

A Proof

A.1 Proof for 3.1

Proof. Let η_j^k be the objective function value of the j th step in sub-problem 2, defined below:

$$\min_{\beta} \tilde{l}^k(\beta) + \sum_{i=1}^{j-1} \tilde{f}_i^{k+1}(\beta) + f_j(\beta) + \sum_{i=j+1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2.$$

From the successive linearization of $l(\beta)$ and $f_j(\beta)$, it can be shown:

$$\eta^k \leq \eta_1^k \leq \eta_2^k \leq \dots \leq \eta_N^k \leq \eta^{k+1}.$$

We explain why $\eta^k \leq \eta_1^k$ where

$$\begin{aligned} \eta^k &= \min_{\beta} l(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2, \\ z^k &= \arg \min_{\beta} l(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2, \\ \eta_1^k &= \min_{\beta} \tilde{l}^k(\beta) + f_1(\beta) + \sum_{i=2}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2. \\ z_1^{k+1} &= \arg \min_{\beta} \tilde{l}^k(\beta) + f_1(\beta) + \sum_{i=2}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2, \end{aligned}$$

where $\tilde{l}^k(\beta) = l(z^k) + \langle g^k, \beta - z^k \rangle$, and $g^k \in \partial l(z^k)$. Since:

$$z^k = \arg \min_{\beta} l(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2.$$

according to optimality condition and $g^k \in \partial l(z^k)$.:

$$z^k = \arg \min_{\beta} \tilde{l}^k(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2.$$

Replace the term $\tilde{f}_1^k(\beta)$ of the problem above with $f_1(\beta)$, since $\tilde{f}_1^k(\beta) \leq f_1(\beta)$, we always have:

$$\tilde{l}^k(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \leq \tilde{l}^k(\beta) + f_1(\beta) + \sum_{i=2}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2$$

Therefore, $\eta^k \leq \eta_1^k$.

We approximate the increase made from η^k to η_N^k where:

$$\begin{aligned} \eta^k &= \min_{\beta} l(\beta) + \sum_{i=1}^N \tilde{f}_i^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2, \\ \eta_N^k &= \min_{\beta} \tilde{l}^k(\beta) + \sum_{i=1}^{N-1} \tilde{f}_i^{k+1}(\beta) + f_N(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2. \end{aligned}$$

Consider a family of relaxation, given by

$$Q_k(\beta, \mu) = \sum_{j=1}^N \left[(1-\mu)(f_j(z_j^k) + \langle g_j^k, \beta - z_j^k \rangle) + \mu(f_j(z^k) + \langle s_j^k, \beta - z^k \rangle) \right] + \tilde{t}^k(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|^2,$$

where $g_j^k \in \partial f_j(z_j^k)$ and s_j^k is any element of $\partial f_j(z^k)$. We can see that $Q_k(\beta, \mu)$ is a relaxation of η_N^k for $\mu \in [0, 1]$ and when $\mu = 0$, $\min_{\beta} Q_k(\beta, 0) = \eta^k$. Denote $\hat{Q}_k(\beta, \mu) = \min_{\beta} Q_k(\beta, \mu)$. The increase from η^k to η_N^k can be bounded by

$$\eta_N^k - \eta^k \geq \max_{\mu \in [0, 1]} \hat{Q}_k(\beta, \mu) - \hat{Q}_k(\beta, 0).$$

It can be easily seen that the optimal solution $\hat{\beta}(\mu)$ to the preceding maximization can be solved in closed-form:

$$\arg \min_{\beta} Q_k(\beta, \mu) = \beta^k - (g^k + \sum_{i=1}^N g_i^k + \mu(s_j^k - g_j^k)) / \rho.$$

We also have:

$$\nabla_{\mu} \hat{Q}_k(\mu) = -\mu \left\| \sum_{j=1}^N s_j^k - g_j^k \right\|_2^2 / \rho + F(z^k) - \tilde{F}^k(z^k).$$

For a fixed $\tilde{\mu} \in [0, 1]$, we have:

$$\hat{Q}_k(\tilde{\mu}) - \hat{Q}_k(0) = \int_0^{\tilde{\mu}} \hat{Q}'_k(\mu) d\mu = \tilde{\mu} (F(z^k) - \tilde{F}^k(z^k)) - \frac{1}{2\rho} \tilde{\mu} \left\| \sum_{j=1}^N s_j^k - g_j^k \right\|_2^2.$$

Choosing $\tilde{\mu} = \min(1, \frac{\rho(F(z^k) - \tilde{F}^k(z^k))}{\left\| \sum_{j=1}^N s_j^k - g_j^k \right\|_2^2})$, we obtain: $\eta_N^k - \eta^k \geq \frac{1}{2} \tilde{\mu} (F(z^k) - \tilde{F}^k(z^k))$. \square

A.2 Proof for Theorem 3.2

Proof.

