4B1-R-2-1

DAL–ADMM アルゴリズムによるスパース共分散選択

Sparse Inverse Covariance Selection via DAL–ADMM

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

Sparse Inverse Covariance Selection (SICS) is a popular tool identifying an intrinsic relationship between continuous random variables. In this paper, we propose a DAL–ADMM algorithm for SICS based on two fundamental techniques, DAL and ADMM. We also provide empirical comparisons of DAL–ADMM and existing algorithms.

1. はじめに

グラフィカルモデルの構造推定はデータから確率変数間の条件付き独立性を推定する問題である．特に，正規分布 $\mathcal{N}(\mathbf{0}_d, \Lambda^{-1})$ に従う確率変数 $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathbb{R}^d$ については 2 変数 x_i と x_j が残りの $d-2$ 個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）Λ の (i,j) 要素について $\Lambda_{ij} = 0$ が成り立つことが同値であることが知られており，構造推定は Λ 中の零成分を推定する問題へと帰着される．このような正規分布に従う確率変数を表現するグラフィカルモデルの構造推定問題は [Dempster 72] 以降，共分散選択として広く研究がなされてきた．

近年，共分散選択へ ℓ_1 正則化を用いた以下のような定式化 [Meinshausen 06, Yuan 07, Banerjee 08] が提案され，効率的な推定を行うことが可能となった：

$$\min_{\Lambda \succ 0} \ f(\Lambda) \equiv -\log\det\Lambda + \mathrm{tr}\,(S\Lambda) + \rho\|\Lambda\|_1 \,. \qquad (1)$$

ここで S は標本共分散行列，前 2 項は正規分布の負の対数尤度，$\rho \geq 0$ は正則化パラメータ，そして $\|\Lambda\|_1 = \sum_{i,j=1}^d |\Lambda_{ij}|$ は ℓ_1 正則化項であり推定値 Λ^* を疎にする効果を有している．また，この ℓ_1 正則化項は問題に応じて他のものへと置き換えることも可能である．例えば [Duchi 08, Schmidt 09] では Λ 中の要素間にグループ構造を導入，グループ正則化 [Turlach 05, Yuan 06] を用いた問題 (1) のより一般的な定式化が与えられている．また [Honorio 09] では Λ 中の一部の要素間の値の変動に正則化を与える定式化が提案されている．上記の定式化は全て凸最適化問題であり，解を一意に決定することができる．特に問題 (1) については様々な解法 [Friedman 08, Duchi 08, Yuan 09, Scheinberg 10a, Scheinberg 10b, Hsieh 11] が提案されており，中でも QUIC[Hsieh 11] は現時点において実用・理論の両面で最も優れた手法であろう．しかし，QUIC は ℓ_1 正則化項の特殊な構造を利用して効率化を図っているため，異なる正則化を用いる問題 [Duchi 08, Schmidt 09, Honorio 09] への適用は困難である．

本研究では特に [Duchi 08] による定式化に着目，QUIC よりも広いクラスの正則化項を扱うことができる解法を提案する．提案法は問題 (1) において QUIC に次ぐ性能を有する ADMM[*1](Alternating Direction Method of Multipliers)[Yuan 09, Scheinberg 10a] を基礎とする手法であり，我々はこの ADMM を双対問題へと適用することで更なる効率化を図る．

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

∗1 [Hsieh 11] の Fig.1 参照.

2. 問題設定

本章では [Duchi 08] による問題 (1) の一般的定式化について紹介する．通常の共分散選択では各 2 変数間の条件付き独立性の有無が興味の対象であるが，遺伝子ネットワークなどでは各 2 変数間よりも各モジュール（確率変数の集合）間の条件付き独立性の有無が解析の対象となる [Duchi 08]．このような場合，各 Λ_{ij} が零か否かでなく添え字集合 G によって指定されるグループ $\{\Lambda_{ij}; \{i,j\} \in G\}$ が零か否かを推定する問題となる．以降では行列から G によって指定された要素を取りだしたものを添え字 G により $\Lambda_G = \{\Lambda_{ij}; \{i,j\} \in G\}$ として表記する．[Duchi 08, Schmidt 09] は上記の問題をグループ正則化項 [Turlach 05, Yuan 06] を用いて以下のように定式化した：

$$\min_{\Lambda \succ 0} \ g(\Lambda) \equiv -\log\det\Lambda + \mathrm{tr}\,(S\Lambda) + \sum_{m=1}^{M} \rho_m \|\Lambda_{G_m}\|_p \,. \quad (2)$$

