Sparse Recovery by means of Nonnegative Least Squares

Citation	Foucart, S., & Koslicki, D. (2014). Sparse Recovery by Means of Nonnegative Least Squares. IEEE Signal Processing Letters, 21(4), 498-502. doi:10.1109/LSP.2014.2307064
DOI	10.1109/LSP.2014.2307064
Publisher	Institute of Electrical and Electronics Engineers
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Sparse Recovery by means of Nonnegative Least Squares

Simon Foucart and David Koslicki

***Abstract*—This short note demonstrates that sparse recovery can be achieved by an ℓ_1-minimization ersatz easily implemented using a conventional nonnegative least squares algorithm. A connection with orthogonal matching pursuit is also highlighted. The preliminary results call for more investigations on the potential of the method and on its relations to classical sparse recovery algorithms.**

***Index Terms*—compressive sensing, sparse recovery, ℓ_1-minimization, orthogonal matching pursuit, nonnegative least squares, adjacency matrices of bipartite graphs, k-mer frequency matrices, Gaussian matrices.**

Throughout this note, we consider the standard compressive sensing problem, i.e., the recovery of a sparse (possibly just almost sparse) vector $\mathbf{x} \in \mathbb{R}^N$ from the mere knowledge of its measurement vector (possibly corrupted by noise) $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ with $m < N$. The measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is perfectly known. Among popular methods to produce an estimate for $\mathbf{x}$, we single out the very popular

• *basis pursuit:*

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}; \tag{BP}$$

• *orthogonal matching pursuit:* construct sequences (S^n) of support sets and $(\mathbf{x}^n)$ of vectors from $S^0 = \emptyset$, $\mathbf{x}^0 = \mathbf{0}$, and iterate, for $n \geq 1$,

$$S^n = S^{n-1} \cup \{j^n\}, \text{ with } j^n \text{ chosen as} \tag{OMP$_1$}$$
$$|\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1})_{j^n}| = \max_{1 \leq j \leq N} |\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1})_j|,$$
$$\mathbf{x}^n = \operatorname*{argmin}_{\mathbf{z} \in \mathbb{R}^N} \{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^n\}. \tag{OMP$_2$}$$

We will introduce a seemingly (and surprisingly) new method which has analogies with both (BP) and (OMP). We start by dealing with nonnegative sparse vectors before considering arbitrary sparse vectors.

I. Nonnegative Sparse Recovery

A. Exact Measurements

In a number of practical situations, the sparse vector $\mathbf{x} \in \mathbb{R}^N$ to be recovered from $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ has nonnegative entries. Although ordinary basis pursuit is sometimes used (see [2]), it can be modified to incorporate the nonnegativity constraint, leading to

• *nonnegative basis pursuit:*

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y} \text{ and } \mathbf{z} \geq \mathbf{0}. \tag{NNBP}$$

It is folklore[1] — but not well-known enough — that, for certain matrices $\mathbf{A}$ (including frequency matrices, see (1) below), if $\mathbf{x}$ is the unique solution of (BP) or the unique solution of (NNBP), then it is in fact the unique solution of the feasibility problem

$$\text{find } \mathbf{z} \in \mathbb{R}^N \text{ such that } \mathbf{A}\mathbf{z} = \mathbf{y} \text{ and } \mathbf{z} \geq \mathbf{0}. \tag{F}$$

Appendix A contains the proofs of this fact and of a few others. For frequency matrices, it is therefore irrelevant to try and recover nonnegative vectors via ℓ_1-minimization, since we might as well minimize a constant function subject to the constraints $\mathbf{A}\mathbf{z} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$. We may also consider solving the problem

• *nonnegative least squares:*

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{subject to } \mathbf{z} \geq \mathbf{0}. \tag{NNLS}$$

This minimization program is directly implemented in MATLAB as the function `lsqnonneg`. It executes the active-set algorithm of Lawson and Hanson (see [14, Chapter 23]), which seems particularly adapted to sparse recovery. Indeed, it is structured as

