2E3-1

# 複数データセットからのガウシアングラフィカルモデルの同時構造推定

Simultaneous Learning of Graphical Structures

原 聡 Satoshi Hara　鷲尾 隆 Takashi Washio

大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

In this paper, we propose a simultaneous estimation technique of graphical structures from several datasets. We formulate the problem as an extension of existing approaches and provide an efficient estimation procedure. The validity of our approach is presented in a numerical simulation.

## 1. はじめに

変数間の条件付き独立性はデータの生成メカニズムと密接に関連しており，そのため観測データから変数間の依存構造を発見することはデータマイニングにおける重要な問題である．中でもガウシアングラフィカルモデル (GGM) は連続変数間の線形な依存関係を表現する基本的なモデルであり，様々な構造推定法が提案されている [Dempster 72, Meinshausen 06, Yuan 07, Banerjee 08, Friedman 08]．特に近年，$\ell_1$ 正則化を用いた手法 [Meinshausen 06, Yuan 07, Banerjee 08, Friedman 08] が提案され，効率的に解を得ることが可能となった．

また，従来は 1 つのデータセットについて 1 つの GGM の推定が行われていたが，近年ではデータセットの類似性を活用して各 GGM の推定精度を向上させる複数同時推定手法 [Honorio 10, Varoquaux 10, Guo 11] が提案されている．しかし，これらの手法の多くは全ての GGM が同じ依存構造を持つことを仮定しているため，構造の異なる GGM の同時推定には適していない．

そこで，本研究では従来法が苦手とする，構造の異なる GGM の同時推定手法を提案する．我々は同時推定問題を 2 つの正則化項を持つ最尤推定問題として定式化し，ブロック座標降下法によるアルゴリズムを与えた．また，数値実験により提案法の有効性を示した．

## 2. 既存研究

確率変数 $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top$ が正規分布に従う時，2 変数 $x_j$ と $x_{j'}$ が他の変数について条件付き独立であることと，精度行列（共分散行列の逆行列）$\Lambda$ の $(j, j')$ 要素について $\Lambda_{jj'} = 0$ が成り立つことが同値であることが知られている．この性質に基づいて確率変数 $\boldsymbol{x}$ の各変数間の依存関係をグラフ表現したものがガウシアングラフィカルモデル (GGM) である．GGM の各頂点は $\boldsymbol{x}$ の各変数に対応し，各頂点間の辺構造は隣接行列 $\Lambda$ により与えられる．

一部の変数間にのみ依存性がある場合，対応する GGM は疎になり，変数間の関係を視覚的に理解しやすくなる．そのため，依存構造の推定はデータをより良く理解するための重要な方法の 1 つであると言える．しかし，通常の最尤推定では精度行列 $\Lambda$ の推定値は標本共分散 $\hat{\Sigma}$ の逆行列であり，これは一般に密行列である．この時 GGM は完全グラフとなり，本質的な依存構造が隠されてしまう．このような問題を避けて疎な構造を推定することが GGM の構造推定の目的である．

### 2.1 $\ell_1$ 正則化最尤推定

GGM の構造推定は [Dempster 72] により導入され，以降様々な手法が提案されてきた．近年，構造推定の $\ell_1$ 正則化最尤推定としての定式化が導入された [Yuan 07, Banerjee 08, Friedman 08]：

$$\max_{\Lambda \in \mathbb{R}} \ \ell(\Lambda; \hat{\Sigma}) - \rho \|\Lambda\|_1 \ \text{ subject to } \ \Lambda \succ 0 \, . \qquad (1)$$

ただし，$\rho$ は正則化パラメータであり，$\ell(\Lambda; \hat{\Sigma})$ は以下で定義される正規分布の対数尤度関数である：

$$\ell(\Lambda; \hat{\Sigma}) = \log\det\Lambda - \mathrm{tr}\left(\Lambda\hat{\Sigma}\right) \, .$$

問題 (1) は凸最適化問題であり，ブロック座標降下法により効率的に解を得ることができる [Friedman 08].

