A PROOF OF THEOREM 3

Define $S'_{t+1} = S_{t+1} \setminus S_*$. In order to prove Theorem 3, it is equivalent to show that there are no more than s entries in $[\widehat{\mathbf{x}}_{t+1}]_{S'_{t+1}}$ with the magnitude larger than τ_t. To this end we define $V_1 = S_t \cup S_*$, $V_2 = V_1 \setminus S_*$. For any subset $S' \subseteq S_*^c$ whose size is small or equal than s, define $S'_1 = S' \cap V_2$, $S'_2 = S' \setminus V_2$. Based on the assumption given in the theorem, we have $|V_1| \le 2s$, $|V_2| \le s$, and $|S'_2| \le s$.

We have

$$\begin{aligned}[\widehat{\mathbf{x}}_{t+1}]_{S'} &= [\mathbf{x}_t]_{S'} - \frac{1}{n}\left(A_{S'}^\top A\mathbf{x}_t - A_{S'}^\top \mathbf{y}\right) = [\mathbf{x}_t]_{S'} - \frac{1}{n}A_{S'}^\top A(\mathbf{x}_t - \mathbf{x}_*) + \frac{1}{n}A_{S'}^\top \mathbf{z} \\ &= [\mathbf{x}_t]_{S'} - \frac{1}{n}A_{S'}^\top A_{S_*}[\mathbf{x}_t - \mathbf{x}_*]_{S_*} - \frac{1}{n}A_{S'}A_{V_2}[\mathbf{x}_t]_{V_2} + \frac{1}{n}A^\top \mathbf{z}\end{aligned}$$

Hence, we have

$$\|[\widehat{\mathbf{x}}_{t+1}]_{S'}\|_2 \le \underbrace{\frac{1}{n}\left\|A_{S'}^\top A_{S_*}[\mathbf{x}_t - \mathbf{x}_*]_{S_*}\right\|_2}_{:=E_1} + \underbrace{\left\|\left(I - \frac{1}{n}A_{V_2}^\top A_{V_2}\right)[\mathbf{x}_t]_{V_2}\right\|_2}_{:=E_2} + \underbrace{\frac{1}{n}\left\|A_{S'_2}A_{V_2}[\mathbf{x}_t]_{V_2}\right\|_2}_{:=E_3} + \underbrace{\frac{1}{n}\|A_{S'}^\top \mathbf{z}\|_2}_{:=E_4}$$

Below, we will bound E_1, E_2, E_3, and E_4, separately. To bound E_1, we use the fact $S' \cap S_* = \emptyset$, $|S_*| \le s$, and $|S'| \le s$, and have

$$E_1 = \max_{\|\mathbf{u}\|_2 \le 1} \frac{1}{n}\mathbf{u}_{S'}^\top A_{S'}^\top A_{S_*}[\mathbf{x}_t - \mathbf{x}_*]_{S_*} \le \max_{\|\mathbf{u}\|_2 \le 1} \frac{1}{n}\mathbf{u}_{S'}^\top A_{S'}^\top A_{S_*}[\mathbf{x}_t - \mathbf{x}_*]_{S_*} \le \delta\|[\mathbf{x}_t - \mathbf{x}_*]_{S_*}\|_2$$

The last inequlaity comes from R.I.P. condition. To bound E_2, we use the fact $|V_2| \le s$ and have

$$E_2 \le \delta\|[\mathbf{x}_t]_{V_2}\|_2$$

To bound E_3, we use the fact $|S'_2| \le s$ and $|V_2| \le s$, and have

$$E_3 \le \delta\|[\mathbf{x}_t]_{V_2}\|_2$$

To bound E_4, we use the fact $\|A^\top \mathbf{z}\|_\infty \le 2\sigma\sqrt{n\log d}$ and therefore

$$E_4 \le 2\sigma\sqrt{\frac{s\log d}{n}}$$

Combining the bounds for E_1, E_2, E_3, and E_4, we have, for any $S' \subset S_*^c$ with $|S'| \le s$,

