文章编号:1000-2367(2016)05-0008-12 DOI:10.16366/j.cnki.1000-2367.2016.05.002

基于 L_1 范数最小化的逆协方差矩阵估计

宋运忠,杨丽英

(河南理工大学 电气工程与自动化学院复杂网络研究室,河南 焦作 454000)

摘 要:由于在高维空间中,基于固定维数的经典方法和结果不再适用,样本协方差矩阵不可逆,估计逆协方差矩阵时存在不稳定、计算成本高和非精确等问题,提出了一种 L_1 范数最小化方法来有效估计高维逆协方差矩阵即精确矩阵.当总体分布满足指数类型条件或者多项式类型条件时,所提估计方法在各种范数下的收敛速率优于其他现存的方法.经分析验证,所提方法为凸优化问题,可采用交替方向乘子算法来解决.之后通过 R 语言在模拟数据和实际数据下进行仿真分析,并与 Glasso 方法对比逆协方差的估计性能和图恢复性能,结果表明所提估计方法准确率高、计算成本低.最后,将所提估计方法用来分析白血病数据集,并运用聚类分析对白血病人进行分类.

关键词:协方差矩阵;高斯图模型;精确矩阵;收敛速率;白血病数据集

中图分类号:TP18 **文献标志码**:A

在数据统计分析中,协方差及其逆协方差估计是一个很重要的问题,例如:主成分分析,线性(二次)判别分析,图模型选择等[1-3].当数据维数比样本大小大很多即高维时,稳定准确的方差估计尤其重要,但在高维背景下,基于固定维数和大样本的经典方法已经不再适用.文中主要研究逆协方差(精确矩阵)估计问题,而在高维背景下,样本协方差通常具有奇异性,不能用来估计逆协方差.精确矩阵估计涉及的领域比较广泛,为了能够一致地估计协方差矩阵,一般要利用特殊的结构如带状结构.文献[3]证明了带状样本协方差矩阵能够获得一致估计,文献[4]提出了样本协方差阈值估计,并且获得了阈值估计的收敛率,文献[5]建立了极小极大收敛速率并引进了一种最优率逐渐缩减的估计方法,文献[3]为估计稀疏精确矩阵引进了惩罚似然估计方法.

本文主要研究精确矩阵及高斯图模型估计,文中引进一种新的估计方法——L_1 范数最小化精确矩阵估计方法(L_1 norm minimization for inverse matrix estimation)简称为 CLIME 估计方法[4],并采取交替方向乘子法((alternating direction method of multiplier 即 ADMM)算法解 CLIME.估计方法不仅仅限制在一种具体的稀疏模型,因此泛化能力较强.计算与仿真结果表明所提方法具有良好的数值性质,在谱范数,无穷范数,F 范数下的收敛速率比现存方法快很多,不仅能够大大减少高维下的计算成本,而且在估计准确性和稳定性方面更有竞争力.

1 L_1 范数最小化估计方法基本概念

设 $X=(X_1,\cdots,X_p)^{\mathrm{T}}$ 为 p 维随机向量,协方差矩阵为 Σ_0,精确矩阵为 $\Omega_0=\Sigma_0^{-1}$,给出独立同分布随机样本 $\{X_1,\cdots,X_n\}$,且分布同 X,Σ_0 的经验估计即样本协方差为:$\Sigma_n=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})(X_k-\overline{X})^{\mathrm{T}}$,其中 $\overline{X}=n^{-1}\sum_{k=1}^{n}X^k$,如果 $p>n$,样本协方差 Σ_n 是奇异的,估计协方差 Σ_0 时不稳定,因而无法用 Σ_0 估计精确矩阵 Ω_0.

收稿日期:2016-01-07;**修回日期**:2016-06-27.

基金项目:国家自然科学基金(61340041;61374079);教育部归国留学人员科研启动项目资助.

第 1 作者简介(通信作者):宋运忠(1968-),男,河南民权人,河南理工大学教授,博士,博士生导师,主要从事复杂网络与多智能体系统等方面的教学与科研工作,E-mail:songhpu@126.com.

给定向量 $a=(a_1,\cdots,a_p)^{\mathrm{T}}\in\mathbf{R}^p$，定义 $|a|_1=\sum_{j=1}^{p}|a_j|$，$|a|_2=\sqrt{\sum_{j=1}^{p}a_j^2}$，给定矩阵 $A=(a_{ij})\in\mathbf{R}^{p\times q}$，定义逐元 L_∞ 无穷范数 $|A|_\infty=\max_{1\le i\le p,1\le j\le q}|a_{ij}|$，谱范数 $\|A\|_2=\sup_{|x|_2\le 1}|Ax|_2$，$L_1$ 范数为 $\|A\|_{L_1}=\max_{1\le j\le q}\sum_{i=1}^{p}|a_{ij}|$，$F$ 范数 $\|A\|_F=\sqrt{\sum_{i,j}a_{ij}^2}$.

下面定义 CLIME 估计方法[5-9]，设$\{\hat{\Omega}_1\}$ 为如下最优化问题的解集：

$$\min\|\Omega\|_1,$$
$$\text{s.t. }|\Sigma_n\Omega-I|_\infty\le\lambda_n,\Omega\in\mathbf{R}^{p\times p},\tag{1}$$

其中 λ_n 为可调参数，在(1)式中，如果 Ω 为不对称的，通常情况下解集也为不对称的，因此要进行对称化，设 $\hat{\Omega}_1=(\hat{\omega}_{ij}^1=(\hat{\omega}_1^1,\cdots,\hat{\omega}_p^1)$，$\Omega_0$ 的 CLIME 的估计值 $\hat{\Omega}$ 对称化为：

$$\hat{\Omega}=(\hat{\omega}_{ij}),$$
$$\hat{\omega}_{ij}=\hat{\omega}_{ji}=\hat{\omega}_{ij}^1I\{|\hat{\omega}_{ij}^1|\le|\hat{\omega}_{ji}^1|\}+\hat{\omega}_{ji}^1I\{|\hat{\omega}_{ij}^1|\ge|\hat{\omega}_{ji}^1|\}.\tag{2}$$

凸优化问题(1)可以分解为 p 最小化问题. 令 e_i 为 $\mathbf{R}^p$ 空间的标准单位向量，其第 i 个元素为 1，其余元素为 0. 令 $\hat{\beta}_i,1\le i\le p$ 为以下凸优化问题的解：

$$\min|\beta|_1,$$
$$\text{s.t. }|\Sigma_n\beta-e_i|_\infty\le\lambda_n,\tag{3}$$

其中 $\beta\in\mathbf{R}^p$.

引理 1　令$\{\hat{\Omega}_1\}$ 为(1)式的解集，$\{\hat{B}\}:=\{(\hat{\beta}_1,\cdots,\hat{\beta}_p)\}$，其中 $\hat{\beta}_i,1\le i\le p$ 是(3)式的解，那么$\{\hat{\Omega}_1\}=\{\hat{B}\}$.

证明

$\Omega=(\omega_1,\cdots,\omega_p),\omega_i\in R^p$，约束条件 $|\Sigma_n\Omega-I|_\infty\le\lambda_n$ 等于 $|\Sigma_n\omega_i-e_i|_\infty\le\lambda_n$，$1\le i\le p$，又因 $\hat{\beta}_i$ 是最优化问题(3)的解，故有

$$|\hat{\omega}_i|_1\ge|\hat{\beta}_i|_1,1\le i\le p,\tag{4}$$

因为 $|\Sigma_n\hat{B}-I|_\infty\le\lambda_n$，根据$\{\hat{\Omega}_1\}$ 是最小化问题(1)的解集，得

$$\|\hat{\Omega}_1\|_1\le\|\hat{B}\|_1,\tag{5}$$

由(4)和(5)式得出 $\hat{B}\in\{\hat{\Omega}_1\}$，反过来如果 $\hat{\Omega}_1\notin\{\hat{B}\}$，那么存在 i 使得 $|\hat{\omega}_i|_1>|\hat{\beta}_i|_1$，因此由(4)式有 $\|\hat{\Omega}_1\|_1>\|\hat{B}\|_1$，与(5)式矛盾，因此两者相等，证毕.

