

Ps. X. 小结:

#### IV. 贪婪策略算法

A. 匹配追踪(Matching Pursuit)算法 (ps. 不涉及系数值, 只涉及索引)

B. 正交匹配追踪算法 (ps. 涉及具体索引的系数值)

#### V. 包含约束的优化策略

A. 基于梯度映射的稀疏重构 (GPSR) (ps. 系数正负分割 + 梯度下降)

B. 基于牛顿法的内点法模型的稀疏重构

C. 基于方向交替法 (ADM) 的稀疏表示策略 (ps. 无约束到约束, 然后对增广拉格朗日用交替策略求解)

#### VI. 基于邻近算法的优化策略

A. 软阈值与收缩算子

B. 迭代收缩阈值算法 (ISTA) (ps. 通过泰勒近似转化到标准的 VI. 6 形式, 并套用收缩算子求解)

C. 快速迭代收缩算法(FISTA) (ps. 将 ISTA 中的黑塞矩阵的近似值改为其最大值, i.e.  $L(f) = 2\lambda_{\max}(X^T X)$ )

D. 通过可分离近似来实现稀疏重构 (SpaRSA) (ps. 通过参数设计, 实现 ISTA 的改进)

#### E. 对 $l_{1/2}$ 范数正则化的稀疏优化

F. 从基于增广拉格朗日数乘的优化策略 (ALM)  $\rightarrow$  对偶形式的增广拉格朗日方法 (DALM)

#### VII. 基于同伦算法的稀疏表示

A. Lasso 同伦算法 (ps. 不断减小  $\lambda$ , 利用子梯度, 求到临界点正负最小步长, 更新同伦参数与支持集合)

B. 基追踪去噪 (BPDN) 的同伦算法 (ps. 直接计算同伦解  $\alpha$ )

C. 基于同伦算法的交替再加权的  $l_{1/2}$  范数最小化 (ps. 从权重  $w_i$  过渡到权重  $\hat{w}_i$ )

#### VIII. 稀疏表示方法的应用 (字典学习, 图像处理, 图像分类, 目标跟踪)

##### A. 字典学习中的稀疏表示

###### 1. 无监督字典学习

a. 基于 KSVD 的无监督字典学习 ( $l_0$  范数, 大环: OMP 求系数位置与值; 小环: KSVD 求字典和仅更新系数值)

b. 基于局部约束线性编码的无监督字典学习 (ps. 局部必稀疏, 稀疏未必局部, 以此建立  $l_2$  约束)

###### 2. 监督的字典学习 (ps. 训练中字典更新 with KSVD + 分类中稀疏表示求解 with OMP)

a. 基于判别性 KSVD (DKSVD) 的字典学习 (ps. 字典学习+稀疏表示+分类模型  $\rightarrow$  整合到 KSVD 模型)

b. 基于标签一致性的 KSVD (LC-KSVD) 的判别性字典学习 (ps. D 和 Y 的标签一致性矩阵 L; KSVD 模型)

c. Fisher 判别性字典学习 (FDDL) (ps. 引入标签关联信息到  $f$ , 引入 Fisher 信息到  $g$ , 按类更新系数和字典)

##### B. 基于稀疏表示的图像处理

1. 基于稀疏表示的图像超分辨率重构 (SRSR) (ps. 块; 低分辨率重构+高分辨率邻近一致; 全局一致性增强)

2. 基于字典学习的稀疏表示模型 (SRMLD) 的图像去噪 (ps. 全局极大似然约束+局部重构的贝叶斯先验约束)

3. 基于集中稀疏表示 (CSR) 模型的单幅退化图像恢复 (ps. 块; 退化图像重构+原始图像稀疏系数期望近似)

C. 基于稀疏表示的图像分类 (ps. 归类到稀疏表示的残差最小的一类)

# 稀疏表示概述的简单梳理

## 摘要:

文章将稀疏表示算法分为了 4 大类做了详细阐述, 分别为 1 无约束的基于贪婪策略的优化, 2 含有约束的优化, 3 基于邻近算法的优化策略, 4 基于同伦算法的优化策略, 并从字典学习, 图像处理等角度给予了实例。

## I. 介绍:

稀疏表示(SR)和压缩感知(CS)关系密切。压缩感知理论表示, 如果信号本质上是稀疏或者可压缩的, 那么原始信号就可以通过少量的基来进行重构。CS 理论通常包括 3 大块, 稀疏表示, 编码度量, 重构算法。

稀疏表示应用很多。图像分类就是一个例子。自然图像可以通过一些视觉单元的进行系数表示。基于稀疏表示的分类(SRC)就是利用不同类的训练样本进行测试样本的稀疏表示, 然后计算样本在不同类中的重构残差。然后测试样本会被分类到残差最小的一类中

稀疏表示包含了信号抽样和编码压缩的过程。

### A. 稀疏表示的分类

不同分类基准产生不同分类:

通过 atom 的类型来分: 基于自然样本的重构, 基于字典的重构;

通过 atom 的标签来分: 监督, 半监督, 无监督;

通过稀疏约束来分: 结构约束的, 稀疏约束的

通过利用 atom 的方法来分: 基于所有类别样本的稀疏, 基于局部类的样本的稀疏

通过稀疏模型和解决方法分: 贪婪算法, 凸松弛优化

通过对 l1 范数的优化方法分: 子梯度优化, 无约束的近似优化, 约束优化

本文从稀疏表示的模型求解方法来分:

1. 贪婪算法: 主要解决 l0 范数的优化问题, 通过迭代, 每次获得一个最优的局部系数。并且最终只获得最优的 k 个系数。
2. 基于约束的优化策略: 核心是通过一个可导、凸、平滑的优化项, 来代替 l1 范数的最小化项。
3. 基于邻近算法的优化: 是一个解决非平滑, 约束的, 大规模的或者分布式的优化问题的有效策略。主要任务是将原始问题规范化到一个特殊的模型, 这个模型可以用邻近算子求解。邻近算子通常包括软阈值算子、硬阈值算子、预解算子。(ps.每次对系数向量的每一位单独/分开处理, 收缩算法)
4. 同伦算法: 通过调整同伦参数, 不断逼近最优解。(ps.维持两个不同功能的集合, 分别更新)

## II. 规范及范数

样本维度 d, 样本数量 N, 样本矩阵 X, 字典大小 n (n>>d, 过完备字典), 字典矩阵 D。

p 范数:

$$\|v\|_p = \left( \sum_{i=1}^n |v_i|^p \right)^{1/p}$$

p=0, 非零样本的数量: 非凸, 非平滑, 不连续, 全局不可导

0<p<1, 以 p=1/2 居多: 非凸, 非平滑, 全局不可导

p=1, 绝对值之和: 凸, 非平滑, 连续, 全局不可导

p=2, 欧式范数: 凸, 平滑, 连续, 全局可导 (不是严格稀疏, 但实际有效)

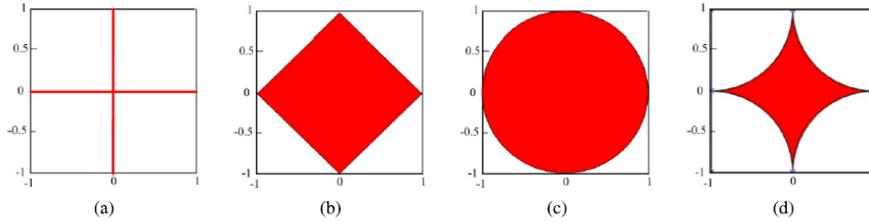


FIGURE 2. Geometric interpretations of different norms in 2-D space [7]. (a), (b), (c), (d) are the unit ball of the  $l_0$ -norm,  $l_1$ -norm,  $l_2$ -norm,  $l_p$ -norm ( $0 < p < 1$ ) in 2-D space, respectively. The two axes of the above coordinate systems are  $x_1$  and  $x_2$ .

### III. 不同范数的稀疏表示公式

#### A. 基于 $l_0$ 范数的稀疏表示:

最稀疏:

$$\hat{\alpha} = \arg \min \|y - X\alpha\|_2^2 \quad s.t. \quad \|\alpha\|_0 \leq \varepsilon \quad (III.7)$$

$$\hat{\alpha} = L(\alpha, \lambda) = \arg \min \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_0 \quad (III.8)$$

#### B. 基于 $l_1$ 范数的稀疏表示:

Lasso 问题:

$$\hat{\alpha} = \arg \min_{\alpha} \|y - X\alpha\|_2^2 \quad s.t. \quad \|\alpha\|_1 \leq \tau \quad (III.11)$$

$$\hat{\alpha} = L(\alpha, \lambda) = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (III.12)$$

#### C. 基于 $l_p$ ( $0 < p < 1$ ) 范数的稀疏:

$$\hat{\alpha} = L(\alpha, \lambda) = \arg \min_{\alpha} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_p^p \quad (III.14)$$

#### D. 基于 $l_2$ 范数和 $l_2.1$ 范数的稀疏:

$l_2$  范数是有限稀疏,  $l_2$  范数的解有判别性, 但不是真正意义上的稀疏。

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_2^2 \quad s.t. \quad \|y - X\alpha\|_2^2 \leq \varepsilon \quad (III.15)$$

$$\hat{\alpha} = L(\alpha, \lambda) = \arg \min_{\alpha} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_2^2 \quad (III.16)$$

### IV. 贪婪策略算法

由于  $l_0$  范数问题是没有多项式复杂度的解, 所以通常通过贪婪算法求解。即, 每一步获得一个局部最优解, 从而获得近似地全局最优解。

#### A. 匹配追踪(Matching Pursuit)算法 (ps.不涉及系数值, 只涉及索引)

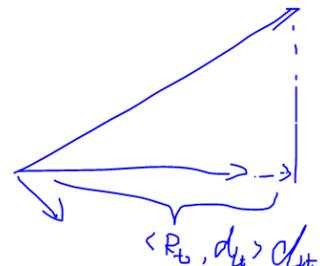
MP 算法主要根据一个相似性度量的目标函数, 通过迭代, 每一步从字典中选择一个当前最优的 atom 作为局部最优解。

假设近似目标为  $y$ , 用  $R$  表示稀疏表示的残差。那我们就有:

$$R_0 = y, D = [d_1, d_2, \dots, d_N] \in \mathbb{R}^{d \times N}$$

根据 MP 的思想, 我们的目标是:

$$|\langle R_0, d_{l_0} \rangle| = \sup |\langle R_0, d_i \rangle| \quad (IV.1)$$



$$y = \langle y, d_{l_0} \rangle d_{l_0} + R_1 \quad (IV.2)$$

并且把当前步映射后的残差当作下一步的初始残差。因而，在  $t$  时刻的当前最优解为：

$$|\langle R_t, d_{l_t} \rangle| = \sup |\langle R_t, d_i \rangle| \quad (IV.5)$$

显然， $d_{l_t}$  是与  $R_{t+1}$  垂直的。不断迭代，直到最终残差到达可接受的阈值范围内，我们获得 ( $n \ll N$ ):

$$y = \sum_{j=1}^{n-1} \langle R_j, d_{l_j} \rangle d_{l_j} + R_n$$

## B. 正交匹配追踪算法 (ps.涉及具体索引的系数值)

**Algorithm 1** Orthogonal Matching Pursuit Algorithm

**Task:** Approximate the constraint problem:

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_0 \quad s.t. \quad y = X\alpha$$

**Input:** Probe sample  $y$ , measurement matrix  $X$ , sparse coefficients vector  $\alpha$

**Initialization:**  $t = 1$ ,  $r_0 = y$ ,  $\alpha = 0$ ,  $D_0 = \phi$ , index set  $\Lambda_0 = \phi$  where  $\phi$  denotes empty set,  $\tau$  is a small constant.

