# Sparse Inverse Covariance Selection via Alternating Linearization Methods (Supplementary Material)

**Katya Scheinberg**
Department of ISE
Lehigh University
katyas@lehigh.edu

**Shiqian Ma, Donald Goldfarb**
Department of IEOR
Columbia University
{sm2756,goldfarb}@columbia.edu

## Appendix

This appendix is mainly devoted to the proof of Theorem 2.1 in [1]. For clarity, we recall Algorithm 2 and Theorem 2.1 in [1].

The problem considered in [1] is

$$\min \quad F(x) \equiv f(x) + g(x). \tag{1}$$

**Algorithm 2** Alternating linearization method with skipping step

**Input:** $x^0 = y^0$
**for** $k = 0, 1, \cdots$ **do**
1. Solve $x^{k+1} := \arg\min_x Q(x, y^k) \equiv f(x) + g(y^k) - \langle \lambda^k, x - y^k \rangle + \frac{1}{2\mu}\|x - y^k\|_2^2$;
2. If $F(x^{k+1}) > Q(x^{k+1}, y^k)$ then $x^{k+1} := y^k$.
3. Solve $y^{k+1} := \arg\min_y Q_f(x^{k+1}, y)$;
4. $\lambda^{k+1} = \nabla f(x^{k+1}) - (x^{k+1} - y^{k+1})/\mu$.

**end for**

**Theorem 2.1.** *Assume $\nabla f$ is Lipschitz continuous with constant $L(f)$. For $\beta/L(f) \le \mu \le 1/L(f)$ where $0 < \beta \le 1$, Algorithm 2 satisfies*

$$F(y^k) - F(x^*) \le \frac{\|x^0 - x^*\|^2}{2\mu(k + k_n)}, \forall k, \tag{2}$$

*where $x^*$ is an optimal solution of* (1) *and $k_n$ is the number of iterations until the $k-th$ for which $F(x^{k+1}) \le Q(x^{k+1}, y^k)$. Thus Algorithm 2 produces a sequence which converges to the optimal solution in function value, and the number of iterations needed is $O(1/\epsilon)$ for an $\epsilon$-optimal solution.*

To prove Theorem 2.1, we need the following definitions and a lemma which is a generalization of Lemma 2.3 in [2]. Let $\Psi : \mathbb{R}^n \to \mathbb{R}$ and $\Phi : \mathbb{R}^n \to \mathbb{R}$ be convex functions and define

$$Q_\psi(u, v) := \phi(u) + \psi(v) + \langle \gamma_\psi(v), u - v \rangle + \frac{1}{2\mu}\|u - v\|_2^2,$$

where $\gamma_\psi(v)$ is any subgradient in the subdifferential $\partial\psi(v)$ of $\psi(v)$ at the point $v$, and

$$p_\psi(v) := \arg\min_u Q_\psi(u, v). \tag{3}$$

**Lemma A-1.** *Let $\Phi(\cdot) = \phi(\cdot) + \psi(\cdot)$. For any $v$, if*

$$\Phi(p_\psi(v)) \le Q_\psi(p_\psi(v), v), \tag{4}$$

*then for any $u$,*

$$2\mu(\Phi(u) - \Phi(p_\psi(v))) \geq \|p_\psi(v) - u\|^2 - \|v - u\|^2. \tag{5}$$

*Proof.* From (4), we have

$$\begin{aligned} \Phi(u) - \Phi(p_\psi(v)) &\geq \Phi(u) - Q_\psi(p_\psi(v), v) \\ &= \Phi(u) - \left(\phi(p_\psi(v)) + \psi(v) + \langle \gamma_\psi(v), p_\psi(v) - v\rangle + \tfrac{1}{2\mu}\|p_\psi(v) - v\|_2^2\right). \end{aligned} \tag{6}$$

Now since $\phi$ and $\psi$ are convex we have

$$\phi(u) \geq \phi(p_\psi(v)) + \langle u - p_\psi(v), \gamma_\phi(p_\psi(v))\rangle, \tag{7}$$

$$\psi(u) \geq \psi(v) + \langle u - v, \gamma_\psi(v)\rangle, \tag{8}$$

where $\gamma_\phi(\cdot)$ is a subgradient of $\phi(\cdot)$ and $\gamma_\phi(p_\psi(v))$ satisfies the first-order optimality conditions for (3), i.e.,

$$\gamma_\phi(p_\psi(v)) + \gamma_\psi(v) + \frac{1}{\mu}(p_\psi(v) - v) = 0. \tag{9}$$

Summing (7) and (8) yields

$$\Phi(u) \geq \phi(p_\psi(v)) + \langle u - p_\psi(v), \gamma_\phi(p_\psi(v))\rangle + \psi(v) + \langle u - v, \gamma_\psi(v)\rangle. \tag{10}$$

Therefore, from (6), (9) and (10) it follows that

$$\begin{aligned} \Phi(u) - \Phi(p_\psi(v)) &\geq \langle \gamma_\psi(v) + \gamma_\phi(p_\psi(v)), u - p_\psi(v)\rangle - \frac{1}{2\mu}\|p_\psi(v) - v\|_2^2 \\ &= \langle -\frac{1}{\mu}(p_\psi(v) - v), u - p_\psi(v)\rangle - \frac{1}{2\mu}\|p_\psi(v) - v\|_2^2 \\ &= \frac{1}{2\mu}\left(\|p_\psi(v) - u\|^2 - \|u - v\|^2\right). \end{aligned}$$

□

*Proof.* **[Proof of Theorem 2.1]** Let $I$ be the set of all iteration indices until $k-1$-st for which no skipping occurs and let $I_c$ be its complement. Let $I = \{n_i\},\ i = 0, \ldots, k_n - 1$. It follows that for all $n \in I_c$ $x^{n+1} = y^n$.

