\boldsymbol{A} = [\mathrm{Vec}(\boldsymbol{A}_1), \cdots, \mathrm{Vec}(\boldsymbol{A}_n)]$。计算 $\boldsymbol{M} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \dfrac{\lambda}{\mu}\boldsymbol{I}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}$；
2. $\boldsymbol{E}^0 = -\boldsymbol{B}$, $\boldsymbol{Z}^k = \boldsymbol{0}$, $k=0$；
3. 更新 $\boldsymbol{x}$：$\boldsymbol{x}^{k+1} = \boldsymbol{M}\boldsymbol{g}$，其中 $\boldsymbol{g} = \mathrm{Vec}\left(\boldsymbol{B} - \boldsymbol{E}^k - \dfrac{1}{\mu}\boldsymbol{Z}^k\right)$
4. 更新 $\boldsymbol{E}$：$\boldsymbol{E}^{k+1} = D_{1/\mu}\left(\boldsymbol{B} - A(\boldsymbol{x}^{k+1}) - \dfrac{1}{\mu}\boldsymbol{Z}^k\right)$；
5. 更新 $\boldsymbol{Z}$：$\boldsymbol{Z}^{k+1} = \boldsymbol{Z}^k + \mu(A(\boldsymbol{x}^{k+1}) + \boldsymbol{E}^{k+1} - \boldsymbol{B})$；

输出：最优回归系数向量 $\boldsymbol{x}^{k+1}$。

但是,传统的 ADMM 方法仅能获得一个 $O(1/k)$ 的收敛率[19],其中 k 代表迭代次数。所以,我们将使用一种加速的 ADMM[20] 来获得模型(5)的最优解,此算法能达到平方收敛。考虑到加速的(Fast)ADMM 的收敛条件要求目标函数是强凸的,我们在模型(5)上添加一个平方项,从而得到一个近似的 NMR 模型如下:

$$\begin{aligned} &\min_{\boldsymbol{E},\boldsymbol{x}} \|\boldsymbol{E}\|_* + \gamma \|\boldsymbol{E}\|_F^2 + \frac{1}{2}\theta \|\boldsymbol{x}\|_2^2 \\ &\text{s.t.} \quad \boldsymbol{B} = A(\boldsymbol{x}) + \boldsymbol{E} \end{aligned} \tag{6}$$

其中,$\gamma,\theta>0$。J. Yang 等人[17]已经说明,在一定条件下,模型(5)与模型(6)的解是一致的。算法 2 列出了解决公式(6)的详细过程,其终止条件类似于算法 1。

算法 2　使用 Fast ADMM 解决 NMR

输入:一列图像矩阵 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n$ 和一张图像矩阵 $\boldsymbol{B}\in R^{p\times q}$,模型参数 λ,μ 和 γ。

1. 令 $A=[\mathrm{Vec}(\boldsymbol{A}_1),\cdots,\mathrm{Vec}(\boldsymbol{A}_n)]$。计算 $\boldsymbol{M}=\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\frac{\theta}{\mu}\boldsymbol{I}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}$,其中 $\theta=\lambda(1+\gamma)$;
2. $\boldsymbol{x}^0=\hat{\boldsymbol{x}}^0=\boldsymbol{0},\boldsymbol{Z}^0=\hat{\boldsymbol{Z}}^0=\boldsymbol{0},\alpha^0=1,\ k=0$;
3. 更新 $\boldsymbol{E}$:$\boldsymbol{E}^{k+1}=\frac{\mu}{\mu+2\gamma}D_{1/\mu}\left(\boldsymbol{B}-A(\hat{\boldsymbol{x}}^k)-\frac{1}{\mu}\hat{\boldsymbol{Z}}^k\right)$;
4. 更新 $\boldsymbol{x}$:$\boldsymbol{x}^{k+1}=\boldsymbol{M}\boldsymbol{g}$,其中 $\boldsymbol{g}=\mathrm{Vec}\left(\boldsymbol{B}-\boldsymbol{E}^{k+1}-\frac{1}{\mu}\hat{\boldsymbol{Z}}^k\right)$
5. 更新 $\boldsymbol{Z}$:$\boldsymbol{Z}^{k+1}=\hat{\boldsymbol{Z}}^k+\mu(A(\boldsymbol{x}^{k+1})+\boldsymbol{E}^{k+1}-\boldsymbol{B})$;
6. 更新 α:$\alpha^{k+1}=\frac{1+\sqrt{1+4\,(\alpha^k)^2}}{2}$;
7. 更新 $\hat{\boldsymbol{x}}$:$\hat{\boldsymbol{x}}^{k+1}=\boldsymbol{x}^{k+1}+\frac{\alpha^k-1}{\alpha^{k+1}}(\boldsymbol{x}^{k+1}-\boldsymbol{x}^k)$;
8. 更新 $\hat{\boldsymbol{Z}}$:$\hat{\boldsymbol{Z}}^{k+1}=\boldsymbol{Z}^{k+1}+\frac{\alpha^k-1}{\alpha^{k+1}}(\boldsymbol{Z}^{k+1}-\boldsymbol{Z}^k)$;

