Supplementary Material for Signal recovery from Pooling Representations

A. Comparison between phaselift and the various alternating minimization algorithms

Here we give a brief comparison between the phaselift algorithm and the algorithms we use in the main text. Our main goal is to show that the similarities between the ℓ_1, ℓ_2, ℓ_∞ recovery results are not just due to the alternating minimization algorithm performing poorly on all three tasks; however we feel that the quality of the recovery with a regressed initialization is interesting in itself, especially considering that it is much faster than either phaselift or phasecut.

In figures 2, and 3 we compare phaselift against alternating minimization with a random initialization and alternating minimization with a nearest neghbor/locally linear regressed initialization. Because we are comparing against phasecut, here we only show inversion of ℓ_2 pooling.

In figure of 2, we use random data and a random dictionary. As the data has no structure, we only compare against random initialization, with and without half rectification. We can see from figure 2 in this case, where we do not know a good way to initialize the alternating minimization, alternating minimization is significantly worse than phasecut. On the other hand, recovery after rectified pooling with alternating minimization does almost as well as phasecut.

In the examples where we have training data, shown in figure 3, alternating minimization with the nearest neighbor regressor (red curve) performs significantly better than phasecut (green and blue curves). Of course phasecut does not get the knowledge of the data distribution used to generate the regressor.

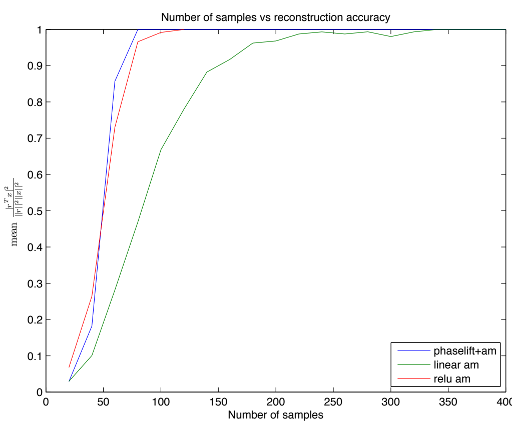

Figure 2. Average recovery angle using phaselift and alternating minimization on random data, Gaussian i.i.d. points in $\mathbb{R}^{40}$. The blue curve is phaselift followed by alternating minimization; the green curve is alternating minimization, and the red is alternating minimization on pooling following half rectification.

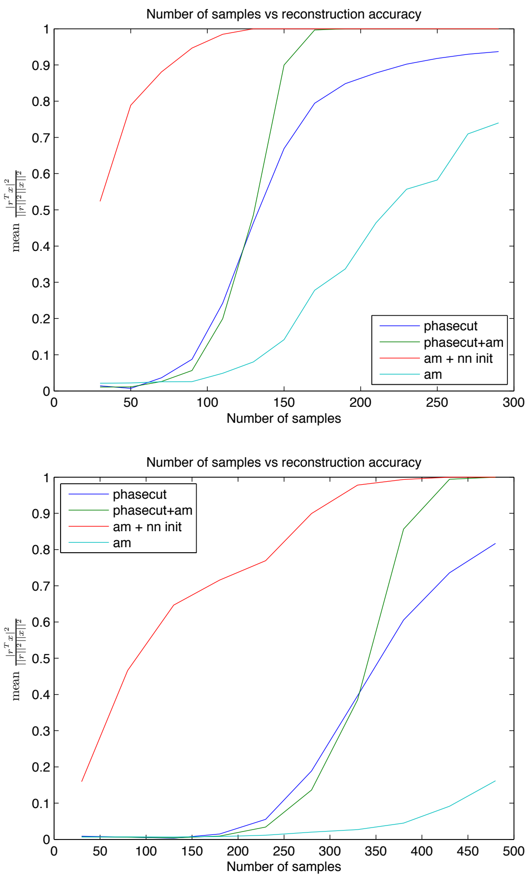

Figure 3. Average recovery angle using phaselift and alternating minimization on MNIST and patches data sets. Top: MNIST digits, projected via PCA to $\mathbb{R}^{100}$. Bottom: 16x16 image patches with mean removed. The red curve is alternating minimization with nearest neighbor initialization, the green is alternating minimization initialized by phasecut (this is the recommended usage of phasecut), the blue is phasecut with no alternating minimization, and the aqua is alternating minimization with a random initialization.

