

A Supplementary material

A.1 Lipschitz Continuity of $\nabla f(X)$

Lemma 2. For any $X, Y \in \mathbb{S}_{++}^p$,

$$\frac{1}{b^2} \|X - Y\|_2 \leq \|X^{-1} - Y^{-1}\|_2 \leq \frac{1}{a^2} \|X - Y\|_2,$$

where $a = \min\{\lambda_{\min}(X), \lambda_{\min}(Y)\}$ and $b = \max\{\lambda_{\max}(X), \lambda_{\max}(Y)\}$.

Proof. To prove the right-hand side inequality, notice that

$$X^{-1} - Y^{-1} = X^{-1}(Y - X)Y^{-1}.$$

Thus,

$$\begin{aligned} \|X^{-1} - Y^{-1}\|_2 &= \|X^{-1}(Y - X)Y^{-1}\|_2 \\ &\leq \|X^{-1}\|_2 \|X - Y\|_2 \|Y^{-1}\|_2 \\ &= \lambda_{\max}(X^{-1})\lambda_{\max}(Y^{-1}) \|X - Y\|_2 \\ &= \frac{1}{\lambda_{\min}(X)} \frac{1}{\lambda_{\min}(Y)} \|X - Y\|_2 \\ &\leq \frac{1}{a^2} \|X - Y\|_2. \end{aligned}$$

To prove the left inequality, note first that

$$Y - X = X(X^{-1} - Y^{-1})Y.$$

Therefore,

$$\begin{aligned} \|X - Y\|_2 &= \|X(X^{-1} - Y^{-1})Y\|_2 \\ &\leq \|X\|_2 \|X^{-1} - Y^{-1}\|_2 \|Y\|_2 \\ &= \lambda_{\max}(X)\lambda_{\max}(Y) \|X^{-1} - Y^{-1}\|_2 \\ &\leq b^2 \|X^{-1} - Y^{-1}\|_2. \end{aligned}$$

This shows that

$$\|X^{-1} - Y^{-1}\|_2 \geq \frac{1}{b^2} \|X - Y\|_2$$

and concludes the proof. \square

The function $\nabla f(X) = S - X^{-1}$ is Lipschitz continuous on any compact domain, since for $X, Y \in \mathbb{S}_{++}^p$ such that $aI \preceq X, Y \preceq bI$,

$$\begin{aligned} \|\nabla f(X) - \nabla f(Y)\|_F &= \|X^{-1} - Y^{-1}\|_F \\ &\leq \sqrt{p} \|X^{-1} - Y^{-1}\|_2 \\ &\leq \frac{\sqrt{p}}{a^2} \|X - Y\|_2 \\ &\leq \frac{\sqrt{p}}{a^2} \|X - Y\|_F. \end{aligned}$$

A.2 Proof of Theorem 1

We now provide the proof of Theorem 1.

Lemma 3. Let Θ_t be as in Algorithm 1 and let Θ_ρ^* be the optimal point of problem (1). Also, define

$$b := \max\{\lambda_{\max}(\Theta_t), \lambda_{\max}(\Theta_\rho^*)\}, \quad a := \min\{\lambda_{\min}(\Theta_t), \lambda_{\min}(\Theta_\rho^*)\}.$$

Then

$$\|\Theta_{t+1} - \Theta_\rho^*\|_F \leq \max\left\{\left|1 - \frac{\zeta_t}{b^2}\right|, \left|1 - \frac{\zeta_t}{a^2}\right|\right\} \|\Theta_t - \Theta_\rho^*\|_F.$$

Proof. By construction in Algorithm 1,

$$\Theta_{t+1} = \eta_{\zeta_t \rho} \left((\Theta_t - \zeta_t (S - \Theta_t^{-1})) \right)$$

Moreover, as Θ_ρ^* is a fixed point of the ISTA iteration [8, Prop. 3.1], it satisfies

$$\Theta_\rho^* = \eta_{\zeta_t \rho} \left(\Theta_\rho^* - \zeta_t (S - (\Theta_\rho^*)^{-1}) \right).$$

