Tutorial on Compressed Sensing

Exercises

1. Exercise

Let $\eta \geq 0$. Show that the ℓ_0 minimization problem

$$(P_{0,\eta}) \qquad x^{\#} = \arg\min \|z\|_0 \text{ s.t. } \|Az - y\|_2 \leq \eta$$

for general $m \times N$-matrices A and $y \in \mathbb{R}^m$ is an NP-hard problem.

Hint: You can use the fact that the exact cover problem is NP-hard.
Exact Cover Problem: Given as the input a natural number m divisible by 3 and a system $\{T_j : j = 1, ..., N\}$ of subsets of $\{1, ..., m\}$ with $|T_j| = 3$ for all $j \in [N]$. Decide, if there is a subsystem of mutually disjoint sets $\{T_j : j \in J\}$, $J \subset [N]$, such that $\cup_{j \in J} T_j = \{1, ..., m\}$.

Solution

We show that any algorithm solving the ℓ_0-problem can be transformed in polynomial time into an algorithm solving the exact cover problem.
Let therefore $\{T_j : j = 1, \dots, N\}$ be a system of subsets of $\{1, \dots, m\}$ with $|T_j| = 3$. We construct a matrix $A \in \mathbb{R}^{m \times N}$ by putting

$$a_{ij} := \begin{cases} 1 & \text{if } i \in T_j \\ 0 & \text{if } i \notin T_j \end{cases},$$

i.e. the j-th column of A is the indicator function of T_j denoted by $\mathcal{X}_{T_j}$ and

$$Ax = \sum_{j=1}^{N} x_j \mathcal{X}_{T_j}. \tag{1}$$

This construction can be of course done in polynomial time. Let now x be the solution to the minimization problem

$$\min \|x\|_0 \text{ s.t. } Ax = y = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{P_0}$$

By (1) follows:

$$m = \|y\|_0 = \|Ax\|_0 = \|\sum_{i=1}^{N} x_j \mathcal{X}_{T_j}\|_0 \leq \sum_{i=1}^{N} \|x_j \mathcal{X}_{T_j}\|_0 \leq 3\|x\|_0, \tag{2}$$

i.e. $\|x\|_0 \geq m/3$, where the last step in the inequality follows, because every $\mathcal{X}_{T_j}$ has at most three nonzero entries and $\|x_j \mathcal{X}_{T_j}\|_0 = 0$, if $x_j = 0$.
We show, that the exact cover problem has a unique solution if and only if $\|x\|_0 = m/3$. This shows that after solving (P_0) we can decide if the exact cover problem has a positive solution or not by computing the ℓ_0-norm of the solution x.

Let us first assume that the exact cover problem has a positive solution. Then there is a set $J \subset \{1, \dots, N\}$ with $|J| = m/3$ and $y = \mathcal{X}_{\{1,\dots,m\}} = \sum_{j \in J} \mathcal{X}_{T_j}$. Hence $y = Ax$ for $x = \mathcal{X}_J$ and $\|x\|_0 = |J| = m/3$, which is indeed the minimizer of (P_0), because of (2).
If on the other hand $y = Ax$ and $\|x\|_0 = m/3$, then $\{T_j : j \in \operatorname{supp} x\}$ solves the exact cover problem.

2. Exercise
Let $q > 1$ and let A be a $m \times N$-matrix with $m < N$. Show that, there is a 1-sparse vector x, which is not a solution of the optimization problem

$$(P_q) \qquad x^* = \arg\min \|z\|_q \text{ s.t. } Az = Ax.$$

Solution

For $j = 1, \dots, N$ let $e_j \in \mathbb{R}^N$ be a 1-sparse vector. Now suppose that for all $z \in \mathbb{R}^N$ with $Az = Ae_j$ and $z \neq e_j$ we have $\|z\|_q^q > \|e_j\|_q^q = 1$. Let $v \in \ker(A) \setminus \{0\}$ and $t \neq 0$ with $|t| < 1/\|v\|_\infty$, then we obtain

$$1 < \|e_j + tv\|_q^q = |1 + tv_j|^q + \sum_{k \neq j} |tv_k|^q = |1 + tv_j|^q + |t|^q \sum_{k \neq j} |v_k|^q \sim_{t \to 0} 1 + qtv_j,$$

where the last estimation follows from the multi-binomial theorem.
This inequality shows that $v_j = 0$ because it is in particular true for $-1/\|v\|_\infty \leq t < 0$. But this is in fact true for all $j \in \{1, \dots, N\}$, and therefore it follows $v = 0$, which yields a contradiction.