- (i) Can be deduced from Lemma 1 and the condition for update step.
- (ii) The first part is straight forward using Lemma 1 and the condition for update step. By the construction of the sub-gradients in sub-problem 1, we have:

$$-\rho(z^k - \beta^k) = g^k + \sum_{i=j}^N g_j^{k-1} \in \partial \tilde{F}^k(z^k).$$

By the definition of sub-gradients, we obtain:

$$\vartheta^k = F(\beta^k) - \tilde{F}^k(z^k) \geq \tilde{F}(\beta^k) - \tilde{F}^k(z^k) \geq \rho \|z^k - \beta^k\|_2^2.$$

If $\vartheta = 0$ then $z^k = \beta^k$. Let β^k be the last estimate.

Let $p(\beta^k) = \arg \min_{\beta} F(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2$. Because $F(\beta) \geq \tilde{F}^k(\beta)$ and $z^k = \arg \min_{\beta} \tilde{F}^k + \frac{\rho}{2} \|\beta - \beta^k\|_2^2$, we have:

$$F(p(\beta^k)) + \frac{\rho}{2} \|p(\beta^k) - \beta^k\|_2^2 \geq \tilde{F}^k(z^k) + \frac{\rho}{2} \|p(\beta^k) - z^k\|_2^2 + \frac{\rho}{2} \|z^k - \beta^k\|_2^2.$$

Similarly, we have:

$$F(z^k) + \frac{\rho}{2} \|z^k - \beta^k\|_2^2 \geq F(p(\beta^k)) + \frac{\rho}{2} \|z^k - p(\beta^k)\|_2^2 + \frac{\rho}{2} \|p(\beta^k) - \beta^k\|_2^2.$$

Adding the two inequalities we obtain:

$$F(z^k) - \tilde{F}^k(z^k) \geq \rho \|z^k - p(\beta^k)\|_2^2.$$

The $LHS \rightarrow 0$ therefore $z^k \rightarrow p(\beta^k)$. Since $z^k = \beta^k$, we have $\beta^k = \arg \min_{\beta} F(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2$ or $\beta^k \in \arg \min F$.

(iii) Can be deduced from Lemma 1 and the condition for update step.

□

A.3 Proof for Theorem 3.3

Proof. We have:

$$\begin{aligned} \|\beta^{k+1} - \tilde{\beta}\|_2^2 - \|\beta^k - \tilde{\beta}\|_2^2 &= \langle \beta^{k+1} - \beta^k, \beta^k - \beta^{k+1} + 2\beta^{k+1} - 2\tilde{\beta} \rangle \\ &\leq 2 \frac{\tilde{F}^k(\tilde{\beta}) - \tilde{F}^k(\beta^{k+1})}{\rho} \leq 2 \frac{F(\tilde{\beta}) - F(\beta^k) + \vartheta_k}{\rho} \\ &\leq 2 \frac{\vartheta_k}{\rho} \leq \frac{2}{\rho} (F(\beta^k) - F(\beta^{k+1}))/\gamma. \end{aligned} \quad (1)$$

Sum these inequalities up in terms of k , using the fact that F is finite-valued, we have $\{\beta^k\}$ is bounded, thus, it has a limit point $\tilde{\beta}$. Since F is a closed function, $F(\tilde{\beta}) \leq \liminf_{\beta^k \rightarrow \tilde{\beta}} F(\beta^k)$, hence, the same inequality applies for $\tilde{\beta}$:

$$\|\beta^{k+1} - \tilde{\beta}\|_2^2 - \|\beta^k - \tilde{\beta}\|_2^2 \leq 2 \frac{\vartheta_k}{\rho}. \quad (2)$$

