Lecture Notes 9: Constrained Optimization

1 Compressed sensing

1.1 Underdetermined linear inverse problems

Linear inverse problems model measurements of the form

$$A\vec{x} = \vec{y} \tag{1}$$

where the data $\vec{y} \in \mathbb{R}^n$ are the result of applying a linear operator represented by the matrix $A \in \mathbb{R}^{m \times n}$ to a signal $\vec{x} \in \mathbb{R}^m$. The aim is to recover $\vec{x}$ from $\vec{y}$, assuming we know A. Mathematically, this is exactly equivalent to the linear-regression problem discussed in Lecture Notes 6. The difference is that in linear regression the matrix consists of measured features, whereas in inverse problems the linear operator usually has a physical interpretation. For example, in imaging problems the operator depends on the optical system used to obtain the data.

Each entry of $\vec{y}$ can be interpreted as a separate measurement of $\vec{x}$

$$\vec{y}[i] = \langle A_{i:}, \vec{x} \rangle, \qquad 1 \leq i \leq n, \tag{2}$$

where $A_{i:}$ is the ith row of A. In many applications, it is desirable to reduce the number of measurements as much as possible. However, by basic linear algebra, the number of measurements must be at least equal to m. If $m > n$ the system of equations (1) is underdetermined. Even if A is full rank, its null space has dimension $m - n$ by Corollary 1.16 in Lecture Notes 2. Any signal of the form $\vec{x} + \vec{w}$ where $\vec{w}$ belongs to the null space of A is a solution to the system.

As we discussed in Lecture Notes 4 and 5, natural images, speech and other signals are often *compressible*: they can be represented as sparse combinations of predefined atoms such as sinusoids or wavelets. The goal of compressed sensing is to exploit the compressibility of signals in order to reconstruct them from a smaller number of measurements. The idea is that although it is impossible to recover an arbitrary m-dimensional signal from n measurements if $m > n$, it may be possible to recover an m-dimensional signal that is parametrized by an s-dimensional vector, as long as $s < n$. The simplest example of compressible structure is sparsity. We will mostly focus on this case to illustrate the main ideas behind compressed sensing.

Example 1.1 (Compressed sensing in magnetic-resonance imaging)**.** Magnetic resonance imaging (MRI) is a popular medical-imaging technique that measures the response of the atomic nuclei of body tissues to high-frequency radio waves when placed in a strong magnetic field. MRI measurements can be modeled as samples from the 2D or 3D Fourier transform of the object that is being imaged, for example a slice of a human brain. An estimate of the corresponding image

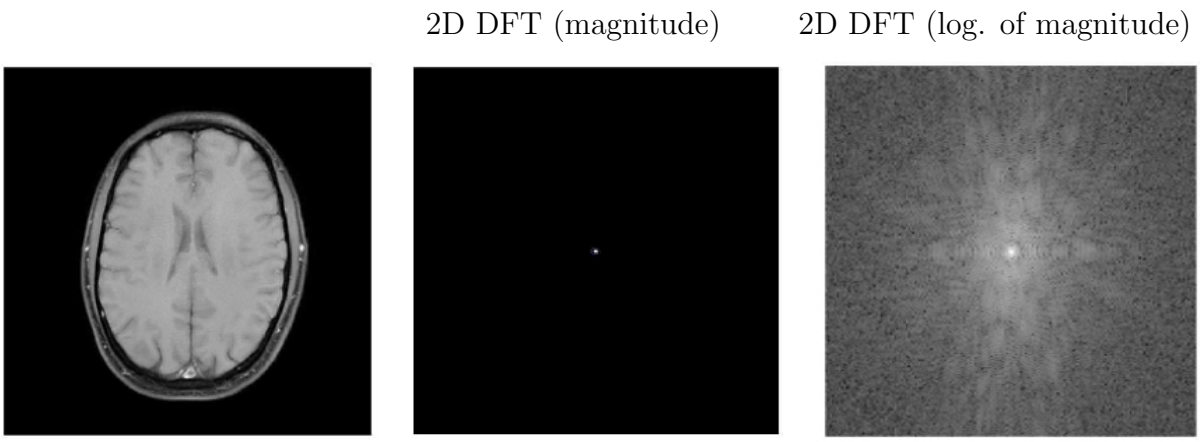

Figure 1: Image of a brain obtained by MRI, along with the magnitude of its 2D discrete Fourier transform (DFT) and the logarithm of this magnitude.

can be obtained by computing the inverse Fourier transform of the data, as shown in Figure 1. An important challenge in MRI is to reduce measurement time: long acquisition times are expensive and bothersome for the patients, especially for those that are seriously ill and for infants. Gathering less measurements, or equivalently undersampling the 2D or 3D Fourier transform of the image of interest, results in shorter data-acquisition times, but poses the challenge of recovering the image from undersampled data. Fortunately, MR images tend to be compressible in the wavelet domain. Compressed sensing of MR images consists of recovering the sparse wavelet coefficients from a small number of Fourier measurements. △

Example 1.2 (1D subsampled Fourier measurements)**.** This cartoon example is inspired by compressed sensing in MRI. We consider the problem of recovering a sparse signal from undersampled Fourier data. The rows of the measurement matrix are a subset of the rows of a DFT matrix, extracted following two strategies: regular and random subsampling. In regular subsampling we select the odd rows of the matrix, whereas in random subsampling we just select the rows uniformly at random. Figure 2 shows the real part of the matrices. Figure 3 shows the underdetermined linear system corresponding to each of the subsampling strategies for a simple example where the signal has sparsity 3. △

1.2 When is sparse estimation well posed?

A first question that arises when we consider sparse recovery from underdetermined measurements is under what conditions the problem is well posed. In other words, is it possible that there may be other sparse signals that produce the same measurements? If that is the case, it is impossible to determine which sparse signal actually generated the data and the problem is ill posed. Whether this situation may arise or not depends on the spark of the measurement matrix.

Definition 1.3 (Spark)**.** *The spark of a matrix is the smallest subset of columns that is linearly dependent.*

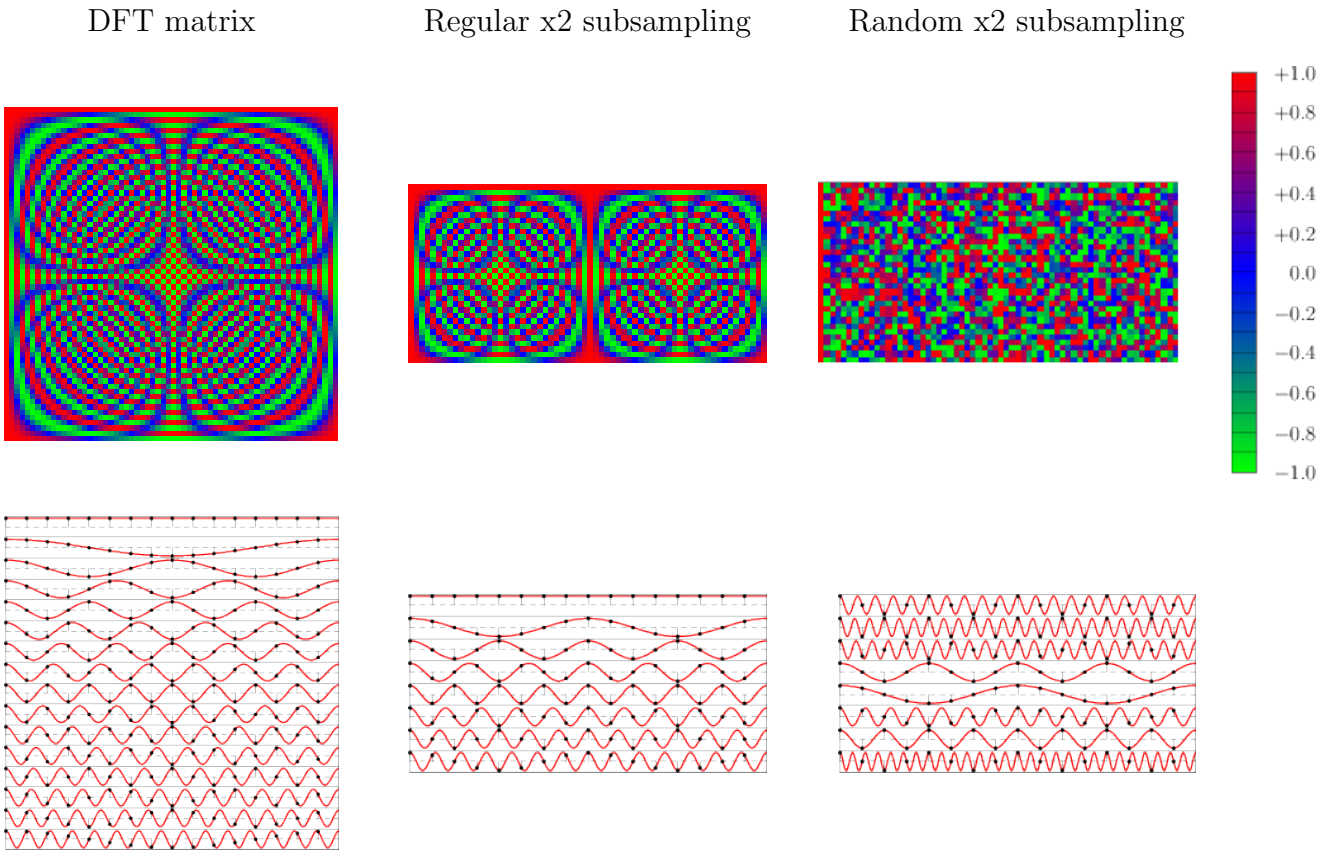

Figure 2: Real part of the DFT matrix, as well as the corresponding regularly-subsampled and randomly-subsampled measurement matrix, represented as a heat map (above) and as samples from continuous sinusoids (below).