• *Lawson–Hanson algorithm:* construct sequences (S^n) of support sets and $(\mathbf{x}^n)$ of vectors from $S^0 = \emptyset$, $\mathbf{x}^0 = \mathbf{0}$, and iterate, for $n \geq 1$,

$$S^n = S^{n-1} \cup \{j^n\}, \text{ with } j^n \text{ chosen as} \tag{LH$_1$}$$
$$\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1})_{j^n} = \max_{1 \leq j \leq N} \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1})_j,$$
$$\widetilde{\mathbf{x}}^n = \operatorname*{argmin}_{\mathbf{z} \in \mathbb{R}^N} \{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^n\}, \tag{LH$_2$}$$

if $\widetilde{\mathbf{x}}^n \geq \mathbf{0}$, then one sets $\mathbf{x}^n$ to $\widetilde{\mathbf{x}}^n$,

otherwise, an inner loop reduces the active set S^n until $\widetilde{\mathbf{x}}^n \geq \mathbf{0}$ and then one sets $\mathbf{x}^n$ to $\widetilde{\mathbf{x}}^n$.

Aside from the inner loop, there is a peculiar analogy between this algorithm (dated from 1974) and orthogonal matching pursuit (usually attributed to [16] in 1993). In the reproducible MATLAB file accompanying this note, the benefit of the inner loop is revealed by the frequencies of successful nonnegative recovery via (NNLS) and (OMP).

This work is supported by the NSF under grant DMS-1120622.
S. Foucart is with the Department of Mathematics, University of Georgia, Athens, GA 30602, USA (e-mail: foucart@math.uga.edu).
D. Koslicki is with the Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA (e-mail: david.koslicki@math.oregonstate.edu).
Manuscript last updated February 14, 2014.

[1] see for instance [5] which contains results along these lines

Usual conditions ensuring sparse recovery via (BP) and (OMP) involve restricted isometry constants of the matrix $\mathbf{A}$ (see e.g. [10, chapter 6]). The guarantees are valid for all sparse vectors $\mathbf{x} \in \mathbb{R}^N$ simultaneously, in particular for all nonnegative sparse vectors $\mathbf{x} \in \mathbb{R}^N$. Under these conditions, from the strict point of view of recovery success (i.e., disregarding computational effort), there is nothing to be gained from solving (NNBP) or (F) instead of (BP). Even for (renormalized) adjacency matrices of lossless expanders, which do not possess the restricted isometry property (see [7]), the sparse recovery of all sparse vectors $\mathbf{x} \in \mathbb{R}^N$ via (BP) is ensured (see [3] or [10, chapter 13]). Since these are frequency matrices, the recovery of all nonnegative sparse vectors $\mathbf{x} \in \mathbb{R}^N$ is guaranteed via (NNBP) and (F), too, but again no gain in terms of recovery success arises. Nonetheless, there are empirical examples where nonnegative sparse recovery via (NNBP) and (F) succeeds, while arbitrary sparse recovery via (BP) fails. One such example is provided by the k-mer frequency matrix of DNA sequences. Its columns are indexed by organisms in a database and its rows are indexed by all 4^k possible DNA subwords of length k. It is obtained after column-normalizing the matrix whose (i,j)th entry counts the number of occurrence of the ith subword in the DNA sequence of the jth organism. We supply such a k-mer frequency matrix in the reproducible MATLAB file and we verify (using the optimization package CVX [11]) that all nonnegative sparse vectors supported on a given set S are recovered via (NNBP) — or equivalently via (F) or (NNLS) — while arbitrarily signed sparse vectors supported on S may not be recovered by (BP).

