指导教师：杨涛　　　　提交时间：2016/3/17

# CVPR2015 Paper Translation

No: 01

姓名：孙志文

学号：2013302506

班号：10011302

# 鲁棒子空间学习的奇异值的弹性网络正则化

**摘要**：对低维结构的研究在计算机视觉中扮演着重要的角色。进来，一系列新的方法，例如 L1 范数最小化和鲁棒主生成分析，已经被用来解决低阶矩阵逼近问题并且表现出对异常值和缺失数据的鲁棒性。但是这些方法要求巨大的计算量，而且当数据损失时可能会找不出解决方法。在这篇论文中，一种为了子空间学习而采取的基于低阶矩阵分解法的弹性网络正则化被提及。该方法找到了一种有效的鲁棒算法，它用强突约束来增强算法的稳定性和维护算法的低等级属性。它表明，该算法中的任何固定点都满足库恩塔克最优性条件。该方法已经应用到大部分的低阶矩阵逼近问题，在各种情况下展现了它的效率，表明了它在现有方法中的有效性和鲁棒性。

## 1.介绍

低阶矩阵逼近问题吸引了很多领域的注意，例如数据重建、图像去噪、协同过滤、背景建模、运动结构和光度立体等。它通常假定矩阵的秩是固定的或事先知道的，并且被认为是一个子空间学习问题。

尽管真实世界的数据通常是高维的，在很多情况下它们也可用较少的参数表示。因此，将数据降维到一个主结构可减少计算时间并删除不想要的噪音成分。解决该问题较好的方法是主成分分析（PCA）。PCA 基于二语规范的数据方差最大化将数据转换到低维空间。为了处理丢失的数据，求解加权低秩矩阵逼近问题采用期望最大化算法。林提出用投影梯度法来解决图像和文本数据的非负矩阵分解问题。Mi-tra et al 提出用增加正则化项的矩阵分解技术防止过度拟合数据和采用半定规划解决问题。这些传统的二语规范的近似方法已经在许多问题中使用，但是众所周知，这些方法对异常和腐败非常敏感，因为 L2 范数放大了数据损坏的负面影响。

作为代替，利用 L1 范数的低秩逼近方法被用来保证对异常值和缺失数据的鲁棒性。另外，已经有低阶矩阵分解的几个概率扩展被用来进行鲁棒逼近。Ke 和 Kanade 提出一种低秩矩阵近似法，该方法采用另外一个基于 L1 范数最小化代价函数的凸规划。Eriksson 和 Hengel 提出 l1-Wiberg 方法来进行缺失数据的加权低秩矩阵逼近。Kwak 提出 L1 最大化方法在特征空间使用贪心方法找到连续的主成分。Kim et al. 提出两种基于梯度的方法，使用整流表示 L1 最小化问题。

近来，许多许多有效的使用增强拉格朗日的方法被用来解决 L1 极小化问题。Shen et al. 提出一种基于增广拉

格朗日交替方向法的L1范数逼近低秩矩阵的方法（ALADM）。Zheng et al. 提出一个利用核范数正则化L1成本函数的正交性约束的实用的加权低秩逼近法(Regl1-ALM)。Cabral et al.提出一种统一的方法，它结合了核范数最小化和双线性分解的核规范的替代定义。他们的方法不需要高的计算复杂度就可得到高效的算法，这些方法已经成功应用到有丢失的数据和异常值存在的低阶因子化方法问题，优于其他秩最小化方法。但是由于问题是非凸的，很难通过分解的方法找到全局最优解。

有一些其他的名为鲁棒主成分分析的基于核范数最小化（RPCA）的最新进展的方法，已经成功地运用到一些问题上。RPCA试图用L1范数正则化核范数成本函数来解决非固定秩矩阵逼近问题和方法，如增强拉格朗日方法（ALM）和加速近梯度（APG）。已经证明，RPCA适合阴影去除和背景建模等问题。然而，为RPCA提出的算法有高计算复杂度，尤其是在大规模的问题上，因为这些算法在每次迭代时执行奇异值分解（SVD）。近来，Shu et al.提出了一种新的秩测度的低阶恢复方法。但是，上述方法有时在变异值下只能找到一个次优解，这是实践难题。

在本文中，我们提出了一种基于在严重污染下的鲁棒子空间学习问题的弹性网正则化高效低秩矩阵分解方法，包括异常值和失败的尝试。我们的方法是一种利用核范数最小化和双线性分解的整体方法。为了防止算法因高度损坏的数据而不稳定，我们为奇异值引入弹性网正规化来介绍强凸性套索式核范数最小化问题。该方法的强凸性使其在不必要的噪声的存在下，通过缩小和纠正不准确的奇异值来缓解稳定性问题。我们也表明使该算法的任何限制点满足的必要条件是一个局部最优解。我们在重构误差和计算速度方面使用知名的基准数据集演示所提出的方法的性能，例如非刚体运动估计，光度立体，背景建模。