While  $\|r_t\| > \tau$  do

Step 1: Find the best matching sample, i.e. the biggest inner product between  $r_{t-1}$  and  $x_j$  ( $j \notin \Lambda_{t-1}$ ) by exploiting  $\lambda_t = \arg \max_{j \notin \Lambda_{t-1}} |\langle r_{t-1}, x_j \rangle|$ .

Step 2: Update the index set  $\Lambda_t = \Lambda_{t-1} \cup \lambda_t$  and reconstruct data set  $D_t = [D_{t-1}, x_{\lambda_t}]$ .

Step 3: Compute the sparse coefficient by using the least square algorithm  $\tilde{\alpha} = \arg \min \|y - D_t \alpha\|_2^2$ .

Step 4: Update the representation residual using  $r_t = y - D_t \tilde{\alpha}$ .

Step 5:  $t = t + 1$ .

End

**Output:**  $D, \alpha$

## V. 包含约束的优化策略

约束优化策略主要是用来获得包含  $l_1$  范数正则项的稀疏表示解。它通过把非平滑不可导的  $l_1$  范数项问题转化为平滑可导的约束优化问题。

### A. 基于梯度映射的稀疏重构 (GPSR) (ps. 系数正负分割 + 梯度下降)

GPSR 算法是一种基于梯度下降的优化算法。针对  $l_1$  范数，GPSR 将稀疏的系数值分为+和-的两部分

$$\alpha = \alpha_+ - \alpha_-, \quad \alpha_+ \geq 0, \quad \alpha_- \geq 0 \quad (V.1)$$

于是， $l_1$  范数问题就转化为：

$$\arg \min L(\alpha) = \arg \min \frac{1}{2} \|y - [X_+, X_-][\alpha_+ - \alpha_-]\|_2^2 + \lambda (1_d^T \alpha_+ + 1_d^T \alpha_-) \quad s.t. \quad \alpha_+ \geq 0, \alpha_- \geq 0$$

进一步，可以转化为：

$$\arg \min G(z) = c^T z + \frac{1}{2} z^T A z \quad s.t. \quad z \geq 0 \quad (V.4)$$

where  $z = [\alpha_+; \alpha_-]$ ,  $c = \lambda 1_{2d} + [-X^T y; X^T y]$ ,  $1_{2d} = [1, \dots, 1]^T$ ,  $A = \begin{pmatrix} X^T X & -X^T X \\ -X^T X & X^T X \end{pmatrix}$ .

使用梯度下降方法以及 line search, 我们有:

$$\arg \min z^{t+1} = z^t - \sigma \nabla G(z^t) \quad \sigma^t = \arg \min G(z^t - \sigma g^t)$$

特别地, 为了确保  $z > 0$ , 上式中  $g^t$  有一个特殊的条件:  $g_i^t = \begin{cases} (\nabla G(z^t))_i, & \text{if } z_i^t > 0 \text{ or } (\nabla G(z^t))_i < 0 \\ 0, & \text{otherwise.} \end{cases}$   
最终我们有:

$$\sigma^t = \frac{(g^t)^T (g^t)}{(g^t)^T A (g^t)} \quad (V.8)$$

同时我们依据 backtracking linear search 的步长更新条件对步长更新给出约束条件

$$G((z^t - \sigma^t \nabla G(z^t))_+) > G(z^t) - \beta \nabla G(z^t)^T \times (z^t - (z^t - \sigma^t \nabla G(z^t))_+) \quad (V.9)$$

**Algorithm 2** Gradient Projection Sparse Reconstruction (GPSR)

**Task:** To address the unconstrained problem:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1$$

**Input:** Probe sample  $y$ , the measurement matrix  $X$ , small constant  $\lambda$

**Initialization:**  $t = 0$ ,  $\beta \in (0, 0.5)$ ,  $\gamma \in (0, 1)$ , given  $\alpha$  so that  $z = [\alpha_+, \alpha_-]$ .

**While** not converged **do**

Step 1: Compute  $\sigma^t$  exploiting Eq. V.8 and  $\sigma^t \leftarrow \text{mid}(\sigma_{min}, \sigma^t, \sigma_{max})$ , where  $\text{mid}(\cdot, \cdot, \cdot)$  denotes the middle value of the three parameters.

Step 2: **While** Eq. V.9 not satisfied **do**  $\sigma^t \leftarrow \gamma \sigma^t$  **end**

Step 3:  $z^{t+1} = (z^t - \sigma^t \nabla G(z^t))_+$  and  $t = t + 1$ .

**End**

**Output:**  $z^{t+1}, \alpha$

**B. 基于牛顿法的内点法模型的稀疏重构**

内点法不是一个迭代算法, 但是是一个有效的模型。它内部集成了牛顿法, 从而有效地解决无约束的平滑优化问题。

主要思路: 无约束非平滑 - (转化) -> 约束平滑 - (内点法) -> 无约束平滑 - (牛顿法) -> 最优解

由于  $|\alpha| = \arg \min_{-\sigma \leq \alpha \leq \sigma} \sigma$ ,  $l_1$  正则化式子  $\hat{\alpha} = L(\alpha, \lambda) = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1$  可以写为:

1.5!  
min  $\alpha = \delta$   
-  $\alpha < + \alpha$

$$\begin{aligned} \hat{\alpha} &= \arg \min \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1 \\ &= \arg \min \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \sum_{i=1}^N \min_{-\sigma_i \leq \alpha_i \leq \sigma_i} \sigma_i \\ &= \arg \min \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \sum_{i=1}^N \min_{-\sigma_i \leq \alpha_i \leq \sigma_i} \sigma_i \\ &= \arg \min_{-\sigma_i \leq \alpha_i \leq \sigma_i} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \sum_{i=1}^N \sigma_i \quad (V.11) \end{aligned}$$

(ps1. 从无约束非平滑到约束平滑) -- 约束:  $s.t. \sigma_i + \alpha_i \geq 0, \sigma_i - \alpha_i \geq 0$

为了将上述“约束问题”转为“非约束问题”, 引入一个 barrier function  $B$ , 故有:

$$\hat{\alpha} = \arg \min_{\alpha, \sigma \in \mathbb{R}^N} G(\alpha, \sigma) = \frac{\nu}{2} \|y - X\alpha\|_2^2 + \lambda \nu \sum_{i=1}^N \sigma_i - B(\alpha, \sigma) \quad (\text{V.14})$$

$$B(\alpha, \sigma) = \sum_{i=1}^N \log(\sigma_i + \alpha_i) + \sum_{i=1}^N \log(\sigma_i - \alpha_i)$$

(ps2. 约束平滑到无约束平滑)

下面，就是解决一个常见的牛顿问题了。

1. 构建牛顿系统的等式，求得  $[\Delta\alpha, \Delta\sigma]$ ;

$$H \begin{bmatrix} \Delta\alpha \\ \Delta\sigma \end{bmatrix} = -\nabla G(\alpha, \sigma) \in \mathbb{R}^{2N} \quad (\text{V.15})$$

$$G(\alpha + \nu) = G(\alpha) + \nu G'(\alpha) + \frac{1}{2} \nu^2 G''(\alpha) \nu$$

$$\Rightarrow \frac{\partial G(\alpha + \nu)}{\partial \nu} = \nu G'(\alpha) + \nu^2 G''(\alpha) \nu = 0$$

其中H为黑塞矩阵 (2Nx2N)，通过预条件共轭梯度 (preconditioned conjugate gradient) 算法求得。

2. 构建  $l_1$  范数优化函数的对偶可行解和对偶鸿沟:

a).  $l_1$  函数的拉格朗日函数为:

$$L(\alpha, z, u) = z^T z + \lambda \|\alpha\|_1 + u(X\alpha - y - z) \quad (\text{V.16})$$

$$\frac{1}{2} \|z\|_2^2 + \lambda \|\alpha\|_1$$

s.t.  $y - X\alpha = z$

于是拉格朗日对偶函数为:

$$\hat{\alpha} = \arg \max F(u) = -\frac{1}{4} u^T u - u^T y \quad \left\{ \begin{array}{l} z = -u \\ \text{s.t. } |(X^T u)_i| \leq \lambda_i \quad (i = 1, 2, \dots, N) \end{array} \right. \quad (\text{V.17})$$

b). 相应的步长 s 和对偶可行解 u 为:

$$u = 2s(y - X\alpha), \quad s = \min\{\lambda/|2y_i - 2(X^T X\alpha)_i|\} \forall i \quad (\text{V.18})$$

c). 原始问题和对偶问题之间的对偶鸿沟为:

$$g = \|y - X\alpha\| + \lambda \|\alpha\|_1 - F(u) \quad (\text{V.19})$$

3. 同样的，牛顿法使用的 backtracking linear search 的终止条件为 ( $\eta$  为牛顿线性搜索的步长):

$$G(\alpha + \eta^t \Delta\alpha, \sigma + \eta^t \Delta\sigma) > G(\alpha, \sigma) + \rho \eta^t \nabla G(\alpha, \sigma) [\Delta\alpha, \Delta\sigma] \quad (\text{V.20})$$

4. 牛顿线性搜索的终止约束条件为:

$$\zeta = \min\{0.1, \beta g / \|h\|_2\}$$

其中  $h = \nabla G(\alpha, \sigma)$ ,  $\beta$  is a small constant

**Algorithm 3** Truncated Newton Based Interior-Point Method (TNIPM) for  $l_1$ - $l_2$

**Task:** To address the unconstrained problem:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1$$

**Input:** Probe sample  $y$ , the measurement matrix  $X$ , small constant  $\lambda$

**Initialization:**  $t = 1, \nu = \frac{1}{\lambda}, \rho \in (0, 0.5), \sigma = 1_N$

Step 1: Employ preconditioned conjugate gradient algorithm to obtain the approximation of  $H$  in Eq. V.15, and then obtain the descent direction of linear search  $[\Delta\alpha^t, \Delta\sigma^t]$ .

Step 2: Exploit the algorithm of backtracking linear search to find the optimal step size of Newton linear search  $\eta^t$ , which satisfies the Eq. V.20.

Step 3: Update the iteration point utilizing  $(\alpha^{t+1}, \sigma^{t+1}) = (\alpha^t, \sigma^t) + (\Delta\alpha^t + \Delta\sigma^t)$ .

Step 4: Construct feasible point using eq. V.18 and duality gap in Eq. V.19, and compute the termination tolerance  $\zeta$  in Eq. V.21.

Step 5: If the condition  $g/F(u) \leq \zeta$  is satisfied, stop; Otherwise, return to step 1, update  $\nu$  in Eq. V.14 and  $t = t + 1$ .