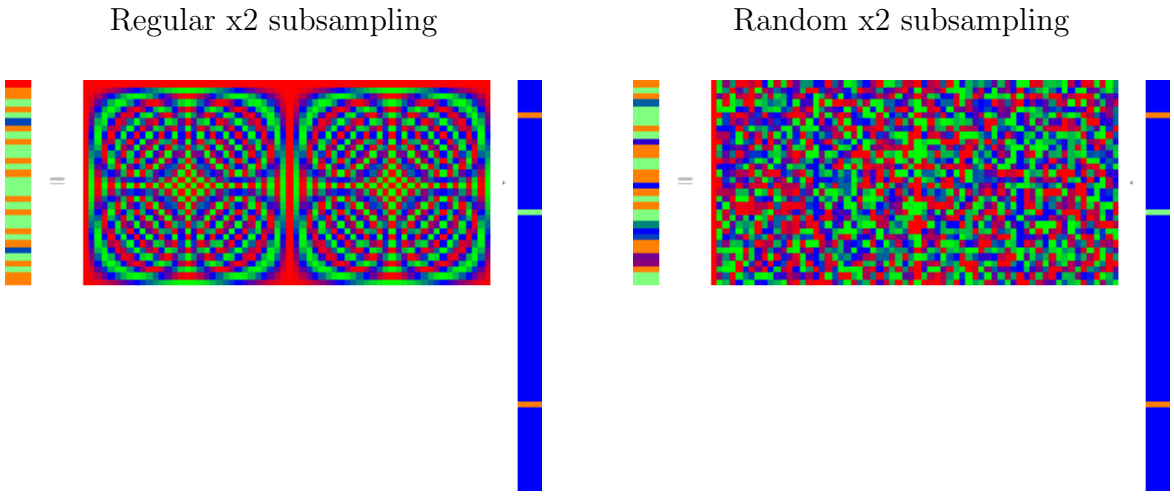

Figure 3: Underdetermined linear system of equations corresponding to the subsampled Fourier matrices in Figure 2.

The spark sets a fundamental limit to the sparsity of vectors that can be recovered uniquely from linear measurements.

Theorem 1.4. *Let $\vec{y} := A\vec{x}^*$, where $A \in \mathbb{R}^{m\times n}$, $\vec{y} \in \mathbb{R}^n$ and $\vec{x}^* \in \mathbb{R}^m$ is a sparse vector with s nonzero entries. The vector $\vec{x}^*$ is guaranteed to be the only vector with sparsity level equal to s consistent with the measurements, i.e. the solution of*

$$\min_{\vec{x}} ||\vec{x}||_0 \qquad \textit{subject to} \qquad A\vec{x} = \vec{y}, \tag{3}$$

for any choice of $\vec{x}^*$ if and only if

$$\text{spark}\,(A) > 2s. \tag{4}$$

Proof. $\vec{x}^*$ is the only sparse vector consistent with the data if and only if there is no other vector $\vec{x}'$ with sparsity s such that $A\vec{x}^* = A\vec{x}'$. This occurs for any choice of $\vec{x}^*$ if and only if for any pair of vectors $\vec{x}_1$ and $\vec{x}_2$ with sparsity level s, we have

$$A\,(\vec{x}_1 - \vec{x}_2) \neq \vec{0}. \tag{5}$$

Let T_1 and T_2 denote the support of the nonzero entries of $\vec{x}_1$ and $\vec{x}_2$. Equation (5) can be written as

$$A_{T_1\cup T_2}\vec{\alpha} \neq \vec{0} \qquad \text{for any } \vec{\alpha} \in \mathbb{R}^{|T_1\cup T_2|}. \tag{6}$$

This is equivalent to all submatrices with at most $2s$ columns (the difference between 2 s-sparse vectors can have at most $2s$ nonzero entries) having no nonzero vectors in their null space and therefore being full rank, which is exactly the meaning of $\text{spark}\,(A) > 2s$. □

If the spark of a matrix is greater than $2s$ then the matrix represents a linear operator that is invertible when restricted to act upon s-sparse signals. However, it may still be the case that two different sparse vectors could generate data that are extremely close, which would make it challenging to distinguish them if the measurements are noisy. In order to ensure that stable inversion is possible, we must in addition require that the *distance* between sparse vectors is preserved, so that if $\vec{x}_1$ is far from $\vec{x}_2$ then $A\vec{x}_1$ is guaranteed to be far from $A\vec{x}_2$. Mathematically, the linear operator should be an isometry when restricted to act upon sparse vectors.

Definition 1.5 (Restricted-isometry property)**.** *A matrix A satisfies the restricted-isometry property with constant κ_s if for any s-sparse vector $\vec{x}$*

$$(1-\kappa_s)\,||\vec{x}||_2 \leq ||A\vec{x}||_2 \leq (1+\kappa_s)\,||\vec{x}||_2\,. \tag{7}$$

If a matrix A satisfies the restricted-isometry property (RIP) for a sparsity level $2s$ then for any pair of vectors $\vec{x}_1$ and $\vec{x}_2$ with sparsity level s, the distance between their corresponding measurements $\vec{y}_1$ and $\vec{y}_2$ is lower bounded by the difference between the two vectors

$$||\vec{y}_2 - \vec{y}_1||_2 = A\,(\vec{x}_1 - \vec{x}_2) \tag{8}$$

$$\geq (1-\kappa_{2s})\,||\vec{x}_2 - \vec{x}_1||_2\,. \tag{9}$$

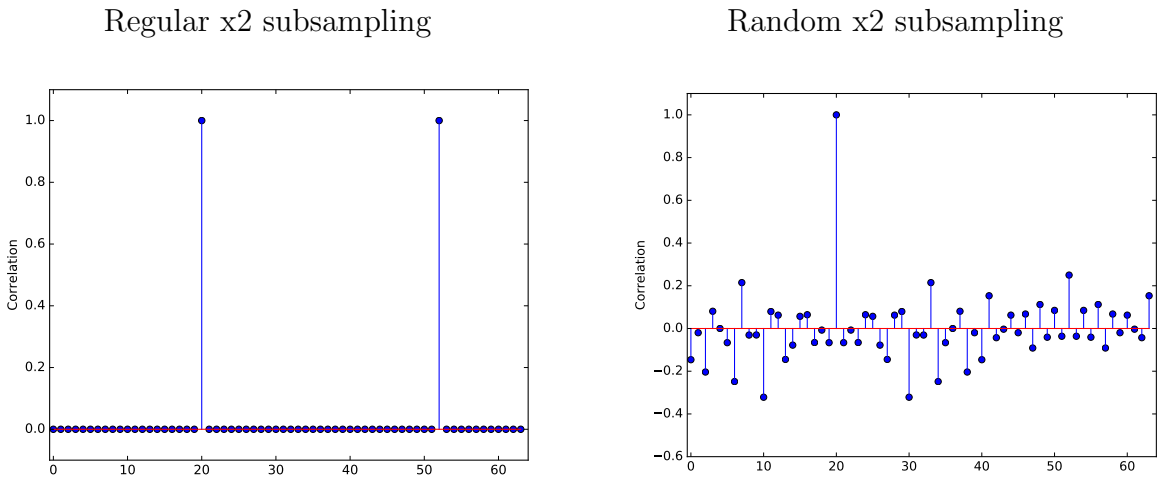

Figure 4: Correlation between the 20th column and the rest of the columns for the matrices described in Example 1.2.

Figure 4 shows the correlation between one of the columns in the matrices matrices described in Example 1.2 and the rest of the columns. For the regularly-subsampled Fourier matrix, there exists another column that is exactly the same. No method will be able to distinguish the data corresponding to even 1-sparse vectors, since the contributions of these two columns will be impossible to distinguish. The matrix does not even satisfy the RIP for a sparsity level equal to two.

In the case of the randomly-subsampled Fourier matrix, column 20 is not highly correlated with any other column. This does not immediately mean that the matrix satisfies the restricted-isometry property. Unfortunately, verifying that a matrix satisfies the spark or the restricted-isometry property is not computationally tractable (essentially, one has to check all possible sparse submatrices). However, we can prove that the RIP holds with high probability for random matrices. In the following theorem we prove this statement for Gaussian iid matrices. The proof for random Fourier measurements is more complicated [8, 10].

Theorem 1.6 (Restricted-isometry property for Gaussian matrices)**.** *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be a random matrix with iid standard Gaussian entries.* $\frac{1}{\sqrt{m}}\mathbf{A}$ *satisfies the restricted-isometry property for a constant* κ_s *with probability* $1 - \frac{C_2}{n}$ *as long as the number of measurements*

$$m \geq \frac{C_1 s}{\kappa_s^2} \log\left(\frac{n}{s}\right) \tag{10}$$

for two fixed constants $C_1, C_2 > 0$.

Proof. Let us fix an arbitrary support T of size s. The $m \times s$ submatrix $\mathbf{A}_T$ of $\mathbf{A}$ that contains the columns indexed by T has iid Gaussian entries, so by Theorem 3.7 in Lecture Notes 3 (in particular equation (81)), its singular values are bounded by

$$\sqrt{m}\,(1 - \kappa_s) \leq \boldsymbol{\sigma}_\mathbf{s} \leq \boldsymbol{\sigma}_\mathbf{1} \leq \sqrt{m}\,(1 + \kappa_s) \tag{11}$$

with probability at least

$$1 - 2\left(\frac{12}{\kappa_s}\right)^s \exp\left(-\frac{m\kappa_s^2}{32}\right). \tag{12}$$

This implies that for any vector $\vec{x}$ with support T

$$\sqrt{1-\kappa_s}\,||\vec{x}||_2 \leq \frac{1}{\sqrt{m}}\,||\mathbf{A}\vec{x}||_2 \leq \sqrt{1+\kappa_s}\,||\vec{x}||_2\,. \tag{13}$$

This is not enough for our purposes, we need this to hold for *all* supports of size s, i.e. on all possible combinations of s columns selected from the n columns in $\mathbf{A}$. A simple bound on the binomial coefficient yields the following bound on the number of such combinations

$$\binom{n}{s} \leq \left(\frac{en}{s}\right)^s. \tag{14}$$

By the union bound (Theorem 3.4 in Lecture Notes 3), we consequently have that the bounds (13) hold for any sparse-s vector with probability at least

$$\begin{aligned} 1 - 2\left(\frac{en}{s}\right)^s \left(\frac{12}{\kappa_s}\right)^s \exp\left(-\frac{m\kappa_s^2}{32}\right) &= 1 - \exp\left(\log 2 + s + s\log\left(\frac{n}{s}\right) + s\log\left(\frac{12}{\kappa_s}\right) - \frac{m\kappa_s^2}{2}\right) \\ &\leq 1 - \frac{C_2}{n} \end{aligned} \tag{15}$$

for some constant C_2 as long as m satisfies (10). □

1.3 Sparse recovery via ℓ_1-norm minimization

Choosing the sparsest vector consistent with the available data is computationally intractable, due to the nonconvexity of the ℓ_0 "norm" ball. Instead, we can minimize the ℓ_1 norm in order to promote sparse solutions.