# 2. 提出的方法

## 2.1 问题制定

在本文中，我们认为基于秩和稀疏的凸包络功能的低秩矩阵和稀疏矩阵分离问题如下：

$$\min_{P,X} f_1(P,X) + \lambda\|PX\|_*, \qquad (1)$$

当 f1(P, X) = ||W ⊙ (Y - P X)||时，Y是一个观测矩阵，λ是一个预先定义的加权参数。||.||1和||.||*分别表示进入明智的L 1范数和核范数，这是L0范数和秩函数凸松弛。在这里，⊙是分量相乘或Hadamard积，W是一个加权矩阵，当$y_{ij}$已知时$w_{ij}$为1，未知时为0。这个问题在低秩矩阵D和稀疏误差矩阵E分别代替PX和Y-PX时非常相似。一般而言，（1）是一个非凸和非光滑问题，使它很难找到一个有效和准确的解决方案。为了

有效地解决这个问题，一个共同的策略是使用交替最小化的方法解决一个变量，而其他变量是固定的。

注意（1）正则化项，$\|PX\|_*$可以解释为一个奇异值的总和$\sum_i^r|\sigma_i|$，$\sigma_i$是低秩矩阵 PX，奇异值 r 是 PX 的秩。这导致了一个套索的问题，其中有一个阈值效应对奇异值。但是，基于 Lasso 的方法由于其弱凸性，缺乏一个收缩的影响，这使得数据高度损坏时算法不稳定。为了提高算法的稳定性，我们介绍了奇异值强凸化与 L 2 奇异值范数法：$\lambda_1\sum_i^r|\sigma_i|+\frac{\lambda}{2}\sum_i^r|\sigma_i|^2$。基于事实$\|D\|_F^2=tr(V\sum U^TU\sum V^T)=tr(\sum^2)=\sum_i|\sigma_i|^2$，且$D=U\sum V^T$是 D 的 SVD，我们引入一个新的处罚优化问题如下：

$$\min_{P,X} f_1(P,X)+\lambda_1\|PX\|_*+\frac{\lambda_2}{2}\|PX\|_F^2. \tag{2}$$

在这个公式中，我们有奇异值的弹性网络化，在许多应用中与套索相比已经显示出其优势。由于 Frobenius 范数，它能基于强凸性稳定一个套索式方法。另外，我们都有一个阈值效应对奇异值从 L1 和 L2 正则化定期形成一个收缩效应产生一个吝啬且稳定的模型。

请注意，没有这些正则化条款，问题（2）可以使用增广拉格朗日铝交流方向法(ALADM)求解。有使用核范数正则化的 L1 范数成本函数的另一种方法。它是利用核范数的另一种定义的扩展，但不包含给定的平滑项。然而，这些方法可以找到一个次优的解决方案，由于这些交替最小化方法没有一个适当的反应条件，可能会在数据严重损坏时导致该方法变得低效。

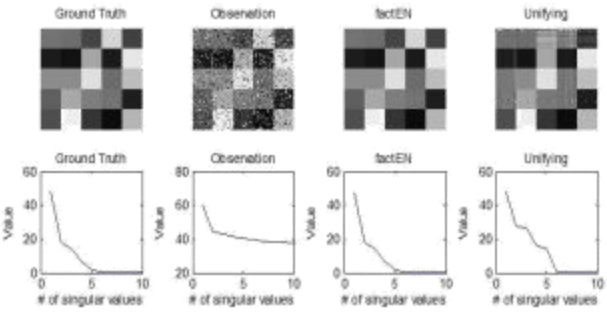



图 1.为一个玩具的例子所提出的套索法的评价

图 1 显示了将所给方法与套索法和地面上的一个简单的例子（100×100）20%离群值相比较的结果，地面真理等级为五。由图可知，所提出的方法对异常值给出了一个稳定的结果，并通过抑制奇异值来消除噪声，而统一的国际统一的发现相对不准确且有更高的奇异值，相比于临提出的方法和地面真理，产生了很差的重建结果。

不幸的是，（2）对于大规模问题来说可能会产生很高的计算复杂度，因为问题是在每一次迭代进行 SVD 分解时得到解决，SVD 是用于解决一个核规范为基础的成本函数。要在实践中解决（2），可以利用以下属性的核规范。

**引理 1**：任何矩阵$D\in R^{m\times n}$，以下是：

$$\|D\|_* = \min_{P,X} \frac{1}{2}(\|P\|_F^2 + \|X\|_F^2) \quad \text{s.t. } D=PX$$

(3)

如果 D 的秩为 r≤min（M，N），则以上最小的方法是因子分解 $D = P_{m\times r} X_{r\times n}$。

使用引理 1，我们把形式（2）等价如下:

$$\min_{P,X,D} f_2(D) + \frac{\lambda_1}{2}(\|P\|_F^2 + \|X\|_F^2) + \frac{\lambda_2}{2}\|D\|_F^2$$

（4）

这样 D=PX，$f2(D) = \|W \odot (Y - D)\|_1$。由于解决问题（4）在实践中有些困难，我们介绍另一个辅助变量 $\hat{D}$ 来解决以下问题

$$\min_{P,X,D,\hat{D}} f_2(\hat{D}) + \frac{\lambda_1}{2}(\|P\|_F^2 + \|X\|_F^2) + \frac{\lambda_2}{2}\|D\|_F^2$$

，s.t.　D=PX, $\hat{D} = D$　(5)