The soft-thresholding operator $\eta_\rho(\cdot)$ is a proximity operator corresponding to $\rho \|\cdot\|_1$. Since prox operators are non-expansive [8, Lemma 2.2], it follows that:

$$\begin{aligned} \|\Theta_{t+1} - \Theta_\rho^*\|_F &= \|\eta_{\zeta_t \rho} (\Theta_t - \zeta_t (S - \Theta_t^{-1})) - \eta_{\zeta_t \rho} (\Theta_\rho^* - \zeta_t (S - (\Theta_\rho^*)^{-1}))\|_F \\ &\leq \|\Theta_t - \zeta_t (S - \Theta_t^{-1}) - (\Theta_\rho^* - \zeta_t (S - (\Theta_\rho^*)^{-1}))\|_F \\ &= \|(\Theta_t + \zeta_t \Theta_t^{-1}) - (\Theta_\rho^* + \zeta_t (\Theta_\rho^*)^{-1})\|_F \end{aligned}$$

To bound the latter expression, recall that if $h : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a differentiable mapping, with $x, y \in U$, and $cx + (1-c)y \in U$ for all $c \in [0, 1]$, then

$$\|h(x) - h(y)\| \leq \sup_{c \in [0,1]} \{ \|J_h(cx + (1-c)y)\| \|x - y\| \}$$

where $J_h(\cdot)$ is the Jacobian of h . Define $h_\gamma : \mathbb{S}_{++}^p \rightarrow \mathbb{R}^{p^2}$ by

$$h_\gamma(X) = \text{vec}(X) + \text{vec}(\gamma X^{-1}),$$

where $\text{vec}(\cdot) : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p^2}$ is the vectorization operator defined by

$$\text{vec}(A) = (A_{1,\cdot}, A_{2,\cdot}, \dots, A_{p,\cdot})^T$$

with A_i , the i^{th} row of A . Note that for $X \in \mathbb{S}_{++}^p$,

$$\frac{\partial X}{\partial X} = I_{p^2} \quad \text{and} \quad \frac{\partial X^{-1}}{\partial X} = -X^{-1} \otimes X^{-1},$$

where \otimes is the Kronecker product and I_{p^2} is the $p^2 \times p^2$ identity matrix. Then the Jacobian of h_γ is given by:

$$J_{h_\gamma}(X) = I_{p^2} - \gamma X^{-1} \otimes X^{-1}.$$

Application of the mean value theorem to h_{ζ_t} over $Z_{t,c} = \text{vec}(c\Theta_t + (1-c)\Theta_\rho^*)$, $c \in [0, 1]$ yields

$$\begin{aligned} \|h_{\zeta_t}(\Theta_t) - h_{\zeta_t}(\Theta_\rho^*)\|_F &\leq \sup_c \left\{ \|I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1}\|_2 \right\} \|\text{vec}(\Theta_t) - \text{vec}(\Theta_\rho^*)\|_2 \\ &= \sup_c \left\{ \|I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1}\|_2 \right\} \|\Theta_t - \Theta_\rho^*\|_F. \end{aligned}$$

Denoting the eigenvalues of $Z_{t,c}$ for given values of t and c as $0 < \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_p$, the eigenvalues of $I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1}$ are $\{1 - \zeta_t (\gamma_i \gamma_j)^{-1}\}_{i,j=1}^p$. By Weyl's inequality,

$$\begin{aligned} \gamma_p &= \lambda_{\max}(Z_{t,c}) \leq \max \{ \lambda_{\max}(\Theta_t), \lambda_{\max}(\Theta_\rho^*) \} \\ \gamma_1 &= \lambda_{\min}(Z_{t,c}) \geq \min \{ \lambda_{\min}(\Theta_t), \lambda_{\min}(\Theta_\rho^*) \}, \end{aligned}$$

and therefore

$$\begin{aligned} \lambda_{\min} \left(I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1} \right) &= 1 - \frac{\zeta_t}{\gamma_1^2} \geq 1 - \frac{\zeta_t}{a^2} \\ \lambda_{\max} \left(I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1} \right) &= 1 - \frac{\zeta_t}{\gamma_p^2} \leq 1 - \frac{\zeta_t}{b^2}. \end{aligned}$$

