

稀疏低秩矩阵优化模型及其在 语音去噪中的应用

马婷婷¹, 刘 潇¹, 沈春根¹, 薛文娟²

(1. 上海理工大学 理学院, 上海 200093;

2. 上海电力大学 数理学院, 上海 200090)

摘 要: 为了更好地恢复低秩稀疏结构的矩阵, 提出了一种改进的稀疏低秩矩阵优化模型. 不同于单纯地在目标中加入核范数和 l_1 范数来保证目标矩阵的低秩性和稀疏性, 该模型一是通过加入稀疏性约束以更加严格地保证了目标矩阵的稀疏性, 二是通过加入线性约束以刻画目标矩阵中可能存在的先验信息. 为求解该非凸非光滑模型, 首先利用精确罚函数将原模型转化为一个约束DC(两个凸函数之差)优化模型, 其次将其线性化为凸子问题, 并用交替方向乘子法求解该子问题. 在一定的假设条件下, 该算法具有全局收敛性. 数值实验表明该模型在语音去噪问题上具有良好的效果.

关键词: 低秩; 稀疏; 两个凸函数之差; 交替方向乘子法

中图分类号: O221.2

文献标识码: A **文章编号:** 1000-4424(2022)02-0127-15

§1 相关工作

近年来, 随着大数据时代的到来, 数据的处理也随之成为一个令人棘手的问题, 其中数据的恢复是很多学者比较关心的问题. 目前许多问题的核心就是如何根据已知的样本矩阵来恢复目标矩阵, 并且这个目标矩阵一般是具有稀疏, 低秩结构的. 稀疏, 低秩矩阵的恢复在机器学习, 稀疏编码^[1], 压缩传感^[2-3], 信号处理^[4], 推荐系统^[5-6], 基因学^[7]等领域均有广泛应用. 为估计具有稀疏和低秩结构的目标矩阵, 文献[8]中提出模型SLR(Sparse Low-Rank Model)

$$\arg \min_{X \in \mathbf{R}^{m \times n}} \left\{ \frac{1}{2} \|Y - X\|_F^2 + \lambda_0 \|X\|_* + \lambda_1 \|X\|_1 \right\}, \quad (1)$$

其中 $Y \in \mathbf{R}^{m \times n}$ 是给定的数据矩阵, $X \in \mathbf{R}^{m \times n}$ 是待估计矩阵, $\|\cdot\|_*$ 是矩阵的核范数(即 $\|X\|_* = \sum_{i=1}^{\text{rank}(X)} \sigma_i(X)$, 其中 $\sigma_i(X)$ 是矩阵 X 的第 i 大奇异值), 它是用来促进待估计矩阵 X 的低秩性, $\|\cdot\|_1$ 表示矩阵的 l_1 范数(即 $\|X\|_1 = \sum_{i=1}^m \sum_{j=1}^n |X_{ij}|$), 它是用来促进矩阵 X 的稀疏性, $\lambda_i \geq 0, i =$

0, 1 是正则化参数. 在Richard等人^[8]提出的SLR模型的基础上, 文献[9]给出了改进模型ISLR (Improved Sparse Low-Rank Model)

$$\arg \min_{X \in \mathbf{R}^{m \times n}} \left\{ \frac{1}{2} \|Y - X\|_F^2 + \lambda_0 \sum_{i=1}^k \phi(\sigma_i(X); a_0) + \lambda_1 \sum_{i=1}^m \sum_{j=1}^n \phi(X_{i,j}; a_1) \right\}, \quad (2)$$

其中 $k = \min(m, n)$, $\phi: \mathbf{R} \rightarrow \mathbf{R}$ 是一个参数化的非凸惩罚函数, 选取恰当的参数 a_0, a_1 以保证模型(2)是强凸的.

稀疏低秩优化模型也是语音去噪问题的重要模型^[10-11]. 在如今的信息化时代, 语音信号^[12-13]是人类最基本的交流方式. 生活中, 从移动电话到微信聊天都可以看到语音信号的广泛应用, 但是通常语音信号会受到噪音的干扰, 那么如何对含有噪音的语音进行高效的去噪就显得尤为重要了.

实际上, 语音去噪的本质就是将语音信号从带噪信号中恢复出来. 本文数值实验部分的主要目标是从如下的含有高斯白噪声的矩阵 $Y \in \mathbf{R}^{m \times n}$ 中经过一定的方法估计出低秩稀疏目标矩阵 $X \in \mathbf{R}^{m \times n}$,

$$Y = X + W,$$

其中 $W \in \mathbf{R}^{m \times n}$ 为高斯白噪声矩阵.

§2 模型建立及文章贡献

2.1 模型建立

虽然运用模型(1), (2)逼近的目标矩阵是低秩稀疏的, 但对于一些实际问题, 目标矩阵可能存在一些先验信息, 或者对目标矩阵的稀疏度有更严格的要求. 基于此, 提出优化模型SSLR (Strictly Sparse Low-Rank Model)

$$\begin{aligned} \min \quad & \frac{1}{2} \|Y - X\|_F^2 + \rho \|X\|_* \\ \text{s.t.} \quad & \mathcal{A}(X) = b, \\ & \|X\|_0 \leq r, \end{aligned} \quad (3)$$

其中 $Y \in \mathbf{R}^{m \times n}$ 是给定的矩阵, $X \in \mathbf{R}^{m \times n}$ 是待估计的低秩稀疏矩阵, $\|X\|_*$ 的目的是保证待估计矩阵 X 的低秩性, $\rho \in \mathbf{R}$ 是非负正则化参数, $\|X\|_0$ 表示 X 中非零元素个数, $r \leq m \times n$ 是给定的正整数, 模型中的不等式 $\|X\|_0 \leq r$ 的成立严格保证 X 的稀疏性. 线性映射 $\mathcal{A}: \mathbf{R}^{m \times n} \rightarrow \mathbf{R}^{|J|}$ 定义为

$$\mathcal{A}(X) = \begin{pmatrix} \langle A_1, X \rangle \\ \langle A_2, X \rangle \\ \vdots \\ \langle A_{|J|}, X \rangle \end{pmatrix},$$

其中基数 $|J| = t$, 且正整数 $t \leq m \times n$, $\langle A_d, X \rangle := X_{i_d, j_d}$, $d = 1, \dots, t$, $A_d \in \mathbf{R}^{m \times n}$ 的具体形式为

$$(A_d)_{ij} = \begin{cases} 1, & (i, j) = (i_d, j_d), \\ 0, & \text{其它}. \end{cases} \quad (4)$$

右端向量 $b = (b_1, \dots, b_{|J|})^T \in \mathbf{R}^{|J|}$, 指标集 J 为

$$J = \{(i_d, j_d) | d = 1, \dots, t\}.$$

2.2 贡献及文章结构

本文提出的模型(3)与文献[8], 文献[9]提出的模型(1), (2) 相比, 不仅在目标中加入核范数惩罚项以促进目标矩阵的低秩性, 还在约束条件中加入稀疏约束 $\|X\|_0 \leq r$ 以严格保证 X 的稀疏性. 与此同时, 考虑到目标语音信号中存在某些特殊的结构, 即目标矩阵存在一些先验信息, 如某些位置的元素是已知的, 所以在模型(3)中加入了线性约束 $\mathcal{A}(X) = b$.

为求解该非凸非光滑模型(3), 利用精确罚函数将原模型转化为一个约束DC(两个凸函数之差)优化模型, 并证明了DC罚函数的精确性. 用交替方向乘子法解决线性化后的凸子问题并证明了此算法具有全局收敛性, 数值实验表明了该模型和算法的优越性.