3. Exercise
Let A be a $m \times N$-matrix, $y \in \mathbb{R}^m$, $\eta > 0$ and let $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^m$. Show that the solution of the optimization problem

$$(P_{1,\eta}) \qquad x^* = \arg\min \|z\|_1 \text{ s.t. } \|Az - y\| \leq \eta$$

is m-sparse in the case of the uniqueness of the solution.
Hint: Show that the system of columns $\{a_j : j \in \text{supp } x^*\}$ is linearly independent.

Solution

We show that, if x^* is a unique solution, with $K := \operatorname{supp} x^*$, then the columns $\{a_j : j \in K\}$ have to be linearly independent. Since at most m columns can be linearly independent the statement follows. Suppose $\{a_j : j \in K\}$ is not linearly independent, then there is a $v \in \mathbb{R}^N$ with $Av = 0$ and $v \neq 0$, i.e. $v \in \ker(A)$, and $\operatorname{supp}(v) \subset K$. But because of the uniqueness we have for every $t \in \mathbb{R}$ small enough (in absolute value):

$$\begin{aligned}\|x^*\|_1 < \|x+tv\|_1 &= \sum_{j\in K} |x_j + tv_j| = \sum_{j\in K} \operatorname{sign}(x_j+tv_j)(x_j+tv_j) \\ &= \sum_{j\in K} \operatorname{sign}(x_j+tv_j)x_j + t\sum_{j\in K} v_j \operatorname{sign}(x_j+tv_j) \\ &= \sum_{j\in K} \operatorname{sign}(x_j)x_j + t\sum_{j\in K} v_j \operatorname{sign}(x_j) \\ &= \|x^*\|_1 + t\sum_{j\in K} v_j \operatorname{sign}(x_j),\end{aligned}$$

which is a contradiction, since we can choose t such that $t\sum_{j\in K} v_j \operatorname{sign}(x_j)$ becomes smaller than zero.

4. Exercise

Let A be an $m\times N$ matrix and $2s \le m$. Show that the following statements are equivalent:

i) There is a mapping $\Lambda : \mathbb{R}^m \to \mathbb{R}^N$ such that $\Lambda(Ax) = x$ for all $x \in \Sigma_s$. We call such a mapping Λ a decoder.

ii) $\Sigma_{2s} \cap \ker(A) = \{0\}$.

iii) For any set T with $\#T = 2s$, the matrix A_T has rank $2s$.

iv) The symmetric non-negative matrix $A_T^t A_T$ is invertible, i.e. positive definite.

Solution

The equivalence between ii), iii) and iv) is linear algebra.
For example ii)⇒iii): If $\Sigma_{2k} \cap \ker(A) = \{0\}$, we can deduce that for every $T \subset \{1,\dots,N\}$ with $|T| \le 2k$ it holds $\ker(A_T) = \{0\}$. And therefore that A_T has full rank.

i)⇒ ii): Let $x \in \Sigma_{2k} \cap \ker(A)$, then we can write $x = x_1 - x_0$, where both x_1 and x_0 lie in Σ_k. Since $x \in \ker(A)$ it holds $Ax = 0$ and therefore $Ax_1 = Ax_0$, which implies by assumption i) that $x_1 = x_0$ and therefore $x = 0$.

ii)⇒ i): For any $y \in \mathbb{R}^m$ we define the decoder Λ as $\Lambda(y)$ to be the element with the smallest support in the set of solutions $\{x \in \mathbb{R}^N : Ax = y\}$. Suppose there is $x_1 \in \Sigma_k$ such that $\Lambda(Ax_1) \ne x_1$. This implies that there is a x_0 with $Ax_0 = Ax_1$ and $\|x_0\|_0 \le \|x_1\|_0 = k$, and hence that $x_1 - x_0 \in \Sigma_{2k} \cap \ker(A)$. By assumption this implies $x_1 = x_0$.

