Robust Principal Component Analysis for Generalized Multi-View Models (Supplementary Material)

Frank Nussbaum[1,2] **Joachim Giesen**[1]

[1]Friedrich-Schiller-Universität, Jena, Germany
[2]DLR Institute of Data Science, Jena, Germany

1 PROOFS OF LEMMAS

1.1 PROOF OF LOCAL IDENTIFIABILITY

Proof of Lemma 1.. Let $r = \text{rank}(L^\star)$ and $s = |\text{gsupp}(S^\star)|$. For local identifiability we must find a (small) ball such that for all $\Delta \neq 0$ from this ball it holds that $(L^\star - \Delta, S^\star + \Delta) \notin \mathcal{L}(r) \times \mathcal{S}(s)$. We already motivated in the main paper that the points that are close to $S^\star \in \mathcal{S}(s)$ and $L^\star \in \mathcal{L}(r)$ in the varieties can be characterized using tangent spaces. Specifically, it can only hold $S^\star + \Delta \in \mathcal{S}(s)$ for small $\Delta \neq 0$ *if* $\Delta \in \mathcal{Q}(S^\star)$. Likewise, it can only hold $L^\star - \Delta \in \mathcal{L}(r)$ for small $\Delta \neq 0$ *if* $\Delta \in \mathcal{T}(L')$ for some tangent space $\mathcal{T}(L')$ to $\mathcal{L}(r)$ at a (smooth) point $L' \in \mathcal{L}(r)$ that is close to $L^\star$. Note again that due to the local curvature of the low-rank matrix variety, we also need to consider nearby tangent spaces. Hence, to prove local identifiability it is sufficient to show that the tangent spaces $\mathcal{Q}(S^\star)$ and $\mathcal{T}(L')$ are transverse for all smooth $L' \in \mathcal{L}(r)$ from some small ball around L.

By definition, transversality of $\mathcal{Q}(S^\star)$ and $\mathcal{T}(L')$ means that $T(L') \cap \mathcal{Q}(S^\star) = \{0\}$, which is equivalent to

$$\min_{M \in \mathcal{Q}(S^\star), \|M\|=1} \|M - P_{\mathcal{T}(L')}M\| > 0. \qquad (1)$$

This is because $M = P_{\mathcal{T}(L')}M$ if and only if $M \in T(L') \cap \mathcal{Q}(S^\star)$. In the following, we want to verify Condition (1) for smooth $L' \in \mathcal{L}(r)$ from a small ball around $L^\star$. We start by calculating that for any $M \in \mathcal{Q}(S^\star)$ with $\|M\| = 1$ it holds that

$$\begin{aligned}
&\|M - P_{\mathcal{T}(L')}M\| \\
&\quad = \|M - P_{\mathcal{T}(L^\star)}M + \left[P_{\mathcal{T}(L^\star)} - P_{\mathcal{T}(L')}\right] M\| \\
&\quad \geqslant \|M - P_{\mathcal{T}(L^\star)}M\| - \| \left[P_{\mathcal{T}(L^\star)} - P_{\mathcal{T}(L')}\right] M\| \\
&\quad \geqslant \kappa - \rho(\mathcal{T}(L^\star), \mathcal{T}(L')),
\end{aligned}$$

where the first inequality is the triangle inequality, and for the second inequality we defined

$$\kappa = \min_{M \in \mathcal{Q}(S^\star), \|M\|=1} \|M - P_{\mathcal{T}(L^\star)}M\|$$

and the twisting between subspaces

$$\rho(\mathcal{T}(L^\star), \mathcal{T}(L')) = \max_{\|M\|=1} \left\| \left[P_{\mathcal{T}(L^\star)} - P_{\mathcal{T}(L')}\right] M \right\|.$$

Here, the assumed transversality of the tangent spaces $\mathcal{Q}(S^\star)$ and $\mathcal{T}(L^\star)$ implies that $\kappa > 0$. Hence, a sufficient condition for the transversality of $\mathcal{Q}(S^\star)$ and $\mathcal{T}(L')$ is that $\rho(\mathcal{T}(L^\star), \mathcal{T}(L')) < \kappa$ since then Condition (1) is satisfied. We show that $\rho(\mathcal{T}(L^\star), \mathcal{T}(L')) < \kappa$ holds whenever L' is sufficiently close to $L^\star$. This proof is technical, but the main idea is to show that the map from smooth $L' \in \mathcal{L}(r)$ to $\rho(\mathcal{T}(L^\star), \mathcal{T}(L'))$ is continuous and since it maps $L^\star$ onto zero, there exists a small ball around $L^\star$ for which $\rho(\mathcal{T}(L^\star), \mathcal{T}(L')) < \kappa$.

We now dive into the technical details. For that, we consider the function f that maps (L', M) with domain restricted to $L' \in \mathcal{L}(r)$, $\|L^\star - L'\| \leqslant 1$, and $\|M\| = 1$ onto $\mathbb{R}$ as follows

$$\begin{aligned}
f(L', M) = \|P_{\mathcal{T}(L^\star)}M - (&P_{U(L')}M \\
&+ MP_{V(L')} - P_{U(L')}MP_{V(L')})\|,
\end{aligned}$$

where $P_{U(L')}$ is the projection matrix that projects onto the column space $U(L')$ of L', and $P_{V(L')}$ is the projection matrix that projects onto the row space $V(L')$ of L'. Note that for a rank-r matrix L', that is, $L' \in \mathcal{L}(r)$ is smooth, it holds that

$$P_{\mathcal{T}(L')}M = P_{U(L')}M + MP_{V(L')} - P_{U(L')}MP_{V(L')},$$

see, for example, Candès et al. [2011]. Consequently, for smooth L' we have $f(L', M) = \|[P_{\mathcal{T}(L^\star)} - P_{\mathcal{T}(L')}]M\|$ and in particular $f(L^\star, M) = 0$ for all M. We now argue that f is continuous as a composition of continuous functions: First, L' maps continuously onto the projection matrices $P_{U(L')}, P_{V(L')}$ because small changes to L' only cause small changes to the row and column spaces of L' and hence to the corresponding projections. Second, the remaining composite functions in the definition of f above are additions, norm functions or matrix products of $P_{U(L')}, P_{V(L')}$, and M. All these operations are continuous, thus overall f is continuous.