ここで M はグループの個数，$\rho_m \geq 0$ は各グループの正則化パラメータ，G_m は $\cup_{m=1}^M G_m = \{\{i,j\}; 1 \leq i,j \leq d\}$ を満たす互いに素な添え字集合である．また，Λ_{G_m} の ℓ_p–ノルムは

$$\|\Lambda_{G_m}\|_p = \begin{cases} \left(\sum_{\{i,j\}\in G_m} |\Lambda_{ij}|^p\right)^{\frac{1}{p}} & (p \in [1,\infty)) \\ \max_{\{i,j\}\in G_m} |\Lambda_{ij}| & (p = \infty) \end{cases}$$

により定義される．上記の定式化が問題 (1) の一般化となっていることは $\rho_m = \rho$ かつ $p = 1$ と置くことで問題 (1) に帰着されることから確認できる．また，$p > 1$ ではグループ毎に正則化により値が零へと丸めこまれるので，最適解 Λ^* はグループ単位で零成分を持つ行列となる．特に $p = 2, \infty$ は比較的計算が容易なため，グループ正則化によく使われる値である．本稿でも以降では $p = 2$ 及び ∞ の場合を扱う．

3. 提案法

本章では問題 (1) の ADMM による解法 [Yuan 09, Scheinberg 10a] に基づき，提案法 DAL–ADMM を導く．

ADMM は古典的な凸最適化テクニックであるが，近年その有用性（大規模化・並列化が容易，など）から改めて脚光をあ

びている手法である [Gabay 76, Boyd 11]．特に機械学習における正則化技術との相性が良く，ℓ_1 正則化に限らずグループ正則化などにも柔軟に対応することが可能である．

また，凸最適化問題では対象とする主問題と双対問題の間の関係性を用いて，これら 2 つのうちより性質の良い方（解きやすい，数値的に安定，など）を解くことで効率的な解法を作ることができる．一例として，AL 法 [Hestenes 69, Powell 67] を双対問題へと適用した Dual Augmented Lagrangian(DAL)[Tomioka 11] が挙げられる．DAL は一定条件下では AL 法よりも優れた計算効率，収束性能を示すことが経験的に知られている．ADMM は AL 法の簡略化とも解釈できる [Boyd 11] ため，DAL の ADMM による簡略的解法を構築することで問題 (1), (2) のより効率的な解法となることが期待できる．ここでは一般的な枠組みである問題 (2) の双対問題を ADMM により解く DAL–ADMM を提案する．

3.1 DAL–ADMM

ここで扱う問題 (2) の双対問題は以下で与えられる [Banerjee 08, Duchi 08]：

$$\begin{aligned}&\min_{W\succ 0} -\log\det W\\ &\text{s.t. } \|W_{G_m}-S_{G_m}\|_q\le\rho_m\ \ (1\le m\le M)\,.\end{aligned}\qquad(3)$$

ただし $W\in\mathbb{R}^{d\times d}$ は双対パラメータ，q は $p^{-1}+q^{-1}=1$ を満たす値である．また問題 (2), (3) の双対性からそれぞれの最適解 Λ^*, W^* について $\Lambda^*=W^{*-1}$ が成り立つ [Banerjee 08]．まず初めに，上記の双対問題を ADMM で扱うために以下の線形制約を有する等価な問題へと書き換える：

$$\begin{aligned}&\min_{W\succ 0,Y} -\log\det W\\ &\text{s.t. } \|Y_{G_m}\|_q\le\rho_m\ \ (1\le m\le M)\,,\\ &\qquad W-Y-S=0_{d\times d}\,.\end{aligned}$$

次に，これを用いて以下のような Augmented Lagrangian (AL) 関数を用意する：

$$\begin{aligned}\mathcal{L}_\beta(W,Y,Z)=&-\log\det W+\sum_{m=1}^{M}\delta_{\rho_m}(Y_{G_m})\\ &+\operatorname{tr}\left(Z^\top(W-Y-S)\right)+\frac{\beta}{2}\|W-Y-S\|_2^2\,.\end{aligned}$$

ここで $\delta_{\rho_m}(Y_{G_m})$ は指示関数であり

$$\delta_{\rho_m}(Y_{G_m})=\begin{cases}0 & \text{if } \|Y_{G_m}\|_q\le\rho_m\\ \infty & \text{otherwise}\end{cases}$$

