ℓ_1-norm minimization

Charles J. Garfinkle

Theorem 1 (Noiseless recovery)**.** *Suppose the $n \times m$ matrix $\mathbf{A}$ has $\|\mathbf{A}_j\|_2 = 1$ for $j = 1, \ldots, m$ and $L_2(\mathbf{A}) > 0$. Let $\mathbf{x}_i \in \mathbb{R}^m$ be such that $\mathbf{x}_i = c_i \mathbf{e}_i$ ($c_i \neq 0$) for $i = 1, \ldots, m$. Suppose $n \times m'$ matrix $\mathbf{B}$ with $\|\mathbf{B}_j\|_2 = 1$ for $j = 1, \ldots, m'$ and vectors $\overline{\mathbf{x}}_i, \ldots, \overline{\mathbf{x}}_m$ together solve:*

$$\min \sum_{i=1}^{m} \|\overline{\mathbf{x}}_i\|_1 \quad \textit{subject to} \quad \mathbf{B}\overline{\mathbf{x}}_i = \mathbf{A}\mathbf{x}_i \quad \textit{for } i = 1, \ldots, m. \tag{1}$$

Then $\mathbf{A} = \mathbf{B}_S \mathbf{P}$ for some $S \subseteq [m']$ of size m and $m \times m$ permutation matrix $\mathbf{P}$.

Proof. Fixing i, we have

$$\|\mathbf{x}_i\|_1 = |c_i| \|\mathbf{A}\mathbf{e}_i\|_2 = \|\mathbf{A}\mathbf{x}_i\|_2 = \|\mathbf{B}\overline{\mathbf{x}}_i\|_2 \leq \sum_j |\bar{c}_j| \|\mathbf{B}\mathbf{e}_j\|_2 = \|\overline{\mathbf{x}}_i\|_1 \tag{2}$$

so that $\|\overline{\mathbf{x}}_i\|_1 = \|\mathbf{x}_i\|_1 + \varepsilon_i$ for some $\varepsilon_i \geq 0$. But also,

$$\sum_i \|\mathbf{x}_i\|_1 \geq \sum_i \|\overline{\mathbf{x}}_i\|_1 = \sum_i \left(\|\mathbf{x}_i\|_1 + \varepsilon_i\right) \tag{3}$$

Hence $\sum_i \varepsilon_i \leq 0$. But since every ε_i is non-negative, it must but that $\varepsilon_i = 0$ for $i = 1, \ldots, m$. Thus,

$$\|\overline{\mathbf{x}}_i\|_1 = \|\mathbf{x}_i\|_1 = \|\mathbf{B}\overline{\mathbf{x}}_i\|_2. \tag{4}$$

Lemma 1. *If $\|\mathbf{T}\mathbf{v}\|_2 = \|\mathbf{v}\|_1$ for $\mathbf{T}$ with $\|\mathbf{T}_j\|_2 = 1$ for all j and $|\mathbf{T}_i \cdot \mathbf{T}_j| < 1$ for all i, j then $\mathbf{v}$ has at most one non-zero entry.*

Proof. Let $\mathbf{v} = \sum_j c_j e_j$. Then,

$$\mathbf{T}\mathbf{v} = \sum_j c_j \mathbf{T}_j = \sum_j c_j \sum_i \mathbf{T}_{ij} \mathbf{e}_i = \sum_i \sum_j c_j \mathbf{T}_{ij} \mathbf{e}_i = \sum_i d_i \mathbf{e}_i \quad \text{where } d_i = \sum_j c_j \mathbf{T}_{ij} \tag{5}$$