**Output:**  $\alpha$

C. 基于方向交替法 (ADM) 的稀疏表示策略 (ps. 无约束到约束, 然后对增广拉格朗日用交替策略求解)

拉格朗日乘法的目标函数是:  $P(x, \alpha) = f(x) - \alpha * g(x)$

增广拉格朗日乘法的目标函数是:  $P(x, \alpha, r) = f(x) - \alpha * g(x) + r/2 * g(x)^2$

增广拉格朗日乘子法的求解: 从一个相对较小的罚函数因子  $u$  和选定的初始的乘子  $\lambda$  出发, 在迭代中不断地求出  $x$ , 在把因子  $u$  调大, 同时更新乘子  $\lambda$  参数使其逼近最优的乘子  $\lambda^*$ 。然后在罚函数因子  $u$  变得非常大之前, 一般就能把最优解得出来了。

引入辅助变量  $s$ , 将无约束的优化问题转化为约束的优化问题。

$$\arg \min_{\alpha, s} \frac{1}{2\tau} \|s\|_2 + \|\alpha\|_1 \quad s.t. \quad s = y - X\alpha \quad (V.22)$$

ADM 使用增广拉格朗日乘子进行进一步改进目标函数:

$$\arg \min_{\alpha, s, \lambda} L(\alpha, s, \lambda) = \frac{1}{2\tau} \|s\|_2 + \|\alpha\|_1 - \lambda^T \times (s + X\alpha - y) + \frac{\mu}{2} \|s + X\alpha - y\|_2^2$$

然后 ADM 通过交替计算的方法求解上述问题:

$$\begin{cases} s^{t+1} = \arg \min L(s, \alpha^t, \lambda^t) & (a) \\ \alpha^{t+1} = \arg \min L(s^{t+1}, \alpha, \lambda^t) & (b) \\ \lambda^{t+1} = \lambda^t - \mu(s^{t+1} + X\alpha^{t+1} - y) & (c) \end{cases} \quad (V.24)$$

其中, (a) 式的  $s$  有如下解:

$$\begin{aligned} \arg \min L(s, \alpha^t, \lambda^t) &= \frac{1}{2\tau} \|s\|_2 + \|\alpha^t\|_1 - (\lambda^t)^T \times (s + X\alpha^t - y) + \frac{\mu}{2} \|s + X\alpha^t - y\|_2^2 \\ &= \frac{1}{2\tau} \|s\|_2 - (\lambda^t)^T s + \frac{\mu}{2} \|s + X\alpha^t - y\|_2^2 + \|\alpha^t\|_1 - (\lambda^t)^T (X\alpha^t - y) \end{aligned} \quad (V.25) \rightarrow s^{t+1} = \frac{\tau}{1 + \mu\tau} (\lambda^t - \mu(y - X\alpha^t)) \quad (V.26)$$

(b) 式的  $\alpha$  有如下解:

$$\begin{aligned} \arg \min L(s^{t+1}, \alpha, \lambda^t) &= \frac{1}{2\tau} \|s^{t+1}\|_2 + \|\alpha\|_1 - (\lambda^t)^T (s^{t+1} + X\alpha - y) + \frac{\mu}{2} \|s^{t+1} + X\alpha - y\|_2^2 \\ &= \|\alpha\|_1 + \frac{\mu}{2} \|s^{t+1} + X\alpha - y - \lambda^t/\mu\|_2^2 \end{aligned} \rightarrow \arg \min \{ \|\alpha\|_1 - (\lambda^t)^T (s^{t+1} + X\alpha - y) + \frac{\mu}{2} \|s^{t+1} + X\alpha - y\|_2^2 \} = \|\alpha\|_1 + f(\alpha) \quad (V.27)$$

用二阶泰勒展开近似  $f(\alpha)$ , 并引入软阈值算子:

$$\arg \min \{ \|\alpha\|_1 + (\alpha - \alpha^t)^T X^T (s^{t+1} + X\alpha^t - y - \lambda^t/\mu) + \frac{1}{2\tau} \|\alpha - \alpha^t\|_2^2 \} \quad (V.28) \quad \alpha^{t+1} = \text{soft} \{ \alpha^t - \tau X^T (s^{t+1} + X\alpha^t - y - \lambda^t/\mu), \frac{\tau}{\mu} \} \quad (V.29)$$

**Algorithm 4** Alternating Direction Method (ADM) Based Sparse Representation Strategy

**Task:** To address the unconstrained problem:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \tau \|\alpha\|_1$$

**Input:** Probe sample  $y$ , the measurement matrix  $X$ , small constant  $\lambda$

**Initialization:**  $t = 0, s^0 = 0, \alpha^0 = 0, \lambda^0 = 0$  ( $\tau = 1.01, \mu$  is a small constant.

Step 1: Construct the constraint optimization problem of problem III.12 by introducing the auxiliary parameter

and its augmented Lagrangian function, i.e. problem (V.22) and (V.23).

While not converged do

Step 2: Update the value of the  $s^{t+1}$  by using Eq. (V.25).

Step 2: Update the value of the  $\alpha^{t+1}$  by using Eq. (V.29).

Step 3: Update the value of the  $\lambda^{t+1}$  by using Eq. (V.24(c)).

Step 4:  $\mu^{t+1} = \tau \mu^t$  and  $t = t + 1$ .

End While

**Output:**  $\alpha^{t+1}$

## VI. 基于邻近算法的优化策略

利用邻近算子 (proximal operator), 通过迭代的方法, 解决原始问题的一系列子问题。

### A. 软阈值与收缩算子

对于如下形式的简单  $l_1$  范数问题有 shrinkage operator 的通解:

VI.6

解的目标函数形式

$$\alpha^* = \min_{\alpha} h(\alpha) = \lambda \|\alpha\|_1 + \frac{1}{2} \|\alpha - s\|^2$$

$$= \sum_{j=1}^N \lambda |\alpha_j| + \sum_{j=1}^N \frac{1}{2} (\alpha_j - s_j)^2$$

where  $\alpha^*$  is the optimal solution of problem VI.6

(1) if  $\alpha_j > 0$ , then  $h(\alpha) = \lambda\alpha + \frac{1}{2}\|\alpha - s\|^2$  and its derivative is  $h'(\alpha_j) = \lambda + \alpha_j^* - s_j$ .  
Let  $h'(\alpha_j) = 0 \Rightarrow \alpha_j^* = s_j - \lambda$ , where it indicates  $s_j > \lambda$ ;

(2) if  $\alpha_j < 0$ , then  $h(\alpha) = -\lambda\alpha + \frac{1}{2}\|\alpha - s\|^2$  and its derivative is  $h'(\alpha_j) = -\lambda + \alpha_j^* - s_j$ .  
Let  $h'(\alpha_j) = 0 \Rightarrow \alpha_j^* = s_j + \lambda$ , where it indicates  $s_j < -\lambda$ ;

(3) if  $-\lambda \leq s_j \leq \lambda$ , and then  $\alpha_j^* = 0$ .

$$\alpha_j^* = \begin{cases} s_j - \lambda, & \text{if } s_j > \lambda \\ s_j + \lambda, & \text{if } s_j < -\lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{or} \quad \alpha^* = \text{shrink}(s, \lambda), \text{ where the } j\text{-th component is } \text{shrink}(s, \lambda)_j = \text{sign}(s_j) \max\{|s_j| - \lambda, 0\}$$

### B. 迭代收缩阈值算法 (ISTA) (ps. 通过泰勒近似转化到标准的 VI. 6 形式, 并套用收缩算子求解)

通常的稀疏表示的目标函数如下:

$$\arg \min F(\alpha) = \frac{1}{2} \|X\alpha - y\|_2^2 + \lambda \|\alpha\|_1 = f(\alpha) + \lambda g(\alpha)$$

ISTA 通过将它转化为 VI. 6 形式的简单  $l_1$  范数形式来实现求解。

1. 用泰勒展开近似  $f(\alpha)$ , 先将希望求解的  $\alpha^{t+1}$  从  $l_2$  范数的项中转化出来:

$$f(\alpha) = f(\alpha^t) + (\alpha - \alpha^t)^T \nabla f(\alpha^t) + \frac{1}{2} (\alpha - \alpha^t)^T H_f(\alpha^t) (\alpha - \alpha^t) + \dots, \text{ 其中 } \nabla f(\alpha) = X^T(X\alpha - y) \text{ and } H_f(\alpha) = X^T X$$

$$f(\alpha) = \frac{1}{2} \|X\alpha^t - y\|_2^2 + (\alpha - \alpha^t)^T X^T(X\alpha^t - y) + \frac{1}{2} (\alpha - \alpha^t)^T X^T X (\alpha - \alpha^t) \quad (\text{VI.10})$$

→ 同构

2. 通过不同方法来实现对  $H_f(\alpha)$  的近似, 获得不同的求解。

$$\begin{aligned} Q_t(\alpha, \alpha^t) &= \frac{1}{2} \|X\alpha^t - y\|_2^2 + (\alpha - \alpha^t)^T X^T(X\alpha^t - y) \\ &\quad + \frac{1}{2\tau} \|\alpha - \alpha^t\|_2^2 \\ &= \frac{1}{2} \|X\alpha^t - y\|_2^2 + \frac{1}{2\tau} \|\alpha - \alpha^t + \tau X^T(X\alpha^t - y)\|_2^2 \\ &\quad - \frac{\tau}{2} \|X^T(X\alpha^t - y)\|_2^2 \\ &= \frac{1}{2\tau} \|\alpha - (\alpha^t - \tau X^T(X\alpha^t - y))\|_2^2 + B(\alpha^t) \end{aligned}$$

If the Hessian matrix  $H_f(\alpha)$  is replaced or approximated in the third term above by using a scalar  $\frac{1}{\tau}I$ , and then

$$f(\alpha) \approx \frac{1}{2} \|X\alpha^t - y\|_2^2 + (\alpha - \alpha^t)^T X^T(X\alpha^t - y) + \frac{1}{2\tau} (\alpha - \alpha^t)^T (\alpha - \alpha^t) = Q_t(\alpha, \alpha^t) \quad (\text{VI.11}) \rightarrow$$

$$\alpha^{t+1} = \arg \min \frac{1}{2\tau} \|\alpha - \theta(\alpha^t)\|_2^2 + \lambda \|\alpha\|_1 \quad (\text{VI.14})$$

→ where  $\theta(\alpha^t) = \alpha^t - \tau X^T(X\alpha^t - y)$ . (ps. 用  $\alpha^{t+1} = \text{shrink}(\theta(\alpha^t), \lambda\tau)$  实现求解)

### C. 快速迭代收缩算法(FISTA) (ps. 将 ISTA 中的黑塞矩阵的近似值改为其最大值, i.e. $L(f) = 2\lambda_{\max}(X^T X)$ )

$$f(\alpha) \approx \frac{1}{2} \|X\alpha^t - y\|_2^2 + (\alpha - \alpha^t)^T X^T(X\alpha^t - y) + \frac{L}{2} (\alpha - \alpha^t)^T (\alpha - \alpha^t) = P_t(\alpha, \alpha^t) \quad (\text{VI.15}) \rightarrow \text{where } \theta(\alpha^t) = \alpha^t - \frac{1}{L} X^T(X\alpha^t - y), \quad \alpha^{t+1} = \arg \min \frac{L}{2} \|\alpha - \theta(\alpha^t)\|_2^2 + \lambda \|\alpha\|_1 \quad (\text{VI.16})$$

同时, FISTA 为进一步算法的收敛速度, 有  $\alpha^t = \alpha^t + \frac{\mu^t - 1}{\mu^{t+1}}(\alpha^t - \alpha^{t-1})$  (VI.17)

**Algorithm 5** Fast Iterative Shrinkage Thresholding Algorithm (FISTA)

**Task:** To address the problem  $\hat{\alpha} = \arg \min F(\alpha) = \frac{1}{2}\|X\alpha - y\|_2^2 + \lambda\|\alpha\|_1$

**Input:** Probe sample  $y$ , the measurement matrix  $X$ , small constant  $\lambda$

**Initialization:**  $t = 0, \mu^0 = 1, L = 2\Lambda_{\max}(X^T X)$ , i.e. Lipschitz constant of  $\nabla f$ .

While not converged do

Step 1: Exploit the shrinkage operator in equation VI.7 to solve problem VI.16.