5. Exercise

[NSP] Given a matrix $A \in \mathbb{R}^{m\times N}$, every vector $x \in \mathbb{R}^N$ supported on a set T is the unique solution of (P_1) with $y = Ax$ if and only if A satisfies the null space property relative to T.

Reminder: A is said to satisfy the null space property relative to the set T if for all $v \in \ker(A)$ holds

$$\|v_T\|_1 < \|v_{T^C}\|_1,$$

where $(v_T)_i = v_i$ if $i \in T$ and $(v_T)_i = 0$ otherwise.

Solution

Given a index set T and assume that every vector $x \in \mathbb{R}^N$ supported on T is the unique minimizer. Thus for every $v \in \ker(A) \setminus \{0\}$, v_T is the unique minimizer of (P_1) with $Ax = Av_k$. But because of $A(v_T + v_{T^C}) = Av = 0$, we can deduce that $A(-v_{T^C}) = Av_k$ and hence by assumption $\|v_T\|_1 < \|v_{T^C}\|$.

Conversely let us assume that the NSP relative to T holds. Given a vector x supported on T, for every $z \in \mathbb{R}^N$ with $Az = Ax$ and $z \neq x$, we have $v := x - z \in \ker(A) \setminus \{0\}$. We obtain by assumption

$$\begin{aligned}\|x\|_1 &\leq \|x - z_T\|_1 + \|z_T\|_1 = \|v_T\| + \|z_T\|_1 < \|v_{T^C}\|_1 + \|z_T\|_1 \\ &= \|z_{T^C}\|_1 + \|z_T\|_1 = \|z\|_1,\end{aligned}$$

where we used in the third step the assumption.

6. Exercise

Given a matrix $A \in \mathbb{R}^{m\times N}$, a vector $x \in \mathbb{R}^N$ with support T is the unique minimizer of (P_1) if and only if $|\sum_{j\in T} \operatorname{sign}(x_j)v_j| < \|v_{T^C}\|_1$ for all $v \in \ker A \setminus \{0\}$.

Solution

Let us start by proving that the inequality implies that $x \in \mathbb{R}^N$ with support T is the unique minimizer of (P_1). For a vector $z \in \mathbb{R}^N$, $z \neq x$, with $Az = Ax$ we write, with $v = x - z \in \ker(A) \setminus \{0\}$,

$$\begin{aligned}\|z\|_1 &= \|z_T\|_1 + \|z_{T^C}\|_1 = \|(x - v)_T\|_1 + \|v_{T^C}\|_1 \\ &> |\langle x - v, \operatorname{sign}(x)_T\rangle| + |\langle v, \operatorname{sign}(x)_T\rangle| \geq |\langle x, \operatorname{sign}(x)_T\rangle| \\ &= \|x\|_1.\end{aligned}$$

It remains to show that the inequality holds as soon as x, supported on T, is the unique minimizer of (P_1). In this situation for $v \in \ker(A) \setminus \{0\}$, the vector $z = x - v$ satisfies $Ax = Az$ and $\|x\|_1 < \|z\|_1$. From this we can deduce

$$\begin{aligned}\langle x, \operatorname{sign}(z)_T\rangle \leq \|x\|_1 < \|z\|_1 = \|z_T\|_1 + \|z_{T^C}\|_1 &= \langle z, \operatorname{sign}(z)_T\rangle + \|z_{T^C}\|_1 \\ &\Leftrightarrow \langle v, \operatorname{sign}(z)_T\rangle < \|z_{T^C}\|_1 \\ &\Leftrightarrow \langle v, \operatorname{sign}(x - v)_T\rangle < \|z_{T^C}\|_1 = \|v_{T^C}\|_1.\end{aligned}$$

But since this holds true for every $v \in \ker(A) \setminus \{0\}$, it holds for $t > 0$ that

$$\begin{aligned}&\langle tv, \operatorname{sign}(x - tv)_T\rangle < t\|v_{T^C}\|_1 \\ \Leftrightarrow\ &\langle v, \operatorname{sign}(x - tv)_T\rangle < \|v_{T^C}\|_1.\end{aligned}$$