2　CLIME 估计的收敛速率分析

设 $\Sigma_n=(\hat{\sigma}_{ij})=(\hat{\sigma}_1,\cdots,\hat{\sigma}_p),\Sigma_0=(\sigma_{ij}^0),EX=(u_1,\cdots,u_p)^{\mathrm{T}}$，根据 X 的矩条件，通常分为两种情况.

(C1) 指数类型的尾条件：假设存在 $0<\eta<1/4$ 使 $\frac{\lg p}{n}\le\eta$，那么：$Ee^{t(X_i-u_i)^2}\le K<\infty$，$|t|\le\eta$，$1\le i\le p$，其中 K 为有界常数.

(C2) 多项式类型的尾条件：假设对于 $\gamma,c_1>0,p\le c_1n^\gamma,\delta>0$，那么：$E|X_i-u_i|^{4\gamma+4+\gamma}\le K,1\le i\le p$. 其中 K 为有界常数.

2.1　谱范数收敛速率分析

首先定义参数空间[3,10-12]：

$$U=U(q,s_0(p))=\left\{\Omega:\Omega\ge0,\|\Omega\|_{L_1}\le M,\max_{1\le i\le p}\sum_{j=1}^{p}|\omega_{ij}|^q\le s_0(p)\right\},\tag{6}$$

其中 $0\le q<1,\Omega=(\omega_{ij})=(\omega_1,\cdots,\omega_p)$. 当 $q=0,U(0,s_0(p))$ 时，满足参数空间的矩阵为 $s_0(p)$ 稀疏矩阵.

分析 $\sup_{\Omega_0\in U}E\|\hat{\Omega}-\Omega_0\|_2^2$ 的收敛速率也是一个很重要的问题，而证明 $\|\hat{\Omega}-\Omega_0\|_2^2$ 期望的存在性比较困难，可以通过修正 $\hat{\Omega}$ 来保证期望的存在性并能够得到收敛速率. 设$\{\hat{\Omega}_{1\rho}\}$ 为如下最优化问题的解集：

$$\min\|\Omega\|_1,\text{s.t. }|\Sigma_{n,\rho}-I|_\infty\le\lambda_n,\Omega\in\mathbf{R}^{p\times p},\tag{7}$$

其中 $\Sigma_{n,\rho}=\Sigma_n+\rho I,\rho>0.\hat{\Omega}_{1,\rho}=(\hat{\omega}_{ij}^1)$,对称化方法同(2),估计值为

$$\hat{\Omega}_\rho=(\omega_{ij\rho}),\hat{\omega}_{ij\rho}=\hat{\omega}_{ji\rho}=\hat{\omega}_{ij\rho}^1 I\{|\hat{\omega}_{ij\rho}^1|\leq|\hat{\omega}_{ji\rho}^1|\}+\hat{\omega}_{ji\rho}^1 I\{|\hat{\omega}_{ij\rho}^1|\geq|\hat{\omega}_{ji\rho}^1|\}. \tag{8}$$

引理 2 假设 $\Omega_0\in U(q,s_0(p))$ 和 $\rho\geq 0$ 如果 $\lambda_n\geq\|\Omega_0\|_{L_1}(\max_{ij}|\hat{\sigma}_{ij}-\sigma_{ij}^0|+\rho)$,那么有:

$$|\hat{\Omega}_\rho-\Omega_0|_\infty\leq 4\|\Omega\|_{L_1}\lambda_n, \tag{9}$$

$$\|\hat{\Omega}_\rho-\Omega_0\|_2\leq C_4 s_0(p)\lambda_n^{1-q}, \tag{10}$$

$$\frac{1}{p}\|\hat{\Omega}_\rho-\Omega_0\|_F^2\leq C_5 s_0(p)\lambda_n^{2-q}, \tag{11}$$

其中 $C_4\leq 2(1+2^{1-q}+3^{1-q})(4\|\Omega_0\|_{L_1})^{1-q},C_5\leq 4\|\Omega_0\|_{L_1}C_4$.

证明

对于优化问题(3),用 $\Sigma_{n,\rho}$ 代替 Σ_n,然后令 $\hat{\beta}_{i,\rho}$ 为问题(3) 的解,注意对于 $\hat{\Omega}_{n,\rho}$ 与 $\{\hat{\beta}_{i,\rho}\}$ 引理 1 仍然满足,$\rho\geq 0$. 为了使问题简洁,仅仅证明 $\rho=0$ 时的情况,$\rho>0$ 与 $\rho=0$ 证明一样.

根据引理 2 的条件,$\Sigma_0=(\sigma_{ij}^0),\Sigma_n=(\hat{\sigma}_{ij}),\rho=0$ 时,

$\lambda_n\geq\|\Omega_0\|_{L_1}(\max_{ij}|\hat{\sigma}_{ij}-\sigma_{ij}^0|)=\|\Omega_0\|_{L_1}(\max_{ij}|\sigma_{ij}^0-\hat{\sigma}_{ij}|)=\|\Omega_0\|_{L_1}|\Sigma_0-\Sigma_n|_\infty$,即

$$|\Sigma_0-\Sigma_n|_\infty\leq\lambda_n/\|\Omega_0\|_{L_1}. \tag{12}$$

由 $\Omega_0=\Sigma_0^{-1}$,则 $|I-\Sigma_n\Omega_0|_\infty=|\Sigma_0\Omega_0-\Sigma_n\Omega_0|_\infty=|(\Sigma_0-\Sigma_n)\Omega_0|_\infty$. 由不等式 $|AB|_\infty\leq|A|_\infty\|B\|_{L_1}$ 得

$|(\Sigma_0-\Sigma_n)\Omega_0|_\infty\leq\|\Omega_0\|_{L_1}|\Sigma_0-\Sigma_n|_\infty\leq\|\Omega_0\|_{L_1}\lambda_n/\|\Omega_0\|_{L_1}\leq\lambda_n$,即

$$|I-\Sigma_n\Omega_0|_\infty\leq\lambda_n. \tag{13}$$

根据 $\hat{\beta}_i$ 的定义,可以得出 $|\hat{\beta}_1|_\infty\leq\|\Omega_0\|_{L_1},1\leq i\leq p$,由引理 1$\{\hat{\Omega}_1\}=\{(\hat{\beta}_1,\cdots,\hat{\beta}_p)\}$ 有

$$\|\hat{\Omega}_1\|_{L_1}\leq\|\Omega_0\|_{L_1}, \tag{14}$$

$|\Sigma_n(\hat{\Omega}-\Omega_0)|_\infty=|\Sigma_n\hat{\Omega}-\Sigma_n\Omega_0|_\infty=|\Sigma_n\hat{\Omega}_1-I+I-\Sigma_n\Omega_0|_\infty\leq$

$|\Sigma_n\hat{\Omega}_1-I|_\infty+|I-\Sigma_n\Omega_0|_\infty\leq 2\lambda_n$,即

$$|\Sigma_n(\hat{\Omega}_1-\Omega_0)|_\infty\leq 2\lambda_n. \tag{15}$$

由(12)、(13)、(14)、(15) 式得

$|\Sigma_0(\hat{\Omega}_1-\Omega_0)|_\infty=|[\Sigma_n-(\Sigma_n-\Sigma_0)](\hat{\Omega}_1-\Omega_0)|_\infty=|\Sigma_n(\hat{\Omega}_1-\Omega_0)-$

$(\Sigma_n-\Sigma_0)(\hat{\Omega}_1-\Omega_0)|_\infty\leq|\Sigma_n(\hat{\Omega}_1-\Omega_0)|_\infty+|(\Sigma_n-\Sigma_0)(\hat{\Omega}_1-\Omega_0)|_\infty\leq$

$2\lambda_n+\|\hat{\Omega}_1-\Omega_0\|_{L_1}|\Sigma_n-\Sigma_0|_\infty\leq 2\lambda_n+(\|\hat{\Omega}_1\|_{L_1}+\|\Omega_0\|_{L_1})|\Sigma_n-$

$\Sigma_0|_\infty\leq 2\lambda_n+2\|\Omega_0\|_{L_1}|\Sigma_n-\Sigma_0|_\infty\leq 2\lambda_n+2\lambda_n=4\lambda_n$,

即 $|\Sigma_0(\hat{\Omega}_1-\Omega_0)|_\infty\leq 4\lambda_n$,因此

$$|\hat{\Omega}_1-\Omega_0|_\infty\leq|\Omega_0\Sigma_0(\hat{\Omega}_1-\Omega_0)|_\infty\leq\|\Omega_0\|_{L_1}|\Sigma_0(\hat{\Omega}_1-\Omega_0)|_\infty\leq 4\|\Omega_0\|_{L_1}\lambda_n.$$

即引理 2 中的(9) 式成立.