Step 2: Update the value of  $\mu$  using  $\mu^{t+1} = \frac{1 + \sqrt{1 + 4(\mu^t)^2}}{2}$ .

Step 3: Update iteration sequence  $\alpha^t$  using equation VI.17

End

**Output:**  $\alpha$

D. 通过可分离近似来实现稀疏重构 (SpaRSA) (ps. 通过参数设计, 实现 ISTA 的改进)

SpaRSA 可以看做是一个 ISTA 的改进版本, 它通过 ISTA 中的优化参数  $\lambda$  (by warm-starting technique) 和构建更加有效的黑塞矩阵  $H_f(\alpha)$  的近似值 (by Barzilai-Borwein spectral method) 来实现。

1. 通过 Worm-Starting Technique 来优化参数  $\lambda$

stam. However, Hale *et al.* [86] concluded that the technique that exploits a decreasing value of  $\lambda$  from a warm-starting point can more efficiently solve the sub-problem VI.14 than  $\rightarrow$  where  $\gamma$  is a small constant.

$$\lambda = \max\{\gamma \|X^T y\|_\infty, \lambda\}$$

$$\alpha^{t+1} = y - X\alpha^{t+1}$$

2. BB Spectral method 方法来近似黑塞矩阵值  $H_f(\alpha)$

$$\frac{1}{\tau^{t+1}} = \arg \min \frac{1}{\tau} (\alpha^{t+1} - \alpha^t) - (\nabla f(\alpha^{t+1}) - \nabla f(\alpha^t))\|_2^2$$

$$= \frac{(\alpha^{t+1} - \alpha^t)^T (\nabla f(\alpha^{t+1}) - \nabla f(\alpha^t))}{(\alpha^{t+1} - \alpha^t)^T (\alpha^{t+1} - \alpha^t)} \quad (VI.20)$$

For problem VI.14, SpaRSA requires that the value of  $\lambda$  is a decreasing sequence using the Eq. VI.18 and the value of  $\tau$  should meet the condition of Eq. VI.20. The

**Algorithm 6** Sparse Reconstruction by Separable Approximation (SpaRSA)

**Task:** To address the problem

$$\hat{\alpha} = \arg \min F(\alpha) = \frac{1}{2}\|X\alpha - y\|_2^2 + \lambda\|\alpha\|_1$$

**Input:** Probe sample  $y$ , the measurement matrix  $X$ , small constant  $\lambda$

**Initialization:**  $t = 0, i = 0, y^0 = y, \frac{1}{\tau^0} I \approx H_f(\alpha) = X^T X$ , tolerance  $\varepsilon = 10^{-5}$ .

Step 1:  $\lambda_t = \max\{\gamma \|X^T y^t\|_\infty, \lambda\}$ .

Step 2: Exploit shrinkage operator to solve problem VI.14, i.e.  $\alpha^{i+1} = \text{shrink}(\alpha^i - \tau^i X^T (X^T \alpha^i - y), \lambda_t \tau^i)$ .

Step 3: Update the value of  $\frac{1}{\tau^{i+1}}$  using the Eq. VI.20.

Step 4: If  $\frac{\|\alpha^{i+1} - \alpha^i\|}{\alpha^i} \leq \varepsilon$ , go to step 5; Otherwise, return to step 2 and  $i = i + 1$ .

Step 5:  $y^{t+1} = y - X\alpha^{i+1}$ .

Step 6: If  $\lambda_t = \lambda$ , stop; Otherwise, return to step 1 and  $t = t + 1$ .

**Output:**  $\alpha^i$

E. 对 11/2 范数正则化的稀疏优化

F. 从基于增广拉格朗日数乘的优化策略 (ALM) → 对偶形式的增广拉格朗日方法 (DALM)

—ALM: 对于目标函数, 额外引入一个等式约束到其拉格朗日函数中进行模型改进。

$$L(\alpha, \lambda) = \|\alpha\|_1 + \frac{\lambda}{2} \|y - X\alpha\|_2^2 \quad s.t. \quad y - X\alpha = 0 \quad \rightarrow \quad \arg \min L_\lambda(\alpha, z) = \|\alpha\|_1 + \frac{\lambda}{2} \|y - X\alpha\|_2^2 + z^T (y - X\alpha)$$

然后和方向交替算法 (ADM) 类似, 有如下迭代求解过程:

$$\begin{aligned} \alpha^{t+1} &= \arg \min L_\lambda(\alpha, z^t) \\ &= \arg \min (\|\alpha\|_1 + \frac{\lambda}{2} \|y - X\alpha\|_2^2 + (z^t)^T X\alpha) \end{aligned} \quad (VI.34)$$

$$z^{t+1} = z^t + \lambda(y - X\alpha^{t+1}) \quad (VI.35)$$

特别地, 与 ADM 的  $\alpha^{t+1}$  求解方法不同, ALM 用 FISTA 算法求解  $\alpha^{t+1}$ 。

---DALM: 下面, 我们使用 ALM 方法来解决  $\hat{\alpha} = \arg \min \|\alpha\|_1 \quad s.t. \quad y = X\alpha$  (III.9) 的对偶优化问题。

1. 首先定义两个记号如下:

$$B_p^\lambda = \{x \in R^N \mid \|x\|_p \leq \lambda\} \quad \text{和} \quad I_\Omega(x) = \begin{cases} 0, & x \in \Omega \\ \infty, & x \notin \Omega \end{cases}$$

2. 建立 (III.9) 的对偶函数, 并引入上述记号:

$$g(\lambda) = \inf_{\alpha} \{\|\alpha\|_1 + \lambda^T (y - X\alpha)\} = \lambda^T y - \sup_{\alpha} \{\lambda^T X\alpha - \|\alpha\|_1\}$$

$$\rightarrow g(\lambda) = \lambda^T y - I_{B_\infty^1}(X^T \lambda), \quad \text{where the conjugate function of } I_{B_\infty^1}(\theta) \text{ is } \|\alpha\|_1$$

$$\rightarrow \min_{\lambda, z} -\lambda^T y \quad s.t. \quad z - X^T \lambda = 0, \quad z \in B_\infty^1$$

3. 构造相应的增广拉格朗日函数:

$$\begin{aligned} \arg \min_{\lambda, z, \mu} L(\lambda, z, \mu) &= -\lambda^T y - \mu^T (z - X^T \lambda) \\ &\quad + \frac{\tau}{2} \|z - X^T \lambda\|_2^2 \quad s.t. \quad z \in B_\infty^1 \end{aligned} \quad (VI.43)$$

4. 通过交替迭代求解增广拉格朗日的优化变量 (其中对于 z, 先因式分解, 再用局部算子投影)

$$\begin{aligned} z^{t+1} &= \arg \min_{z \in B_\infty^1} L_\tau(\lambda^t, z, \mu^t) \\ &= \arg \min_{z \in B_\infty^1} \{-\mu^T (z - X^T \lambda^t) + \frac{\tau}{2} \|z - X^T \lambda^t\|_2^2\} \\ &= \arg \min_{z \in B_\infty^1} \{\frac{\tau}{2} \|z - (X^T \lambda^t + \frac{1}{\tau} \mu^T)\|_2^2\} \\ &= P_{B_\infty^1}(X^T \lambda^t + \frac{1}{\tau} \mu^T) \end{aligned} \quad (VI.44)$$

$$\begin{aligned} \lambda^{t+1} &= \arg \min_{\lambda} L_\tau(\lambda, z^{t+1}, \mu^t) \\ &= \arg \min_{\lambda} \{-\lambda^T y + (\mu^t)^T X^T \lambda + \frac{\tau}{2} \|z^{t+1} - X^T \lambda\|_2^2\} \\ &= Q(\lambda) \end{aligned} \quad (VI.45)$$

where  $P_{B_\infty^1}(u)$  is a projection, or called a proximal operator, onto  $B_\infty^1$  and it is also called group-wise soft-thresholding. For example, let  $x = P_{B_\infty^1}(u)$ , then the  $i$ -th component of solution  $x$  satisfies  $x_i = \text{sign}(u_i) \min\{|u_i|, 1\}$

Take the derivative of  $Q(\lambda)$  with respect to  $\lambda$  and obtain

$$\lambda^{t+1} = (\tau X X^T)^{-1} (\tau X z^{t+1} + y - X \mu^t) \quad (VI.46)$$

$$\mu^{t+1} = \mu^t - \tau (z^{t+1} - X^T \lambda^{t+1}) \quad (VI.47)$$

<p><b>Algorithm 8</b> Dual Augmented Lagrangian Method for <math>l_1</math>-Norm Regularization</p> <p><b>Task:</b> To address the dual problem of <math>\hat{\alpha} = \arg \min_{\alpha} \ \alpha\ _1 \quad s.t. \quad y = X\alpha</math></p> <p><b>Input:</b> Probe sample <math>y</math>, the measurement matrix <math>X</math>, a small constant <math>\lambda^0</math>.</p> <p><b>Initialization:</b> <math>t = 0, \varepsilon = 0.01, \tau = \frac{1-\varepsilon}{\ X\ _2^2}, \mu^0 = 0</math>.</p> <p>While not converged do</p>	<p>Step 1: Apply the projection operator to compute <math>z^{t+1} = P_{B_\infty^1}(X^T \lambda^t + \frac{1}{\tau} \mu^T)</math> (投影 + 投影)</p> <p>Step 2: Update the value of <math>\lambda^{t+1} = (\tau X X^T)^{-1} (\tau X z^{t+1} + y - X \mu^t)</math>.</p> <p>Step 3: Update the value of <math>\mu^{t+1} = \mu^t - \tau (z^{t+1} - X^T \lambda^{t+1})</math>.</p> <p>Step 4: <math>t = t + 1</math>.</p> <p>End While</p> <p><b>Output:</b> <math>\alpha = \mu[1 : N]</math></p>
--	--

### VII. 基于同伦算法的稀疏表示

维持两个互补集合，依据相应准则不断调整同伦参数，并通过添加或者移除集合中元素，更新集合。

#### A. Lasso 同伦算法 (ps. 不断减小 $\lambda$ ，利用子梯度，求到临界点正负最小步长，更新同伦参数与支持集合)

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_1 \quad s.t. \quad y = X\alpha \quad (III.9) \quad \hat{\alpha} = L(\alpha, \lambda) = \arg \min_{\alpha} \frac{1}{2} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (III.12)$$

令  $\lambda$  从一个很大的值逐渐减小，变化到 0。III. 12 的解就会逐渐收敛到 III. 9 的解。同伦算法跟踪了在  $\lambda$  在这个变化过程中的同伦解的轨迹。

首先，对 III. 12 求导，我们有

$$\frac{\partial L}{\partial \alpha} = -X^T(y - X\alpha) + \lambda \partial \|\alpha\|_1, \quad \text{其中 } \partial \|\alpha\|_1 = \begin{cases} \theta_i = \text{sgn}(\alpha_i), & \alpha_i \neq 0 \\ \theta_i \in [-1, 1], & \alpha_i = 0 \end{cases}$$