により定義される．また，$\beta\ge 0$ はアルゴリズムパラメータで，$\beta=0$ において AL 関数は通常の Lagrangian に一致する．DAL–ADMM ではこの AL 関数を用いて，以下の 3 ステップを繰り返すことにより最適化を行う：

$$\begin{cases}W^{(k+1)}\in\operatorname*{argmin}_{W\succ 0}\mathcal{L}_\beta(W,Y^{(k)},Z^{(k)})\\ Y^{(k+1)}\in\operatorname*{argmin}_{Y}\mathcal{L}_\beta(W^{(k+1)},Y,Z^{(k)})\\ Z^{(k+1)}=Z^{(k)}+\beta(W^{(k+1)}-Y^{(k+1)}-S)\end{cases}.$$

以下の 2 節では W, Y の更新が効率的に実行可能であることを示す．

3.2 W の更新

W の更新式は以下の最適化問題として与えられる：

$$\min_{W\succ 0} -\log\det W+\operatorname{tr}\left(Z^{(k)\top}W\right)+\frac{\beta}{2}\|W-Y^{(k)}-S\|_2^2\,.$$

また W に関する微分を 0 と置くことにより，その最適性条件 $W-(Y^{(k)}-Z^{(k)}/\beta+S)-W^{-1}/\beta=0_{d\times d}$ が得られる．ここで固有値分解 $Y^{(k)}-Z^{(k)}/\beta+S=UDU^\top$，$D=\operatorname{diag}(\sigma_1,\sigma_2,\ldots,\sigma_d)$ を用いると解は $W=U\tilde{D}U^\top$，$\tilde{D}=\operatorname{diag}(\tilde{\sigma}_1,\tilde{\sigma}_2,\ldots,\tilde{\sigma}_d)$ と与えられることがわかる．ここで $\tilde{\sigma}_i$ は 2 次方程式 $\tilde{\sigma}_i-\sigma_i-\sigma_i^{-1}/\beta=0$ の解であり，$\tilde{\sigma}_i=(\sigma_i+\sqrt{\sigma_i^2+4/\beta})/2$ である．なお，W の正定値性はこの式から自動的に満たされていることが確認できる．

3.3 Y の更新

Y の更新式は以下の通りである：

$$\min_{Y}\sum_{m=1}^{N}\delta_{\rho_m}(Y_{G_m})-\operatorname{tr}\left(Z^{(k)\top}Y\right)+\frac{\beta}{2}\|W^{(k+1)}-Y-S\|_2^2\,.$$

これは各添え字集合 G_m に関する問題へと分割可能であり，

$$\min_{Y_{G_m}}\ \frac{1}{2}\|Y_{G_m}-B_{G_m}\|_2^2\ \text{ s.t. }\ \|Y_{G_m}\|_q\le\rho_m\qquad(4)$$

で与えられる．ただし $B_{jj'}=W_{jj'}^{(k+1)}+Z_{jj'}^{(k)}/\beta-S_{jj'}$ である．$p=1,2,\infty$ にはそれぞれ $q=\infty,2,1$ が対応しており，どの場合も上記の計算は簡単に実行可能である [Schmidt 09]．

3.4 収束性

DAL–ADMM は以下の 2 つの収束性を有している．

1. 点列 $\{Z^{(k)}\}_{k=1}^{\infty}$ は最適解 $Z^*=\Lambda^*$ に収束する．
2. 関数値 $\tilde{g}(W,Y)=-\log\det W+\sum_{m=1}^{M}\delta_{\rho_m}(Y_{G_m})$ は最適値 $\tilde{g}(W^*,Y^*)$ に 1 次収束する．

これらは Lagrangian $\mathcal{L}_0(W,Y,Z)$ から最適性条件 $Z=W^{-1}$ を得，そこに ADMM の一般的な収束性定理 [Boyd 11, He 12] 及び $\Lambda^*=W^{*-1}$ を適用することで導かれる．