下面证明引理 2 中的(10) 式.

证明 令 $t_n=|\hat{\Omega}-\Omega_0|_\infty$,由已知 $\hat{\Omega}=(\hat{\omega}_j),\Omega_0=(\omega_j^0)$,定义 $h_j=\hat{\omega}_j-\omega_j^0,h_j^1=(\hat{\omega}_{ij}I\{|\hat{\omega}_{ij}|\geq 2t_n\};1\leq i\leq p)^{\mathrm{T}}-\omega_j^0,h_j^2=h_j-h_j^1$.

根据(2) 中 $\hat{\Omega}$ 的定义,有 $|\hat{\omega}_j|\leq|\hat{\omega}_j^1|_1\leq|\omega_j^0|_1,\hat{\omega}_j=h_j+\omega_j^0=h_j^1+h_j^2+\omega_j^0$,那么

$$|\omega_j^0|_1-|h_j^1|_1+|h_j^2|_1\leq|\omega_j^0+h_j^1|_1+|h_j^2|_1=|\hat{\omega}_j|_1\leq|\omega_j^0|_1,$$

即 $|\omega_j^0|_1+|h_j^2|_1-|h_j^1|_1\leq|\omega_j^0|_1$,可以得出 $|h_j^2|_1-|h_j^1|_1\leq 0$,即 $|h_j^2|_1\leq|h_j^1|_1$,由 $h_j=h_j^1+h_j^2$,得 $|h_j|_1\leq|h_j^1|_1+|h_j^2|_1\leq 2|h_j^1|_1$,因此只需求 $|h_j^1|_1$ 的上界,下面要利用如下不等式:对于 $\forall a,b,c\in\mathbf{R}$,有

$$|I\{a<c\}-I\{b<c\}|\leq I\{|b-c|<|a-b|\},$$

$$|h_j^1|_1=\sum_{i=1}^p|\hat{\omega}_{ij}I\{|\hat{\omega}_{ij}|\geq 2t_n\}-\omega_{ij}^0|=\sum_{i=1}^p|\hat{\omega}_{ij}I\{|\hat{\omega}_{ij}|\geq 2t_n\}-\omega_{ij}^0[I\{|\omega_{ij}^0|\geq 2t_n\}+$$

$$I\{|\omega_{ij}^0|\leq 2t_n\}]|=\sum_{i=1}^p|\hat{\omega}_{ij}I\{|\hat{\omega}_{ij}|\geq 2t_n\}-\omega_{ij}^0 I\{|\omega_{ij}^0|\geq 2t_n\}-\omega_{ij}^0 I\{|\omega_{ij}^0|\leq 2t_n\}|\leq$$

$$\sum_{i=1}^{p} | \omega_{ij}^0 I\{| \omega_{ij}^0 | \le 2t_n\} | + \sum_{i=1}^{p} | \hat{\omega}_{ij} I\{| \hat{\omega}_{ij} | \ge 2t_n\} - \omega_{ij}^0 I\{| \omega_{ij}^0 | \ge 2t_n\} | \le$$

$$(2t_n)^{1-q} s_0(p) + t_n \sum_{i=1}^{p} I\{| \hat{\omega}_{ij} | \ge 2t_n\} + \sum_{i=1}^{p} | \omega_{ij}^0 | | I\{| \hat{\omega}_{ij} | \ge 2t_n - I\{| \omega_{ij}^0 | \ge$$

$$2t_n\} | \le (2t_n)^{1-q} s_0(p) + t_n \sum_{i=1}^{p} I\{| \hat{\omega}_{ij} | \ge t_n\} + \sum_{i=1}^{p} | \omega_{ij}^0 | I\{| | \omega_{ij}^0 | -$$

$$2t_n | \le | \hat{\omega}_{ij} - \omega_{ij}^0 | \} \le (2t_n)^{1-q} s_0(p) + (t_n)^{1-q} s_0(p) + (3t_n)^{1-q} s_0(p) \le$$

$$(1 + 2^{1-q} + 3^{1-q}) t_n^{1-q} s_0(p).$$

即

$$| h_j^1 |_1 \le (1 + 2^{1-q} + 3^{1-q}) t_n^{1-q} s_0(p). \tag{16}$$

引理 2 中(10) 式结论为 $\| \hat{\Omega}_p - \Omega_0 \|_2 \le C_4 s_0(p) \lambda_n^{1-q}$，且 $C_4 \le 2(1 + 2^{1-q} + 3^{1-q})(4 \| \Omega_0 \|_{L_1})^{1-q}$.

$$| h_j |_1 \le 2 | h_j^1 |_1 \le 2(1 + 2^{1-q} + 3^{1-q}) t_n^{1-q} s_0(p) = 2(1 + 2^{1-q} + 3^{1-q}) | \hat{\Omega} - \Omega_0 |_\infty^{1-q} s_0(p) \le$$

$$2(1 + 2^{1-q} + 3^{1-q})(4 \| \Omega_0 \|_{L_1} \lambda_n)^{1-q} s_0(p) = 2(1 + 2^{1-q} + 3^{1-q})(4 \| \Omega_0 \|_{L_1}) \lambda_n^{1-q} s_0(p),$$

即引理 2 中(10) 式成立.

由(9)、(16) 式以及不等式 $\| A \|_F^2 \le p \| A \|_{L_1} | A |_\infty$，有 $\| \hat{\Omega} - \Omega_0 \|_F^2 \le p \| \hat{\Omega} - \Omega_0 \|_{L_1} | \hat{\Omega} - \Omega_0 |_\infty$，即

$$\frac{1}{p} \| \hat{\Omega} - \Omega_0 \|_F^2 \le \| \hat{\Omega} - \Omega_0 \|_{L_1} | \hat{\Omega} - \Omega_0 |_\infty \le \| \hat{\Omega} - \Omega_0 \|_{L_1} 4 \| \Omega_0 \|_{L_1} \lambda_n \le$$

$$C_4 s_0(p) \lambda_n^{1-q} 4 \| \Omega_0 \|_{L_1} \lambda_n = C_4 s_0(p) 4 \| \Omega_0 \|_{L_1} \lambda_n^{2-q} \le C_5 s_0(p) \lambda_n^{2-q},$$

引理 2 中的(11) 式证毕.

定理 1　假设 $\Omega_0 \in U(q, s_0(p))$，

(a) 假设条件(C1) 成立，设 $\lambda_n = C_0 M \sqrt{\frac{\lg p}{n}}$，$C_0 = 2\eta^{-2}(2 + \tau + \eta^{-1} e^2 K^2)^2$，$\tau > 0$ 则有：

$$\| \hat{\Omega} - \Omega \|_2 \le C_1 M^{2-2q} s_0(p) \left(\frac{\lg p}{n} \right)^{(1-q)/2}, \tag{17}$$

其中(17) 成立的概率大于 $1 - 4p^{-\tau}$，$C_1 \le 2(1 + 2^{1-q} + 3^{1-q}) 4^{1-q} C_0^{1-q}$.