为了解决（5），我们利用增广拉格朗日框架将（5）转换为无约束问题：

$$\begin{aligned}\mathcal{L}(P, X, D, \widehat{D}, \Lambda_1, \Lambda_2) &= f_2(\widehat{D}) + \frac{\lambda_1}{2}\left(\|P\|_F^2 + \|X\|_F^2\right) \\ &+ \frac{\lambda_2}{2}\|D\|_F^2 + \operatorname{tr}\left(\Lambda_1^T(D - PX)\right) + \operatorname{tr}\left(\Lambda_2^T(\widehat{D} - D)\right) \\ &+ \frac{\beta}{2}\left(\|D - PX\|_F^2 + \|\widehat{D} - D\|_F^2\right),\end{aligned} \tag{6}$$

$\Lambda_1, \Lambda_2 \in R^{m\times n}$ 的拉格朗日乘子，β＞ 0 是一个小小的惩罚参数。

## 2.2 算法

根据之前的描述，我们提出了一种基于增强拉格朗日框架的方法，用一种交替最小化技术解决它。为了解决问题 P，我们修复了其他变量并解决了接下来的优化问题：

$$\begin{aligned}P_+ = \arg\min_P \frac{\lambda_1}{2}\|P\|_F^2 &+ \operatorname{tr}\left(\Lambda_1^T(D - PX)\right) \\ &+ \frac{\beta}{2}\|D - PX\|_F^2.\end{aligned} \tag{7}$$

这个优化问题是一个最小二乘问题，解决方案是：

$$P_+ = (\Lambda_1 + \beta D)X^T(\lambda_1 I + \beta X X^T)^{-1}, \tag{8}$$

I 表示身份矩阵。对于 X，我们解决以下优化问题：

$$\begin{aligned}X_+ = \arg\min_X \frac{\lambda_1}{2}\|X\|_F^2 &+ \operatorname{tr}\left(\Lambda_1^T(D - PX)\right) \\ &+ \frac{\beta}{2}\|D - PX\|_F^2,\end{aligned} \tag{9}$$

可以解决类似于（7）的问题，它的解决方案是：

$$X_+ = (\lambda_1 I + \beta P^T P)^{-1} P^T(\Lambda_1 + \beta D). \tag{10}$$

为了寻找 D，我们考虑以下优化问题：

$$\begin{aligned}D_+ = \arg\min_D &\frac{\lambda_2}{2}\|D\|_F^2 \\ &+ \operatorname{tr}\left(\Lambda_1^T(D - PX)\right) + \operatorname{tr}\left(\Lambda_2^T(\widehat{D} - D)\right) \\ &+ \frac{\beta}{2}\left(\|D - PX\|_F^2 + \|\widehat{D} - D\|_F^2\right),\end{aligned} \tag{11}$$

它的解决方案是：

$$D_+ = \frac{\beta PX + \beta\widehat{D} + \Lambda_2 - \Lambda_1}{\lambda_2 + 2\beta}. \tag{12}$$

我们得到下面的方程来解决问题 $\hat{D}$：

$$\widehat{D} = \arg\min_{\widehat{D}} f_2(\widehat{D}) + \operatorname{tr}\left(\Lambda_2^T(\widehat{D} - D)\right) + \frac{\beta}{2}\|\widehat{D} - D\|_F^2 \tag{13}$$

并利用绝对值阈值算子计算该解：

$$\begin{cases} W \odot \widehat{D}_{+} \leftarrow W \odot \left(Y - \mathcal{S}\left(Y - D + \frac{\Lambda_2}{\beta}, \frac{1}{\beta}\right)\right), \\ \overline{W} \odot \widehat{D}_{+} \leftarrow \overline{W} \odot \left(D - \frac{\Lambda_2}{\beta}\right), \end{cases} \tag{14}$$

最后，我们更新的拉格朗日乘子为：

$$\begin{aligned} \Lambda_1 &= \Lambda_1 + \beta(D - PX), \\ \Lambda_2 &= \Lambda_2 + \beta(\widehat{D} - D). \end{aligned} \tag{15}$$

**Algorithm 1** factEN by ALM for optimizing (5)

```
1: Input: Y ∈ ℝ^{m×n}, r, β, ρ, and λ1, λ2 = 10^{-3}
2: Output: P ∈ ℝ^{m×r}, X ∈ ℝ^{r×n}, and D ∈ ℝ^{m×n}
3: while not converged do
4:   while not converged do
5:     Update P using (8)
6:     Update X using (10)
7:     Update D using (12)
8:     Update D̂ using (14)
9:   end while
10:  Update the Lagrange multipliers Λ1, Λ2 using (15)
11:  β = min(ρβ, 10^{20})
12: end while
```

在前面分析的基础上，我们推导出一个强大的弹性网络化的低秩矩阵分解算法，该算法总结了算法 1。由于算法是基于弹性网络正规化且使用了矩阵分解的方法，所以该方法被命名为 factEN。在算法中，我们假设了一个归一化的观测矩阵，因此输出矩阵 P 和 X 可以通过重新缩放使用缩放因子得到。我们对优化变量进行高斯 N 的初始化。

该算法的内循环的计算复杂度（算法 1 中的 4-9 行）为 $O(mnr)$，与 Unifying 和 ALADM 相同。IALM [17] and Regl1-ALM [34]在每一次迭代进行奇异值分解操作，它们的计算复杂性是 $O(\min(m,n)\max(m,n)^2)$ 和 $O(r\max(m,n)^2)$，比 factEN 需要更多的努力。注意，该方法可以很容易地扩展到通过从测量矩阵的子矩阵采样来描述每个迭代线性复杂度的算法速度。