Because f is continuous on a compact domain, it is also uniformly continuous. Hence, there exists $\delta > 0$ (w.l.o.g. $\delta \leqslant 1$) such that for all L_1', L_2' with $\|L_1' - L_2'\| < \delta$ and for all M_1, M_2 with $\|M_1 - M_2\| < \delta$ it holds that $|f(L_1', M_1) - f(L_2', M_2)| < \frac{\kappa}{2}$. Consequently, it holds for L' with $\|L^\star - L'\| < \delta$ independently of M that

$$f(L', M) < f(L^\star, M) + \frac{\kappa}{2} = \frac{\kappa}{2}.$$

We can take the supremum over M with $\|M\| = 1$ on the left-hand side of this equation. If we only consider smooth L', this implies that

$$\begin{aligned}\rho(\mathcal{T}(L^\star), \mathcal{T}(L')) &= \sup_{M:\|M\|=1} \|[P_{\mathcal{T}(L^\star)} - P_{\mathcal{T}(L')}]M\| \\ &= \sup_{M:\|M\|=1} f(L', M) \leqslant \frac{\kappa}{2} < \kappa.\end{aligned}$$

This completes the proof because we have shown that for all smooth L' from the spectral-norm ball with radius δ around $L^\star$ the tangent spaces $\mathcal{Q}(S^\star)$ and $\mathcal{T}(L')$ are transverse. Particularly, there do not exist small non-zero $\Delta \in \mathcal{T}(L') \cap \mathcal{Q}(S^\star)$ for any L' from that ball, hence locally around $(L^\star, S^\star)$ there are no alternative decompositions $(L^\star - \Delta, S^\star + \Delta) \in \mathcal{L}(r) \times \mathcal{S}(s)$. This establishes local identifiability of $(L^\star, S^\star)$. □

1.2 PROJECTIONS ON TANGENT AND NORMAL SPACES

Throughout the remaining proofs, we frequently need the following lemma that bounds the norms of projections onto tangent and normal spaces. Here, the $\ell_{\infty,2}$-norm on $\mathbb{R}^{m\times n}$, given by $\|A\|_{\infty,2} = \max_{i\in[d],j\in[n]} \|a_{ij}\|_2$, is the dual norm of the $\ell_{1,2}$-norm.

Lemma 5. *Let $\mathcal{Q}(S)$ be the tangent space at $S \in \mathcal{S}(|\operatorname{gsupp}(S)|)$. Then, for $M \in \mathbb{R}^{m\times n}$, it holds that*

$$\|P_{\mathcal{Q}(S)}M\|_{\infty,2} \leqslant \|M\|_{\infty,2} \quad \textit{and}$$
$$\|P_{\mathcal{Q}(S)^\perp}M\|_{\infty,2} \leqslant \|M\|_{\infty,2}.$$

Next, let $\mathcal{T}(L)$ be the tangent space at $L \in \mathcal{L}(\operatorname{rank}(L))$. Then, for $N \in \mathbb{R}^{m\times n}$, it holds that

$$\|P_{\mathcal{T}(L)}N\| \leqslant 2\|N\| \quad \textit{and} \quad \|P_{\mathcal{T}(L)^\perp}N\| \leqslant \|N\|.$$

Proof. The claims for the projections on low-rank tangent spaces have been proven in Nussbaum and Giesen [2020]. The claims for projections on group-sparse tangent and normal spaces are simple. □

1.3 WEAKER ASSUMPTIONS USING NORM-COMPATIBILITY CONSTANTS

Proof of Lemma 3. We prove the bound for $\mu(\mathcal{Q}(S))$ first. Remember the definition

$$\mu(\mathcal{Q}(S)) = \max_{M\in\mathcal{Q}(S),\|M\|_{\infty,2}=1} \|M\|,$$

and recall that the *group-sign* function gsign maps a matrix $A \in \mathbb{R}^{m\times n}$ onto the matrix $\operatorname{gsign}(A) \in \mathbb{R}^{m\times n}$ with

$$\operatorname{gsign}(A)_{ij} = \begin{cases} a_{ij}/\|a_{ij}\|_2, & a_{ij} \not\equiv 0 \\ 0, & \text{else} \end{cases}$$

for $i \in 1, \dots, d$ and $j = 1, \dots, n$. Now, consider M with $\|M\|_{\infty,2} = 1$. Then, the matrix $|\operatorname{gsign}(M)|$ has normalized groups and non-negative entries. Moreover, it satisfies the element-wise inequality $|M| \leqslant |\operatorname{gsign}(M)|$. As a consequence of the Perron-Frobenius theorem Horn and Johnson [2012] it follows that $\|M\| \leqslant \||\operatorname{gsign}(M)|\|$. This allows us to conclude that to obtain $\mu(\mathcal{Q}(S))$ we only need to consider non-negative matrices $0 \leqslant M \in \mathcal{Q}(S)$ that are *normalized* in the sense that its non-zero groups precisely have norm 1, that is, $M = \operatorname{gsign}(M)$. Now, we bound the spectral norm of a matrix M, see Schur [1911], as follows

$$\|M\|^2 \leqslant \|M\|_1 \|M\|_\infty,$$

where $\|M\|_1$ is the maximum ℓ_1-norm of a column of M and $\|M\|_\infty$ is the maximum ℓ_1-norm of a row of M. We bound $\|M\|_1$. W.l.o.g. let c be a column of M with $\|M\|_1 = \|c\|_1$. Then, it holds

$$\begin{aligned}\|M\|_1 = \|c\|_1 \leqslant \sqrt{\eta}\|c\|_{1,2} &\leqslant \sqrt{\eta}\operatorname{gdeg}_{\max}(M) \\ &\leqslant \sqrt{\eta}\operatorname{gdeg}_{\max}(S),\end{aligned}$$

where the first inequality follows since $\eta = \max_{i\in[d]} m_i$ is the maximum number of elements that belong to a group from the column and hence that $\sqrt{\eta}$ is a norm compatibility constant for the column (vector) ℓ_1-norm and the column (vector) $\ell_{1,2}$-norm. The second inequality follows because the vector $\ell_{1,2}$-norm of c is equal to the number of non-zero groups of c. This number is bounded by $\operatorname{gdeg}_{\max}(M)$. Finally, the last inequality follows from $\operatorname{gsupp}(M) \subseteq \operatorname{gsupp}(S)$ as $M \in \mathcal{Q}(S)$.

Similar reasoning for rows instead of columns leads us to conclude that $\|M\|_\infty \leqslant \operatorname{gdeg}_{\max}(S)$ since this time comparison of the $\ell_{1,2}$- and ℓ_1 row (vector) norms is not necessary because the intersections of the groups of M with a row respectively contain at most one element. Therefore, we get the upper bound

$$\|M\| \leqslant \sqrt{\|M\|_1 \|M\|_\infty} \leqslant \eta^{1/4}\, \mathrm{gdeg}_{\max}(S).$$

This establishes the claim $\mu(Q(S)) \leqslant \eta^{1/4}\, \mathrm{gdeg}_{\max}(S)$.

Proof of (b). The claim about $\xi(\mathcal{T}(L))$ follows from

$$\begin{aligned}\xi(\mathcal{T}) &= \max_{M\in\mathcal{T}(L),\, \|M\|=1} \|M\|_{\infty,2} \\ &\leqslant \sqrt{\eta} \max_{M\in\mathcal{T}(L),\, \|M\|=1} \|\operatorname{vec}(M)\|_\infty \\ &\leqslant 2\sqrt{\eta}\, \mathrm{coh}(L),\end{aligned}$$

where the first inequality follows from the general norm bound between the (vector) ℓ_∞-norm and the $\ell_{\infty,2}$-norm which holds because η is the maximum number of elements of a group. Finally, the last inequality is a consequence of [Chandrasekaran et al., 2011, Proposition 4]. □

1.4 SUFFICIENT CONDITIONS FOR TRANSVERSALITY

Using the norm-compatibility constants introduced in the previous section, we can prove the following result that is stronger than Lemma 2 from the main paper.