证明　由于定理 1(a) 的证明依据是引理 2，所以根据引理 2 知，只需证明

$$\max_{ij} | \hat{\sigma}_{ij} - \sigma_{ij}^0 | \le C_0 \sqrt{\frac{\lg p}{n}}. \tag{18}$$

在(C1) 条件下(18) 式成立的概率为 $1 - 4p^{-\tau}$，不失一般性，假设 $EX = 0$，$\Sigma_n^0 := n^{-1} \sum_{k=1}^{n} X_k X_k^{\mathrm{T}}$，$Y_{kij} = X_{ki} X_{kj} - E X_{ki} X_{kj}$，那么 $\Sigma_n = \Sigma_n^0 - \overline{X}\overline{X}^{\mathrm{T}}$，令 $t = \eta \sqrt{\frac{\lg p}{n}}$，由不等式 $| e^s - 1 - s | \le s^2 e^{\max(s,0)}$，$s \in \mathbf{R}$，令 $C_{K1} = 2 + \tau + \eta^{-1} K^2$，经过简单的计算得

$$P\left(\sum_{k=1}^{n} Y_{ij} \ge \eta^{-1} C_{K1} \sqrt{n \lg p} \right) \le e^{-C_{K1} \lg p} (E \exp(t Y_{kij}))^n \le \exp(- C_{K1} \lg p +$$

$$n t^2 E Y_{kij}^2 e^{t | Y_{kij} |}) \le \exp(- C_{K1} \lg p + \eta^{-1} K^2 \lg p) \le \exp(- (\tau + 2) \lg p),$$

因此有

$$P(| \Sigma_n^0 - \Sigma_0 |) \ge \eta^{-1} C_{K1} \sqrt{\frac{\lg p}{n}} \le 2p^{-\tau}. \tag{19}$$

由不等式 $e^s \le e^{s^2+1}$，$s > 0$，有 $E e^{t | X_j |} \le eK$，$t \le \eta^{1/2}$. 令 $C_{K2} = 2 + \tau + \eta^{-1} e^2 K^2$，$a_n = C_{K2}^2 (\frac{\lg p}{n})^{1/2}$ 有：

$$P(| \overline{X}\overline{X}^{\mathrm{T}} |_\infty \ge \eta^{-2} a_n \sqrt{\frac{\lg p}{n}}) \le p \max_i P\left(\sum_{k=1}^{n} X_{ki} \ge \eta^{-1} C_{K2} \sqrt{\frac{\lg p}{n}} \right) +$$

$$p \max_i P\left(- \sum_{k=1}^{n} X_{ki} \ge \eta^{-1} C_{K2} \sqrt{\frac{\lg p}{n}} \right) \le 2p^{-\tau-1}, \tag{20}$$

由(19)、(20)式与不等式 $C_0 > \eta^{-1}C_{K1} + \eta^{-1}a_n$ 得(18)式成立,即定理 1(a) 证毕.

(b) 假设条件(C2)成立:设 $\lambda_n = C_2 M\sqrt{\frac{\lg p}{n}}, C_2 = \sqrt{(5+\tau)(\theta+1)}$,则有:

$$\| \hat{\Omega} - \Omega \|_2 \leq C_3 M^{2-2q} s_0(p)\left(\frac{\lg p}{n}\right)^{\frac{1-q}{2}}, \tag{21}$$

其中(21)式成立概率大于 $1 - O(n^{-\delta/8} + p^{-\tau/2})$, $C_3 \leq 2(1 + 2^{1-q} + 3^{1-q})4^{1-q}C_2^{1-q}$.

证明

不失一般性,假设 $EX = 0, \Sigma_n^0 := n^{-1}\sum_{k=1}^{n} X_k X_k^{\mathrm{T}}, Y_{kij} = X_{ki}X_{kj} - EX_{ki}X_{kj}$,则 $\Sigma_n = \Sigma_n^0 - \bar{X}\bar{X}^{\mathrm{T}}$. 令 $\overset{\vee}{Y}_{kij} = X_{ki}X_{kj}I\{|X_{ki}X_{kj}| \leq \sqrt{n/(\lg p)^3}\} - EX_{ki}X_{kj}I\{|X_{ki}X_{kj}| \leq \sqrt{n/(\lg p)^3}\}$, $\overset{\vee}{Y}_{kij} = Y_{kij} - \bar{Y}_{kij}$,由 $b_n := \max_{i,j} E|X_{ki}X_{kj}|I\{|X_{ki}X_{kj}| \geq \sqrt{n/(\lg p)^3}\} = O(1)n^{-r-1/2}$,又根据条件(C2)得:

$$P\left(\max_{i,j}\left|\sum_{k=1}^{n}\overset{\vee}{Y}_{kij}\right| \geq 2nb_n\right) \leq P\left(\max_{i,j}\left|\sum_{k=1}^{n} X_{ki}X_{kj}I\{|X_{ki}X_{kj}| > \sqrt{n/(\lg p)^3}\}\right| \geq nb_n\right) \leq$$

$$P\left(\max_{i,j}\sum_{k=1}^{n}|X_{ki}X_{kj}|I\{X_{ki}^2 + X_{kj}^2 \geq 2\sqrt{n/(\lg p)^3}\} \geq nb_n\right) \leq P(\max_{k,i} X_{ki}^2 \geq$$

$$\sqrt{n/(\lg p)^3}) \leq pnP(X_1^2 \geq \sqrt{n/(\lg p)^3} = O(1)n^{-\delta/8}.$$

由伯恩斯坦不等式 $\max_{|z|\leq 1}(|P'(z)|) \leq n.\max_{|z|\leq 1}(|P(z)|)$,得

$$P\left(\max_{i,j}\left|\sum_{k=1}^{n}\bar{Y}_{kij}\right| \geq \sqrt{(\theta+1)(4+\tau)n\lg p}\right) \leq p^2 \max_{i,j} P\left(\left|\sum_{k=1}^{n}\bar{Y}_{kij}\right| \geq \sqrt{(\theta+1)(4+\tau)n\lg p}\right) \leq$$

$$2p^2 \max_{i,j}\exp(-(\theta+1)(4+\tau)n\lg p/(2nE\bar{Y}_{1ij}^2 + \sqrt{(\theta+1)(64+16\tau)}n/(3\lg p))) = O(1)p^{-\tau/2},$$

因此得

$$P(|\Sigma_n^0 - \Sigma_0|_\infty \geq \sqrt{(\theta+1)(4+\tau)}\frac{\lg p}{n} + 2b_n) = O(n^{-\delta/8} + p^{-\tau/2}). \tag{22}$$

由相同的截断参数与伯恩斯坦不等式得:$P\left(\max_i\left|\sum_{k=1}^{n}X_{ki}\right| \geq \sqrt{\max_i \sigma_{ii}^0(4+\tau)n\lg p}\right) = O(n^{-\delta/8} + p^{-\tau/2})$,因此有

$$P(|\bar{X}\bar{X}^{\mathrm{T}}|_\infty \geq \max_i \sigma_{ii}^0(4+\tau)\frac{\lg p}{n}) = O(n^{-\delta/8} + p^{-\tau/2}), \tag{23}$$

由(22)和(23)式得

$$\max_{ij}|\hat{\sigma}_{ij} - \sigma_{ij}^0| \leq \sqrt{(\theta+1)(5+\tau)}\frac{\lg p}{n} \tag{24}$$

成立的概率大于 $1 - O(n^{-\delta/8} + p^{-\tau/2})$,证毕.

定理 2 假设 $\Omega_0 \in U(q, s_0(p))$ 且条件(C1)成立. 设 $\lambda_n = C_0 M\sqrt{\frac{\lg p}{n}}$, C_0 同定理 1(a), τ 充分大, $\rho = \sqrt{\frac{\lg p}{n}}$. 如果 $\xi > 0, p \geq n^{\xi}$ 成立,那么有以下结论:

$$\sup_{\Omega_0 \in U} E\|\hat{\Omega}_\rho - \Omega_0\|_2^2 = O(M^{4-4q}s_0^2(p)\left(\frac{\lg p}{n}\right)^{1-q}). \tag{25}$$

推论 1 如果没有条件 $\rho = \sqrt{\frac{\lg p}{n}}$. 通过证明表明满足以下条件时,定理 2 仍然成立:

$$\min\left(\sqrt{\frac{\lg p}{n}}, p^{-\alpha}\right) \leq \rho \leq \sqrt{\frac{\lg p}{n}}\alpha > 0 \tag{26}$$

证明

因为 $\Sigma_{n,\rho}^{-1}$ 为一可行点,由(26)式有:$\|\hat{\Omega}_\rho\|_1 \leq \|\hat{\Omega}_{1\rho}\|_1 \leq \|\Sigma_{n,\rho}^{-1}\|_1 \leq p^2\max\left(\sqrt{\frac{n}{\lg p}}, p^{\alpha}\right)$.