2.3 符号和记号说明

为了方便起见, 这里给出下文所需的一些符号定义. 定义 J 的补集 $\bar{J} = \{(i, j) | (i, j) \notin J, i = 1, \dots, m, j = 1, \dots, n\}$. 对于任意的向量 $x, y \in \mathbf{R}^n$, 任意矩阵 $A, B \in \mathbf{R}^{m \times n}$, 定义内积

$$\langle x, y \rangle = x^T y, \quad \langle A, B \rangle = \text{trace}(A^T B).$$

文中其余部分安排如下: §3给出了DC函数的基本知识并将本文模型转化为DC罚函数问题, 给出转化后的问题的KKT条件以及DC罚函数的精确性的详细证明; §4采用交替方向乘子法解决子问题; §5给出了收敛性分析; §6进行了数值实验及数值结果分析; 总结在§7给出.

§3 预备知识

3.1 DC函数

对于稀疏结构的向量版优化问题也已经被很多学者广泛关注^[14-15], 其稀疏约束一般表示为

$$\|w\|_0 \leq r,$$

其中 $w \in \mathbf{R}^n$, $r \leq n$. 由于 l_0 范数的非凸性和不连续性, 导致很多稀疏结构的优化问题非常难于求解. 为了使得所求问题便于求解, 通常将 l_0 范数约束转化 l_1 或者 l_p 范数约束的优化问题. 最近Gotoh等人在文献[14]中提出了用一个DC函数(两个凸函数的差)来代替 l_0 范数的不等式约束, 具体来讲就是写成两个范数差的等式约束:

$$\|w\|_1 - \|w\|_{(r)} = 0,$$

其中

$$\|w\|_1 = \sum_{i=1}^n |w_i|$$

是向量 w 的 l_1 范数, $\|w\|_{(r)}$ 是向量 w 中元素的绝对值的前 r 大的和, 也就是说, 如果向量 w 的元素的绝对值按降序排列为

$$|w_1| \geq |w_2| \geq \dots \geq |w_n|,$$

则

$$\|w\|_{(r)} = \sum_{i=1}^r |w_i|.$$

与文献[14]类似, 本文将模型(3)中的约束 $\|X\|_0 \leq r$ 等价的转化成DC函数约束

$$\|X\|_1 - \|X\|_{(r)} = 0. \quad (5)$$

再利用精确罚函数来处理变换后的等式约束(5), 将原问题(3)转化为惩罚DC函数问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|Y - X\|_F^2 + \rho \|X\|_* + c(\|X\|_1 - \|X\|_{(r)}) \\ \text{s.t.} \quad & \mathcal{A}(X) = b, \end{aligned} \quad (6)$$

其中罚参数 $c > 0$. 设 X^* 是问题(6)的最优解, 当选取合适的罚参数 c , 那么 X^* 也是本文提出的初始模型(3)的一阶稳定点, 详细的证明将在后文给出.

3.2 最优性条件

模型(6)的拉格朗日函数为:

$$\mathcal{L}(X, Z, \lambda) = \frac{1}{2} \|Y - X\|_F^2 + \rho \|X\|_* + c \|X\|_1 - c \|X\|_{(r)} + \langle \mathcal{A}(X) - b, \lambda \rangle,$$

其中 $\lambda \in \mathbf{R}^t$ 是问题(6)等式约束的拉朗格朗日乘子, 下面定理给出问题(6)的一阶必要条件.

定理3.1 设矩阵 X^* 是问题(6)的全局最优解, 则存在 $\xi^* \in \partial(\rho \|X^*\|_*)$, $\eta^* \in \partial(c \|X^*\|_1)$, $\gamma^* \in \partial(c \|X^*\|_{(r)})$, 拉格朗日乘子 $\lambda^* \in \mathbf{R}^t$ 满足KKT条件

$$\begin{cases} 0 = X^* - Y + \xi^* + \eta^* - \gamma^* + \mathcal{A}^*(\lambda^*), \\ \mathcal{A}(X^*) = b. \end{cases}$$

根据文献[14], 这里次微分集合

$$\partial(\|X\|_{(r)}) = \arg \max_S \left\{ \sum_{i=1}^m \sum_{j=1}^n X_{ij} S_{ij} : \sum_{i=1}^m \sum_{j=1}^n |S_{ij}| = r, -1 \leq S_{ij} \leq 1 \right\}, \quad (7)$$

极大化上式中的变量 S 便可得到次梯度, 但并非唯一的. 本文算法里每次迭代用到其中一个次梯度. 具体地, 先将矩阵 X 的元素按照绝对值降序排列

$$|X_{i_1 j_1}| \geq \cdots \geq |X_{i_r j_r}| \geq \cdots \geq |X_{i_{mn} j_{mn}}|,$$

则可得一个次梯度 S 为

$$S_{i_k j_k} = \begin{cases} 1, & k \leq r \text{ 且 } X_{i_k j_k} \geq 0, \\ -1, & k \leq r \text{ 且 } X_{i_k j_k} < 0, \\ 0, & k > r. \end{cases} \quad (8)$$

需要指出的是后文算法中所用次梯度正是由(8)给出. 当 $r = m \times n$ 时, $S \in \partial(c \|X\|_1)$.

3.3 DC罚函数的精确性

首先令集合 $C = \{X \in \mathbf{R}^{m \times n} | \mathcal{A}(X) = b\}$, 记 $\Psi(X) = \frac{1}{2} \|Y - X\|_F^2$.

定理3.2 设 X^* 是DC罚问题(6)的最优解. 若存在常数 θ 满足 $\|X^*\|_F \leq \theta$, 且罚参数 c 满足如下的不等式

$$c > \|\nabla \Psi(\mathbf{0})\| + \frac{3}{2} \theta + \rho,$$

其中 $\mathbf{0}$ 为 $m \times n$ 维全0矩阵, 则 X^* 也是原问题(3)的最优解.

证 由于 X^* 是DC罚问题(6)的最优解, 则 $X^* \in C$ 且满足 $\mathcal{A}(X^*) = b$. 不妨设向量 b 中非零元素的个数为 l , 则 $0 \leq l \leq t$, 存在 $\langle A_d, X^* \rangle := X_{i_d j_d}^* \neq 0$, $d = 1, \dots, l$ 且 $(i_d, j_d) \in J$. 将矩阵 X^* 的元素去掉 $X_{i_d j_d}^*$ 之后降序排列如下:

$$|X_{i_1 j_1}^*| \geq \cdots \geq |X_{i_s j_s}^*| \geq |X_{i_{s+1} j_{s+1}}^*| \geq \cdots \geq |X_{i_\ell j_\ell}^*|,$$

其中 $s = r - l$, $\ell = m \times n - l$. 假设 $\|X^*\|_0 > r$, 则 $|X_{i_{s+1} j_{s+1}}^*| > 0$. 令 $\tilde{X} = X^* - X_{i_{s+1} j_{s+1}}^* \tilde{E}$, 其中 $\tilde{E}$ 表示第 (i_{s+1}, j_{s+1}) 位置元素为1, 其它位置元素为0的 $m \times n$ 维矩阵, 显然 $\tilde{X}$ 也是罚问题(6)的一个可行解.

容易观察出 $\Psi(X) = \frac{1}{2} \|Y - X\|_F^2$ 的梯度 $\nabla \Psi(X)$ 是Lipschitz连续的, 且Lipschitz常数 $L = 1$.