For $n \in I$ we can apply Lemma A-1 to obtain the following inequalities. In (5), by letting $\psi = f$, $\phi = g$, $u = x^*$ and $u = x^{n+1}$, we get $p_\psi(v) = y^{n+1}$, $\Phi = F$ and

$$2\mu(F(x^*) - F(y^{n+1})) \geq \|y^{n+1} - x^*\|^2 - \|x^{n+1} - x^*\|^2. \tag{11}$$

Similarly, by letting $\psi = g$, $\phi = f$, $u = x^*$ and $v = y^n$ in (5) we get $p_g(v) = x^{n+1}$, $\Phi = F$ and

$$2\mu(F(x^*) - F(x^{n+1})) \geq \|x^{n+1} - x^*\|^2 - \|y^n - x^*\|^2. \tag{12}$$

Taking the summation of (11) and (12) we get

$$2\mu(2F(x^*) - F(x^{n+1}) - F(y^{n+1})) \geq \|y^{n+1} - x^*\|^2 - \|y^n - x^*\|^2. \tag{13}$$

For $n \in I_c$, (11) holds, and we get

$$2\mu(F(x^*) - F(y^{n+1})) \geq \|y^{n+1} - x^*\|^2 - \|y^n - x^*\|^2, \tag{14}$$

due to the fact that $x^{n+1} = y^n$ in this case.

Summing (13) and (14) over $n = 0, 1, \ldots, k-1$ we get

$$\begin{aligned} &2\mu((2|I| + |I_c|)F(x^*) - \sum_{n\in I} F(x^{n+1}) - \sum_{n=0}^{k-1} F(y^{n+1})) \qquad (15) \\ &\geq \sum_{n=0}^{k-1}\left(\|y^{n+1} - x^*\|^2 - \|y^n - x^*\|^2\right) \\ &= \|y^k - x^*\|^2 - \|y^0 - x^*\|^2 \\ &\geq -\|x^0 - x^*\|^2. \end{aligned}$$

For any $n$, since Lemma A-1 holds for any $u$, letting $u = x^{n+1}$ instead of $x^*$ we get from (11) that

$$2\mu(F(x^{n+1}) - F(y^{n+1})) \geq \|y^{n+1} - x^{n+1}\|^2 \geq 0, \tag{16}$$

or, equivalently,

$$2\mu(F(x^n) - F(y^n)) \geq \|y^n - x^n\|^2 \geq 0. \tag{17}$$

Similarly, for $n \in I$ by letting $u = y^n$ instead of $x^*$ we get from (12) that

$$2\mu(F(y^n) - F(x^{n+1})) \geq \|x^{n+1} - y^n\|^2 \geq 0. \tag{18}$$

On the other hand, for $n \in I_c$, (18) also holds because $x^{n+1} = y^n$, and hence holds for all $n$.

Adding (16) and (18) we obtain

$$2\mu(F(y^n) - F(y^{n+1})) \geq 0. \tag{19}$$

and adding (17) and (18) we obtain

$$2\mu(F(x^n) - F(x^{n+1})) \geq 0. \tag{20}$$

(19) and (20) show that the sequences $F(y^n)$ and $F(x^n)$ are non-increasing. Thus we have,

$$\sum_{n=0}^{k-1} F(y^{n+1}) \geq kF(y^k) \quad \text{and} \quad \sum_{n \in I} F(x^{n+1}) \geq k_n F(x^k). \tag{21}$$

Combining (15) and (21) yields

$$2\mu\left((k + k_n)F(x^*) - k_n F(x^k) - kF(y^k)\right) \geq -\|x^0 - x^*\|^2. \tag{22}$$

From (17) we know that $F(x^k) \geq F(y^k)$. Thus (22) implies that

$$2\mu(k + k_n)\left(F(y^k) - F(x^*)\right) \leq \|x^0 - x^*\|^2,$$

which gives us the desired result (2).

Also, for any given $\epsilon > 0$, as long as $k \geq \frac{L(f)\|x^0 - x^*\|^2}{2\beta\epsilon}$, we have from (2) that $F(y^k) - F(x^*) \leq \epsilon$; i.e., the number of iterations needed is $O(1/\epsilon)$ for an $\epsilon$-optimal solution. □

## References


[1] K. Scheinberg, S. Ma, and D. Goldfarb. Sparse inverse covariance selection via alternating linearization methods. In *Proceedings of the Neural Information Processing Systems (NIPS)*, 2010.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.