B. Inaccurate Measurements

Let us now consider the more realistic situation where the acquisition of the vector $\mathbf{x} \in \mathbb{R}^N$ is not perfect, so that the measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{R}^m$ contains an error term $\mathbf{e} \in \mathbb{R}^m$. If we know a reasonable bound $\|\mathbf{e}\|_2 \le \eta$, a common strategy for the approximate recovery of $\mathbf{x}$ is to solve the inequality-constrained basis pursuit consisting of minimizing $\|\mathbf{z}\|_1$ subject to $\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \le \eta$. Another common strategy is the LASSO (see [17]) consisting of minimizing $\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2$ subject to $\|\mathbf{z}\|_1 \le \tau$. Ordinary LASSO is sometimes used for nonnegative recovery (as in [1]), but one can also incorporate the constraint $\mathbf{z} \ge \mathbf{0}$, leading to the so-called nonnegative garrote (see [4]). A third strategy (intimately linked to the previous two) is basis pursuit denoising (see [8]) consisting of minimizing $\|\mathbf{z}\|_1 + \nu\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2$. For nonnegative recovery, incorporating an additional nonnegativity constraint has been employed e.g. in [15], [13]. Here, we replace the ℓ_1-norm in the objective function by a squared ℓ_1-norm (which is arguably more natural), i.e., we consider the program

- *ℓ_1-squared nonnegative regularization:*

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{z}\|_1^2 + \lambda^2 \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2 \quad \text{subject to } \mathbf{z} \ge \mathbf{0} \quad \text{(NNREG)}$$

for some large $\lambda > 0$. Our main observation (already exploited fruitfully in [12]) is that (NNREG) can be recast as a nonnegative least squares problem, for which the algorithm of Lawson and Hanson is particularly well adapted. Indeed, since $\mathbf{z} \ge \mathbf{0}$, we have $\|\mathbf{z}\|_1 = \sum_{j=1}^N z_j$, so (NNREG) takes the form

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\widetilde{\mathbf{A}}\mathbf{z} - \widetilde{\mathbf{y}}\|_2^2 \quad \text{subject to } \mathbf{z} \ge \mathbf{0}$$

where $\widetilde{\mathbf{A}} \in \mathbb{R}^{(m+1)\times N}$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^{m+1}$ are defined by

$$\widetilde{\mathbf{A}} = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline & \lambda\mathbf{A} & \end{array}\right] \text{ and } \widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda\mathbf{y} \end{array}\right].$$

As exhibited in the reproducible file, the execution time compares favorably to state-of-the-art algorithms for ℓ_1-minimization (we selected YALL1 [20]) but the choice of the parameter λ is critical in our current implementation.

C. Theoretical Justifications

To support the previous heuristics, we prove that, as $\lambda \to \infty$, the solutions $\mathbf{x}^\lambda$ of (NNREG) converge to the solution $\mathbf{x}^\sharp$ of (NNBP) — assumed to be unique. This is done for matrices $\mathbf{A}$ with nonnegative entries and all column-sums equal to one, i.e.,

$$A_{i,j} \ge 0 \text{ for all } i,j \quad \text{and} \quad \sum_{i=1}^m A_{i,j} = 1 \text{ for all } j. \quad (1)$$

We call such matrices frequency matrices. Examples are given by k-mer frequency matrices and by (renormalized) adjacency matrices of left-regular bipartite graphs. For $t \in [0,1]$, the vector $(1-t)\mathbf{x}^\sharp + t\mathbf{x}^\lambda$ is nonnegative, so the minimality of $\mathbf{x}^\lambda$ yields

$$\begin{aligned} &\|\mathbf{x}^\lambda\|_1^2 + \lambda^2\|\mathbf{A}\mathbf{x}^\lambda - \mathbf{y}\|_2^2 \\ &\quad \le \|(1-t)\mathbf{x}^\sharp + t\mathbf{x}^\lambda\|_1^2 + \lambda^2\|\mathbf{A}((1-t)\mathbf{x}^\sharp + t\mathbf{x}^\lambda) - \mathbf{y}\|_2^2. \end{aligned}$$

The triangle inequality and the fact that $\mathbf{y} = \mathbf{A}\mathbf{x}^\sharp$ give

$$\begin{aligned} &\|\mathbf{x}^\lambda\|_1^2 + \lambda^2\|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_2^2 \\ &\quad \le ((1-t)\|\mathbf{x}^\sharp\|_1 + t\|\mathbf{x}^\lambda\|_1)^2 + \lambda^2 t^2 \|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_2^2. \end{aligned}$$

