# Duality

## 1 Motivation

### 1.1 Compressed sensing

The goal of compressed sensing is to recover signals from a small number of linear measurements. An idealized version of the problem is estimating a sparse signal from underdetermined linear measurements, i.e. finding a sparse signal $\vec{x}_{\text{true}} \in \mathbb{R}^d$ such that

$$A\vec{x}_{\text{true}} = \vec{y} \tag{1}$$

where $\vec{y} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times d}$, and $m < d$. In the notes on randomization we established that if $A$ is randomized, then the sparse-recovery problem is often well posed, in the sense that there is a single sparse vector consistent with the data. However, we did not discuss any algorithms to perform recovery. As discussed in the notes on convex optimization, minimizing the number of nonzero entries subject to equality constraints is not computationally tractable, but minimizing the $\ell_1$ norm is often an effective surrogate. Figure 1 shows the result of solving

$$\min_{\vec{x}} ||\vec{x}||_1 \quad \text{subject to } A\vec{x} = \vec{y} \tag{2}$$

where $A$ contains random rows from the DFT matrix, a randomized operator inspired by magnetic-resonance imaging. The solution is perfect! In contrast, minimizing the $\ell_2$ norm produces a dense estimate that is very different from the original signal. In order to analyze this phenomenon we will study constrained optimization problems, where a cost function is minimized over a fixed set.

### 1.2 An algorithm for sparse recovery

Consider the problem of recovering a sparse vector $\vec{x}_{\text{true}} \in \mathbb{R}^d$ if we have access to inner products with any vector of our choice. A way to perform recovery is to solve the problem

$$\max_{\vec{u} \in \mathbb{R}^n} \langle \vec{u}, \vec{x}_{\text{true}} \rangle \quad \text{subject to } ||\vec{u}||_\infty \leq 1. \tag{3}$$

The inner product is maximized by setting each entry $\vec{u}[i]$, $1 \leq i \leq d$, to the sign of $\vec{x}_{\text{true}}[i]$, unless $\vec{x}_{\text{true}}[i]$ is zero. If $\vec{x}_{\text{true}}[i]$ is zero, then $\vec{u}[i]$ can equal any value, without affecting the cost function. As a result, for most solutions, the locations at which $\vec{u}$ equals $-1$ or $1$ reveal the nonzero support of $\vec{x}_{\text{true}}$.

As mentioned in the previous section, in compressed sensing we have access to measurements of the form $\vec{y} = A\vec{x}_{\text{true}}$. Interestingly, this allows us to compute inner products with vectors belonging

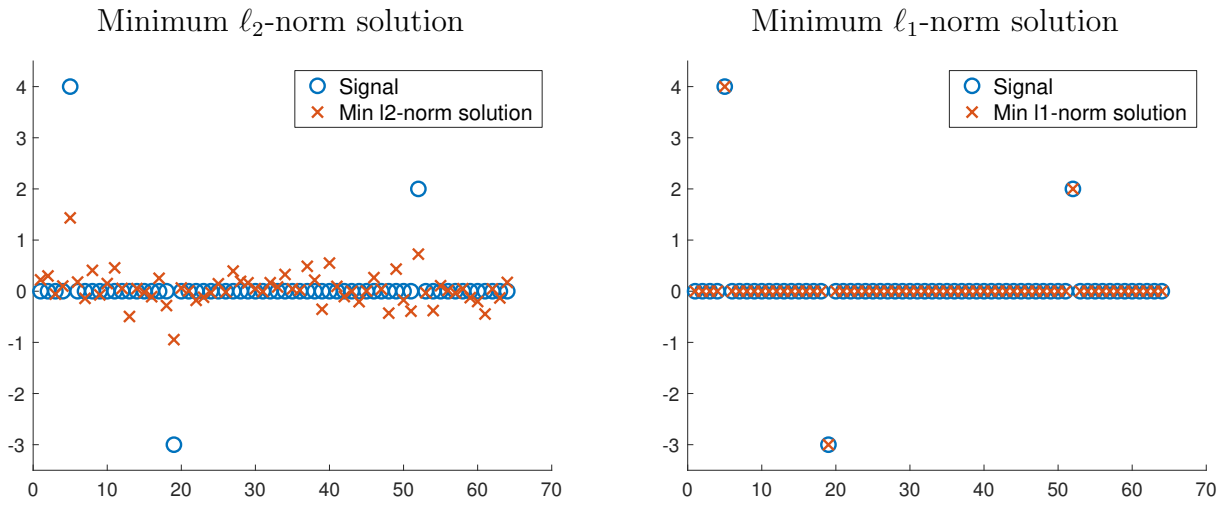


**Figure 1:** Minimum $\ell_2$-norm and $\ell_1$-norm solutions for a problem where the data are underdetermined random frequency measurements from a sparse signal.

to a fixed subspace: the row space of $A$. For any $\vec{v} \in \mathbb{R}^m$

$$\langle \vec{v}, \vec{y} \rangle = \langle \vec{v}, A\vec{x}_{\text{true}} \rangle \tag{4}$$

$$= \langle A^T\vec{v}, \vec{x}_{\text{true}} \rangle. \tag{5}$$

This suggests the following algorithm for estimating the support of $\vec{x}_{\text{true}}$. Solve the problem

$$\max_{\vec{v} \in \mathbb{R}^m} \langle \vec{v}, \vec{y} \rangle \quad \text{subject to } \left|\left|A^T\vec{v}\right|\right|_\infty \leq 1, \tag{6}$$

and then check where $A^T\vec{v}$ equals $-1$ or 1. Perhaps surprisingly, the approach is equivalent to the $\ell_1$-norm minimization in Eq. (2). To understand why we will study Lagrangian duality.

## 1.3 Matrix completion

Completing a low-rank matrix from a subset of its entries is an important problem in collaborative filtering (see the notes on the SVD). As discussed in the notes on convex optimization, minimizing the rank is not a viable strategy to solve this problem. A tractable alternative is to minimize the nuclear norm instead. Let $\Omega$ be a subset of $m$ entries of a $n_1 \times n_2$ matrix, and let $\vec{y} \in \mathbb{R}^m$. The idea is to solve the constrained optimization problem

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y}. \tag{7}$$

In practice, one would take into account noise by either using an inequality constraint or a regularized least-squares cost function as in the notes on convex optimization. However, analyzing this version of the problem is useful to analyze theoretically when the problem is well posed and nuclear-norm minimization is able to achieve recovery.


**Figure 2:** An example of a nonconvex set (left) and a convex set (right).

# 2 Lagrangian duality

## 2.1 Convex sets

Consider the constrained optimization problem

$$\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) \qquad \text{subject to } \vec{x} \in \mathcal{S}, \tag{8}$$

where $\mathcal{S}$ is a subset of $\mathbb{R}^n$ and $f$ is a convex function. If $\vec{x} \in \mathcal{S}$ we say that $\vec{x}$ is a *feasible* point for the optimization problem. Recall that the convexity of $f$ is manifested in the fact that for any two points $\vec{x}$ and $\vec{y}$, and any $\theta \in (0, 1)$,

$$f(\theta\vec{x} + (1-\theta)\,\vec{y}) \leq \theta f(\vec{x}) + (1-\theta)\, f(\vec{y}). \tag{9}$$

In order to preserve this property in the constrained optimization problem, every segment connecting feasible points should belong to $\mathcal{S}$. Sets satisfying this property are called convex.

**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 $\theta \in (0, 1)$*

$$\theta\vec{x} + (1-\theta)\,\vec{y} \in \mathcal{S}. \tag{10}$$

Figure 2 shows a simple example of a convex and a nonconvex set. The following theorem establishes an important fact: disjoint convex sets can always be separated by a hyperplane.