## 2.3 收敛性分析

在这一节中,我们分析了所提出的方法的收敛性。虽然很难保证其收敛到局部极小，但实证证据表明，提出的算法具有很强的收敛行为。然而，我们通过在温和的条件下，算法产生的迭代序列的极限点是一个满足库恩-塔克（KKT）条件下的固定点来提供 factEN 弱收敛性证明。值得证明任何收敛点必须是一个满足 KKT 的点，因为这是成为一个局部最优解的必要条件。这一结果为该算法的行为提供了一个保证。

我们通过假设完全观测数据模型（5）i.e.来改写 factEN 成本函数，对所有的 i、j，$w_{i,j}=1$，如下所示：

$$\min_{P,X,D,\widehat{D}} f_3(\widehat{D}) + \frac{\lambda_1}{2}\left(\|P\|_F^2 + \|X\|_F^2\right) + \lambda_2\|D\|_F^2 \quad \text{s.t.} \;\; D = PX,\ \widehat{D} = D. \tag{16}$$

其中 $f_3(\hat{D}) = \|Y - \hat{D}\|_1$，然而，类似的结果只可以得出部分观测到的数据模型。

让我们假设该算法达到一个固定点。KKT 条件如下：

$$D - PX = 0,\ \widehat{D} - D = 0,\ \frac{\partial \mathcal{L}}{\partial P} = \lambda_1 P - \Lambda_1 X^T$$
$$\frac{\partial \mathcal{L}}{\partial X} = \lambda_1 X - P^T \Lambda_1 = 0,\ \frac{\partial \mathcal{L}}{\partial D} = \lambda_2 D + \Lambda_1 - \Lambda_2$$
$$\Lambda_2 \in -\partial_{\widehat{D}}(\|Y - \widehat{D}\|_1).$$

在这里，我们可以从（17）得到下面的方程的最后关系：

$$\begin{aligned} Y - D + \frac{\Lambda_2}{\beta} &\in Y - D - \frac{1}{\beta}\partial_{\widehat{D}}(\|Y - \widehat{D}\|_1) \\ &= Y - \widehat{D} - \frac{1}{\beta}\partial_{\widehat{D}}(\|Y - \widehat{D}\|_1) \triangleq Q_\beta(Y - \widehat{D} \end{aligned}$$

从（23）我们可以得出如下关系：

$$Y - \widehat{D} = Q_\beta^{-1}\left(Y - D + \frac{\Lambda_2}{\beta}\right) \equiv \mathcal{S}\left(Y - D + \frac{\Lambda_2}{\beta}\right.$$

其中 $S(x,r) = sign(x)\max(|x| - \tau, 0)$。

因此，KKT 条件可以改写为：

$$\begin{aligned} &D - PX = 0,\ \widehat{D} - D = 0,\ \lambda_1 P - \Lambda_1 X^T = 0, \\ &\lambda_1 X - P^T \Lambda_1 = 0,\ \lambda_2 D + \Lambda_1 - \Lambda_2 = 0, \\ &Y - \widehat{D} = \mathcal{S}\left(Y - D + \frac{\Lambda_2}{\beta}, \frac{1}{\beta}\right). \end{aligned} \tag{20}$$

基于这些条件，我们证明了满足 KKT 条件的收敛点。

**定理 1**：让 $G \stackrel{\Delta}{=} (P, X, D, \hat{D}, \Lambda_1, \Lambda_2)$，$\{G^j\}_{j=1}^{\infty}$ 由 fastEn 生成。假设 $\{G^j\}_{j=1}^{\infty}$ 是有界的且 $\lim_{j\to\infty}\{G^{j+1} - G^j\} = 0$。然后 $\{G^j\}_{j=1}^{\infty}$ 的任何计算点满足 KKT 条件。特别地，不管 $\{G^j\}_{j=1}^{\infty}$ 是否收敛，它都收敛为一个 KKT 点。

**证明**：首先，我们从该算法中得到拉格朗日乘子 $\Lambda_1, \Lambda_2$ 如下：

$$\begin{aligned} \Lambda_{1+} &= \Lambda_1 + \beta(D - PX) \\ \Lambda_{2+} &= \Lambda_2 + \beta(\widehat{D} - D), \end{aligned} \tag{21}$$

其中在序列 中 $\Lambda_{i+}$ 是 $\Lambda_i$ 的下一个点。

如果序列中的变量 $\{\Lambda_1^j\}_{j=1}^{\infty}$ 和 $\{\Lambda_2^j\}_{j=1}^{\infty}$ 收敛到一个固定点，$(\Lambda_{1+} - \Lambda_1) \to 0$ 和 $(\Lambda_{2+} - \Lambda_2) \to 0$，然后 $(D - PX) \to 0, (\hat{D} - D) \to 0$，这满足 KKT 条件中的第一个条件。

其次，我们从 P+来推导算法，可以得到：

$$P_+ - P = (\Lambda_1 + \beta D)X^T(\lambda_1 I + \beta XX^T)^{-1} - P, \tag{22}$$

其中 I 表示身份矩阵，可以用乘法 $(\lambda_1 I + \beta XX^T)$ 改写为：

$$\begin{aligned} &(P_+ - P)(\lambda_1 I + \beta XX^T) \\ &\quad = \Lambda_1 X^T - \lambda_1 P + \beta(D - PX)X^T. \end{aligned} \tag{23}$$

