A Double Weighted Model for Real Noisy Image Denoising

Anonymous ICCV submission

Paper ID ****

Abstract

Motivated by the weighted Orthogonal Procrustes Problem, we propose a noval weighted Frobenious norm based weighted sparse coding model for non-Gaussian error modeling. We solve this model in an alternative manner. Updating of each variable has closed-form solutions and the overall model converges to a stationary point. The proposed model is applied in real image denoising problem and extensive experiments demonstrate that the proposed model can much better performance (over 1.0dB improvement on PSNR) than state-of-the-art image denoising methods, including some excellant commercial software. The noval weighted Frobeniius norm can perfectly fit the non-Gaussian property of real noise.

1. Introduction

Given a patch extracted from the tested image, assume $d = 64$)的维度，拉成列向量$\mathbf{y} \in \mathbb{R}^{d}$，那么我们在其周围找$m$个相似块，得到相似块矩阵$\mathbf{Y} \in \mathbb{R}^{d\times m}$。我们希望在这个数据里学习到一个正交字典，同时利用稀疏表达这个工具。很自然，我想到了K-SVD这个模型，给定数据矩阵$\mathbf{Y}$,KSVD不断迭代更新，从而得到最适合表达数据的字典和稀疏。但是与KSVD不同的是，我希望我的模型里，字典是正交的并且有闭合解，系数也可以有闭合解。从而，我们想到的模型主体是这样的：

$$\min_{\mathbf{D},\mathbf{C}} \frac{1}{2}\|\mathbf{Y}-\mathbf{D}\mathbf{C}\mathbf{W}\|_F^2+\lambda\|\mathbf{C}\|_{1,\mathbf{s}} \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D}=\mathbf{I}. \tag{1}$$

或者

$$\min_{\mathbf{D},\mathbf{C}} \frac{1}{2}\|(\mathbf{Y}-\mathbf{D}\mathbf{C})\mathbf{W}\|_F^2+\lambda\|\mathbf{C}\|_{1,\mathbf{s}} \quad \text{s.t.} \quad \mathbf{D}^\top\mathbf{D}=\mathbf{I}. \tag{2}$$

其中$\|\mathbf{C}\|_{1,\mathbf{s}}$表示用的是加权L1范数(对每一行加不同的权值)。

这两个模型是类似的解法，但是物理意义稍有不同。我们先对模型进行初始化：

2. 初始化

- 正交字典$\mathbf{D}^{(0)}$由$\mathbf{Y}=\mathbf{D}\mathbf{\Sigma}^{(0)}\mathbf{V}^\top$做SVD分解得到;
- 稀疏系数$\mathbf{C}$的行方向上的权值$\mathbf{s}^{(0)} = \text{diag}(\mathbf{\Sigma}^{(0)})$是weighted sparse coding的权重。$\mathbf{C}$每一行的权值不同，意味着正交字典$\mathbf{D}$的每一列对$\mathbf{C}$的重要性不同；
- $\mathbf{W}^{(0)}$是对角矩阵，初始化为恒等矩阵$\mathbf{I}_{m\times m}$(原因是一开始把所有块平等对待，视为受到相同的高斯噪声的影响)。或者初始化为一个对角矩阵，对角线上的每个元素对应于$\mathbf{Y}$的每一列（图像块）上的噪声水平σ，这个噪声水平可以用现有估计高斯噪声水平的方法估计出来);

3. 反复迭代求解D,C直到收敛,更新W

3.1. 反复迭代求解D,C直到收敛

$k = 0, 1, 2, ...$:

a. update $\mathbf{C}$

$$\mathbf{C}^{(k+1)} = \arg\min_{\mathbf{C}} \frac{1}{2}\|(\mathbf{Y}-\mathbf{D}^{(k)}\mathbf{C})\mathbf{W}^{(k)}\|_F^2+\lambda\|\mathbf{C}\|_{1,\mathbf{s}}. \tag{3}$$

有闭合解，每一列单独求解：

$$\mathbf{c}_i^{(k+1)} = \arg\min_{\mathbf{c}_i} \frac{1}{2}\|(\mathbf{y}_i - \mathbf{D}^{(k)}\mathbf{c}_i)\mathbf{W}_{ii}^{(k)}\|_2^2 + \lambda\|\mathbf{s}^T\mathbf{c}_i\|_1. \tag{4}$$