3.5 実装上の詳細

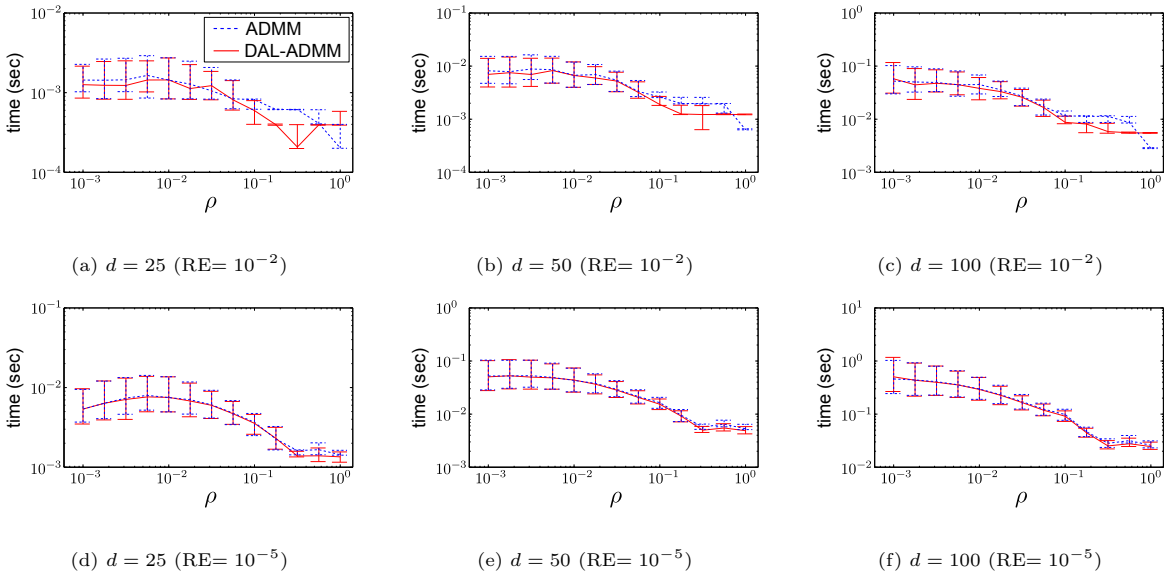
上記の理論的な収束性とは別に，アルゴリズム上では逐次更新を停止する収束判定条件が必要である．ここでは判定条件として [Boyd 11](3.3.1 節) に従い以下の 2 つの誤差指標を導入し，

$$\begin{aligned}r_{\mathrm{p}}^{(k+1)}&=\|W^{(k+1)}-Y^{(k+1)}-S\|_2\,,\\ r_{\mathrm{d}}^{(k+1)}&=\beta\|Y^{(k+1)}-Y^{(k)}\|_2\,,\end{aligned}$$

これらが閾値 ϵ よりも小さくなった時 $\max(r_{\mathrm{p}}^{(k+1)},r_{\mathrm{d}}^{(k+1)})\le\epsilon$ を収束と見なす．ここで $r_{\mathrm{p}}^{(k+1)}$, $r_{\mathrm{d}}^{(k+1)}$ はそれぞれ制約条件，解の収束性がどれだけ満たされているかを測る指標である．

ADMM, DAL–ADMM ではアルゴリズムパラメータ β の選択も収束の速さに影響を与える重要な問題である．提案法では [Boyd 11](3.4.1 節) のヒューリスティックを導入する．これは上記 2 つの誤差指標どちらか一方のみが極端に小さくなることを避け，バランスを取りながら最適化を行う方法である．具体的には以下により各ステップ毎に β の更新を行う：

$$\beta^{(k+1)}\leftarrow\begin{cases}2\beta^{(k)} & \text{if } r_{\mathrm{p}}^{(k+1)}\ge 10r_{\mathrm{d}}^{(k+1)}\\ 0.5\beta^{(k)} & \text{if } r_{\mathrm{d}}^{(k+1)}\ge 10r_{\mathrm{p}}^{(k+1)}\\ \beta^{(k)} & \text{otherwise}\end{cases}.$$



(a) $d = 25$ (RE= 10^{-2}) (b) $d = 50$ (RE= 10^{-2}) (c) $d = 100$ (RE= 10^{-2})

(d) $d = 25$ (RE= 10^{-5}) (e) $d = 50$ (RE= 10^{-5}) (f) $d = 100$ (RE= 10^{-5})

図 1: 実験 1 の結果：相対誤差 (RE) が $10^{-2}, 10^{-5}$ に達するまでの計算時間の中央値，25%点及び 75%点

4. 数値実験

本章では数値実験により提案法 DAL–ADMM の有効性を確認する．ここでは以下の 2 つの実験を行う．

1. DAL–ADMM と ADMM の比較.
2. DAL–ADMM と PSM[Duchi 08, Schmidt 09] *2の比較.

1 つ目の実験では双対問題へ ADMM を適用することでどの程度性能が改善されたかを評価する．既存の ADMM[Yuan 09, Scheinberg 10a] は問題 (1) のもとで導かれているので，比較にも問題 (1) を用いる．2 つ目の実験では一般化された問題 (2) を解く手法として，既存の PSM に対しての優位性を確認する．なお，これらの実験はともに Windows 7(64bit), Intel Xeon W3565 CPU 上で MATLAB を用いて行った．