从第一个条件，我们可以得出 $\Lambda_1 X^T - \lambda_1 P \to 0$ 当 $(P_+ - P) \to 0$。

第三，在算法中运用 $X_+ = (\lambda_1 I + \beta P^T P)^{-1} P^T(\Lambda_1 + \beta D)$，我们可以得到：

$$\begin{aligned} &(\lambda_1 I + \beta P^T P)(X_+ - X) \\ &\quad = P^T \Lambda_1 - \lambda_1 X + \beta P^T(D - PX). \end{aligned} \tag{24}$$

如果 $(X_+ - X) \to 0$，$(P^T \Lambda_1 - \lambda_1 X) \to 0$。

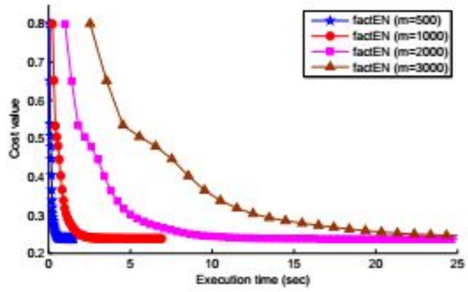



图 2. 在每次迭代中的四个合成实例的成本值。

同样，使用算法得到的 D+法，我们可以得到下面的方程：

$$(\lambda_2 + 2\beta)(D_+ - D) = \beta(PX - D + \widehat{D} - D) - \Lambda_1 + \Lambda_2 - \lambda_2 D.$$

最后，从（19）我们可以得到如下方程：

$$\widehat{D}_+ - \widehat{D} = Y - \mathcal{S}\left(Y - D + \frac{\Lambda_2}{\beta}, \beta\right) - D.$$

我们假设$\{G^j\}_{j=1}^{\infty}$是有界的，$\{X_+X_+^T\}_{j=1}^{\infty}$和$\{P_+^T P_+\}_{j=1}^{\infty}$也是有界的，因此方程(21)，(23)，(24)，(25)和(26)在 j 趋于无穷时趋于零。因此，结果渐近满足 KKT 条件（16），命题得证。

在我们的算法中，我们设置停止标准为：

$$\frac{\|D^{(t)} - P^{(t)}X^{(t)}\|_1}{\|Y\|_1} < \theta,$$

其中 T 是迭代次数， θ 是一个很小的正数。这里，我们在包括元素对应的未知项全要素的条件下计算 D。当相邻迭代的终止成本之间的差异变小时，对算法来说能够找到一个近似固定点，在第三节的实验中我们设定的停止条件为 θ = 10−5。图 2 在每次迭代的 500*500 到 3000*3000 的四个例子中显示缩放成本值，与 3.1 节所描述的异常值相符，每一点都表示每次迭代的成本值。如图所示，factEN 所花费的价值减少的非常快，且在很少的迭代次数内收敛到一个固定点。

## 3.实验结果

我们通过尝试各种合成和真实世界的问题来评价所提出的方法 fastEN 的性能，例如非刚性运动估计[ 30,34 ]，光度立体[ 4，33 ]和背景建模[ 24，32 ]。我们将 fastEN 同国家的最先进的低秩逼近方法相比较，例如 ALADM[23]，Regl1 -ALM [34]，统一法[ 4 ]，秩估计方法 IALM [17] 和 ROSL [24]。我们将 factEN 参数设置如下：长颈鹿和静态人脸数据集时$\rho$=1.05，其余情况下$\rho$=1.2；非刚性运动估计问题时$\beta = 10^{-2}$，其余情况下$\beta = 0.5$。注意$\beta = \beta_0 / \|Y\|_\infty$。

### 3.1. 综合数据

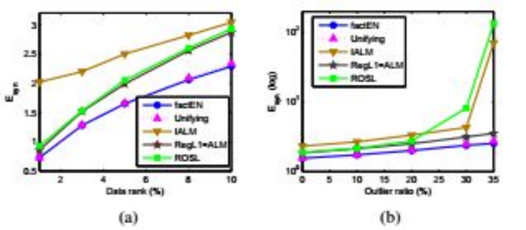



图 3. 各种条件下的合成实例（500×500）的平均性能。(a)对不同观测数据的平均重建误差（5%）。(b)不同的异常值比（5%数据秩）的平均重建误差。

首先，我们将提出的方法应用到合成的例子。我们在 500*500 到 1000*1000 的范围内产生了六个有高斯噪声的测试集，是从 $N(0,10^{-2})$ 采样的。实验中，平均重构误差 $E_{syn}$ 被计算，其中 $M^{gt}$ 是基本真相，$\widehat{M}$ 是低秩矩阵近似的应用算法。

图像 3 显示一个合成（500*500）实例的平均性能，列出了各种数据的行列和各种异常率，以验证在各种条件下的鲁棒性。总体，所提出的方法和统一给出了这两种情况下的重建误差的最佳平均性能。从图像 3（b）可知，当离群值比较小时大多数方法是强大的，但是 ROSL 和 IALM 当异常值增加时性能较差，限制他们在实践中的应用。