### 2.2 マルチタスク構造推定

複数の類似したタスクを，その類似性を利用して同時に解くことで解の精度を向上させることがマルチタスク学習 [Caruana 97] の目的である．このマルチタスク学習手法の 1 つ，Group Lasso [Bach 08] の手法を構造推定へと適用した手法 [Honorio 10, Varoquaux 10] が提案されている．マルチタスク構造推定では $N$ 個の標本共分散行列 $\hat{\Sigma}_1, \hat{\Sigma}_2, \ldots, \hat{\Sigma}_N$ から精度行列 $\Lambda_1, \Lambda_2, \ldots, \Lambda_N$ を同時に推定することを目的とする．[Honorio 10] らは，$N$ 個の精度行列が全て同じ疎な構造を持つと仮定し，以下の問題を導入した：

$$\max_{\Lambda_1, \Lambda_2, \ldots, \Lambda_N} \sum_{i=1}^{N} t_i \ell(\Lambda_i; \hat{\Sigma}_i) - \rho \sum_{j \neq j'} \max_{1 \leq i \leq N} |\Lambda_{i,jj'}|$$
$$\text{subject to } \ \Lambda_1, \Lambda_2, \ldots, \Lambda_N \succ 0 \, . \qquad (2)$$

ここで，$\rho$ は正則化パラメータ，$t_1, t_2, \ldots, t_N$ は非負の重みである．この問題は正則化の影響により同時構造 $\tilde{\Lambda}_{jj'} = \max_{1 \leq i \leq N} |\Lambda_{i,jj'}|$ が疎になる．つまり，一部の要素については $\Lambda_{1,jj'} = \Lambda_{2,jj'} = \ldots = \Lambda_{N,jj'} = 0$ が成り立つ．一方，各個別の GGM については疎制約がないため，上記が成り立たない要素については一般に $\Lambda_{i,jj'} \neq 0$ であり，全ての GGM は同じグラフ構造を持つ．

連絡先: 大阪大学産業科学研究所，〒 567-0047 大阪府茨木市美穂ヶ丘 8-1，{hara, washio}@ar.sanken.osaka-u.ac.jp

## 3. 問題設定

マルチタスク構造推定 (2) はタスク間の類似性を有効活用することができる一方で，全ての GGM が同じグラフ構造を有するという強い仮定を有している．この仮定を排除し，異なる構造を持つ GGM の同時構造推定を可能とするために，(2) に各 GGM を疎にする正則化項を加えた以下の問題を導入する：

$$\max_{\{\Lambda_i\}_{i=1}^N} \sum_{i=1}^N t_i \left( \ell(\Lambda_i; \hat{\Sigma}_i) - \rho \|\Lambda_i\|_1 \right) - \gamma \sum_{j \neq j'} \max_{1 \leq i \leq N} |\Lambda_{i,jj'}|$$
$$\text{subject to } \Lambda_1, \Lambda_2, \ldots, \Lambda_N \succ 0 \qquad (3)$$

ここで，$\rho, \gamma$ は正則化パラメータ，$t_1, t_2, \ldots, t_N$ は $\sum_{i=1}^N t_i = 1$ を満たす非負の重みである．この問題は $\gamma \to 0$ で個別推定 (1) と，$\rho \to 0$ で同時推定 (2) と等価となり，$\rho > 0$, $\gamma > 0$ では両者の中間的な問題となっている．

## 4. アルゴリズム

本章では問題 (3) のブロック座標降下法 [Tseng 01] による解法を紹介する．なお，提案法の最適解への収束性は [Tseng 01] の定理 4.1 から保証される．