**Theorem 2.2.** *There exists a hyperplane separating any pair of nonempty disjoint convex sets $\mathcal{S}_1$, $\mathcal{S}_2 \subset \mathbb{R}^n$. More precisely, there exists $\vec{a} \neq \vec{0} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that for all $\vec{x}_1 \in \mathcal{S}_1$ $\langle \vec{a}, \vec{x}_1 \rangle \leq b$ and for all $\vec{x}_2 \in \mathcal{S}_2$ $\langle \vec{a}, \vec{x}_2 \rangle \geq b$. Equivalently, the function*

$$h(\vec{x}) := \langle \vec{a}, \vec{x} \rangle + b \tag{11}$$

*is nonpositive on $\mathcal{S}_1$ and nonnegative on $\mathcal{S}_2$.*

*Proof.* Following Section 2.5.1 in [1], we prove the result under the assumption that there exist two points $\vec{y}_1 \in \mathcal{S}_1$ and $\vec{y}_2 \in \mathcal{S}_2$ which achieve the minimum distance between the sets:

$$||\vec{y}_2 - \vec{y}_1||_2 = \min_{\vec{x}_1\in\mathcal{S}_1,\, \vec{x}_2\in\mathcal{S}_2} ||\vec{x}_2 - \vec{x}_1||_2 \,. \tag{12}$$

See Exercise 2.22 in [1] for the extension to the general case.

We consider the hyperplane orthogonal to $\vec{y}_2 - \vec{y}_1$ that lies exactly between $\vec{y}_1$ and $\vec{y}_2$. The hyperplane contains the points where the linear function

$$h(\vec{x}) := \left\langle \vec{y}_2 - \vec{y}_1, \vec{x} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle \tag{13}$$

equals zero. We now show that for all $\vec{x}_2 \in \mathcal{S}_2$ $h(\vec{x}_2) \geq 0$, the same argument can be used to prove that for all $\vec{x}_1 \in \mathcal{S}_1$ $h(\vec{x}_1) \leq 0$.

Let us assume that there exists a point $\vec{u} \in \mathcal{S}_2$ such that $h(\vec{u}) < 0$. This implies that the inner product

$$\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle < 0 \tag{14}$$

because

$$\begin{aligned} h(\vec{u}) &= \left\langle \vec{y}_2 - \vec{y}_1, \vec{u} - \frac{\vec{y}_1 + \vec{y}_2}{2} \right\rangle && (15)\\ &= \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle + \left\langle \vec{y}_2 - \vec{y}_1, \frac{\vec{y}_2 - \vec{y}_1}{2} \right\rangle && (16)\\ &= \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle + \frac{1}{2} ||\vec{y}_2 - \vec{y}_1||_2^2 . && (17) \end{aligned}$$

Now consider the point $\vec{y}_\theta := \theta\vec{u} + (1-\theta)\vec{y}_2 \in \mathcal{S}_2$. Its squared distance to $\vec{y}_1$ is given by

$$\begin{aligned} ||\vec{y}_\theta - \vec{y}_1||_2^2 &= ||\theta(\vec{u} - \vec{y}_2) + \vec{y}_2 - \vec{y}_1||_2^2 && (18)\\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + \theta^2 ||\vec{u} - \vec{y}_2||_2^2 + 2\theta \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle && (19)\\ &= ||\vec{y}_2 - \vec{y}_1||_2^2 + g(\theta). && (20) \end{aligned}$$

We have $g(0) = 0$ and $g'(0) = \langle \vec{y}_2 - \vec{y}_1, \vec{u} - \vec{y}_2 \rangle < 0$ so for small enough $\theta$ $\vec{y}_\theta$ is closer to $\vec{y}_1$ than $\vec{y}_2$. This is a contradiction because $\vec{y}_2$ is the point in $\mathcal{S}_2$ that is closest to $\vec{y}_1$ by assumption.

□

A hyperplane is a convex set.

**Lemma 2.3.** *The hyperplane* $\mathcal{H} := \left\{ \vec{x} \mid A\vec{x} = \vec{b} \right\}$ *– where* $\vec{x} \in \mathbb{R}^n$, $\vec{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ *– is a convex set.*

*Proof.* For any $\vec{x}, \vec{y} \in \mathcal{H}$ and any $\theta \in (0, 1)$

$$\begin{aligned} A\left(\theta\vec{x} + (1-\theta)\,\vec{y}\right) &= \theta A\vec{x} + (1-\theta)\,A\vec{y} && (21)\\ &= \vec{b} && (22) \end{aligned}$$

so $\theta\vec{x} + (1-\theta)\,\vec{y} \in \mathcal{H}$. □

Convex sets are often described as the sublevel sets of a convex function.

**Definition 2.4** (Sublevel set)**.** *The $\gamma$-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 $\gamma$,*

$$\mathcal{S}_\gamma := \{\vec{x} \mid f(\vec{x}) \leq \gamma\}. \tag{23}$$

**Lemma 2.5** (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 $\gamma$-sublevel set of a convex function $f$ then for any $\theta \in (0, 1)$

$$f(\theta\vec{x} + (1-\theta)\vec{y}) \leq \theta f(\vec{x}) + (1-\theta) f(\vec{y}) \quad \text{by convexity of } f \tag{24}$$

$$\leq \gamma \tag{25}$$

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

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

**Lemma 2.6** (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 $\cap_{i=1}^m \mathcal{S}_i$. □

Using the previous lemmas, any optimization problem of the form,

$$\min_{\vec{x} \in \mathbb{R}^n} f_0(\vec{x}) \qquad \text{subject to } f_i(\vec{x}) \leq 0, \quad 1 \leq i \leq k, \tag{26}$$

$$A\vec{x} = \vec{b}, \tag{27}$$

where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, has a convex feasibility set as long as the functions $f_1, \ldots, f_k$ are all convex.

## 2.2 The Lagrangian function

Lagrangian duality is an important tool for the analysis of constrained convex optimization problems. The basic idea is to augment the cost function with additive terms that encode the constraints. To simplify the exposition, we will consider an optimization problem with equality constraints,

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \qquad \text{subject to } A\vec{x} = \vec{b}, \tag{28}$$

where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$. Essentially the same results apply for problems with additional inequality constraints.

**Definition 2.7.** *The Lagrangian of the optimization problem in Eq. (28) is*

$$L(\vec{x}, \vec{\alpha}) := f(\vec{x}) + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right), \tag{29}$$

*where the vector $\vec{\alpha} \in \mathbb{R}^m$ is called a Lagrange multiplier.*

By definition, at any feasible point $\vec{x}$ the Lagrangian is equal to the cost function

$$L\left(\vec{x}, \vec{\alpha}\right) = f\left(\vec{x}\right). \tag{30}$$

**Example 2.8** (Constrained $\ell_1$-norm minimization in 2D)**.** Consider the optimization problem:

$$\min_{\vec{x} \in \mathbb{R}^2} ||\vec{x}||_1 \quad \text{subject to } 2\vec{x}[1] - 3\vec{x}[2] = 4. \tag{31}$$

Figure 3 shows heatmaps of the cost function in Example 2.8 and the corresponding Lagrangian function

$$\mathcal{L}\left(\vec{x}, \alpha\right) = ||\vec{x}||_1 + \alpha(4 - 2\vec{x}[1] + 3\vec{x}[2]) \tag{32}$$

for different values of the Lagrange multiplier $\alpha$. △

Minimizing the Lagrangian over $\vec{x}$ yields a function that only depends on $\vec{\alpha}$. We call the function the Lagrange dual function of the optimization problem. This motivates calling the Lagrange multiplier a dual variable. In contrast, we refer to $\vec{x}$ as the primal variable.

**Definition 2.9** (Lagrange dual function)**.** *The Lagrange dual function is the infimum of the Lagrangian over the primal variable* $\vec{x}$

$$g\left(\vec{\alpha}\right) := \inf_{\vec{x} \in \mathbb{R}^n} L\left(\vec{x}, \vec{\alpha}\right). \tag{33}$$

The Lagrange dual function provides a lower bound on the solution to the optimization problem for any value of the dual variable.

**Theorem 2.10** (Lagrange dual function as a lower bound of the primal optimum)**.** *Let* $p^*$ *denote a minimum of the optimization problem in Eq.* (28)*,*

$$g\left(\vec{\alpha}\right) \leq p^*. \tag{34}$$

*Proof.* The result follows directly from (30). Let $\vec{x}^*$ be a feasible point that attains the minimum,

$$\begin{aligned} p^* &= f\left(\vec{x}^*\right) && (35)\\ &= L\left(\vec{x}^*, \vec{\alpha}\right) && (36)\\ &\geq g\left(\vec{\alpha}\right). && (37) \end{aligned}$$

□

## 2.3 The dual problem

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.