And for t small enough it holds $\text{sign}(x - tv)_j = \text{sign}(x_j)$ and therefore:

$$\langle v, \text{sign}(x)\rangle < \|v_{T^C}\|_1.$$

7. Exercise

Show that the RIP implies the NSP.
More explicit: Let $A \in \mathbb{R}^{m,d}$ satisfy the restricted isometry porperty (RIP) of order $2s$ with constant $0 < \delta_{2s} < 1/3$, i.e.

$$(1 - \delta_{2s})\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_{2s})\|x\|_2^2$$

holds for all $2s$-sparse vectors x, i.e. for all

$$x \in \Sigma_{2s} = \{v \in \mathbb{R}^d \mid \|v\|_0 = \#\{i \mid v_i \neq 0\} \leq 2s\}.$$

Show that A satisfies the null space property of order s (NSP), i.e. for any $T \subset [d]$ with $\#T \leq s$ and any $v \in \ker A \backslash \{0\}$ it holds

$$2\|v_T\|_1 < \|v\|_1,$$

where $(v_T)_i = v_i$ if $i \in T$ and $(v_T)_i = 0$ otherwise.
Hint:

1. First show that

$$\langle Ax, Ay\rangle \leq \delta_{2s}\|x\|_2\|y\|_2$$

if x, y are s-sparse with disjoint support.

2. For $v \in \ker A\backslash\{0\}$ let $T_0 \subset [d]$ denote the set of indices corresponding to the s-largest entries of v (in magnitude). Further let $T^c = T_1 \cup T_2 \cup \ldots$ be a partition of T^c such that T_1 contains the indices of s-largest entries of v_{T^c}, T_2 contains the s-largest entries of $V_{T^c\backslash T_1}$ etc.

Solution

1. step Let $x, y \in \Sigma_s$ with disjoint support and with $\|x\|_2 = \|y\|_2 = 1$. Then it holds $x \pm y \in \Sigma_{2s}$ and $\|x \pm y\|_2^2 = 2$. Using the RIP of A of order $2s$ we obtain

$$2(1 - \delta_{2s}) = (1 - \delta_{2s})\|x \pm y\|_2^2 \leq \|A(x \pm y)\|_2^2 \leq (1 + \delta_{2s})\|x \pm y\|_2^2 = 2(1 + \delta_{2s}).$$

Now the claim follows from the polarization identity, since

$$|\langle Ax, Ay\rangle| = \frac{1}{4}\left|\|A(x+y)\|_2^2 - \|A(x-y)\|_2^2\right| \leq \frac{1}{4}\big(2(1+\delta_{2s}) - 2(1-\delta_{2s})\big) = \delta_{2s}.$$

2. step Let $v \in \ker A\backslash\{0\}$ and let $T_0 \subset [d] = \{1, 2, \ldots, d\}$ denote the set of indices corresponding to the largest s entries of v (in magnitude). Further divide $T^c = [d]\backslash T$ into sets

$$\begin{aligned} T_1 &\quad - \quad s - \text{largest indices of } v_{T^c} \\ T_2 &\quad - \quad s - \text{largest indices of } v_{T^c\backslash T_1} \\ &\vdots \end{aligned}$$

In total we splittet the support T of v into disjoint sets $T_0, T_1, \dots$ such that T_0 contains indices of s largest entries of v, T_1 contains indices of remaining s-largest entries etc., hence

$$T = T_0 \cup T_1 \cup T_2 \cup \dots$$

Since we v is an element of the kernel of A we get

$$0 = Av = A(v_{T_0} + v_{T_1} + v_{T_2} + \dots) \quad \Rightarrow \quad Av_{T_0} = -A(v_{T_1} + v_{T_2} + \dots).$$

Now we can apply the RIP (since $\# T_0 \le s$) to arrive at

$$\begin{aligned}(1-\delta_{2s})\|v_{T_0}\|_2^2 &\le \|Av_{T_0}\|_2^2 = \langle Av_{T_0}, Av_{T_0}\rangle = \langle Av_{T_0}, -A(v_{T_1} + v_{T_2} + \dots)\rangle \\ &= \sum_{i\ge 1} \langle Av_{T_0}, -Av_{T_i}\rangle.\end{aligned}$$

Here we can apply our first step to get

$$(1-\delta_{2s})\|v_{T_0}\|_2^2 \le \sum_{i\ge 1} \langle Av_{T_0}, -Av_{T_i}\rangle \le \sum_{i\ge 1} \delta_{2s}\|v_{T_0}\|_2 \|v_{T_i}\|_2.$$