即对于任意矩阵 $U, V \in C$, 有

$$\|\nabla \Psi(U) - \nabla \Psi(V)\|_F \leq \|U - V\|_F$$

成立. 当 $\|X^*\|_F \leq \theta$ 成立时, 不等式

$$\|\nabla \Psi(X^*)\| \leq \|\nabla \Psi(\mathbf{0})\| + \|\nabla \Psi(X^*) - \nabla \Psi(\mathbf{0})\| \leq \|\nabla \Psi(\mathbf{0})\| + \theta, \quad (9)$$

也是显然成立的. 由于 X^* 是 DC 罚问题(6)的最优解, $\tilde{X}$ 是罚问题(6)的可行解, 可以得到

$$\Psi(X^*) + \rho\|X^*\|_* + c(\|X^*\|_1 - \|X^*\|_{(r)}) - \Psi(\tilde{X}) - \rho\|\tilde{X}\|_* - c(\|\tilde{X}\|_1 - \|\tilde{X}\|_{(r)}) \quad (10)$$

$$= \Psi(X^*) - \Psi(\tilde{X}) - \rho(\|\tilde{X}\|_* - \|X^*\|_*) + c|X_{i_{s+1}j_{s+1}}^*| \quad (11)$$

$$\geq X_{i_{s+1}j_{s+1}}^* \langle \nabla \Psi(X^*), \tilde{E} \rangle - \frac{1}{2}(X_{i_{s+1}j_{s+1}}^*)^2 - \rho|X_{i_{s+1}j_{s+1}}^*| + c|X_{i_{s+1}j_{s+1}}^*| \quad (12)$$

$$\geq |X_{i_{s+1}j_{s+1}}^*|(c - \text{sign}(X_{i_{s+1}j_{s+1}}^*) \langle \nabla \Psi(X^*), \tilde{E} \rangle - \frac{\theta}{2} - \rho) \quad (13)$$

$$\geq |X_{i_{s+1}j_{s+1}}^*|(c - \|\nabla \Psi(X^*)\| - \frac{\theta}{2} - \rho) \quad (14)$$

$$\geq |X_{i_{s+1}j_{s+1}}^*|(c - \|\nabla \Psi(\mathbf{0})\| - \frac{3}{2}\theta - \rho), \quad (15)$$

其中不等式(12)可由[16, Theorem 2.1.5]推出, 不等式(13)可由 $\|X^*\|_F \leq \theta$ 推出, 不等式(15)可由(9)推出. 当 $c > \|\nabla \Psi(\mathbf{0})\| + \frac{3}{2}\theta + \rho$ 时, (10)显然大于零, 与 X^* 是 DC 罚问题(6)的最优解相矛盾, 所以 $\|X^*\|_0 \leq r$. 因此 X^* 也是原问题(3)的最优解.

定理3.1成立的前提是 $\|X^*\|_F \leq \theta$ 对任意 $c > 0$ 成立. 下面的定理证明了这样的 $\theta > 0$ 是存在的.

定理3.3 若 X_c^* 是 DC 罚问题(6)的最优解, 则对任何大于零的罚参数 c , $\|X_c^*\|_F$ 是有界的.

证 因为 X_c^* 是 DC 罚问题(6)的最优解, 故存在 $\xi_c^* \in \partial(\rho\|X_c^*\|_*)$, $\eta_c^* \in \partial(c\|X_c^*\|_1)$, $\gamma_c^* \in \partial(c\|X_c^*\|_{(r)})$, $\lambda^* \in \mathbf{R}^t$, 使得

$$\begin{cases} X_c^* - Y + \xi_c^* + \eta_c^* - \gamma_c^* - \mathcal{A}^*(\lambda_c^*) = 0, \\ \mathcal{A}(X_c^*) = b \end{cases} \quad (16)$$

成立.

假设 $\|X_c^*\|_F$ 关于 c 并非有界的, 则存在一个序列 $\{c_k\}$, 使当 $c_k \rightarrow +\infty$ 时, $\|X_{c_k}^*\|_F \rightarrow +\infty$. 对于上述(16)的第一个等式, 两边同时除以 $\|X_{c_k}^*\|_F$ 可得

$$\lim_{c_k \rightarrow +\infty} \frac{X_{c_k}^* - Y + \xi_{c_k}^* + \eta_{c_k}^* - \gamma_{c_k}^* - \mathcal{A}^*(\lambda_{c_k}^*)}{\|X_{c_k}^*\|_F} = 0. \quad (17)$$

由(7)知 $-1 \leq \gamma_{c_k}^* \leq 1$, $-1 \leq \eta_{c_k}^* \leq 1$, 所以 $\gamma_{c_k}^*$ 和 $\eta_{c_k}^*$ 均是有界的. 由[17, Theorem 2]可知, 核范数的次梯度 ξ_c^* 也是有界的. 从(17)可得

$$\lim_{c_k \rightarrow +\infty} \left(\frac{X_{c_k}^*}{\|X_{c_k}^*\|_F} - \frac{\mathcal{A}^*(\lambda_{c_k}^*)}{\|X_{c_k}^*\|_F} \right) = 0.$$

不妨设

$$\lim_{c_k \rightarrow +\infty} \frac{X_{c_k}^*}{\|X_{c_k}^*\|_F} = \tilde{X}_c^*,$$

则

$$\lim_{c_k \rightarrow +\infty} \frac{\mathcal{A}^*(\lambda_{c_k}^*)}{\|X_{c_k}^*\|_F} = \tilde{X}_c^*. \quad (18)$$

取上述(16)的第一个等式中 $c = c_k$, 再与 $X_{c_k}^*$ 作内积并将(16)的第二个等式代入可得

$$\|X_{c_k}^*\|_F^2 - \langle X_{c_k}^*, Y - \xi_{c_k}^* - \eta_{c_k}^* + \gamma_{c_k}^* \rangle - \langle b, \lambda_{c_k}^* \rangle = 0, \quad (19)$$

对于上述等式(19)两边同时除以 $\|X_{c_k}^*\|_F^2$ 可得

$$\lim_{c_k \rightarrow +\infty} \frac{\|X_{c_k}^*\|_F^2 - \langle X_{c_k}^*, Y - \xi_{c_k}^* - \eta_{c_k}^* + \gamma_{c_k}^* \rangle - \langle b, \lambda_{c_k}^* \rangle}{\|X_{c_k}^*\|_F^2} = 0.$$

由于 $\xi_{c_k}^*$, $\eta_{c_k}^*$, $\gamma_{c_k}^*$ 均是有界的, 有

$$\lim_{c_k \rightarrow +\infty} \frac{\langle b, \lambda_{c_k}^* \rangle}{\|X_{c_k}^*\|_F^2} = 1.$$

因此 $\lim_{c_k \rightarrow +\infty} \frac{\lambda_{c_k}^*}{\|X_{c_k}^*\|_F} \neq 0$. 结合定义式(4)中 $A_1, A_2, \dots, A_t$ 的线性无关性, 可以得到

$$\lim_{c_k \rightarrow +\infty} \frac{\mathcal{A}^*(\lambda_{c_k}^*)}{\|X_{c_k}^*\|_F} = \lim_{c_k \rightarrow +\infty} \mathcal{A}^*\left(\frac{\lambda_{c_k}^*}{\|X_{c_k}^*\|_F}\right) \|X_{c_k}^*\|_F \rightarrow +\infty.$$

这与(18)矛盾, 因此对任何大于零的罚参数 c , $\|X_c^*\|_F$ 是有界的.

定理3.4 如果 X^* 是DC罚问题(6)的最优解, 且罚参数 c 满足

$$c > \|\nabla \Psi(\mathbf{0})\| + \frac{3}{2}\theta + \rho,$$

则 X^* 也是原问题(3)的最优解.

定理3.4的证明可由定理3.2和定理3.3立即得到.