用  $\Lambda$  和  $u$  分别代表稀疏表示  $\alpha$  的 support index 和 sign。  $\Lambda^c$  则是其互补集合。  $X_{\Lambda}$  代表 support atom。根据 III. 12 的 KKT 优化条件，我们有：

子梯度：取了字典  $X$  的其

中的 support 部分的导数=0  $\rightarrow X_{\Lambda}(y - X\alpha) = \lambda u; \quad \|X_{\Lambda^c}^T(y - X\alpha)\|_{\infty} \leq \lambda \quad (VII.2)$

如果我们 decrease the value of  $\lambda$  to  $\lambda - \tau$ ，那么下述约束必须满足（临界点判定条件）

$$\begin{aligned} X_{\Lambda}^T(y - X\alpha) + \tau X_{\Lambda}^T X \delta &= (\lambda - \tau)u \quad (a) \\ \|p + \tau q\|_{\infty} &\leq \lambda - \tau \quad (b) \end{aligned}$$

where  $p = X^T(y - X\alpha)$ ,  $q = X^T X \delta$  ← *the vector of residual correlations.*

下面，我们给出同伦算法的过程：

1. 计算更新方向  $\delta_i$ ：

$$X_{\Lambda}^T X_{\Lambda} \delta_i = u \rightarrow \delta_i = \begin{cases} (X_{\Lambda}^T X_{\Lambda})^{-1} u, & \text{on } \Lambda \\ 0, & \text{otherwise} \end{cases}$$

2. 通过跟踪同伦解路径，计算更新步长  $\tau$ （导致通向临界点的最小步长  $\tau_i^* = \min(\tau_i^+, \tau_i^-)$ ）：

$$\tau_i^+ = \min_{i \in \Lambda^c} \left( \frac{\lambda - p_i}{1 - x_i^T X_{\Lambda} \delta_i}, \frac{\lambda + p_i}{1 + x_i^T X_{\Lambda} \delta_i} \right)_+ \quad (VII.6) \quad \text{and} \quad \tau_i^- = \min_{i \in \Lambda} \left( \frac{-\alpha_i}{\delta_i} \right)_+ \quad (VII.7)$$

其中，  $p_i = x_i^T(y - x_i \alpha_i)$ ，  $\min(\cdot)_+$  表示只对元素中的正值取  $\min(\cdot)$ ，  $\tau_i^+$  表示添加元素  $i^+$  的最小步长，  $\tau_i^-$  表示删去元素  $i^-$  的最小步长。

3. 最终的当前同伦解更新为： $\alpha_{l+1} = \alpha_l + \tau_l^* \delta$ 。（方向\*步长）。

*更新  $\alpha$   $\Rightarrow$  解调  $\alpha$  的非零值  
 $\rightarrow$   $\lambda, u \Rightarrow$  解调  $\alpha$  的非零值*

#### Algorithm 9 Lasso Homotopy Algorithm

**Task:** To address the Lasso problem:

$$\hat{\alpha} = \arg \min_{\alpha} \|y - X\alpha\|_2^2 \quad s.t. \quad \|\alpha\|_1 \leq \varepsilon$$

**Input:** Probe sample  $y$ , measurement matrix  $X$ .

**Initialization:**  $l = 1$ , initial solution  $\alpha_l$  and its support set  $\Lambda_l$ .

Repeat:

Step 1: Compute update direction  $\delta_l$  by using Eq. (VII.5).

Step 2: Compute  $\tau_l^+$  and  $\tau_l^-$  by using Eq. (VII.6) and Eq. (VII.7).

Step 3: Compute the optimal minimum step size  $\tau_l^*$  by using  $\tau_l^* = \min\{\tau_l^+, \tau_l^-\}$ .

Step 4: Update the solution  $\alpha_{l+1}$  by using  $\alpha_{l+1} = \alpha_l + \tau_l^* \delta_l$ .

Step 5: Update the support set:

If  $\tau_l^* == \tau_l^-$  then

~~Remove the  $i^-$  from the support set, i.e.~~

$$\Lambda_{l+1} = \Lambda_l \setminus i^-.$$

else

(Add the  $i^+$  into the support set, i.e.

$$\Lambda_{l+1} = \Lambda_l \cup i^+$$

End if

Step 6:  $l = l + 1$ .

Until  $\|X^T(y - X\alpha)\|_{\infty} = 0$

**Output:**  $\alpha_{l+1}$

*必+/- 1 为 2 到 的 临界点*

*$\tau$  的解是动态 临界点步长*

### B. 基追踪去噪 (BPDN) 的同伦算法 (ps. 直接计算同伦解 $\alpha$ )

BPDN 同伦与 Lasso 同伦中的计算子梯度的方向大小, 最小正负步长, 支持集合更新方法均相同。两者的区别在于, BPDN 同伦每步用直接计算同伦解的解析解, 而不是每步去迭代逼近:

$$\alpha = \begin{cases} (X_{\Lambda}^T X_{\Lambda})^{-1} (X_{\Lambda}^T y - \lambda u), & \text{on } \Lambda \\ 0, & \text{otherwise} \end{cases} \quad (\text{VII.10})$$

---

#### Algorithm 10 BPDN Homotopy Algorithm

**Task:** To address the Lasso problem:

$$\hat{\alpha} = \arg \min_{\alpha} \|y - X\alpha\|_2^2 + \lambda \|\alpha\|_1$$

---

**Input:** Probe sample  $y$ , measurement matrix  $X$ .

**Initialization:**  $l = 0$ , initial solution  $\alpha_0$  and its support set  $\Lambda_0$ , a large value  $\lambda_0$ , step size  $\tau$ , tolerance  $\varepsilon$ .

Repeat:

Step 1: Compute update direction  $\delta_{l+1}$  by using

$$\delta_{l+1} = (X_{\Lambda}^T X_{\Lambda})^{-1} u_l.$$

Step 2: Update the solution  $\alpha_{l+1}$  by using Eq. (VII.10).

Step 3: Update the support set and the sign sequence set.

Step 6:  $\lambda_{l+1} = \lambda_l - \tau, l = l + 1$ . *Handwritten: 同伦参数更新*

Until  $\lambda \leq \varepsilon$

**Output:**  $\alpha_{l+1}$

---

### C. 基于同伦算法的交替再加权的 $l_1$ 范数最小化 (ps. 从权重 $w_i$ 过渡到权重 $\hat{w}_i$ )

普通的加权的  $l_1$  范数最小化问题目标函数为:

$$\operatorname{argmin} \frac{1}{2} \|X\alpha - y\|_2^2 + \sum_{i=1}^N w_i |\alpha_i|$$

其权重的更新方法为  $w_i = \frac{\lambda}{|\alpha_i| + \sigma}$ , where parameters  $\lambda$  and  $\sigma$  are both small constants.

新的基于同伦算法的再加权  $l_1$  范数最小化问题目标函数为:

$$\operatorname{argmin} \frac{1}{2} \|X\alpha - y\|_2^2 + \sum_{i=1}^N ((1 - \sigma)w_i + \sigma \hat{w}_i) |\alpha_i|$$

其中  $w_i$  and  $\hat{w}_i$  分别为给定的权重和新获得的权重。

同样, 由 KKT 条件, 有:

$$x_i^T (X\alpha - y) = -((1 - \sigma)w_i + \sigma \hat{w}_i) u_i \quad \text{for } i \in \Lambda \quad (a)$$

$$|x_i^T (X\alpha - y)| < (1 - \sigma)w_i + \sigma \hat{w}_i \quad \text{for } i \in \Lambda^c \quad (b)$$

如果我们 parameter changes from  $\sigma$  to  $\sigma + \tau$  in the update direction  $\delta$ , 则相应的临界点判定条件为:

$$\begin{aligned} X_{\Lambda}^T (X\alpha - y) + \tau X_{\Lambda}^T X \delta & \quad \text{where } u \text{ is the sign sequence of } \alpha \text{ on its support } \Lambda, \\ = -((1 - \sigma)W + \sigma \hat{W})u + \tau(W - \hat{W})u & \quad (a) \quad p_i = x_i^T (X\alpha - y), q_i = x_i^T X \delta, r_i = (1 - \sigma)w_i + \sigma \hat{w}_i \\ |p - \tau q| \leq r + \tau s & \quad (b), \text{ and } s_i = \hat{w}_i - w_i. \end{aligned}$$

对同伦参数  $\delta_l$  求导, 获得子梯度的方向  $\delta_l$  和大小  $\tau_l^* = \min(\tau_l^+, \tau_l^-)$ .

$$\tau_l^+ = \min_{i \in \Lambda^c} \left( \frac{r_i - p_i}{q_i - s_i}, \frac{-r_i - p_i}{q_i + s_i} \right)_+$$

$$\text{梯度方向: } \delta_l = \begin{cases} (X_{\Lambda}^T X_{\Lambda})^{-1} (W - \hat{W})u, & \text{on } \Lambda \\ 0, & \text{otherwise} \end{cases}; \text{ 梯度步长: } \tau_l^- = \min_{i \in \Lambda} \left( \frac{-\alpha_l^i}{\delta_l^i} \right)_+$$

**Algorithm 11** Iterative Reweighting Homotopy Algorithm for Weighted  $l_1$ -Norm Minimization

**Task:** To address the weighted  $l_1$ -norm minimization:

$$\hat{\alpha} = \arg \min \frac{1}{2} \|X\alpha - y\|_2^2 + W\|\alpha\|_1$$

**Input:** Probe sample  $y$ , measurement matrix  $X$ .

**Initialization:**  $l = 1$ , initial solution  $\alpha_l$  and its support set  $\Lambda_l, \sigma_l = 0$ , *开始*

Repeat:

Step 1: Compute update direction  $\delta_l$  by using Eq. (VII.17).

Step 2: Compute  $p, q, r$  and  $s$  by using Eq. (VII.16).

Step 2: Compute  $\tau_l^+$  and  $\tau_l^-$  by using Eq. (VII.18) and Eq. (VII.19).

Step 3: Compute the step size  $\tau_l^*$  by using

$$\tau_l^* = \min\{\tau_l^+, \tau_l^-\}.$$

Step 4: Update the solution  $\alpha_{l+1}$  by using  $\alpha_{l+1} = \alpha_l + \tau_l^* \delta_l$ .

Step 5: Update the support set:

If  $\tau_l^+ == \tau_l^-$  then

Shrink the value to zero at the index  $i^-$  and remove  $i^-$ , i.e.  $\Lambda_{l+1} = \Lambda_l \setminus i^-$ .

else

Add the  $i^+$  into the support set, i.e.  $\Lambda_{l+1} = \Lambda_l \cup i^+$

End if

Step 6:  $\sigma_{l+1} = \sigma_l + \tau_l$  and  $l = l + 1$ .