### 4.1 ブロック座標降下法

ブロック座標降下法では行列全体を同時に最適化するのではなく，一部の値を固定し，残りの要素について逐次的に最適化を行う．今，精度行列 $\Lambda_i$ の要素のうち変数 $x_1, \ldots, x_{m-1}, x_{m+1}, \ldots, x_d$ に対応する要素を固定し，$x_m$ に関する成分だけの最適化問題を考える．問題 (3) は行・列の並び替えに対して不変なため，常に $x_m$ に対応する要素を行列の末尾へと並び替えることができる．この時，精度行列 $\Lambda_i$ 及び標本共分散 $\hat{\Sigma}_i$ を以下のように分割する：

$$\Lambda_i = \begin{bmatrix} Z_i & \boldsymbol{z}_i \\ \boldsymbol{z}_i^\top & \omega_i \end{bmatrix}, \quad \hat{\Sigma}_i = \begin{bmatrix} P_i & \boldsymbol{p}_i \\ \boldsymbol{p}_i^\top & q_i \end{bmatrix}.$$

さらに今 $Z_1, Z_2, \ldots, Z_N$ を固定しているため，$\{\boldsymbol{z}_i, \omega_i\}_{i=1}^N$ についての最適化問題は以下で与えられる：

$$\max_{\{\boldsymbol{z}_i, \omega_i\}_{i=1}^N} \sum_{i=1}^N t_i \left\{ \log\left(\omega_i - \boldsymbol{z}_i^\top Z_i^{-1} \boldsymbol{z}_i\right) - 2\boldsymbol{p}_i^\top \boldsymbol{z}_i \right.$$
$$\left. - (q_i + \rho)\omega_i - 2\rho\|\boldsymbol{z}_i\|_1 \right\} - 2\gamma \sum_j \max_{1 \leq i \leq N} |z_{ij}| . \quad (4)$$

ここで，$z_{ij}$ は $\boldsymbol{z}_i$ の第 $j$ 要素である．まず $\omega_i$ について最適化を行い以下を得る：

$$\omega_i = \boldsymbol{z}_i^\top Z_i^{-1} \boldsymbol{z}_i + (q_i + \rho)^{-1} . \qquad (5)$$

これをさらに (4) へと代入し，以下の問題を得る：

$$\min_{\{\boldsymbol{z}_i\}_{i=1}^N} \sum_{i=1}^N t_i \left\{ \frac{q_i + \rho}{2} \boldsymbol{z}_i^\top Z_i^{-1} \boldsymbol{z}_i + \boldsymbol{p}_i^\top \boldsymbol{z}_i + \rho\|\boldsymbol{z}_i\|_1 \right\}$$
$$+ \gamma \sum_j \max_{1 \leq i \leq N} |z_{ij}| .$$

ここで，さらに以下のように変数 $x_{m'}$ $(m' \neq m)$ に関する要素とそれ以外の要素へとベクトル・行列を分解する：

$$\boldsymbol{z}_i = \begin{bmatrix} \boldsymbol{v}_i \\ w_i \end{bmatrix}, \quad \boldsymbol{p}_i = \begin{bmatrix} \boldsymbol{r}_i \\ s_i \end{bmatrix}, \quad Z_i = \begin{bmatrix} H_i & \boldsymbol{h}_i \\ \boldsymbol{h}_i^\top & g_i \end{bmatrix}.$$

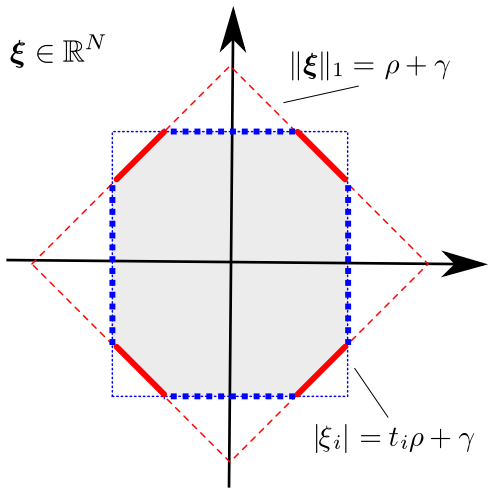


図 1: 問題 (7) の実行可能領域（彩色部）：解が境界上にある場合は太線破線部 ($|\xi_i| = t_i\rho + \gamma$) または実線部 ($\|\boldsymbol{\xi}\|_1 = \rho + \gamma$) のどちらか（それぞれ 4.2.1, 4.2.2 に対応）