3.4 相关问题

在介绍算法之前, 先考虑一个与模型(6)的解析解相关的无约束优化问题,

$$Z^* = \arg \min_Z \frac{1}{2} \|Z - W\|_F^2 + \lambda \|Z\|_*, \quad (20)$$

其中 $Z \in \mathbf{R}^{m \times n}$ 且 $W \in \mathbf{R}^{m \times n}$ 是已知的, $\text{rank}(W) = \bar{r}$. 根据文献[18]中的Lemma 1可知, 问题(20)具有封闭解 $Z^* = S_\lambda(W)$, 其中 $S_\lambda(W) := UD_\lambda V^T$, $D_\lambda = \text{diag}[(d_1 - \lambda)_+, \dots, (d_{\bar{r}} - \lambda)_+]$, UDV^T 是 W 的奇异值分解, $D = \text{diag}[d_1, \dots, d_{\bar{r}}]$, $U \in \mathbf{R}^{m \times \bar{r}}$, $V \in \mathbf{R}^{n \times \bar{r}}$, $t_+ = \max(t, 0)$. 这里问题(20)的封闭解将为下文变量 Z 的求解提供一种思路.

§4 算法研究

4.1 模型线性化

由于问题(6)的目标函数中含有非凸函数项 $-c\|X\|_{(r)}$, 不但导致凸优化对偶理论不能直接应用, 而且与非光滑函数项 $c\|X\|_1$ 相耦合, 导致临近梯度类方法不可以直接处理该非光滑项, 从而给问题求解带来了较大困难. 为了克服问题(6)的目标函数中非凸函数项 $-c\|X\|_{(r)}$ 带来的求解困难, 可以通过线性化该非凸函数以降低问题难度. 由于 $c\|X\|_{(r)}$ 是凸函数, 有凸函数的次梯度不等式

$$c\|X\|_{(r)} \geq c\|X^l\|_{(r)} + \langle S^l, X - X^l \rangle$$

成立, 其中 $S^l \in \partial(c\|X\|_{(r)})$. 由于 l_1 范数和核范数的同时存在, 问题的求解并不容易. 因此通过

引入一个新的变量 $Z \in \mathbf{R}^{m \times n}$ 使其分离, 于是可以把较难处理的问题(6)转化成子问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|Y - X\|_F^2 + \rho \|Z\|_* + c \|X\|_1 - \langle S^l, X \rangle \\ \text{s.t.} \quad & \mathcal{A}(X) = b, \\ & X = Z. \end{aligned} \quad (21)$$

4.2 交替方向乘子法(ADMM)

交替方向乘子法是一种求解有约束问题的最优化算法, 通过分解过程将比较大的难求解的全局问题分解为多个较小的, 易于求解的局部子问题, 并通过协调子问题的解得到全局问题的解. 交替方向乘子法最早由Gabay, Mercier等人^[19]提出, 后来有很多学者对其相关理论进行了深入的研究, 如何炳生等人^[20]研究了一种自适应罚参数的ADMM算法. 2011年, Boyd等人^[21]对ADMM方法及其应用进行全面综述.

针对本文要优化的原始问题(3), 已在4.1节转化为要解决含有多个待估计变量的子问题(21), 因此可以采用交替方向乘子法来解决此子问题. 其增广拉格朗日函数 $\mathcal{L}$ 定义为

$$\begin{aligned} \mathcal{L}_{\sigma_1, \sigma_2}(X, Z, \lambda, \Lambda) = & \frac{1}{2} \|Y - X\|_F^2 + \rho \|Z\|_* + c \|X\|_1 - \langle S^l, X \rangle \\ & + \langle \lambda, \mathcal{A}(X) - b \rangle + \langle \Lambda, X - Z \rangle + \frac{\sigma_1}{2} \|\mathcal{A}(X) - b\|_2^2 + \frac{\sigma_2}{2} \|X - Z\|_F^2, \end{aligned}$$

其中 $\Lambda \in \mathbf{R}^{m \times n}$, $\lambda \in \mathbf{R}^t$ 是拉格朗日乘子, σ_1, σ_2 是大于零的常数. 子问题(21)用交替方向乘子法(ADMM)迭代求解, 迭代形式为

$$\begin{cases} X^{k+1} = \arg \min_X \mathcal{L}_{\sigma_1 \sigma_2}(X, Z^k, \lambda^k, \Lambda^k), \\ Z^{k+1} = \arg \min_Z \mathcal{L}_{\sigma_1 \sigma_2}(X^{k+1}, Z, \lambda^k, \Lambda^k), \\ \lambda^{k+1} = \lambda^k + \sigma_1 (\mathcal{A}(X^{k+1}) - b), \\ \Lambda^{k+1} = \Lambda^k + \sigma_2 (X^{k+1} - Z^{k+1}). \end{cases} \quad (22)$$

对于式(22)中变量 X 的更新, 可以分为以下的两个部分考虑: 对于2.1节中定义的指标集 $J = \{(i_d, j_d) | d = 1, \dots, t\}$, 当 $(i, j) \in J$ 时

$$X_{i_d j_d}^{k+1} = \arg \min_X \left\{ c |X_{i_d j_d}| + \frac{\sigma_1 + \sigma_2 + 1}{2} (X_{i_d j_d} - \frac{1}{\sigma_1 + \sigma_2 + 1} P_{i_d j_d})^2 \right\}, \quad (23)$$

由 l_1 收缩算子可得式(23)的封闭解

$$X_{i_d j_d}^{k+1} = \frac{1}{\sigma_1 + \sigma_2 + 1} \text{Shrink}(P_{i_d j_d}, c) = \frac{1}{\sigma_1 + \sigma_2 + 1} (\max\{|P_{i_d j_d}| - c, 0\} \text{sign}(P_{i_d j_d})), \quad (24)$$

其中 $P_{i_d j_d} = S_{i_d j_d}^l + \sigma_1 b_d + \sigma_2 Z_{i_d j_d}^k + Y_{i_d j_d} - \lambda_d^k - \Lambda_{i_d j_d}^k$, $\text{sign}(\cdot)$ 是符号函数; 当 $(i, j) \in \bar{J}$ 时, 类似可得封闭解

$$X_{ij}^{k+1} = \frac{1}{\sigma_2 + 1} (\max\{|Q_{ij}| - c, 0\} \text{sign}(Q_{ij})), \quad (25)$$

其中 $Q_{ij} = S_{ij}^l + \sigma_2 Z_{ij}^k + Y_{i_d j_d} - \Lambda_{ij}^k$. 对于式(22)中变量 Z 的更新, 把二次项放在一起并忽略和变量 Z 无关的项, 整理可以写成

$$Z^{k+1} = \arg \min_Z \left\{ \frac{1}{2} \|Z - (\frac{1}{\sigma_2} \Lambda^k + X^{k+1})\|_F^2 + \frac{\rho}{\sigma_2} \|Z\|_* \right\}, \quad (26)$$

由3.4节问题(20)的封闭解形式, 可以求得类似于 $S_\lambda(W)$ 的问题(26)的封闭解.

4.3 算法框架

对于算法1, 有以下两点需要注意.