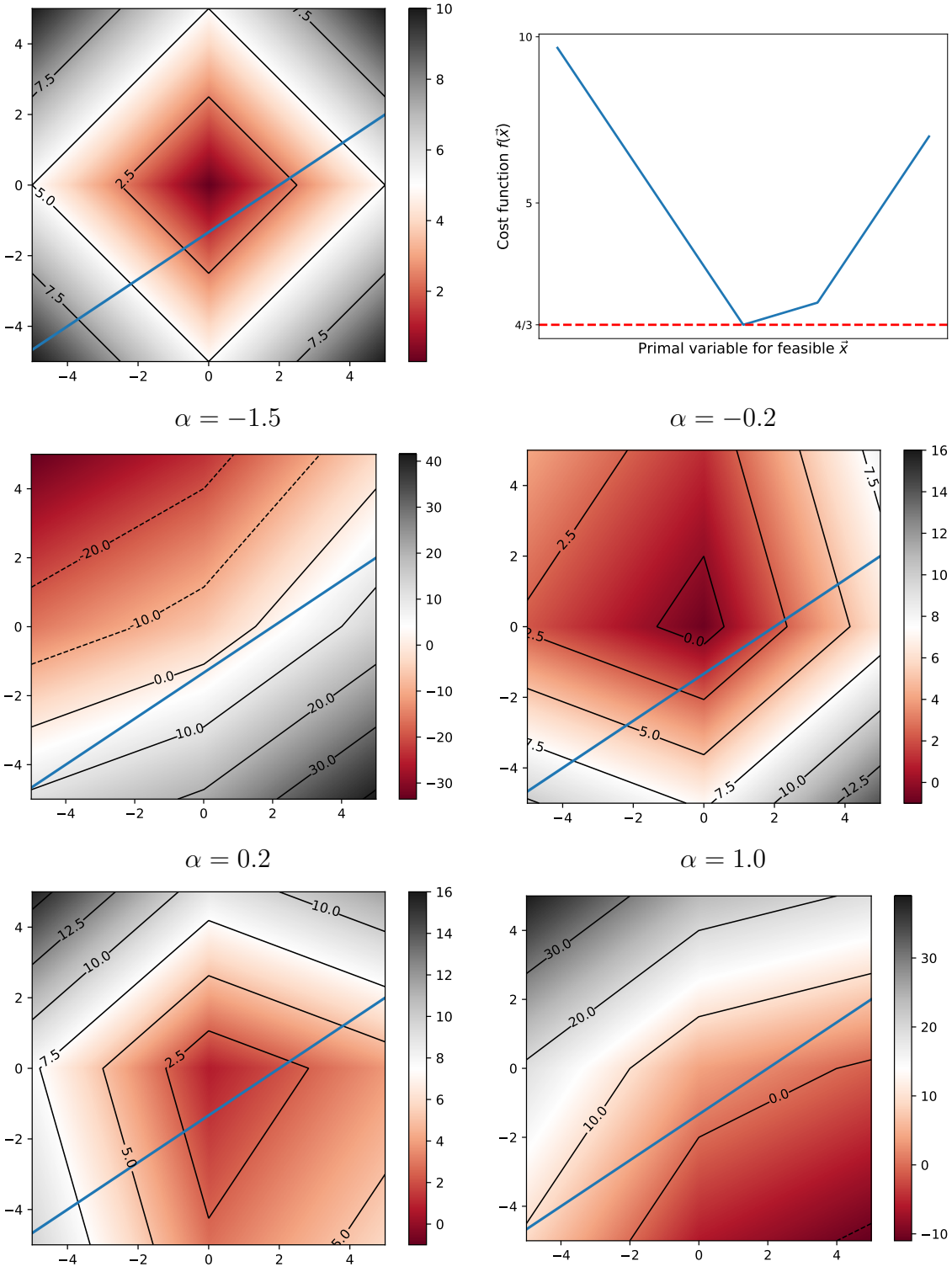


**Figure 3:** The left image in the first row shows the heatmap of the cost function in Example 2.8. The feasible set is indicated with a blue line. The right image shows the cost function restricted to the feasible set. The second and third rows show heatmaps of the Lagrangian function for different values of the Lagrange multiplier $\alpha$.

**Definition 2.11** (Dual problem)**.** *The dual problem of the optimization problem in Eq. (28) is*

$$\max_{\alpha \in \mathbb{R}^m} g\left(\alpha\right). \tag{38}$$

The function $-g\left(\alpha\right) := \sup_{\vec{x} \in \mathbb{R}^n} L\left(\vec{x}, \vec{\alpha}\right)$ is a pointwise supremum of linear functions. The following lemma establishes that it is therefore convex.

**Lemma 2.12** (Supremum of convex functions)**.** *Pointwise supremum of a family of convex functions indexed by a set $\mathcal{I}$*

$$f_{\sup}\left(\vec{x}\right) := \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right). \tag{39}$$

*is convex.*

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

$$\begin{aligned}
f_{\sup}\left(\theta\vec{x} + \left(1-\theta\right)\vec{y}\right) &= \sup_{i \in \mathcal{I}} f_i\left(\theta\vec{x} + \left(1-\theta\right)\vec{y}\right) && (40)\\
&\leq \sup_{i \in \mathcal{I}} \theta f_i\left(\vec{x}\right) + \left(1-\theta\right) f_i\left(\vec{y}\right) \quad \text{by convexity of the } f_i && (41)\\
&\leq \theta \sup_{i \in \mathcal{I}} f_i\left(\vec{x}\right) + \left(1-\theta\right) \sup_{j \in \mathcal{I}} f_j\left(\vec{y}\right) && (42)\\
&= \theta f_{\sup}\left(\vec{x}\right) + \left(1-\theta\right) f_{\sup}\left(\vec{y}\right) && (43)
\end{aligned}$$

□

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 2.10, states that the optimum of the dual problem is a lower bound for the primal optimum. This is known as weak duality. Note that it does not require convexity of the cost function.

**Corollary 2.13** (Weak duality)**.** *Let $p^*$ denote a minimum of the optimization problem in Eq. 28 and $d^*$ a maximum of the corresponding dual problem, then*

$$d^* \leq p^*. \tag{44}$$

For many convex problems a much stronger statement holds: the values obtained by minimizing the primal and maximizing the dual are the same! We defer the proof of the theorem to Section 2.5.

**Theorem 2.14** (Strong duality)**.** *Let $p^*$ denote a minimum of the optimization problem in Eq. 28 and $d^*$ a maximum of the corresponding dual problem, then if $f$ is convex and $A$ is full rank,*

$$d^* = p^*. \tag{45}$$

## 2.4 Norm minimization with equality constraints

The following theorem derives the dual problem of norm-minimization problems subject to equality constraints. The dual norm defined in the theorem can easily be shown to satisfy the properties of a norm. The dual norm of the $\ell_1$ norm is the $\ell_\infty$ norm, and the dual norm of the operator norm is the nuclear norm.

**Theorem 2.15.** *Let the dual of a norm $||\cdot||$ be defined by*

$$||\vec{y}||_d := \max_{||\vec{x}|| \leq 1} \langle \vec{y}, \vec{x} \rangle . \tag{46}$$

*Then the Lagrange dual function of the optimization problem*

$$\min_{\vec{x} \in \mathbb{R}^n} ||\vec{x}|| \qquad \text{subject to } A\vec{x} = \vec{b}, \tag{47}$$

*where $A \in \mathbb{R}^{m \times n}$, $\vec{b} \in \mathbb{R}^m$, equals*

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{b} \rangle \qquad \text{subject to } \left|\left|A^T \vec{\alpha}\right|\right|_d \leq 1. \tag{48}$$

*Proof.* The Lagrangian equals

$$\begin{aligned} L(\vec{x}, \vec{\alpha}) &:= ||\vec{x}|| + \vec{\alpha}^T \left(\vec{b} - A\vec{x}\right) && (49)\\ &= ||\vec{x}|| - \langle A^T \vec{\alpha}, \vec{x} \rangle + \vec{\alpha}^T \vec{b} && (50)\\ &= \left(1 - \left\langle A^T \vec{\alpha}, \frac{\vec{x}}{||\vec{x}||} \right\rangle\right) ||\vec{x}|| + \vec{\alpha}^T \vec{b}. && (51) \end{aligned}$$