Hence,

$$\sup_c \left\{ \|I_{p^2} - \zeta_t Z_{t,c}^{-1} \otimes Z_{t,c}^{-1}\|_2 \right\} \leq \max \left\{ \left| 1 - \frac{\zeta_t}{b^2} \right|, \left| 1 - \frac{\zeta_t}{a^2} \right| \right\}$$

which completes the proof. \square

It follows from Lemma 3 that Algorithm 1 converges linearly if

$$s_t(\zeta_t) := \max \left\{ \left| 1 - \frac{\zeta_t}{b^2} \right|, \left| 1 - \frac{\zeta_t}{a^2} \right| \right\} \in (0, 1), \forall t. \quad (23)$$

Since the minimum of

$$s(\zeta) = \max \left\{ \left| 1 - \frac{\zeta}{a^2} \right|, \left| 1 - \frac{\zeta}{b^2} \right| \right\}$$

is at $\zeta = \frac{2}{a^{-2}+b^{-2}}$, Theorem 1 follows directly from Lemma 3. It now remains to show that the eigenvalues of the G-ISTA iterates remain bounded in eigenvalue. A more general convergence result for strongly convex functions exists in the literature; this result is stated below.

Theorem 4. *Let f be strongly convex with convexity constant μ , and ∇f be Lipschitz continuous with constant L . Then for constant step size $0 < \zeta < \frac{2}{L}$, the iterates of the ISTA iteration (equation (8)), $\{x_t\}_{t \geq 0}$ to minimize $f + g$ as in (4), satisfy*

$$\|x_{t+1} - x^*\|_F \leq \max \{ |1 - \zeta L|, |1 - \zeta \mu| \} \|x_t - x^*\|_F,$$

which is to say that they converge linearly with rate $\max \{ |1 - \zeta L|, |1 - \zeta \mu| \}$. Furthermore,

1. The step size which yields an optimal worst-case contraction bound is $\zeta = \frac{2}{\mu+L}$.

2. The optimal worst-case contraction bound corresponding to $\zeta = \frac{2}{\mu+L}$ is given by

$$\begin{aligned} s(\zeta) &:= \max \{ |1 - \zeta L|, |1 - \zeta \mu| \} \\ &= 1 - \frac{2}{1 + \frac{\mu}{L}}. \end{aligned}$$

Proof. See [7, 21] and references therein. □

A.3 Proof of Theorem 2

In this section, the eigenvalues of $\Theta_t, \forall t$ are bounded. To begin, the eigenvalues of $\Theta_{t+\frac{1}{2}} := \Theta_t - \zeta_t(S - \Theta_t^{-1})$ are bounded.

Lemma 4. Let $0 < a < b$ be given positive constants and let $\zeta_t > 0$. Assume $aI \preceq \Theta_t \preceq bI$. Then the eigenvalues of $\Theta_{t+\frac{1}{2}} := \Theta_t - \zeta_t(S - \Theta_t^{-1})$ satisfy:

$$\lambda_{\min}(\Theta_{t+\frac{1}{2}}) \geq \begin{cases} 2\sqrt{\zeta_t} - \zeta_t \lambda_{\max}(S) & \text{if } a \leq \sqrt{\zeta_t} \leq b \\ \min \left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b} \right) - \zeta_t \lambda_{\max}(S) & \text{otherwise} \end{cases} \quad (24)$$

and

$$\lambda_{\max}(\Theta_{t+\frac{1}{2}}) \leq \max \left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b} \right) - \zeta_t \lambda_{\min}(S).$$

Proof. Denoting the eigenvalue decomposition of Θ_t by $\Theta_t = U\Gamma U^T$,

$$\begin{aligned} \Theta_{t+\frac{1}{2}} &= \Theta_t - \zeta_t(S - \Theta_t^{-1}) \\ &= U\Gamma U^T - \zeta_t(S - U\Gamma^{-1}U^T) \\ &= U(\Gamma - \zeta_t(U^T S U - \Gamma^{-1}))U^T \end{aligned}$$