Lemma 6. *Let* $S \in \mathcal{S}(|\mathrm{gsupp}(S)|)$ *with tangent space* $\mathcal{Q}(S)$ *at* S, *and let* $L \in \mathcal{L}(\mathrm{rank}(L))$ *with tangent space* $\mathcal{T}(L)$ *at* L. *Suppose that it holds* $\mu(\mathcal{Q}(S))\xi(\mathcal{T}(L)) < 1$. *Then, the tangent spaces are transverse, that is, it holds* $\mathcal{Q}(S) \cap \mathcal{T}(L) = \{0\}$.

Proof. Let $0 \neq M \in \mathcal{T}(L)$. We calculate

$$\begin{aligned}\|P_{\mathcal{Q}(S)}M\| &\leqslant \mu(\mathcal{Q}(S))\|P_{\mathcal{Q}(S)}M\|_{\infty,2} \\ &\leqslant \mu(\mathcal{Q}(S))\|M\|_{\infty,2} \\ &\leqslant \mu(\mathcal{Q}(S))\xi(\mathcal{T}(L))\|M\| < \|M\|,\end{aligned}$$

where the first inequality uses the definition of $\mu(\mathcal{Q}(S))$, the second inequality uses the projection Lemma 5, the third inequality follows from $M \in \mathcal{T}(L)$ and the definition of $\xi(T(L))$, and the last inequality follows from the assumption. It follows that $P_{\mathcal{Q}(S)}(M) \neq M$ such that M cannot be contained in $\mathcal{Q}(S)$. This implies transversality of the tangent spaces. □

Lemma 2 can now be proven as a simple corollary of Lemma 6.

Proof of Lemma 2.. By Lemma 3 and the assumption of Lemma 2 it holds that

$$\mu(\mathcal{Q}(S^\star))\xi(\mathcal{T}(L^\star)) < 2\eta^{3/4}\, \mathrm{gdeg}_{\max}(S^\star)\, \mathrm{coh}(L^\star) < 1.$$

Hence, the claim follows from Lemma 6. □

2 PROOFS OF THEOREM 2 AND THEOREM 1

In this section, we prove the theorems from the main paper. The first two section contain the proof of Theorem 2, which is done by studying the optimality conditions of Problem (1). The third section contains the proof of Theorem 1 as a corollary of Theorem 2.

2.1 OPTIMALITY CONDITIONS

Any solution (L, S) to Problem (1) must satisfy the first-order optimality conditions of Problem (1) that can be derived from the Lagrangian

$$\mathcal{L}(L, S, Z) = \|L\|_* + \gamma\|S\|_{1,2} + \langle Z, X - L - S\rangle\,,$$

where Z are the dual variables for the constraint $X = L + S$, and $\langle\cdot,\cdot\rangle$ denotes the standard scalar product on matrices. The first-order optimality conditions w.r.t. S and L require that Z is a subgradient from the $\ell_{1,2}$-norm and the nuclear norm *subdifferentials*, that is, it must hold $Z \in \gamma\partial\|S\|_{1,2}$ and $Z \in \partial\|L\|_*$. The norm subdifferentials can be characterized using dual norms Watson [1992]. First, it holds $Z \in \gamma\partial\|S\|_{1,2}$ if and only if

$$P_{\mathcal{Q}(S)}(Z) = \gamma\, \mathrm{gsign}(S) \quad \text{and} \quad \|P_{\mathcal{Q}^\perp(S)}(Z)\|_{\infty,2} \leqslant \gamma.$$

Next, it holds $Z \in \partial\|L\|_*$ if and only if

$$P_{\mathcal{T}(L)}(Z) = UV^\top \quad \text{and} \quad \|P_{\mathcal{T}(L)^\perp}(Z)\| \leqslant 1,$$

where $L = UDV^\top$ is a singular value decomposition of L, and $\|\cdot\|$ denotes the spectral norm, which is dual to the nuclear norm.

Based on the first-order optimality conditions, the following result states that $(L^\star, S^\star)$ uniquely solves Problem (1) provided that the tangent spaces $\mathcal{Q} = \mathcal{Q}(S^\star)$ and $\mathcal{T} = \mathcal{T}(L^\star)$ are transverse and given a dual Z that *strictly* satisfies the subgradient conditions above.

Proposition 1. *Suppose that* $X = L^\star + S^\star$. *Then,* $(L^\star, S^\star)$ *is the unique minimizer of Problem* (1) *if the following conditions are satisfied:*

1. *It holds* $\mathcal{Q}(S^\star) \cap \mathcal{T}(L^\star) = \{0\}$.

2. *There exists a subgradient* $Z \in \gamma\partial\|S^\star\|_{1,2} \cap \partial\|L^\star\|_*$ *that satisfies the strict dual-feasible conditions*

$$\|P_{\mathcal{Q}^\perp(S^\star)}(Z)\|_{\infty,2} < \gamma \quad \textit{and} \quad \|P_{\mathcal{T}^\perp(L^\star)}(Z)\| < 1.$$

The idea for the proof of uniqueness, which can be done similarly as in Candès et al. [2011], is to assume the existence of another minimizer $(L^\star - M, S^\star + M)$ with the goal of showing that $M = 0$. For that, the proof uses the subgradient property and the subgradient characterizations to show that $M \in \mathcal{Q}(S^\star) \cap \mathcal{T}(L^\star)$. In the spirit of the primal-dual witness proof technique that was originally introduced in Wainwright [2009] this implies that $S^\star + M \in \mathcal{Q}(S^\star)$ and $L^\star - M \in \mathcal{T}(L^\star)$. However, from the transversality $\mathcal{Q}(S^\star) \cap \mathcal{T}(L^\star) = \{0\}$ it follows that M must be zero.

Proof of Proposition 1. First, it follows that $(L^\star, S^\star)$ is *an* optimum since by the second condition from the assumption there exists a dual Z that satisfies both optimality conditions. Now, for some matrix M, let $(L^\star - M, S^\star + M)$ be another minimizer of Problem (1). The minimizer must have this form in order to be feasible. Our goal is to show that the components of M in the normal spaces $\mathcal{Q}^\perp$ and $\mathcal{T}^\perp$ vanish, respectively. We begin by using the subgradient property:

$$\begin{aligned} 0 &= \gamma\|S^\star + M\|_{1,2} + \|L^\star - M\|_* - \gamma\|S^\star\|_{1,2} - \|L^\star\|_* \\ &\geqslant \langle Z_{1,2}, M\rangle - \langle Z_*, M\rangle \\ &= \langle P_{\mathcal{Q}^\perp}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}^\perp}(Z_*), M\rangle \\ &\quad + \langle P_{\mathcal{Q}}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}}(Z_*), M\rangle, \end{aligned}$$

where $Z_{1,2} \in \gamma\,\partial\|S^\star\|_{1,2}$ and $Z_* \in \partial\|L^\star\|_*$ are subgradients whose choices we make precise later. The idea is to chose them such that the right hand side of the inequality is maximized. In the last line, we decomposed the terms into their tangential and normal components. Note that the tangential components $\langle P_{\mathcal{Q}}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}}(Z_*), M\rangle$ do not depend on the choice of the subgradients since by the subgradient characterizations the pair $(Z_{1,2}, Z_*)$ must satisfy