Algorithm 1.7 (Sparse recovery via ℓ_1-norm minimization)**.** *Given data $\vec{y} \in \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{m\times n}$, the minimum-$\ell_1$-norm estimate is the solution to the optimization problem*

$$\min_{\vec{x}} ||\vec{x}||_1 \qquad \textit{subject to} \qquad A\vec{x} = \vec{y}. \tag{16}$$

Figure 5 shows the minimum ℓ_2- and ℓ_1-norm estimates of the sparse vector in the sparse recovery problem described in Example 1.2. In the case of the regularly-subsampled matrix, both methods yield erroneous solutions that are sparse. As discussed previously, for that matrix the sparse-recovery problem is ill posed. In the case of the randomly-subsampled matrix, ℓ_2-norm minimization promotes a solution that contains a lot of small entries. The reason is that large entries are very expensive because we are minimizing the square of the magnitudes. Such large entries are not as expensive for the ℓ_1-norm cost function. As a result, the algorithm produces a sparse solution that is exactly equal to the original signal. Figure 6 provides some geometric intuition as to why the ℓ_1-norm minimization problem promotes sparse solutions.

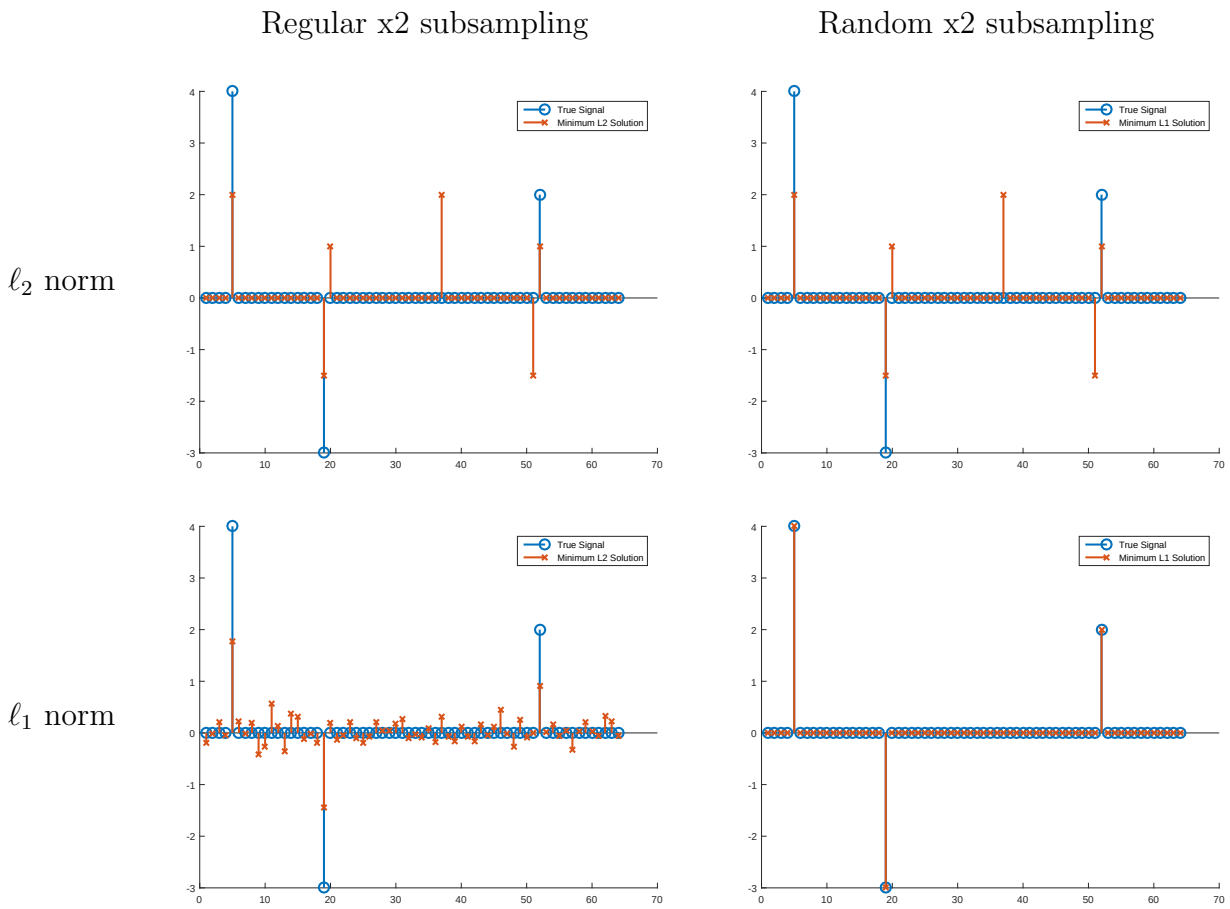

Figure 5: Minimum ℓ_2- and ℓ_1-norm estimates of the sparse vector in the sparse recovery problem described in Example 1.2.

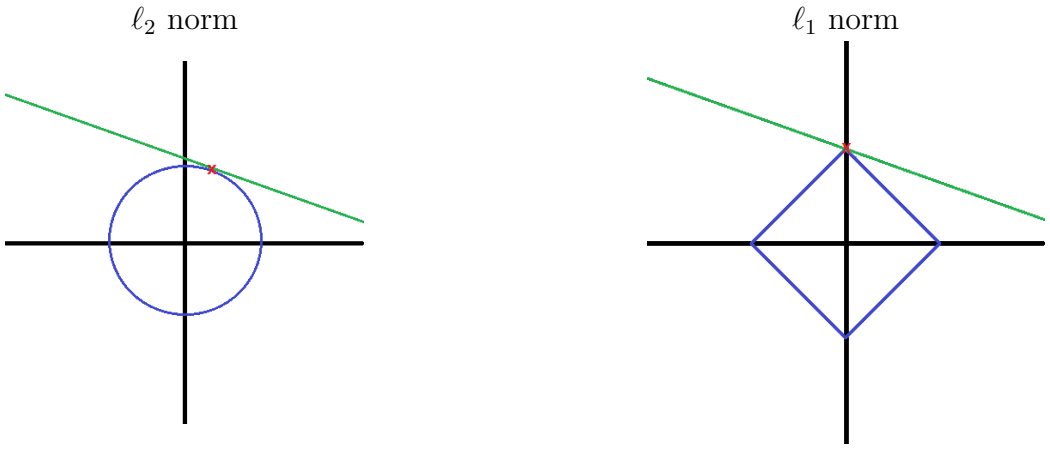

Figure 6: Cartoon of the ℓ_2- and ℓ_1-norm minimization problems for a two-dimensional signal. The lines represent the hyperplane of signals such that $A\vec{x} = \vec{y}$. The ℓ_1-norm ball is spikier, so that as a result the solution lies on a low-dimensional face of the norm ball. In contrast, the ℓ_2-norm ball is rounded and this does not occur.

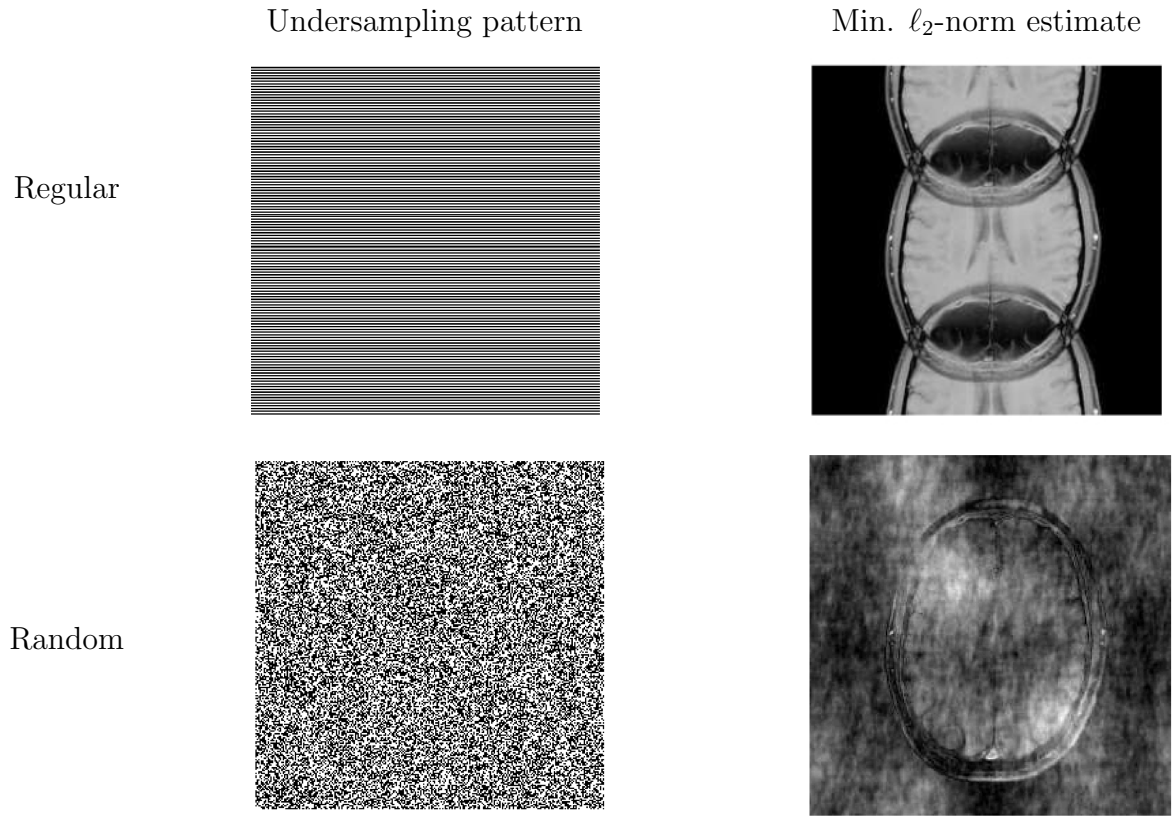

Figure 7: Two different sampling strategies in 2D k space: regular undersampling in one direction (top) and random undersampling (bottom). The original data is the same as in Figure 1. On the right we see the corresponding minimum-ℓ_2-norm estimate for each undersampling pattern.