根据(18)式以及引理2，令 $p \geq n^{\xi}$，τ 足够大，有：

$$\sup_{\Omega_0 \in U} E \| \hat{\Omega}_\rho - \Omega_0 \|_2^2 = \sup_{\Omega_0 \in U} E \| \hat{\Omega}_\rho - \Omega_0 \|_2^2 \times I\{\max_{ij} | \hat{\sigma}_{ij} - \sigma_{ij}^0 | + \rho \leq C_0 \sqrt{\frac{\lg p}{n}}\} +$$

$$\sup_{\Omega_0 \in U} E \| \hat{\Omega}_\rho - \Omega_0 \|_2^2 \times I\{\max_{ij} | \hat{\sigma}_{ij} - \sigma_{ij}^0 | + \rho > C_0 \sqrt{\frac{\lg p}{n}}\} = O\left(M^{4-4q} s_0^2(p) \left(\frac{\lg p}{n}\right)^{1-q}\right),$$

定理2证毕.

当随机向量 X 中变量有序时，可以得到更优收敛速率，与文献[3]相似，研究满足如下参数空间的精确矩阵：

$$U_0(\alpha, B) = \left\{ \begin{array}{l} \Omega : \Omega \geq 0, \\ k \geq 0, \\ \max\limits_j \sum\limits_i \{ | \omega_{ij} | : | i-j | \geq k \} \leq B(k+1)^{-\alpha} \end{array} \right\}, \tag{27}$$

其中 $\alpha > 0$，$\Omega_0 \in U_0(\alpha, B)$.

定理3　设 $\Omega_0 \in U_0(\alpha, B)$，$\lambda_n = CB\sqrt{\frac{\lg p}{n}}$ 且 C 充分大.

(a) 如果(C1)和(C2)成立，那么概率大于 $1 - O(n^{-\delta/8} + p^{-\tau/2})$ 时(28)式成立：

$$\| \hat{\Omega} - \Omega_0 \| n. \tag{28}$$

(b) 假设 $\xi > 0$，$p \geq n^{\xi}$. 如果(C1)成立且 $\rho = \sqrt{\frac{\lg p}{n}}$ 有：

$$\sup_{\Omega_0 \in U(\alpha, B)} E \| \hat{\Omega}_\rho - \Omega_0 \|_2^2 = O\left(B^4 \left(\frac{\lg p}{n}\right)^{\alpha/(\alpha+1)}\right). \tag{29}$$

证明

令 k_n 是满足 $1 \leq k_n \leq n$ 的整数，定义：

$$h_j = \hat{\omega} - \omega_j^0, h_j^1 = (\hat{\omega}_{ij} I\{1 \leq i \leq k_n\}; 1 \leq i \leq p)^{\mathrm{T}} - \omega_j^0, h_j^2 = h_j - h_j^1.$$

由引理2的证明得到 $| h_j |_1 \leq 2 | h_j^1 |_1$，有 $\Omega_0 \in U(\alpha, M)$，有 $\sum\limits_{j \geq k_n} | \omega_j^0 | \leq M k_n^{-\alpha}$.

由定理1的证明得 $\sum\limits_{j=1}^{k_n} | \hat{\omega}_{ij} - \omega_{ij}^0 | = O(k_n \sqrt{\frac{\lg p}{n}})$，成立的概率大于 $1 - O(n^{-\delta/8} + p^{-\tau/2})$.

令 $k_n = [(n/\lg p)^{1/(2\alpha+2)}]$ 从而证明定理3(a)，定理3(b)的证明同定理2.

2.2　在 l_∞ 范数和 F 范数下的收敛速率

以上分析了谱范数下的估计性能，l_∞ 范数与 F 范数下的收敛速率同样也可以得到.

定理4　(a) 在定理1(a)的条件下，有：

$$| \hat{\Omega} - \Omega_0 |_\infty \leq 4C_0 M^2 \sqrt{\frac{\lg p}{n}},$$

$$\frac{1}{p} \| \hat{\Omega} - \Omega_0 \|_F^2 \leq 4C_1 M^{4-2q} s_0(p) \left(\frac{\lg p}{n}\right)^{1-q/2}. \tag{30}$$

(30)式成立的概率大于 $1 - 4p^{-\tau}$.

证明同定理1(a).

(b) 在定理1(b)的条件下，有

$$| \hat{\Omega} - \Omega_0 |_\infty \leq 4C_2 M^2 \sqrt{\frac{\lg p}{n}},$$

$$\frac{1}{p} \| \hat{\Omega} - \Omega_0 \|_F^2 \leq 4C_3 M^{4-2q} s_0(p) \left(\frac{\lg p}{n}\right)^{1-q/2}, \tag{31}$$

其中(31)式成立的概率大于 $1 - O(n^{-\delta/8} + p^{-\tau/2})$.

在定理4(b)中收敛速率明显优于文献[8]中得到的收敛速率.

证明同定理1(b).

定理 5 在定理 2 的条件下,有:

$$\sup_{\Omega_0 \in U} E \mid \hat{\Omega} - \Omega_0 \mid_\infty^2 \leq O(M^4 \sqrt{\frac{\lg p}{n}},$$

$$\frac{1}{p} \sup_{\Omega_0 \in U} E \parallel \hat{\Omega} - \Omega_0 \parallel_F^2 \leq O(M^{4-2q} s_0(p) \left(\frac{\lg p}{n}\right)^{1-q/2}, \tag{32}$$

证明同定理 2.

3 数值分析

CLIME 估计方法主要研究解决线性规划问题(Linear Program)(1):

$$\min \parallel \Omega \parallel_1 \text{s. t. } \mid \Sigma_n \Omega - I \mid_\infty \leq \lambda_n, \Omega \in R^{p \times p},$$

CLIME 估计方法不仅有很强的统计上的良好性质,而且具有内在的计算上的优势. 首先,(1) 中的线性规划(LP) 没有明确地给出 $\hat{\Omega}$ 的正定性,其在高维背景下是一个很大的挑战;第二,可以看到(1) 可以分解为 p 个独立的 LP 问题,每 LP 对应着 $\hat{\Omega}$ 的每一列. 这个可分结构使得(1) 可以用内点法一列一列的解决 LP 或者交替方向乘子法(ADMM)[13-15]. 但是这些方法不能很好地处理极高维问题:上述两种算法不能用来处理成千上万节点的情况,而且特殊情况下,需要整个样本协方差矩阵来输入单机内存,这对于中等规模的问题都是不切合实际的. 因此提出一种有效地 CLIME-ADMM,提出的 CLIME-ADMM 算法可以按比例增加到成千上万的维数,拓展性强. 首先,以 CLIME 为目标提出了列块 ADMM 算法.