So that

$$\begin{aligned}
\|\mathbf{T}\mathbf{v}\|_2^2 = \sum_i d_i^2 &= \sum_i \left(\sum_j c_j \mathbf{T}_{ij} \right)^2 && (6)\\
&= \sum_i \left(\sum_j \left(c_j \mathbf{T}_{ij}\right)^2 + \sum_j \sum_{k \neq j} \left(c_j \mathbf{T}_{ij}\right)\left(c_k \mathbf{T}_{ik}\right) \right) && (7)\\
&= \sum_j c_j^2 \sum_i \mathbf{T}_{ij}^2 + \sum_j \sum_{k \neq j} c_j c_k \sum_i \mathbf{T}_{ij} \mathbf{T}_{ik} && (8)\\
&= \sum_j c_j^2 + \sum_j \sum_{k \neq j} c_j c_k \left(\mathbf{T}_j \cdot \mathbf{T}_k\right) && (9)
\end{aligned}$$

where we have applied the formula $\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_i \sum_{j \neq i} a_i a_j$. Next, observe

$$\|\mathbf{v}\|_1^2 = \left(\sum_j |c_j| \right)^2 = \sum_j c_j^2 + \sum_j \sum_{k \neq j} |c_j||c_k| \tag{10}$$

so that$\|\mathbf{T}\mathbf{v}\|_2^2 = \|\mathbf{v}\|_1$ implies $\sum_i \sum_{j \neq i} \alpha_{ij} = 0$ for $\alpha_{ij} = |c_i||c_j| - c_i c_j \left(\mathbf{T}_i \cdot \mathbf{T}_j\right)$. Now,

$$\alpha_{ij} \geq |c_i||c_j| - |c_i c_j \left(\mathbf{T}_i \cdot \mathbf{T}_j\right)| = |c_i||c_j| \left(1 - |\mathbf{T}_i \cdot \mathbf{T}_j|\right) \geq 0 \tag{11}$$

since $\|\mathbf{T}_j\|_2 = 1$ for all j. Thus $\alpha_{ij} = 0$ for all i, j and we have $|c_i||c_j| = |c_i||c_j||\mathbf{T}_i \cdot \mathbf{T}_j|$, or $|c_i||c_j| \left(1 - |\mathbf{T}_i \cdot \mathbf{T}_j|\right) = 0$. Since $|\mathbf{T}_i \cdot \mathbf{T}_j| < 1$, this implies $|c_i||c_j| = 0$ for all i, j. Thus, $\mathbf{v}$ has at most one nonzero entry, since if $c_i \neq 0$ then $c_j = 0$ for all $j \neq i$. $\square$

By Lemma 1, either $\mathbf{B}$ has colinear columns or the $\overline{\mathbf{x}}_i$ are all 1-sparse vectors.

It is easy to see that if $\mathbf{B}$ has colinear columns, there exists some $S \subset [m']$ such that the submatrix $\mathbf{B}_S$ has no colinear columns and satisfies $\mathbf{B}_S \overline{\mathbf{x}}'_i = \mathbf{B}\overline{\mathbf{x}}_i$ for some $\overline{\mathbf{x}}'_i \in \mathbb{R}^{|S|}$ with $\|\overline{\mathbf{x}}'_i\|_1 = \|\overline{\mathbf{x}}_i\|_1$ for all i. Simply set $S = [m']$ and $\overline{\mathbf{x}}'_i = \overline{\mathbf{x}}_i$ and let $i = 1$. Iterating through $j \neq i$, if $|\mathbf{B}_i \cdot \mathbf{B}_j| = 1$, let $S \to S \setminus \{j\}$ and $\overline{c}'_i \to \overline{c}'_i + \overline{c}_j \cdot \text{sgn}(\mathbf{B}_i \cdot \mathbf{B}_j)$. Induct on i.

By the above claim, there is some $S \subseteq [m']$ such that $\mathbf{B}_S$ has no colinear columns (i.e. $L_2(\mathbf{B}) > 0$) and satisfies the assumptions of the theorem for some $\overline{\mathbf{x}}'_i \in \mathbb{R}^{|S|}$ for which $\|\overline{\mathbf{x}}'_i\|_1 = \|\overline{\mathbf{x}}_i\|_1$ for all i. By the lemma, then, the $\overline{\mathbf{x}}'_i$ must all have at most one non-zero entry (i.e. they are 1-sparse) and the result follows by application of the ℓ_0-norm theorem.