输出:最优的回归系数向量 $\hat{\boldsymbol{x}}^{k+1}$。

在 NMR 的建模动机中,J. Yang 等人[17]使用了“低秩”的概念,这是源于对由遮挡造成的误差图像的一个直观的描述。实际上,NMR 并不局限于“低秩”的误差图像,因为它不直接优化秩函数而是优化误差的核范数。比起秩函数,核范数能对误差图像提供一种更加可行的刻画,因为它在刻画近似低秩的误差图像的时候仍然是有效的。这些误差图像也许是代数满秩的,但它的奇异值接近于 0。例如,在光照改变的情形下,假如人脸是几何光滑的,那么误差图像中的像素则是高度相关的。所以,由光照改变造成的误差图像一般是近似低秩的。

我们知道一个矩阵的核范数是它的所有奇异值的和，也就是奇异值向量的 L_1 范数。从概率分布的角度来讲，L_1 范数对满足拉普拉斯（Laplacian）分布的随机向量提供了一种最优的刻画，而 L_2 范数对高斯（Gaussian）分布是最优的[10,21]。所以，如果一个误差图像的所有奇异值满足 Laplacian 分布，那么核范数对误差图像提供了一种最优的刻画方式。幸运的是，J. Yang 等人[17]在人脸图像上证实了这一现象确实存在。在分类器的设计中，NMR 用核范数来度量测试图像与重构图像之间的误差。一系列的实验结果证明了 NMR 对结构性的噪声具备一定的鲁棒性。

3 基于推广幂指数分布的稳健矩阵回归

前一节的分析表明了，基于核范数的矩阵回归可有效地反映噪声图像的二维结构，特别是低秩结构。我们知道，这种低秩结构仅是二维空间结构的一种特殊情形。而在实际应用中，图像水平噪声中像素之间的依赖关系则更普遍。因此，本节将从这类噪声分布的本质出发，寻求一种更一般的稳健回归方法。

实际上，以上所涉及的回归方法都属于多变量分析的范畴。众所周知，多变量分析所考察的对象主要是独立正态分布的数据。但是，在实际应用中，当数据的基本分布具有长尾属性的时候，误差矩阵的正态性和独立性假设也许是不适合的。于是，B. M. G. Kibria 和 M. S. Haq[21]提出了线性的多元 t-errors 模型来拟合实际的数据。S. Basu 等人[22]在语音识别领域使用多元幂指数分布作为一种长尾的分布。通过假设观测值是依赖的，M. H. Liu[23]将向量的幂指数回归模型推广到矩阵变量的形式。这些研究表明：为了描述实际应用中的观测数据，依赖且长尾的分布假设是十分重要的。

对于许多实际中的人脸变化，诸如遮挡、伪装或者由光照改变所带来的阴影，它们所产生的误差图像的像素之间一般是高度相关的（文献[30]中举出了一个例子）。由于独立性假设不再成立，RSC[11]和 CESR[14]无法显示出优势。为解决这一问题，X. X. Li 等人[24]探索了连续性遮挡的本质结构并提出了一种结构的稀疏误差编码模型。K. Jia 等人[25]将一类结构性稀疏诱导的范数引入到 SRC 框架中来拟合一些结构性的噪声。Y. Deng 等人[26]通过先发现遮挡的最佳匹配块，然后使用拉普拉斯方法来完成人脸识别。W. Deng 等人[27]使用一个辅助的类内变化字典来表示训练图像和测试图像之间的变化，并将这一方法应用在每类中含有很少的甚至一个训练图像的情形。此外，一些学者从字典学习的角度来研究结构性的噪声。例如，M. Yang 等人[28]通过预先学习的余差映像来探测人脸遮挡的像素，然后使用不包含遮挡像素位置的新的字典来完成分类任务。W. Ou 等人[29]提出的途径与文献[28]类似，不同之处在于他们同时学习了一个干净的字典和一个含有噪声的字典，然后将这个干净的字典应用于分类任务中。