闭合解为：

$$\mathbf{c}_i^{(k+1)} = \mathrm{sgn}(\mathbf{D}^\top\mathbf{y}) \odot \max(|\mathbf{D}^\top\mathbf{y}| - \frac{\lambda}{(\mathbf{W}_{ii}^{(k)})^2}\mathbf{s}, 0), \tag{5}$$

b. update $\mathbf{D}$

$$\begin{aligned} \mathbf{D}^{(k+1)} = \arg\min_{\mathbf{D}} & \frac{1}{2}\|(\mathbf{Y} - \mathbf{D}\mathbf{C}^{(k+1)})\mathbf{W}^{(k)}\|_F^2 \\ \text{s.t.} \quad & \mathbf{D}^\top\mathbf{D} = \mathbf{I}. \end{aligned} \tag{6}$$

闭合解为：

$$\mathbf{D}^{(k+1)} = \mathbf{U}\mathbf{V}^\top \tag{7}$$

其中，$\mathbf{U}$和$\mathbf{V}$由$\mathbf{C}^{(k+1)}(\mathbf{W}^{(k)})^2\mathbf{Y}^\top = \mathbf{U}\mathbf{\Sigma}^{(k+1)}\mathbf{V}^\top$做SVD分解得到。$(\mathbf{W}^{(k)})^2$是指$\mathbf{W}$的对角线上的元素分别平方后得到的对角矩阵。

c. update $\mathbf{s}$

$\mathbf{s}^{k+1} = \mathbf{\Sigma}^{(k+1)}$

3.2. (3),(6)收敛后，更新$\mathbf{W}$

d. update $\mathbf{W}$

$\mathbf{W}$是对角矩阵，只需要更新迭代对角元即可：

$$\mathbf{W}_{ii}^{new} = \exp(-\lambda_2\|\mathbf{y}_i - \mathbf{D}^{(k+1)}\mathbf{c}_i^{(k+1)}\|_2) \tag{8}$$

λ_2是参数；

或

$$\mathbf{W}_{ii}^{new} = \lambda_2(\sigma_i - \|\mathbf{y}_i - \mathbf{D}^{(k+1)}\mathbf{c}_i^{(k+1)}\|_2) \tag{9}$$

或者其它可能性。

待说明问题：模型(1)或(2)的收敛性，收敛速率等问题，收敛值的上下界等问题。

实际中，我采用模型(2)，没有多次迭代（因为多次迭代速度太慢），权重的设计采用(8)，权重的物理意义，设计与效果值得探讨。

4. 孟老师的模型以及区别

孟老师的模型：

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{U}\mathbf{V}^\top)\|_1 \tag{10}$$

孟老师的模型是coordinate-wise的迭代，其不断迭代$\mathbf{U}, \mathbf{V}$直到收敛。这个模型应用在Low-Rank Matrix Factorization with Missing Entries这个问题里，$\mathbf{W}$指矩阵的元素是否是missing entry，其元素只有0或1。

区别：

- 权重的设计和迭代方式不同，物理含义不一样。孟老师的模型里的权重主要是指矩阵的元素是否是missing entry，也就是$\mathbf{W}$里只有0或1；
- 模型不同；
- 应用不同；

5. 杨猛的模型以及区别

杨猛的模型：

$$\min_{\mathbf{c}} \|\mathbf{W}^{\frac{1}{2}}(\mathbf{y} - \mathbf{D}\mathbf{c})\|_2^2 \quad \text{s.t.} \quad \|\mathbf{c}\|_1 < \sigma \tag{11}$$

区别：

- 关于加权，我的方法里，F范数的权重是加在列（sample）上，但是杨猛是加在向量2范的每一行（variable）上；设计出发点不同（我的模型是从weighted orthogonal Procrustes problem里的模型里演化过来的）；权重迭代方式不同；
- 我的模型里，字典是正交的，是通过数据学到的，有闭合解；杨的方法是训练集里的sample直接做字典；
- 我的模型里系数1范的权重是加在每一行上，所以，相当于我的模型里，每行每列的每个元素都带权重；而且系数有闭合解。