$$\|[\widehat{\mathbf{x}}_{t+1}]_{S'}\|_2 \le 2\delta\|[\mathbf{x}_* - \mathbf{x}_t]_{V_2}\|_2 + \delta\|[\mathbf{x}_* - \mathbf{x}_t]_{S_*}\|_2 + 2\sigma\sqrt{\frac{s\log d}{n}} \le 3\delta\|[\mathbf{x}_* - \mathbf{x}_t]\|_2 + 2\sigma\sqrt{\frac{s\log d}{n}}$$

which leads to the fact that no more than s entries in $[\widehat{\mathbf{x}}_{t+1}]_{S'_{t+1}}$ is larger than

$$\frac{3\delta}{\sqrt{s}}\|\mathbf{x}_* - \mathbf{x}_t\|_2 + 2\sigma\sqrt{\frac{\log d}{n}}$$

B PROOF OF THEOREM 4

First, according to Theorem 3, we have $\mathbf{x}_{t+1}$ is $2s$ sparse with $|S_{t+1} \setminus S_*| \le s$.

Since

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_{t+1}\|_2^2 + \Omega_t(\mathbf{x}),$$

we have

$$
\begin{aligned}
0 &\geq \frac{1}{2}\|\mathbf{x}_{t+1} - \widehat{\mathbf{x}}_{t+1}\|_2^2 + \Omega_t(\mathbf{x}_{t+1}) - \frac{1}{2}\|\mathbf{x}_* - \widehat{\mathbf{x}}_{t+1}\|_2^2 - \Omega_t(\mathbf{x}_*) \\
&= \frac{1}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 + (\mathbf{x}_{t+1} - \mathbf{x}_*)^\top(\mathbf{x}_* - \widehat{\mathbf{x}}_{t+1}) + \Omega_t(\mathbf{x}_{t+1}) - \Omega_t(\mathbf{x}_*) \\
&= \frac{1}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 - (\mathbf{x}_{t+1} - \mathbf{x}_*)^\top(\mathbf{x}_t - \mathbf{x}_*) + \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_*)^\top A^\top(A\mathbf{x}_t - \mathbf{y}) + \Omega_t(\mathbf{x}_{t+1}) - \Omega_t(\mathbf{x}_*) \qquad (14) \\
&= \frac{1}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 - (\mathbf{x}_{t+1} - \mathbf{x}_*)^\top(\mathbf{x}_t - \mathbf{x}_*) + \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_*)^\top A^\top A(\mathbf{x}_t - \mathbf{x}_*) \\
&+ \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_*)^\top A^\top \mathbf{z} + \Omega_t(\mathbf{x}_{t+1}) - \Omega_t(\mathbf{x}_*)
\end{aligned}
$$

Using the fact that $\Omega_t(\mathbf{x})$ is concave in $|\mathbf{x}|$, we have

$$
\Omega_t(\mathbf{x}_{t+1}) \leq \Omega(\mathbf{x}_*) + \tau_t \sum_{i=1}^{d}(|\mathbf{x}_t|_i - |\mathbf{x}_*|_i) \leq \Omega(\mathbf{x}_*) + \tau_t|\mathbf{x}_t - \mathbf{x}_i|_1
$$

Using the fact

$$
\left\|\frac{1}{n}A^\top \mathbf{z}\right\|_\infty \leq 2\sigma\sqrt{\frac{\log d}{n}} \ ,
$$

from inequality (14), we have

$$
\frac{1}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|^2 \leq (\mathbf{x}_{t+1} - \mathbf{x}_*)^\top(\mathbf{x}_t - \mathbf{x}_*) - \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_*)^\top A^\top A(\mathbf{x}_t - \mathbf{x}_*) + \left(2\sigma\sqrt{\frac{\log d}{n}} + \tau_t\right)|\mathbf{x}_{t+1} - \mathbf{x}_*|_1 \ . \qquad (15)
$$

For the first two terms on the right hand side of Eq. (15), we have,

$$
\begin{aligned}
& (\mathbf{x}_{t+1} - \mathbf{x}_*)^\top(\mathbf{x}_t - \mathbf{x}_*) - \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_*)^\top A^\top A(\mathbf{x}_t - \mathbf{x}_*) \\
\leq \quad & \frac{\delta}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 + \frac{\delta}{2}\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 + \frac{\delta}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \leq \delta\left(\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 + \|\mathbf{x}_t - \mathbf{x}_*\|_2^2\right) \ .
\end{aligned}
$$