Let $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_p)$ with $\gamma_1 \leq \dots \leq \gamma_p$. By Weyl's inequality, the eigenvalues of $\Theta_{t+\frac{1}{2}}$ are bounded below by

$$\lambda_i \left(\Theta_{t+\frac{1}{2}} \right) \geq \gamma_i + \frac{\zeta_t}{\gamma_i} - \zeta_t \lambda_{\max}(S),$$

and bounded above by

$$\lambda_i \left(\Theta_{t+\frac{1}{2}} \right) \leq \gamma_i + \frac{\zeta_t}{\gamma_i} - \zeta_t \lambda_{\min}(S)$$

The function $f(x) = x + \frac{\zeta_t}{x}$ over $a \leq x \leq b$ has only one extremum which is a global minimum at $x = \sqrt{\zeta_t}$. Therefore,

$$\min_{a \leq x \leq b} x + \frac{\zeta_t}{x} = \begin{cases} 2\sqrt{\zeta_t} & \text{if } a \leq \sqrt{\zeta_t} \leq b \\ \min\left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b}\right) & \text{otherwise} \end{cases},$$

and

$$\max_{a \leq x \leq b} x + \frac{\zeta_t}{x} = \max\left(a + \frac{\zeta_t}{b}, b + \frac{\zeta_t}{a}\right).$$

Since $a \leq \gamma_1 \leq b$,

$$\begin{aligned} \lambda_{\min}(\Theta_{t+\frac{1}{2}}) &\geq \gamma_1 + \frac{\zeta_t}{\gamma_1} - \zeta_t \lambda_{\max}(S) \\ &\geq \min_{a \leq x \leq b} \left(x + \frac{\zeta_t}{x}\right) - \zeta_t \lambda_{\max}(S) \\ &= \begin{cases} 2\sqrt{\zeta_t} - \zeta_t \lambda_{\max}(S) & \text{if } a \leq \sqrt{\zeta_t} \leq b \\ \min\left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b}\right) - \zeta_t \lambda_{\max}(S) & \text{otherwise} \end{cases} \end{aligned}$$

Similarly,

$$\begin{aligned} \lambda_{\max}(\Theta_{t+\frac{1}{2}}) &\leq \gamma_p + \frac{\zeta_t}{\gamma_p} - \zeta_t \lambda_{\min}(S) \\ &\leq \max_{a \leq x \leq b} \left(x + \frac{\zeta_t}{x}\right) - \zeta_t \lambda_{\min}(S) \\ &= \max\left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b}\right) - \zeta_t \lambda_{\min}(S). \end{aligned}$$

□

It remains to demonstrate that the soft-thresholded iterates Θ_{t+1} remain bounded in eigenvalue.

Lemma 5. Let $0 < a < b$ and $\zeta_t > 0$. Then:

$$\min\left(a + \frac{\zeta_t}{a}, b + \frac{\zeta_t}{b}\right) = a + \frac{\zeta_t}{a}$$

if and only if $\zeta_t \leq ab$.

Proof. Under the stated assumptions,

$$\begin{aligned} a + \frac{\zeta_t}{a} \leq b + \frac{\zeta_t}{b} &\Leftrightarrow \zeta_t \left(\frac{1}{a} - \frac{1}{b}\right) \leq b - a \\ &\Leftrightarrow \zeta_t \leq \frac{b - a}{\frac{1}{a} - \frac{1}{b}} \\ &\Leftrightarrow \zeta_t \leq ab. \end{aligned}$$

□

Lemma 6. Let A be a symmetric $p \times p$ matrix. Then the soft-thresholded matrix $\eta_\epsilon(A)$ satisfies

$$\lambda_{\min}(A) - p\epsilon \leq \lambda_{\min}(\eta_\epsilon(A))$$

In particular, A_ϵ is positive definite if $\lambda_{\min}(A) > p\epsilon$.