3.1 列块 ADMM 算法更新

提出一种列块 ADMM 算法来解决估计精确矩阵问题,这种算法中用列块代替一列一列. 假设一个列块包括 $k(1 \leq k \leq p)$ 列,那么(1) 问题等于解决$[p/n]$个独立的线性规划问题. $X \in \mathbf{R}^{p \times k}$ 表示 $\hat{\Omega}$ 的 k 列,(1) 式可以写为如下形式:

$$\min \parallel X \parallel_1, \text{s. t. } \mid \Sigma_n X - E \mid_\infty \leq \lambda_n. \tag{33}$$

(2) 式可以写成如下的等式约束形式:

$$\min \parallel X \parallel_1, \text{s. t. } \parallel Z - X \parallel_\infty \leq \lambda, \Sigma_n X = Z. \tag{34}$$

通过分离变量 $Z \in \mathbf{R}^{p \times k}$,无穷范数约束变成为一个箱子约束且从 L_1 范数目标函数中分离. 接下来用 ADMM 算法去解决问题(34). (34) 式的增广拉格朗日式为:

$$L_\rho = \parallel X \parallel_1 + \rho \langle Y, \Sigma_n X - Z \rangle + \frac{\rho}{2} \parallel \Sigma_n X - Z \parallel_2^2, \tag{35}$$

其中 $Y \in \mathbf{R}^{p \times k}$ 是对偶变量,$\rho > 0$. ADMM 算法产生出下面的迭代:

$$X^{t+1} = \text{argmin}_X \parallel X \parallel_1 + \frac{\rho}{2} \parallel \Sigma_n X - Z^t + Y^t \parallel_2^2, \tag{36}$$

$$Z^{t+1} = \underset{\parallel Z - E \parallel_\infty \leq \lambda}{\text{argmin}} \frac{\rho}{2} \parallel \Sigma_n X^{t+1} - Z + Y^t \parallel_2^2, \tag{37}$$

$$Y^{t+1} = Y^t + \Sigma_n X^{t+1} - Z^{t+1}. \tag{38}$$

作为一个 Lasso 问题,(36) 式可以用现存的 Lasso 算法来解决,但是将会产生一个双重循环的问题. (36) 式没有一个封闭解因为在二次规划惩罚中的 Σ_n 使 X 耦合. 这里用过线性化二次惩罚和增加一个临近的方法对 X 进行去耦:

$$X^{t+1} = \text{argmin}_X \parallel X \parallel_1 + \eta \langle V^t, X \rangle + \frac{\eta}{2} \parallel X - X^t \parallel, \tag{39}$$

其中 $V^t = \frac{\rho}{\eta} \Sigma_n (Y^t + \Sigma_n X^t - Z^t)$, $\eta > 0$, (39) 式称为非精确 ADMM 迭代. 用(38) 式中 $V^t = \frac{\rho}{\eta} \Sigma_n (2Y^t - Y^{t-1})$,

令 $V^t = \Sigma_n Y^t$,有 $V^t = \frac{\rho}{\eta}(2\hat{V}^t - \hat{V}^{t-1})$, (39) 式有以下封闭解:

$$X^{t+1} = s_f\left(X^t - V^t, \frac{1}{\eta}\right) \tag{40}$$

其中 s_f 代表软阈值.

令 $U^{t+1}=\Sigma_n X^{t+1}$,(37) 式是一个箱子约束二次规划问题且有如下形式的封闭解:

$$Z^{t+1}=b_x(U^{t+1}+Y^t,E,\lambda), \tag{41}$$

其中 b_x 代表在无穷范数约束 $\|Z-E\|_\infty\leq\lambda$ 上的投影. 特别地,如果 $\|U^{t+1}+Y^t-E\|_\infty\leq\lambda$,$Z^{t+1}=U^{t+1}+Y^t$,那么 $Y^{t+1}=Y^t+U^{t+1}-Z^{t+1}=0$.

3.2 CLIME-ADMM 算法流程

提出的 ADMM 算法程序流程如下.

表 1　CLIME-ADMM 算法

算法 CLIME 列块 ADMM 算法
1. 输入:$\Sigma_n,\lambda,\rho,\eta$
2. 输出:X
3. 初始化:$X^0,Z^0,Y^0,V^0,\hat{V}^0=0$
4. for $t=0$ to $T-1$ do
4.1 X-update:$X^{t+1}=s_f\left(X^t-V^t,\frac{1}{\eta}\right)$, 其中 $s_f(X,\gamma)=\begin{cases}X_{ij}-\gamma,X_{ij}>\gamma,\\X_{ij}+\gamma,X_{ij}<-\gamma,\\0,\text{其他}.\end{cases}$
4.2 矩阵乘法:$\begin{cases}U^{t+1}=\Sigma_n X^{t+1}\\U^{t+1}=A(A'X^{t+1})\end{cases}$,
4.3 Z-update:$Z^{t+1}=b_x(U^{t+1}+Y^t,\lambda)$, $b_x(X,E,\lambda)=\begin{cases}E_{ij}+\lambda,X_{ij}-E_{ij}>\lambda,\\X_{ij},\mid X_{ij}-E_{ij}\mid\leq\lambda,\\E_{ij}-\lambda,\mid X_{ij}-E_{ij}\mid<-\lambda\end{cases}$
4.4 Y-update:$Y^{t+1}=Y^t+U^{t+1}-Z^{t+1}$
4.5 矩阵乘法:$\begin{cases}\hat{V}^{t+1}=\Sigma_n Y^{t+1}\\\hat{V}^{t+1}=A(A'Y^{t+1})\end{cases}$
4.6 V-update:$V^{t+1}=\frac{\rho}{\eta}(2\hat{V}^{t+1}-\hat{V}^t)$
5. end

4　CLIME 估计方法仿真分析

在仿真中使用 R 语言 flare 包与 Glasso 包,并在 R3.2.1 软件环境下进行,在 flare 包中使用 ADMM 算法解 CLIME. 针对 6 种情况分析 CLIME 估计方法的性能:(i)$n=200,d=100$;(ii)$n=200,d=200$;(iii)$n=200,d=400$;(iv)$n=400,d=100$;(v)$n=400,d=200$;(vi)$n=400,d=400$. 用两种模型来产生无向图和精确矩阵[16],两种模型依次为带状模型和毂状模型[17],如图 1 所示.

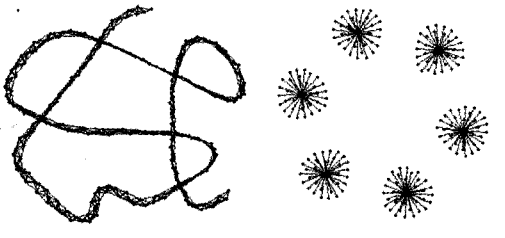


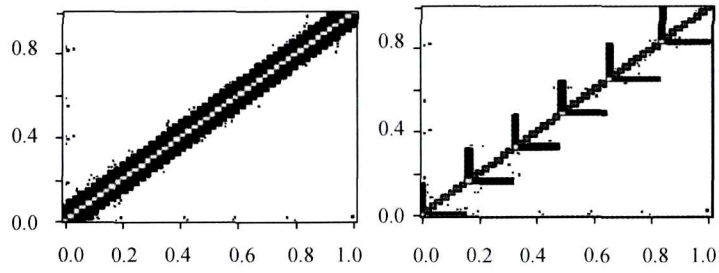
图1 在仿真过程中用到的带状和毂状模型,仅仅给出 d=100时的图例

为了简化分析,只给出 $d=100$ 时的两种图模型. 为了更好的说明 CLIME 估计方法的恢复性能,用 Glasso 方法与 CLIME 方法进行对比. 图 2(a)－(c)给出了带状模型(左)和毂状模型(右)在两种方法下的精确矩阵估计结果热图. 热图的对比说明了 CLIME 比 Glasso 更稀疏,经过观察 CLIME 方法恢复的模型比 Glasso 更接近真实模型. 值得注意的是 Glasso 因包含错误的非零元素而具有更多的噪声,而且当真实模型有大量的非零元素分布在非对角线上时,Glasso 往往包含更多的错误的非零元素.

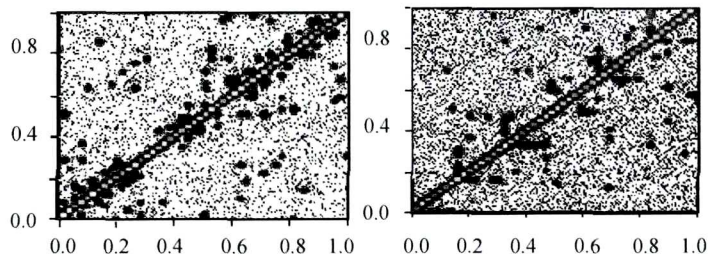
4.1 图恢复性能

高斯图像化模型的选择是一个很重要的问题. 令 $G=(V,E)$ 为代表 X 各元素之间条件独立关系的图.