1. 算法1中 $S^l \in (\partial c \|X^l\|_{(r)}), l = 0, 1, 2, \dots$ 是通过式(8)计算得到;
2. 在数值实验中, 设置算法1的停止条件为: 达到最大迭代次数或问题(21)内外循环的解达到

算法 1: 基于ADMM的DC优化算法

```

1  初始化:  $X^0, Z^0, \Lambda^0, \lambda^0, \varepsilon_1, \varepsilon_2, S^0 \in \partial(c\|X^0\|_{(r)});$ 
2  for  $l = 0, 1, 2, \dots$  do
3      计算次梯度  $S^l \in \partial(c\|X^l\|_{(r)});$ 
4      for  $k = 0, 1, 2, \dots$  do
5           $X^{k+1}$  的更新由式(24), (25)可得;
6           $Z^{k+1}$  的更新由式(26)可得;
7           $\lambda^{k+1} = \lambda^k + \sigma_1(\mathcal{A}(X^{k+1}) - b);$ 
8           $\Lambda^{k+1} = \Lambda^k + \sigma_2(X^{k+1} - Z^{k+1});$ 
9          if  $\|X^{k+1} - X^k\|_F^2 \leq \varepsilon_1$  且  $\|Z^{k+1} - Z^k\|_F^2 \leq \varepsilon_2$  then
10             输出:  $X^{k+1}, Z^{k+1};$ 
11         end
12     end
13      $X^{l+1} = X^{k+1}, Z^{l+1} = Z^{k+1};$ 
14     if  $\|X^{l+1} - X^l\|_F^2 \leq \varepsilon_1$  且  $\|Z^{l+1} - Z^l\|_F^2 \leq \varepsilon_2$  then
15         输出:  $X^{l+1}, Z^{l+1};$ 
16     end
17 end

```

预先给定的精度.

§5 收敛性分析

本节给出算法1的全局收敛性. 算法内循环收敛性的证明主要基于He等人^[22]和Zhou等人^[23]的工作.

引理5.1 假设 (X^*, Z^*) 是模型(21)的最优解, 且 λ^*, Λ^* 分别是对应的等式约束 $\mathcal{A}(X) = b$, $X = Z$ 的最优对偶变量, 则由算法1步骤5-8产生的序列 $\{(X^k, Z^k, \lambda^k, \Lambda^k)\}$ 满足

$$\|V^{k+1} - V^*\|_H^2 \leq \|V^k - V^*\|_H^2 - \|V^k - V^{k+1}\|_H^2, \quad (27)$$

其中 $V^k = (Z^k, \lambda^k, \Lambda^k)^T$, $V^* = (Z^*, \lambda^*, \Lambda^*)^T$.

需要指出这里 V^k 和 V^* 不是传统意义上分块矩阵, 而是类似于Matlab中元胞结构, 因为这里面混合了向量与矩阵, 其中的转置运算类似矩阵转置. 分块矩阵 H 为

$$\begin{pmatrix} \sigma_2 I_{m \times n} & 0 & 0 \\ 0 & \frac{1}{\sigma_1} I_{t \times 1} & 0 \\ 0 & 0 & \frac{1}{\sigma_2} I_{m \times n} \end{pmatrix},$$

加权范数以及内积的定义分别为

$$\|V\|_H := \langle V, HV \rangle, \quad \langle U, V \rangle_H := \langle U, HV \rangle.$$

证 首先将 $\|V^k - V^*\|_H^2$ 展开成如下形式

$$\begin{aligned}\|V^k - V^*\|_H^2 &= \|(V^{k+1} - V^*) + (V^k - V^{k+1})\|_H^2 \\ &= \|V^{k+1} - V^*\|_H^2 + 2\langle V^{k+1} - V^*, V^k - V^{k+1} \rangle_H + \|V^k - V^{k+1}\|_H^2,\end{aligned}\quad (28)$$

由(28)可知, 要保证不等式(27)的成立, 只需证明不等式

$$\langle V^{k+1} - V^*, V^k - V^{k+1} \rangle_H \geq 0$$

成立.

为了便于描述, 记模型(21)的目标函数为

$$\begin{cases} \Theta_1(X) = \frac{1}{2}\|Y - X\|_F^2 + c\|X\|_1 - \langle S^l, X \rangle, \\ \Theta_2(Z) = \rho\|Z\|_*.\end{cases}$$

现假设 (X^*, Z^*) 是模型(21)的最优解, λ^*, Λ^* 分别为对应等式约束 $\mathcal{A}(X) = b, X = Z$ 的最优对偶变量. 于是对于任意的 (X, Z, λ, Λ) , 有变分不等式

$$\begin{cases} \Theta_1(X) - \Theta_1(X^*) + \langle X - X^*, \mathcal{A}^*(\lambda^*) \rangle + \langle X - X^*, \Lambda^* \rangle \geq 0, \\ \Theta_2(Z) - \Theta_2(Z^*) + \langle Z - Z^*, -\Lambda^* \rangle \geq 0, \\ \langle \lambda - \lambda^*, -\mathcal{A}(X) + b \rangle \geq 0, \\ \langle \Lambda - \Lambda^*, -X + Z \rangle \geq 0\end{cases}\quad (29)$$

成立.

为了更简单的表示上述的变分不等式, 定义

$$\mathcal{F}(X, Z, \lambda, \Lambda) := \begin{pmatrix} \mathcal{A}^*(\lambda) + \Lambda \\ -\Lambda \\ -\mathcal{A}(X) + b \\ -X + Z \end{pmatrix} = F(W),$$

其中

$$F = \begin{pmatrix} 0 & 0 & \mathcal{A}^* & I \\ 0 & 0 & 0 & -I \\ -\mathcal{A} & 0 & 0 & 0 \\ -I & I & 0 & 0 \end{pmatrix}, \quad W = (X, Z, \lambda, \Lambda)^T, \quad W \in \Omega (\Omega = \mathbf{R}^{m \times n} \times \mathbf{R}^{m \times n} \times \mathbf{R}^t \times \mathbf{R}^{m \times n}).$$

注意, 由于线性算子 $\mathcal{A}$ 的出现, 这里的 F 并非传统意义上的分块矩阵, 而是类似于Matlab的元胞结构, 乘法类似于分块矩阵乘积. 显然 F 是反对称的. 另记 $\Theta(X, Z) = \Theta_1(X) + \Theta_2(Z)$, 则可将(29)写成如下的形式:

$$\Theta(X, Z) - \Theta(X^*, Z^*) + \langle F(W^*), W - W^* \rangle \geq 0, \quad (30)$$

其中 $W^* = (X^*, Z^*, \lambda^*, \Lambda^*)^T, W^* \in \Omega$.

根据(22), 存在 $\eta^{k+1} \in \partial\|X^{k+1}\|_1, \xi^{k+1} \in \partial\|Z^{k+1}\|_*$, 使得

$$\begin{cases} X^{k+1} - Y + c\eta^{k+1} - S^l + \mathcal{A}^*(\lambda^k) + \Lambda^k + \sigma_1\mathcal{A}^*(\mathcal{A}(X^{k+1}) - b) \\ \quad + \sigma_2(X^{k+1} - Z^k) = 0, \\ \rho\xi^{k+1} - \Lambda^k + \sigma_2(Z^{k+1} - X^{k+1}) = 0, \\ \lambda^{k+1} = \lambda^k + \sigma_1(\mathcal{A}(X^{k+1}) - b), \\ \Lambda^{k+1} = \Lambda^k + \sigma_2(X^{k+1} - Z^{k+1}) \end{cases}$$

成立. 可将它等价地写成如下的变分不等式

$$\begin{cases} \theta_1(X) - \theta_1(X^{k+1}) + \langle X - X^{k+1}, \mathcal{A}^*(\lambda^k) + \sigma_1 \mathcal{A}^*(\mathcal{A}(Z^{k+1}) - b) \rangle \\ + \langle X - X^{k+1}, \Lambda^k + \sigma_2(X^{k+1} - Z^k) \rangle \geq 0, \\ \theta_2(Z) - \theta_2(Z^{k+1}) + \langle Z - Z^{k+1}, -\Lambda^k \rangle + \sigma_2 \langle Z - Z^{k+1}, Z^{k+1} - X^{k+1} \rangle \geq 0, \\ \lambda^{k+1} = \lambda^k + \sigma_1(\mathcal{A}(X^{k+1}) - b), \\ \Lambda^{k+1} = \Lambda^k + \sigma_2(X^{k+1} - Z^{k+1}). \end{cases} \quad (31)$$