Let us define $\vec{u}$ as

$$\vec{u} := \arg\max_{||\vec{x}|| \leq 1} \left\langle A^T \vec{\alpha}, \vec{x} \right\rangle \tag{52}$$

so that $\left\langle A^T \vec{\alpha}, \vec{u} \right\rangle = \left|\left|A^T \vec{\alpha}\right|\right|_d$. By definition of the dual norm and Eq. (51), for any $a := ||\vec{x}|| \neq 0$

$$\begin{aligned} L(\vec{x}, \vec{\alpha}) &\geq a \left(1 - \left|\left|A^T \vec{\alpha}\right|\right|_d\right) + \vec{\alpha}^T \vec{b} && (53)\\ &= a \left(1 - \langle A^T \vec{\alpha}, \vec{u} \rangle\right) + \vec{\alpha}^T \vec{b} && (54)\\ &= L(a\vec{u}, \vec{\alpha}). && (55) \end{aligned}$$

If $\left|\left|A^T \vec{\alpha}\right|\right|_d > 1$ the value can be made arbitrarily small by letting $a \to \infty$. If $\left|\left|A^T \vec{\alpha}\right|\right|_d \leq 1$ then the minimum is achieved by setting $a = 0$. The dual function therefore equals

$$g(\vec{\alpha}) = \begin{cases} \vec{\alpha}^T \vec{b} & \text{if } \left|\left|A^T \vec{\alpha}\right|\right|_d \leq 1, \\ -\infty & \text{otherwise.} \end{cases} \tag{56}$$

□

The following corollary shows that the problem described in Section 1.2 is the dual of the $\ell_1$-norm minimization problem with equality constraints.

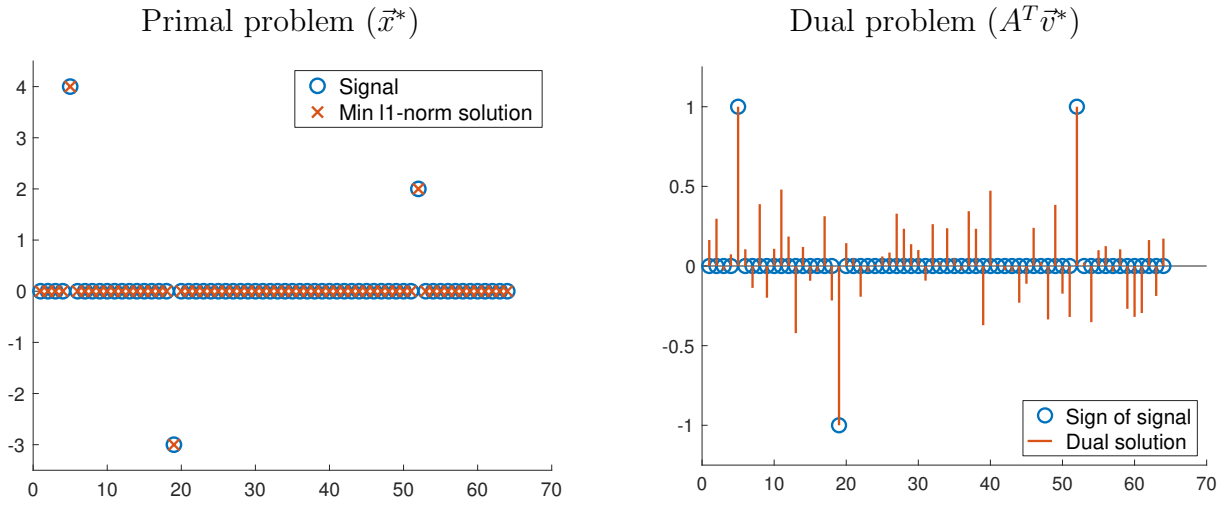


**Figure 4:** The left image shows the minimum $\ell_1$-norm solution for a problem where the data are underdetermined random frequency measurements from a sparse signal. The image on the right shows $A^T\vec{v}^*$ where $A$ is the measurement operator and $\vec{v}^*$ is a solution to the dual. As established in Lemma 2.18 the dual solution reveals the support of the primal solution.

**Corollary 2.16.** *Let $A \in \mathbb{R}^{m\times n}$, $\vec{b} \in \mathbb{R}^m$. The dual of the optimization problem*

$$\min_{\vec{x}\in\mathbb{R}^n} ||\vec{x}||_1 \qquad \text{subject to} \qquad A\vec{x} = \vec{b} \tag{57}$$

*is*

$$\max_{\vec{\alpha}\in\mathbb{R}^m} \langle \vec{\alpha}, \vec{b} \rangle \qquad \text{subject to } \left|\left|A^T\vec{\alpha}\right|\right|_\infty \leq 1. \tag{58}$$

**Example 2.17** (Constrained $\ell_1$-norm minimization in 2D (continued))**.** By Corollary 4.2 the dual function of the optimization problem in Example 2.8 equals

$$g(\alpha) = \begin{cases} 4\alpha & \text{if } |\alpha| \leq \frac{1}{3}, \\ -\infty & \text{otherwise,} \end{cases} \tag{59}$$

because in that case $A^T\alpha = (2\alpha, -3\alpha)$. Figure 5 shows the dual function alongside the cost function restricted to the feasibility set. The maximum of the dual function equals the minimum of the primal as dictated by strong duality. △

Strong duality has an interesting consequence for the $\ell_1$-norm minimization problem with equality constraints: dual solutions can be used to reveal the support of the primal solution (as described intuitively in Section 1.2). This is illustrated in Figure 4.

**Lemma 2.18.** *If there exists a feasible vector for the primal problem, then the solution $\vec{\alpha}^*$ to Problem (154) satisfies*

$$(A^T\vec{\alpha}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0 \tag{60}$$

*for any solution $\vec{x}^*$ to the primal problem.*

*Proof.* By strong duality

$$||\vec{x}^*||_1 = \vec{y}^T\vec{\alpha}^* \tag{61}$$
$$= (A\vec{x}^*)^T\vec{\alpha}^* \tag{62}$$
$$= (\vec{x}^*)^T(A^T\vec{\alpha}^*) \tag{63}$$
$$= \sum_{i=1}^{m}(A^T\vec{\alpha}^*)[i]\vec{x}^*[i]. \tag{64}$$

By Hölder's inequality

$$||\vec{x}^*||_1 \geq \sum_{i=1}^{m}(A^T\vec{\alpha}^*)[i]\vec{x}^*[i] \tag{65}$$

with equality if and only if

$$(A^T\vec{\alpha}^*)[i] = \text{sign}(\vec{x}^*[i]) \quad \text{for all } \vec{x}^*[i] \neq 0. \tag{66}$$

□

## 2.5 Proof of Theorem 2.14

The proof follows Section 5.3.2 of [1]. To simplify the argument we assume $A$ has full row rank. We begin by defining the set

$$\mathcal{A} := \left\{(\vec{v}, t) \mid \vec{b} - A\vec{x} = \vec{v} \quad \text{and} \quad f(\vec{x}) \leq t \quad \text{for some } \vec{x} \in \mathbb{R}^n\right\}. \tag{67}$$

Notice that the solution of the optimization problem over the feasible set is

$$p^* = \inf\left\{t \mid (\vec{0}, t) \in \mathcal{A}\right\}. \tag{68}$$

If $(\vec{v}, t) \in \mathcal{A}$ then there exists an $\vec{x}$ such that

$$\langle\vec{\alpha}, \vec{v}\rangle + t \geq \langle\vec{\alpha}, \vec{b} - A\vec{x}\rangle + f(\vec{x}) \tag{69}$$
$$= \mathcal{L}(\vec{x}, \vec{\alpha}) \tag{70}$$

for any $\vec{\alpha}$, with equality if we set $t := f(\vec{x})$. This implies,

$$g(\vec{\alpha}) := \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\alpha}) \tag{71}$$
$$= \inf\left\{\langle\vec{\alpha}, \vec{v}\rangle + t \mid (\vec{v}, t) \in \mathcal{A}\right\}. \tag{72}$$

Geometrically, the hyperplane

$$\langle\vec{\alpha}, \vec{v}\rangle + t = g(\vec{\alpha}) \tag{73}$$

is a supporting hyperplane to $\mathcal{A}$. See Figure 5 for an illustration. This implies that weak duality holds, since

$$p^* = \langle\vec{\alpha}, \vec{0}\rangle + p^* \tag{74}$$
$$\geq g(\vec{\alpha}), \tag{75}$$

for any $\vec{\alpha}$ because $(\vec{0}, p^*) \in \mathcal{A}$.