$$P_{\mathcal{Q}}(Z_{1,2}) = \gamma\,\mathrm{gsign}(S^\star) \quad \text{and} \quad P_{\mathcal{T}}(Z_*) = UV^\top,$$

where $L^\star = UDV^\top$ is the singular value decomposition of $L^\star$. Hence, this constant part can be bounded by

$$\begin{aligned} &\langle P_{\mathcal{Q}}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}}(Z_*), M\rangle \\ &= \langle Z - P_{\mathcal{Q}^\perp}(Z), M\rangle - \langle Z - P_{\mathcal{T}^\perp}(Z), M\rangle \\ &= -\langle P_{\mathcal{Q}^\perp}(Z), M\rangle + \langle P_{\mathcal{T}^\perp}(Z), M\rangle \\ &= -\langle P_{\mathcal{Q}^\perp}(Z), P_{\mathcal{Q}^\perp}(M)\rangle + \langle P_{\mathcal{T}^\perp}(Z), P_{\mathcal{T}^\perp}(M)\rangle \\ &\geqslant -\left|\langle P_{\mathcal{Q}^\perp}(Z), P_{\mathcal{Q}^\perp}(M)\rangle\right| - \left|\langle P_{\mathcal{T}^\perp}(Z), P_{\mathcal{T}^\perp}(M)\rangle\right| \\ &\geqslant -\left\|P_{\mathcal{Q}^\perp}(Z)\right\|_{\infty,2}\left\|P_{\mathcal{Q}^\perp}(M)\right\|_{1,2} \\ &\quad - \|P_{\mathcal{T}^\perp}(Z)\|\,\|P_{\mathcal{T}^\perp}(M)\|_*, \end{aligned}$$

where the first equality uses that Z satisfies the subgradient conditions as well, and the final inequality applies the generalized Hoelder inequality (respectively for the $\|\cdot\|_{1,2}$, $\|\cdot\|_{\infty,2}$ and the $\|\cdot\|$, $\|\cdot\|_*$ dual norm pairs).

Next, we calculate $\langle P_{\mathcal{Q}^\perp}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}^\perp}(Z_*), M\rangle$ after choosing the normal components of $Z_{1,2}$ and Z_* in $\mathcal{Q}^\perp$ and $\mathcal{T}^\perp$, respectively. First, we select $P_{\mathcal{Q}^\perp}(Z_{1,2}) = \gamma\,\mathrm{gsign}\left(P_{\mathcal{Q}^\perp}(M)\right)$. This yields a valid subgradient because then $\|P_{\mathcal{Q}^\perp}(Z_{1,2})\|_{\infty,2} = \gamma$. Moreover, it holds that

$$\begin{aligned} \langle P_{\mathcal{Q}^\perp}(Z_{1,2}), M\rangle &= \langle P_{\mathcal{Q}^\perp}(Z_{1,2}), P_{\mathcal{Q}^\perp}(M)\rangle \\ &= \gamma\left\langle \mathrm{gsign}\left(P_{\mathcal{Q}^\perp}(M)\right), P_{\mathcal{Q}^\perp}(M)\right\rangle \\ &= \gamma\|P_{\mathcal{Q}^\perp}(M)\|_{1,2}. \end{aligned}$$

Second, we select $P_{\mathcal{T}^\perp}(Z_*) = -\tilde{U}\,\mathrm{sign}(\tilde{\Sigma})\tilde{V}^\top$ based on a singular value decomposition $P_{\mathcal{T}^\perp}(M) = \tilde{U}\tilde{\Sigma}\tilde{V}^\top$ of $P_{\mathcal{T}^\perp}(M)$. This forms a valid subgradient since $\|P_{\mathcal{T}^\perp}(Z_*)\| = 1$. Besides, we have

$$\begin{aligned} -\langle P_{\mathcal{T}^\perp}(Z_*), M\rangle &= -\langle P_{\mathcal{T}^\perp}(Z_*), P_{\mathcal{T}^\perp}(M)\rangle \\ &= -\left\langle -\tilde{U}\,\mathrm{sign}(\tilde{\Sigma})\tilde{V}^\top, \tilde{U}\tilde{\Sigma}\tilde{V}^\top\right\rangle \\ &= \mathrm{tr}\left(\left(\tilde{U}\,\mathrm{sign}(\tilde{\Sigma})\tilde{V}^\top\right)^\top \tilde{U}\tilde{\Sigma}\tilde{V}^\top\right) \\ &= \mathrm{tr}\left(\tilde{V}\,\mathrm{sign}(\tilde{\Sigma})\tilde{U}^\top\tilde{U}\tilde{\Sigma}\tilde{V}^\top\right) \\ &= \mathrm{tr}\left(\tilde{V}|\tilde{\Sigma}|\tilde{V}^\top\right) = \mathrm{tr}\left(\tilde{V}^\top\tilde{V}|\tilde{\Sigma}|\right) \\ &= \mathrm{tr}\left(|\tilde{\Sigma}|\right) = \|P_{\mathcal{T}^\perp}(M)\|_*. \end{aligned}$$

In summary, the specific choices of the subgradients yield

$$\begin{aligned} 0 &\geqslant \langle P_{\mathcal{Q}^\perp}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}^\perp}(Z_*), M\rangle \\ &\quad + \langle P_{\mathcal{Q}}(Z_{1,2}), M\rangle - \langle P_{\mathcal{T}}(Z_*), M\rangle \\ &\geqslant \gamma\|P_{\mathcal{Q}^\perp}(M)\|_{1,2} + \|P_{\mathcal{T}^\perp}(M)\|_* \\ &\quad - \left\|P_{\mathcal{Q}^\perp}(Z)\right\|_{\infty,2}\left\|P_{\mathcal{Q}^\perp}(M)\right\|_{1,2} \\ &\quad - \|P_{\mathcal{T}^\perp}(Z)\|\,\|P_{\mathcal{T}^\perp}(M)\|_* \\ &= \left(\gamma - \|P_{\mathcal{Q}^\perp}(Z)\|_{\infty,2}\right)\|P_{\mathcal{Q}^\perp}(M)\|_{1,2} \\ &\quad + \left(1 - \|P_{\mathcal{T}^\perp}(Z)\|\right)\|P_{\mathcal{T}^\perp}(M)\|_*. \end{aligned}$$

By the strictly-dual-feasible condition we have that $\|P_{\mathcal{Q}^\perp}(Z)\|_{\infty,2} < \gamma$ and $\|P_{\mathcal{T}^\perp}(Z)\| < 1$. Thus, if any of $P_{\mathcal{Q}^\perp}(M)$ *or* $P_{\mathcal{T}^\perp}(M)$ are non-zero, then the right-hand side becomes strictly positive, which would be a contradiction. Therefore, we must have $P_{\mathcal{Q}^\perp}(M) = 0$ and $P_{\mathcal{T}^\perp}(M) = 0$. This means that M must be contained in both $\mathcal{Q}$ and T. However, since we assumed transversality, we have $\mathcal{Q} \cap T = \{0\}$. It follows that $M = 0$, which implies the uniqueness of the solution to Problem (1). □

2.2 PROOF OF THEOREM 2

Here, we prove Theorem 2. The main idea of the proof is to show that for any $\gamma \in (\gamma^{\circ}_{\min}, \gamma^{\circ}_{\max})$ there exists a strictly dual feasible Z as required by Proposition 1.