After rearrangement, we obtain

$$\begin{aligned} \lambda^2(1-t^2)\|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_2^2 &\le (1-t)(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1) \\ &\quad \times ((1-t)\|\mathbf{x}^\sharp\|_1 + (1+t)\|\mathbf{x}^\lambda\|_1). \quad (2) \end{aligned}$$

The left-hand side being nonnegative, we notice that $\|\mathbf{x}^\lambda\|_1 \le \|\mathbf{x}^\sharp\|_1$. We also take into account that

$$\begin{aligned} &\|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_2 \ge \frac{1}{\sqrt{m}}\|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_1 \\ &= \frac{1}{\sqrt{m}} \sum_{i=1}^m \Big| \sum_{j=1}^N A_{i,j}(x_j^\lambda - x_j^\sharp) \Big| \ge \frac{1}{\sqrt{m}} \sum_{j=1}^N \sum_{i=1}^m A_{i,j}(x_j^\sharp - x_j^\lambda) \\ &= \frac{1}{\sqrt{m}} \sum_{j=1}^N (x_j^\sharp - x_j^\lambda) = \frac{1}{\sqrt{m}}(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1). \end{aligned}$$

Substituting the latter inequality into (2) and dividing by $(1-t)(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1) \geq 0$ implies that

$$\frac{\lambda^2(1+t)}{m}(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1) \leq (1-t)\|\mathbf{x}^\sharp\|_1 + (1+t)\|\mathbf{x}^\lambda\|_1,$$

that is to say

$$\Big(\frac{\lambda^2}{m} - \frac{1-t}{1+t}\Big)\|\mathbf{x}^\sharp\|_1 \leq \Big(1+\frac{\lambda^2}{m}\Big)\|\mathbf{x}^\lambda\|_1.$$

With the optimal choice $t=1$, we derive the estimates

$$\frac{\lambda^2/m}{1+\lambda^2/m}\|\mathbf{x}^\sharp\|_1 \leq \|\mathbf{x}^\lambda\|_1 \leq \|\mathbf{x}^\sharp\|_1, \tag{3}$$

which shows that $\|\mathbf{x}^\lambda\|_1 \to \|\mathbf{x}^\sharp\|_1$ as $\lambda \to \infty$. From here, we can prove that $\mathbf{x}^\lambda \to \mathbf{x}^\sharp$ as $\lambda \to \infty$. Indeed, if not, there would exist a number of $\varepsilon > 0$ and an increasing sequence (λ_n) such that $\|\mathbf{x}^{\lambda_n} - \mathbf{x}^\sharp\|_1 \geq \varepsilon$ for all n. Since the sequence $(\mathbf{x}^{\lambda_n})$ is bounded, we can extract a subsequence $(\mathbf{x}^{\lambda_{n_k}})$ converging to some $\mathbf{x}^\star \in \mathbb{R}^N$. We note that $\|\mathbf{x}^\star - \mathbf{x}^\sharp\|_1 \geq \varepsilon$. We also note that $\mathbf{x}^\star \geq \mathbf{0}$ and $\|\mathbf{x}^\star\|_1 = \|\mathbf{x}^\sharp\|_1$. Moreover, the inequality

$$\begin{aligned}\|\mathbf{x}^{\lambda_{n_k}}\|_1^2 + \lambda_{n_k}^2\|\mathbf{A}\mathbf{x}^{\lambda_{n_k}} - \mathbf{y}\|_2^2 &\leq \|\mathbf{x}^\sharp\|_1^2 + \lambda_{n_k}^2\|\mathbf{A}\mathbf{x}^\sharp - \mathbf{y}\|_2^2 \\ &= \|\mathbf{x}^\sharp\|_1^2\end{aligned}$$

