Corollary 1. *Let $A \in \mathbb{R}^{m\times n}$ and let k be a natural number with $k \le n/2$. If $\delta_{2k}(A) < 1/3$, then every k-sparse vector x is the unique solution of (P_1) with $y = Ax$.*

1.3.4 RIP for random matrices

From what was said up to now, we know that matrices with small restricted isometry constants fulfill the null space property, and sparse solutions of underdetermined linear equations involving such matrices can be found by ℓ_1-minimization (P_1). We discuss in this chapter a class of matrices with small RIP constants. It turns out that the most simple way is to construct these matrices by taking its entries to be independent standard normal variables.

We denote until the end of this section

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} \omega_{1,1} & \dots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{m1} & \dots & \omega_{mn} \end{pmatrix}, \tag{1.14}$$

where $\omega_{ij}, i = 1,\dots,m, j = 1,\dots,n$, are i.i.d. standard normal variables. We shall show that such a matrix satisfies the RIP with reasonably small constants with high probability.

1.3.4.1 Concentration inequalities

Before we come to the main result of this chapter, we need some properties of independent standard normal variables.

Lemma 1. *(i) Let ω be a standard normal variable. Then $\mathbb{E}\,(e^{\lambda\omega^2}) = 1/\sqrt{1-2\lambda}$ for $-\infty < \lambda < 1/2$.*

(ii) (2-stability of the normal distribution) Let $m \in \mathbb{N}$, let $\lambda = (\lambda_1,\dots,\lambda_m) \in \mathbb{R}^m$ and let $\omega_1,\dots,\omega_m$ be i.i.d. standard normal variables. Then $\lambda_1\omega_1 + \dots + \lambda_m\omega_m \sim (\sum_{i=1}^m \lambda_i^2)^{1/2} \cdot \mathcal{N}(0,1)$, i.e. it is equidistributed with a multiple of a standard normal variable.

Proof. The proof of (i) follows from the substitution $s := \sqrt{1-2\lambda}\cdot t$ in the following way.

$$\begin{aligned}\mathbb{E}\,(e^{\lambda\omega^2}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda t^2}\cdot e^{-t^2/2}dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{(\lambda-1/2)t^2}dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-s^2/2}\cdot\frac{ds}{\sqrt{1-2\lambda}} = \frac{1}{\sqrt{1-2\lambda}}.\end{aligned}$$

Although the property (ii) is very well known (and there are several different ways to prove it), we provide a simple geometric proof for the sake of completeness. It is enough to consider the case $m = 2$. The general case then follows by induction.

Let therefore $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2, \lambda \neq 0$, be fixed and let ω_1 and ω_2 be i.i.d. standard normal random variables. We put $S := \lambda_1\omega_1 + \lambda_2\omega_2$. Let $t \geq 0$ be an arbitrary non-negative real number. We calculate

$$\begin{aligned}\mathbb{P}(S \leq t) &= \frac{1}{2\pi}\int_{(u,v):\lambda_1 u+\lambda_2 v\leq t} e^{-(u^2+v^2)/2}dudv = \frac{1}{2\pi}\int_{u\leq c;v\in\mathbb{R}} e^{-(u^2+v^2)/2}dudv\\ &= \frac{1}{\sqrt{2\pi}}\int_{u\leq c} e^{-u^2/2}du.\end{aligned}$$

We have used the rotational invariance of the function $(u,v) \to e^{-(u^2+v^2)/2}$. The value of c is given by the distance of the origin from the line $\{(u,v) : \lambda_1 u + \lambda_2 v = t\}$. It follows by elementary geometry and Pythagorean theorem that (cf. $\Delta OAP \simeq \Delta BAO$ in Figure 1.3)

$$c = |OP| = |OB| \cdot \frac{|OA|}{|AB|} = \frac{t}{\sqrt{\lambda_1^2+\lambda_2^2}}.$$

We therefore get

$$\mathbb{P}(S \leq t) = \frac{1}{\sqrt{2\pi}}\int_{\sqrt{\lambda_1^2+\lambda_2^2}\cdot u\leq t} e^{-u^2/2}du = \mathbb{P}\left(\sqrt{\lambda_1^2+\lambda_2^2}\cdot\omega \leq t\right).$$

The same estimate holds for negative t's by symmetry and the proof is finished. □