Until  $\sigma_{l+1} = 1$

**Output:**  $\alpha_{l+1}$  *开始 = 有 N*

**VIII. 稀疏表示方法的应用 (字典学习, 图像处理, 图像分类, 目标跟踪)**

**A. 字典学习中的稀疏表示**

稀疏表示的 atoms 来源有两种, 一是基于事先指定的集合, 另一种是基于学习的集合。但由于后者效果比较好, 这儿主要关注后者。通常, 过程包括两个交替更新环节: 1.稀疏近似; 2.字典学习。

字典学习的目标函数如下:

$$\arg \min_{D \in \Omega, x_i} \left\{ \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{2} \|y_i - Dx_i\|_2^2 + \lambda P(x_i) \right) \right\} \quad \text{where } \Omega = \{D = [d_1, d_2, \dots, d_M] : d_i^T d_i = 1, i = 1, 2, \dots, M\} \text{ (M here may not be equal to N)}$$

**1. 无监督字典学习**

**a. 基于 KSVD 的无监督字典学习 (IO 范数, 大环: MP 求系数位置与值; 小环: KSVD 求字典和仅更新系数值)**

通常对 IO 范数处理, 其优化目标函数为:

$$\arg \min_{D, X} \{\|Y - DX\|_F^2\} \quad \text{s.t. } \|x_i\|_0 \leq k, i = 1, 2, \dots, N$$

下面, 迭代优化 D 和 X。首先, 固定字典 D, 那么目标函数就变为:

$$\arg \min_{x_i} \|y_i - Dx_i\|_2^2 \quad \text{s.t. } \|x_i\|_0 \leq k, i = 1, 2, \dots, N$$

显然这个 IO 范数约束的稀疏表示的求解可用贪婪策略中的(正交)匹配追踪算法 **OMP** 实现。然后固定稀疏矩阵 X, 优化字典 D。这个过程的目标函数就简化为:

$$\hat{D} = \arg \min_D \|Y - DX\|_F^2$$

那么最终求解的字典为:  $\hat{D} = YX^\dagger = YX^T(XX^T)^{-1}$ 。(ps.复杂度大, 为  $O(n^3)$ )。

为此，用 KSVD 求解字典，并更新相应 atoms 的对应的系数值（过程不增加非零系数数量）。目标函数为：

$$\begin{aligned} \hat{D} &= \arg \min_D \|Y - DX\|_F^2 = \arg \min_D \|Y - \sum_{j=1}^M d_j x_j^T\|_F^2 \\ &= \arg \min_D \|(Y - \sum_{j \neq l} d_j x_j^T) - d_l x_l^T\|_F^2 \end{aligned} \quad (\text{VIII.5})$$

$x_j \in \{V, X\}$   
 $\leftarrow (1, 2, \dots, M)$

KSVD 首先计算残差  $E_l = Y - \sum_{j \neq l} d_j x_j^T$ ，然后更新字典的 atom  $d_l$  和该 atom 对应的系数  $x_l$ 。特别的，为了保证  $x_l$  的稀疏性，KSVD 在 SVD 操作中，仅提取原本  $x_l$  中非零位置的残差  $E_l^p$  和系数（即  $E_l^p$  属于  $R(dx|s|)$ ， $x_l$  属于  $R(1x|s|)$ ）。s 为当前系数  $x_l$  中，包含第 l 个 atom 的样本的索引。然后对相关残差矩阵  $E_l^p$  进行 SVD 分解， $E_l^p = U \Lambda V^T$ 。最终更新  $d_l$  为 U 的第一列，而  $x_l$  相应位置（s 处）的值更新为 V 的第一列  $x \wedge(1, 1)$ 。

**Algorithm 12** The K-SVD Algorithm for Dictionary Learning

**Task:** Learning a dictionary  $D$ :  $\arg \min_{D, X} \|Y - DX\|_F^2$   
 s.t.  $\|x_i\|_0 \leq k, i = 1, 2, \dots, N$

**Input:** The matrix composed of given samples  $Y = [y_1, y_2, \dots, y_m]$ .

**Initialization:** Set the initial dictionary  $D$  to the  $l_2$ -norm unit matrix  $i = 1$ .

**While not converged do**

Step 1: For each given example  $y_i$ , employing the classical sparse representation with  $l_0$ -norm regularization solve problem VIII.3 for further estimating  $X^i$ , set  $l = 1$ .

While  $l$  is not equal to  $k$  do

Step 2: Compute the overall representation residual  $E_l = Y - \sum_{j \neq l} d_j x_j^T$ .

Step 3: Extract the column items of  $E_l$  which corresponds to the nonzero elements of  $x_l^T$  and obtain  $E_l^p$ .

Step 4: SVD decomposes  $E_l^p$  into  $E_l^p = U \Lambda V^T$ .

Step 5: Update  $d_l$  to the first column of  $U$  and update corresponding coefficients in  $x_l^T$  by  $\wedge(1, 1)$  times the first column of  $V$ .

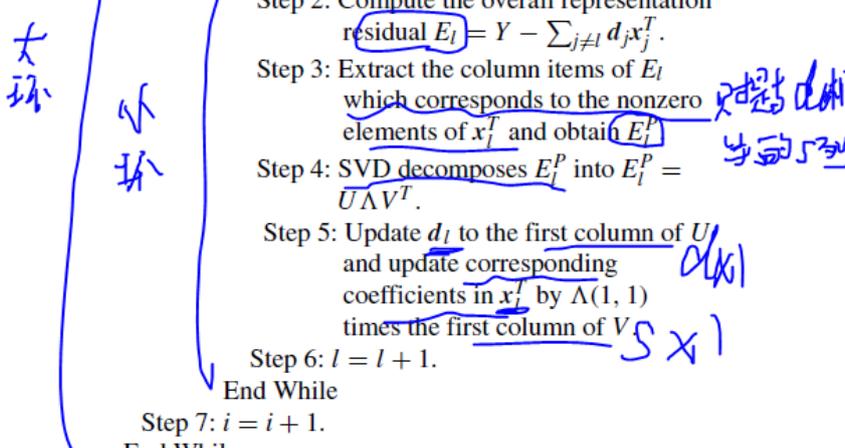
Step 6:  $l = l + 1$ .

End While

Step 7:  $i = i + 1$ .

End While

**Output:** dictionary  $D$



**b. 基于局部约束线性编码的无监督字典学习 (ps. 局部必稀疏，稀疏未必局部，以此建立 l2 约束)**

LLC 引入了局部导致稀疏的思想，从而建立基于 l2 范数的有效约束，其最终目标函数为

$$\begin{aligned} \arg \min_{x_i, D} \sum_{i=1}^N \|y_i - Dx_i\|_2^2 + \mu \|b_i \odot x_i\|_2^2 \\ \text{s.t. } 1^T x_i = 1, i = 1, 2, \dots, N \end{aligned} \quad (\text{VIII.6}), \text{ 其中 } b_i = \exp\left(\frac{\text{dist}(y_i, D)}{\sigma}\right), \text{ 属于 } R(M \times 1).$$

首先，如果直接使用 K-Means 聚类获得的聚类中心作为字典 D，那么我们可以获得  $x_i$  的解析解：

$$\hat{x}_i = (C_i + \mu \text{diag}^2(b)) \setminus 1 \quad \text{where the operator } a \setminus b \text{ denotes } a^{-1}b, \text{ and } C_i = (D^T - 1y_i^T)$$

$$x_i = \hat{x}_i / 1^T \hat{x}_i, \quad (D^T - 1y_i^T)^T \text{ is the covariance matrix with respect to } y_i.$$

进一步，如果进行字典的优化学习，我们有如下目标函数：

$$\arg \min_{x_i, D} \sum_{i=1}^N \|y_i - Dx_i\|_2^2 + \mu \|b_i \odot x_i\|_2^2$$

$$s.t. \quad 1^T x_i = 1, \quad \forall i; \quad \|d_j\|_2^2 \leq 1, \quad \forall j$$

**Algorithm 13** The Incremental Codebook Optimization Algorithm

**Task:** Learning a dictionary  $D$ :  $\arg \min_{x_i, D} \sum_{i=1}^N \|y_i - Dx_i\|_2^2 + \mu \|b_i \odot x_i\|_2^2$  s.t.  $1^T x_i = 1, \forall i; \|d_j\|_2^2 \leq 1, \forall j$

**Input:** The matrix composed of given samples  $Y = [y_1, y_2, \dots, y_N]$ .

**Initialization:**  $i = 1, \varepsilon = 0.01, D$  initialized by  $K$ -Means clustering algorithm.

**While**  $i$  is not equal to  $N$  **do**

Step 1: Initialize  $b$  with  $1 \times N$  zero vector.

Step 2: Update locality constraint parameter  $b_j$  with

$$b_j = \exp\left(-\frac{\text{dist}(y_i, d_j)}{\sigma}\right) \text{ for } \forall j.$$

Step 3: Normalize  $b$  using the equation  $b = \frac{b - b_{\min}}{b_{\max} - b_{\min}}$ .

Step 4: Exploit the LLC coding algorithm to obtain  $x_i$ .

Step 5: Keep the set of  $D^i$ , whose corresponding entries of the code  $x_i$  are greater than  $\varepsilon$ , and drop out other elements, i.e.  $\text{index} \leftarrow \{j \mid \text{abs}\{x_i(j)\} > \varepsilon\} \forall j$  and  $D^i \leftarrow D(:, \text{index})$ .

Step 6: Update  $x_i$  exploiting  $\arg \max \|y_i - D^i x_i\|_2^2$  s.t.  $1^T x_i = 1$ .

Step 7: Update dictionary  $D$  using a classical gradient descent method with respect to problem VIII.6.

Step 8:  $i = i + 1$ .

**End While**

**Output:** dictionary  $D$

Handwritten notes and diagrams for Algorithm 13:

- Step 2:  $b_j = \exp\left(-\frac{\text{dist}(y_i, d_j)}{\sigma}\right)$  for  $\forall j$ . (Note:  $\text{dist}$  is circled in blue)
- Step 3: Normalize  $b$  using the equation  $b = \frac{b - b_{\min}}{b_{\max} - b_{\min}}$ . (Note:  $b$  is circled in blue)
- Step 5: Keep the set of  $D^i$ , whose corresponding entries of the code  $x_i$  are greater than  $\varepsilon$ , and drop out other elements, i.e.  $\text{index} \leftarrow \{j \mid \text{abs}\{x_i(j)\} > \varepsilon\} \forall j$  and  $D^i \leftarrow D(:, \text{index})$ . (Note: "drop out" is written in blue)
- Step 6: Update  $x_i$  exploiting  $\arg \max \|y_i - D^i x_i\|_2^2$  s.t.  $1^T x_i = 1$ . (Note: "更新字典" is written in blue)
- Step 7: Update dictionary  $D$  using a classical gradient descent method with respect to problem VIII.6. (Note: "更新字典" is written in blue)
- Step 7:  $\Delta B_i \leftarrow -2\tilde{c}_i(x_i - B_i \tilde{c}_i), \mu \leftarrow \sqrt{1/i}$   
 $B_i \leftarrow B_i - \mu \Delta B_i / \|\tilde{c}_i\|_2$   
 $B(:, id) \leftarrow \text{proj}(B_i)$  (Note:  $\Delta B_i$  is circled in blue)

## 2. 监督的字典学习 (ps. 训练中字典更新 with KSVD + 分类中稀疏表示求解 with OMP)

无监督的字典学习能够实现很好的数据重构，但是不一定有助于改进分类。而监督学习能够将数据的标签集成到稀疏表示和字典学习的过程中去，这就使得学到的字典有助于后续分类。

### a. 基于判别性 KSVD (DKSVD) 的字典学习 (ps. 字典学习+稀疏表示+分类模型 $\rightarrow$ 整合到 KSVD 模型)

DKSVD 与 KSVD 的区别在于，它额外引入了标签信息到目标函数中。并以此在字典学习，稀疏表示的同时，建立了一个分类模型。另  $H$  为标签矩阵，DKSVD 的目标函数如下：

$$(D, C, X) = \arg \min_{D, C, X} \|Y - DX\|_F^2 + \mu \|H - CX\|_F^2 + \eta \|C\|_F^2 \quad s.t. \quad \|x_i\|_0 \leq k \quad \text{(VIII.11)}$$

Handwritten notes for (VIII.11): "判别性" (discriminative), "字典学习" (dictionary learning), "稀疏表示" (sparse representation), "分类模型" (classification model).

$$(D, C, X) = \arg \min_{D, C, X} \left\| \begin{pmatrix} Y \\ \sqrt{\mu} H \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\mu} C \end{pmatrix} X \right\|_F^2 + \eta \|C\|_F^2 \quad s.t. \quad \|x_i\|_0 \leq k \quad \text{(VIII.12)}$$

Handwritten notes for (VIII.12):  $(D, C) \times N$ ,  $(D, C) \times M$ ,  $M \times N$ .