Figure 8: Minimum-ℓ_1-norm for the two undersampling patterns shown in Figure 7.

1.4 Sparsity in a transform domain

If the signal is sparse in a transform domain, then we can modify the optimization problem to take this into account. Let W represent a wavelet transform, such that we assume that the corresponding wavelet coefficients of the image are sparse. In that case, we solve the optimization problem,

$$\min_{\vec{x}} ||\vec{c}||_1 \qquad \text{subject to} \qquad AW\vec{c} = \vec{y}. \tag{17}$$

If we want to recover the original $\vec{c}^*$ then we would need to verify that AW should satisfy the RIP, which would require analyzing the inner products between the rows of the measurement A (the measurement vectors) and the columns of W (the sparsifying basis functions). However, we might be fine with any $\vec{c}'$ such that $A\vec{c}' = \vec{x}^*$. In that case, characterizing when the problem is well posed is more challenging.

Figure 8 shows the result of applying ℓ_1-norm minimization to recover an image from the data corresponding to the images shown in Figure 7. For regular undersampling, then the estimate is essentially the same as the minimum-ℓ_2-norm estimate. This is not surprising, since the minimum-ℓ_2-norm estimate is also sparse in the wavelet domain because it is equal to a superposition of two shifted copies of the image. In contrast, ℓ_1-norm minimization recovers the original image perfectly when coupled with random projections. Intuitively, ℓ_1-norm minimization *cleans up* the noisy aliasing caused by random undersampling.

2 Constrained optimization

2.1 Convex sets

A set is convex if it contains all segments connecting points that belong to it.

Definition 2.1 (Convex set)**.** *A convex set $\mathcal{S}$ is any set such that for any $\vec{x}, \vec{y} \in \mathcal{S}$ and $\theta \in (0, 1)$*

$$\theta\vec{x} + (1 - \theta)\,\vec{y} \in \mathcal{S}. \tag{18}$$

Figure 9 shows a simple example of a convex and a nonconvex set.

The following lemma establishes that the intersection of convex sets is convex.

Lemma 2.2 (Intersection of convex sets)**.** *Let $\mathcal{S}_1, \ldots, \mathcal{S}_m$ be convex subsets of $\mathbb{R}^n$, $\cap_{i=1}^m \mathcal{S}_i$ is convex.*

Proof. Any $\vec{x}, \vec{y} \in \cap_{i=1}^m \mathcal{S}_i$ also belong to S_1. By convexity of $\mathcal{S}_1$ $\theta\vec{x} + (1 - \theta)\,\vec{y}$ belongs to S_1 for any $\theta \in (0, 1)$ and therefore also to $\cap_{i=1}^m \mathcal{S}_i$. □

The following theorem shows that projection onto non-empty closed convex sets is unique.

Theorem 2.3 (Projection onto convex set)**.** *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a non-empty closed convex set. The projection of any vector $vx \in \mathbb{R}^n$ onto $\mathcal{S}$*

$$\mathcal{P}_{\mathcal{S}}\left(\vec{x}\right) := \arg\min_{\vec{y} \in \mathcal{S}} ||\vec{x} - \vec{y}||_2 \tag{19}$$

exists and is unique.

Proof. Existence

Since $\mathcal{S}$ is non-empty we can choose an arbitrary point $\vec{y}\,' \in \mathcal{S}$. Minimizing $||\vec{x} - \vec{y}||_2$ over $\mathcal{S}$ is equivalent to minimizing $||\vec{x} - \vec{y}||_2$ over $\mathcal{S} \cap \{\vec{y} \mid ||\vec{x} - \vec{y}||_2 \leq ||\vec{x} - \vec{y}\,'||_2\}$. Indeed, the solution cannot be a point that is farther away from $\vec{x}$ than $\vec{y}\,'$. By Weierstrass's extreme-value theorem, the optimization problem

$$\begin{aligned} &\text{minimize} && ||\vec{x} - \vec{y}||_2^2 && (20)\\ &\text{subject to} && s \in \mathcal{S} \cap \{\vec{y} \mid ||\vec{x} - \vec{y}||_2 \leq ||\vec{x} - \vec{y}\,'||_2\} && (21) \end{aligned}$$

has a solution because $||\vec{x} - \vec{y}||_2^2$ is a continuous function and the feasibility set is bounded and closed, and hence compact. Note that this also holds if $\mathcal{S}$ is not convex.

Uniqueness

Assume that there are two distinct projections $\vec{y}_1 \neq \vec{y}_2$. Consider the point

$$\vec{y}\,' := \frac{\vec{y}_1 + \vec{y}_2}{2}, \tag{22}$$

which belongs to $\mathcal{S}$ because $\mathcal{S}$ is convex. The difference between $\vec{x}$ and $\vec{y}\,'$ and the difference between $\vec{y}_1$ and $\vec{y}\,'$ are orthogonal vectors,

$$\begin{aligned} \langle \vec{x} - \vec{y}\,', \vec{y}_1 - \vec{y}\,' \rangle &= \left\langle \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2}, \vec{y}_1 - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle && (23)\\ &= \left\langle \frac{\vec{x} - \vec{y}_1}{2} + \frac{\vec{x} - \vec{y}_2}{2}, \frac{\vec{x} - \vec{y}_1}{2} - \frac{\vec{x} - \vec{y}_2}{2} \right\rangle && (24)\\ &= \frac{1}{4}\left(||\vec{x} - \vec{y}_1||^2 + ||\vec{x} - \vec{y}_2||^2\right) && (25)\\ &= 0, && (26) \end{aligned}$$

because $||\vec{x} - \vec{y}_1|| = ||\vec{x} - \vec{y}_2||$ by assumption. By Pythagoras' theorem this implies

$$\begin{aligned} ||\vec{x} - \vec{y}_1||_2^2 &= ||\vec{x} - \vec{y}\,'||_2^2 + ||\vec{y}_1 - \vec{y}\,'||_2^2 && (27)\\ &= ||\vec{x} - \vec{y}\,'||_2^2 + \left|\left|\frac{\vec{y}_1 - \vec{y}_2}{2}\right|\right|_2^2 && (28)\\ &> ||\vec{x} - \vec{y}\,'||_2^2 && (29) \end{aligned}$$

because $\vec{y}_1 \neq \vec{y}_2$ by assumption. We have reached a contradiction, so the projection is unique. □

A convex combination of n points is any linear combination of the points with nonnegative coefficients that add up to one. In the case of two points, this is just the segment between the points.

Figure 9: An example of a nonconvex set (left) and a convex set (right).

Definition 2.4 (Convex combination)**.** *Given n vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n \in \mathbb{R}^n$,*

$$\vec{x} := \sum_{i=1}^{n} \theta_i \vec{x}_i \tag{30}$$

is a convex combination of $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ as along as the real numbers $\theta_1, \theta_2, \ldots, \theta_n$ are nonnegative and add up to one,

$$\theta_i \geq 0, \quad 1 \leq i \leq n, \tag{31}$$

$$\sum_{i=1}^{n} \theta_i = 1. \tag{32}$$

The convex hull of a set $\mathcal{S}$ contains all convex combination of points in $\mathcal{S}$. Intuitively, it is the smallest convex set that contains $\mathcal{S}$.

Definition 2.5 (Convex hull)**.** *The convex hull of a set $\mathcal{S}$ is the set of all convex combinations of points in $\mathcal{S}$.*

A possible justification of why we penalize the ℓ_1-norm to promote sparse structure is that the ℓ_1-norm ball is the convex hull of 1-sparse vectors with unit norm, which form the intersection between the ℓ_0 "norm" ball and the ℓ_∞-norm ball. The lemma is illustrated in 2D in Figure 10.

Lemma 2.6 (ℓ_1-norm ball)**.** *The ℓ_1-norm ball is the convex hull of the intersection between the ℓ_0 "norm" ball and the ℓ_∞-norm ball.*

Proof. We prove that the ℓ_1-norm ball $\mathcal{B}_{\ell_1}$ is equal to the convex hull of the intersection between the ℓ_0 "norm" ball $\mathcal{B}_{\ell_0}$ and the ℓ_∞-norm ball $\mathcal{B}_{\ell_\infty}$ by showing that the sets contain each other.