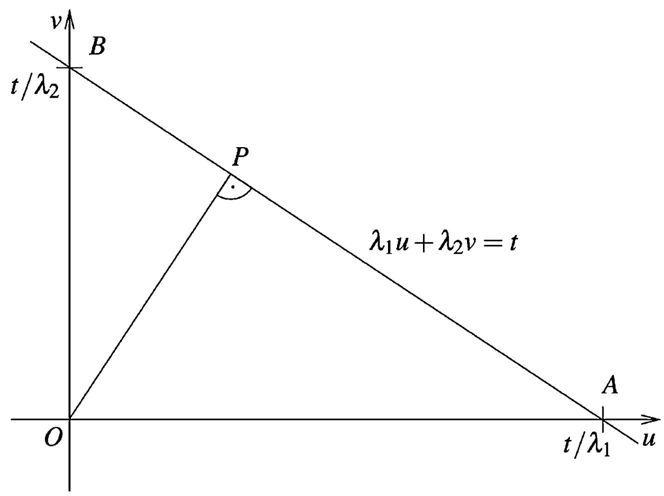

Fig. 1.3 Calculating $c = |OP|$ by elementary geometry for $\lambda_1, \lambda_2 > 0$

If $\omega_1,\dots,\omega_m$ are (possibly dependent) standard normal random variables, then $\mathbb{E}(\omega_1^2+\dots+\omega_m^2)=m$. If $\omega_1,\dots,\omega_m$ are even independent, then the value of $\omega_1^2+\dots+\omega_m^2$ concentrates very strongly around m. This effect is known as *concentration of measure*, cf. [49, 50, 55].

Lemma 2. *Let $m\in\mathbb{N}$ and let $\omega_1,\dots,\omega_m$ be i.i.d. standard normal variables. Let $0<\varepsilon<1$. Then*

$$\mathbb{P}(\omega_1^2+\dots+\omega_m^2\ge(1+\varepsilon)m)\le e^{-\frac{m}{2}[\varepsilon^2/2-\varepsilon^3/3]}$$

and

$$\mathbb{P}(\omega_1^2+\dots+\omega_m^2\le(1-\varepsilon)m)\le e^{-\frac{m}{2}[\varepsilon^2/2-\varepsilon^3/3]}.$$

Proof. We prove only the first inequality. The second one follows in exactly the same manner. Let us put $\beta:=1+\varepsilon>1$ and calculate

$$\begin{aligned}\mathbb{P}(\omega_1^2+\dots+\omega_m^2\ge\beta m)&=\mathbb{P}(\omega_1^2+\dots+\omega_m^2-\beta m\ge 0)\\&=\mathbb{P}(\lambda(\omega_1^2+\dots+\omega_m^2-\beta m)\ge 0)\\&=\mathbb{P}(\exp(\lambda(\omega_1^2+\dots+\omega_m^2-\beta m))\ge 1)\\&\le\mathbb{E}\exp(\lambda(\omega_1^2+\dots+\omega_m^2-\beta m)),\end{aligned}$$

where $\lambda>0$ is a positive real number, which shall be chosen later on. We have used the Markov's inequality (1.3) in the last step. Further we use the elementary properties of exponential function and (1.5) for the independent variables $\omega_1,\dots,\omega_m$. This leads to

$$\mathbb{E}\exp(\lambda(\omega_1^2+\dots+\omega_m^2-\beta m))=e^{-\lambda\beta m}\cdot\mathbb{E}e^{\lambda\omega_1^2}\cdots e^{\lambda\omega_m^2}=e^{-\lambda\beta m}\cdot(\mathbb{E}e^{\lambda\omega_1^2})^m$$

and with the help of Lemma 1 we get finally (for $0<\lambda<1/2$)

$$\mathbb{E}\exp(\lambda(\omega_1^2+\dots+\omega_m^2-\beta m))=e^{-\lambda\beta m}\cdot(1-2\lambda)^{-m/2}.$$

We now look for the value of $0<\lambda<1/2$, which would minimize the last expression. Therefore, we take the derivative of $e^{-\lambda\beta m}\cdot(1-2\lambda)^{-m/2}$ and put it equal to zero. After a straightforward calculation, we get

$$\lambda=\frac{1-1/\beta}{2},$$

which obviously satisfies also $0 < \lambda < 1/2$. Using this value of λ we obtain

$$\mathbb{P}(\omega_1^2 + \cdots + \omega_m^2 \geq \beta m) \leq e^{-\frac{1-1/\beta}{2}\cdot\beta m} \cdot (1-(1-1/\beta))^{-m/2} = e^{-\frac{\beta-1}{2}m} \cdot \beta^{m/2}$$
$$= e^{-\frac{\varepsilon m}{2}} \cdot e^{\frac{m}{2}\ln(1+\varepsilon)}.$$

The result then follows from the inequality

$$\ln(1+t) \leq t - \frac{t^2}{2} + \frac{t^3}{3}, \quad -1 < t < 1. \qquad \square$$

Using 2-stability of the normal distribution, Lemma 2 shows immediately that A defined as in (1.14) acts with high probability as isometry on one fixed $x \in \mathbb{R}^n$.