Consider two subsequences: $\{\beta_i^{k1}\} \rightarrow \tilde{\beta}$ and $\{\beta_j^{k2}\} \rightarrow \tilde{\beta}$. For some $i \geq j$, combine the two inequalities above we have:

$$\|\beta_i^{k1} - \tilde{\beta}\|_2^2 \leq \|\beta_j^{k2} - \tilde{\beta}\|_2^2 + \sum_{k \geq j} \frac{2}{\rho} \vartheta_k. \quad (3)$$

Since $\vartheta^k \rightarrow 0$, $\sum_{k \in \mathcal{K}} \vartheta^k \leq \infty$, so $\sum_{k \geq j} \vartheta^k \rightarrow 0$. The RHS of the inequality $\rightarrow 0$ or $\{\beta_i^{k1}\} \rightarrow \tilde{\beta}$. Therefore there is only a unique limit point $\tilde{\beta}$.

By construction, the subgradient of function F at the point z^k can be written as $g^k + \sum_{i=1}^N g_i^k = -\rho(z^k - \beta^k)$. Thus, for all β , we obtain:

$$\begin{aligned} F(\beta) &\geq l(z^k) + \sum_{i=1}^N f_i(z^k) + \langle g^k + \sum_{i=1}^N g_i^k, \beta - z^k \rangle \\ &= F(\beta^k) - \vartheta^k + \langle g^k + \sum_{i=1}^N g_i^k, \beta - z^k \rangle. \end{aligned}$$

Before we had $\rho \|z^k - \beta^k\| \leq \vartheta^k$ and $\vartheta^k \rightarrow 0$, it follows that $z^k - \beta^k \rightarrow 0$. Furthermore, the sequence $z^k \rightarrow \tilde{\beta}$. This implies that $g^k + \sum_{i=1}^N g_i^{k-1} \rightarrow 0$, and $F(\beta) \geq \lim F(\beta^k) \geq F(\tilde{\beta})$. \square

A.4 Proof for Lemma 3.4

Proof. Since iteration k is a null step we have:

$$\begin{aligned} F(z^k) &> F(\beta^k) - \gamma(F(\beta^k) - \tilde{F}(z^k)) \\ F(z^k) - \tilde{F}(z^k) &> (1 - \gamma)(F(\beta^k) - \tilde{F}(z^k)) = (1 - \gamma)\vartheta^k \end{aligned}$$

Thus,

$$\begin{aligned} F(\beta^k) - \eta^{k+1} &\leq F(\beta^k) - \eta^k - \frac{1}{2}\tilde{\mu}(F(z^k) - \tilde{F}(z^k)) \\ &\leq F(\beta^k) - \eta^k - \frac{1}{2}\tilde{\mu}(1 - \gamma)(F(\beta^k) - \tilde{F}(z^k)) \\ &\leq (1 - \frac{1}{2}\tilde{\mu}(1 - \gamma))(F(\beta^k) - \eta^k) \end{aligned}$$

By the definition of $\tilde{\mu} = \min(1, \frac{\rho(F(z^k) - \tilde{F}(z^k))}{\|\sum_{i=1}^N s_i^k - g_i^{k-1}\|^2})$ and since f_i 's are finite-valued convex functions, so the norm of the sub-gradients is bounded so there exists $M > 0$ such that

$$\rho \|\sum_{i=1}^N s_i^k - g_i^{k-1}\|_2^2 \leq M.$$

Thus, we obtain:

$$F(\beta^k) - \eta^{k+1} \leq (1 - \frac{(1 - \gamma)^2 \varepsilon}{M})(F(\beta^k) - \eta^k).$$

\square

A.5 Proof for Lemma 3.6

Proof. By definition of $F_\rho(\beta^k)$ we have:

$$\begin{aligned} \min_{\beta} \{F(\beta) + \frac{\rho}{2}\|\beta - \beta^k\|_2^2\} &\geq \min_{\beta} \{\tilde{F}(\beta) + \frac{\rho}{2}\|\beta - \beta^k\|_2^2\} \\ &= \tilde{F}(z^k) + \frac{\rho}{2}\|z^k - \beta^k\|_2^2 > \tilde{F}(z^k) \end{aligned}$$