$\underline{\mathcal{B}_{\ell_1} \subseteq \mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)}$

Let $\vec{x}$ be an n-dimensional vector in $\mathcal{B}_{\ell_1}$. If we set $\theta_i := |\vec{x}[i]|$, where $\vec{x}[i]$ is the ith entry of $\vec{x}$ by

$\vec{x}[i]$, and $\theta_0 = 1 - \sum_{i=1}^{n} \theta_i$ we have $\sum_{i=0}^{n} \theta_i = 1$ by construction, $\theta_i = |\vec{x}[i]| \geq 0$ and

$$\theta_0 = 1 - \sum_{i=1}^{n+1} \theta_i \tag{33}$$

$$= 1 - ||\vec{x}||_1 \tag{34}$$

$$\geq 0 \qquad \text{because } \vec{x} \in \mathcal{B}_{\ell_1}. \tag{35}$$

We can express now $\vec{x}$ as a convex combination of the standard basis vectors multiplied by the sign of the entries of $\vec{x}$ $\text{sign}\left(\vec{x}[1]\right) \vec{e}_1$, $\text{sign}\left(\vec{x}[2]\right) \vec{e}_2$, $\ldots$, $\text{sign}\left(\vec{x}[n]\right) \vec{e}_n$, which belong to $\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}$ since they have a single nonzero entry with magnitude equal to one, and the zero vector $\vec{0}$, which also belongs to $\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}$,

$$\vec{x} = \sum_{i=1}^{n} \theta_i \, \text{sign}\left(\vec{x}[i]\right) \vec{e}_i + \theta_0 \vec{0}. \tag{36}$$

$\underline{\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right) \subseteq \mathcal{B}_{\ell_1}}$

Let $\vec{x}$ be an n-dimensional vector in $\mathcal{C}\left(\mathcal{B}_{\ell_0} \cap \mathcal{B}_{\ell_\infty}\right)$. By the definition of convex hull, we can write

$$\vec{x} = \sum_{i=1}^{m} \theta_i \vec{y}_i, \tag{37}$$

where $m > 0$, $\vec{y}_1, \ldots, \vec{y}_m \in \mathbb{R}^n$ have a single entry bounded by one, $\theta_i \geq 0$ for all $1 \leq i \leq m$ and $\sum_{i=1}^{m} \theta_i = 1$. This immediately implies $\vec{x} \in \mathcal{B}_{\ell_1}$, since

$$||\vec{x}||_1 \leq \sum_{i=1}^{m} \theta_i \, ||\vec{y}_i||_1 \quad \text{by the Triangle inequality} \tag{38}$$

$$\leq \sum_{i=1}^{m} \theta_i \, ||\vec{y}_i||_\infty \quad \text{because each } \vec{y}_i \text{ only has one nonzero entry} \tag{39}$$

$$\leq \sum_{i=1}^{m} \theta_i \tag{40}$$

$$\leq 1. \tag{41}$$

□

2.2 Constrained convex programs

In this section we discuss convex optimization problems, which are problems in which a convex function is minimized over a convex set.

Definition 2.7 (Convex optimization problem)**.** *An optimization problem is a convex optimization problem if it can be written in the form*

$$\begin{aligned} &\text{minimize} && f_0\left(\vec{x}\right) && (42)\\ &\text{subject to} && f_i\left(\vec{x}\right) \leq 0, \quad 1 \leq i \leq m, && (43)\\ &&& h_i\left(\vec{x}\right) = 0, \quad 1 \leq i \leq p, && (44) \end{aligned}$$

where $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ are functions satisfying the following conditions

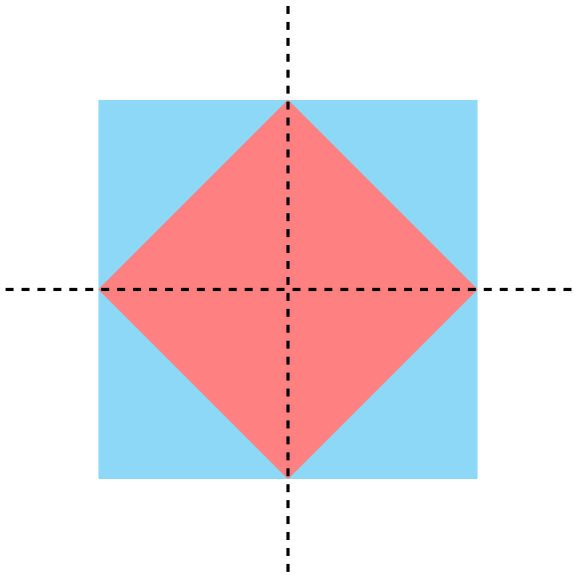

Figure 10: Illustration of Lemma (2.6) The ℓ_0 "norm" ball is shown in black, the ℓ_∞-norm ball in blue and the ℓ_1-norm ball in a reddish color.

- *The cost function f_0 is convex.*
- *The functions that determine the inequality constraints $f_1, \ldots, f_m$ are convex.*
- *The functions that determine the equality constraints $h_1, \ldots, h_p$ are affine, i.e. $h_i(\vec{x}) = \vec{a}_i^T \vec{x} + b_i$ for some $\vec{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.*

Any vector that satisfies all the constraints in a convex optimization problem is said to be *feasible*. A solution to the problem is any vector $\vec{x}^*$ such that for all feasible vectors $\vec{x}$

$$f_0(\vec{x}) \geq f_0(\vec{x}^*). \tag{45}$$

If a solution exists $f(\vec{x}^*)$ is the optimal value or optimum of the optimization problem.

Under the conditions in Definition 2.7 we can check that the feasibility set of the optimization problem is indeed convex. Indeed, it corresponds to the intersection of several convex sets: the 0-sublevel sets of $f_1, \ldots, f_m$, which are convex by Lemma 2.9 below, and the hyperplanes $h_i(\vec{x}) = \vec{a}_i^T \vec{x} + b_i$. The intersection is convex by Lemma 2.2.

Definition 2.8 (Sublevel set)**.** *The γ-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$, where $\gamma \in \mathbb{R}$, is the set of points in $\mathbb{R}^n$ at which the function is smaller or equal to γ,*

$$C_\gamma := \{\vec{x} \mid f(\vec{x}) \leq \gamma\}. \tag{46}$$

Lemma 2.9 (Sublevel sets of convex functions)**.** *The sublevel sets of a convex function are convex.*

Proof. If $\vec{x}, \vec{y} \in \mathbb{R}^n$ belong to the γ-sublevel set of a convex function f then for any $\theta \in (0, 1)$

$$f(\theta\vec{x} + (1-\theta)\vec{y}) \leq \theta f(\vec{x}) + (1-\theta) f(\vec{y}) \quad \text{by convexity of } f \tag{47}$$

$$\leq \gamma \tag{48}$$

because both $\vec{x}$ and $\vec{y}$ belong to the γ-sublevel set. We conclude that any convex combination of $\vec{x}$ and $\vec{y}$ also belongs to the γ-sublevel set. □

If both the cost function and the constraint functions of a convex optimization problem are affine, the problem is a linear program.

Definition 2.10 (Linear program)**.** *A linear program is a convex optimization problem of the form*

$$\begin{aligned} \text{minimize} \quad & \vec{a}^T \vec{x} && (49)\\ \text{subject to} \quad & \vec{c}_i^T \vec{x} \leq d_i, \quad 1 \leq i \leq m, && (50)\\ & A\vec{x} = \vec{b}. && (51) \end{aligned}$$

It turns out that ℓ_1-norm minimization can be cast as a linear program.

Theorem 2.11 (ℓ_1-norm minimization as a linear program)**.** *The optimization problem*

$$\begin{aligned} \text{minimize} \quad & ||\vec{x}||_1 && (52)\\ \text{subject to} \quad & A\vec{x} = \vec{b} && (53) \end{aligned}$$

can be recast as the linear program

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} \vec{t}[i] && (54)\\ \text{subject to} \quad & \vec{t}[i] \geq \vec{e_i}^T \vec{x}, && (55)\\ & \vec{t}[i] \geq -\vec{e_i}^T \vec{x}, && (56)\\ & A\vec{x} = \vec{b}. && (57) \end{aligned}$$

Proof. To show that the linear problem and the ℓ_1-norm minimization problem are equivalent, we show that they have the same set of solutions.

Let us denote an arbitrary solution of the linear program by $\left(\vec{x}^{\,\mathrm{lp}}, \vec{t}^{\,\mathrm{lp}}\right)$. For any solution $\vec{x}^{\,\ell_1}$ of the ℓ_1-norm minimization problem, we define $\vec{t}^{\,\ell_1}$ such that $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$. $\left(\vec{x}^{\,\ell_1}, \vec{t}^{\,\ell_1}\right)$ is feasible for the LP so

$$\begin{aligned} \left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 &= \sum_{i=1}^{n} \vec{t}^{\,\ell_1}[i] && (58)\\ &\geq \sum_{i=1}^{n} \vec{t}^{\,\mathrm{lp}}[i] \quad \text{by optimality of } \vec{t}^{\,\mathrm{lp}} && (59)\\ &\geq \left|\left|\vec{x}^{\,\mathrm{lp}}\right|\right|_1 \quad \text{by constraints (55) and (56).} && (60) \end{aligned}$$

This implies that any solution of the linear program is also a solution of the ℓ_1-norm minimization problem.

To prove the converse, we fix a solution $\vec{x}^{\,\ell_1}$ of the ℓ_1-norm minimization problem. Setting $\vec{t}^{\,\ell_1}[i] := \left|\vec{x}^{\,\ell_1}[i]\right|$, we show that $\left(\vec{x}^{\,\ell_1}, \vec{t}^{\,\ell_1}\right)$ is a solution of the linear program. Indeed,

$$\sum_{i=1}^{n} t_i^{\ell_1} = \left|\left|\vec{x}^{\,\ell_1}\right|\right|_1 \tag{61}$$

$$\leq \left|\left|\vec{x}^{\,\mathrm{lp}}\right|\right|_1 \quad \text{by optimality of } \vec{x}^{\,\ell_1} \tag{62}$$

$$\leq \sum_{i=1}^{n} \vec{t}^{\,\mathrm{lp}}[i] \quad \text{by constraints (55) and (56).} \tag{63}$$

□

If the cost function is a positive semidefinite quadratic form and the constraints are affine a convex optimization problem is called a quadratic program (QP).

Definition 2.12 (Quadratic program)**.** *A quadratic program is a convex optimization problem of the form*

$$\begin{aligned} \text{minimize} \quad & \vec{x}^T Q \vec{x} + \vec{a}^T \vec{x} && (64)\\ \text{subject to} \quad & \vec{c}_i^{\,T} \vec{x} \leq d_i, \quad 1 \leq i \leq m, && (65)\\ & A\vec{x} = \vec{b}, && (66) \end{aligned}$$

where $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite.

A corollary of Theorem 2.11 is that ℓ_1-norm regularized least squares can be cast as a QP.