passed to the limit yields $\mathbf{A}\mathbf{x}^\star = \mathbf{y}$. By uniqueness of $\mathbf{x}^\sharp$ as a minimizer of $\|\mathbf{z}\|_1$ subject to $\mathbf{A}\mathbf{z} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$, we deduce that $\mathbf{x}^\star = \mathbf{x}^\sharp$, which contradicts $\|\mathbf{x}^\star - \mathbf{x}^\sharp\|_1 \geq \varepsilon$. We conclude that $\mathbf{x}^\lambda$ does converge to $\mathbf{x}^\sharp$, as announced. The reproducible file includes experimental evidence that the left-hand side of (3) captures the correct behavior of $(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1)/\|\mathbf{x}^\sharp\|_1$ as a function of the parameter λ.

II. Arbitrarily Signed Sparse Recovery

We consider in this section the sparse recovery of arbitrary vectors, not necessarily nonnegative ones. We build on the previous considerations to introduce a nonnegative-least-squares surrogate for basis pursuit. The method does not require knowledge of convex optimization to be implemented — it is easily written in MATLAB and its execution time is rather fast.

A. Rationale

The crucial point is again to replace the ℓ_1-norm in basis pursuit denoising by its square, thus introducing

- *ℓ_1-squared regularization:*

$$\min_{\mathbf{z}\in\mathbb{R}^N} \|\mathbf{z}\|_1^2 + \lambda^2\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2. \tag{REG}$$

Next, decomposing any $\mathbf{z} \in \mathbb{R}^N$ as $\mathbf{z} = \mathbf{z}_+ - \mathbf{z}_-$ with $\mathbf{z}_+ \geq \mathbf{0}$ and $\mathbf{z}_- \geq \mathbf{0}$, (REG) is transformed into the nonnegative least squares problem

$$\min_{\widetilde{\mathbf{z}}\in\mathbb{R}^{2N}} \|\widetilde{\mathbf{A}}\widetilde{\mathbf{z}} - \widetilde{\mathbf{y}}\|_2 \quad \text{subject to } \widetilde{\mathbf{z}} \geq \mathbf{0},$$

where $\widetilde{\mathbf{A}} \in \mathbb{R}^{(m+1)\times 2N}$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^{m+1}$ are defined by

$$\widetilde{\mathbf{A}} = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \hline & \lambda\mathbf{A} & & & -\lambda\mathbf{A} & \end{array}\right], \quad \widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda\mathbf{y}\end{array}\right].$$

To see the equivalence of the optimization problems, we simply observe that the objective function in (REG) is, with $\widetilde{\mathbf{z}} = \left[\begin{array}{c}\mathbf{z}_+ \\ \hline \mathbf{z}_-\end{array}\right]$,

$$(\|\mathbf{z}_+\|_1 + \|\mathbf{z}_-\|_1)^2 + \lambda^2\|\mathbf{A}\mathbf{z}_+ - \mathbf{A}\mathbf{z}_- - \mathbf{y}\|_2^2 = \|\widetilde{\mathbf{A}}\widetilde{\mathbf{z}} - \widetilde{\mathbf{y}}\|_2^2.$$

B. Theoretical Justifications

We support the previous heuristics by a result valid for random matrices $\mathbf{A} \in \mathbb{R}^{m\times N}$ whose entries are independent Gaussian variables with mean zero and variance $1/m$. With high probability, these matrices satisfy both the ℓ_1-quotient property and the restricted isometry property of optimal order $s \asymp m\ln(N/m)$. This implies (see [18] or [10, chapter 11]) that, for any $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$ in $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, the solution $\mathbf{x}^\sharp$ of (BP) approximates $\mathbf{x}$ with an error controlled as