これにより得られた以下の $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^\top$ についての最適化問題が (3) の部分問題である：

$$\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}^\top \mathrm{diag}(\boldsymbol{a})\boldsymbol{w} - \boldsymbol{b}^\top \boldsymbol{w} + \rho\|\boldsymbol{t} * \boldsymbol{w}\|_1 + \gamma\|\boldsymbol{w}\|_\infty \qquad (6)$$

ここで $\boldsymbol{a}$, $\boldsymbol{b}$ はそれぞれ $a_i = t_i g_i (q_i + \rho)$, $b_i = -t_i\{(q_i + \rho)\boldsymbol{h}_i^\top \boldsymbol{v}_i + s_i\}$ を要素とする $N$ 次元のベクトルであり，また $*$ はベクトルの要素積である．さらに，この問題の双対問題は以下で与えられる：

$$\min_{\boldsymbol{\xi}} \frac{1}{2} (\boldsymbol{b} - \boldsymbol{\xi})^\top \mathrm{diag}(\boldsymbol{a})^{-1} (\boldsymbol{b} - \boldsymbol{\xi})$$
$$\text{subject to } \|\boldsymbol{\xi}\|_1 \leq \rho + \gamma , \ |\xi_i| \leq t_i\rho + \gamma . \qquad (7)$$

ここで $\boldsymbol{w} = \mathrm{diag}(\boldsymbol{a})^{-1} (\boldsymbol{b} - \boldsymbol{\xi})$ である．

### 4.2 部分問題の解法

図 1 に部分問題 (7) の実行可能領域を示す．$\boldsymbol{b}$ が領域の内側にある，つまり $\|\boldsymbol{b}\|_1 \leq \rho + \gamma$, $|b_i| \leq t_i\rho + \gamma$ を満たす時，問題 (7) の解は明らかに $\boldsymbol{\xi} = \boldsymbol{b}$ である．それ以外の場合，解は実行可能領域の境界上にある．これはさらに図 1 中の太線破線部 ($|\xi_i| = t_i\rho + \gamma$) と実線部 ($\|\boldsymbol{\xi}\|_1 = \rho + \gamma$) の 2 通りの場合に分類される．以下でそれぞれの場合についての効率的な解法を紹介する．また，全体の手続きをまとめた提案法の疑似コードを Algorithm 1 に記す．

#### 4.2.1 解が境界 $|\xi_i| = t_i\rho + \gamma$ の上にある場合

この時，問題 (7) の 2 つ目の制約条件のみからなる問題を解くことで解が得られる．さらにこの場合，問題は以下の $N$ 個の 1 変数最適化問題へと分割される：

$$\min_{\xi_i} \frac{1}{2a_i} (b_i - \xi_i)^2 \quad \text{subject to } |\xi_i| \leq t_i\rho + \gamma . \qquad (8)$$

これは制約付きの最小二乗推定であり，解析的に解が得られる：

$$\xi_i = \begin{cases} b_i & (\text{if } |b_i| \leq t_i\rho + \gamma) \\ t_i\rho + \gamma & (\text{if } b_i > t_i\rho + \gamma) \\ -t_i\rho - \gamma & (\text{if } b_i < -t_i\rho - \gamma) . \end{cases}$$