Corollary 2.13 (ℓ_1-norm regularized least squares as a QP)**.** *The optimization problem*

$$\text{minimize} \quad ||A\vec{x} - y||_2^2 + \lambda \left|\left|\vec{x}\right|\right|_1 \tag{67}$$

can be recast as the quadratic program

$$\begin{aligned} \text{minimize} \quad & \vec{x}^T A^T A \vec{x} - 2\vec{y}^T \vec{x} + \lambda \sum_{i=1}^{n} \vec{t}[i] && (68)\\ \text{subject to} \quad & \vec{t}[i] \geq \vec{e_i}^{\,T} \vec{x}, && (69)\\ & \vec{t}[i] \geq -\vec{e_i}^{\,T} \vec{x}. && (70) \end{aligned}$$

2.3 Duality

The Lagrangian of an optimization problem combines the cost function and the constraints.

Definition 2.14. *The Lagrangian of the optimization problem in Definition 2.7 is defined as the cost function augmented by a weighted linear combination of the constraint functions,*

$$L\left(\vec{x}, \vec{\alpha}, \vec{\nu}\right) := f_0\left(\vec{x}\right) + \sum_{i=1}^{m} \vec{\alpha}[i]\, f_i\left(\vec{x}\right) + \sum_{j=1}^{p} \vec{\nu}[j]\, h_j\left(\vec{x}\right), \tag{71}$$

where the vectors $\vec{\alpha} \in \mathbb{R}^m, \vec{\nu} \in \mathbb{R}^p$ are called Lagrange multipliers or dual variables, whereas $\vec{x}$ is the primal variable.

The Lagrangian yields a family of lower bounds to the cost function of the optimization problem at every feasible point.

Lemma 2.15. *As long as $\vec{\alpha}[i] \geq 0$ for $1 \leq i \leq m$, the Lagrangian of the optimization problem in Definition 2.7 lower bounds the cost function at all feasible points, i.e. if $\vec{x}$ is feasible then*

$$L\left(\vec{x}, \vec{\alpha}, \vec{\nu}\right) \leq f_0\left(\vec{x}\right). \tag{72}$$

Proof. If $\vec{x}$ is feasible and $\vec{\alpha}[i] \geq 0$ for $1 \leq i \leq m$ then

$$\vec{\alpha}[i]\, f_i\left(\vec{x}\right) \leq 0, \tag{73}$$
$$\vec{\nu}[j]\, h_j\left(\vec{x}\right) = 0, \tag{74}$$

which immediately implies (72). □

Minimizing over the primal variable yields a family of lower bounds that only depends on the dual variables. We call the corresponding function the Lagrange dual function.

Definition 2.16 (Lagrange dual function)**.** *The Lagrange dual function is the infimum of the Lagrangian over the primal variable $\vec{x}$*

$$l\left(\vec{\alpha}, \vec{\nu}\right) := \inf_{\vec{x} \in \mathbb{R}^n} L\left(\vec{x}, \vec{\alpha}, \vec{\nu}\right). \tag{75}$$

Theorem 2.17 (Lagrange dual function as a lower bound of the primal optimum)**.** *Let $\vec{x}^*$ denote an optimal value of the optimization problem in Definition 2.7,*

$$l\left(\vec{\alpha}, \vec{\nu}\right) \leq \vec{x}^*, \tag{76}$$

as long as $\vec{\alpha}[i] \geq 0$ for $1 \leq i \leq n$.

Proof. The result follows directly from (72),

$$\vec{x}^* = f_0\left(\vec{x}^*\right) \tag{77}$$
$$\geq L\left(\vec{x}^*, \vec{\alpha}, \vec{\nu}\right) \tag{78}$$
$$\geq l\left(\vec{\alpha}, \vec{\nu}\right). \tag{79}$$

□

Optimizing the lower bound provided by the Lagrange dual function yields an optimization problem that is called the *dual* problem of the original optimization problem. The original problem is called the primal problem in this context.

Definition 2.18 (Dual problem)**.** *The dual problem of the optimization problem from Definition 2.7 is*

$$\text{maximize} \quad l\left(\vec{\alpha}, \vec{\nu}\right) \tag{80}$$
$$\text{subject to} \quad \vec{\alpha}[i] \geq 0, \quad 1 \leq i \leq m. \tag{81}$$

Note that the cost function is a pointwise supremum of linear (and hence convex) functions.

Lemma 2.19 (Supremum of convex functions)**.** *Pointwise supremum of a family of convex functions indexed by a set* $\mathcal{I}$

$$f_{\text{sup}}\left(\vec{x}\right) := \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right). \tag{82}$$

is convex.

Proof. For any $0 \leq \theta \leq 1$ and any $\vec{x}, \vec{y} \in \mathbb{R}$,

$$\begin{aligned} f_{\text{sup}}\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \sup_{i \in \mathcal{I}} f_i\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) && (83)\\ &\leq \sup_{i \in \mathcal{I}} \theta f_i\left(\vec{x}\right) + (1-\theta)\, f_i\left(\vec{y}\right) \quad \text{by convexity of the } f_i && (84)\\ &\leq \theta \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right) + (1-\theta) \sup_{j \in \mathcal{I}} f_j\left(\vec{y}\right) && (85)\\ &= \theta f_{\text{sup}}\left(\vec{x}\right) + (1-\theta)\, f_{\text{sup}}\left(\vec{y}\right) && (86) \end{aligned}$$

□

As a result of the lemma, the dual problem is a convex optimization problem even if the primal is nonconvex! The following result, which is an immediate corollary to Theorem 2.17, states that the optimum of the dual problem is a lower bound for the primal optimum. This is known as weak duality.

Corollary 2.20 (Weak duality)**.** *Let* $\vec{x}^*$ *denote an optimum of the optimization problem in Definition 2.7 and* d^* *an optimum of the corresponding dual problem,*

$$d^* \leq \vec{x}^*. \tag{87}$$

In the case of convex functions, the optima of the primal and dual problems are often equal, i.e.

$$d^* = \vec{x}^*. \tag{88}$$

This is known as strong duality. A simple condition that guarantees strong duality for convex optimization problems is Slater's condition.

Definition 2.21 (Slater's condition)**.** *A vector* $x \in \mathbb{R}^n$ *satisfies Slater's condition for a convex optimization problem if*

$$\begin{aligned} f_i\left(\vec{x}\right) < 0, \quad 1 \leq i \leq m, && (89)\\ Ax = b. && (90) \end{aligned}$$

A proof of strong duality under Slater's condition can be found in Section 5.3.2 of [2].

The following theorem derives the dual problem for the ℓ_1-norm minimization problem with equality constraints.

Theorem 2.22 (Dual of ℓ_1-norm minimization)**.** *The dual of the optimization problem of*

$$\min_{\vec{x}} ||\vec{x} \in \mathbb{R}^m||_1 \qquad \text{subject to} \qquad A\vec{x} = \vec{y} \tag{91}$$

is

$$\max_{\vec{\nu} \in \mathbb{R}^n} \quad \vec{y}^T \vec{\nu} \qquad \text{subject to} \qquad \left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1. \tag{92}$$

Proof. The Lagrangian is equal to

$$L\left(\vec{x}, \vec{\nu}\right) = ||\vec{x}||_1 + \vec{\nu}^T \left(\vec{y} - A\vec{x}\right), \tag{93}$$

so the Lagrange dual function equals

$$l\left(\vec{\alpha}, \vec{\nu}\right) := \inf_{\vec{x} \in \mathbb{R}^n} ||\vec{x}||_1 - (A^T \vec{\nu})^T \vec{x} + \vec{\nu}^T \vec{y}. \tag{94}$$

If $(A^T \vec{\nu})[i] > 1$ then one can set $\vec{x}[i]$ arbitrarily large so that $l\left(\vec{\alpha}, \vec{\nu}\right) \to -\infty$. The same happens if $(A^T \vec{\nu})[i] < 1$. If $\left|\left|A^T \vec{\nu}\right|\right|_\infty \leq 1$, by Hölder's inequality (Theorem 3.16 in Lecture Notes 1)

$$\begin{aligned} \left|(A^T \vec{\nu})^T \vec{x}\right| &\leq ||\vec{x}||_1 \left|\left|A^T \vec{\nu}\right|\right|_\infty & (95) \\ &\leq ||\vec{x}||_1, & (96) \end{aligned}$$

so the Lagrangian is minimized by setting $\vec{x}$ to zero and $l\left(\vec{\alpha}, \vec{\nu}\right) = \vec{\nu}^T \vec{y}$. This completes the proof. □

Interestingly, the solution to the dual of the ℓ_1-norm minimization problem can often be used to estimate the support of the primal solution. Figure 11 shows that the vector $A^T \vec{\nu}^*$, where A is the underdetermined linear operator and $\vec{\nu}^*$ is a solution to Problem (92), reveals the support of the original signal for the randomly-subsampled data in Example 1.2.

Lemma 2.23. *If the exists a feasible vector for the primal problem, then the solution $\vec{\nu}^*$ to Problem (92) satisfies*

$$(A^T \vec{\nu}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0 \tag{97}$$

for all solutions $\vec{x}^*$ to the primal problem.

Proof. If there is a feasible vector for the primal problem, then strong duality holds because the optimization problem is a linear program with finite cost function. By strong duality

$$\begin{aligned} ||\vec{x}^*||_1 &= \vec{y}^T \vec{\nu}^* & (98) \\ &= (A\vec{x}^*)^T \vec{\nu}^* & (99) \\ &= (\vec{x}^*)^T (A^T \vec{\nu}^*) & (100) \\ &= \sum_{i=1}^{m} (A^T \vec{\nu}^*)[i] \vec{x}^*[i]. & (101) \end{aligned}$$

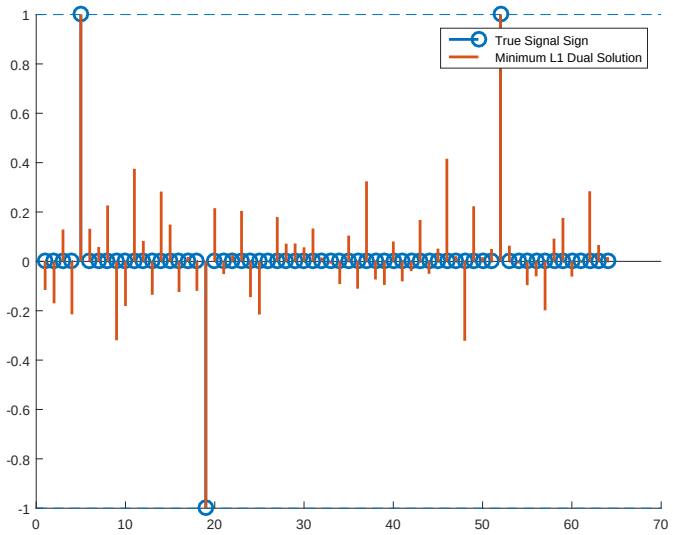

Figure 11: The vector $A^T\vec{\nu}^*$, where A is the underdetermined linear operator and $\vec{\nu}^*$ is a solution to Problem (92), reveals the support of the original signal for the randomly-subsampled data in Example 1.2.