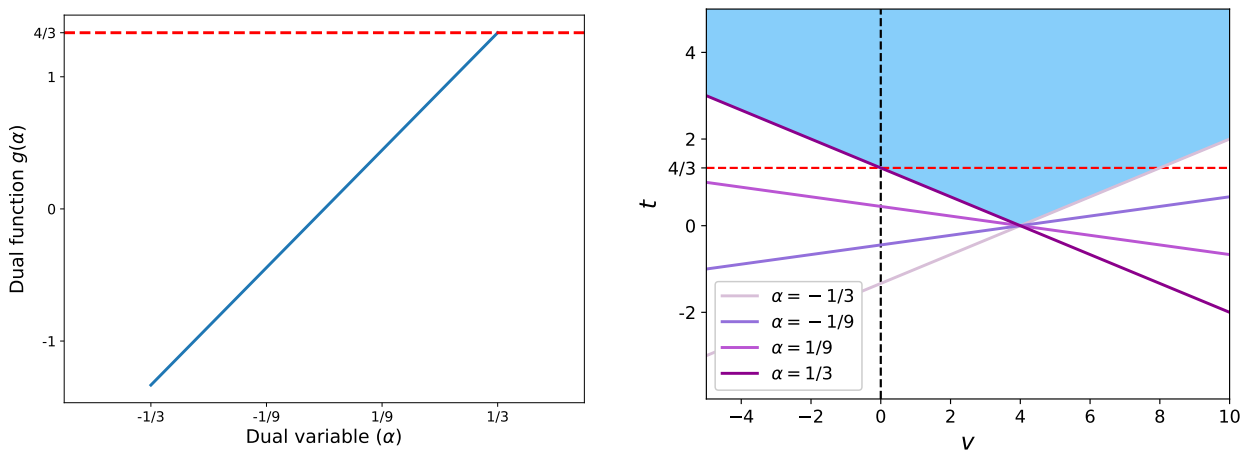


**Figure 5:** The left image shows the dual function derived in Example 2.17. The right image shows the set $\mathcal{A}$ defined in Eq. (67). The supporting hyperplanes (lines) correspond to $\langle \vec{\alpha}, \vec{v} \rangle + t = g(\vec{\alpha})$ for different values of $\alpha$. Note that their intersection with the line $v = 0$ equals $g(\alpha)$.

**Example 2.19** (Constrained $\ell_1$-norm minimization in 2D (continued))**.** For the problem in Example 2.8, if we fix $v := 4 - 2\vec{x}[1] + 3\vec{x}[2]$ then

$$||\vec{x}||_1 = |\vec{x}[1]| + \left| \frac{v - 4 + 2\vec{x}[1]}{3} \right|. \tag{76}$$

This is a piecewise linear function with two kinks at $\vec{x}[1] = 0$ and $\vec{x}[1] = (4 - v)/2$. Since for $\vec{x}[1] \to \pm\infty$ we have $||\vec{x}||_1 \to \infty$ the minimum is given by

$$\min_{v=4-2\vec{x}[1]+3\vec{x}[2]} ||\vec{x}||_1 = \min \left\{ \left| \frac{v-4}{3} \right|, \left| \frac{v-4}{2} \right| \right\}, \tag{77}$$

so that

$$\mathcal{A} := \left\{ (v, t) \mid t \geq \min \left\{ \left| \frac{v-4}{3} \right|, \left| \frac{v-4}{2} \right| \right\} \right\}. \tag{78}$$

The set is depicted in Figure 5. △

A crucial observation is that $\mathcal{A}$ is convex as long as the cost function of the primal problem is convex.

**Lemma 2.20.** *The set $\mathcal{A}$ defined in Eq. (67) is convex if $f$ is convex.*

*Proof.* Let $(\vec{v}_1, t_1)$ and $(\vec{v}_2, t_2)$ belong to $\mathcal{A}$. For any $\theta \in (0, 1)$, the question is whether $\theta(\vec{v}_1, t_1) + (1 - \theta)(\vec{v}_2, t_2)$ belongs to $\mathcal{A}$. Since $(\vec{v}_1, t_1)$ and $(\vec{v}_2, t_2)$ are in $\mathcal{A}$, there exists $\vec{x}_1$ and $\vec{x}_2$ such that

$$\vec{v}_1 = \vec{b} - A\vec{x}_1, \tag{79}$$
$$f(\vec{x}_1) \leq t_1, \tag{80}$$
$$\vec{v}_2 = \vec{b} - A\vec{x}_2, \tag{81}$$
$$f(\vec{x}_2) \leq t_2. \tag{82}$$

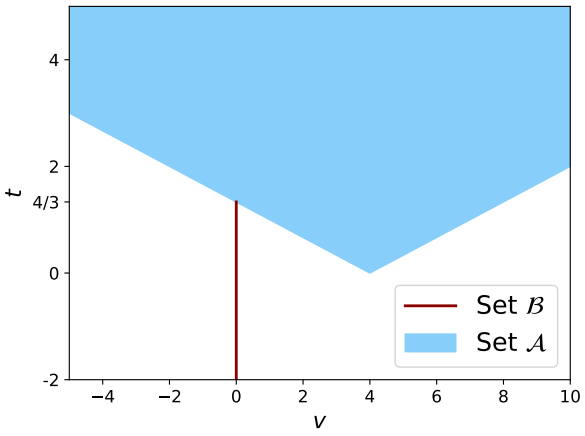


**Figure 6:** The sets $\mathcal{A}$ and $\mathcal{B}$ defined in Eqs. (67) and (67) respectively for the optimization problem in Example 2.8.

This implies

$$\theta\vec{v}_1 + (1-\theta)\vec{v}_2 = \vec{b} - A(\vec{v}_1\vec{x}_1 + (1-\theta)\vec{x}_2) \tag{83}$$

and by convexity of $f$

$$\begin{aligned} f(\theta\vec{x}_1 + (1-\theta)\vec{x}_2) &\leq \theta f(\vec{x}_1) + (1-\theta)f(\vec{x}_2) && (84)\\ &\leq \theta t_1 + (1-\theta)t_2, && (85)\end{aligned}$$

so $\theta(\vec{v}_1, t_1) + (1-\theta)(\vec{v}_2, t_2) \in \mathcal{A}$. □

We assume $p^*$ is finite. If $p^* = -\infty$ strong duality holds trivially because by weak duality $d^* = -\infty$. The set

$$\mathcal{B} := \left\{(\vec{0}, t) \mid t < p^*\right\} \tag{86}$$

is convex. It is just a line ending right below $p^*$ as depicted in Figure 6. Notice that the sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint. If $t \in \mathcal{A} \cap \mathcal{B}$ there exists $\vec{x}$ such that $f(\vec{x}) \leq t < p^*$, which contradicts the assumption that $p^*$ is the optimum.

By Theorem 2.2 there exists a hyperplane separating $\mathcal{A}$ and $\mathcal{B}$. In particular, there exist $\vec{w} \in \mathbb{R}^m$ and $z \in \mathbb{R}$ with $(\vec{w}, z) \neq \vec{0}$ such that

$$\begin{aligned} \vec{w}^T\vec{v} + zt &\geq q \quad \text{for all } (\vec{v}, t) \in \mathcal{A}, && (87)\\ \vec{w}^T\vec{v} + zt &\leq q \quad \text{for all } (\vec{v}, t) \in \mathcal{B}. && (88)\end{aligned}$$

Note that $z \geq 0$. Indeed, if $t \in \mathcal{A}$ all $t' > t$ belong to $\mathcal{A}$, so if $z < 0$ Eq. (87) cannot hold.