B. Proofs of results in Section 2

B.1. Proof of Proposition 2.2

Let us first show that $A_0 > 0$ is sufficient to construct an inverse of M_α. Let $x \in \mathbb{R}^N$. By definition, the coordinates of $M_\alpha(x) > \alpha$ correspond to

$$s(x) = \{i \ s.t. \langle x, f_i \rangle > \alpha_i\} \subset \{1, \ldots, M\} ,$$

which in particular implies that x is known to lie in $V_{S(x)}$, the subspace generated by $s(x)$. But the restriction $\mathcal{F}_{s(x)}$ is a linear operator, which can be inverted in V_S as long as $\lambda_-(\mathcal{F}_{s(x)}\big|_{V_S}) \geq A_0 > 0$.

Let us now show that $A_0 > 0$ is also necessary. Let us suppose that for some S, $\mathcal{F}_S$ is such that $\lambda_-(\mathcal{F}_S\big|_{V_S}) = 0$. It results that there exists $\eta \in V_S$ such that $\|\eta\| > 0$ but $\|\mathcal{F}_S \eta\| = 0$. Since S is a cone, we can find $x \in S$ and $\epsilon \neq 0$ small enough such that $x + \epsilon e \in S$. It results that $M_\alpha(x) = M_\alpha(x + \epsilon e)$ which implies that M_α cannot be injective.

Finally, let us prove (9). If x, x' are such that $S = s(x) = s(x')$, then

$$\|M_\alpha(x) - M_\alpha(x')\| = \|\mathcal{F}_S(x - x')\| \geq A_0(x - x') .$$

If $s(x) \neq s(x')$, let us denote $sI = s(x) \cap s(x')$. Then we have that

$$\begin{aligned} \| & M_\alpha(x) - M_\alpha(x')\|^2 = \|M_\alpha(x) - M_\alpha(x')\|^2_{sI} + \\ + & \|M_\alpha(x)\|^2_{s(x)\setminus sI} + \|M_\alpha(x')\|^2_{s(x')\setminus sI} . \end{aligned}$$

By denoting $\gamma = \lambda_-^2(\mathcal{F}_{sI}) + \min(\lambda_-^2(\mathcal{F}_{s(x)\setminus sI}), \lambda_-^2(\mathcal{F}_{s(x')\setminus sI}))$, and using the parallelogram identity $\|x\|^2 + \|x'\|^2 = \frac{1}{2}(\|x+x'\|^2 + \|x-x'\|^2) \geq \min(\|x - x'\|, \|x + x'\|)^2$, it results that

$$\|M_\alpha(x) - M_\alpha(x')\|^2 \geq \gamma \min(\|x - x'\|, \|x + x'\|)^2 ,$$

which implies (9). $\square$.

B.2. Proof of Proposition 2.4

The upper Lipschitz bound is obtained by observing that, in dimension d,

$$\forall\, y \in \mathbb{R}^d \ , \ \|y\|_1 \leq \sqrt{d}\|y\|_2 \ , \ \|y\|_\infty \leq d\|y\|_2 .$$

It results that

$$\begin{aligned} \|P_p(x) - P_p(x')\| &\leq \alpha_p \|P_2(x) - P_2(x')\| \qquad (21) \\ &= \alpha_p \|M(x) - M(x')\| \leq \alpha_p \lambda_+(\mathcal{F}) . \end{aligned}$$

Let us now concentrate on the lower Lipschitz bound. Given $x, x' \in \mathbb{R}^n$, we first consider a rotation $\tilde{\mathcal{F}}_k$ on each subspace $\mathcal{F}_k$ such that $\langle x, \tilde{f}_{k,j}\rangle = \langle x', \tilde{f}_{k,j}\rangle = 0$ for $j > 2$, which always exists. If now we modify $\tilde{\mathcal{F}}_k$ by applying a rotation of the remaining two-dimensional subspace such that x and x' are bisected, one can verify that

$$\begin{aligned} (\|\mathcal{F}_k x\|_2 - \|\mathcal{F}_k x'\|_2)^2 &= (\|\widetilde{\mathcal{F}}_k x\|_2 - \|\widetilde{\mathcal{F}}_k x'\|_2)^2 \\ &= (|\langle x, \tilde{f}_{k,1}\rangle| - |\langle x', \tilde{f}_{k,1}\rangle|)^2 \\ &\quad + (|\langle x, \tilde{f}_{k,2}\rangle| - |\langle x', \tilde{f}_{k,2}\rangle|)^2 , \end{aligned}$$

which implies, by denoting $M(x) = (|\langle x, \widetilde{f}_{k,j}\rangle|)_{k,j}$, that $\|P_2(x) - P_2(x')\| = \|M(x) - M(x')\|$. Since $\widetilde{\mathcal{F}} \in \mathcal{Q}_2$, it results from Proposition 2.1 that

$$\begin{aligned} \|P_2(x) - P_2(x')\| &\geq d(x, x') \min_{S \subset \{1 \ldots m\}} \sqrt{\lambda_-^2(\widetilde{\mathcal{F}}_S) + \lambda_-^2(\widetilde{\mathcal{F}}_{S^c})} \\ &\geq d(x, x') A_2 \ \square. \qquad (22) \end{aligned}$$