Proof. Let

$$\mathcal{A} := \{M \in \mathbb{M}_p : M_{i,j} \in \{0, 1, -1\}\}.$$

For every $\epsilon > 0$, the matrix A_ϵ can be written as

$$\eta_\epsilon(A) = A + \epsilon_1 A_1 + \epsilon_2 A_2 + \cdots + \epsilon_k A_k,$$

for some $k \leq \binom{p}{2} + p$ where $A_i \in \mathcal{A}$, $\epsilon_i > 0$ and $\sum_{i=1}^k \epsilon_i = \epsilon$. Now let

$$c_p := \max\{|\lambda_{\min}(M)| : M \in \mathcal{A}\}.$$

The constant c_p is finite since \mathcal{A} is a finite set. Since $-A \in \mathcal{A}$ for every $A \in \mathcal{A}$, and since $|\lambda_{\min}(-A)| = |\lambda_{\max}(A)|$, it follows that

$$c_p = \max\{|\lambda_{\max}(M)| : M \in \mathcal{A}\}.$$

Applying the Gershgorin circle theorem [see, e.g., 12] gives $c_p \leq p$. Since p is an eigenvalue of the matrix B such that $B_{i,j} = 1$ for all i, j , it follows that $c_p = p$.

Recursive application of Weyl's inequality gives that

$$\begin{aligned} \lambda_{\min}(\eta_\epsilon(A)) &\geq \lambda_{\min}(A) - \epsilon|\lambda_{\max}(A_1)| - \cdots - \epsilon_k|\lambda_{\max}(A_k)| \\ &\geq \lambda_{\min}(A) - c_p \sum_{i=1}^k \epsilon_i \\ &= \lambda_{\min}(A) - c_p \epsilon. \end{aligned}$$

□

Recall from Lemma 1 that the eigenvalues of the optimal solution to problem (1) are bounded below by $\frac{1}{\|S\|_{2+p\rho}}$. The following theorem shows that $\alpha = \frac{1}{\|S\|_{2+p\rho}}$ is a valid bound to ensure that $\alpha I \preceq \Theta_{t+1}$ if $\alpha I \preceq \Theta_t$.

Lemma 7. Let $\rho > 0$ and $\alpha = \frac{1}{\|S\|_{2+p\rho}} < b'$. Assume $\alpha I \preceq \Theta_t \preceq b' I$ and consider

$$\Theta_{t+1} = \eta_{\zeta_t, \rho}(\Theta_t - \zeta_t(S - \Theta_t^{-1}))$$

Then for every $0 < \zeta_t \leq \alpha^2$, $\alpha I \preceq \Theta_{t+1}$.

Proof. The result follows by combining Lemma 4 and Lemma 6. Notice first that the hypothesis $\zeta_t \leq \alpha^2$ guarantees that $\sqrt{\zeta_t} \notin [\alpha, b']$. Also, from Lemma 5, we have

$$\min\left(\alpha + \frac{\zeta_t}{\alpha}, b' + \frac{\zeta_t}{b'}\right) = \alpha + \frac{\zeta_t}{\alpha}$$

since $\zeta_t \leq \alpha^2 \leq \alpha b'$. Hence, by Lemma 4,

$$\begin{aligned} \lambda_{\min}(\Theta_{t+\frac{1}{2}}) &\geq \min\left(\alpha + \frac{\zeta_t}{\alpha}, b' + \frac{\zeta_t}{b'}\right) - \zeta_t \lambda_{\max}(S) \\ &= \alpha + \frac{\zeta_t}{\alpha} - \zeta_t \lambda_{\max}(S). \end{aligned}$$

Now, applying Lemma 6 to $\Theta_{t+1} = \eta_{\zeta_t, \rho}(\Theta_{t+\frac{1}{2}})$, we obtain

$$\begin{aligned} \lambda_{\min}(\Theta_{t+1}) &= \lambda_{\min}\left(\eta_{\zeta_t, \rho}(\Theta_{t+\frac{1}{2}})\right) \\ &\geq \lambda_{\min}(\Theta_{t+\frac{1}{2}}) - p\rho\zeta_t \\ &\geq \alpha + \frac{\zeta_t}{\alpha} - \zeta_t \lambda_{\max}(S) - p\rho\zeta_t. \end{aligned}$$

We therefore have $\alpha I \preceq \Theta_{t+1}$ whenever

$$\alpha + \frac{\zeta_t}{\alpha} - \zeta_t \lambda_{\max}(S) - p\rho\zeta_t \geq \alpha.$$

