1.3.6 *Optimality of bounds*

When recovering k-sparse vectors one obviously needs at least $m \geq k$ linear measurements. Even when the support of the unknown vector would be known, this number of measurements would be necessary to identify the value of the nonzero coordinates. Therefore, the dependence of the bound (1.17) on k can possibly only be improved in the logarithmic factor. We shall show that even that is not possible and that this dependence is already optimal as soon as a stable recovery of k-sparse vectors is requested. The approach presented here is essentially taken over from [40].

The proof is based on the following combinatorial lemma.

Lemma 5. *Let $k \leq n$ be two natural numbers. Then there are N subsets $T_1, \dots, T_N$ of $\{1, \dots, n\}$, such that*

(i) $N \geq \left(\frac{n}{4k}\right)^{k/2}$,
(ii) $|T_i| = k$ *for all* $i = 1, \dots, N$ *and*
(iii) $|T_i \cap T_j| < k/2$ *for all* $i \neq j$.

Proof. We may assume that $k \leq n/4$, otherwise one can take $N = 1$ and the statement becomes trivial. The main idea of the proof is straightforward (and similar to the proof of Lemma 3). We choose the sets $T_1, T_2, \dots$ inductively one after another as long as possible, satisfying (ii) and (iii) on the way, and then we show that this process will run for at least N steps with N fulfilling (i).

Let $T_1 \subset \{1, \dots, n\}$ be any set with k elements. The number of subsets of $\{1, \dots, n\}$ with exactly k elements, whose intersection with T_1 has at least $k/2$ elements is bounded by the product of 2^k (i.e., the number of all subsets of T_1) and $\binom{n-k}{\lfloor k/2 \rfloor}$, which is the number of all subsets of T_1^c with at most $k/2$ elements. Therefore there are at least

$$\binom{n}{k} - 2^k \binom{n-k}{\lfloor k/2 \rfloor}$$

sets $T \subset \{1, \dots, n\}$ with k elements and $|T \cap T_1| < k/2$. We select T_2 to be any of them. After the jth step, we have selected sets $T_1, \dots, T_j$ with (ii) and (iii) and there are still

$$\binom{n}{k} - j 2^k \binom{n-k}{\lfloor k/2 \rfloor}$$

to choose from. The process stops if this quantity is not positive any more, i.e. after at least

$$N \geq \frac{\binom{n}{k}}{2^k\binom{n-k}{\lfloor k/2\rfloor}} \geq 2^{-k}\frac{\binom{n}{k}}{\binom{n-\lceil k/2\rceil}{\lfloor k/2\rfloor}} = 2^{-k}\frac{n!}{(n-k)!k!}\cdot\frac{(\lfloor k/2\rfloor)!(n-k)!}{(n-\lceil k/2\rceil)!}$$
$$= 2^{-k}\frac{n(n-1)\dots(n-\lceil k/2\rceil+1)}{k(k-1)\dots(k-\lceil k/2\rceil+1)} \geq 2^{-k}\left(\frac{n}{k}\right)^{\lceil k/2\rceil} \geq \left(\frac{n}{4k}\right)^{k/2}$$

steps.

The following theorem shows that any stable recovery of sparse solutions requires at least m number of measurements, where m is of the order $k\ln(en/k)$.

Theorem 7. *Let $k \leq m \leq n$ be natural numbers, let $A \in \mathbb{R}^{m\times n}$ be a measurement matrix, and let $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ be an arbitrary recovery map such that for some constant $C > 0$*

$$\|x - \Delta(Ax)\|_2 \leq C\frac{\sigma_k(x)_1}{\sqrt{k}} \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{1.26}$$

Then

$$m \geq C' k \ln(en/k) \tag{1.27}$$

with some other constant C' depending only on C.

Proof. We may assume that $C \geq 1$. Furthermore, if k is proportional to n (say $k \geq n/8$), then (1.27) becomes trivial. Hence we may also assume that $k \leq n/8$.

By Lemma 5, there exist index sets $T_1,\dots,T_N$ with $N \geq (n/4k)^{k/2}$, $|T_i| = k$ and $|T_i \cap T_j| < k/2$ if $i \neq j$. We put $x_i = \chi_{T_i}/\sqrt{k}$. Then $\|x_i\|_2 = 1$, $\|x_i\|_1 = \sqrt{k}$ and $\|x_i - x_j\|_2 > 1$ for $i \neq j$.

Let

$$\mathscr{B} = \left\{ z \in \mathbb{R}^n : \|z\|_1 \leq \frac{\sqrt{k}}{4C} \quad \text{and} \quad \|z\|_2 \leq 1/4 \right\}.$$

Then $x_i \in 4C\cdot\mathscr{B}$ for all $i = 1,\dots,N$.