B.3. Proof of Corollary 2.5

Given x, x', let I denote the groups I_k, $k \leq K$ such that $s(x) \cap s(x') \cap I_k = I_k$. It results that

$$\begin{aligned} & \|R_p(x) - R_p(x')\|^2 \\ = & \sum_{k \in I} |R_p(x)_k - R_p(x')_k|^2 + \sum_{k \notin I} |R_p(x)_k - R_p(x')_k|^2 \\ \geq & \sum_{k \in I} |R_p(x)_k - R_p(x')_k|^2 + \sum_{k \notin I} (\|M_0(x)\big|_{I_k} - M_0(x')\big|_{I_k}\|)^2 , \end{aligned}$$

where $M_0(x)\big|_{I_k}$ denotes the restriction of the modulus operator $M_0(x) = |\mathcal{F}^T x|$ to the indices I_k.

On the groups in I we can apply the same arguments as in theorem 2.4, and hence find a frame $\tilde{\mathcal{F}}$ from the family $\widetilde{\mathcal{Q}}_{p,x,x'}$ such that

$$\|R_p(x) - R_p(x')\|_I = \|M(x) - M(x')\|_I ,$$

with $M(x) = (|\langle x, \widetilde{f}_{k,j}\rangle|)_{k \in I, j}$ and $\{\widetilde{f}_{k,j}\} \in \widetilde{\mathcal{Q}}_{p,x,x'}$, and where $\|\cdot\|_I$ denotes the restriction to the set of coordinates given by I. Then, by following the same arguments used previously, it results from the definition of $\tilde{A}_p$ that

$$\|R_p(x) - R_p(x')\| \geq \tilde{A}_p d(x, x') .$$

Finally, the upper Lipschitz bound is obtained by noting that

$$\|M_\alpha(x) - M_\alpha(x')\| \leq \|\mathcal{F}(x - x')\| ,$$

and using the same argument as in (21) $\square$.

B.4. Proof of Proposition 2.6

Let $x, x' \in \mathbb{R}^N$, and let $\mathcal{J} = s(x) \cap s(x')$. Suppose first that $\mathcal{C}_{s(x)} \cap \mathcal{C}_{s(x')} \neq \emptyset$. Since $\|P_\infty x - P_\infty\| \geq \||\mathcal{F}_s x| - |\mathcal{F} s x'|\big|_{\mathcal{J}}\|$, it results that

$$d(x, x') A_{s(x), s(x')} \leq \|P_\infty x - P_\infty x'\| \qquad (23)$$

by Proposition 2.1 and by definition (17).

Let us now suppose $\mathcal{C}_{s(x)} \cap \mathcal{C}_{s(x')} = \emptyset$, and let $z = P_\infty x - P_\infty x'$. It results that $z = |\mathcal{F}_{s(x)} x| - |\mathcal{F}_{s(x')} x'| \in \mathbb{R}^K$, and hence we can split the coordinates $(1 \dots K)$ into Ω, Ω^c such that

$$\begin{aligned} z\big|_\Omega &= \mathcal{F}_{s(x)}\big|_\Omega(x) - \mathcal{F}_{s(x')}\big|_\Omega(x') \,, \\ z\big|_{\Omega^c} &= \mathcal{F}_{s(x)}\big|_{\Omega^c}(x) + \mathcal{F}_{s(x')}\big|_{\Omega^c}(x') \,. \end{aligned}$$

We shall concentrate in each restriction independently. Since $\mathcal{F}_{s(x')}\big|_\Omega(x') \in \mathcal{F}_{s(x')}\big|_\Omega(\mathcal{C}_{s(x')})$, it results that

$$\begin{aligned} \|z\big|_\Omega\| &\geq \inf_{y \in \mathcal{F}_{s(x')}\big|_\Omega} \|\mathcal{F}_{s(x)}\big|_\Omega(x) - y\| \\ &\geq \|\mathcal{F}_{s(x)}\big|_\Omega(x)\| \cdot |\sin(\beta(s(x), s(x'), \Omega))| \quad (24) \end{aligned}$$