If $z > 0$ then Eq. (87) implies

$$\begin{aligned} \mathcal{L}(z^{-1}\vec{w}, \vec{x}) &= f(\vec{x}) + z^{-1}\vec{w}^T(\vec{b} - A\vec{x}) && (89)\\ &\geq z^{-1}q && (90)\end{aligned}$$

for all $\vec{x}$. Eq. (88) implies

$$p^* \leq z^{-1} q. \tag{91}$$

Combining the inequalities yields

$$\mathcal{L}(z^{-1}\vec{w}, \vec{x}) \geq p^*, \tag{92}$$

which implies

$$g(z^{-1}\vec{w}) := \inf_{\vec{x}} \mathcal{L}(z^{-1}\vec{w}, \vec{x}) \geq p^*. \tag{93}$$

Since by weak duality $g(\vec{\alpha}) \leq p^*$ for any $\vec{\alpha}$, we conclude that $g(z^{-1}\vec{w}) = d^* = p^*$.

If $z = 0$ then $\vec{w} \neq 0$ and $\vec{w}^T \vec{v} \geq q$ for all $(\vec{v}, t) \in \mathcal{A}$. But this contradicts the assumption that $A$ has full row rank, since for every $\vec{v}$ there is a corresponding $\vec{x}$ with $\vec{b} - A\vec{x} = \vec{v}$, and thus a corresponding point $(\vec{v}, t) \in \mathcal{A}$. But $\vec{w}^T \vec{v} \geq q$ cannot hold for all $\vec{v} \in \mathbb{R}^m$.

# 3 Compressed sensing

## 3.1 Exact recovery via $\ell_1$-norm minimization

In the lecture notes on randomization, we established that compressed sensing, i.e. recovery of a sparse vector from linear underdetermined measurements, is a well-posed problem for randomized measurement matrices. However, we did not propose a tractable algorithm to perform recovery. The following theorem establishes that $\ell_1$-norm minimization reconstructs sparse vectors exactly from random data with high probability. We study Gaussian measurements for simplicity, but similar results can be extended to random Fourier matrices [6] and other measurements [2, 4].

**Theorem 3.1** (Exact recovery via $\ell_1$-norm minimization)**.** *Let $\mathbf{A} \in \mathbb{R}^{m \times d}$ be a random matrix with iid standard Gaussian entries and $\vec{x}_{\text{true}} \in \mathbb{R}^d$ a vector with $s$ nonzero entries such that $\mathbf{A}\vec{x}_{\text{true}} = \vec{y}$. Then $\vec{x}_{\text{true}}$ is the unique solution to the $\ell_1$-norm minimization problem*

$$\min_{\vec{x} \in \mathbb{R}^d} ||\vec{x}||_1 \qquad \textit{subject to} \qquad A\vec{x} = \vec{y} \tag{94}$$

*with probability at least $1 - \frac{1}{d}$ as long as the number of measurements satisfies*

$$m \geq C s \log d, \tag{95}$$

*for a fixed constant $C$.*

The proof of the theorem combines insights from convex duality with tools from probability theory, as explained in Sections 3.2 and 3.3 below. It is worth mentioning that the result can also be proved directly using the restricted-isometry property [3].

The guarantees in Theorem 3.1 are essentially optimal in the following sense: If we know the location of the nonzero entries of the vector, then recovery can be achieved with just $s$ entries

Undersampling pattern | Direct reconstruction | Min. $\ell_1$-norm estimate (wavelet coefficients)

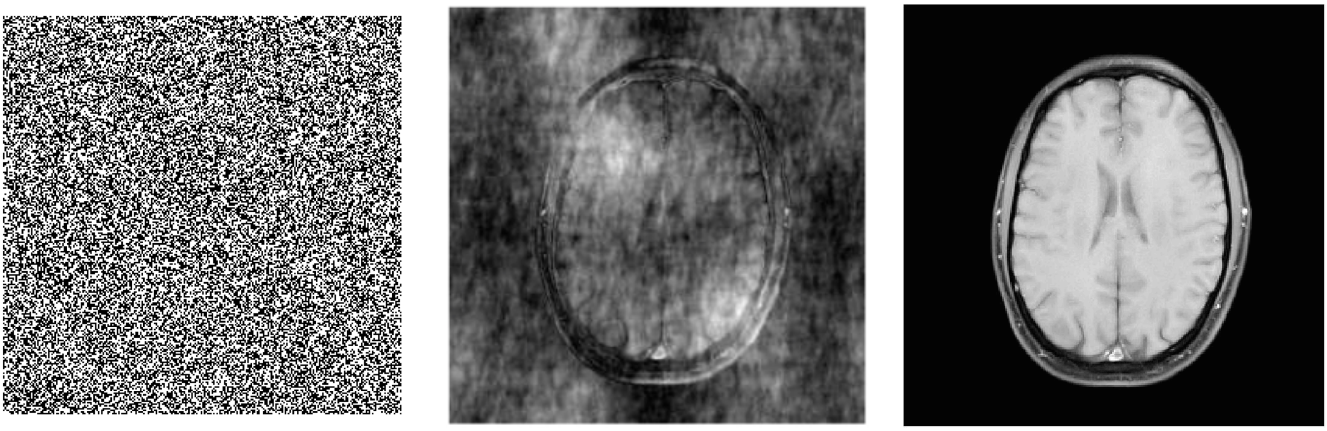

**Figure 7:** The left image shows an idealized randomized 2D Fourier sampling pattern (in practice the 2D Fourier space is samples along lines). Reconstructing directly produces artifacts due to aliasing. Minimizing the $\ell_1$ norm of the wavelet coefficients produces an almost perfect reconstruction.

(by inverting the corresponding submatrix of $A$). Surprisingly, $\ell_1$-norm minimization achieves recovery for $s \approx m$ with no dependence on $d$ up to logarithmic factors.

In practice, signals are usually not sparse, but they may have a sparse representation in a specific domain. For example, images are approximately sparse in the wavelet domain, i.e. $\vec{x}_{\text{true}} = W\vec{c}_{\text{true}}$, where $\vec{c}_{\text{true}}$ is sparse. In that case, recovery can be performed by minimizing the $\ell_1$ norm of the signal coefficients

$$\min_{\vec{c}} ||\vec{c}||_1 \qquad \text{subject to} \qquad AW\vec{c} = \vec{y}, \tag{96}$$

where $\vec{y}$ denotes the data. Figure 7 shows an example using simulated MRI data.

## 3.2 Dual certificate for $\ell_1$-norm minimization

In this section we consider the question of how to show that a fixed vector $\vec{x}_{\text{true}}$ is the solution to the $\ell_1$-norm minimization problem (94). We need to prove that no other vector $\vec{x}$ compatible with the data (i.e. such that $A\vec{x} = \vec{y}$) has smaller $\ell_1$ norm than $\vec{x}_{\text{true}}$. This can be achieved using duality. Assume there exists a feasible vector $\vec{\alpha}\,'$ for the dual problem

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to} \quad \left|\left|A^T\vec{\alpha}\right|\right|_\infty \leq 1 \tag{97}$$

such that

$$||\vec{x}_{\text{true}}||_1 = \langle \vec{\alpha}\,', \vec{y} \rangle. \tag{98}$$

Then by weak duality, for any $\vec{x}$

$$\begin{aligned} ||\vec{x}||_1 &\geq \langle \vec{\alpha}\,', \vec{y} \rangle & (99) \\ &= ||\vec{x}_{\text{true}}||_1 \,. & (100) \end{aligned}$$

This suggests the following proof strategy: Show that for any sparse vector $\vec{x}_{\text{true}}$, there exists a corresponding dual variable $\vec{\alpha}\,'$ such that $||\vec{x}_{\text{true}}||_1 = \langle \vec{\alpha}\,', \vec{y} \rangle$. By Lemma 2.18, we know that this occurs if $A^T\vec{\alpha}\,'$ is equal to the sign of $\vec{x}_{\text{true}}$ on its support,

$$(A^T\vec{\alpha}\,')[i] = \text{sign}(\vec{x}_{\text{true}}[i]) \quad \text{for all } \vec{x}_{\text{true}}[i] \neq 0. \tag{101}$$