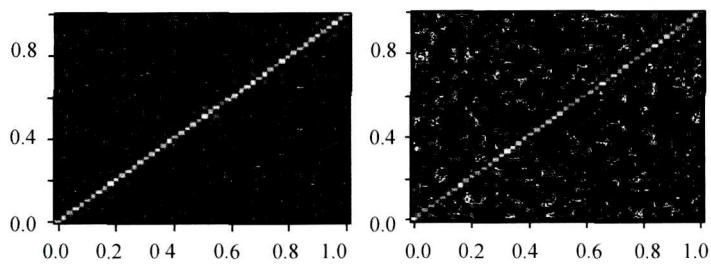
顶点集 V 有 p 个元素 $X_1,\cdots,X_P$，边集 $|E|$ 包含有序对 (i,j)，如果在 X_i 和 X_j 条件不独立，那么 $(i,j)\in E$. 设 $X\in(u_0,\Sigma_0)\Omega_0=(\omega_{ij}^0)$，$X_i$ 和 X_j 之间条件不独立，有 $\omega_{ij}^0\neq 0$. 当随机向量 X 服从高斯分布时，高斯模型下图估计[16] 等于恢复精确矩阵的估计.



带状（左）与毂状（右）真值



带状（左）与毂状（右）GLIME估计值



带状（左）与毂状（右）Glasso估计值

图2 d=100时CLIME, Glasso在带状（左）和毂状（右）种典型模型下的估计值

首先对比 CLIME 与 Glasso 的图恢复性能. 用正误识率和负误识率来评价图恢复性能. 图 3 为带状与毂状模型下的 ROC 曲线. 从图 3 可以看出，在两种模型下 CLIME 图恢复性能明显优于 Glasso，这意味着 CLIME 适用于更广泛的模型，从总体数据来看，CLIME 在高维图估计中更有竞争力. 最后给出 CLIME 在毂状模型 F 范数下的误差分析图，如图 4，两种方法在不同维数下的运行时间在图 5 中给出.

4.2 在白血病数据集上的应用

基因表达数据集为慢性淋巴细胞白血病(Chronic lymphocytic leukemia)数据集[18]，采用了 Affymetrix 公司的 HG_U95Av2 表达谱芯片(含有 12 625 个探针组)，使用 Bioconductor 软件，并在 R3.2.1 环境下平台下进行仿真，共测量了 24 个病人，每个样品来自一个癌症病人，所有病人根据健康状态分为两组：稳定期(stable)组：CLL1，CLL17，CLL18，CLL24，CLL22，CLL9，CLL20，CLL12，CLL2，进展组(progressive)也称为恶化期组. 因此经过预处理之后得到的对象是一个 12 625 行，24 列的矩阵，原始分布图与预处理之后的分布图如图 6.

接着对数据进行马氏距离聚类分析. 首先对预处理之后的矩阵进行转置，然后用所提出的 CLIME 估计方法估计出精确矩阵，马氏距离定义为 $D_{ij}=(X_i-X_j)^{\mathrm{T}}\hat{\Omega}(X_i-X_j)$，基于马氏距离聚类分析方法，给出癌症病人样本的聚类图，如图 7.

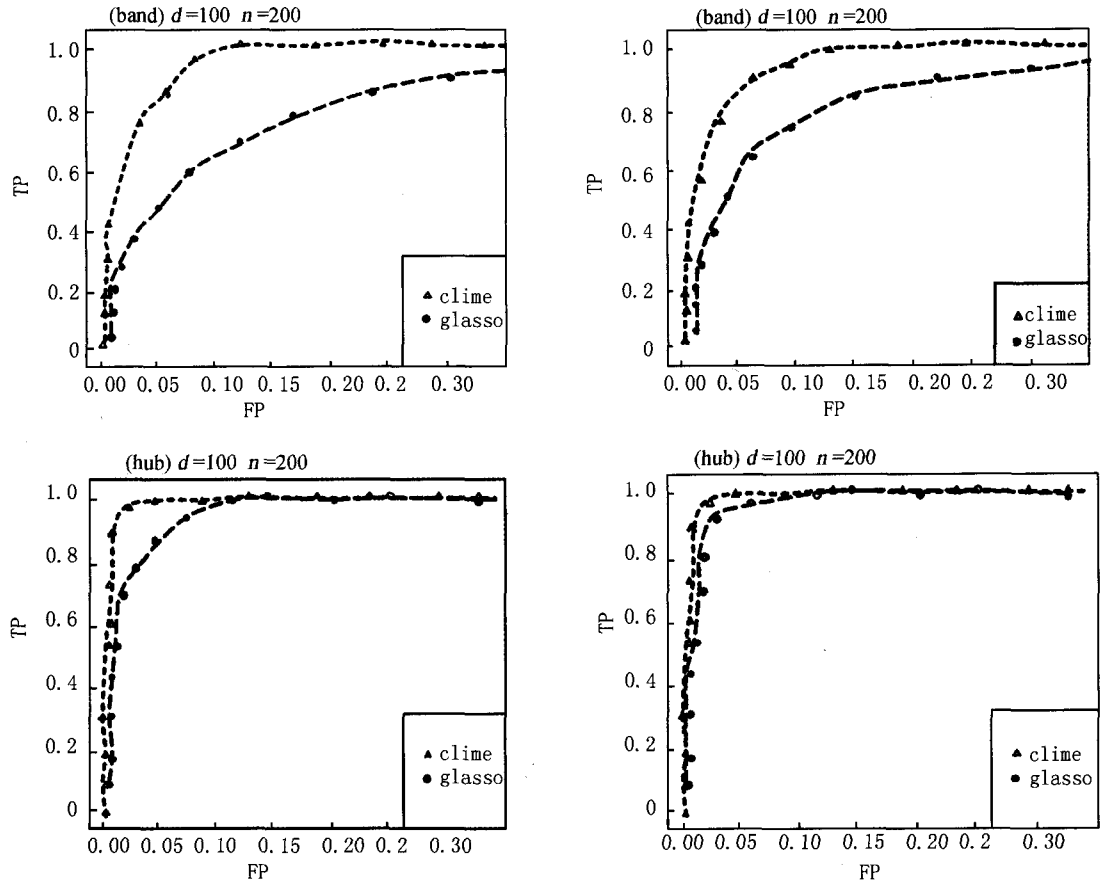


图3 带状模型和毂状模型下的ROC曲线

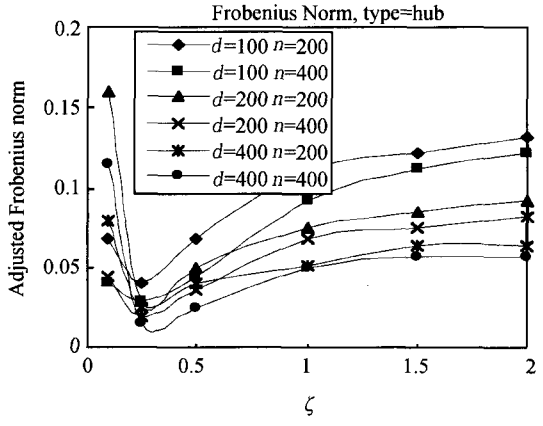


图4 CLIME-F范数误差

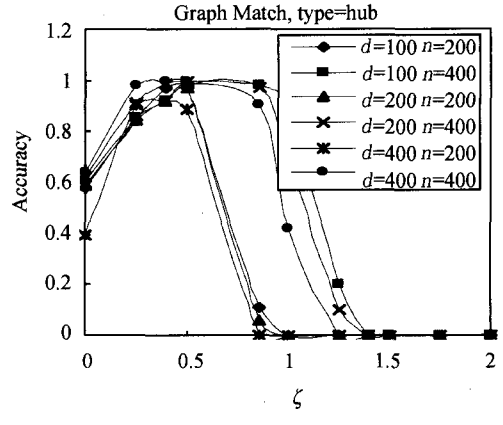


图5 CLIME在毂状图模型下图恢复准确率

从图 6 可以看出，经过 gcRMA 预处理之后，不但所有的曲线很好地重合在一起，而且分布更加接近高斯分布，这样就可以通过估计基因表达数据的精确矩阵来估计高斯图. 从聚类分析的整体结果图 7 来看，分类效果良好，但是稳定组和恶化组并没有完全分开，这并不是说明实验失败，理论上讲，如果总体上两组数据是分开的，那么说明导致癌症从稳定到恶化的因素起主导作用，如果不是，很可能其他因素起主导作用，要具体问题具体分析. 所以从聚类结果图可以得出 CLL 数据的实验样本来自不同的个体，而不是细胞，很可能个体差异起到了主导作用，由此可以根据聚类结果图分析出病因所在.