Using our construction of the T_i's we further estimate for $i \ge 1$

$$\begin{aligned}\|v_{T_i}\|_2 &= \left(\sum_{j\in T_i} v_j^2\right)^{1/2} \le \left(\sum_{j\in T_i}\left(\max_{k\in T_i}|v_k|\right)^2\right)^{1/2} = \sqrt{s}\max_{k\in T_i}|v_k| \le \sqrt{s}\min_{k\in T_{i-1}}|v_k| \\ &\le \sqrt{s}\frac{\sum\limits_{j\in T_{i-1}}|v_j|}{s} = \frac{\|v_{T_{i-1}}\|_1}{\sqrt{s}}.\end{aligned}$$

Hence,

$$(1-\delta_{2s})\|v_{T_0}\|_2^2 \le \delta_{2s}\sum_{i\ge 1}\|v_{T_0}\|_2\|v_{T_i}\|_2 \le \delta_{2s}\|v_{T_0}\|_2 \sum_{i\ge 1}\frac{\|v_{T_{i-1}}\|_1}{\sqrt{s}} = \frac{\delta_{2s}\|v_{T_0}\|_2}{\sqrt{s}}\|v\|_1.$$

Dividing by $\|v_{T_0}\|_2$ and $(1-\delta_{2s})$ and using $\delta_{2s} < 1/3$ we end up with

$$\|v_{T_0}\|_2 \le \frac{1}{\sqrt{s}}\underbrace{\frac{\delta_{2s}}{1-\delta_{2s}}}_{<1/2}\|v\|_1 \le \frac{\|v\|_1}{2\sqrt{s}}$$

which yields the claim, since by Cauchy-Schwartz inequality it holds for any $x \in \mathbb{R}^s$

$$\|x\|_1 = \langle x, \operatorname{sign}(x)\rangle \le \|x\|_2\|\operatorname{sign}(x)\|_2 = \sqrt{s}\|x\|_2.$$

8. Exercise

Let $s \in \mathbb{N}$, $0 < \delta < 1$ and let

$$m \ge c\delta^{-2}s\log(ed/s).$$

Further let $A = \tilde{A}/\sqrt{m} \in \mathbb{R}^{m,d}$ with i.i.d. entries $\tilde{a}_{ij} \sim \mathcal{N}(0,1)$. Show that A satisfies the RIP of order s with RIP constant $\delta_s \leq \delta$ with probability at least

$$1 - 2\exp\left(-C\delta^2 m\right).$$

Hint: First use a Bernstein inequality to show that

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq t\|x\|_2^2\right) \leq 2\exp\left(-ct^2 m\right)$$

holds for all $t > 0$ and $x \in \mathbb{R}^d$. Then show the desired RIP inequality for a fixed s-dimensional subspace using a covering argument.

Solution

We use the *Bernstein inequality:*
Let $X_1, \ldots, X_m$ be independent mean zero (i.e. $\mathbb{E}X_i = 0$) subexponential random variables, i.e. it holds

$$\mathbb{P}(|X_i| \geq t) \leq \beta \exp(-\kappa t)$$

holds for any $t > 0$ and constants $\beta, \kappa > 0$. Then it holds

$$\mathbb{P}\left(\left|\sum_{i=1}^m X_i\right| \geq t\right) \leq 2\exp\left(\frac{-\kappa^2 t^2}{4\beta m + \kappa t}\right).$$

1. step With Bernstein's inequality we first want to show that

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq t\|x\|_2^2\right) \leq 2\exp(-ct^2 m).$$

Therefore let $x \in \mathbb{R}^d$ with $\|x\|_2 = 1$ and let $\tilde{a}_i$ denote the i-th row of $\tilde{A}$. Consider the random variable

$$X_i = |\langle \tilde{a}_i, x\rangle|^2 - \|x\|_2^2 = |\langle \tilde{a}_i, x\rangle|^2 - 1.$$