**Algorithm 1** 提案法の疑似コード

**Input :** 標本共分散行列 $\hat{\Sigma}_1, \hat{\Sigma}_2, \ldots, \hat{\Sigma}_N$
正則化パラメータ $\rho, \gamma \geq 0$
重み $t_1, t_2, \ldots, t_N > 0$, $\sum_{i=1}^N t_i = 1$
**Output :** 精度行列 $\Lambda_1, \Lambda_2, \ldots, \Lambda_N$

初期化： $\Lambda_i \leftarrow \left(\hat{\Sigma}_i + \rho I_d\right)^{-1}$;
以下を $\Lambda_1, \Lambda_2, \ldots, \Lambda_N$ が収束するまで繰り返す;
**for** $x_m : m = 1$ to $d$ **do**
  **for** $x_{m'} : m' \neq m$ **do**
    **if** $\|\boldsymbol{b}\|_1 \leq \rho + \gamma$ and $|b_i| \leq t_i\rho + \gamma$ **then**
      $\boldsymbol{\xi} \leftarrow \boldsymbol{b}$;
    **else**
      部分問題 (8) を解く;
      **if** (8) の解が $\|\boldsymbol{\xi}\|_1 \leq \rho + \gamma$ を満たさない **then**
        部分問題 (9) を解く;
      **end if**
    **end if**
    $\boldsymbol{w} \leftarrow \mathrm{diag}(\boldsymbol{a})^{-1}(\boldsymbol{b} - \boldsymbol{\xi})$;
    $\Lambda_i$ の $(m, m')$ 及び $(m', m)$ 成分を $w_i$ に更新する;
  **end for**
  $\Lambda_i$ の $(m, m)$ 成分を (5) により更新する;
**end for**

ここでは 1 つ目の制約条件を無視したため，得られた解が $\|\boldsymbol{\xi}\|_1 \leq \rho + \gamma$ を満たさない可能性がある．その場合，解は境界 $\|\boldsymbol{\xi}\|_1 = \rho + \gamma$ 上にある．

#### 4.2.2 解が境界 $\|\boldsymbol{\xi}\|_1 = \rho + \gamma$ の上にある場合

ここでは問題 (7) の 1 つ目の制約条件を等式で置き換えた問題を解く．これは $\xi_i = \mathrm{sgn}(b_i) y_i$ と置いた，以下の連続二次ナップザック問題と等価である：

$$\min_{\boldsymbol{y}} \sum_{i=1}^N \frac{1}{2a_i} \left(|b_i| - y_i\right)^2$$
$$\text{subject to } \sum_{i=1}^N y_i = \rho + \gamma \,,\ 0 \leq y_i \leq t_i\rho + \gamma \,. \quad (9)$$

問題 (9) は以下の手続きにより効率的に解くことができる．まず，KKT 条件より解が $y_i(\nu) = \min(\max(|b_i| - \nu a_i, 0), t_i\rho + \gamma)$ の形で与えられ，かつ最適なパラメータ $\nu$ は $\sum_{i=1}^N y_i(\nu) = \rho + \gamma$ を満たすものであることがわかる．さらに $\sum_{i=1}^N y_i(\nu)$ が $\{(|b_i| - t_i\rho - \gamma)/a_i, |b_i|/a_i\}_{i=1}^N$ を breakpoint にもつ区分線形な単調減少関数であることから，breakpoint をソートすることにより $\sum_{i=1}^N y_i(\nu_0) \leq \rho + \gamma$ を満たす最小の breakpoint $\nu_0$ を効率的に見つけることができる．この $\nu_0$ を用いて最適なパラメータ $\nu$ は以下で与えられる：

$$\nu = \frac{\sum_{i \in I_1} (t_i\rho + \gamma) + \sum_{i \in I_2} |b_i| - \rho - \gamma}{\sum_{i \in I_2} a_i} \,.$$

ただし，$I_1$，$I_2$ はそれぞれ $I_1 = \{i; |b_i| - \nu_0 a_i \geq t_i\rho + \gamma\}$，$I_2 = \{i; 0 \leq |b_i| - \nu_0 a_i < t_i\rho + \gamma\}$ により定義されるインデックス集合である．