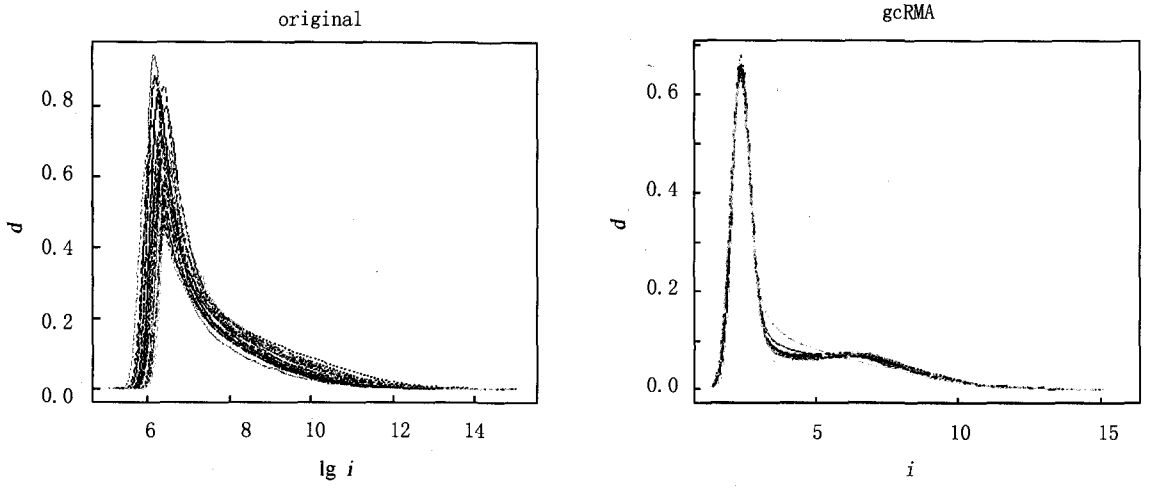


图6 原始(左)与预处理后(右)信号强度直方图

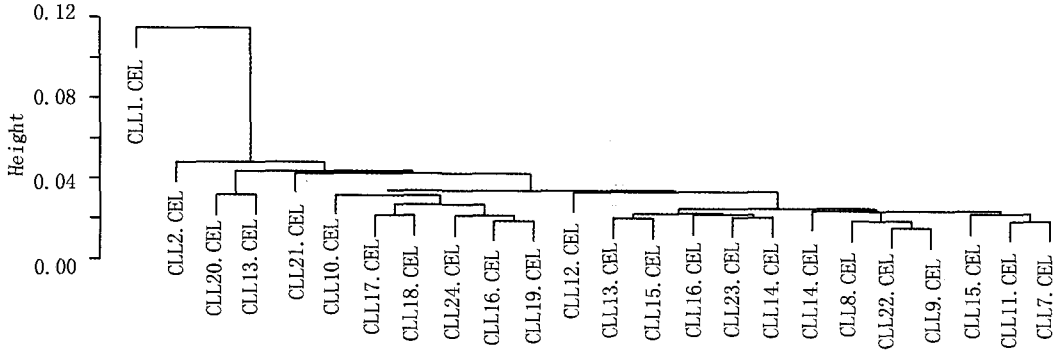


图7 样本聚类分析图

5 结 论

针对高维精确矩阵与图模型估计,提出了基于 L_1 范数最小化的估计方法即 CLIME 方法. 在精确矩阵估计过程中,列块乘子方向交替法是一种快速有效的算法,而且所提方法在不同范数下的收敛速率皆优于现存方法. 计算上,所提方法比现有方法快很多. 通过对模拟数据与真实数据仿真分析表明 CLIME 方法在估计高维精确矩阵时拓展性好,运行速度快,准确率高,在基因大数据上得到广泛而有效的应用.

参 考 文 献

[1] Onureena B. Model Selection Through Sparse Maximum Likelihood Estimation[J]. Machine Learning Research,2008,9:485-516.

[2] Ricardo U L, Florent D, Florence D. Out-of-core adaptive iso-surface extraction from binary volume data[J]. Graphical Models,2014,76:593-608.

[3] Peter J, Bickel E L. Regularized Estimation of Large Covariance Matrices[J]. The Annals of Statistics,2008,36:199-227.

[4] Peter J, Bickel E L. Covariance Regularization by Thresholding[J]. The Annals of Statistics,2008,36: 2577-2604.

[5] Emmanuel C, Terence T. The Dantzig Selector Statistical Estimation When p Is Much Larger Than n[J]. The Annals of Statistics,2007,35:2313-2351.

[6] Fan Jianqing, Feng Yang, Wu Yichao. Network Exploration via the Adaptive Lasso and SCAD Penalties[J]. The Annals of Applied Statistics,2009,2:521-541.

[7] Stephen B, Lieven V. Convex Optimization[M]. Cambridge: Cambridge University Press,2004.

[8] Clifford L, Fan Jianqing. Sparsistency and Rates of Convergence in Large Covariance Matrix Estimation[J]. The Annals of Statistics,2009,37:4254-4278.

[9] Karoui N. Operator Norm Consistent Estimation of Large-Dimensional Sparse Covariance Matrices[J]. The Annals of Statistics, 2008, 36:2717-2756.

[10] Fan Jianqing, Li Runze. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties[J]. American Statistical Association, 2001, 96:1348-1360.

[11] Friedman J, Hastie T, Robert T. Sparse Inverse Covariance Estimation with the Graphical Lasso[J]. Biostatistics, 2008, 9:432-441.

[12] Hess K R, Anderson K, Symmans W F, et al. Pharmacogenomic Predictor of Sensitivity to Preoperative Chemotherapy With Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide in Breast Cancer[J]. Clinical Oncology, 2006, 24:4236-4244.

[13] Ravikumar P, Martin J, Wainwright G R, et al. High-dimensional covariance estimation by minimizing L_1-penalized log-determinant divergence[J]. Electronic Journal of Statistics, 2011, 5:935-980.

[14] Cai T T, Liu W, Harrison H Z. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation[J]. The Annals of Statistics, 2012, 40: 2389-2420.

[15] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundation and Trends Machine Learning, 2011, 3(1):1-122.

[16] Yuan Ming, Lin Yi. Model Selection and Estimation in the Gaussian Graphical Model[J]. Biometrika, 2007, 94:19-35.

[17] Ji Jian, Li Xiao, Xu Shuangxing, et al. SAR image despeckling by sparse reconstruction based on shearlets[J]. Acta Automatica Sinica, 2015, 41(8):1495-1501.

[18] Nicolai M, Peter B. High-Dimensional Graphs and Variable Selection With the Lasso[J]. The Annals of Statistics, 2006, 34:1436-1462.

A approach to Precision Matrix Estimation Based on L_1 Norm Minimization

SONG Yunzhong, YANG Liying

(Complex Networks Lab; School of Electrical Engineering & Automation, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Because it is irreversible for sample covariance matrix in high-dimensional situation, and it is not stable to estimate the inverse covariance matrix. What's more, the classic methods and results based on fixed dimension is no longer applicable. A L_1 norm minimization method was proposed to estimate a high-dimensional inverse covariance matrix, and the related problem namely Gaussian graphical model selection were analyzed. The convergence rates under the various norm were given when the population distribution has either exponential-type tails or polynomial-type tails. The convergence rates are superior to other existing methods. The methods is convex optimization problem, and can be converted into linear programming, then apply alternating direction method of multiplier algorithm to solve it. Numerical performance of the estimator was investigated using both simulated and real data by R language. The precision matrix estimation performance and recovery performance of the various models was compared. The results show that the proposed method has high accuracy, low computational cost and rapid running speed. In addition, the procedure was applied to analyze a Leukemia dataset and used cluster analysis to classify the patients.

Keywords: covariance matrix; Gaussian graphical model; precision matrix; rate of convergence; leukemia dataset