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_1 \leq C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\|\mathbf{e}\|_2 \tag{4}$$

for some constants $C, D > 0$. Here, the sparsity defect appears in $\sigma_s(\mathbf{x})_1 = \min\{\|\mathbf{x}-\mathbf{z}\|_1, \mathbf{z} \text{ is } s\text{-sparse}\}$. Unlike Subsection I-C, we do not prove that the solution $\mathbf{x}^\lambda$ of (REG) converges to $\mathbf{x}^\sharp$, but rather that it approximates $\mathbf{x}$ with an error controlled almost as in (4), namely

$$\|\mathbf{x} - \mathbf{x}^\lambda\|_1 \leq C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\|\mathbf{e}\|_2 + \frac{E\,s}{\lambda^2}\|\mathbf{x}\|_1 \tag{5}$$

for some constants $C, D, E > 0$ when λ is chosen large enough. First, we use the minimality of $\mathbf{x}^\lambda$ to write

$$\begin{aligned}\|\mathbf{x}^\lambda\|_1^2 + \lambda^2\|\mathbf{A}\mathbf{x}^\lambda - \mathbf{y}\|_2^2 &\leq \|\mathbf{x}^\sharp\|_1^2 + \lambda^2\|\mathbf{A}\mathbf{x}^\sharp - \mathbf{y}\|_2^2 \\ &= \|\mathbf{x}^\sharp\|_1^2,\end{aligned} \tag{6}$$

so $\|\mathbf{x}^\lambda\|_1 \leq \|\mathbf{x}^\sharp\|_1$ follows in particular. Next, the restricted isometry property implies the ℓ_2-robust null space property (see [10, Theorem 6.13]) and we derive in turn (see [10, Theorem 4.20]) that

$$\begin{aligned}&\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1 \\ &\quad\leq c(\|\mathbf{x}^\lambda\|_1 - \|\mathbf{x}^\sharp\|_1 + 2\sigma_s(\mathbf{x}^\sharp)_1) + d\sqrt{s}\|\mathbf{A}(\mathbf{x}^\lambda - \mathbf{x}^\sharp)\|_2 \\ &\quad\leq 2c\sigma_s(\mathbf{x}^\sharp)_1 + d\sqrt{s}\|\mathbf{A}\mathbf{x}^\lambda - \mathbf{y}\|_2\end{aligned}$$

for some constant $c, d > 0$. Therefore,

$$\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1^2 \leq 8c^2\sigma_s(\mathbf{x}^\sharp)_1^2 + 2d^2 s\|\mathbf{A}\mathbf{x}^\lambda - \mathbf{y}\|_2^2. \tag{7}$$

But (6) also implies that

$$\begin{aligned}\|\mathbf{A}\mathbf{x}^\lambda - \mathbf{y}\|_2^2 &\leq \frac{(\|\mathbf{x}^\sharp\|_1 + \|\mathbf{x}^\lambda\|_1)(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}^\lambda\|_1)}{\lambda^2} \\ &\leq \frac{2\|\mathbf{x}^\sharp\|_1(\|\mathbf{x}^\sharp - \mathbf{x}^\lambda\|_1)}{\lambda^2}.\end{aligned}$$

This inequality substituted in (7) gives

$$\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1^2 \leq 8c^2\sigma_s(\mathbf{x}^\sharp)_1^2 + \frac{4d^2 s\|\mathbf{x}^\sharp\|_1(\|\mathbf{x}^\sharp - \mathbf{x}^\lambda\|_1)}{\lambda^2}.$$

Solving this quadratic inequality in $\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1$ yields

$$\begin{aligned}\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1 &\leq \frac{2d^2 s\|\mathbf{x}^\sharp\|_1}{\lambda^2} + \sqrt{\frac{4d^4 s^2\|\mathbf{x}^\sharp\|_1^2}{\lambda^4} + 8c^2\sigma_s(\mathbf{x}^\sharp)_1^2} \\ &\leq \frac{4d^2 s\|\mathbf{x}^\sharp\|_1}{\lambda^2} + \sqrt{8}c\sigma_s(\mathbf{x}^\sharp)_1.\end{aligned}$$