进一步, 将 $\lambda^k = \lambda^{k+1} - \sigma_1(\mathcal{A}(X^{k+1}) - b)$ 和 $\Lambda^k = \Lambda^{k+1} - \sigma_2(X^{k+1} - Z^{k+1})$ 分别代入到(31)的前两个不等式可得

$$\begin{cases} \theta_1(X) - \theta_1(X^{k+1}) + \langle X - X^{k+1}, \mathcal{A}^*(\lambda^{k+1}) + \Lambda^{k+1} \rangle + \sigma_2 \langle X - X^{k+1}, Z^{k+1} - Z^k \rangle \geq 0, \\ \theta_2(Z) - \theta_2(Z^{k+1}) + \langle Z - Z^{k+1}, -\Lambda^{k+1} \rangle \geq 0, \end{cases}$$

把上面两式相加可得

$$\begin{aligned} & \theta(X, Z) - \theta(X^{k+1}, Z^{k+1}) + \langle F(W^{k+1}), W - W^{k+1} \rangle + \sigma_2 \langle X - X^{k+1}, Z^{k+1} - Z^k \rangle \\ & + \frac{1}{\sigma_1} \langle \lambda - \lambda^{k+1}, \lambda^{k+1} - \lambda^k \rangle + \frac{1}{\sigma_2} \langle \Lambda - \Lambda^{k+1}, \Lambda^{k+1} - \Lambda^k \rangle \geq 0. \end{aligned}$$

于是, 将上述不等式移项可得

$$\begin{aligned} & \theta(X, Z) - \theta(X^{k+1}, Z^{k+1}) + \langle F(W^{k+1}), W - W^{k+1} \rangle \\ & + \sigma_2 \langle X - X^{k+1}, Z^{k+1} - Z^k \rangle \\ & \geq \frac{1}{\sigma_1} \langle \lambda - \lambda^{k+1}, \lambda^k - \lambda^{k+1} \rangle + \frac{1}{\sigma_2} \langle \Lambda - \Lambda^{k+1}, \Lambda^k - \Lambda^{k+1} \rangle. \end{aligned} \quad (32)$$

令 $W = W^*$, 上述不等式(32)可以等价地写成

$$\begin{aligned} & \frac{1}{\sigma_1} \langle \lambda^{k+1} - \lambda^*, \lambda^k - \lambda^{k+1} \rangle + \frac{1}{\sigma_2} \langle \Lambda^{k+1} - \Lambda^*, \Lambda^k - \Lambda^{k+1} \rangle \\ & \geq \theta(X^{k+1}, Z^{k+1}) - \theta(X^*, Z^*) + \langle F(W^{k+1}), W^{k+1} - W^* \rangle \\ & + \sigma_2 \langle X^{k+1} - X^*, Z^{k+1} - Z^k \rangle \\ & \geq \theta(X^{k+1}, Z^{k+1}) - \theta(X^*, Z^*) + \langle F(W^*), W^{k+1} - W^* \rangle \\ & + \langle F(W^{k+1}) - F(W^*), W^{k+1} - W^* \rangle + \sigma_2 \langle X^{k+1} - X^*, Z^{k+1} - Z^k \rangle. \end{aligned}$$

令 $W = W^{k+1}$, 于是由(30)知

$$\theta(X^{k+1}, Z^{k+1}) - \theta(X^*, Z^*) + \langle F(W^*), W^{k+1} - W^* \rangle \geq 0.$$

因为 F 是反对称的, 所以 F 是单调的, 因此有

$$\langle F(W^{k+1}) - F(W^*), W^{k+1} - W^* \rangle \geq 0$$

成立. 所以不等式

$$\frac{1}{\sigma_1} \langle \lambda^{k+1} - \lambda^*, \lambda^k - \lambda^{k+1} \rangle + \frac{1}{\sigma_2} \langle \Lambda^{k+1} - \Lambda^*, \Lambda^k - \Lambda^{k+1} \rangle \geq \sigma_2 \langle X^{k+1} - X^*, Z^{k+1} - Z^k \rangle \quad (33)$$

显然成立. 在不等式(33)两边同时加上 $\sigma_2 \langle Z^{k+1} - Z^*, Z^k - Z^{k+1} \rangle$ 可得

$$\begin{aligned} & \sigma_2 \langle Z^{k+1} - Z^*, Z^k - Z^{k+1} \rangle + \frac{1}{\sigma_1} \langle \lambda^{k+1} - \lambda^*, \lambda^k - \lambda^{k+1} \rangle \\ & + \frac{1}{\sigma_2} \langle \Lambda^{k+1} - \Lambda^*, \Lambda^k - \Lambda^{k+1} \rangle \\ & \geq \sigma_2 \langle Z^{k+1} - Z^*, Z^k - Z^{k+1} \rangle + \sigma_2 \langle X^{k+1} - X^*, Z^{k+1} - Z^k \rangle \\ & = \sigma_2 \langle Z^{k+1} - Z^* - (X^{k+1} - X^*), Z^k - Z^{k+1} \rangle \\ & = \sigma_2 \langle Z^{k+1} - X^{k+1}, Z^k - Z^{k+1} \rangle \\ & = \langle \Lambda^k - \Lambda^{k+1}, Z^k - Z^{k+1} \rangle. \end{aligned} \quad (34)$$

由(29)可推出以下两个不等式

$$\begin{cases} \Theta_2(Z^k) - \Theta_2(Z^{k+1}) + \langle Z^k - Z^{k+1}, -\Lambda^{k+1} \rangle \geq 0, \\ \Theta_2(Z^{k+1}) - \Theta_2(Z^k) + \langle Z^{k+1} - Z^k, -\Lambda^k \rangle \geq 0 \end{cases}$$

显然成立. 将上述两个不等式相加得

$$\langle \Lambda^k - \Lambda^{k+1}, Z^k - Z^{k+1} \rangle \geq 0. \quad (35)$$

由不等式(34)和(35)可以得到

$$\begin{aligned} & \sigma_2 \langle Z^{k+1} - Z^*, Z^k - Z^{k+1} \rangle + \frac{1}{\sigma_1} \langle \lambda^{k+1} - \lambda^*, \lambda^k - \lambda^{k+1} \rangle \\ & + \frac{1}{\sigma_2} \langle \Lambda^{k+1} - \Lambda^*, \Lambda^k - \Lambda^{k+1} \rangle \\ & = \langle V^{k+1} - V^*, V^k - V^{k+1} \rangle_H \geq \langle \Lambda^k - \Lambda^{k+1}, Z^k - Z^{k+1} \rangle \geq 0. \end{aligned}$$

因此不等式(27)的成立得证. 至此引理5.1证毕.

基于引理5.1, 可得下面两个收敛性定理.

定理5.1 对于任意给定的初始点, 由算法1中步骤5-8产生的序列 $\{(X^k, Z^k, \lambda^k, \Lambda^k)\}$ 都会收敛到模型(21)的最优解.

证 引理5.1的成立可以得到 $\|V^{k+1} - V^k\|_H^2 \rightarrow 0$, $\|V^{k+1} - V^*\|_H^2$ 是单调非增且收敛的. 由 $\|V^{k+1} - V^k\|_H^2 \rightarrow 0$ 可知 $\lambda^{k+1} - \lambda^k \rightarrow 0$, $\Lambda^{k+1} - \Lambda^k \rightarrow 0$, $Z^{k+1} - Z^k \rightarrow 0$. 又因为 $\Lambda^{k+1} = \Lambda^k + \sigma_2(X^{k+1} - Z^{k+1})$ 的成立, 因此有 $X^{k+1} - X^k \rightarrow 0$, $X^k - Z^k \rightarrow 0$. 所以序列 $\{(X^k, Z^k, \lambda^k, \Lambda^k)\}$ 必然存在一个极限点 $\tilde{V}^* = (\tilde{X}^*, \tilde{Z}^*, \tilde{\lambda}^*, \tilde{\Lambda}^*)^T$. 其中 $(\tilde{X}^*, \tilde{Z}^*)$ 是模型(21)的最优解, $(\tilde{\lambda}^*, \tilde{\Lambda}^*)$ 分别是对应的等式约束 $\mathcal{A}(X) = b$, $X = Z$ 的最优对偶变量. 所以定理5.1得证.