Since by definition

$$\forall\, k \,, \; \sum_{j \in I_k} |\langle x, f_j \rangle|^2 \leq \frac{1}{|I_k|} |\langle x, f_{s(x)_k} \rangle|^2 \,,$$

it results, assuming without loss of generality that all pools have equal size ($|I_k| = \frac{M}{K}$),

$$\begin{aligned} \forall\, x \in \mathcal{C}_s \,, \; \|\mathcal{F}_{s(x)}\big|_\Omega(x)\| &\geq \sqrt{\frac{K}{M}} \|\mathcal{F}\big|_\Omega(x)\| \\ &\geq \sqrt{\frac{K}{M}} \lambda_-(\mathcal{F}\big|_\Omega) \|x\| \quad (25) \end{aligned}$$

Equivalently, since $\mathcal{F}_{s(x)}\big|_\Omega(x) \in \mathcal{F}_{s(x)}\big|_\Omega(\mathcal{C}_{s(x)})$ we also have

$$\|z\big|_\Omega\| \geq \sqrt{\frac{K}{M}} \lambda_-(\mathcal{F}\big|_\Omega) \cdot |\sin(\beta(s(x), s(x'), \Omega))| \cdot \|x'\| \,. \quad (26)$$

It follows that

$$\begin{aligned} \|z\big|_\Omega\| &\geq \sqrt{\frac{K}{M}} \lambda_-(\mathcal{F}\big|_\Omega) \,|\sin(\beta(s(x), s(x'), \Omega))| \max(\|x\|, \|x'\|) \\ &\geq \sqrt{\frac{K}{4M}} \lambda_-(\mathcal{F}\big|_\Omega) \,|\sin(\beta(s(x), s(x'), \Omega))| \, d(x, x') \,. \quad (27) \end{aligned}$$

By aggregating the bound for Ω and Ω^c we obtain (29) $\square$.

B.4.1. Maxout

These results easily extend to the so-called Maxout operator (Goodfellow et al., 2013), defined as $x \mapsto MO(x) = \{\max_{j \in I_k} \langle x, f_j \rangle \,;\, k = 1 \dots K\}$. By redefining the switches of x as

$$s(x) = \{j \,;\, \langle x, f_j \rangle > \max(\langle x, f_{j'} \rangle \,;\, \forall\, j' \in \text{pool}(j)\} \,, \quad (28)$$

the following corollary computes a Lower Lipschitz bound of $MO(x)$:

Corollary B.1 *The Maxout operator MO satisfies (20) with $A(s, s')$ defined using the switches (28).*

B.4.2. ℓ_1 Pooling

Propostion 2.6 can be used to obtain a bound of the lower Lipschitz constant of the ℓ_1 pooling operator.

Observe that for $x \in \mathbb{R}^n$,

$$\|x\|_1 = \sum_i |x_i| = \max_{\epsilon_i = \pm 1} |\langle x, \epsilon \rangle| \,.$$

It results that $P_1(x; \mathcal{F}) \equiv P_\infty(x; \widetilde{\mathcal{F}})$, with

$$\widetilde{\mathcal{F}} = (\tilde{f}_{k,\epsilon} = \sum_i \epsilon(i) f_{k,i} \,;\, k = 1 \dots, K \,;\, \epsilon \in \{-1, 1\}^L\} \,.$$

Each pool $\widetilde{\mathcal{F}}_k$ can be rewritten as $\widetilde{\mathcal{F}}_k = H_L \mathcal{F}_k$, where H_L is the $L \times 2^L$ Hadamard matrix whose rows contain the ϵ vectors. One can verify that $H_L^T H_L = 2^L \mathbf{1}$, which implies that for any $\Omega \subseteq \{1 \dots K\}$, $\lambda_-(\widetilde{\mathcal{F}}\big|_\Omega) = 2^{L/2} \lambda_-(\mathcal{F}\big|_\Omega)$. It results that

Corollary B.2 *The ℓ_1 pooling operator P_1 satisfies*

$$\forall\, x, x' \,,\; d(x, x') \left(\min_{s, s'} \tilde{A}(s, s') \right) \leq \|P_1(x) - P_1(x')\| \,, \quad (29)$$

where $d(x, x') = \min(\|x - x'\|, \|x + x'\|)$ and

$$\begin{aligned} \tilde{A}(s, s') = \max \Big\{ &\min_{\Omega \subseteq \mathcal{J}(s,s')} \sqrt{\lambda_-^2(\widetilde{\mathcal{F}}_\Omega) + \lambda_-^2(\widetilde{\mathcal{F}}_{\mathcal{J}-\Omega})} \,, \\ &\frac{1}{2} \min_{\Omega \subseteq \{1 \dots K\}} \sqrt{\Lambda_{s,s',\Omega}^2 + \Lambda_{s,s',\Omega^c}^2} \Big\} \,, \end{aligned}$$

with s, s' and $\beta(s, s')$ are defined on the frame $\widetilde{\mathcal{F}}$.