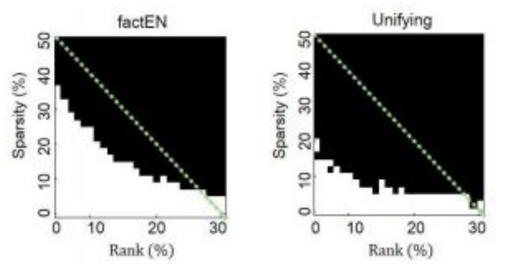



图 4. 利用所提出的方法和统一的一个合成的例子（1000 ×1000）的秩和稀疏的相变。当一个恢复低秩矩阵满足 $\| M^{gt} - \widehat{M} \|_1 / \| M^{gt} \|_1 \le 5 \times 10^{-4}$ 时可实现正确的恢复（白色区域）

要验证所提出的方法与关于秩和稀疏性的统一相比较的能力，我们进行了一个 1000*1000 合成的例子。图 4 表明了在不同的等级和稀疏比率的情况下正确回收率的部分。所提出的方法正确地恢复的区域似乎是比统一更广泛的。由图可知，本文提出的方法更能够比统一更能处理腐败。

图 5（a）和 5（b）分别显示了平均重建误差和执行时间不同的算法，各种 8%固定数据等级和 4%离群值的矩阵均匀分布在[-20,20 ]范围内。我们不能根据大规模的问题（1000*1000）来评价 IALM 和 Regl 1 -ALM，因为计算复杂度太高。在所有情况下，所提出的方法在误差重建方面优于其他方法。尽管 Regl 1 -ALM 与所提出的方法在数据规模较小时显示出了相似的性能，但它需要花费的时间较长且在数据规模较大时性能较差。ALADM 的计算时间比 fastEn 要短，但性能较差。

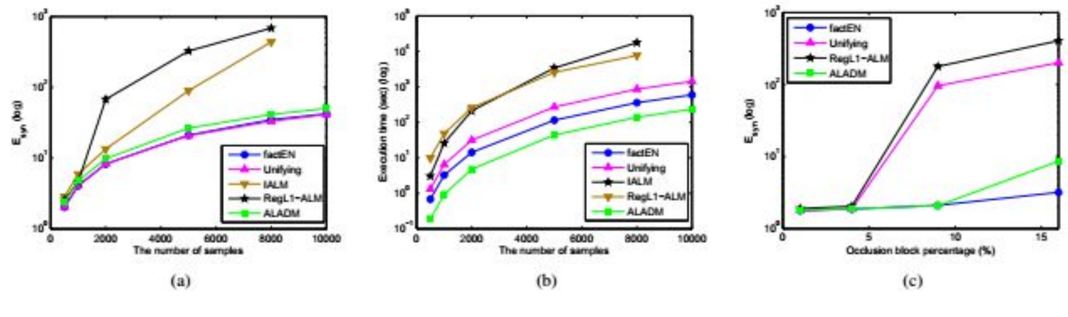



图 5. 数据损坏问题存在时的平均合成性能。（a）各种数据大小随机值的平均重构误差。（b）各种数据大小的平均执行时间。（c）不同区块腐败规模的平均重构误差以及 300×300 的例子 20%的缺失度。

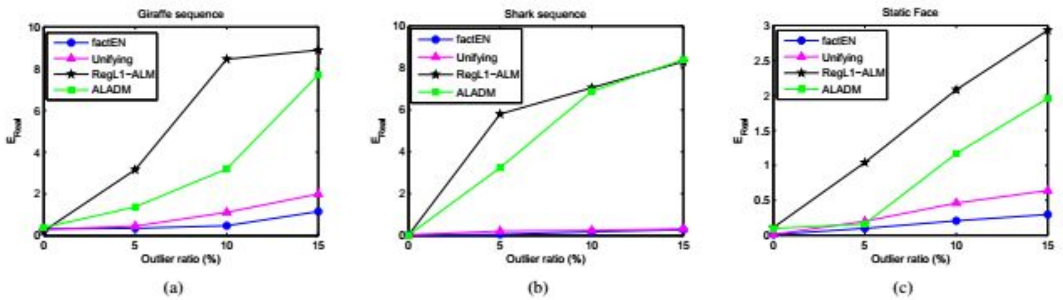



图 6. 在存在的异常值和丢失的数据时现实世界问题的平均表现（非刚体运动估计，光度立体）。（a)长颈鹿序列，(b)鲨鱼序列，(c)静态面。

为了在现实条件下衡量所提出的算法，我们在一个合成的例子中通过改变异常值来检查丢失条目的数据块损坏。对于近似结构的 300*300 的例子，我们增加了各种大小的闭塞与 20%缺失的数据。图 5（c）显示了不同方法的重建误差。如图所示，所给方法显示了重建的鲁棒性，而除了 ALADM 的其他方法在有大规模的数据块损坏时只能给出差的重建结果。