□

Theorem 2 (Noisy recovery). *Suppose the $n \times m$ matrix $\mathbf{A}$ has $\|\mathbf{A}_j\|_2 = 1$ for $j = 1, \ldots, m$ and $L_2(\mathbf{A}) > 0$. Let $\mathbf{x}_i \in \mathbb{R}^m$ be such that $\mathbf{x}_i = c_i \mathbf{e}_i$ ($c_i \neq 0$) for $i = 1, \ldots, m$. Suppose $n \times m'$ matrix $\mathbf{B}$ with $\|\mathbf{B}_j\|_2 = 1$ for $j = 1, \ldots, m'$ and vectors $\overline{\mathbf{x}}_i, \ldots, \overline{\mathbf{x}}_m$ together solve:*

$$\min \sum_{i=1}^{m} \|\overline{\mathbf{x}}_i\|_1 \quad \textit{subject to} \quad \|\mathbf{B}\overline{\mathbf{x}}_i - \mathbf{A}\mathbf{x}_i\|_2 \leq \varepsilon \quad \textit{for } i = 1, \ldots, m. \tag{12}$$

Then $\mathbf{A} = \mathbf{B}_S \mathbf{P}$ for some $S \subseteq [m']$ of size m and $m \times m$ permutation matrix $\mathbf{P}$.

Proof. By the reverse triangle inequality, we have for all i:

$$\varepsilon \geq \|\mathbf{A}\mathbf{x}_i - \mathbf{B}\overline{\mathbf{x}}_i\|_2 \geq |\|\mathbf{A}\mathbf{x}_i\| - \|\mathbf{B}\overline{\mathbf{x}}_i\|_2| \,. \tag{13}$$

So,

$$\|\overline{\mathbf{x}}_i\|_1 \geq \|\mathbf{B}\overline{\mathbf{x}}_i\|_2 \geq \|\mathbf{A}\mathbf{x}_i\|_2 - \varepsilon = |c_i| - \varepsilon. \tag{14}$$

It is trivial to show that letting $\mathbf{B} = \mathbf{A}$ and $\overline{\mathbf{x}}_i = \left(1 - \frac{\varepsilon}{|c_i|}\right) \mathbf{x}_i$ for all i is a particular solution satisfying the constraints, since then $\|\mathbf{B}\overline{\mathbf{x}}_i - \mathbf{A}\mathbf{x}_i\|_2 = \varepsilon$. Thus, a solution to the minimization problem must satisfy:

$$\sum_{i=1}^{m} \|\overline{\mathbf{x}}_i\|_1 \leq \sum_i (|c_i| - \varepsilon) \,. \tag{15}$$

Taken together with Eq. 14, this implies that $\|\overline{\mathbf{x}}_i\|_1 = |c_i| - \varepsilon$ for all i, thus $\|\overline{\mathbf{x}}_i\|_1 = \|\mathbf{B}\overline{\mathbf{x}}_i\|_2$ for all i. By Lemma 1, either $\mathbf{B}$ has colinear columns or the $\overline{\mathbf{x}}_i$ are all 1-sparse vectors.

It is easy to see that if $\mathbf{B}$ has colinear columns, there exists some $S \subset [m']$ such that the submatrix $\mathbf{B}_S$ has no colinear columns and satisfies $\mathbf{B}_S \overline{\mathbf{x}}'_i = \mathbf{B}\overline{\mathbf{x}}_i$ for some $\overline{\mathbf{x}}'_i \in \mathbb{R}^{|S|}$ with $\|\overline{\mathbf{x}}'_i\|_1 = \|\overline{\mathbf{x}}_i\|_1$ for all i. Simply set $S = [m']$ and $\overline{\mathbf{x}}'_i = \overline{\mathbf{x}}_i$ and let $i = 1$. Iterating through $j \neq i$, if $|\mathbf{B}_i \cdot \mathbf{B}_j| = 1$, let $S \to S \setminus \{j\}$ and $\overline{c}'_i \to \overline{c}'_i + \overline{c}_j \cdot \text{sgn}(\mathbf{B}_i \cdot \mathbf{B}_j)$. Induct on i.