## 5. 数値実験

本章では数値実験により提案法の有効性を実証する．

### 5.1 データの生成

本実験では従来の同時構造推定手法 [Honorio 10, Varoquaux 10] が不得手とする，構造の異なる複数の GGM の同時構造推定問題を扱う．特に，個別推定 (1) と同時推定 (2) の中間となる「グラフ構造を部分的に共有する複数の GGM」（例：図 2）を考える．

上記の要請を満たすデータを以下の手続きにより生成した．まず，$N$ 個のデータそれぞれについて $d$ 個の変数 $x_1, x_2, \ldots, x_d$ を重複のない $k$ 個の小集合へと分割する．そして各小集合ごとに小さな精度行列をランダム*1に生成する．この時,「グラフ構造を部分的に共有する複数の GGM」を実現するために，一部の小集合が $N$ 個のデータ間で共通となるようにした．最後に，これら小さな精度行列の間に辺を加えることで N 個の疎な精度行列 $\Lambda_1, \Lambda_2, \ldots, \Lambda_N$ を生成した．実験ではデータセットの個数 $N = 5$，変数の数 $d = 20$，そして小集合の個数 $k = 5$ とし（図 2），各データセットごとに精度行列を $\Lambda_i$ とする正規分布から 100 個ずつ標本を生成した．また，前処理として各データセットごとに各変数の標準偏差が 1 になるように正規化を行った．

実験結果の評価には ROC 曲線を用いる．ここで，GGM 構造推定についての TPF(真陽性率) 及び FPF(偽陽性率) は以下により定義した：

$$\mathrm{TPF} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j \neq j'} I(\Lambda_{i,jj'} \neq 0 \wedge \hat{\Lambda}_{i,jj'} \neq 0)}{\sum_{j \neq j'} I(\Lambda_{i,jj'} \neq 0)}$$
$$\mathrm{FPF} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j \neq j'} I(\Lambda_{i,jj'} = 0 \wedge \hat{\Lambda}_{i,jj'} \neq 0)}{\sum_{j \neq j'} I(\Lambda_{i,jj'} = 0)} \,.$$

ただし，$\Lambda_i, \hat{\Lambda}_i$ はそれぞれ真の精度行列，推定された精度行列であり，$I(p)$ は命題 $p$ が真の時に 1 を，偽の時に 0 をとる関数である．

### 5.2 実験結果

上記の手続きで 100 回ランダムにデータを生成して提案法を適用し，結果を個別推定 (1) [Friedman 08] 及び同時推定 (2) [Honorio 10] と比較した結果を図 3 に示す．ここでは提案法のハイパーパラメータ $\gamma = 0.1$，また提案法及び同時推定法の重み $t_i = 0.2$ とし，それぞれの手法について $\rho$ を変化させてグラフを描画した．ROC 曲線では低い FPF で高い TPF が達成可能な良い手法ほどグラフがより左上にある．図 3 では個別推定よりも同時推定が，そして同時推定よりも提案法の曲線がより左上にある．これは提案法が個別推定と同時推定の中間の手法となっており，同時推定による精度向上を図りつつも各 GGM の個別のスパース構造推定を可能とすることで，各々の手法を単一で用いた場合に比べて良い GGM の構造推定性能を有していることを示している．

## 6. まとめと今後の課題

本研究では異なるグラフ構造を持つ GGM の同時構造推定手法を提案した．提案法は個別推定と同時推定の中間の定式化となっており，ブロック座標降下法により解を得ることができる．また，座標降下法の部分問題が 2 通りの解を持つことを示し，それぞれについて効率的な解法を与えた．数値実験により，提案法の従来法への優位性を実証した．

*1 本実験では精度行列の対角要素を 1 とし，非対角要素を $[-0.8, -0.1] \cup [0.1, 0.8]$ の上で定義される一様部分から生成した．