Substitute back into Eq. (15),

$$
\begin{aligned}
(1 - 2\delta)\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2^2 \quad &\leq \quad 2\delta\|\mathbf{x}_t - \mathbf{x}_*\|_2^2 + \left(2\sigma\sqrt{\frac{\log d}{n}} + \tau_t\right)|\mathbf{x}_{t+1} - \mathbf{x}_*|_1 \\
&\leq \quad 2\delta\Delta_t^2 + \sqrt{s}\left(2\sigma\sqrt{\frac{\log d}{n}} + \tau_t\right)\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2.
\end{aligned}
$$

Recall that

$$
x^2 \leq ax + b \Rightarrow x \leq \max(2a, \sqrt{2b}) \ ,
$$

using $\|\mathbf{x}_t - \mathbf{x}_*\|_2 \leq \Delta_t$, the above quadratic inequality gives us

$$
\|\mathbf{x}_{t+1} - \mathbf{x}_*\|_2 \leq 4\sigma\sqrt{\frac{s\log d}{n}} + \max\left(2\sqrt{\frac{\delta}{1 - 2\delta}}, 3\delta\right)\Delta_t
$$

C PROOF OF THEOREM 2

According to Theorem 4, with $t \geq T_0$, we have

$$
\|\mathbf{x}_t - \mathbf{x}_*\|_2 \leq \frac{5\sigma}{1 - q}\sqrt{\frac{s\log d}{n}} \qquad (16)
$$

$$
2\sigma\sqrt{\frac{\log d}{n}} \leq \tau_t \leq 3\sigma\sqrt{\frac{\log d}{n}} \qquad (17)
$$

We first show the selection consistency for $t > t_0$. Following the analysis of Theorem 4, we have

$$\begin{aligned}(1-2\delta)\|\mathbf{x}_{t+1}-\mathbf{x}_*\|_2^2 &\leq 2\delta\|\mathbf{x}_t-\mathbf{x}_*\|_2^2 + 2\sigma\sqrt{\frac{\log d}{n}}\|\mathbf{x}_{t+1}-\mathbf{x}_*\|_2 + \Omega(\mathbf{x}_*) - \Omega(\mathbf{x}_{t+1}) \\ &\overset{[1]}{\leq} 2\delta\|\mathbf{x}_t-\mathbf{x}_*\|_2^2 + \frac{2\sigma^2}{1-\delta}\frac{\log d}{n} + \frac{1-\delta}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_*\|_2^2 + \Omega(\mathbf{x}_*) - \Omega(\mathbf{x}_{t+1}) \ .\end{aligned}$$

Inequality [1] is because $ab \leq \frac{1}{2(1-\delta)}a^2 + \frac{1-\delta}{2}b^2$. Since

$$\lambda_{\min}(\mathbf{x}_*) \geq \frac{4\sigma}{1-\delta}\sqrt{\frac{2\log d}{n}} \geq \tau_t$$

we have

$$\Omega(\mathbf{x}_*) = \frac{\tau_t^2}{2}|S_*|$$

Based on the updating equation for $\mathbf{x}_{t+1}$, we have $\Omega_t(\mathbf{x}_{t+1}) = \frac{\tau_t^2}{2}|S_{t+1}|$. We thus have

$$\frac{1-3\delta}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_*\|_2^2 \leq 2\delta\|\mathbf{x}_t-\mathbf{x}_*\|_2^2 + \frac{2\sigma^2\log d}{(1-\delta)n} + \frac{\tau_t^2}{2}\left(|S_*| - |S_{t+1}|\right)$$

Define $A = |S_{t+1} \setminus S_*|$ and $B = |S_* \setminus S_{t+1}|$. We have

$$\frac{1-3\delta}{2}\left(\tau_t^2 A + \lambda_{\min}^2(\mathbf{x}_*)B\right) \leq \frac{50\delta s\log d\sigma^2}{(1-q)^2 n} + \frac{2\sigma^2\log d}{(1-\delta)n} + \frac{\tau_t^2}{2}(B-A) \ .$$

The first term on the right hand side is based on Eq. (16).