Then it holds

- X_i are independent, since $\tilde{a}_i$ are independent,
- X_i are subexponential, since $\tilde{a}_i$ (and $\langle \tilde{a}_i, x\rangle$) are Gaussians,
- X_i have mean zero, since

  $$\mathbb{E}X_i = E\,|\langle \tilde{a}_i, x\rangle|^2 - \|x\|_2^2 = \|x\|_2^2 \mathbb{E}\underbrace{\left\langle \tilde{a}_i\,,\, \frac{x}{\|x\|_2}\right\rangle}_{\sim\mathcal{N}(0,1)}^2 - \|x\|_2^2 = \|x\|_2^2 \underbrace{\mathbb{E}\,|g|^2}_{=1} - \|x\|_2^2 = 0$$

  with $g \sim \mathcal{N}(0,1)$,
- it holds

  $$\frac{1}{m}\sum_{i=1}^m X_i = \frac{1}{m}\sum_{i=1}^m (|\langle \tilde{a}_i, x\rangle|^2 - \|x\|_2^2) = \sum_{i=1}^m (\left|\left\langle \frac{\tilde{a}_i}{\sqrt{m}}\,,\, x\right\rangle\right|^2 - \|x\|_2^2) = \|Ax\|_2^2 - \|x\|_2^2.$$

Now we apply Bernstein's inequality to get

$$\mathbb{P}\left(\left|\|Ax\|_2^2-\|x\|_2^2\right|\geq t\right)=\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^m X_i\right|\geq t\right)=\mathbb{P}\left(\left|\sum_{i=1}^m X_i\right|\geq mt\right)\leq 2\exp\left(\frac{-\kappa^2t^2m^2}{2\beta m+\kappa tm}\right).$$

2. step We fix an s-dimensional subspace. Therefore let $T\subset[d]$ with $\#T=s$ and let

$$X_T=\{x\in\Sigma_s \mid \operatorname{supp} x\subset T\}.$$

We want to show that

$$(1-\delta)\|x\|_2\leq\|Ax\|_2\leq(1+\delta)\|x\|_2$$

holds fpr all $x\in X_T$ with probability at least

$$1-2\left(\frac{12}{\delta}\right)^s\exp(-c\delta^2m).$$

Let $\delta>0$ and let $Q\subset X_T$ be a $\delta/4$-net of $X_T\cap B^d$, i.e. it holds

- $\|q\|_2=1$ for all $q\in Q$ and
- for any $x\in X_T$ with $\|x\|_2=1$ there exists some $q\in Q$ with $\|x-q\|_2\leq\delta/4$.

It is known that we can choose Q with $\#Q\leq(12/\delta)^s$ (a proof is given below). For any $q\in Q$ and $t=\delta/2$ we now use the first step to get

$$\mathbb{P}\left(\left|\|Aq\|_2^2-\|q\|_2^2\right|\geq\delta/2\right)\leq 2\exp(-c\delta^2m)$$

which is equivalent to

$$\left(1-\frac{\delta}{2}\right)\|q\|_2^2\leq\|Aq\|_2^2\leq\left(1+\frac{\delta}{2}\right)\|q\|_2^2$$

with probability at least $1-2\exp(-c\delta^2m)$. Hence, for any (fixed) $q\in Q$ it also holds

$$\left(1-\frac{\delta}{2}\right)\|q\|_2\leq\|Aq\|_2\leq\left(1+\frac{\delta}{2}\right)\|q\|_2$$

with probability at least $1-2\exp(-c\delta^2m)$. Hence, this inequality holds (simultaneously) far all $q\in Q$ with probability at least

$$1-2\#Q\exp(-c\delta^2m)\geq1-2(12/\delta)^s\exp(-c\delta^2m).$$

Now we want to prove that the desired inequality also holds for all $x\in X_T$. Let $\hat\delta>0$ be the smallest constant such that

$$\|Ax\|_2\leq(1+\hat\delta)\|x\|_2$$

holds for all $x \in X_T$ and let $v \in X_T$ be fixed with $\|v\|_2 = 1$. Then there is some $q \in Q$ with $\|v - q\|_2 \leq \delta/4$ and we get

$$\|Av\|_2 \leq \|Aq\|_2 + \|A(v-q)\|_2 \leq \left(1 + \frac{\delta}{2}\right) + (1+\hat{\delta})\frac{\delta}{4}$$

which implies

$$\hat{\delta} \leq \left(1 + \frac{\delta}{2}\right) + (1+\hat{\delta})\frac{\delta}{4} \quad \Rightarrow \quad \hat{\delta} < \delta.$$