Using Theorem 2.5, we have:

$$F_\rho(\beta^k) \leq F(\beta^k) - \rho \|\beta^k - \beta^*\|_2^2 \varphi\left(\frac{F(\beta^k) - F(\beta^*)}{\rho \|\beta^k - \beta^*\|_2^2}\right).$$

Case 1: $\alpha\rho\|\beta^k - \beta^*\|^2 \leq F(\beta^k) - F(\beta^*) \leq \rho\|\beta^k - \beta^*\|_2^2$,

$$F_\rho(\beta^k) \leq F(\beta^k) - \frac{(F(\beta^k) - F(\beta^*))^2}{\rho\|\beta^k - \beta^*\|_2^2}.$$

Combine the above inequalities we have:

$$\alpha(F(\beta^k) - F(\beta^*)) \leq F(\beta^k) - \tilde{F}(z^k).$$

Case 2: $F(\beta^k) - F(\beta^*) \geq \rho\|\beta^k - \beta^*\|_2^2$.

Similarly we obtain a simple relation:

$$\begin{aligned} F_\rho(\beta^k) &\leq F(\beta^k) - 2(F(\beta^k) - F(\beta^*)) + \rho\|\beta^k - \beta^*\|_2^2. \\ 2(F(\beta^k) - F(\beta^*)) - \rho\|\beta^k - \beta^*\|_2^2 &\leq F(\beta^k) - \tilde{F}(z^k) \\ F(\beta^k) - F(\beta^*) &\leq F(\beta^k) - \tilde{F}(z^k). \end{aligned}$$

When a *descent step* happens for Case 1:

$$\begin{aligned} F(z^k) &\leq (1 - \gamma)F(\beta^k) + \gamma\tilde{F}(z^k) \\ &\leq (1 - \gamma)F(\beta^k) + \gamma((1 - \alpha)F(\beta^k) + \alpha F(\beta^*)) \\ &\leq (1 - \gamma\alpha)(F(\beta^k) - F(\beta^*)) + F(\beta^*). \end{aligned}$$

Same argument applied for Case 2, we obtained linear convergence rate for *descent steps*:

$$F(z^k) - F(\beta^*) \leq (1 - \min(\alpha, 1)\gamma)(F(\beta^k) - F(\beta^*)).$$

□

A.6 Proof for Lemma 3.7

Proof.

$$\begin{aligned}
\eta^k &= \min_{\beta} l(\beta) + \sum_{i=1}^N f_i(z_i^k) + \langle g_i^k, \beta - z_i^k \rangle + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&\geq \min_{\beta} l(\beta^k) + \langle g^k, \beta - \beta^k \rangle + \sum_{i=1}^N f_i(z_i^k) + \langle g_i^k, \beta - z_i^k \rangle + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&= \min_{\beta} l(\beta^k) + \langle g^k, \beta - \beta^k \rangle + \sum_{i=1}^{N-1} f_i(z_i^k) + \langle g_i^k, \beta - z_i^k \rangle + f_N(\beta) + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&\geq \min_{\beta} l(\beta^k) + \langle g^k, \beta \rangle + \sum_{i=1}^{N-1} f_i(z_i^k) + \langle g_i^k, \beta \rangle + f_N(z_i^{k-1}) + \langle g_i^{k-1}, \beta - z_i^{k-1} \rangle + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&= \dots \\
&\geq \min_{\beta} l(\beta^k) + \langle g^k, \beta - \beta^k \rangle + \sum_{i=1}^N f_i(z_i^{k-1}) + \langle g_i^{k-1}, \beta - z_i^{k-1} \rangle + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&= \eta^{k-1} - \langle g^k, \beta^k \rangle - \sum_{i=1}^N \langle g_i^{k-1}, \beta^k \rangle + \langle g^k, \beta \rangle + \sum_{i=1}^N \langle g_i^{k-1}, \beta \rangle - \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
&= \eta^{k-1} + \min_{\beta} \langle g^k + \sum_{i=1}^N g_i^{k-1}, \beta - \beta^k \rangle - \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 + \frac{\rho}{2} \|\beta - \beta^k\|_2^2
\end{aligned}$$