From Eq.(17) and the assumption of lower bound of $\lambda_{\min}(\mathbf{x}_*)$, it is easy to check that

$$\frac{1-3\delta}{2}\lambda_{\min}^2(\mathbf{x}_*) \geq \tau_t^2 \ .$$

We have

$$A + B \leq \frac{1}{2}\left(\frac{50\delta s}{(1-q)^2} + \frac{1}{1-\delta}\right) < 1 \ .$$

Since both A and B are integers, we have $A = B = 0$, which implies $S_{t+1} = S_*$

We then proceed to show $\|\mathbf{x}_t - \mathbf{x}_o\|_2$ will converge to zero geometrically. Following the analysis of Theorem 4, we have

$$0 \geq \frac{1}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_o\|_2^2 - (\mathbf{x}_{t+1}-\mathbf{x}_o)^\top(\mathbf{x}_t-\mathbf{x}_o) + \frac{1}{n}(\mathbf{x}_{t+1}-\mathbf{x}_o)^\top A^\top(A\mathbf{x}_t - \mathbf{y}) + \Omega_t(\mathbf{x}_{t+1}) - \Omega_t(\mathbf{x}_o)$$

Since for any $i \in S_*$,

$$|\mathbf{x}_o|_i \geq \lambda_{\min}(\mathbf{x}_*) - \frac{\sigma}{1-\delta}\sqrt{\frac{\log d}{n}} \geq 3\sigma\sqrt{\frac{\log d}{n}} \geq \tau_t$$

we have $\Omega_t(\mathbf{x}_o) = \tau_t^2|S_*|/2 = \tau_t^2/2|S_{t+1}| = \Omega_t(\mathbf{x}_{t+1})$, and therefore

$$0 \geq \frac{1}{2}\|\mathbf{x}_{t+1}-\mathbf{x}_o\|_2^2 - (\mathbf{x}_{t+1}-\mathbf{x}_o)^\top(\mathbf{x}_t-\mathbf{x}_o) + \frac{1}{n}(\mathbf{x}_{t+1}-\mathbf{x}_o)^\top A^\top(A\mathbf{x}_t - \mathbf{y})$$

Using the fact $S_{t+1} = S_*$, we have

$$\begin{aligned}&\frac{1}{n}(\mathbf{x}_{t+1}-\mathbf{x}_o)^\top A^\top(A\mathbf{x}_t-\mathbf{y}) \\ =\ &\frac{1}{n}([\mathbf{x}_{t+1}-\mathbf{x}_o]_{S_*})^\top A_{S_*}^\top(A_{S_*}[\mathbf{x}_t]_{S_*} - \mathbf{y}) \\ =\ &\frac{1}{n}([\mathbf{x}_{t+1}-\mathbf{x}_o]_{S_*})^\top A_{S_*}^\top A_{S_*}[\mathbf{x}_t-\mathbf{x}_o]_{S_*} + \frac{1}{n}([\mathbf{x}_{t+1}-\mathbf{x}_o]_{S_*})^\top A_{S_*}^\top(A_{S_*}\mathbf{x}_o - \mathbf{y})\end{aligned}$$

Using the fact $A_{S_*}\mathbf{x}_o - \mathbf{y}$ is orthogonal to the subspace spanned by the column vectors in A_{S_*}, we have

$$\frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_o)^\top A^\top (A\mathbf{x}_t - \mathbf{y}) = \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_o)^\top A^\top A(\mathbf{x}_t - \mathbf{x}_o)$$

Hence, we have

$$\begin{aligned} 0 &\geq \frac{1}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_o\|_2^2 - (\mathbf{x}_{t+1} - \mathbf{x}_o)^\top (\mathbf{x}_t - \mathbf{x}_o) + \frac{1}{n}(\mathbf{x}_{t+1} - \mathbf{x}_o)^\top A^\top A(\mathbf{x}_t - \mathbf{x}_o) \\ &\geq \frac{1 - 3\delta}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_o\|_2^2 - \frac{3\delta}{2}\|\mathbf{x}_t - \mathbf{x}_o\|_2^2 \end{aligned}$$

which proves the theorem.