不同于以往的方法，本节我们所提出的 $S_P L_Q$[30] 使用了一种依赖的矩阵分布来刻画由遮挡、伪装或者光照改变所产生的结构性噪声。这种依赖的矩阵分布可写成：

$$P(\boldsymbol{B}|\boldsymbol{x})=C\exp\left(-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{B}-A(\boldsymbol{x}))^{\mathrm{T}}(\boldsymbol{B}-A(\boldsymbol{x}))\right)^{p/2}\right) \tag{7}$$

其中，$C>0$ 是一个正比例常数，$p>0$ 是给定的参数。在文献[30]中，Luo 等人证实了分布(7)能够刻画噪声像素之间的依赖关系。因此，误差矩阵的空间结构信息得以充分保留。此外，该方法假设表示系数是独立的且服从相同的高斯分布($q=2$)或者拉普拉斯分布($q=1$)。通过最大后验概率估计，并引入一个辅助变量 $\boldsymbol{E}$，我们最终可获得如下模型：

$$\min_{\boldsymbol{x},\boldsymbol{E}} \|\boldsymbol{E}\|_{S_p}^p+\frac{1}{2}\lambda\|\boldsymbol{x}\|_q^q \tag{8}$$

其中，$\|\cdot\|_{S_p}^p$ 代表一个矩阵的 shatten-p 范数。这里，我们主要讨论三种情形：$p=1/2$，$2/3$ 和 1。解决模型(8)的主要迭代过程被总结在算法 3 中，其中步骤 2 中的具体形式可参考文献[30]。

算法 3 使用 ADMM 解决 $\boldsymbol{S}_{\mathrm{p}}\boldsymbol{L}_{\mathrm{q}}$

输入：二维字典 $\boldsymbol{A}=\{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n\}$，其中每一个 $\boldsymbol{A}_i\in R^{l\times m}$，测试图像 $\boldsymbol{B}\in R^{l\times m}$，以及模型参数 $\lambda,\mu,\tau>0$。

初始化：$\boldsymbol{Y}=\boldsymbol{Z}=\boldsymbol{0}$.

1. 如果 $q=2$，固定其他变量，更新 $\boldsymbol{x}$：

$$\boldsymbol{x}^{k+1}=\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+\frac{\lambda}{\mu}\boldsymbol{I}_n\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}^k,$$

如果 $q=1$，固定其他变量，更新 $\boldsymbol{x}$：

$$\boldsymbol{x}^{k+1}=\mathrm{sgn}(\boldsymbol{g}^k)\circ\max\left\{|\boldsymbol{g}^k|-\frac{\lambda}{2\mu\tau},0\right\};$$

其中，$\boldsymbol{b}^k=\mathrm{Vec}\left(\boldsymbol{B}-\boldsymbol{E}^k-\frac{1}{\mu}\boldsymbol{Z}^k\right)$，$\boldsymbol{g}^k=\boldsymbol{x}^k-\frac{1}{\tau}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}^k-\boldsymbol{b}^k)$.

2. 固定其他变量，更新 $\boldsymbol{E}$：

$$\boldsymbol{E}^{k+1}=D_{\frac{1}{\mu}}^{p}\left(\boldsymbol{B}-A(\boldsymbol{x}^{k+1})-\frac{1}{\mu}\boldsymbol{Z}^k\right);$$

3. 固定其他变量，更新乘子 $\boldsymbol{Z}$：

$$\boldsymbol{Z}^{k+1}=\boldsymbol{Z}^k+\mu(A(\boldsymbol{x}^{k+1})+\boldsymbol{E}^{k+1}-\boldsymbol{B}).$$

输出：最优的回归系数 $\boldsymbol{x}^{k+1}$.