## 3.2. 现实世界问题

我们评估了所提出的方法对表 1 中现实世界问题的性能。对于这些问题，我们对以上所观察到的条目计算的平均绝对误差为

$$E_{\mathrm{Real}} = \frac{\| W \otimes M^{gt} - \widehat{M} \|_1}{\| W \|_1}$$

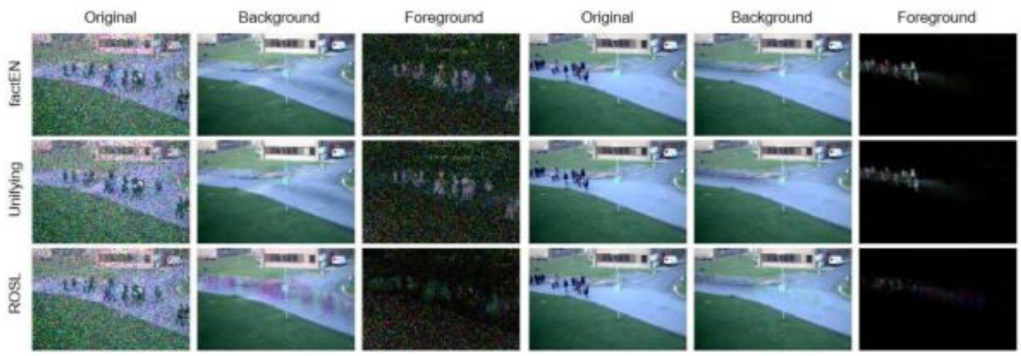
。

首先，我们使用长颈鹿序列[ 3 ]进行了一个非刚性运动估计实验。为了证明所提出的方法的鲁棒性，我们在一个框架中取代了 5%的随机选择的点，异常值的范围为[ 0，100 ]，而数据点是在[ 127，523 ]的范围内。在这个设置中，我们通过改变离群值的数据进行了几个实验。在各种异常值的存在下，长颈鹿序列的结果如图 6（a）所示，该图还包括在没有异常值增加时的情况。如图所示，fastEn 在不管离群值比值的情况下给出了最佳性能。虽然统一在离群值比值较小时给出了类似的重建性能，但在异常值比值增加时性能变差。 Regl 1 -ALM 和 ALADM 相比其他国家的最先进的方法表现出较差的性能。图 7 显示了对于 fastEn 和统一法来说 $\lambda_1$ 的选择如何影响平均重构误差。所提出的方法在 $\lambda_1$ 和 $\lambda_2$ 取不同值时显示了更稳定的性能，而统一法对 $\lambda_1$ 的选取较敏感。

表 1. 已知秩的实际问题总结

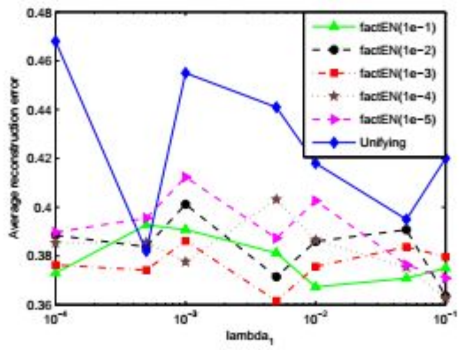
| Datasets | Size | Rank $r$ | Missing |
|---|---|---|---|
| Giraffe [3] | 91 × 240 | 6 | 30 % |
| Shark [31] | 240 × 167 | 6 | 10 % |
| Static Face [3] | 4,096 × 20 | 4 | 42 % |
| PETS 2009 [1] | 110,592 × 221 | 2 | 0 % |

我们还使用鲨鱼序列进行了运动估计问题。在这里，我们在每一帧中随机丢弃 10%个点作为数据丢失。我们在[−1000, 1000]的范围内对每一帧从 0%到 15%的跟踪点作

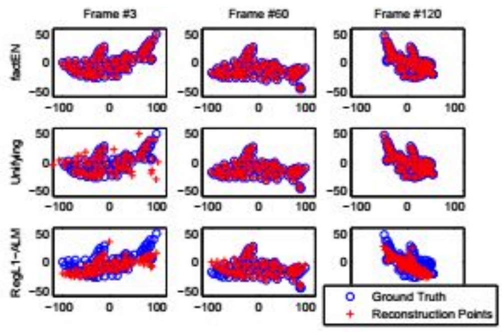



图(9). 在 PEST2009 数据集中选定的两帧的背景建模方法的结果。每个算法将原始图像分解为背景和前景图像。

为离群值，数据点在 [-105,105]的范围内。各种不同方法的异常值比的平均重构误差如图 6（b）所示，由图可知，fastEn 和统一法都给出了较好的重建结果。然而，所提出的方法在平均值方面给出了比统一法更好的重建结果。三个选定的算法的重建结果如图 8 所示，由图可以发现与其他方法相比，所给方法能观察到更好的重建效果。




图(7). 所提出的方法和统一法在长颈鹿序列不同 $\lambda_1$ 时的比较。$(\cdot)$ 表示 $\lambda_2$ 的值。




图(8). fastEn、统一法和 RegL1-ALM 对鲨鱼序列的重建结果。

对于光度立体问题，我们使用静态人脸序列。我们研究所提出的方法在存在缺失数据的条件下对异常值的稳定性如何。我们在[ 0，100 ]的范围内对每一个框架从 0%到 15%的跟踪点作为离群值，总的结果如图 6（c）所示。由图可知，无论离群值比如何，所提出的方法与其它方法相比都有明显的区别。

为完成背景建模任务，我们用 PETS2009，调整每帧为 288 × 384。我们将所提出的方法与国家最先进的方法：统一法和 ROSL 进行了对比，增加了 30%个随机选择帧的随机噪声，图 9 显示了选定帧的背景建模结果。由图可知，fastEn 和统一法正确分离了前景。秩估计法 ROSL 未能在大量数据损