In order to be a feasible dual vector, the magnitude of $A^T\vec{\alpha}\,'$ must be smaller than one in the remaining entries. By Theorem 3.18 in the notes on convex optimization, a vector $\vec{g} := A^T\vec{\alpha}\,'$ that satisfies these conditions is a subgradient of the $\ell_1$ norm at $\vec{x}_{\text{true}}$. This provides an alternative proof that the existence of $\vec{\alpha}\,'$ implies optimality of $\vec{x}_{\text{true}}$. For any $\vec{x}$ such that $A\vec{x} = \vec{y}$

$$\begin{aligned}
||\vec{x}||_1 &\geq ||\vec{x}_{\text{true}}||_1 + \langle \vec{g}, \vec{x} - \vec{x}_{\text{true}} \rangle && (102)\\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', A(\vec{x} - \vec{x}_{\text{true}}) \rangle && (103)\\
&= ||\vec{x}_{\text{true}}||_1 + \langle \vec{\alpha}\,', \vec{y} - \vec{y} \rangle && (104)\\
&= ||\vec{x}_{\text{true}}||_1 . && (105)
\end{aligned}$$

Geometrically, the subgradient $\vec{g}$ is orthogonal to the difference vector $\vec{x} - \vec{x}_{\text{true}}$, for any feasible vector $\vec{x}$. This means that the difference vector cannot cross the supporting hyperplane to the $\ell_1$-norm function corresponding to $\vec{g}$. The following lemma uses a slight modification of this argument to show that if we add an additional condition to $\vec{\alpha}\,'$ (it must be strictly smaller than one on the off-support) then its existence implies that $\vec{x}_{\text{true}}$ is the unique solution of the primal problem. We call such a dual variable a dual certificate for the $\ell_1$-norm minimization problem, because it certifies optimality of a fixed primal vector.

**Theorem 3.2** (Dual certificate for $\ell_1$-norm minimization)**.** *Let $\vec{x}_{\text{true}}$ be a $d$-dimensional vector with support $\mathcal{S}$ such that $A\vec{x}_{\text{true}} = \vec{y}$ and the submatrix $A_{\mathcal{S}}$ containing the columns of $A$ indexed by $\mathcal{S}$ is full rank. If there exists a vector $\vec{\alpha}_{\text{cert}} \in \mathbb{R}^m$ such that $\vec{g}_{\text{cert}} := A^T\vec{\alpha}_{\text{cert}}$ satisfies*

$$\begin{aligned}
\vec{g}_{\text{cert}}[i] &= \text{sign}(\vec{x}_{\text{true}}[i]) \qquad \text{if } \vec{x}_{\text{true}}[i] \neq 0 && (106)\\
|\vec{g}_{\text{cert}}[i]| &< 1 \qquad\qquad\quad \text{if } \vec{x}_{\text{true}}[i] = 0 && (107)
\end{aligned}$$

*then $\vec{x}_{\text{true}}$ is the unique solution to the $\ell_1$-norm minimization problem (94).*

*Proof.* For any feasible $\vec{x} \in \mathbb{R}^d$, let $\vec{h} := \vec{x} - \vec{x}_{\text{true}}$. If $A_{\mathcal{S}}$ is full rank then $\vec{h}_{\mathcal{S}^c} \neq 0$ unless $\vec{h} = 0$ because otherwise $\vec{h}_{\mathcal{S}}$ would be a nonzero vector in the null space of $A_{\mathcal{S}}$. Condition (107) implies

$$\|\vec{h}_{\mathcal{S}^c}\|_1 > \vec{g}_{\text{cert}}^T\vec{h}_{\mathcal{S}^c}, \tag{108}$$

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

$$\begin{aligned}
||\vec{x}||_1 &= \|\vec{x}_{\text{true}} + \mathcal{P}_{\mathcal{S}}(\vec{h})\|_1 + \|\vec{h}_{\mathcal{S}^c}\|_1 \qquad \text{because } \vec{x}_{\text{true}} \text{ is supported on } \mathcal{S} && (109)\\
&> ||\vec{x}_{\text{true}}||_1 + \vec{g}_{\text{cert}}^T\mathcal{P}_{\mathcal{S}}(\vec{h}) + (A^T\vec{\alpha}_{\text{cert}})^T\mathcal{P}_{\mathcal{S}^c}(\vec{h}) \qquad \text{by (108)} && (110)\\
&= ||\vec{x}_{\text{true}}||_1 + \vec{\alpha}_{\text{cert}}^T A\vec{h} && (111)\\
&= ||\vec{x}_{\text{true}}||_1 . && (112)
\end{aligned}$$

□

## 3.3 Proof of Theorem 3.1

Let us fix $\vec{x}_{\text{true}} \in \mathbb{R}^d$ and denote the $s$ indices corresponding to its nonzero support by $\mathcal{S} \subset \{1, \ldots, d\}$. By Theorem 3.2, to prove Theorem 3.1 we need to show that there exists a vector in the row space of $\mathbf{A}$ which interpolates the sign of $\vec{x}_{\text{true}}$ on $\mathcal{S}$ and has magnitude strictly smaller than one on $\mathcal{S}^c$.

An important property of randomized matrices is that their columns are not correlated (in the notes on randomization we proved a stronger statement: small groups of columns are almost orthogonal). Consider the vector $\vec{c}_i$ containing the correlations of the $i$th column $A_i \in \mathbb{R}^m$ of $\mathbf{A}$ with every other column

$$\vec{\mathbf{c}}_i := \mathbf{A}^T \mathbf{A}_i, \quad 1 \leq i \leq m. \tag{113}$$

We have $\vec{\mathbf{c}}_i[i] = ||\mathbf{A}_i||_2^2 \approx d$ (see Theorem 2.10 in the notes on randomization). In contrast, the remaining entries are small with high probability. They contain the inner product between two standard Gaussian vectors. Since the inner product only depends on the relative angle between the vectors, to get a rough estimate we can fix one of them and assume its $\ell_2$ norm equals $\sqrt{d}$. By Lemma 3.1 the inner product between a standard Gaussian vector and vector with $\ell_2$ norm equal to $\sqrt{d}$ is a standard Gaussian variable with variance $d$. We conclude that $\vec{\mathbf{c}}_i[j]$ behaves like a Gaussian random variable with standard deviation close to $\sqrt{d}$ and is consequently significantly smaller than $\vec{\mathbf{c}}_i[i]$ (this is made precise below).

The vector $\vec{\mathbf{c}}_i$ belongs to the row space of $\mathbf{A}$ and is highly concentrated on its $i$th entry. This is perfect for our purposes! We use $\vec{\mathbf{c}}_i$, $i \in \mathcal{S}$, to build $\vec{\mathbf{g}}_{\text{cert}}$ (which is a random vector because it depends on $\mathbf{A}$):

$$\vec{\mathbf{g}}_{\text{cert}} := \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i, \tag{114}$$

where the weights $\mathbf{w}_i$, $i \in \mathcal{S}$ are adjusted so that for all $j \in \mathcal{S}$

$$\text{sign}\left(\vec{x}_{\text{true}}\right)[j] = \vec{\mathbf{g}}_{\text{cert}}[j]. \tag{115}$$

Figure 8 shows examples of correlation vectors corresponding to two randomized matrices (a randomly-undersampled DFT and an iid Gaussian matrix), as well as a deterministic matrix (a regularly-undersampled DFT). In the case of the randomized measurements, one entry of the correlation is large and the rest are small. As a result, the certificate in Eq. (114) is valid, which means $\ell_1$-norm minimization achieves recovery. For the deterministic measurements, there exists an additional column that is completely correlated with each column (in fact they are equal). As a result, the construction fails (there are other points at which $\vec{\mathbf{g}}_{\text{cert}}$ equals one), which is not surprising since $\ell_1$-norm minimization does not achieve perfect recovery.