This is equivalent to

$$\zeta_t \left(\frac{1}{\alpha} - \lambda_{\max}(S) - p\rho \right) \geq 0.$$

Since $\zeta_t > 0$, this is equivalent to

$$\frac{1}{\alpha} - \lambda_{\max}(S) - p\rho \geq 0.$$

Reorganizing the terms of the previous equation, we obtain that $\alpha I \preceq \Theta_{t+1}$ if

$$\alpha \leq \frac{1}{\lambda_{\max}(S) + p\rho} = \frac{1}{\|S\|_2 + p\rho}.$$

□

It remains to show that the eigenvalues of the iterates Θ_t remain bounded above, for all t .

Lemma 8. Let $\alpha = \frac{1}{\|S\|_2 + p\rho}$ and let $\zeta_t \leq \alpha^2, \forall t$. Then the G-ISTA iterates Θ_t satisfy $\Theta_t \preceq b'I, \forall t$, with $b' = \|\Theta_\rho^*\|_2 + \|\Theta_0 - \Theta_\rho^*\|_F$.

Proof. By Lemma 7, $\alpha I \preceq \Theta_t$ for every t . As $\alpha I \preceq \Theta^*$ (Lemma 1),

$$\Lambda_t^- := \min\{\lambda_{\min}(\Theta_t), \lambda_{\min}(\Theta_\rho^*)\}^2 \geq \alpha^2.$$

for all t . Also, since $\Lambda_t^+ \geq \Lambda_t^-$ and $\zeta_t \leq \alpha^2$,

$$\max \left\{ \left| 1 - \frac{\zeta_t}{b^2} \right|, \left| 1 - \frac{\zeta_t}{\alpha^2} \right| \right\} \leq 1.$$

Therefore, by Lemma 3,

$$\|\Theta_t - \Theta_\rho^*\|_F \leq \|\Theta_{t-1} - \Theta_\rho^*\|_F.$$

Applying this result recursively gives

$$\|\Theta_t - \Theta_\rho^*\|_F \leq \|\Theta_0 - \Theta_\rho^*\|_F.$$

Since $\|\cdot\|_2 \leq \|\cdot\|_F$, we therefore have

$$\|\Theta_t\|_2 - \|\Theta_\rho^*\|_2 \leq \|\Theta_t - \Theta_\rho^*\|_2 \leq \|\Theta_t - \Theta_\rho^*\|_F \leq \|\Theta_0 - \Theta_\rho^*\|_F,$$

and so,

$$\lambda_{\max}(\Theta_t) = \|\Theta_t\|_2 \leq \|\Theta_\rho^*\|_2 + \|\Theta_0 - \Theta_\rho^*\|_F$$

which completes the proof. □

A.4 Additional timing comparisons

This section provides additional synthetic timing comparisons for $p = 500$ and $p = 5000$. In addition, two real datasets were investigated. The “estrogen” dataset [22] contains $p = 652$ dimensional gene expression data from $n = 158$ breast cancer patients. The “temp” dataset [6] consists of average annual temperature measurements from $p = 1732$ locations over $n = 157$ years (1850-2006).

		ρ	0.05	0.10	0.15	0.20
problem	algorithm	time/iter	time/iter	time/iter	time/iter	time/iter
$p = 500$ $n = 100$ $\text{nnz}(\Omega) = 3\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	31.61%/42.76	19.61%/18.23	11.08%/8.13	5.02%/3.06	
	glasso	28.34/11	10.91/8	7.08/7	5.57/6	
	QUIC	8.33/23	1.98/13	0.96/11	0.38/10	
	G-ISTA	4.44/402	1.14/110	0.30/38	0.14/18	
$p = 500$ $n = 600$ $\text{nnz}(\Omega) = 3\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	20.73%/6.62	3.93%/2.44	0.90%/1.49	0.13%/1.20	
	glasso	7.44/6	4.53/5	3.45/4	2.62/3	
	QUIC	1.08/9	0.17/7	0.06/5	0.04/5	
	G-ISTA	0.28/31	0.10/13	0.07/9	0.03/5	
$p = 500$ $n = 100$ $\text{nnz}(\Omega) = 15\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	31.36%/46.83	19.74%/19.93	11.65%/8.95	5.45%/3.25	
	glasso	28.61/11	11.27/8	7.22/7	5.34/6	
	QUIC	8.47/23	2.01/13	0.73/9	0.22/7	
	G-ISTA	4.80/466	1.09/115	0.28/34	0.15/20	
$p = 500$ $n = 600$ $\text{nnz}(\Omega) = 15\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	24.81%/9.78	6.36%/2.64	0.79%/1.28	0.03%/1.08	
	glasso	8.52/6	4.59/5	3.55/4	2.54/3	
	QUIC	1.56/10	0.25/7	0.05/5	0.03/5	
	G-ISTA	0.50/51	0.10/13	0.06/7	0.02/3	