From the optimality condition of the first sub-problem, we have $g^k + \sum_{i=1}^N g_i^{k-1} = -\rho(\beta^k - \beta^{k-1})$. Replace this relation into the above inequality to obtain:

$$\begin{aligned}
\eta^k &\geq \eta^{k-1} + \min_{\beta} \langle -\rho(\beta^k - \beta^{k-1}), \beta - \beta^k \rangle - \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 \\
\langle -\rho(\beta^k - \beta^{k-1}), \beta - \beta^k \rangle + \frac{\rho}{2} \|\beta - \beta^k\|_2^2 &\text{ minimizes at } \beta = 2\beta^k - \beta^{k-1} \text{ with value } \\
&-\frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2. \text{ Thus we have:}
\end{aligned}$$

$$\eta^k \geq \eta^{k-1} - \rho \|\beta^k - \beta^{k-1}\|_2^2$$

By the update step rule:

$$\begin{aligned}
F(\beta^k) &\leq F(\beta^{k-1}) - \gamma(F(\beta^{k-1}) - \eta^k) + \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 \\
\frac{1}{\gamma}(F(\beta^{k-1}) - F(\beta^k)) &\geq F(\beta^{k-1}) - \eta^k + \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 \\
\frac{1}{\gamma}(F(\beta^{k-1}) - F(\beta^k)) &\geq \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 + \frac{\rho}{2} \|\beta^k - \beta^{k-1}\|_2^2 = \rho \|\beta^k - \beta^{k-1}\|_2^2
\end{aligned}$$

The last inequality is due to Theorem 2.2. \square

A.7 Proof for Lemma 3.8

Proof. From Lemma 3.3, we have:

$$\begin{aligned}
F(\beta^k) - \eta^k &\leq \frac{1}{\gamma}(F(\beta^{k-1}) - F(\beta^k)) - \frac{\rho}{2}\|\beta^k - \beta^{k-1}\|_2^2 \\
&\leq \frac{1}{\gamma}(F(\beta^{k-1}) - F(\beta^k)) + \frac{1}{2\gamma}(F(\beta^{k-1}) - F(\beta^k)) \\
&= \frac{3}{2\gamma}(F(\beta^{k-1}) - F(\beta^k))
\end{aligned}$$

□

A.8 Proof for Lemma 3.9

Proof. By Theorem 3.1:

$$F_\rho(\beta^k) \leq F(\beta^k) - \rho\|\beta^k - \beta^*\|_2^2 \varphi\left(\frac{F(\beta^k) - F(\beta^*)}{\rho\|\beta^k - \beta^*\|_2^2}\right) \leq F(\beta^k) - \rho\|\beta^k - \beta^*\|_2^2 \varphi(\alpha)$$

Also by definition of $F_\rho(\beta^k)$:

$$F(\beta^*) + \frac{\rho}{2}\|\beta^* - \beta^k\|_2^2 \geq F_\rho(\beta^k)$$

So:

$$\begin{aligned}
F_\rho(\beta^k) &\leq F(\beta^k) - 2(F_\rho(\beta^k) - F(\beta^*))\varphi(\alpha) \\
\eta^k &\leq F_\rho(\beta^k) \leq \frac{F(\beta^k) + 2F(\beta^*)\varphi(\alpha)}{2\varphi(\alpha) + 1}
\end{aligned}$$

□

B Plots and graphs

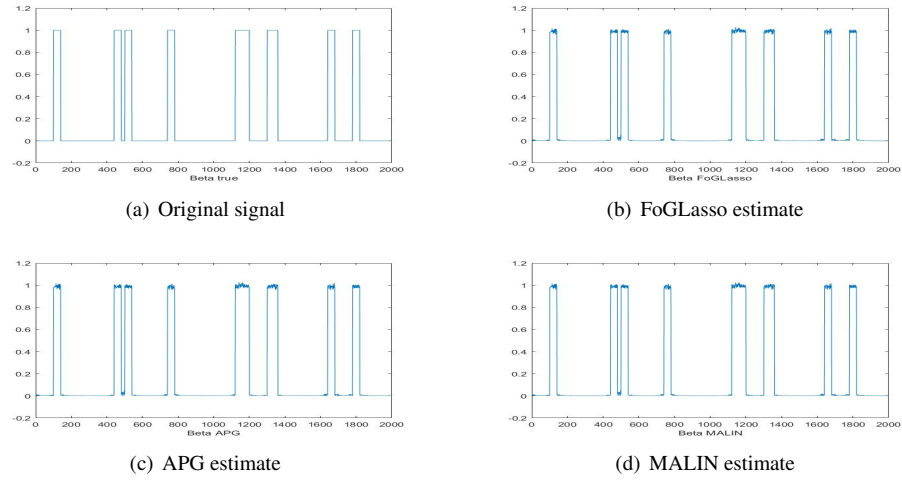


Figure 1: Comparison of solutions' quality.