In consideration of the KSVD algorithm, each column of the dictionary will be normalized to  $l_2$ -norm unit vector and  $\begin{pmatrix} D \\ \sqrt{\mu} C \end{pmatrix}$  will also be normalized, and then the penalty term  $\|C\|_F^2$  will be dropped out and problem VIII.12 will be

$$(Z, X) = \arg \min_{Z, X} \|W - ZX\|_F^2 \quad s.t. \quad \|x_i\|_0 \leq k \quad \text{(VIII.13)}$$

where  $W = \begin{pmatrix} Y \\ \sqrt{\mu} H \end{pmatrix}$ ,  $Z = \begin{pmatrix} D \\ \sqrt{\mu} C \end{pmatrix}$  normalized in KSVD.

Handwritten note: "这是KSVD的求解格式" (This is the solving format of KSVD).

相比于 KSVD, DKSVD 包含了训练和分类两个环节。在训练环节, DKSVD 针对 VIII.13 应用 KSVD 求解后, 能得到  $W$ , 归一化的  $Z$ , 以及  $X$ 。从  $Z$  中提取到  $D$ ,  $C$  后, 我们对  $D$  和  $C$  重新归一化。并计算相应新的  $x_i$ :

$$\begin{cases} D' = [d'_1, d'_2, \dots, d'_M] = \left[ \frac{d_1}{\|d_1\|}, \frac{d_2}{\|d_2\|}, \dots, \frac{d_M}{\|d_M\|} \right] \\ C' = [c'_1, c'_2, \dots, c'_M] = \left[ \frac{c_1}{\|d_1\|}, \frac{c_2}{\|d_2\|}, \dots, \frac{c_M}{\|d_M\|} \right] \\ x'_i = x_i \times \|d_i\| \end{cases} \quad (\text{VIII.14})$$

在分类环节, 对于测试样本, 我们先获得其稀疏表示 (用正交匹配追踪算法 OMP):

$$\hat{x}_i = \arg \min \|y_i - D'x'_i\|_2^2 \quad s.t. \|x'_i\|_0 \leq k$$

然后, 利用分类参数矩阵  $C'$  进行分类投影:

$$\text{label} = C' \times \hat{x}_i$$

### b. 基于标签一致性的 KSVD (LC-KSVD) 的判别性字典学习 (ps. $D$ 和 $Y$ 的标签一致性矩阵 $L$ ; KSVD 模型)

利用联合标签矩阵  $L$ , 建立集成了字典构建和线性分类的一个混合重构判别性模型 (判别性的稀疏编码), 进行字典和分类器的整体的关联学习。LC-KSVD 的目标函数如下:

$$(D, A, C, X) = \arg \min_{D, A, C, X} \|Y - DX\|_F^2 + \mu \|L - AX\|_F^2 + \eta \|H - CX\|_F^2 \quad s.t. \|x_i\|_0 \leq k \quad (\text{VIII.17})$$

and  $L$  is a joint label matrix for labels of  $Y$  and  $D$ . For example, providing that  $Y = [y_1 \dots y_4]$  and  $D = [d_1 \dots d_4]$  where  $y_1, y_2, d_1$  and  $d_2$  are from the first class, and  $y_3, y_4, d_3$  and  $d_4$  are from the second class, and then the joint label

matrix  $L$  can be defined as  $L = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

同样的, 类似 DKSVD 到 KSVD 的转化, 我们也可以将 LC-KSVD 转化到 KSVD 的求解形式:

$$(Z, X) = \arg \min_{Z, X} \|T - ZX\|_2^2 \quad s.t. \|x_i\|_0 \leq k \quad (\text{VIII.18}) \quad \text{where } T = \begin{pmatrix} Y \\ \sqrt{\mu}L \\ \sqrt{\eta}H \end{pmatrix}, Z = \begin{pmatrix} D \\ \sqrt{\mu}A \\ \sqrt{\eta}C \end{pmatrix}$$

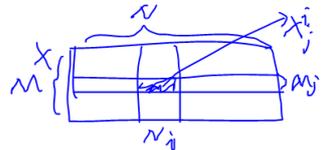
用 KSVD 求解后, 可以分别对  $D$ ,  $A$ ,  $C$  进行归一化。并利用到之后的分类环节中 (OMP)。具体如下:

$$\begin{aligned} D' &= [d'_1, d'_2, \dots, d'_M] = \left[ \frac{d_1}{\|d_1\|}, \frac{d_2}{\|d_2\|}, \dots, \frac{d_M}{\|d_M\|} \right] & \hat{x}_i &= \arg \min \|y_i - D'x'_i\|_2^2 \quad s.t. \|x'_i\|_0 \leq k \quad (\text{VIII.20}) \\ A' &= [a'_1, a'_2, \dots, a'_M] = \left[ \frac{a_1}{\|d_1\|}, \frac{a_2}{\|d_2\|}, \dots, \frac{a_M}{\|d_M\|} \right] & \text{The final classification is based on a simple linear} & \\ C' &= [c'_1, c'_2, \dots, c'_M] = \left[ \frac{c_1}{\|d_1\|}, \frac{c_2}{\|d_2\|}, \dots, \frac{c_M}{\|d_M\|} \right] & \text{predictive function} & \\ & & l &= \arg \max_f \{f = C' \times \hat{x}_i\} \quad (\text{VIII.21}) \end{aligned}$$

### c. Fisher 判别性字典学习 (FDDL) (ps. 引入标签关联信息到 $f$ , 引入 Fisher 信息到 $g$ , 按类更新系数和字典)

集成了类别的标签信息和 Fisher 判别性信息到字典学习框架中。其广义目标函数为:

$$J(D, X) = \arg \min_{D, X} \{f(Y, D, X) + \mu \|X\|_1 + \eta g(X)\}$$



特别的, FDDL 按样本的类, 进行 class by class 的字典更新和稀疏表示求解。如果用  $Y^i R(dxNi)$  代表样本中

属于第  $i$  类的的数据,  $X^i R(MxNi)$  代表第  $i$  类数据的稀疏表示,  $X^j R(MjxNi)$  代表  $X^i$  系数中--基于属于第  $j$  类的 atoms

--重构的部分。  $D^i R(dxMi)$  代表字典中属于第  $i$  类的 atoms。那么可以建立目标函数如下:

$$J(D, X) = \arg \min_{D, X} \left( \sum_{i=1}^c f(Y^i, D, X^i) + \mu \|X\|_1 + \eta (\text{tr}(S_W(X)) - S_B(X)) + \lambda \|X\|_F^2 \right)$$

where  $f(Y^i, D, X^i) = \|Y^i - DX^i\|_F^2 + \|Y^i - D^i X^i\|_F^2 + \sum_{j \neq i} \|D^j X^i\|_F^2$  and  $S_W(X)$  and  $S_B(X)$  are the within-class scatter of  $X$  and between-class scatter of  $X$ , respectively.

要求不同类的样本和字典 atoms 的关联系数尽可能小, 特别是到  $D_i$  的值大时

然后，FDDL 在字典和稀疏表示更新的过程中，按类进行一类接一类的优化。

首先，固定字典  $D$ ，通过相应的子问题，计算第  $i$  (1:  $c$ ) 类的稀疏表示  $X^i$ ：

$$J(X^i) = \arg \min_{X^i} (f(Y^i, D, X^i) + \mu \|X^i\|_1 + \eta g(X^i))$$

其中  $g(X^i) = \|X^i - M_i\|_F^2 - \sum_{l=1}^c \|M_l - M\|_F^2 + \lambda \|X^i\|_F^2$ ， $M_j$  和  $M$  分别是第  $j$  类和所有样本的稀疏表示系数在当前的均值（在优化之前先计算）。上面这个  $l_1$  范数问题可以通过迭代投影算法求解。

然后固定  $X$ ，原优化问题只剩  $f$  项。通过相应的  $l_2$  范数的子问题，计算第  $i$  类 (1:  $c$ ) 类的 atoms  $D_i$ ：

$$J(D) = \arg \min_D (\|Y^i - D^i X^i - \sum_{j \neq i} D^j X^j\|_F^2 + \lambda \|D^i X^i\|_F^2 + \sum_{j \neq i} \|D^j X^j\|_F^2)$$

## B. 基于稀疏表示的图像处理

基于稀疏表示的图像处理方法有如下的通用框架：

S1. 将退化的图像  $X$  分成有重叠的 patches;

S2. 构建字典  $D$ ，然后对每一个 patches 求解如下优化问题，以获得当前 patch 的稀疏表示：

$$\hat{\alpha} = \arg \min \|\alpha\|_p \quad s.t. \|x - HD\alpha\|_2^2 \leq \epsilon, \quad \text{其中 } H \text{ 是退化矩阵, } 0 < p <= 1$$

S3. 利用稀疏表示  $\hat{\alpha}$  和字典  $D$ ，重构相应的 patch:  $\hat{x} = D\hat{\alpha}$ 。

S4. 将重构的 patch 放到原先的位置，然后平均重叠区域，从而获得整个重构的图像  $X'$ 。

S5. 令  $X = X'$ ，以新图像为重构目标，重复 1-4 步

### 1. 基于稀疏表示的图像超分辨率重构 (SRSR) (ps. 块; 低分辨率重构+高分率邻近一致; 全局一致性增强)

假设我们有两个紧致的字典  $D_l$  和  $D_h$ ，分别代表图像 patch 在低分辨率和高分率的字典。 $X$  和  $Y$  分别代表高分率和低分辨率的图像。 $x$  和  $y$  分别代表相应的图像 patch。因此，存在一种映射关系，即  $x = Py$ 。进一步，如果  $Y$  是由  $X$  经下采样和模糊处理产生，那么就有： $Y = SBX$ 。其中， $S$  和  $B$  代表了下采样算子和模糊滤波器（不是直接的矩阵操作）。

SRSR 假设任意的高分辨率图像块可以由高分率字典中的 atom 线性表示，并给出如下的目标函数：

$$x = D_h \alpha \quad s.t. \|\alpha\|_0 \leq k$$

由于上式 NP-Hard，SRSR 将其转化为  $l_1$  范数问题：

$$\arg \min \|\alpha\|_1 \quad s.t. \|x - D_h \alpha\|_2^2 \leq \epsilon$$

SRSR 通过 2 步来实现上式的求解：基于局部模型的稀疏表示 (LMBSR) + 全局重构的一致性增强。

S1. LMBSR 在低分辨率的图像上对每一个图像 patch 进行基于局部模型的稀疏表示：

$$\arg \min \|Fy - FD_l \alpha\|_2^2 + \lambda \|\alpha\|_1, \quad \text{其中 } F \text{ 是特征提取算子。}$$

同时，为了保证当前获得的超分辨率重构 patch 和先前已获得的邻近的超分辨率重构 patches 相一致，SRSR 进一步改进了目标函数：

$$\arg \min \|\alpha\|_1 \quad s.t. \|Fy - FD_l \alpha\|_2^2 \leq \epsilon_1; \|v - LD_h \alpha\|_2^2 \leq \epsilon_2$$

其中  $v$  表示先前已获得的超分辨率图像 patches 与当前要重构的 patch 的重叠部分。 $L$  表示提取当前 patch 的该重叠部分。进一步，上式可以化简为 (ps. 下式  $\hat{y}$  和  $D$  都是优化前可以计算确定的，即仅  $\alpha$  为变量)：

$$\arg \min \|\hat{y} - D\alpha\|_2^2 + \lambda \|\alpha\|_1, \quad \text{其中 } \hat{y} = \begin{bmatrix} Fy \\ v \end{bmatrix} \text{ and } D = \begin{bmatrix} FD_l \\ LD_h \end{bmatrix}$$

$$\begin{matrix} f_1 = [-1, 0, 1], & f_2 = f_1^T \\ f_3 = [1, 0, -2, 0, 1], & f_4 = f_3^T \end{matrix}$$

(优化过程中怎么对  $D_l$  进行  $F$  操作?)