4.1 実験 1

1 つ目の実験では [Hara 12] の Appendix B.1 の手順により疎な精度行列 Λ を生成し，正規分布 $\mathcal{N}(\mathbf{0}_d, \Lambda^{-1})$ から標本 $\{\boldsymbol{x}_n\}_{n=1}^N$ を得た．ここでは精度行列 Λ を生成するにあたり Λ が平均的に 20% の非零要素を持つように設定した．本実験ではデータの次元を $d = 25, 50$ 及び 100，標本数 $n = 5d$ とし，問題 (1) の正則化パラメータ ρ は $10^{-3} \sim 10^{0}$ の間で 13 通りに変えて比較を行った．

ランダムにデータを 1000 回生成し，ρ の各値について ADMM 及び DAL–ADMM を実行した結果を図 1 へとまとめた．各グラフはそれぞれ $d = 25, 50, 100$ の場合について相対誤差 (RE) $(f(\Lambda^{(k)}) - f(\Lambda^*))/f(\Lambda^*)$ が $10^{-2}, 10^{-5}$ になるまでの計算時間の中央値，25%点及び 75%点をプロットしたものである．図から，ρ が比較的大きい場合（結果の精度行列 Λ^* が疎な場合）に提案法が ADMM よりも速く収束する傾向を見てとれる．この差異は RE= 10^{-2} において特に顕著であり，提案法が最適解の近傍までより速く近づいていると言える．

4.2 実験 2

2 つ目の実験では [Hara 12] の Appendix B.1 の手順により T 個の疎な行列 Λ_t $(1 \le t \le T)$ を生成，これらをブロック対角状に並べた行列 Λ を精度行列として正規分布 $\mathcal{N}(\mathbf{0}_d, \Lambda^{-1})$ から標本 $\{\boldsymbol{x}_n\}_{n=1}^N$ を得た．本実験では各行列 Λ_t を 10 次元，$T = 5$ $(d = 50)$，標本数 $n = 5d$ とした．また，変数を 10 個ずつ $\mathcal{X}_t = \{x_{10t-9}, x_{10t-8}, \ldots, x_{10t}\}$ $(1 \le t \le 5)$ へと分割，これらの間のペアをグループ構造 $G_{tt'} = \{\{10t - i, 10t' - j\}; 1 \le i, j \le 10\}$ $(1 \le t, t' \le 5)$ とした．実験ではグループ G_{tt} を $p = 1$ で，$G_{tt'}$ $(t \ne t')$ を $p = \infty$ により正則化，またこの時各グループの正則化パラメータをそれぞれ $\rho_m = \rho, 100\rho$ とした．実験ではパラメータ ρ を $10^{-3} \sim 10^{0}$ の間で 13 通りに変えて PSM と DAL–ADMM の比較を行った．

ランダムにデータを 1000 回生成し，ρ の各値について PSM 及び提案法を実行した結果を図 2 へとまとめた．縦軸は相対誤差 (RE)$(g(\Lambda^{(k)}) - g(\Lambda^*))/g(\Lambda^*)$ が $10^{-2}, 10^{-5}$ になるまでの計算時間，横軸は正則化パラメータ ρ である．図から提案法が PSM よりも各段に速く収束していることを見ることができる．なお，PSM は ρ が大きい場合に収束が遅く RE= 10^{-5} に達しなかったため ρ が小さいときのみをプロットしてある．

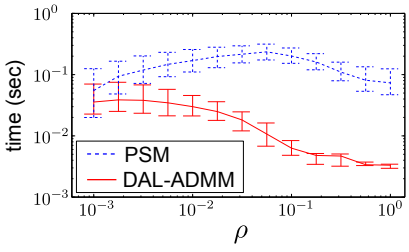
5. まとめと今後の課題

本研究では共分散選択への新たなアルゴリズム DAL–ADMM を提案した．提案法は ℓ_1 正則化を用いた定式化 (1) だけでなく，既存の QUIC[Hsieh 11] が困難とする一般的定式化 (2) へも適用が可能である．また，数値実験により問題 (1) において ρ の値が比較的大きいときに DAL–ADMM が ADMM[Yuan 09, Scheinberg 10a] よりも速く収束することを，また問題 (2) において DAL–ADMM が PSM[Duchi 08] よりも各段に少ない計算時間で高精度な解を与えることをそれぞれ確認した．