Proof of Theorem 2. Let us first check that the range of values for γ given by

$$\left(\frac{\xi(\mathcal{T})}{1-4\mu(\mathcal{Q})\xi(\mathcal{T})}, \frac{1-3\mu(\mathcal{Q})\xi(\mathcal{T})}{\mu(\mathcal{Q})}\right)$$

is non-empty. For that, observe that comparing the borders of the interval leads to the quadratic inequality

$$12\,[\mu(\mathcal{Q})\xi(\mathcal{T})]^2 - 8\,[\mu(\mathcal{Q})\xi(\mathcal{T})] + 1 > 0$$

in $\mu(\mathcal{Q})\xi(\mathcal{T})$. The roots of the quadratic polynomial are $1/6$ and $1/2$, so clearly under the assumption $\mu(\mathcal{Q})\xi(\mathcal{T}) < 1/6$ the given range is non-empty.

Because of the assumption, we also can apply Lemma 6 that yields $\mathcal{Q} \cap \mathcal{T} = \{0\}$. Therefore, there exists a *unique* $Z \in \mathcal{Q} \oplus \mathcal{T}$, where $\oplus$ denotes the direct sum, such that the orthogonal projections of Z onto the tangent spaces $\mathcal{Q}$ and $\mathcal{T}$ are consistent with the subgradient conditions, that is, it holds

$$P_{\mathcal{Q}}(Z) = \gamma\, \mathrm{gsign}(S^{\star}) \quad \text{and} \quad P_{\mathcal{T}}(Z) = UV^{\top}.$$

Remember that $L^{\star} = UDV^{\top}$ is the (restricted) singular value decomposition of $L^{\star}$. The rest of the proof is dedicated to showing that Z also *strictly* satisfies the remaining subgradient conditions that concern the orthogonal projections, that is, we want to show the strict dual-feasible conditions

$$\|P_{\mathcal{Q}^{\perp}}(Z)\|_{\infty,2} < \gamma \quad \text{and} \quad \|P_{\mathcal{T}^{\perp}}(Z)\| < 1$$

that are required by Proposition 1. For that, let $Z = Z_{\mathcal{Q}} + Z_{\mathcal{T}}$ be the unique splitting of Z into its components $Z_{\mathcal{Q}} \in \mathcal{Q}$ and $Z_{\mathcal{T}} \in \mathcal{T}$, see Figure 7. We have $Z_{\mathcal{Q}} = P_{\mathcal{Q}}(Z) - P_{\mathcal{Q}}(Z_{\mathcal{T}}) = \gamma\,\mathrm{gsign}(S^{\star}) - P_{\mathcal{Q}}(Z_{\mathcal{T}})$ and $Z_{\mathcal{T}} = P_{\mathcal{T}}(Z) - P_{\mathcal{T}}(Z_{\mathcal{Q}}) = UV^{\top} - P_{\mathcal{T}}(Z_{\mathcal{Q}})$. Now, we

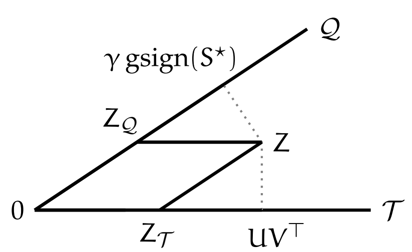

Figure 7: Decomposition of the dual Z in $\mathcal{Q} \oplus \mathcal{T}$.

start bounding the orthogonal components. The component of Z in $\mathcal{Q}^{\perp}$ can be bounded as

$$\begin{aligned}\|P_{\mathcal{Q}^{\perp}}(Z)\|_{\infty,2} &= \|P_{\mathcal{Q}^{\perp}}(Z_{\mathcal{T}})\|_{\infty,2} \leqslant \|Z_{\mathcal{T}}\|_{\infty,2} \\ &\leqslant \xi(\mathcal{T})\|Z_{\mathcal{T}}\| = \xi(\mathcal{T})\|UV^{\top} - P_{\mathcal{T}}(Z_{\mathcal{Q}})\| \\ &\leqslant \xi(\mathcal{T})(1 + \|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|), \end{aligned} \tag{2}$$

where we used the projection Lemma 5 in the first, the definition of $\xi(\mathcal{T})$ in the second, and the triangle inequality in the last inequality. Similarly, we can bound the component of Z in $\mathcal{T}^{\perp}$

$$\begin{aligned}\|P_{\mathcal{T}^{\perp}}(Z)\| &= \|P_{\mathcal{T}^{\perp}}(Z_{\mathcal{Q}})\| \leqslant \|Z_{\mathcal{Q}}\| \\ &\leqslant \mu(\mathcal{Q})\,\|Z_{\mathcal{Q}}\|_{\infty,2} \\ &= \mu(\mathcal{Q})\|\gamma\,\mathrm{gsign}(S^{\star}) - P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2} \\ &\leqslant \mu(\mathcal{Q})\,\left(\gamma + \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2}\right), \end{aligned} \tag{3}$$

where again Lemma 5 was used in the first, the definition of $\mu(\mathcal{Q})$ in the second, and finally the triangle inequality in the last inequality. To continue the calculations we bound the norms of $P_{\mathcal{T}}(Z_{\mathcal{Q}})$ and $P_{\mathcal{Q}}(Z_{\mathcal{T}})$.

$$\begin{aligned}\|P_{\mathcal{T}}(Z_{\mathcal{Q}})\| &\leqslant 2\|Z_{\mathcal{Q}}\| \leqslant 2\mu(\mathcal{Q})\,\left(\gamma + \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2}\right), \\ \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2} &\leqslant \|Z_{\mathcal{T}}\|_{\infty,2} \leqslant \xi(\mathcal{T})\,(1 + \|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|),\end{aligned}$$

where we used the projection Lemma 5, bounded $\|Z_{\mathcal{T}}\|_{\infty,2}$ as in (2), and bounded $\|Z_{\mathcal{Q}}\|$ as in (3). Plugging the bounds on $\|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|$ and $\|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|$ into each other yields

$$\begin{aligned}\|P_{\mathcal{T}}(Z_{\mathcal{Q}})\| &\leqslant 2\mu(\mathcal{Q})\,\left[\gamma + \xi(\mathcal{T})\,(1 + \|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|)\right], \\ \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2} &\leqslant \xi(\mathcal{T})\,\left[1 + 2\mu(\mathcal{Q})\,\left(\gamma + \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2}\right)\right].\end{aligned}$$