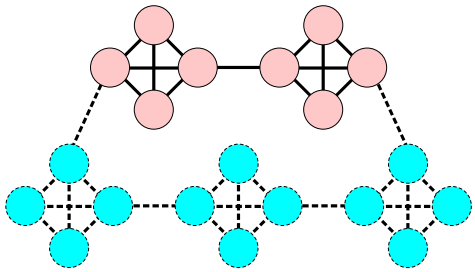

図 2: GGM の構造：実線部が複数 GGM 間で共通のグラフ構造，点線部は各 GGM により異なる構造

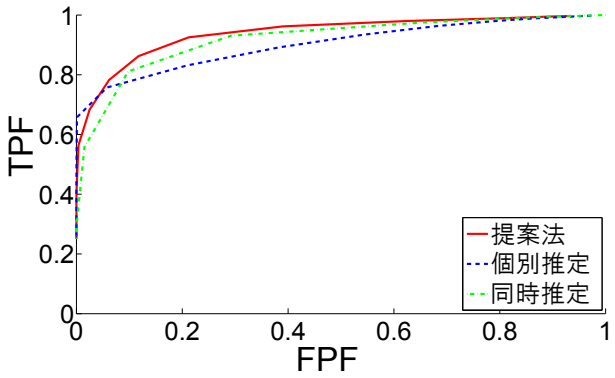


図 3: GGM 構造推定精度の ROC 曲線：データをランダムに 100 回生成し，パラメータ $\rho$ を変えながら提案法 ($\gamma = 0.1$)，個別推定 (1) 及び同時推定 (2) を適用し，得られた TPF, FPF それぞれの平均をプロット

今後の課題として問題 (3) の理論的な解析が挙げられる．特に漸近的な挙動の解析や，オラクル性 [Zou 06] を持つ定式化への拡張などは興味深い問題である．

## 謝辞

## 参考文献

[Bach 08] Bach, F. R.: Consistency of the group Lasso and multiple kernel learning, *The Journal of Machine Learning Research*, Vol. 9, pp. 1179–1225 (2008)

[Banerjee 08] Banerjee, O., El Ghaoui, L., and d'Aspremont, A.: Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data, *The Journal of Machine Learning Research*, Vol. 9, pp. 485–516 (2008)

[Caruana 97] Caruana, R.: Multitask learning, *Machine Learning*, Vol. 28, No. 1, pp. 41–75 (1997)

[Dempster 72] Dempster, A. P.: Covariance selection, *Biometrics*, Vol. 28, No. 1, pp. 157–175 (1972)

[Friedman 08] Friedman, J., Hastie, T., and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, Vol. 9, No. 3, pp. 432–441 (2008)

[Guo 11] Guo, J., Levina, E., Michailidis, G., and Zhu, J.: Joint estimation of multiple graphical models, *Biometrika*, Vol. 98, No. 1, pp. 1–15 (2011)

[Honorio 10] Honorio, J. and Samaras, D.: Multi-task learning of gaussian graphical models, in *Proceedings of the 27th Conference on Machine Learning* (2010)

[Meinshausen 06] Meinshausen, N. and Bühlmann, P.: High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics*, Vol. 34, No. 3, pp. 1436–1462 (2006)

[Tseng 01] Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization, *Journal of Optimization Theory and Applications*, Vol. 109, No. 3, pp. 475–494 (2001)

[Varoquaux 10] Varoquaux, G., Gramfort, A., Poline, J. B., and Thirion, B.: Brain covariance selection: better individual functional connectivity models using population prior, *Arxiv preprint arXiv:1008.5071* (2010)

[Yuan 07] Yuan, M. and Lin, Y.: Model selection and estimation in the Gaussian graphical model, *Biometrika*, Vol. 94, pp. 19–35 (2007)

[Zou 06] Zou, H.: The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1418–1429 (2006)