Now using the facts that $\|\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}\|_1 + \|\mathbf{x}^\sharp - \mathbf{x}\|_1$ and that $\sigma_s(\mathbf{x}^\sharp)_1 \le \sigma_s(\mathbf{x})_1 + \|\mathbf{x}^\sharp - \mathbf{x}\|_1$, we derive

$$\|\mathbf{x}^\lambda - \mathbf{x}^\sharp\|_1 \le \frac{4d^2 s \|\mathbf{x}\|_1}{\lambda^2} + \sqrt{8} c \sigma_s(\mathbf{x})_1 + \left(\frac{4d^2 s}{\lambda^2} + \sqrt{8}c\right)\|\mathbf{x}^\sharp - \mathbf{x}\|_1.$$

We finally deduce (5) by calling upon (4) — provided that λ is chosen large enough to have $\lambda^2 \ge s$, say. The experiment conducted in the reproducible file strongly suggests that (5) reflects the correct behavior as a function of λ when $\mathbf{x}$ is exactly s-sparse and $\mathbf{e} = \mathbf{0}$.

III. Conclusion

We have established a connection between basis pursuit and orthogonal matching pursuit by introducing a simple method to approximate ℓ_1-minimizers by solutions of nonnegative least squares problems. Several computational improvements can be contemplated (e.g. the choice of the parameter λ, the possibility of varying λ at each iteration within Lawson and Hanson's algorithm, the combined reconstruction of positive and nonnegative parts and the implementation of QR-updates in the iterative algorithm). Several theoretical questions can also be raised (e.g. is the number of iterations to solve (NNLS) at most proportional to the sparsity, at least under restricted isometry conditions as in [19], see also [10, Theorem 6.25]?). Importantly, a careful understanding of the relations between ℓ_1-minimization-like and OMP-like algorithms would be extremely valuable.

Appendix A

Let $\mathbf{x} \in \mathbb{R}^N$ be a fixed nonnegative vector and let S be its support. It is not hard to verify that

- $\mathbf{x}$ is the unique solution of (BP) if and only if

$$\text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \Big|\sum_{j \in S} v_j\Big| < \sum_{\ell \in \overline{S}} |v_\ell|; \qquad \text{(BP')}$$

- $\mathbf{x}$ is the unique solution of (NNBP) if and only if

$$\text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \mathbf{v}_{\overline{S}} \ge \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0; \quad \text{(NNBP')}$$

- $\mathbf{x}$ is the unique solution of (F) if and only [2]

$$\text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \mathbf{v}_{\overline{S}} \ge \mathbf{0} \text{ is impossible.} \qquad \text{(F')}$$

The implication (F')⇒(NNBP') is clear ((F)⇒(NNBP) is clearer if the logical statement 'FALSE implies anything' is troublesome). The implication (BP')⇒(NNBP') also holds in general: indeed, for $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, if $\mathbf{v}_{\overline{S}} \ge \mathbf{0}$, then $\big|\sum_{j\in S} v_j\big| < \sum_{\ell \in \overline{S}} |v_\ell|$ yields

$$\sum_{i=1}^N v_i = \sum_{j \in S} v_j + \sum_{\ell \in \overline{S}} v_\ell \ge -\Big|\sum_{j\in S} v_j\Big| + \sum_{\ell \in \overline{S}} |v_\ell| > 0.$$

Let us now assume that the matrix $\mathbf{A}$ satisfies the condition $\sum_{i=1}^N v_i = 0$ for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$. This condition can be imposed by appending a row of ones to any matrix, but it is also fulfilled for frequency matrices, since $\mathbf{A}\mathbf{v} = \mathbf{0}$ gives

$$0 = \sum_{i=1}^m (\mathbf{A}\mathbf{v})_i = \sum_{i=1}^m \sum_{j=1}^N A_{i,j} v_j = \sum_{j=1}^N \sum_{i=1}^m A_{i,j} v_j = \sum_{j=1}^N v_j.$$