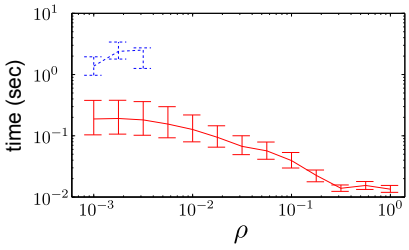
なお，本研究では特にグループ正則化に着目してアルゴリズムを導出したが，より一般のノルムを正則化に用いた場合でも，対応する Y の更新 (4) が効率的に実行可能であれば DAL–ADMM を適用することができる．

今後の課題として適切なアルゴリズムパラメータ β の値の選択方法の検討がある．現状ではこの問題にヒューリスティックを用いることで対応しているが，ADMM の収束性は β を固定したもとでのみ保証されており，値を動的に変化させた場合の収束性についての理論的な保証は与えられていない．この問題の解決は ADMM，DAL–ADMM の理論的性質の解明，実用性の向上の両面から重要な問題である．

*2 http://www.di.ens.fr/~mschmidt/Software/PQN.html



(a) RE= 10^{-2}



(b) RE= 10^{-5}

図 2: 実験 2 の結果：相対誤差 (RE) が 10^{-2}, 10^{-5} に達するまでの計算時間の中央値，25%点及び 75%点

参考文献

[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)

[Boyd 11] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, Vol. 3, No. 1, pp. 1–122 (2011)

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

[Duchi 08] Duchi, J., Gould, S., and Koller, D.: Projected subgradient methods for learning sparse gaussians, *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pp. 145–152 (2008)

[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)

[Gabay 76] Gabay, D. and Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Computers & Mathematics with Applications*, Vol. 2, No. 1, pp. 17–40 (1976)

[Hara 12] Hara, S. and Washio, T.: Learning a Common Substructure of Multiple Graphical Gaussian Models, *Arxiv preprint arXiv:1203.0117* (2012)

[He 12] He, B. and Yuan, X.: On the $O(1/n)$ Convergence Rate of the Douglas–Rachford Alternating Direction Method, *SIAM Journal on Numerical Analysis*, Vol. 50, pp. 700–709 (2012)

[Hestenes 69] Hestenes, M.: Multiplier and gradient methods, *Journal of Optimization Theory and Applications*, Vol. 4, No. 5, pp. 303–320 (1969)

[Honorio 09] Honorio, J., Ortiz, L., Samaras, D., Paragios, N., and Goldstein, R.: Sparse and locally constant Gaussian graphical models, *Advances in Neural Information Processing Systems*, Vol. 22, pp. 745–753 (2009)

[Hsieh 11] Hsieh, C., Sustik, M., Dhillon, I., and Ravikumar, P.: Sparse Inverse Covariance Matrix Estimation Using Quadratic Approximation, *Advances in Neural Information Processing Systems*, Vol. 24, pp. 2330–2338 (2011)

[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)

[Powell 67] Powell, M.: A method for non-linear constraints in minimization problems, *Optimization*, pp. 283–298 (1967)

[Scheinberg 10a] Scheinberg, K., Ma, S., and Goldfarb, D.: Sparse inverse covariance selection via alternating linearization methods, *Advances in Neural Information Processing Systems*, Vol. 23, pp. 2101–2109 (2010)

[Scheinberg 10b] Scheinberg, K. and Rish, I.: Learning sparse Gaussian Markov networks using a greedy coordinate ascent approach, *Machine Learning and Knowledge Discovery in Databases*, pp. 196–212 (2010)

[Schmidt 09] Schmidt, M., Van Den Berg, E., Friedlander, M., and Murphy, K.: Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm, *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, pp. 456–463 (2009)

[Tomioka 11] Tomioka, R., Suzuki, T., and Sugiyama, M.: Super-Linear Convergence of Dual Augmented Lagrangian Algorithm for Sparsity Regularized Estimation, *The Journal of Machine Learning Research*, Vol. 12, pp. 1537–1586 (2011)

[Turlach 05] Turlach, B., Venables, W., and Wright, S.: Simultaneous variable selection, *Technometrics*, Vol. 47, No. 3, pp. 349–363 (2005)

[Yuan 06] Yuan, M. and Lin, Y.: Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B*, Vol. 68, No. 1, pp. 49–67 (2006)

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

[Yuan 09] Yuan, X.: Alternating direction methods for sparse covariance selection, *Preprint available at http://www.optimization-online.org/DB_FILE/2009/09/2390.pdf* (2009)