可以证明，算法 3 在 $p=1$ 处全局收敛，在 $p<1$ 时局部收敛。一系列实验结果证实了算

法 3 能够有效地处理带有结构性噪声的图像分类问题。此外，与 $p=1$ 相比，它在 $p<1$ 时能够取得更好的效果。

4 基于核-L_1范数的联合矩阵回归

前两节已经强调了噪声的结构信息对鲁棒回归建模的重要性。为了有效地处理结构性的噪声，一些学者在建模中融入了一些额外的技巧。例如，Morelli Andrés 等人[31]使用压缩感知提出了一种迭代的遮挡探测算法。X. X. Li 等人[32]引入了部分迭代重加权稀疏编码，它能够精确地探测出与整个训练集相关的遮挡区域。R. Min 等人[33]先实施了显式的遮挡分析，然后从非遮挡区域完成人脸识别任务。J. X. Mi 等人[34]先将一张图像分成几块，其中，遮挡块由一个指示器决定，然后样本非遮挡块用作分类的特征依据。这些方法的共同特点是需提前探测出遮挡的位置。但是，对于许多实际的噪声，准确地探测出它们在图像中的位置是一项艰巨的任务。

不同于以上途径，基于回归的方法直接刻画噪声而不需要考虑噪声所处的位置，例如，SRC[4]使用 L_1 范数来描述稀疏噪声，NMR[17]借助核范数来拟合结构性噪声。可是，这些方法仅适合处理单一结构的噪声。我们知道，实际的噪声结构是复杂的，不仅具备一些稀疏属性，还可能隐藏一些结构信息。我们所提出的 NL_1R（nuclear-L_1 norm joint matrix regression）[35]结合了 SRC 和 NMR 的各自优势，采用核-L_1 范数来刻画一类混合噪声，即稀疏噪声加上结构性噪声。

近年来，关于混合噪声的研究一直是图像处理中的热点问题。例如，Y. Xiao 等人[36]将中值类型的滤波器与字典学习相结合来恢复被高斯和脉冲混合噪声损坏的图像。J. Liu 等人[37]提出了一种加权的字典学习途径来实现混合噪声移除任务。这种方法将稀疏字典学习、图像重构、噪声聚类以及参数估计融入到一个框架中。J. L. Jiang 等人[38]采纳一种加权的编码技巧来同时移除高斯和脉冲噪声。随后，J. L. Jiang 等人[39]借助加权的低秩模型提出一种全新的混合噪声移除方法。这个方法同时保留了图像的全局结构和局部边界信息。但是，这些混合噪声移除模型仅仅适合点状噪声，他们仍然需借助传统的高斯噪声移除方法来完成降噪任务。此外，它们在单一的图像上实施降噪任务，并没考虑其他信息。因此，这些方法不适合模式表示（或分类）任务。

幸运的是，NL_1R[35]能够同时处理混合噪声的移除和模式分类任务。它从两个不同的角度来考虑混合噪声。一方面，它将混合噪声看作一个整体，并使用一种推广的矩阵变量分布来描述这类噪声。这个分布是推广的矩阵 Slash 分布[35]与拉普拉斯分布的线性组合，即

$$P(\boldsymbol{E};C',\boldsymbol{M})=C'\exp(-\operatorname{tr}(\boldsymbol{E})^{\mathrm{T}}(\boldsymbol{E})-\nu\|\boldsymbol{E}\|_1) \tag{9}$$

这里ν是一超参数,C'是正比例常数。借助最大后验概率估计并引入辅助变量$\boldsymbol{E}$和$\boldsymbol{Z}$,我们可获得以下模型:

$$\min_{\boldsymbol{E},\boldsymbol{Z},\boldsymbol{x}} \|\boldsymbol{E}\|_* + \alpha_1 \|\boldsymbol{Z}\|_1 + \frac{1}{2}\beta_1((2-\tau)\|\boldsymbol{z}\|_\tau^\tau + (\tau-1)\|\boldsymbol{x}\|_\tau^\tau)$$
$$\text{s.t.} \quad A_1(\boldsymbol{x}) + B_1(\boldsymbol{z}) + C_1(\boldsymbol{E}) + D_1(\boldsymbol{Z}) = E_1 \tag{10}$$

其中,$A_1(\boldsymbol{x}) = (\mathrm{Vec}(A(\boldsymbol{x}));A\boldsymbol{0};(2-\tau)\boldsymbol{x})$, $B_1(\boldsymbol{z}) = (\boldsymbol{0};\boldsymbol{0};-(2-\tau)\boldsymbol{z})$, $C_1(\boldsymbol{E}) = (\mathrm{Vec}(\boldsymbol{E});\mathrm{Vec}(\boldsymbol{E});\boldsymbol{0})$, $D_1(\boldsymbol{Z}) = (\boldsymbol{0};-\mathrm{Vec}(\boldsymbol{Z});\boldsymbol{0})$, $E_1 = (\mathrm{Vec}(\boldsymbol{B});\boldsymbol{0};\boldsymbol{0})$,且$\tau \in \{1,2\}$, $\alpha_1,\beta_1 > 0$。