3. step We already proved the inequality for every s-dimensional subspace. Since there are

$$(d)\ s \leq \left(\frac{ed}{s}\right)^s$$

possibilities to choose s indices out of d the claim follows.
Covering argument It remains to show the covering argument which we used for the set Q which we will prove in a more general setting:
Let X be an m-dimensional normed space, let $\varepsilon > 0$ and denote $B_X = \{x \in X \mid \|x\| \leq 1\}$. Then the covering number

$$N = \min\{n \in \mathbb{N} \mid \exists q_1, \ldots, q_n \in B_X : B_X \subset \bigcup_{i=1}^{n} (q_i + \varepsilon B_X)\}$$

can be bounded by

$$N \leq \left(\frac{2+2\varepsilon}{\varepsilon}\right)^m.$$

Indeed, let $Q = \{q_1, \ldots, q_k\}$ be (any) maximal set of points in B_X with

$$\|q_i - q_j\| > \varepsilon, \quad \text{for all } i \neq j.$$

Then it holds

- $B_X \subset \bigcup q_i + \varepsilon B_X$, since otherwise there is some $z \in B_X$ with $\|z - q_i\| > \varepsilon$ for alle $i = 1, \ldots, k$ in contradiction to the maximality of Q. Hence, we have $k \geq N$.
- The sets $q_i + \varepsilon/2 B_X$ are mutually disjoint. Assume that there exists i, j and

  $$z \in (q_i + \varepsilon/2 B_X) \cap (q_j + \varepsilon/2 B_X).$$

  It follows $\|q_i - q_j\| \leq \|q_i - z\| + \|q_j - z\| \leq \varepsilon/2 + \varepsilon/2$ which implies $i = j$.

We conclude

$$\bigcup_{i=1}^{k} q_i + \frac{\varepsilon}{2} B_X \subset \bigcup_{i=1}^{k} + \varepsilon B_X \subset \varepsilon B_X$$

and comparing the volumes we arrive at

$$\begin{aligned}\operatorname{vol}\left(\bigcup_{i=1}^{k} q_i + \frac{\varepsilon}{2}B_X\right) &= k \operatorname{vol}\left(\frac{\varepsilon}{2}B_X\right) = k\left(\frac{\varepsilon}{2}\right)^m \operatorname{vol}(B_X) \le \operatorname{vol}((1+\varepsilon)B_X) \\ &= (1+\varepsilon)^m \operatorname{vol}(B_X),\end{aligned}$$

hence

$$k\left(\frac{\varepsilon}{2}\right)^m \le (1+\varepsilon)^m \quad \Rightarrow \quad N \le k \le \left(\frac{2+2\varepsilon}{\varepsilon}\right)^m .$$

Matlab Exercises

1. Matlab Exercise

1. Implement the basis pursuit

$$\min_{x\in\mathbb{R}^d} \|x\|_1 \quad \text{subject to} \quad Ax = y$$

   in the form of a linear optimization problem.
   Hint: Matlab routine `linprog` can be useful.

2. Test your program for noisy measurements of the form $y = Ax + z$, where $z \in \mathbb{R}^m$ is either deterministic noise (i.e. $\|z\|$ is small) or random Gaussian noise (i.e. $z_i \sim \mathcal{N}(0, \sigma^2)$ and $\sigma > 0$ small).

2. Matlab Exercise

Show numerically that the number of measurements m only has to grow logarithmically in the dimension d if we want to recover an s-sparse signal $x_0 \in \mathbb{R}^d$ from linear measurements $y = Ax$ with $A \in \mathbb{R}^{m,d}$.
To show this, calculate for increasing values of d and m the error of your approximation and plot the resulting matrix.

3. Matlab Exercise

1. Implement the 1-Bit Compressed Sensing Algorithm

$$\max_{x\in\mathbb{R}^d} \sum_{i=1}^{m} y_i\langle a_i, x\rangle \quad \text{subject to} \quad \|x\|_1 \le R, \quad \|x\|_2 \le 1,$$

   which recovers the true signal $x_0 \in \mathbb{R}^d$ with $\|x_0\|_1 \le R$ and $\|x_0\|_2 \le 1$ from measurements $y_i = \operatorname{sign}\langle a_i, x_0\rangle$, $i = 1, \dots, m$.
   Hint: Matlab package `CVX` can be useful.