实际 3 倍放大实验中，对低分辨率的提取  $3 \times 3$  的块，然后进行双倍插值，得到  $6 \times 6$  的实际特征块。对高分率图像提取  $9 \times 9$  的特征块。进行  $F$  时，先对  $D_l$  resize，滤波后级联，在当做  $D_l$ 。 ???

由于上式本质是  $l_1$  范数的稀疏表示问题，所以可以用之前的算法求解。之后，超分辨率的图像 patch 就可以获得： $x = D_h \alpha^*$ 。

S2. SRSR 进行全局重构的一致性增强以消除 S1 可能产生的噪声。假设第一步获得的整体图像为  $X_0$ ，有：

$$X^* = \arg \min \|X - X_0\|_2^2 \quad s.t. Y = SBX$$

$X^*$  为最终的重构的高分辨率图像。

**Algorithm 14** Super-Resolution via Sparse Representation

**Input** Training image patches dictionaries  $D_l$  and  $D_h$ , a low-resolution image  $Y$ .

For each overlapped  $3 \times 3$  patches  $y$  of  $Y$  using one-pass algorithm, from left to right and top to bottom

Step 1: Compute optimal sparse representation coefficients  $\alpha^*$  in problem (VIII.32)

Step 2: Compute the high-resolution patch by  $x = D_h \alpha^*$ .

Step 3: Put the patch  $x$  into a high-resolution image  $X_0$  in corresponding location.

End

Step 4: Compute the final super-resolution image  $X^*$  in problem (VIII.33).

**Output:**  $X^*$

迭代求解，低分辨率图像

迭代求解

利用之前的重建的系数

计算高分部分

**2. 基于字典学习的稀疏表示模型 (SRMLD) 的图像去噪 (ps. 全局极大似然约束+局部重构的贝叶斯先验约束)**

主要思路是分离实际信号和噪声信号，从而实现噪声的移除和原始图像的重构。

用  $z$  表示目标图像  $x$  中的每一个 patch，则其局部(patch)重构向量  $\alpha$  应满足如下贝叶斯先验约束条件：

$$\arg \min_{\alpha} \|\alpha\|_0 \quad s.t. D\alpha = z \quad \text{or} \quad \arg \min_{\alpha} \|D\alpha - z\|_2^2 + \lambda \|\alpha\|_0$$

为此，SRMLD 建立了如下目标函数：

$$\arg \min_{D, \alpha_i, x} \delta \|x - y\|_2^2 + \sum_{i=1}^M \|D\alpha_i - P_i x\|_2^2 + \sum_{i=1}^M \lambda_i \|\alpha_i\|_0$$

其中， $P$  代表了提取  $x$  的相应的 patch。上式中第一项是全局的极大似然约束，要求重构的图像  $x$  在整体上与原始有噪声的图像  $y$  相接近，但同时重构图像的每一个 patch 应满足局部重构的贝叶斯先验约束条件。

上式有 3 个优化变量，可以相继进行优化。

首先固定字典  $D$  和系数  $\alpha$ ，简化上式为：

$$\arg \min_x \delta \|x - y\|_2^2 + \sum_{i=1}^M \|D\alpha_i - z\|_2^2, \quad \text{其中 } z = P_i x$$

由于上式是简单的凸优化问题，故有封闭解：

$$x = \left( \sum_{i=1}^M P_i^T P_i + \delta I \right)^{-1} \left( \sum_{i=1}^M P_i^T D\alpha_i + \delta y \right)^{-1} \quad (\text{VIII.38})$$

然后利用求得的初步重构图像  $x$ ，我们有

$$\arg \min_{D, \alpha_i} \sum_{i=1}^M \|D\alpha_i - P_i x\|_2^2 + \sum_{i=1}^M \lambda_i \|\alpha_i\|_0$$

由于各个  $\alpha$  独立，上式可以分成  $M$  个子问题独立求解。并且由于是  $l_0$  范数，我们可以通过 K-SVD 实现  $\alpha$  和  $D$  的一起更新。

**Algorithm 15** Image Denoising via Sparse and Redundant Representation Over Learned Dictionary

**Task:** To denoise a measured image  $y$  from white and additional Gaussian white noise:

$$\arg \min_{D, \alpha_i, x} \delta \|x - y\|_2^2 + \sum_{i=1}^M \|D\alpha_i - P_i x\|_2^2 + \sum_{i=1}^M \lambda_i \|\alpha_i\|_0$$

**Input:** Measured image sample  $y$ , the number of training iteration  $T$ .

**Initialization:**  $t = 1$ , set  $x = y$ ,  $D$  initialized by an overcomplete DCT dictionary.

While  $t \leq T$  do

Step 1: For each image patch  $P_i x$ , employ the KSVD algorithm to update the values of sparse representation solution  $\alpha_i$  and corresponding dictionary  $D$ .   
*KSVD* → For each atom in  $D$ , find the patches that use this atom, ...

Step 2:  $t = t + 1$

End While

Step 3: Compute the value of  $x$  by using Eq. (VIII.38).

**Output:** denoised image  $x$

**3. 基于集中稀疏表示 (CSR) 模型的单幅退化图像恢复 (ps. 块; 退化图像重构+原始图像稀疏系数期望近似)**

对于一副退化图像  $y$ , 我们可以表示为:  $y = Hx + v$ . 其中  $H$  为退化算子,  $x$  为原始图像,  $v$  为高斯声. 假设有如下两个稀疏优化问题:

$$\alpha_x = \arg \min \|\alpha\|_1 \quad s.t. \|x - D\alpha\|_2^2 \leq \varepsilon \quad (\text{VIII.42})$$

$$\alpha_y = \arg \min \|\alpha\|_1 \quad s.t. \|x - HD\alpha\|_2^2 \leq \varepsilon \quad (\text{VIII.43})$$

$\alpha_x$  表示不受退化影响后的原始图像的稀疏表示系数,  $\alpha_y$  表示受退化影响后的原始图像的稀疏表示系数. 自然地, 我们希望两者的差距尽可能小:

$$\alpha_y = \arg \min_{\alpha} \|y - HD\alpha\|_2^2 + \lambda \|\alpha\|_1 + \mu \|\alpha - \alpha_x\|_1$$

但是, 由于  $x$  未知, 我们很难直接获得  $\alpha_x$  的值. 但是, 我们假设一副图像之中, 对于 patch  $i$ , 会有很多与其相似 patches. 我们可以通过块匹配来获得这些 similar patches, 获得 patch  $i$  相似的 patch 聚类 (包括  $i$  本身), 记为  $\Omega_i$ . 其中第  $l$  块已带有的稀疏表示向量为  $\alpha_{il}$ . 然后, 我们利用这些相似 patches 的稀疏表示进行加权, 来近似当前 patch  $i$  的稀疏表示的期望, 即  $E[\alpha_x]$ :

$$\theta_i = \sum_{l \in \Omega_i} w_{il} \alpha_{il}, \quad \text{where } w_{il} = \exp(-\|x_i - x_{il}\|_2^2/h) / N, \quad x_i = D\alpha_i, \quad x_{il} = D\alpha_{il}$$

然后, 相应的最优化目标函数就可以写为:

$$\alpha_y = \arg \min_{\alpha} \|y - HD\alpha\|_2^2 + \lambda \|\alpha\|_1 + \mu \sum_{i=1}^M \|\alpha_i - \theta_i\|_1$$

特别地, 也可以使用局部 PCA 字典 (对每个 similar patches 获得其局部 PCA 字典) 替代上面的全局字典.

**Algorithm 16** Centralized Sparse Representation for Image Restoration

**Initialization:** Set  $x = y$ , initialize regularization parameter  $\lambda$  and  $\mu$ , the number of training iteration  $T$ ,  $t = 0$ ,  $\theta^0 = 0$ .

Step 1: Partition the degraded image into  $M$  overlapped patches.

While  $t \leq T$  do

Step 2: For each image patch, update the corresponding dictionary for each cluster via k-means and PCA.

Step 3: Update the regularization parameters  $\lambda$  and  $\mu$  by using

$$\lambda = \frac{2\sqrt{2}\rho^2}{\sigma_t} \quad \text{and} \quad \mu = \frac{2\sqrt{2}\rho^2}{\eta_t}$$

Step 4: Compute the nonlocal means estimation of the unbiased estimation of  $\alpha_x$ , i.e.  $\theta_i^{t+1}$ , by using Eq. (VIII.46) for each image patch.

Step 5: For a given  $\theta^{t+1}$ , compute the sparse representation solution, i.e.  $\alpha_y^{t+1}$ , in problem (VIII.48) by using the extended iterative shrinkage algorithm in literature [184].

Step 6:  $t = t + 1$

End While

**Output:** Restored image  $x = D\alpha_y^{t+1}$

### C. 基于稀疏表示的图像分类 (ps.归类到稀疏表示的残差最小的一类)

首先用全局样本进行稀疏表示。然后分别提取稀疏表示向量中不同类别的部分，计算测试样本在不同类中的重构残差，最后将测试样本分类到残差最小的一类。

---

#### Algorithm 17 The Scheme of Sparse Representation Based Classification Method

---

Step 1: Normalize all the samples to have unit  $l_2$ -norm.

Step 2: Exploit the linear combination of all the training samples to represent the test sample and the following  $l_1$ -norm minimization problem is satisfied

$$\alpha^* = \arg \min \|\alpha\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{X}\alpha\|_2^2 \leq \varepsilon.$$

Step 3: Compute the representation residual for each class

$$r_i = \|\mathbf{y} - \mathbf{X}_i \alpha_i^*\|_2^2$$

where  $\alpha_i^*$  here denotes the representation coefficients vector associated with the  $i$ -th class.

Step 4: Output the identity of the test sample  $\mathbf{y}$  by judging

$$\text{label}(\mathbf{y}) = \arg \min_i (r_i).$$

---

## IX. 讨论与结论

1. 正则化参数的选择，鲁棒的稀疏表示， $l_1$  范数的高效性和有效性算法等需要进一步的研究。
2. 利用低秩表示和稀疏表示到转换学习的框架中；研究子空间学习和稀疏表示之间的关联性；将当前的稀疏表示和深度学习结合起来；增强稀疏表示的鲁棒性；对  $l_1$ -范数的研究；