Under this condition on $\mathbf{A}$, (NNBP')⇒(F')⇒(BP') holds, so that (BP), (NNBP), and (F) are all equivalent, as announced in Subsection I-A. Indeed, (NNBP') implies (F') because if $\mathbf{v}_{\overline{S}} \ge \mathbf{0}$, then $\sum_{i=1}^N v_i$ could not be zero by (NNBP'), and (F') implies (BP') because the impossibility of $\mathbf{v}_{\overline{S}} \ge \mathbf{0}$ gives $\big|\sum_{\ell \in \overline{S}} v_\ell\big| < \sum_{\ell\in\overline{S}} |v_\ell|$, so

$$\Big|\sum_{j \in S} v_j\Big| = \Big| - \sum_{\ell \in \overline{S}} v_\ell\Big| < \sum_{\ell \in \overline{S}} |v_\ell|.$$

[2] This allows to check that an algorithm's output $\mathbf{x}^\sharp$ satisfying $\mathbf{A}\mathbf{x}^\sharp = \mathbf{y}$ and $\mathbf{x}^\sharp \ge \mathbf{0}$ is the true solution $\mathbf{x}$: verify (F') for $S^\sharp = \mathrm{supp}(\mathbf{x}^\sharp)$ by using e.g. CVX [11] to solve the convex program $\text{minimize} \max_{\ell \in \overline{S^\sharp}} v_\ell$ subject to $\mathbf{A}\mathbf{v} = \mathbf{0}$ (and an additional constraint to exclude the zero solution) — the minimum should be positive

References

[1] M. AlQuraishi and H.H. McAdams, *Direct inference of protein-DNA interactions using compressed sensing methods*, Proc. Natl. Acad. Sci., 108(36), 14819–14824, 2011.
[2] A. Amir and O. Zuk, *Bacterial community reconstruction using compressed sensing*, J. Comput. Biol., 18(11), 1723–1741, 2011.
[3] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, *Combining geometry and combinatorics: A unified approach to sparse signal recovery*, In 'Proc. 46th Annual Allerton Conf. on Communication, Control, and Computing', 798–805, 2008.
[4] L. Breiman, *Better subset regression using the nonnegative garrote*, Technometrics, 37(4), 373–384, 1995.
[5] A. Bruckstein, M. Elad, and M. Zibulevsky, *On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations*, IEEE Trans. Inf. Theory, 54(11), 4813–4820, 2008.
[6] E.J. Candès, *Compressive sampling*, In 'Proc. of the International Congress of Mathematicians', 2006.
[7] V. Chandar, *A negative result concerning explicit matrices with the restricted isometry property*, Preprint, 2008.
[8] S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic Decomposition by Basis Pursuit*, SIAM J. Sci. Comput., 20(1), 33–61, 1998.
[9] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory, 52(4), 1289–1306, 2006.
[10] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhaüser, 2013.
[11] M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab software for disciplined convex programming*, 2008.
[12] D. Koslicki, S. Foucart, and G. Rosen, *Quikr: a method for rapid reconstruction of bacterial communities via compressive sensing*, Bioinformatics, 29(17), 2096–2102, 2013.
[13] A.S. Lan, A.E. Waters, C. Studer, and R.G. Baraniuk, *Sparse Factor Analysis for Learning and Content Analytics*, Preprint.
[14] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974
[15] W. Li, J. Feng, and T. Jiang, *IsoLasso: a LASSO regression approach to RNA-Seq based transcriptome assembly*, J. Comput. Biol., 18(11), 1693–1707, 2011.
[16] S.G. Mallat and Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Proces., 41(12), 3397–3415, 1993.
[17] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Stat. Soc. B, 58(1), 267–288, 1996.
[18] P. Wojtaszczyk, *Stability and instance optimality for Gaussian measurements in compressed sensing*, Found. Comput. Math., 10, 1–13, 2010.
[19] T. Zhang, *Sparse recovery with orthogonal matching pursuit under RIP*, IEEE Trans. Inf. Theory, 57(9), 6215–6221, 2011.
[20] Y. Zhang, J. Yang, and W. Yin, *YALL1: Your ALgorithms for L1*, online at yall1.blogs.rice.edu, 2011.