By the above claim, there is some $S \subseteq [m']$ such that $\mathbf{B}_S$ has no colinear columns (i.e. $L_2(\mathbf{B}) > 0$) and satisfies the assumptions of the theorem for some $\overline{\mathbf{x}}'_i \in \mathbb{R}^{|S|}$ for which $\|\overline{\mathbf{x}}'_i\|_1 = \|\overline{\mathbf{x}}_i\|_1$ for all i. By the lemma, then, these $\overline{\mathbf{x}}'_i$ must all have at most one non-zero entry (i.e. they are 1-sparse). Therefore, there exist $\overline{c}_1, \ldots, \overline{c}_{\overline{m}}$ and a map $\pi : [m] \to [\overline{m}]$ such that

$$\|c_i \mathbf{A}_i - \overline{c}_i \mathbf{B}_{\pi(i)}\|_2 \leq \varepsilon \quad \text{for all } i \tag{16}$$

We could end here by applying the noisy ℓ_0-norm theorem; this would imply dictionary recovery up to an error commensurate with ε. Instead, we wil go further by using the fact that in this case we know $|\overline{c}_i| = |c_i| - \varepsilon$ for all i.

Lemma 2. $\|\mathbf{u} - \mathbf{v}\|_2 \leq \varepsilon$ *for* $\|\mathbf{v}\|_2 = \|\mathbf{u}\|_2 - \varepsilon \implies \|\mathbf{v}\|_2 \mathbf{u} = \|\mathbf{u}\|_2 \mathbf{v}$.

Proof. Observe that

$$\begin{aligned}
\varepsilon^2 &\geq \|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle && (17)\\
&= \|\mathbf{u}\|_2^2 + (\|\mathbf{u}\|_2 - \varepsilon)^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle && (18)\\
&= 2\|\mathbf{u}\|_2^2 - 2\|\mathbf{u}\|_2 \varepsilon + \varepsilon^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle && (19)\\
&\implies \langle \mathbf{u}, \mathbf{v} \rangle \geq \|\mathbf{u}\|_2 (\|\mathbf{u}\|_2 - \varepsilon) = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 && (20)
\end{aligned}$$

But $\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$ always, thus $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$. Therefore,

$$\|\|\mathbf{u}\|_2 \mathbf{v} - \|\mathbf{v}\|_2 \mathbf{u}\|_2^2 = (\|\mathbf{u}\|_2 \|\mathbf{v}\|_2)^2 + (\|\mathbf{v}\|_2 \|\mathbf{u}\|_2)^2 - 2\|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \langle \mathbf{u}, \mathbf{v} \rangle = 0 \tag{21}$$

So $\|\mathbf{u}\|_2 \mathbf{v} = \|\mathbf{v}\|_2 \mathbf{u}$. □

By Lemma 2 (i.e. with $c_i \mathbf{A}_i = \mathbf{u}$ and $\overline{c}_i \mathbf{B}_{\pi(i)} = \mathbf{v}$), we have for all i that $|c_i| \overline{c}_i \mathbf{B}_{\pi(i)} = |\overline{c}_i| c_i \mathbf{A}_i$, or

$$\overline{c}'_i \mathbf{B}_{\pi(i)} = c_i \mathbf{A}_i \quad \text{for all } i \tag{22}$$

for $\overline{c}'_i = \text{sign}(\overline{c}_i) \cdot |c_i|$, and the result follows by application of the noiseless ℓ_0-norm theorem. (Alternatively, we may directly infer $\mathbf{B}_{\pi(i)} = \text{sign}(c_i \overline{c}_j) \mathbf{A}_i$) and apply only the argument of the the ℓ_0-norm proof establishing that π is a permutation.) □