Table 2: Timing comparisons for $p = 500$ dimensional datasets, generated as in Section 5.1

		ρ	0.02	0.04	0.06	0.08
problem	algorithm	time/iter	time/iter	time/iter	time/iter	time/iter
$p = 5000$ $n = 1000$ $\text{nnz}(\Omega) = 3\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	26.22%/54.47	13.68%/23.74	6.36%/8.69	2.03%/2.31	
	glasso	30814.29/11	12612.85/8	9224.79/7	6184.84/5	
	QUIC	22547.70/21	3725.07/11	946.11/8	199.48/6	
	G-ISTA	2651.43/575	417.20/94	93.33/25	39.05/11	
$p = 5000$ $n = 6000$ $\text{nnz}(\Omega) = 3\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	12.89%/15.18	3.23%/3.73	1.11%/1.60	0.16%/1.16	
	glasso	10307.26/7	8725.86/7	4846.58/4	3587.35/3	
	QUIC	3108.14/10	396.60/7	86.66/5	21.56/4	
	G-ISTA	268.28/70	50.17/14	35.67/10	28.82/8	
$p = 5000$ $n = 1000$ $\text{nnz}(\Omega) = 15\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	26.08%/80.04	13.93%/37.12	6.91%/16.52	2.47%/3.08	
	glasso	36302.86/11	13413.57/8	9914.41/7	7408.33/6	
	QUIC	22667.29/21	4649.99/12	1329.20/9	240.25/6	
	G-ISTA	3952.85/849	701.57/170	176.11/45	42.46/12	
$p = 5000$ $n = 6000$ $\text{nnz}(\Omega) = 15\%$	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	18.65%/27.69	5.34%/7.26	0.66%/1.41	0.03%/1.09	
	glasso	13180.47/7	9052.77/7	4842.28/4	3578.05/3	
	QUIC	6600.91/12	795.46/8	59.03/5	16.10/4	
	G-ISTA	804.93/189	103.69/23	36.17/10	18.87/5	

Table 3: Timing comparisons for $p = 5000$ dimensional datasets, generated as in Section 5.1.

		ρ	0.15	0.30	0.45	0.60
problem	algorithm	time/iter	time/iter	time/iter	time/iter	time/iter
$p = 682$ $n = 158$ <i>Dataset: estrogen</i>	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	5.29%/290.03	3.39%/88.55	2.31%/29.69	1.63%/8.96	
	glasso	106.18/24	120.18/34	110.54/35	40.52/13	
	QUIC	12.36/19	2.71/11	1.08/9	0.54/7	
	G-ISTA	43.96/2079	11.99/595	3.23/172	1.00/53	
problem	algorithm	time/iter	time/iter	time/iter	time/iter	
$p = 1732$ $n = 157$ <i>Dataset: temp</i>	$\text{nnz}(\Omega_\rho^*)/\kappa(\Omega_\rho^*)$	2.02%/1075.8	1.77%/289.63	1.34%/23.02	0.22%/2.10	
	glasso	1919.64/31	2535.86/46	1144.07/22	254.14/5	
	QUIC	497.47/18	103.76/13	10.16/8	2.31/7	
	G-ISTA	1221.40/6194	183.20/819	30.01/159	1.78/10	

Table 4: Timing comparisons for the real datasets described above.