定理5.2 对于任意给定的初始点, 由算法1中产生的序列 $\{(X^l, Z^l, \lambda^l, \Lambda^l)\}$ 都会收敛到模型(6)的一阶稳定点.

由文献[14]中DCA的性质可知定理5.2是显然成立的, 更加详尽的证明见文献[24-25]. 另外, 关于模型(6)的解等价于模型(3)的解, 已在本文3.3节中详细阐述. 于是, 由以上定理可知由算法1产生的序列是本文提出的模型(3)的一阶稳定点.

§6 数值实验

本节数值实验主要是解决含有高斯白噪声的语音信号去噪问题. 将稀疏低秩估计模型应用到含噪语音信号的频谱图上, 并将估计的频谱图进行逆变换到时域上, 以得到去噪语音信号. 具体来说, 如果 y 是输入的含有噪音的语音信号, 采用SSLR模型得到去噪语音信号 x^* ,

$$x^* = S^\dagger \{\text{SSLR}(S(y))\},$$

其中 S , $S^\dagger$ 分别表示短时傅里叶变换^[26-27](Short-Time Fourier Transform(STFT))及其逆变换, STFT窗口之间实现50%的重合, 一个窗口设置64个样本, DFT长度设置为512个样本.

针对于本文提出的模型SSLR(3)中的两个约束的参数设置如下, 对于输入的语音信号 y , 经STFT后的矩阵 X 维数为 $m \times n$, 对于目标矩阵 X 中的先验信息的个数设置为 $20\% \times m \times n$, 任意取定先验信息位置, 即得线性映射 $\mathcal{A}(X) = b$ 中向量 $b \in \mathbf{R}^{m \times n/5}$. 对于严格控制目标矩阵 X 稀疏度的不等式约束 $\|X\|_0 \leq r$, 设置参数 $r = 5\% \times m \times n$.

将本文提出的模型SSLR(3)分别与模型SLR(1), ISLR(2)作对比. 对于去噪后的语音信号, 分别从频谱图, 信噪比(SNR), 平方根误差(RSE)三个方面进行对比观察去噪效果.

图1-3中(a)图的语音信号为同一段语音信号, 来自文献[9], 记为语音信号a; 图1-3中(b)图的语音信号是相同的, 来自文献[28]的实验数据集, 记为语音信号b. 相比于语音信号a, 信号b的时间为1秒.

6.1 频谱图

分别在语音信号a, b上分别加入高斯白噪声($\sigma = 0.03, \sigma = 0.1$), 实验中将本文提出的SSLR模型分别和SLR, ISLR模型进行对比, 结果展示在图1.

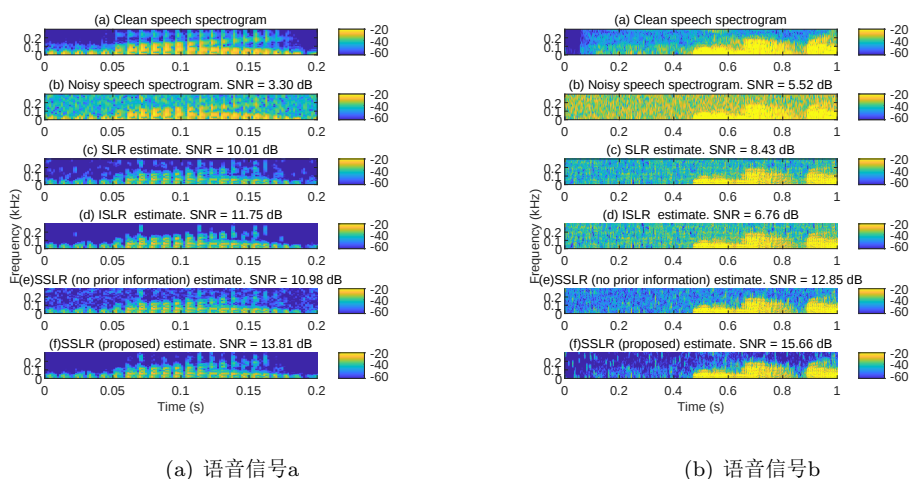


图 1 不同模型的语音信号频谱图

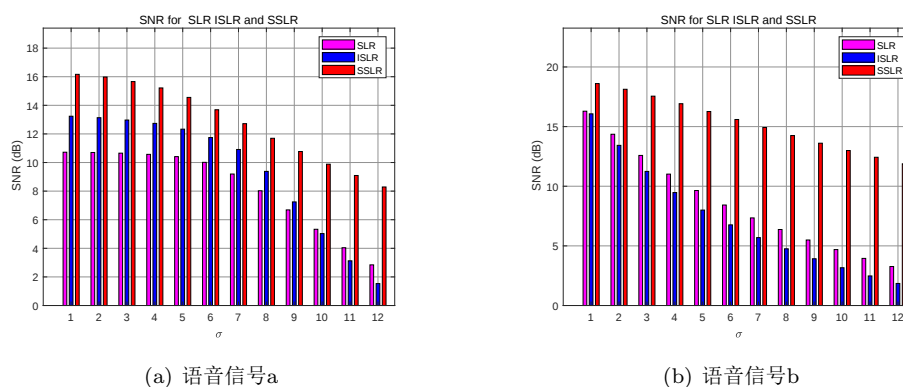


图 2 不同噪音影响下的信噪比(SNR)

图1(a), (b)分别是原始语音信号和加入噪音后的语音信号的频谱图, (c), (d), (f)分别为通过SLR, ISLR, SSLR模型得到的去噪语音信号频谱图, (e)为SSLR模型无先验信息(即不含有等式约束 $A(X) = b$)情况下得到的去噪语音信号频谱图. 从(c), (d), (f)可以看出, 对于两段不同时间的语音信号, 与模型SLR, ISLR相比, SSLR模型得到的去噪语音信号有更高的SNR, 并且此方

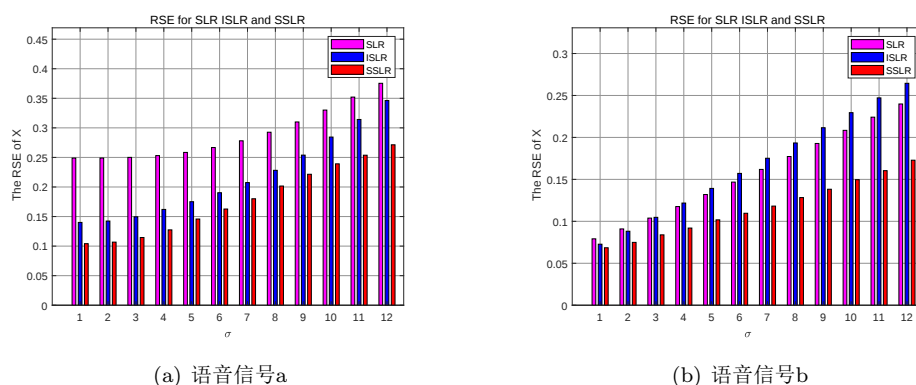


图 3 不同噪音影响下的RSE

法具有更少的伪影. 从(e), (f)可以得到, 与未加入先验信息的(e)相比, (f)具有更高的SNR且伪影相对来说也是更少的. 综上, 文章提出的带有先验信息的SSLR模型和算法的表现都是良好的.