By Hölder's inequality

$$||\vec{x}^*||_1 \geq \sum_{i=1}^{m} (A^T\vec{\nu}^*)[i]\vec{x}^*[i] \tag{102}$$

with equality if and only if

$$(A^T\vec{\nu}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0. \tag{103}$$

□

Consider the following algorithm for sparse recovery. Our goal is to find nonzero locations of a sparse vector $\vec{x}$ from $\vec{y} = A\vec{x}$. We have access to inner products of $\vec{x}$ and $A^T\vec{w}$ for any $\vec{w}$, since

$$\vec{y}^T\vec{w} = (A\vec{x})^T\vec{w} \tag{104}$$

$$= \vec{x}^T(A^T\vec{w}). \tag{105}$$

This suggests maximizing $A^T\vec{w}$, while bounding its magnitude entries by 1. In that case, the entries where $\vec{x}$ is nonzero should saturate to 1 or -1. This is exactly Problem (92)!

3 Analysis of constrained convex programs

3.1 Minimum ℓ_2-norm solution

The best-case scenario for the analysis of constrained convex program is that the optimization problem has a closed-form solution. This is the case for ℓ_2-norm minimization.

Theorem 3.1. *Let $A \in \mathbb{R}^{m \times n}$ be a full rank matrix such that $m < n$. For any $\vec{y} \in \mathbb{R}^n$ the solution to the optimization problem*

$$\arg\min_{\vec{x}} ||\vec{x}||_2 \qquad \text{subject to} \qquad A\vec{x} = \vec{y}. \tag{106}$$

is

$$\vec{x}^* := VS^{-1}U^T\vec{y} = A^T\left(A^TA\right)^{-1}\vec{y}. \tag{107}$$

Proof. Let us decompose $\vec{x}$ into its projection on the row space of A and on its orthogonal complement

$$\vec{x} = \mathcal{P}_{\text{row}(A)}\,\vec{x} + \mathcal{P}_{\text{row}(A)^\perp}\,\vec{x}. \tag{108}$$

Let $A = USV^T$ be the reduced SVD of A where S contains the nonzero singular values. Since A is full rank V contains an orthonormal basis of row (A) and we can write $\mathcal{P}_{\text{row}(A)}\,\vec{x} = V\vec{c}$ for some vector $\vec{c} \in \mathbb{R}^n$. We have

$$\begin{aligned} A\vec{x} &= A\mathcal{P}_{\text{row}(A)}\,\vec{x} && (109)\\ &= USV^TV\vec{c} && (110)\\ &= US\vec{c}. && (111) \end{aligned}$$

So that the equality constraint is equivalent to

$$US\vec{c} = \vec{y}, \tag{112}$$

where US is square and invertible because A is full rank, so that

$$\vec{c} = S^{-1}U^T\vec{y} \tag{113}$$

and hence for all feasible vectors $\vec{x}$

$$\mathcal{P}_{\text{row}(A)}\,\vec{x} = VS^{-1}U^T\vec{y}. \tag{114}$$

By Pythagoras' theorem, minimizing $||\vec{x}||_2$ is equivalent to minimizing

$$||\vec{x}||_2^2 = \left|\left|\mathcal{P}_{\text{row}(A)}\,\vec{x}\right|\right|_2^2 + \left|\left|\mathcal{P}_{\text{row}(A)^\perp}\,\vec{x}\right|\right|_2^2. \tag{115}$$

Since $\mathcal{P}_{\text{row}(A)}\,\vec{x}$ is fixed by the equality constraint, the minimum is achieved by setting $\mathcal{P}_{\text{row}(A)^\perp}\,\vec{x}$ to zero and the solution equals

$$\vec{x}^* := VS^{-1}U^T\vec{y} = A^T\left(A^TA\right)^{-1}\vec{y}. \tag{116}$$

□

The next lemma exploits the closed-form solution of the minimum ℓ_2-norm to explain the aliasing that occurs for the regularly-subsampled data in Figure 5.

Lemma 3.2 (Regular subsampling). *Let A be the regularly-subsampled DFT matrix in Example 1.2 and let*

$$\vec{x} := \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \tag{117}$$

be the original signal. The minimum ℓ_2-norm estimate equals

$$\vec{x}_{\ell_2} := \arg\min_{A\vec{x}=\vec{y}} ||\vec{x}||_2 \tag{118}$$

$$= \frac{1}{2} \begin{bmatrix} \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \\ \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \end{bmatrix}. \tag{119}$$

Proof. As is obvious in Figure 2 and we discussed in Example 1.15 of Lecture Notes 4, the matrix A is equal to two concatenated DFT matrices of size $m/2$ (for simplicity we assume m is even)

$$A := \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix}, \tag{120}$$

where $F^*_{m/2} F_{m/2} = F_{m/2} F^*_{m/2} = I$. By Theorem 3.1

$$\vec{x}_{\ell_2} = A^T \left(A^T A \right)^{-1} \vec{y} \tag{121}$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \right)^{-1} \frac{1}{\sqrt{2}} \begin{bmatrix} F_{m/2} & F_{m/2} \end{bmatrix} \begin{bmatrix} \vec{x}_{\text{up}} \\ \vec{x}_{\text{down}} \end{bmatrix} \tag{122}$$

$$= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} \left(\frac{1}{2} \left[F_{m/2} F^*_{m/2} + F_{m/2} F^*_{m/2} \right] \right)^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \tag{123}$$

$$= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \\ F^*_{m/2} \end{bmatrix} I^{-1} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \tag{124}$$

$$= \frac{1}{2} \begin{bmatrix} F^*_{m/2} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \\ F^*_{m/2} \left(F_{m/2} \vec{x}_{\text{up}} + F_{m/2} \vec{x}_{\text{down}} \right) \end{bmatrix} \tag{125}$$

$$= \frac{1}{2} \begin{bmatrix} \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \\ \vec{x}_{\text{up}} + \vec{x}_{\text{down}} \end{bmatrix}. \tag{126}$$

□

3.2 Minimum ℓ_1-norm solution

Unfortunately, the solution to ℓ_1-norm minimization with linear constraints does not have a closed-form solution. When we considered unconstrained nondifferentiable convex problems without closed-form solutions in Lecture Notes 8, we characterized the solutions by exploiting the fact that the zero vector is a subgradient of a convex cost function at a point if and only if the point is a minimizer. Here we will use a different argument based on the dual problem (which can often also be interpreted geometrically in terms of subgradients as we discuss below). The main idea is to construct a dual feasible vector whose existence implies that the original signal which we aim to recover is the unique solution to the primal.

Consider a certain sparse vector $\vec{x}^* \in \mathbb{R}^m$ with support T such that $A\vec{x}^* = \vec{y}$. If there exists a vector $\vec{\nu} \in \mathbb{R}^n$ such that $A^T\vec{\nu}$ is equal to the sign of $\vec{x}$ on T and has magnitude smaller than one elsewhere, then ν is feasible for the dual problem 92, so by weak duality $||\vec{x}||_1 \geq \vec{y}^T\vec{\nu}$ for any $\vec{x} \in \mathbb{R}^m$ that is feasible for the primal. We then have

$$\begin{aligned} ||\vec{x}||_1 &\geq \vec{y}^T\vec{\nu} && (127)\\ &= (A\vec{x}^*)^T\vec{\nu} && (128)\\ &= (\vec{x}^*)^T(A^T\vec{\nu}) && (129)\\ &= \sum_{i=1}^{m} \vec{x}^*[i]\,\mathrm{sign}(\vec{x}^*[i]) && (130)\\ &= ||\vec{x}^*||_1 . && (131) \end{aligned}$$

Geometrically, $A^T\vec{\nu}$ is a subgradient of the ℓ_1 norm at $\vec{x}^*$. The subgradient is orthogonal to the feasibility hyperplane given by $A\vec{x} = \vec{y}$ (any vector $\vec{v}$ within the hyperplane is the difference between two feasible vectors and therefore satisfies $A\vec{v} = \vec{0}$). As a result, for any other feasible vector $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 &\geq ||\vec{x}^*||_1 + (A^T\vec{\nu})^T(\vec{x} - \vec{x}^*) && (132)\\ &= ||\vec{x}^*||_1 + \vec{\nu}^T(A\vec{x} - A\vec{x}^*) && (133)\\ &= ||\vec{x}^*||_1 . && (134) \end{aligned}$$

These two arguments show that the existence of a certain dual vector can be used to establish that a certain primal feasible vector is a solution, but they do not establish uniqueness. It turns out that requiring that the magnitude of $A^T\vec{\nu}$ be strictly smaller than one on T^c is enough to guarantee it (as long as A is full rank). In that case, we call the dual variable $\vec{\nu}$ a dual certificate for the ℓ_1-norm minimization problem.

Theorem 3.3 (Dual certificate for ℓ_1-norm minimization)**.** *Let $\vec{x}^* \in \mathbb{R}^m$ with support T such that $A\vec{x}^* = \vec{y}$ and the submatrix A_T containing the columns of A indexed by T is full rank. If there exists a vector $\vec{\nu} \in \mathbb{R}^n$ such that*

$$\begin{aligned} (A^T\vec{\nu})[i] &= \mathrm{sign}(\vec{x}^*[i]) \qquad \text{if } \vec{x}^*[i] \neq 0 && (135)\\ \left|(A^T\vec{\nu})[i]\right| &< 1 \qquad \text{if } \vec{x}^*[i] = 0 && (136) \end{aligned}$$

then $\vec{x}^*$ is the unique solution to the ℓ_1-norm minimization problem (16).