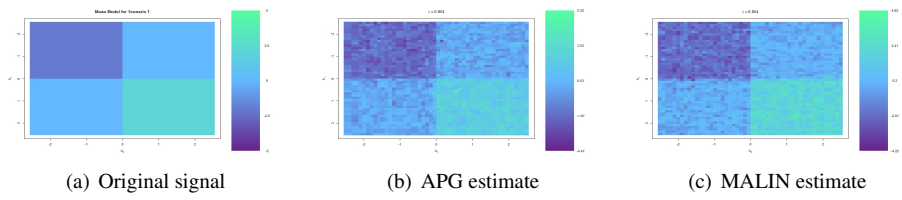


Figure 2: Scenario 1

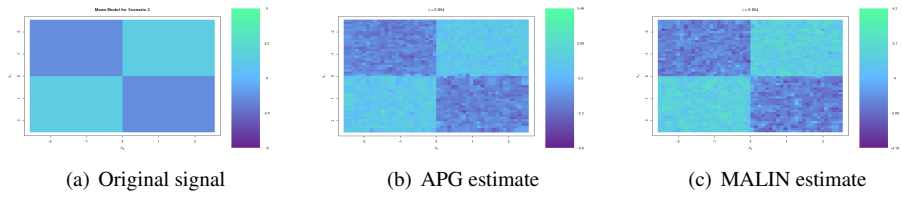


Figure 3: Scenario 2

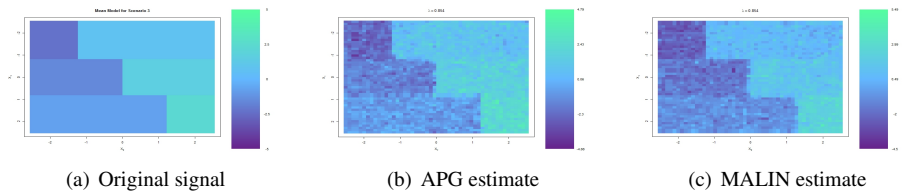


Figure 4: Scenario 3