6.2 信噪比(SNR)

图2为不同高斯白噪声 σ 对SLR, ISLR, SSLR模型去噪后的SNR的影响, 图2(a)中 σ 从0.005到0.06(以0.005递增), 共12组数据; 图2(b)中 σ 从0.05到0.16(以0.01递增), 共12组数据. 图2表明, 随着 σ 的增大, 虽然三种模型去噪后的语音信号的SNR都有所降低, 但对于两个长度不同的语音信号a, b, 利用本文提出的SSLR模型恢复后的语音信号的SNR是远远高于另外两种模型的.

6.3 平方根误差(RSE)

图3为不同高斯白噪声 σ 对SLR, ISLR, SSLR模型去噪后的RSE的影响, 图3(a)中 σ 从0.005到0.06(以0.005递增), 共12组数据; 图3(b)中 σ 从0.05到0.16(以0.01递增), 共12组数据. 首先给出RSE的计算方法

$$RSE = \frac{\|X_{\text{est}} - X_{\text{org}}\|_F}{\|X_{\text{org}}\|_F},$$

其中 X_{org} 是给定的初始矩阵, X_{est} 是目标矩阵. 图3表明, 随着 σ 的增大, ISLR和SLR模型更新的RSE值是有明显上升趋势的, 而且远高于本文提出的模型SSLR得到的估计的矩阵的RSE, 并且随着 σ 的增大, SSLR模型更新得到的RSE依旧是小於另外两个模型求解的结果.

§7 结论

本文提出了一种具有先验信息的严格稀疏低秩矩阵优化模型SSLR, 运用基于交替方向乘子法(ADMM)的DC方法解决此模型. 该算法的思想是先利用DC技术将非凸项函数线性化得到凸子问题, 再利用ADMM求解该凸子问题. 在一定的假设条件下, 所提的算法具有全局收敛性. 此外, 从频谱图, 信噪比(SNR), 平方根误差(RSE)三个方面与同类模型(SLR和ISLR)进行语音去噪实验对比表明: 所提的SSLR模型较其它两个模型更优越.

参考文献:

- [1] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding[J]. Journal of Machine Learning Research, 2009, 11(1): 19-60.

- [2] Candes E J, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. *IEEE Transactions on Information Theory*, 2006, 52(2): 489-509.
- [3] Migliore M D. A simple introduction to compressed sensing/sparse recovery with applications in antenna measurements[J]. *IEEE Antennas and Propagation Magazine*, 2014, 56(2): 14-26.
- [4] Oymak S, Jalali A, Fazel M, et al. Simultaneously structured models with application to sparse and low-rank matrices[J]. *IEEE Transactions on Information Theory*, 2014, 61(5): 2886-2908.
- [5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. *Computer*, 2009, 42(8): 30-37.
- [6] Luo Xin, Zhou Mengchu, Xia Yunni, et al. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems[J]. *IEEE Transactions on Industrial Informatics*, 2014, 10(2): 1273-1284.
- [7] Kim H, Park H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis[J]. *Bioinformatics*, 2007, 23(12): 1495-1502.
- [8] Richard E, Savalle P A, Vayatis N. Estimation of simultaneously sparse and low-rank matrices[A]. //Proc of the 29th International Conference on International Conference on Machine Learning[C]. Edinburgh, Scotland: ACM Press, 2012: 51-58.
- [9] Parekh A, Selesnick I W. Improved sparse low-rank matrix estimation[J]. *Signal Processing*, 2017, 139: 62-69.
- [10] Huang Jianjun, Zhang Xiongwei, Zhang Yafei, et al. Speech denoising via low-rank and sparse matrix decomposition[J]. *ETRI Journal*, 2014, 36(1): 167-170.
- [11] Zhou Chen, Ellis D P W. Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition[A].// 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics[C]. New Paltz: NY, 2013: 1-4.
- [12] Ozerov A, Vincent E, Bimbot F. A general flexible framework for the handling of prior information in audio source separation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(4): 1118-1123.
- [13] 向前, 唐勇. 基于生成对抗网络的汉语语音增强技术研究[J]. *计算机应用研究*, 2020, 37(S2): 150-151, 155.
- [14] Gotoh J, Takeda A, Tono K. DC formulations and algorithms for sparse optimization problems[J]. *Mathematical Programming*, 2018, 169(1): 141-176.
- [15] Shen Chungen, Liu Xiao. Solving nonnegative sparse optimization problems via DC quadratic-piecewise-linear approximations[J]. *Journal of Global Optimization*, 2021: 1-37. // <http://doi.org/10.1007/s10898-021-01028-9>.
- [16] Nesterov Y. *Introductory Lectures on Convex Optimization: A Basic Course*[M]. New York: Springer, 2004.
- [17] Watson G A. Characterization of the subdifferential of some matrix norms-ScienceDirect[J]. *Linear Algebra & its Applications*, 1992, 170(6): 33-45.
- [18] Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices[J]. *Journal of Machine Learning Research*, 2010, 11: 2287-2322.
- [19] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. *Computers Mathematics with Applications*, 1976, 2(1): 17-40.

- [20] He Bingsheng, Yang Hai, Wang Shengli. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities[J]. Journal of Optimization Theory and Applications, 2000, 106(2): 337-356.
- [21] Boyd S, Parikh N, Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning, 2010, 3(1): 1-122.
- [22] He Bingsheng, Ma Feng, Yuan Xiaoming. Convergence study on the symmetric version of ADMM with larger step sizes[J]. SIAM Journal on Imaging Sciences, 2016, 9(3): 1467-1501.
- [23] Zhou Shenglong, Xiu Naihua, Luo Ziyang, et al. Sparse and low-rank covariance matrix estimation [J]. Journal of the Operations Research Society of China, 2015, 3(2): 231-250.
- [24] Phan Dinh T, Le Thi H A. Convex analysis approach to D.C. programming: theory, algorithm and applications[J]. Acta Mathematica Vietnamica, 1997, 22(1): 289-355.
- [25] Phan Dinh T, Le Thi H A. Transactions on Computational Intelligence XIII[A].// Recent advances in DC programming and DCA[C], Volume 8342 of the Lecture Notes in Computer Science, Berlin: Springer, 2014, 1-37.
- [26] 张贤达, 保铮. 非平稳信号分析与处理[M]. 北京:国防工业出版社, 1998.
- [27] 殷晓中, 于盛林. 信号的时频分析理论与应用评述[J]. 现代电子技术, 2006, 21: 118-120.
- [28] 赵力, 梁瑞宇, 魏昕, 等. 语音信号处理[M]. 第3版. 北京: 机械工业出版社, 2016.

Sparse low-rank matrix optimization model with application in speech denoising

MA Ting-ting¹, LIU Xiao¹, SHEN Chun-gen¹, XUE Wen-juan²

(1. College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. School of Mathematics and Physics, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: This paper proposes a sparse low-rank matrix optimization model to recover a matrix with sparse low-rank structure. Different from simply adding the nuclear norm and norm to the objective function to promote the low-rankness and sparsity of the target matrix, this model not only guarantees the sparsity of the target matrix by adding the sparsity constraint, but also takes advantage of the possible prior information in the target matrix via some linear constraints. To solve the non-convex and non-smooth model, the original model is first transformed into a constrained DC (difference between two convex functions) optimization model using the exact penalty function, and then it is linearized into a convex subproblem which is solved by the alternating direction multiplier method. Under some assumptions, the algorithm has global convergence. Numerical experiments show that the model has a good effect on speech denoising.

Keywords: low-rankness; sparsity; difference between two convex functions; alternating direction multiplier method

MR Subject Classification: 90C25; 90C30