By solving these inequalities for $\|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|$ and $\|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2}$, respectively, we obtain

$$\|P_{\mathcal{T}}(Z_{\mathcal{Q}})\| \leqslant \frac{2\gamma\mu(\mathcal{Q}) + 2\mu(\mathcal{Q})\xi(\mathcal{T})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})} \tag{4a}$$

$$\|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2} \leqslant \frac{\xi(\mathcal{T}) + 2\gamma\mu(\mathcal{Q})\xi(\mathcal{T})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})} \tag{4b}$$

Note that since $\mu(\mathcal{Q})\xi(\mathcal{T}) < 1/6 < 1/2$, the denominators are positive. Bringing (2) and (4a) together yields

$$\begin{aligned}\|P_{\mathcal{Q}^{\perp}}(Z)\|_{\infty,2} &\leqslant \xi(\mathcal{T})(1 + \|P_{\mathcal{T}}(Z_{\mathcal{Q}})\|) \\ &\leqslant \xi(\mathcal{T})\left(1 + \frac{2\gamma\mu(\mathcal{Q}) + 2\mu(\mathcal{Q})\xi(\mathcal{T})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) \\ &= \xi(\mathcal{T})\left(\frac{1 + 2\gamma\mu(\mathcal{Q})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) \\ &= \left[\xi(\mathcal{T})\left(\frac{1 + 2\gamma\mu(\mathcal{Q})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) - \gamma\right] + \gamma \\ &= \left[\frac{\xi(\mathcal{T}) + 2\gamma\mu(\mathcal{Q})\xi(\mathcal{T}) - \gamma + 2\gamma\mu(\mathcal{Q})\xi(\mathcal{T})}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})}\right] + \gamma \\ &= \left[\frac{\xi(\mathcal{T}) - \gamma\,(1 - 4\mu(\mathcal{Q})\xi(\mathcal{T}))}{1 - 2\mu(\mathcal{Q})\xi(\mathcal{T})}\right] + \gamma < \gamma,\end{aligned}$$

where the last inequality holds by the assumption

$$\gamma > \xi(\mathcal{T})/[1 - 4\mu(\mathcal{Q})\xi(\mathcal{T})].$$

Next, by (3) and (4b) we have

$$\begin{aligned}\|P_{\mathcal{T}^\perp}(Z)\| &\leqslant \mu(\mathcal{Q})\ \left(\gamma + \|P_{\mathcal{Q}}(Z_{\mathcal{T}})\|_{\infty,2}\right) \\ &\leqslant \mu(\mathcal{Q})\ \left(\gamma + \frac{\xi(\mathcal{T}) + 2\gamma\mu(\mathcal{Q})\xi(\mathcal{T})}{1-2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) \\ &= \mu(\mathcal{Q})\ \left(\frac{\gamma + \xi(\mathcal{T})}{1-2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) \\ &< \mu(\mathcal{Q})\ \left(\frac{[1-3\mu(\mathcal{Q})\xi(\mathcal{T})]/\mu(\mathcal{Q}) + \xi(\mathcal{T})}{1-2\mu(\mathcal{Q})\xi(\mathcal{T})}\right) \\ &= \frac{1-3\mu(\mathcal{Q})\xi(\mathcal{T}) + \mu(\mathcal{Q})\xi(\mathcal{T})}{1-2\mu(\mathcal{Q})\xi(\mathcal{T})} = 1,\end{aligned}$$

where we used the bound $\gamma < [1-3\mu(\mathcal{Q})\xi(\mathcal{T})]/\mu(\mathcal{Q})$ from the assumption in the last inequality. This completes the proof. □

2.3 PROOF OF THEOREM 1

Finally, we prove Theorem 1 from the main paper as a simple corollary of Theorem 2.

Proof of Theorem 1 as a Corollary of Theorem 2. This is straightforward using the lower bounds on $\mathrm{gdeg}_{\max}(S^\star)$ and $\mathrm{coh}(L^\star)$ from Lemma 3. Particularly, it holds

$$\mu(\mathcal{Q}(S^\star))\xi(\mathcal{T}(L^\star)) \leqslant \eta^{\frac{1}{4}}\ \mathrm{gdeg}_{\max}(S^\star) 2\eta^{\frac{1}{2}}\ \mathrm{coh}(L^\star) < \frac{1}{6},$$

where the final inequality follows from the assumption. Hence, we can apply Theorem 2. One can check by plugging in the lower bounds from Lemma 3 that the range $(\gamma_{\min}, \gamma_{\max})$ of values for γ is a non-empty subrange of the range $(\gamma^\circ_{\min}, \gamma^\circ_{\max})$ given in Theorem 2. □

3 ADMM FOR PROBLEM (1)

Similar to Candès et al. [2011], we derive an Alternating Direction Method of Multipliers (ADMM) algorithm for the decomposition problem

$$\min_{S,\,L\,\in\,\mathbb{R}^{m\times n}} \quad \gamma\|S\|_{1,2} + \|L\|_* \quad \text{s.t.} \quad X = S + L$$

(note that we changed the order of S and L). This has augmented Lagrangian

$$\begin{aligned}\mathcal{L}(S,L,Z) &= \gamma\|S\|_{1,2} + \|L\|_* \\ &\quad + \langle Z, X-S-L\rangle + \frac{1}{2\kappa}\|X-S-L\|_F^2 \\ &= \gamma\|S\|_{1,2} + \|L\|_* \\ &\quad + \frac{1}{2\kappa}\|X-S-L+\kappa Z\|_F^2 - \frac{\kappa}{2}\|Z\|_F^2,\end{aligned}$$

where Z are the dual variables for the constraint and $\kappa > 0$. Minimization of the augmented Lagrangian w.r.t. S and L is equivalent to solving proximal operators with known solutions. Consequently, ADMM performs the following updates:

$$\begin{cases} S^{k+1} &= \arg\min_S\ \mathcal{L}(S, L^k, Z^k) \\ &= \mathrm{gShrink}(X - L^k + \kappa Z^k, \gamma\kappa), \\ L^{k+1} &= \arg\min_L\ \mathcal{L}(S^{k+1}, L, Z^k) \\ &= \mathrm{sShrink}(X - S^{k+1} + \kappa Z^k, \kappa), \\ Z^{k+1} &= Z^k + \kappa^{-1}(X - S^{k+1} - L^{k+1}), \end{cases}$$

where the group soft-shrinkage operation acts on the (i,j)-th group as

$$[\mathrm{gShrink}(Z,\kappa)]_{ij} = z_{ij}\cdot\max\left\{1 - \frac{\kappa}{\|z_{ij}\|_2}, 0\right\},$$

remember that z_{ij} is the sub-vector that corresponds to the i-th group of variables in the j-th column of Z. Moreover, the spectral shrinkage operator is given by

$$\mathrm{sShrink}(Z,\kappa) = U\,\mathrm{Shrink}(E,\kappa)V^\top, \quad \text{with}$$
$$\mathrm{Shrink}(z,\kappa) = \mathrm{sign}(z)\ \max\{z-\kappa, 0\},$$

where $Z = UEV^\top$ is the singular value decomposition of Z.