Theorem 4. *Let $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ and let A be as in (1.14). Then*

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geq t\right) \leq 2e^{-\frac{m}{2}[t^2/2 - t^3/3]} \leq 2e^{-Cmt^2} \tag{1.15}$$

for $0 < t < 1$ with an absolute constant $C > 0$.

Proof. Let $x = (x_1, x_2, \ldots, x_n)^T$. Then we get by the 2-stability of normal distribution and Lemma 2

$$\begin{aligned}
&\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geq t\right)\\
&= \mathbb{P}\left(\left|(\omega_{1,1}x_1 + \cdots + \omega_{1n}x_n)^2 + \cdots + (\omega_{m1}x_1 + \cdots + \omega_{mn}x_n)^2 - m\right| \geq mt\right)\\
&= \mathbb{P}\left(\left|\omega_1^2 + \cdots + \omega_m^2 - m\right| \geq mt\right)\\
&= \mathbb{P}\left(\omega_1^2 + \cdots + \omega_m^2 \geq m(1+t)\right) + \mathbb{P}\left(\omega_1^2 + \cdots + \omega_m^2 \leq m(1-t)\right)\\
&\leq 2e^{-\frac{m}{2}[t^2/2 - t^3/3]}.
\end{aligned}$$

This gives the first inequality in (1.15). The second one follows by simple algebraic manipulations (for $C = 1/12$). $\square$

Remark 4. (i) Observe that (1.15) may be easily rescaled to

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq t\|x\|_2^2\right) \leq 2e^{-Cmt^2}, \tag{1.16}$$

which is true for every $x \in \mathbb{R}^n$.

(ii) A slightly different proof of (1.15) is based on the rotational invariance of the distribution underlying the random structure of matrices defined by (1.14). Therefore, it is enough to prove (1.15) only for one fixed element $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$. Taking $x = e_1 = (1,0,\ldots,0)^T$ to be the first canonical unit vector allows us to use Lemma 2 without the necessity of applying the 2-stability of normal distribution.

1.3.4.2 RIP for random Gaussian matrices

The proof of restricted isometry property of random matrices generated as in (1.14) is based on two main ingredients. The first is the concentration of measure phenomenon described in its most simple form in Lemma 2, and reformulated in Theorem 4. The second is the following entropy argument, which allows to extend Theorem 4 and (1.15) from one fixed $x \in \mathbb{R}^n$ to the set Σ_k of all k-sparse vectors.

Lemma 3. *Let $t > 0$. Then there is a set $\mathcal{N} \subset \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ with*

(i) $|\mathcal{N}| \le (1+2/t)^n$ *and*
(ii) *for every $z \in \mathbb{S}^{n-1}$, there is a $x \in \mathcal{N}$ with $\|x - z\|_2 \le t$.*

Proof. Choose any $x^1 \in \mathbb{S}^{n-1}$. If $x^1,\ldots,x^j \in \mathbb{S}^{n-1}$ were already chosen, take $x^{j+1} \in \mathbb{S}^{n-1}$ arbitrarily with $\|x^{j+1} - x^l\|_2 > t$ for all $l = 1,\ldots,j$. This process is then repeated as long as possible, i.e. until we obtain a set $\mathcal{N} = \{x^1,\ldots,x^N\} \subset \mathbb{S}^{n-1}$, such that for every $z \in \mathbb{S}^{n-1}$ there is a $j \in \{1,\ldots,N\}$ with $\|x^j - z\|_2 \le t$. This gives the property (ii).

We will use volume arguments to prove (i). It follows by construction that $\|x^i - x^j\|_2 > t$ for every $i,j \in \{1,\ldots,N\}$ with $i \ne j$. By triangle inequality, the balls $B(x^j, t/2)$ are all disjoint and are all included in the ball with the center in the origin and radius $1 + t/2$. By comparing the volumes we get

$$N \cdot (t/2)^n \cdot V \le (1+t/2)^n \cdot V,$$

where V is the volume of the unit ball in $\mathbb{R}^n$. Hence, we get $N = |\mathcal{N}| \le (1+2/t)^n$.