坏存在时找到一个好的解决方案。所提出的方法的计算时间为 186.37 秒，统一法为 497.46 秒，ROSL 为 145.93 秒。尽管 ROSL 比 fastEn 有稍快的计算时间，但它没有提供令人满意的结果。

## 4.结论

在本文中，我们提出了一种基于奇异值的弹性网正则化的子空间学习的新方法——fastEn。该方法可以处理丢失或未知的条目，以及异常值。随着弹性网正则化方案的引入，所提出的方法可以找到一个更有效的解决方案，而且对丢失的数据、异常值和不同的参数值来说是稳定的。所提出的方法已被应用到各种问题中，例如非刚体运动估计、光度立体、背景建模问题。实验结果表明，该方法在近似误差和执行时间方面优于其他现有的的方法。调查所提出的方法在大规模和更具挑战性的问题上的竞争力将会是很有趣的。

## 参考文献


[1]PETS 2009 dataset. http://www.cvg.rdg.ac.uk/PETS2009.

[2] S. P. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

[3] A. M. Buchanan and A. W. Fitzgibbon. Damped newton Algorithms for matrix factorization with missing data. In CVPR, 2005.

[4] R. Cabral, F. D. la Torre, J. P. Costeira, and A. Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In ICCV, 2013.

[5] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? Journal of the ACM, 58:11:1 – 11:37, 2011.

[6] X. Ding, L. He, and L. Carin. Bayesian robust principal component analysis. IEEE Trans. on Image Processing, 20:3419 – 3430, 2011.

[7] A. Eriksson and A. Hengel. Efficient computation of robust weighted low-rank matrix approximations using the l 1 norm. IEEE Trans. on Pattern Analysis and Machine Intelligence, 34(9):1681 – 1690, 2012.

[8] X. Guo, X. Wang, L. Yang, X. Cao, and Y. Ma. Robust fore-ground detection using smoothness and arbitrariness con-straints. In ECCV, 2014.

[9] Z. Harchaoui, M. Douze, M. Paulin, M. Dudik, and J. Mal-ick. Large-scale image classification with trace-norm regu-larization. In CVPR, 2012.

[10] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans. on Pattern Analysis and Machine Intelligence, 35(9):2117－2130, 2013.

[11] I. T. Jolliffe. Principal Component Analysis. John Wiley and Sons, 1986.

[12] Q. Ke and T. Kanade. Robust l 1 norm factorization in the presence of outliers and missing data by alternative convex programming. In CVPR, 2005.

[13] E. Kim, M. Lee, C.-H. Choi, N. Kwak, and S. Oh. Efficient l 1 -norm-based low-rank matrix approximations for large-scale problems using alternating rectified gradient method. IEEE Trans. on Neural Networks and Learning Systems, 26(2):237－251, 2015.

[14] N. Kwak. Principal component analysis based on L 1 -norm maximization. IEEETrans.onPatternAnalysisandMachine Intelligence, 3(9):1672－1680, 2008.

[15] H. Li, N. Chen, and L. Li. Error analysis for matrix elastic-net regularization algorithms. IEEE Trans. on Neural Networks and Learning Systems, 23(5):737－748, 2012.

[16] C.-J. Lin. Projected gradient methods for nonnegative matrix factorization. Neural Computation, 19:2756－2779, 2007.

[17] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices.Mathematical Programming, 2010.

[18] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. Journal of Machine Learning Research, 11:2287－2322, 2010.

[19] D. Meng and F. D. la Torre. Robust matrix factorization with unknown noise. In ICCV, 2013.

[20] K. Mitra, S. Sheorey, and R. Chellappa. Large-scale matrix factorization with missing data under additional constraints. In NIPS, 2010.

[21] T. Okatani, T. Yoshida, and K. Deguchi. Efficient algorithm for low-rank matrix factorization with missing components and performance comparision of latest algorithms. In ICCV, 2011.

[22] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review, 52:471 – 501, 2010.

[23] Y. Shen, Z. Wen, and Y. Zhang. Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization. Optimization Methods and Software, pages 1 – 26, 2012.

[24] X. Shu, F. Porikli, and N. Ahuja. Robust orthonormal subspace learning: Efficient recovery of corrupted low-rank matrices. In CVPR, 2014.

[25] N. Srebro. Weighted low-rank approximations. In ICML, 2003.

[26] T. Sun and C.-H. Zhang. Calibrated elastic regularization in matrix completion. In NIPS, 2012.

[27] A. Talwalkar, L. Mackey, Y. Mu, S.-F. Chang, and M. I. Jordan. Distributed low-rank subspace segmentation. In ICCV, 2013.

[28] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 58:267 – 288, 1994.

[29] K.-C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. Journal of Optimization, 6(3):615 – 640, 2010.

[30] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: factorization method. International Journal of Computer Vision, 9(2):137 – 154, 1992.

[31] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3d shape from 2d motion. In NIPS, 2003.

[32] N. Wang and D.-Y. Yeung. Bayesian robust matrix factorization for image and video processing. In ICCV, 2013.

[33] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, T. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In ACCV, 2010.

[34] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust l 1 -norm. In CVPR, 2012.

[35] T. Zhou and D. Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In ICML, 2011.

[36] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society,
Series B, 67:301–320, 2005.