Proof. For any feasible $\vec{x} \in \mathbb{R}^m$, let $\vec{h} := \vec{x} - \vec{x}^*$. If A_T is full rank then $\vec{h}_{T^c} \neq 0$ unless $\vec{h} = 0$ because otherwise $\vec{h}_T$ would be a nonzero vector in the null space of A_T. Condition (136) implies

$$\left|\left|\vec{h}_{T^c}\right|\right|_1 > (A^T\vec{\nu})^T\vec{h}_{T^c}, \tag{137}$$

where $\vec{h}_{T^c}$ denotes $\vec{h}$ restricted to the entries indexed by T^c. Let $\mathcal{P}_T(\cdot)$ denote a projection that sets to zero all entries of a vector except the ones indexed by T. We have

$$\begin{aligned} ||\vec{x}||_1 &= \left|\left|\vec{x}^* + \mathcal{P}_T(\vec{h})\right|\right|_1 + \left|\left|\vec{h}_{T^c}\right|\right|_1 \qquad \text{because } \vec{x}^* \text{ is supported on } T && (138)\\ &> ||\vec{x}^*||_1 + (A^T\vec{\nu})^T\mathcal{P}_T(\vec{h}) + (A^T\vec{\nu})^T\mathcal{P}_{T^c}(\vec{h}) \qquad \text{by (137)} && (139)\\ &= ||\vec{x}^*||_1 + \vec{\nu}^T A\vec{h} && (140)\\ &= ||\vec{x}^*||_1 . && (141) \end{aligned}$$

□

A strategy to prove that compressed sensing succeeds for a class of signals is to propose a dual-certificate construction and show that it produces a valid certificate for any signal in the class. We illustrate this with Gaussian matrices, but similar arguments can be extended to random Fourier matrices [7] and other measurements [4] (see also [6] for a more recent proof technique based on approximate dual certificates that provides better guarantees). It is also worth noting that the restricted-isometry property can be directly used to prove exact recovery via ℓ_1-norm minimization [5], but this technique is less general. Dual certificates can also be used to analyze other problems such as matrix completion, super-resolution and phase retrieval.

Theorem 3.4 (Exact recovery via ℓ_1-norm minimization)**.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a random matrix with iid standard Gaussian entries and $\vec{x}^* \in \mathbb{R}^m$ a vector with s nonzero entries such that $\mathbf{A}\vec{x}^* = \vec{y}$. Then $\vec{x}^*$ is the unique solution to the ℓ_1-norm minimization problem (16) with probability at least $1 - \frac{1}{n}$ as long as the number of measurements satisfies*

$$m \geq Cs \log n, \tag{142}$$

for a fixed numerical constant C.

Proof. By Theorem 3.3 all we need to show is that for any support T of size s and any possible sign pattern $\vec{w} := \vec{x}^*_T \in \mathbb{R}^s$ of the nonzero entries of $\vec{x}^*$ there exists a valid dual certificate $\vec{\nu}$. The certificate must satisfy

$$\mathbf{A}_T^T \vec{\nu} = \vec{w}. \tag{143}$$

Ideally we would like to analyze the vector $\vec{\nu}$ satisfying this underdetermined system of s equations such that $\mathbf{A}_{T^c}^T \vec{\nu}$ has the smallest possible ℓ_∞ norm. Unfortunately, the solution to the optimization problem

$$\min_{\nu} \left|\left|\mathbf{A}_{T^c}^T \vec{\nu}\right|\right|_\infty \qquad \text{subject to} \qquad \mathbf{A}_T^T \vec{\nu} = \vec{w} \tag{144}$$

does not have a closed-form solution. However, the solution to

$$\min_{\vec{\nu}} ||\vec{\nu}||_2 \qquad \text{subject to} \qquad \mathbf{A}_T^T \vec{\nu} = \vec{w} \tag{145}$$

does, so we can analyze it instead. By Theorem 3.1 the solution is

$$\vec{\boldsymbol{\nu}}_{\ell_2} := \mathbf{A}_T^T \left(\mathbf{A}_T^T \mathbf{A}_T\right)^{-1} \vec{w}. \tag{146}$$

To control $\vec{\boldsymbol{\nu}}_{\ell_2}$ we resort to the bound on the singular values of a fixed $m \times s$ submatrix in equation (11). Setting $\kappa_s := 0.5$ we denote by $\mathcal{E}$ the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma_s} \leq \boldsymbol{\sigma_1} \leq 1.5\sqrt{m}, \tag{147}$$

where

$$\mathrm{P}\left(\mathcal{E}\right) \geq 1 - \exp\left(-C'\frac{m}{s}\right) \tag{148}$$

for a fixed constant C'. Conditioned on $\mathcal{E}$ $\mathbf{A}_T$ is full rank and $\mathbf{A}_T^T\mathbf{A}_T$ is invertible, so $\vec{\boldsymbol{\nu}}_{\ell_2}$ guarantees condition (135). In order to verify condition (136), we need to bound $\mathbf{A}_i^T\vec{\boldsymbol{\nu}}_{\ell_2}$ for all indices $i \in T^c$. Let $\mathbf{USV}^T$ be the SVD of $\mathbf{A}_T$. Conditioned on $\mathcal{E}$ we have

$$\begin{aligned} ||\vec{\boldsymbol{\nu}}_{\ell_2}||_2 &= \left|\left|\mathbf{VS}^{-1}\mathbf{U}^T\vec{w}\right|\right|_2 && (149)\\ &\leq \frac{||\vec{w}||_2}{\boldsymbol{\sigma}_{\mathbf{s}}} && (150)\\ &\leq 2\sqrt{\frac{s}{m}}. && (151)\end{aligned}$$

For a fixed $i \in T^c$ and a fixed vector $\vec{v} \in R^n$, $\mathbf{A}_i^T\vec{v}/\left|\left|\vec{v}\right|\right|_2$ is a standard Gaussian random variable, which implies

$$\begin{aligned} \mathrm{P}\left(\left|\mathbf{A}_i^T\vec{v}\right| \geq 1\right) &= \mathrm{P}\left(\frac{\left|\mathbf{A}_i^T\vec{v}\right|}{||\vec{v}||_2} \geq ||\vec{v}||_2\right) && (152)\\ &\leq 2\exp\left(-\left|\left|\vec{v}\right|\right|_2^2/2\right) && (153)\end{aligned}$$

by the following lemma.

Lemma 3.5 (Proof in Section 4.1)**.** *For a Gaussian random variable* $\mathbf{u}$ *with zero mean and unit variance and any* $t > 0$

$$\mathrm{P}\left(|\mathbf{u}| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2}\right). \tag{154}$$

Note that if $i \notin T$ then $\mathbf{A}_i$ and $\vec{\boldsymbol{\nu}}_{\ell_2}$ are independent (they depend on different and hence independent entries of $\mathbf{A}$). This means that due to equation 151

$$\begin{aligned} \mathrm{P}\left(\left|\mathbf{A}_i^T\vec{\boldsymbol{\nu}}_{\ell_2}\right| \geq 1 \,|\, \mathcal{E}\right) &= \mathrm{P}\left(\left|\mathbf{A}_i^T\vec{v}\right| \geq 1 \quad \text{for} \quad ||\nu||_2 \leq 2\sqrt{\frac{s}{m}}\right) && (155)\\ &\leq 2\exp\left(-\frac{m}{8s}\right). && (156)\end{aligned}$$

As a result,

$$\begin{aligned} \mathrm{P}\left(\left|\mathbf{A}_i^T\vec{\boldsymbol{\nu}}_{\ell_2}\right| \geq 1\right) &\leq \mathrm{P}\left(\left|\mathbf{A}_i^T\vec{\boldsymbol{\nu}}_{\ell_2}\right| \geq 1 \,|\, \mathcal{E}\right) + \mathrm{P}\left(\mathcal{E}^c\right) && (157)\\ &\leq 2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right). && (158)\end{aligned}$$

We now apply the union bound to obtain a bound that holds for all $i \in T^c$. Since T^c has cardinality at most n

$$\mathrm{P}\left(\bigcup_{i\in T^c}\left\{\left|\mathbf{A}_i^T\vec{\boldsymbol{\nu}}_{\ell_2}\right| \geq 1\right\}\right) \leq n\left(2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right)\right). \tag{159}$$

We can consequently choose a constant C so that if the number of measurements satisfies

$$m \geq Cs\log n \tag{160}$$

we have exact recovery with probability $1 - \frac{1}{n}$. □

4 Proofs

4.1 Proof of Lemma 3.5

By symmetry of the Gaussian probability density function, we just need to bound the probability that $u > t$. Applying Markov's inequality (Theorem 2.6 in Lecture Notes 3) we have

$$\begin{aligned}
\mathrm{P}\left(\mathbf{u} \geq t\right) &= \mathrm{P}\left(\exp\left(\mathbf{u}t\right) \geq \exp\left(t^2\right)\right) && (161)\\
&\leq \mathrm{E}\left(\exp\left(\mathbf{u}t - t^2\right)\right) && (162)\\
&= \exp\left(-\frac{t^2}{2}\right)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-t)^2}{2}\right)\mathrm{d}x && (163)\\
&= \exp\left(-\frac{t^2}{2}\right). && (164)
\end{aligned}$$

References

The proofs of exact recovery via ℓ_1-norm minimization and of the restricted isometry property of Gaussian matrices are based on arguments in [3] and [1]. For further reading on mathematical tools used to analyze compressed sensing we refer to [11] and [9].

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] E. Candès and B. Recht. Simple bounds for recovering low-complexity models. *Mathematical Programming*, 141(1-2):577–589, 2013.

[4] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[5] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[6] E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

[7] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[8] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions in Information Theory*, 52:5406–5425, 2006.

[9] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 1. 2013.

[10] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.

[11] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.