□

With all these tools at hand, we can now state the main theorem of this section, whose proof follows closely the arguments of [4].

Theorem 5. *Let $n \ge m \ge k \ge 1$ be natural numbers and let $0 < \varepsilon < 1$ and $0 < \delta < 1$ be real numbers with*

$$m \ge C\delta^{-2}\Big(k \ln(en/k) + \ln(2/\varepsilon)\Big), \tag{1.17}$$

where $C > 0$ is an absolute constant. Let A be again defined by (1.14). Then

$$\mathbb{P}\big(\delta_k(A) \le \delta\big) \ge 1 - \varepsilon.$$

Proof. The proof follows by the concentration inequality of Theorem 4 and the entropy argument described in Lemma 3. By this lemma, there is a set

$$\mathscr{N} \subset Z := \{z \in \mathbb{R}^n : \text{supp}\,(z) \subset \{1,\dots,k\}, \|z\|_2 = 1\},$$

such that

(i) $|\mathscr{N}| \le 9^k$ and
(ii) $\min_{x\in\mathscr{N}} \|z-x\|_2 \le 1/4$ for every $z \in Z$.

We show that if $\big|\|Ax\|_2^2 - 1\big| \le \delta/2$ for all $x \in \mathscr{N}$, then $\big|\|Az\|_2^2 - 1\big| \le \delta$ for all $z \in Z$.

We proceed by the following bootstrap argument. Let $\gamma > 0$ be the smallest number, such that $\big|\|Az\|_2^2 - 1\big| \le \gamma$ for all $z \in Z$. Then $\big|\|Au\|_2^2 - \|u\|_2^2\big| \le \gamma\|u\|_2^2$ for all $u \in \mathbb{R}^n$ with supp $(u) \subset \{1,\dots,k\}$. Let us now assume that $\|u\|_2 = \|v\|_2 = 1$ with supp $(u) \cup$ supp $(v) \subset \{1,\dots,k\}$. Then we get by polarization identity

$$\begin{aligned}
|\langle Au, Av\rangle - \langle u,v\rangle| &= \frac{1}{4}\Big|(\|A(u+v)\|_2^2 - \|A(u-v)\|_2^2) - (\|u+v\|_2^2 - \|u-v\|_2^2)\Big| \\
&\le \frac{1}{4}\Big|\|A(u+v)\|_2^2 - \|u+v\|_2^2\Big| + \frac{1}{4}\Big|\|A(u-v)\|_2^2 - \|u-v\|_2^2\Big| \\
&\le \frac{\gamma}{4}\|u+v\|_2^2 + \frac{\gamma}{4}\|u-v\|_2^2 = \frac{\gamma}{2}(\|u\|_2^2 + \|v\|_2^2) = \gamma.
\end{aligned}$$

Applying this inequality to $u' = u/\|u\|_2$ and $v' = v/\|v\|_2$, we obtain

$$|\langle Au, Av\rangle - \langle u,v\rangle| \le \gamma\|u\|_2\|v\|_2 \tag{1.18}$$

for all $u, v \in \mathbb{R}^n$ with supp $(u) \cup$ supp $(v) \subset \{1,\dots,k\}$.

Let now again $z \in Z$. Then there is an $x \in \mathscr{N}$, such that $\|z-x\|_2 \le 1/4$. We obtain by triangle inequality and (1.18)

$$\begin{aligned}
\big|\|Az\|_2^2 - 1\big| &= \big|\|Ax\|_2^2 - 1 + \langle A(z+x), A(z-x)\rangle - \langle z+x, z-x\rangle\big| \\
&\le \delta/2 + \gamma\|z+x\|_2\|z-x\|_2 \le \delta/2 + \gamma/2.
\end{aligned}$$

As the supremum of the left-hand side over all admissible z's is equal to γ, we obtain that $\gamma \le \delta$ and the statement follows.