Similarly as in Corollary 2.7, one can obtain a similar bound for the Rectified ℓ_1 pooling.

B.5. Proof of Corollaries 2.7 and B.1

The result follows immediately from Proposition 2.6, by replacing the phaseless invertibility condition of Propostion 2.1 by the one in Proposition 2.2. $\square$.

B.6. Proof of Proposition 2.8

Proposition 2.8 also extends to the maxout case. We restate it here with the extra result:

Proposition B.3 *Let $\mathcal{F} = (f_1, \dots, f_M)$ be a random frame of $\mathbb{R}^N$, organized into K disjoint pools of dimension L. Then these statements hold with probability 1:*

1. *P_p is injective (modulo $x \sim -x$) if $K \geq 4N$ for $p = 1, \infty$, and if $K \geq 2N - 1$ for $p = 2$.*
2. *The Maxout operator MO is injective if $K \geq 2N + 1$.*

Let us first prove (i), with $p=\infty$. Let x, $x' \in \mathbb{R}^N$ such that $P_\infty(x) = P_\infty(x')$, and let $s = s(x)$, $s' = s(x')$. The set of K pooling measurements is divided into the intersection $\mathcal{J}(s,s') = \{k\,;\, s(x)_k = s(x')_k\}$ and its complement $\mathcal{J}(s,s')^c = \{k\,;\, s(x)_k \neq s(x')_k\}$. Suppose first that $|\mathcal{J}(s,s')| \geq 2N-1$. Then it results that we can pick $d = \lceil \frac{|\mathcal{J}(s,s')|}{2} \rceil \geq N$ elements of $\mathcal{J}(s,s')$ to form a frame V, such that either $x - x' \in \text{Ker}(V)$ or $x + x \in \text{Ker}(V)$. Since a random frame of dimension $\geq N$ spans $\mathbb{R}^N$ with probability 1, it results that $x = \pm x'$. Suppose otherwise that $|\mathcal{J}(s,s')| < 2N-1$. It follows that $|\mathcal{J}(s,s')^c| \geq 2N+1$, and hence that any partition of $\mathcal{J}(s,s')^c$ into two frames will contain always a frame $\mathcal{F}\big|_\Omega$ with at least $N+1$ columns. Since two random subspaces of dimension N in $\mathbb{R}^M$ have nonzero largest principal angle with probability 1 as long as $K > N$, it results that $\Lambda_{s,s',\Omega} > 0$ and hence that $\text{Prob}(|P_\infty(x)\big|_{\mathcal{J}(s,s')^c}| = |P_\infty(x')\big|_{\mathcal{J}(s,s')^c}|) = 0$. The case $p=1$ is proved identically thanks to Corollary B.2.

Finally, in order to prove (ii) we follow the same strategy. If $|\mathcal{J}(s,s')| \geq N$, then $MO(x) = MO(x') \Rightarrow x = x'$ with probability 1 since $\mathcal{F}\big|_\mathcal{J}$ spans $\mathbb{R}^N$ with probability 1. Otherwise it results that $|\mathcal{J}(s,s')^c| \geq N+1$, which implies $MO(x) \neq MO(x')$, since two random subspaces of dimension N in $\mathbb{R}^{|\mathcal{J}(s,s')^c|}$ have 0 intersection with probability 1.

Let us now prove the case $p=2$. We start drawing a random basis for each of the pools $F_1, \dots, F_K$. From proposition 2.4, it follows that we have to check that if $M \geq 2N$, the quantity

$$\min_{F' = U\,F\,,\,U^T U = \mathbf{1}} \min_{\Omega \subseteq \{1 \dots M\}} \lambda_-^2(F'_\Omega) + \lambda_-^2(F'_{\Omega^c}) > 0$$

with probability 1. If $M \geq 2N-1$, it follows that either Ω has the property that it intersects at least N pools, either Ω^c intersects N pools. Say it is Ω. Now, for each pool with nonzero intersection, say F_k, we have that

$$\|(F'_k)^T y\| \geq \frac{1}{\sqrt(L)} |\langle f_{k,j}, y \rangle|$$

for some $f_{k,j}$ belonging to the initial random basis of F_k. It results that

$$\lambda_-^2(F'_\Omega) \geq \frac{1}{\sqrt(L)} \lambda_-^2(F^*) \ ,$$

where F^* is a subset of N columns of the original frame $\mathcal{F}$, which means

$$\lambda_-^2(F'_\Omega) \geq \frac{1}{\sqrt(L)} \lambda_-^2(F*) > 0 \ .$$

$\square$.