We claim that the sets $A(x_i + \mathscr{B})$ are mutually disjoint. Indeed, let us assume that this is not the case. Then there is a pair of indices $i,j \in \{1,\dots,n\}$ and $z,z' \in \mathscr{B}$ with $i \neq j$ and $A(x_i+z) = A(x_j+z')$. It follows that $\Delta(A(x_i+z)) = \Delta(A(x_j+z'))$ and we get a contradiction by

$$1 < \|x_i - x_j\|_2 = \|(x_i + z - \Delta(A(x_i+z)) - (x_j + z' - \Delta(A(x_j+z')) - z + z'\|_2$$
$$\leq \|(x_i + z - \Delta(A(x_i+z))\|_2 + \|x_j + z' - \Delta(A(x_j+z'))\|_2 + \|z\|_2 + \|z'\|_2$$

$$\leq C\frac{\sigma_k(x_i+z)_1}{\sqrt{k}} + C\frac{\sigma_k(x_j+z')_1}{\sqrt{k}} + \|z\|_2 + \|z'\|_2$$
$$\leq C\frac{\|z\|_1}{\sqrt{k}} + C\frac{\|z'\|_1}{\sqrt{k}} + \|z\|_2 + \|z'\|_2 \leq 1.$$

Furthermore,

$$A(x_i+\mathscr{B}) \subset A((4C+1)\mathscr{B}), \quad i=1,\dots,N$$

Let $d \leq m$ be the dimension of the range of A. We denote by $V \neq 0$ the d-dimensional volume of $A(\mathscr{B})$ and compare the volumes

$$\sum_{j=1}^{N} \mathrm{vol}\big(A(x_j+\mathscr{B})\big) \leq \mathrm{vol}\big(A((4C+1)\mathscr{B})\big).$$

Using linearity of A, we obtain

$$\left(\frac{n}{4k}\right)^{k/2} V \leq N \cdot V \leq (4C+1)^d V \leq (4C+1)^m V.$$

We divide by V and take the logarithm to arrive at

$$\frac{k}{2}\ln\left(\frac{n}{4k}\right) \leq m\ln(4C+1). \tag{1.28}$$

If $k \leq n/8$, then it is easy to check that there is a constant $c' > 0$, such that

$$\ln\left(\frac{n}{4k}\right) \geq c'\ln\left(\frac{en}{k}\right).$$

Putting this into (1.28) finishes the proof. □

1.4 Extensions

Section 1.3 gives a detailed overview of the most important features of compressed sensing. On the other hand, inspired by many questions coming from application driven research, various additional aspects of the theory were studied in the literature. We present here few selected extensions of the ideas of compressed sensing, which turned out to be the most useful in practice. To keep the presentation reasonable short, we do not give any proofs, and only refer to relevant sources.

1.4.1 Frames and Dictionaries

We have considered in Section 1.3 vectors $x \in \mathbb{R}^n$, which are sparse with respect to the natural canonical basis $\{e_j\}_{j=1}^n$ of $\mathbb{R}^n$. In practice, however, the signal has a sparse representation with respect to a basis (or, more general, with respect to a frame or dictionary). Let us first recall some terminology.

A set of vectors $\{\phi_j\}_{j=1}^n$ in $\mathbb{R}^n$, which is linearly independent and which spans the whole space $\mathbb{R}^n$ is called a basis. It follows easily that such a set necessarily has n elements. Furthermore, every $x \in \mathbb{R}^n$ can be expressed uniquely as a linear combination of the basis vectors, i.e. there is a unique $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$, such that

$$x = \sum_{j=1}^{n} c_j \phi_j. \tag{1.29}$$

A basis is called orthonormal, if it satisfies the orthogonality relations

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{1.30}$$

If $\{\phi\}_{j=1}^n$ is an orthonormal basis and $x \in \mathbb{R}^n$, then the decomposition coefficients c_j in (1.29) are given by $c_j = \langle x, \phi_j \rangle$. Furthermore, the relation

$$\|x\|_2^2 = \sum_{j=1}^{n} |c_j|^2 \tag{1.31}$$

holds true.

Equations (1.29)–(1.30) can be written also in matrix notation. If $\boldsymbol{\Phi}$ is an $n \times n$ matrix with j-th column equal to ϕ_j, then (1.29) becomes $x = \boldsymbol{\Phi} c$ and (1.30) reads $\boldsymbol{\Phi}^T \boldsymbol{\Phi} = I$, where I denoted the $n \times n$ identity matrix. As a consequence, $c = \boldsymbol{\Phi}^T x$. We shall say that x has sparse or compressible representation with respect to the basis $\{\phi_j\}_{j=1}^n$ if the vector $c \in \mathbb{R}^n$ is sparse or compressible, respectively.

To allow for more flexibility in representation of signals, it is often useful to drop the condition of linear independence of the set $\{\phi_j\}_{j=1}^N \subset \mathbb{R}^n$. As before, we represent such a system of vectors by an $n \times N$ matrix $\boldsymbol{\Phi}$. We say that $\{\phi_j\}_{j=1}^N$ is a frame, if there are two positive finite constants $0 < A \leq B$, such that

$$A\|x\|_2^2 \leq \sum_{j=1}^{N} |\langle x, \phi_j \rangle|^2 \leq B\|x\|_2^2. \tag{1.32}$$