2. Test your algorithm with noisy measurements of the form

$$y_i = \begin{cases} \operatorname{sign}\langle a_i, x_0\rangle & \text{with probability } 1-p, \\ -\operatorname{sign}\langle a_i, x_0\rangle & \text{with probability } p \end{cases}$$

for some $0 < p < 1/2$.

4. Matlab Exercise

Let $f\colon B_d \to \mathbb{R}$ be a ridge function with $f(x) = g(\langle a, x\rangle)$ for some (unknown) s-sparse ridge vector $a \in \mathbb{R}^d$ with $\|a\|_2 = 1$ and some differentiable ridge profile $g\colon \mathbb{R} \to \mathbb{R}$ with $g'(0) \neq 0$. The ridge vector a gets recovered by the following algorithm:

- *Input:* Ridge function $f(x) = g(\langle a, x\rangle)$, $h > 0$ small and $m \in \mathbb{N}$
- Take $\Phi \in \mathbb{R}^{m\times d}$ a normalized Bernoulli matrix (i.e. with entries ±1, both with probability 1/2.
- Put $\tilde{b}_j := \frac{f(h\varphi_j)-f(0)}{h}$, $j = 1, \dots, m$
- Put $\tilde{a} := \Delta_1(\tilde{b}) = \arg\min_{w\in\mathbb{R}^d} \ \|w\|_1$ s.t. $\Phi w = \tilde{b}$
- Put $\hat{a} := \frac{\tilde{a}}{\|\tilde{a}\|_2}$
- *Output:* $\hat{a}$

Implement this algorithm and show numerically that it indeed recovers the ridge vector a.

Solution

Matlab implementations are given in the corresponding Matlab files.

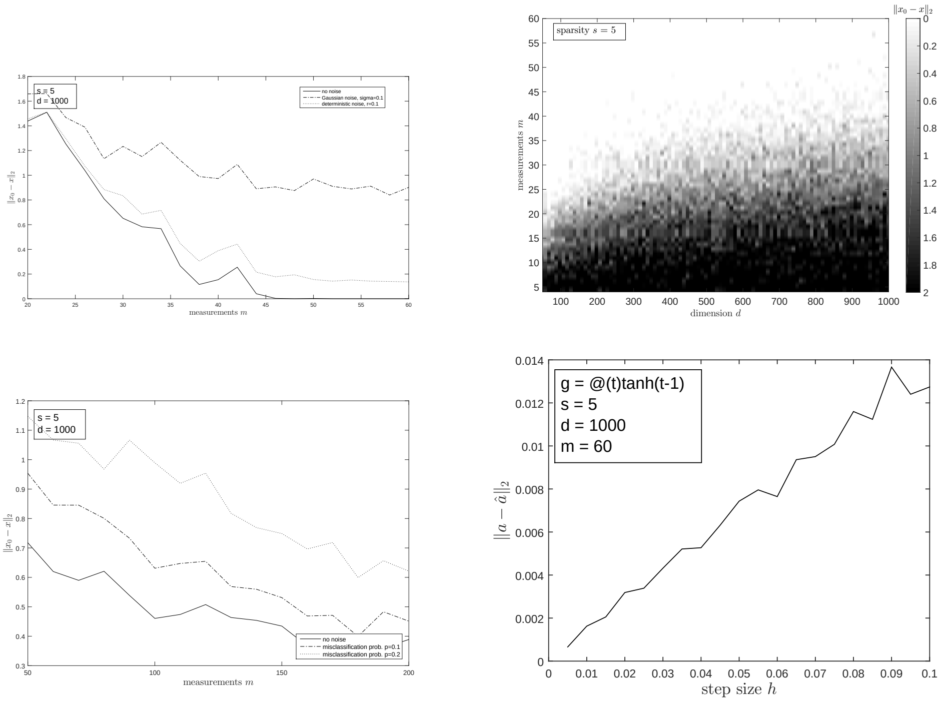

Figure 1: Top: Generated figure of `test_basis_pursuit` (left) and figure generated by `phase_transition` (right). Bottom: Generated figure of `test_one_bit` (left) and generated figure of `ridge_function` (right).