D MORE EXPERIMENTS

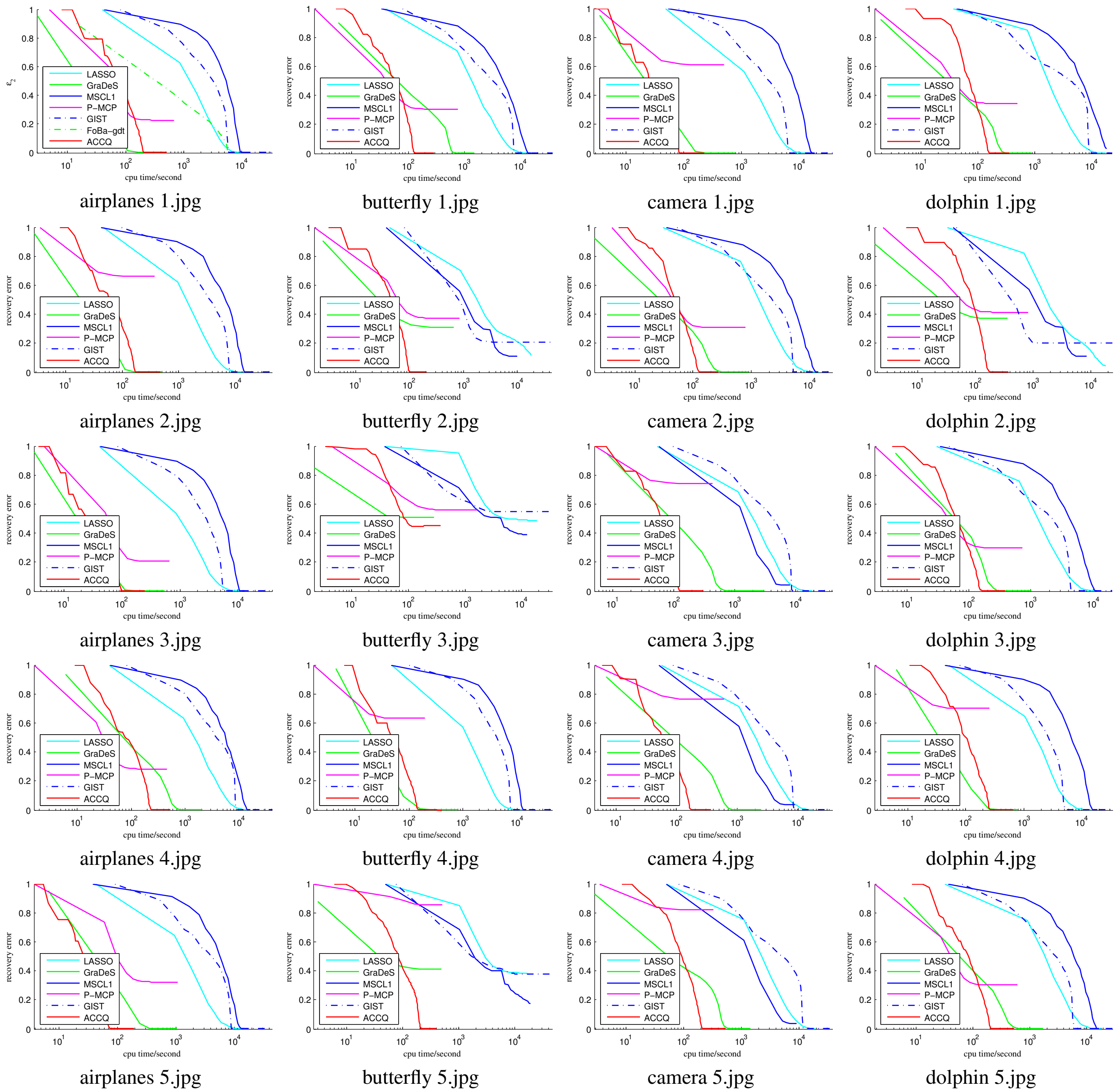


Figure 3: ℓ_2–Norm Recovery Error For All Datasets, $\sigma = 0$

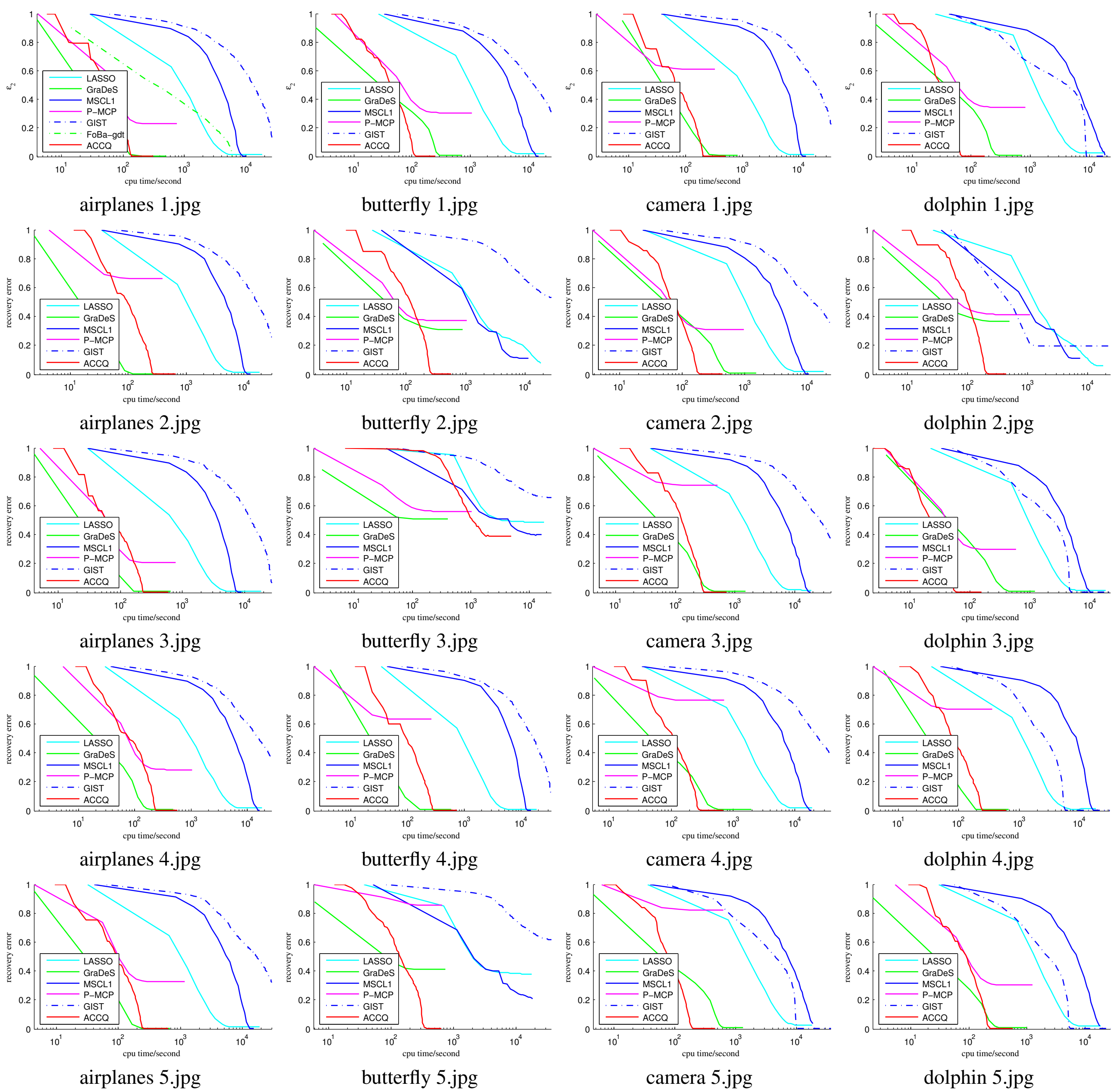
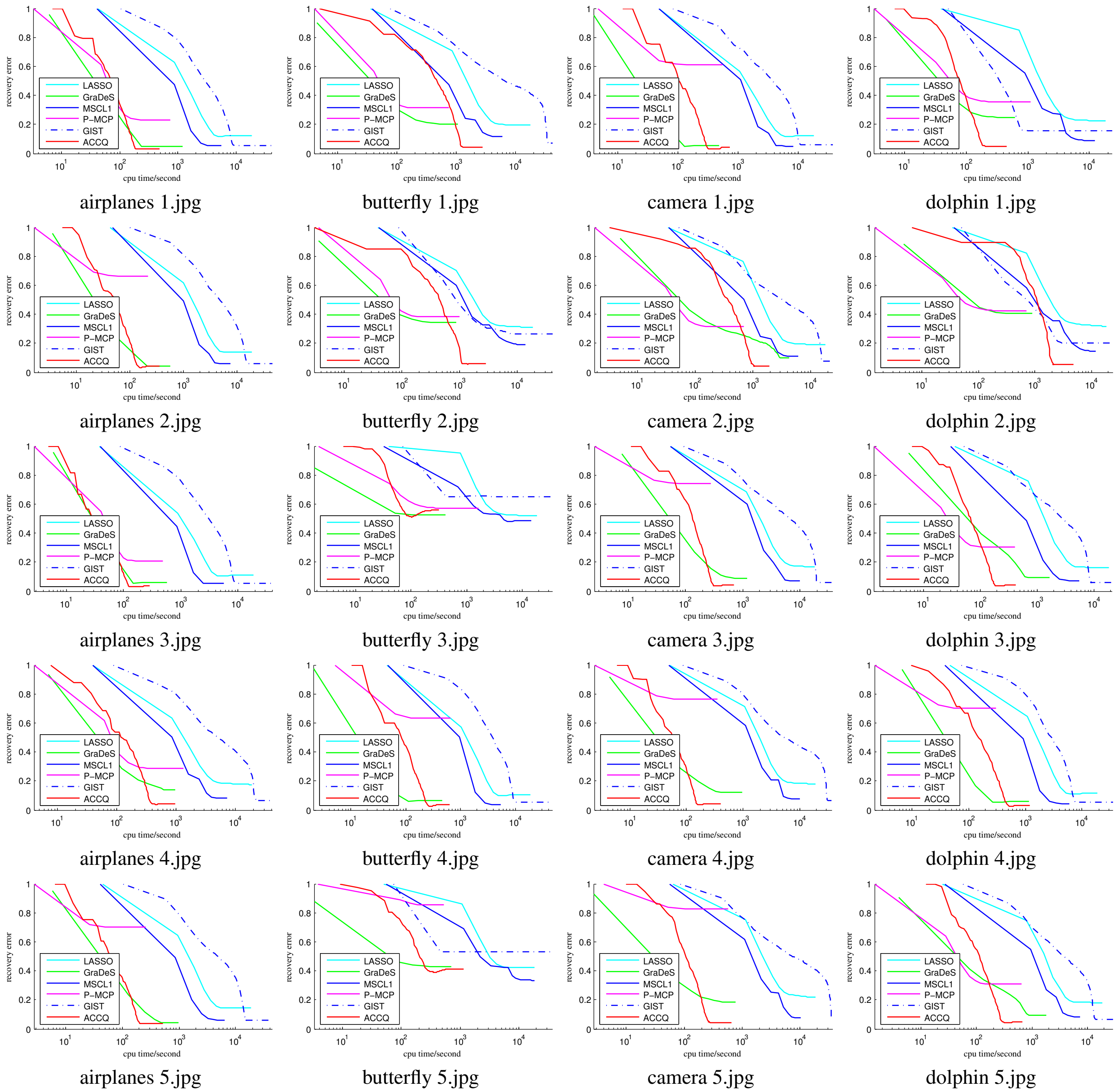


Figure 4: ℓ_2–Norm Recovery Error For All Datasets, $\sigma = 0.01$



airplanes 1.jpg | butterfly 1.jpg | camera 1.jpg | dolphin 1.jpg

airplanes 2.jpg | butterfly 2.jpg | camera 2.jpg | dolphin 2.jpg

airplanes 3.jpg | butterfly 3.jpg | camera 3.jpg | dolphin 3.jpg

airplanes 4.jpg | butterfly 4.jpg | camera 4.jpg | dolphin 4.jpg

airplanes 5.jpg | butterfly 5.jpg | camera 5.jpg | dolphin 5.jpg

Figure 5: ℓ_2–Norm Recovery Error For All Datasets, $\sigma = 0.1$