4 ADDITIONAL EXPERIMENTAL RESULTS

In addition to the materials in the main paper, Figure 8 shows how the solution of Problem (1) for different regularization parameters changes for the electrical grid data. Remember that we reparameterize the objective of Problem (1) as $(1-\alpha)\|L\|_* + \alpha\|S\|_{1,2}$ and search for α in the compact interval $[0,1]$ instead of searching for γ in the unbounded interval $[0,\infty)$. In Figure 8, we see that between roughly $\alpha = 0.01$ and $\alpha = 0.014$ the solution does not change much. The change is not completely zero. Still, the solution is relatively stable, particularly, the same structural discoveries (a low-rank day-and-night pattern, see the main paper) can be made for all values of α (or equivalently γ) in the stable range.

In what follows, we show additional experimental results that concern the tasks reconstruction of RGB images, cloud removal from Sentinel-2 data, and detection of weather anomalies.

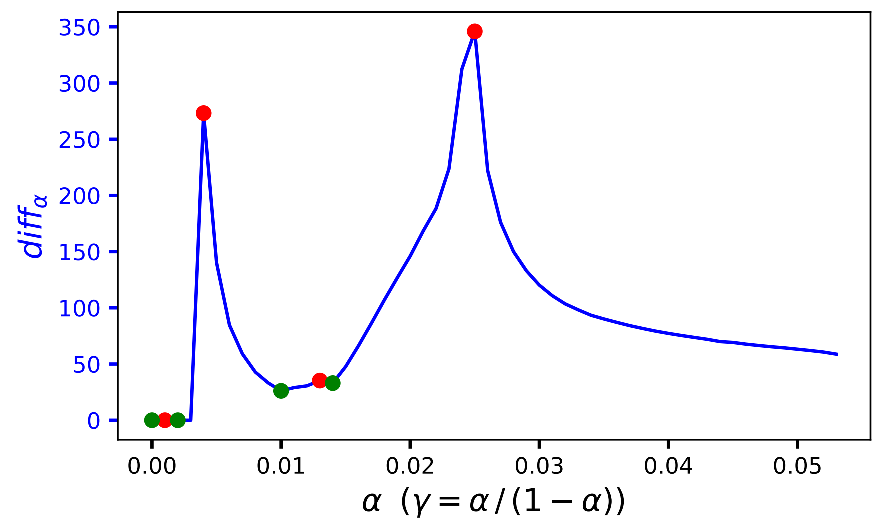

Figure 8: Change of the solution for different regularization parameter values for the electrical loadprofiles data: Here, $\mathrm{diff}_\alpha = \|L_{\alpha-\delta} - L_\alpha\|_F + \|S_{\alpha-\delta} - S_\alpha\|_F$ is as in the main paper, where we use the step-size $\delta = 10^{-3}$. Green dots represent local minima, red dots local maxima.

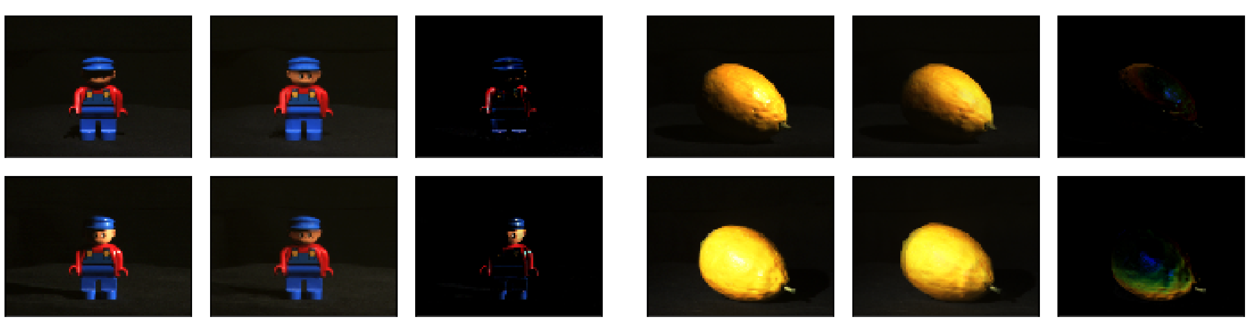

Figure 9: Robust reconstruction of RGB images for two additional objects from the Amsterdam Library of Object Images Geusebroek et al. [2005], respectively using $\gamma = 10^{-2}$. From left to right, the original images, the reconstructed low-rank RGB images, and the identified group-sparse components are shown. In the low-rank components, spotlights have been removed and shadows have been diminished.