模型(10)被称为同时式的核-L_1范数联合回归模型(simultaneous nuclear-L_1 norm joint matrix regression model,$\mathrm{SNL_1R}$)。算法4总结了解决$\mathrm{SNL_1R}$的详细过程。

算法4 使用ADMM解决$\mathrm{SNL_1R}$

输入: 一组训练图像$\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n$,其中$\boldsymbol{A}_i \in R^{l\times m}$,$i=1,2,\cdots,n$,一张测试图像$\boldsymbol{B} \in R^{l\times m}$,模型参数$\alpha_1$, β_1和$\mu=1$。

初始化: $\boldsymbol{Z}=\boldsymbol{0}$, $\boldsymbol{y}=\boldsymbol{0}$, $\boldsymbol{z}=\boldsymbol{x}=\boldsymbol{0}$;

1. 固定其余变量,更新$\boldsymbol{E}$:

$$\boldsymbol{E} = \frac{1}{2}D_{1/\mu}\left(\boldsymbol{B} - A(\boldsymbol{x}) - \frac{1}{\mu}\boldsymbol{Y}_1 + \boldsymbol{Z} - \frac{1}{\mu}\boldsymbol{Y}_2\right);$$

2. 固定其余变量,更新$\boldsymbol{Z}$:

$$\boldsymbol{Z} = \mathrm{sgn}(\boldsymbol{E} + \frac{1}{\mu}\boldsymbol{Y}_2) \circ \max\left\{\left|\boldsymbol{E} + \frac{1}{\mu}\boldsymbol{Y}_2\right| - \frac{\alpha_1}{\mu}, 0\right\},$$

3. 固定其余变量,更新$\boldsymbol{x}$:

$$\boldsymbol{x} = (\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \beta_1\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\mathrm{Vec}(\mu\boldsymbol{B} - \mu\boldsymbol{E} - \boldsymbol{Y}_1)$$

其中,$\boldsymbol{I}$表示单位矩阵,$\boldsymbol{A}=[\mathrm{Vec}(\boldsymbol{A}_1),\mathrm{Vec}(\boldsymbol{A}_2),\cdots,\mathrm{Vec}(\boldsymbol{A}_n)]$。

4. 固定其余变量,更新$\boldsymbol{z}$:

$$\boldsymbol{z} = \mathrm{sgn}(\boldsymbol{x} + \frac{1}{\mu}\boldsymbol{y}_3) \circ \max\left\{\left|\boldsymbol{x} + \frac{1}{\mu}\boldsymbol{y}_3\right| - \frac{\beta_1}{2\mu}, 0\right\}$$

5. 更新乘子$\boldsymbol{y}$:

$$\boldsymbol{y} = \boldsymbol{y} + \mu(A_1(\boldsymbol{x}) + B_1(\boldsymbol{z}) + C_1(\boldsymbol{E}) + D_1(\boldsymbol{Z}) - E_1)。$$

输出: 回归系数$\boldsymbol{x}$,$\boldsymbol{E}$和$\boldsymbol{Z}$。

另一方面,$\mathrm{NL_1R}$假设这类混合噪声是两种独立成分的叠加,即,结构性噪声和稀疏噪声,并分别使用推广的矩阵Slash分布和独立的拉普拉斯分布来刻画每个成分。类似于$\mathrm{SNL_1R}$的推导方式,我们可以获得以下模型:

$$\min_{\boldsymbol{E}_1,\boldsymbol{E}_2,\boldsymbol{x},\boldsymbol{z}} \|\boldsymbol{E}_1\|_* + \alpha_2 \|\boldsymbol{E}_2\|_1 + \frac{1}{2}\beta_2((2-\tau)\|\boldsymbol{z}\|_\tau^\tau + (\tau-1)\|\boldsymbol{x}\|_\tau^\tau)$$
$$\text{s.t.}\quad A_2(\boldsymbol{x}) + B_2(\boldsymbol{z}) + C_2(\boldsymbol{E}_1) + D_2(\boldsymbol{E}_2) = E_2 \tag{11}$$