Equipped with this tool, the rest of the proof follows by a simple union bound.

$$\begin{aligned}\mathbb{P}(\delta_k(A) > \delta) &\le \sum_{\substack{T\subset\{1,\ldots,n\}\\ |T|\le k}} \mathbb{P}\Big(\exists z \in \mathbb{R}^n : \text{supp }(z) \subset T, \|z\|_2 = 1 \text{ and } \big|\|Az\|_2^2 - 1\big| > \delta\Big)\\ &= \binom{n}{k}\mathbb{P}\Big(\exists z \in Z \text{ with } \big|\|Az\|_2^2 - 1\big| > \delta\Big)\\ &\le \binom{n}{k}\mathbb{P}\Big(\exists x \in \mathscr{N} : \big|\|Ax\|_2^2 - 1\big| > \delta/2\Big).\end{aligned}$$

By Theorem 4, the last probability may be estimated from above by $2e^{-C'm\delta^2}$. Hence we obtain

$$\mathbb{P}(\delta_k(A) > \delta) \le 9^k\binom{n}{k}\cdot 2e^{-C'm\delta^2}$$

Hence it is enough to show that the last quantity is at most ε if (1.17) is satisfied. But this follows by straightforward algebraic manipulations and the well-known estimate

$$\binom{n}{k} \le \frac{n^k}{k!} \le \left(\frac{en}{k}\right)^k. \qquad \square$$

1.3.4.3 Lemma of Johnson and Lindenstrauss

Concentration inequalities similar to (1.15) play an important role in several areas of mathematics. We shall present their connection to the famous result from functional analysis called Johnson–Lindenstrauss lemma, cf. [1, 22, 46, 54]. The lemma states that a set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that the mutual distances between the points are nearly preserved. The connection between this classical result and compressed sensing was first highlighted in [4], cf. also [47].

Lemma 4. *Let $0 < \varepsilon < 1$ and let m, N and n be natural numbers with*

$$m \ge 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1}\ln N.$$

Then for every set $\{x^1,\ldots,x^N\} \subset \mathbb{R}^n$ there exists a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, such that

$$(1-\varepsilon)\|x^i - x^j\|_2^2 \le \|f(x^i) - f(x^j)\|_2^2 \le (1+\varepsilon)\|x^i - x^j\|_2^2, \qquad i,j \in \{1,\ldots,N\}. \tag{1.19}$$

Proof. We put $f(x) = Ax$, where again

$$Ax = \frac{1}{\sqrt{m}} \begin{pmatrix} \omega_{1,1} & \dots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{m1} & \dots & \omega_{mn} \end{pmatrix} x,$$

and $\omega_{ij}, i = 1, \dots, m, j = 1, \dots, n$ are i.i.d. standard normal variables. We show that with this choice f satisfies (1.19) with positive probability. This proves the existence of such a mapping.

Let $i, j \in \{1, \dots, N\}$ arbitrary with $x^i \neq x^j$. Then we put $z = \frac{x^i - x^j}{\|x^i - x^j\|_2}$ and evaluate the probability that the right-hand side inequality in (1.19) does not hold. Theorem 4 then implies

$$\mathbb{P}\Big(\Big|\|f(x^i) - f(x^j)\|_2^2 - \|x^i - x^j\|_2^2\Big| > \varepsilon \|x^i - x^j\|_2^2\Big) = \mathbb{P}\Big(\Big|\|Az\|^2 - 1\Big| > \varepsilon\Big)$$
$$\leq 2e^{-\frac{m}{2}[\varepsilon^2/2 - \varepsilon^3/3]}.$$

The same estimate is also true for all $\binom{N}{2}$ pairs $\{i, j\} \subset \{1, \dots, N\}$ with $i \neq j$. The probability that one of the inequalities in (1.19) is not satisfied is therefore at most

$$2 \cdot \binom{N}{2} \cdot e^{-\frac{m}{2}[\varepsilon^2/2 - \varepsilon^3/3]} < N^2 \cdot e^{-\frac{m}{2}[\varepsilon^2/2 - \varepsilon^3/3]} = \exp\Big(2 \ln N - \frac{m}{2}[\varepsilon^2/2 - \varepsilon^3/3]\Big) \leq e^0 = 1$$

for $m \geq 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \ln N$. Therefore, the probability that (1.19) holds for all $i, j \in \{1, \dots, N\}$ is positive and the result follows. □

1.3.5 Stability and Robustness

The ability to recover sparse solutions of underdetermined linear systems by quick recovery algorithms as ℓ_1-minimization is surely a very promising result. On the other hand, two additional features are obviously necessary to extend this results to real-life applications, namely

- Stability: We want to be able to recover (or at least approximate) also vectors $x \in \mathbb{R}^n$, which are not exactly sparse. Such vectors are called *compressible* and mathematically they are characterized by the assumption that their best k-term approximation decays rapidly with k. Intuitively, the faster the decay of the best k-term approximation of $x \in \mathbb{R}^n$ is, the better we should be able to approximate x.
- Robustness: Equally important, we want to recover sparse or compressible vectors from noisy measurements. The basic model here is the assumptions that the