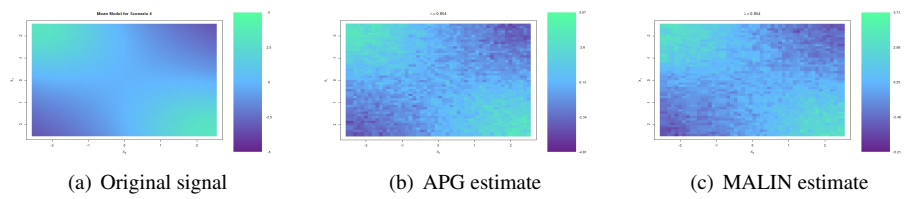
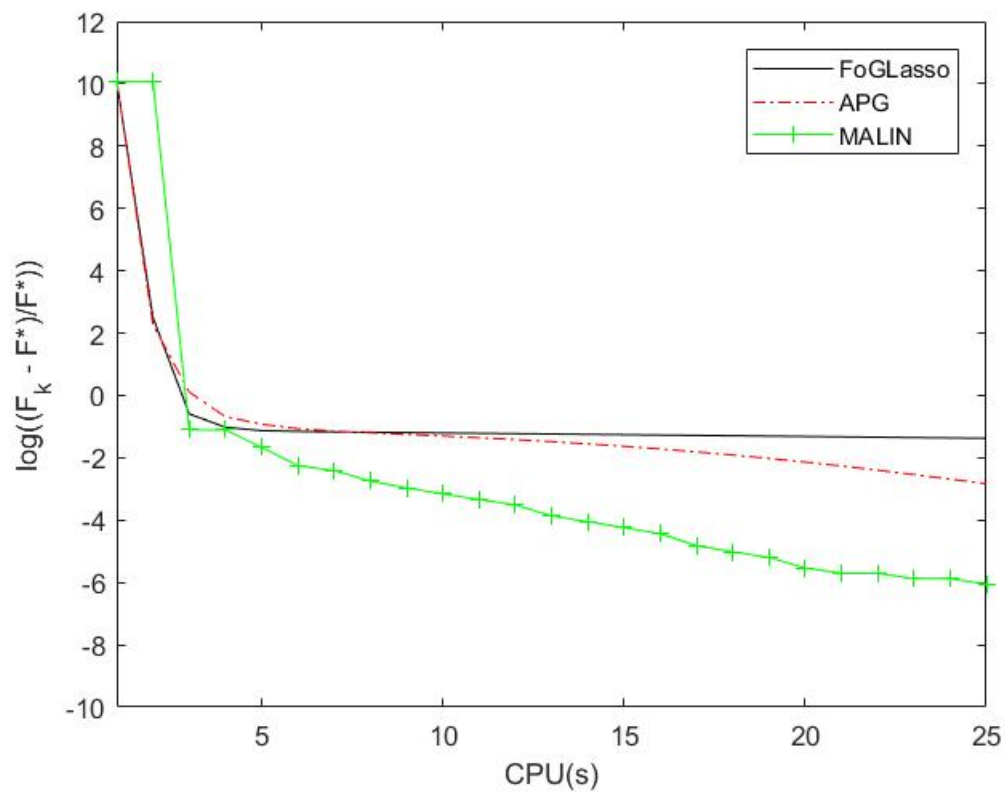
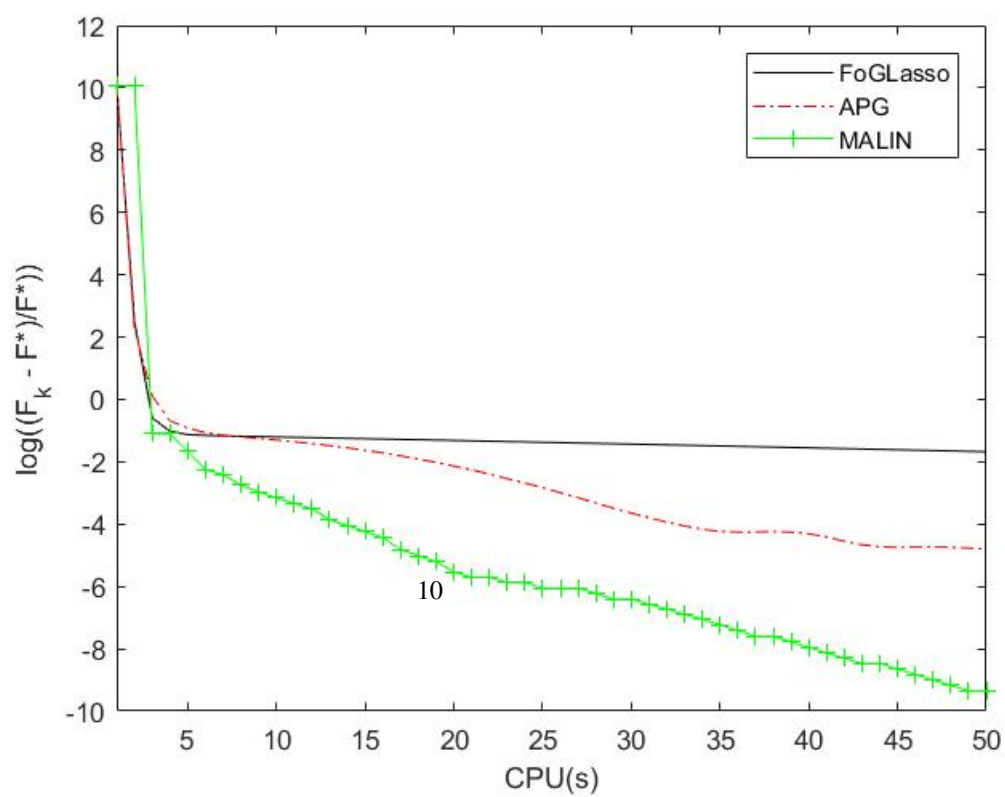


Figure 5: Scenario 4



(a) $n = 1000, p = 2000$



(b) $n = 1000, p = 5000$