其中，$A_2(\boldsymbol{x})=(\mathrm{Vec}(D(\boldsymbol{x}));(2-\tau)\boldsymbol{x})$，$B_2(\boldsymbol{z})=(\boldsymbol{0};-(2-\tau)\boldsymbol{z})$，$C_2(\boldsymbol{E}_1)=(\mathrm{Vec}(\boldsymbol{E}_1);\boldsymbol{0})$，$D_2(\boldsymbol{E}_1)=(\mathrm{Vec}(\boldsymbol{E}_2);\boldsymbol{0})$，$E=(\mathrm{Vec}(\boldsymbol{B});\boldsymbol{0})$且$\tau\in\{1,2\}$，$\alpha_1$，$\beta_1>0$。

因为误差矩阵$\boldsymbol{E}$被分解成两部分，我们称式(11)为分解式的核-L_1范数联合矩阵回归模型(decomposed nuclear-L_1 norm joint matrix regression model，DNL$_1$R)。解决此模型的算法被总结在算法5中。

算法 5　使用 ADMM 解决 SNL$_1$R

输入：一组训练图像$\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n$，其中$\boldsymbol{A}_i\in R^{l\times m}$，$i=1,2,\cdots,n$，一张测试图像$\boldsymbol{B}\in R^{l\times m}$，模型参数$\alpha_2$，$\beta_2$以及$\mu=1$。

初始化：$\boldsymbol{E}_2=\boldsymbol{0}$，$\boldsymbol{y}=0$，$\boldsymbol{x}=\boldsymbol{z}=\boldsymbol{0}$；

1. 固定其余变量，更新$\boldsymbol{E}_1$：

$$\boldsymbol{E}_1 = D_{1/\mu}\left(\boldsymbol{B} - A(\boldsymbol{x}) - \boldsymbol{E}_2 - \frac{1}{\mu}\boldsymbol{Y}_1\right)。$$

2. 固定其余变量，更新$\boldsymbol{E}_2$：

$$\boldsymbol{E}_2 = \mathrm{sgn}(\boldsymbol{B} - A(\boldsymbol{x}) - \boldsymbol{E}_1 - \frac{1}{\mu}\boldsymbol{Y}_1) \circ \max\left\{\left|\boldsymbol{B} - A(\boldsymbol{x}) - \boldsymbol{E}_1 - \frac{1}{\mu}\boldsymbol{Y}_1\right| - \frac{\alpha_2}{\mu}, 0\right\}$$

3. 固定其余变量，更新$\boldsymbol{x}$：

$$\boldsymbol{x} = \begin{cases} (\mu\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \beta_2\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\mathrm{Vec}(\mu\boldsymbol{B} - \mu\boldsymbol{E}_1 - \mu\boldsymbol{E}_2 - \boldsymbol{Y}_1), \text{ if } \tau = 2, \\ (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \boldsymbol{I})^{-1}\left(\boldsymbol{A}^{\mathrm{T}}\mathrm{Vec}\left(\boldsymbol{B} - \boldsymbol{E}_1 - \boldsymbol{E}_2 - \frac{1}{\mu}\boldsymbol{Y}_1\right) + \left(\boldsymbol{z} - \frac{1}{\mu}\boldsymbol{y}_2\right)\right), \text{ if } \tau = 1. \end{cases}$$

4. 固定其余变量，更新$\boldsymbol{z}$：

$$\boldsymbol{z} = \mathrm{sgn}(\boldsymbol{x} + \frac{1}{\mu}\boldsymbol{y}_2) \circ \max\left\{\left|\boldsymbol{x} + \frac{1}{\mu}\boldsymbol{y}_2\right| - \frac{\beta_2}{2\mu}, 0\right\}$$

5. 更新乘子：

$$\boldsymbol{y} = \boldsymbol{y} + \mu(A_2(\boldsymbol{x}) + B_2(\boldsymbol{z}) + C_2(\boldsymbol{E}_1) + D_2(\boldsymbol{E}_2) - E_2).$$

输出：最优的回归系数$\boldsymbol{x}$，$\boldsymbol{E}_1$和$\boldsymbol{E}_2$。

一般情况下，算法4和算法5拥有比SRC或者RSC更低的复杂度。我们在AR[41]、Extended Yale B[42]以及Multi-PIE[43]数据库上一系列人脸识别实验证明：与其他方法相比，NL$_1$R可处理更多类型的噪声且性能更优越。并且我们发现，DNL$_1$R的性能比SNL$_1$R更加稳定。