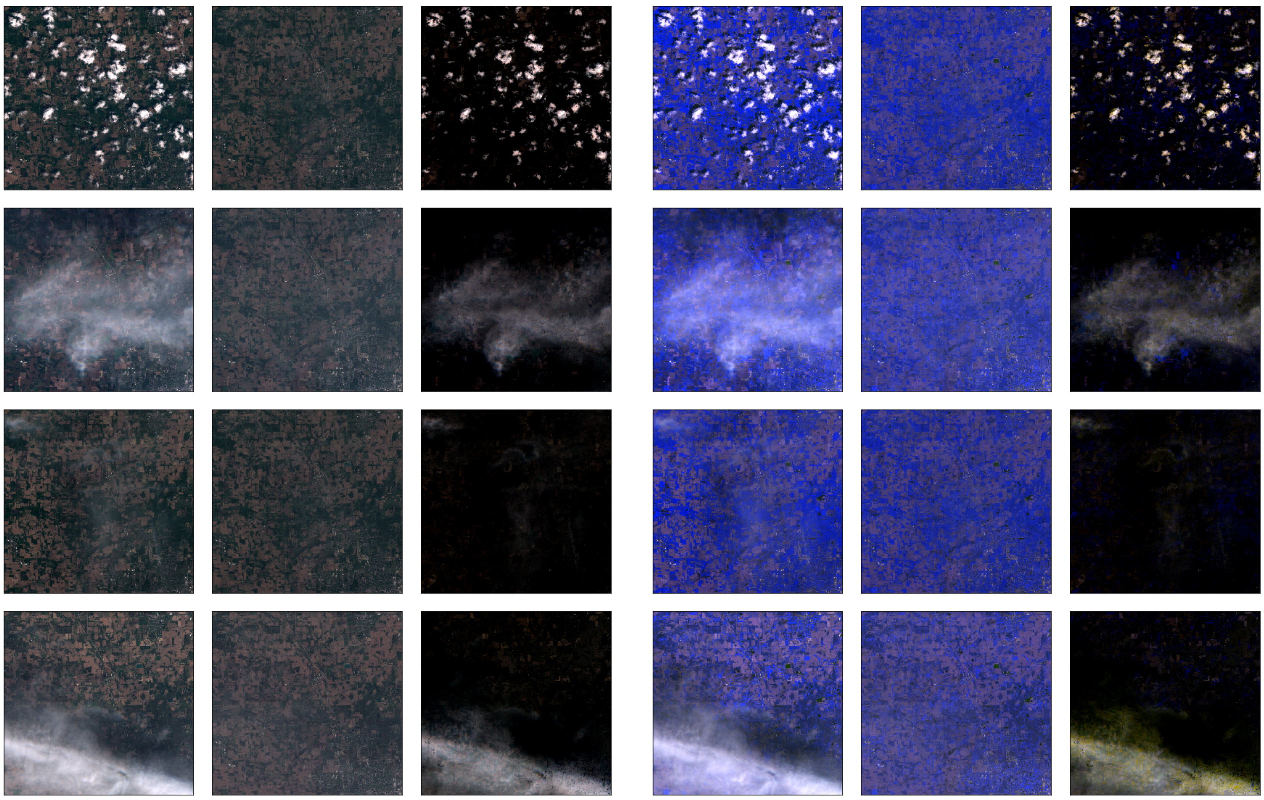

Figure 10: Cloud removal using robust PCA (Sentinel-2 data), $\gamma = 10^{-3}$. In this additional experiment (which is not included in the main paper for space reasons), we investigate the task of detecting/removing clouds from satellite data. For this task, it makes sense to apply robust PCA because the surface does not change much (besides seasonal shifts in vegetation), while clouds cover parts of the surface only temporarily. Experiments were performed on a multi-spectral image time series that consists of 20 observations of Fort Wayne (Indiana, USA) from the years 2019 and 2020. The data was obtained from the *Copernicus Open Access Hub* ESA [2020]. After cropping and downsampling, each image has a size of 1000×1000 pixels and uses four bands: red, green, blue, and near infrared (these correspond to bands 2, 3, 4, and 8 from the 13 available bands of the Sentinel-2 mission). To apply robust PCA, we group the four channels such that each pixel forms a group. In total, the data matrix has dimensions $X \in \mathbb{R}^{4\,000\,000 \times 20}$. In the plot, the three columns on the left show the decompositions with RGB colors, the three columns on the right show the corresponding false color images that are constructed from the red, green, and near infrared bands. As usual, from left to right, the original images, the low-rank components, and the group-sparse components are shown.

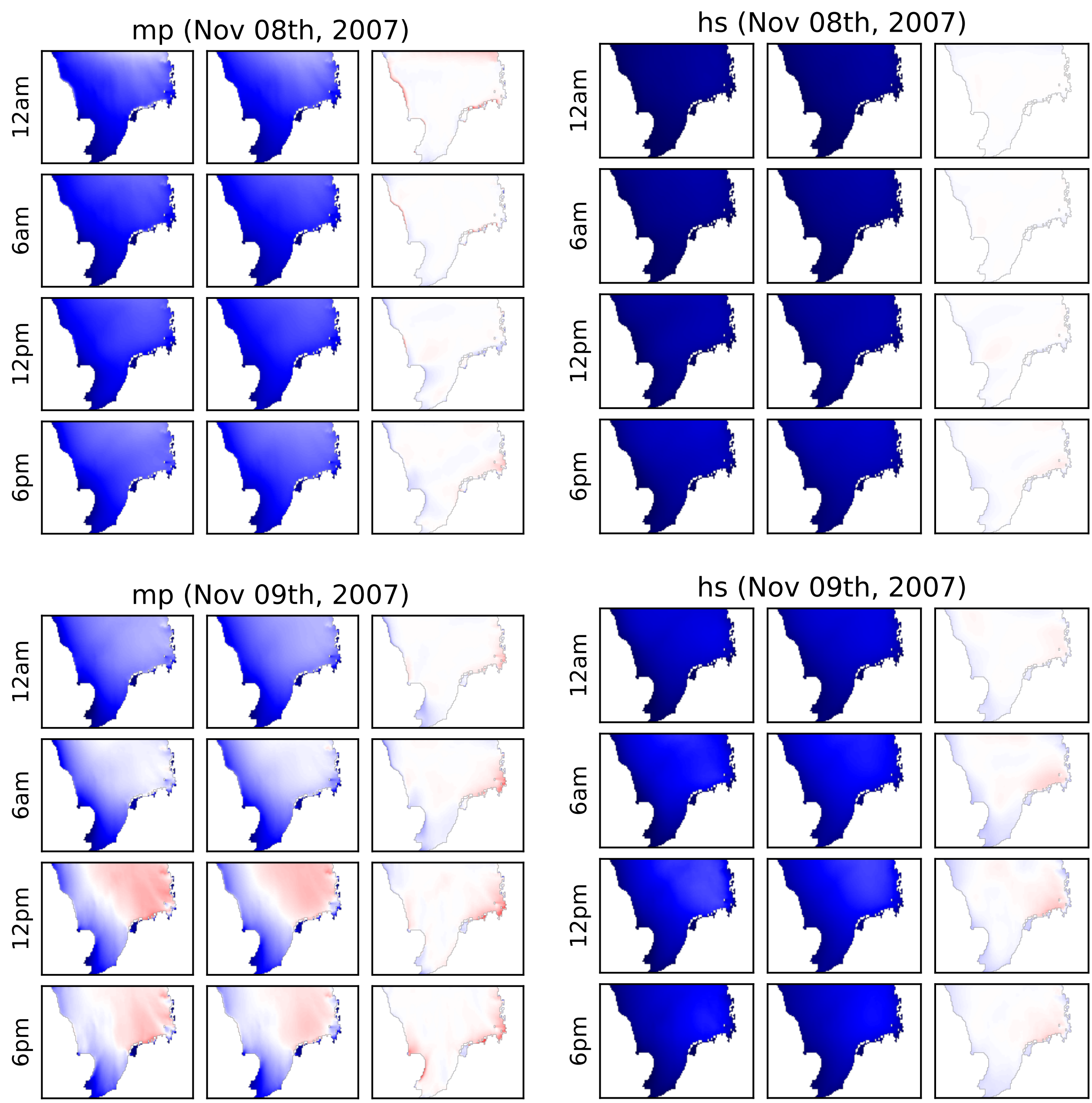

Figure 11: Decomposition of the wave hindcast data, $\gamma = 10^{-3}$: The features *mp* (mean wave period) and *hs* (significant weight height) are shown for four time steps of November 8th, 2007 (top row of plots) and November 9th (bottom row), respectively. In each plot, the left column shows the data matrix, the middle one the low-rank component, and the right one shows the outlier component. Red corresponds to increased energy. November 8th, 2007 was a day without weather anomalies. Consequently, the outlier component is mostly zero (white). The next day cyclone Tilo caused severe North Sea floods. As a consequence, the coastal lines show increased energy (storm surges) in the outlier components.

References

Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

ESA. Sentinel-2 mission, tile t16tfl (2019-2020), accessed: 2020-08-06., 2020. URL `https://scihub.copernicus.eu/`.

Jan-Mark Geusebroek, Gertjan J. Burghouts, and Arnold W. M. Smeulders. The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge university press, 2012.

Frank Nussbaum and Joachim Giesen. Pairwise sparse + low-rank models for variables of mixed type. *Journal of Multivariate Analysis*, 178:104601, 2020. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2020.104601. URL `http://www.sciencedirect.com/science/article/pii/S0047259X19303756`.

Issai Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die reine und Angewandte Mathematik*, 140:1–28, 1911.

Martin J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Information Theory*, 55(5):2183–2202, 2009.

Alistair Watson, G. Characterization of the subdifferential of some matrix norms. *Linear algebra and its applications*, 170:33–45, 1992.