For any vector $\vec{v}$, let us denote by $\vec{v}_{\mathcal{S}}$ the subvector of its entries indexed by $\mathcal{S}$. In matrix form we have

$$\begin{aligned} \text{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}} &= \left(\sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i\right)_{\mathcal{S}} & (116) \\ &= \sum_{i \in \mathcal{S}} \mathbf{w}_i \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_i & (117) \\ &= \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}, & (118) \end{aligned}$$

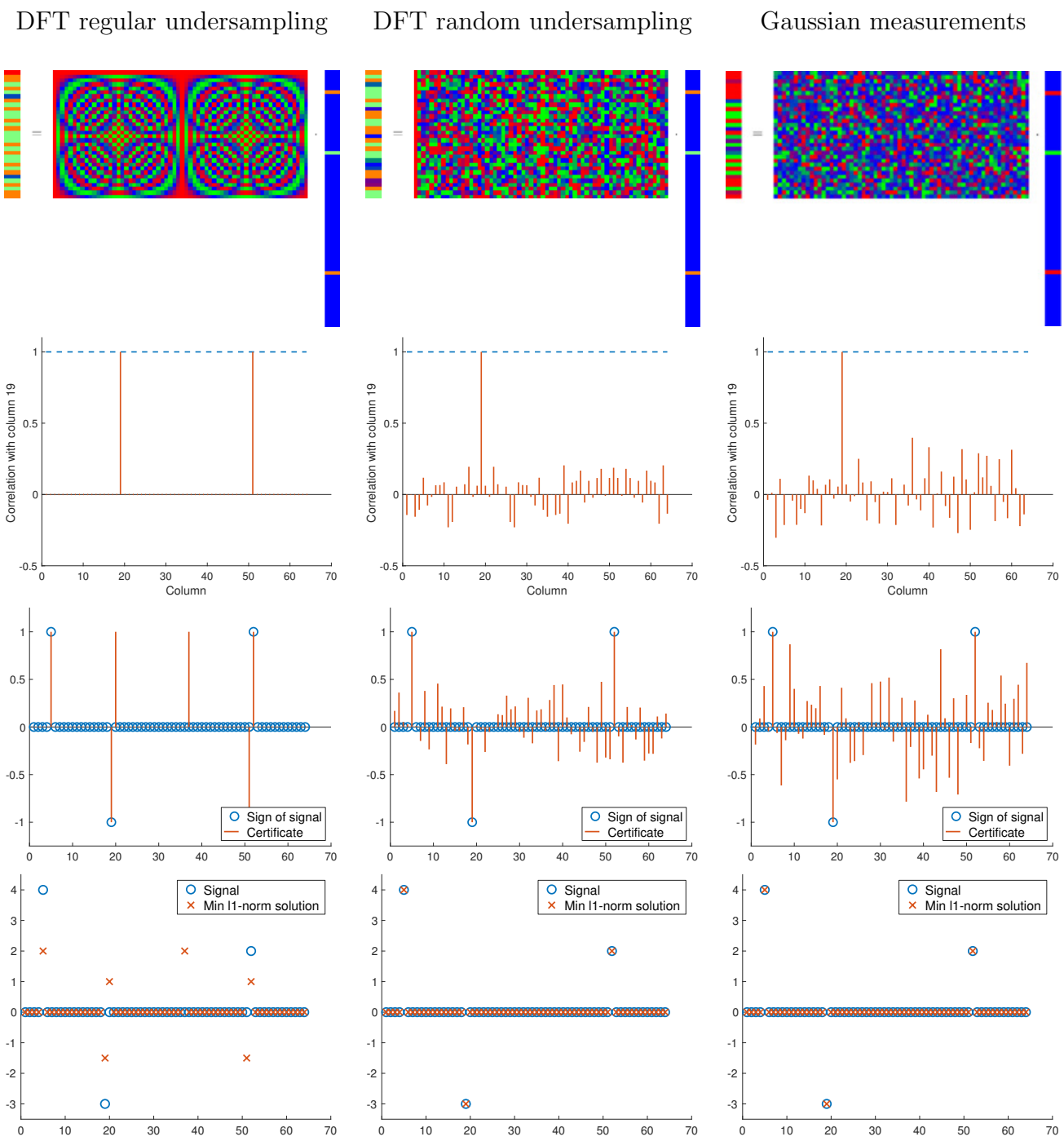


**Figure 8:** The top row illustrates different underdetermined matrices used to obtain compressed measurements of a sparse vector: a regularly-undersampled DFT matrix, a randomly-undersampled DFT matrix and a Gaussian matrix (we only show the real-part of the DFT submatrices). The second row shows correlation vectors defined as in Eq. (113) for the three matrices. The third row shows dual certificates constructed as in the proof of Theorem 3.1 for a concrete example. The fourth row shows the corresponding solution to the $\ell_1$-norm minimization problem.

where the $\vec{\mathbf{w}}$ is an $s$-dimensional vector containing the weights $\mathbf{w}_i$, $i \in \mathcal{S}$. Solving for $\vec{\mathbf{w}}$ yields

$$\vec{\mathbf{w}} := \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}. \tag{119}$$

The corresponding certificate candidate $\vec{\mathbf{g}}_{\text{cert}}$ equals

$$\begin{aligned}
\vec{\mathbf{g}}_{\text{cert}} &= \sum_{i \in \mathcal{S}} \mathbf{w}_i \vec{\mathbf{c}}_i && (120)\\
&= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} && (121)\\
&= \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}. && (122)
\end{aligned}$$

To complete the proof we need to show that verify that $\vec{\mathbf{g}}_{\text{cert}}$ satisfies the conditions of Theorem 3.2.

$A_{\mathcal{S}}$ is full rank and condition (106) holds

Let $\boldsymbol{\sigma}_s$ denote the smallest singular value of $A_{\mathcal{S}}$. Setting $\epsilon := 0.5$ in Theorem 4.4 of the notes on randomization, let $\mathcal{E}$ denote the event that

$$0.5\sqrt{m} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq 1.5\sqrt{m}. \tag{123}$$

Then

$$\mathrm{P}\left(\mathcal{E}\right) \geq 1 - \exp\left(-C' \frac{m}{s}\right) \tag{124}$$

for a fixed constant $C'$. Conditioned on $\mathcal{E}$ $\mathbf{A}_{\mathcal{S}}$ is full rank and $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$ is invertible, so $\vec{\mathbf{g}}_{\text{cert}}$ satisfies condition (106).

Condition (107) holds

Let

$$\begin{aligned}
\vec{\boldsymbol{\alpha}}_{\text{cert}} &:= \mathbf{A}_{\mathcal{S}} \vec{\mathbf{w}}_{\text{cert}} && (125)\\
&= \mathbf{A}_{\mathcal{S}} \left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)^{-1} \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}, && (126)
\end{aligned}$$

so that $\vec{\mathbf{g}}_{\text{cert}} = \mathbf{A}^T \vec{\boldsymbol{\alpha}}_{\text{cert}}$, and let $\mathbf{USV}^T$ be the SVD of $\mathbf{A}_{\mathcal{S}}$. Conditioned on $\mathcal{E}$ we have

$$\begin{aligned}
\left|\left|\vec{\boldsymbol{\alpha}}_{\text{cert}}\right|\right|_2 &= \left|\left|\mathbf{US}^{-1}\mathbf{V}^T \operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2 && (127)\\
&\leq \frac{\left|\left|\operatorname{sign}\left(\vec{x}_{\text{true}}\right)_{\mathcal{S}}\right|\right|_2}{\boldsymbol{\sigma}_{\mathbf{s}}} && (128)\\
&\leq 2\sqrt{\frac{s}{m}}. && (129)
\end{aligned}$$

For a fixed $i \in \mathcal{S}^c$ 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

$$\begin{aligned}
\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1\right) &= \mathrm{P}\left(\frac{\left|\mathbf{A}_i^T \vec{v}\right|}{\left|\left|\vec{v}\right|\right|_2} \geq \frac{1}{\left|\left|\vec{v}\right|\right|_2}\right) && (130)\\
&\leq 2\exp\left(-\frac{1}{2\left|\left|\vec{v}\right|\right|_2^2}\right) && (131)
\end{aligned}$$

by the following lemma.

**Lemma 3.3** (Proof in Section 5.1)**.** *For a Gaussian random variable* $\mathbf{u}$ *with zero mean and unit variance and any* $t > 0$

$$\mathrm{P}\left(|\mathbf{u}| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2}\right). \tag{132}$$

Note that if $i \notin \mathcal{S}$ then $\mathbf{A}_i$ and $\vec{\boldsymbol{\alpha}}_{\text{cert}}$ are independent (they depend on different and hence independent entries of $\mathbf{A}$). This means that due to equation (129)

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) = \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{v}\right| \geq 1 \quad \text{for} \quad ||\vec{v}||_2 \leq 2\sqrt{\frac{s}{m}}\right) \tag{133}$$

$$\leq 2\exp\left(-\frac{m}{8s}\right). \tag{134}$$

As a result,

$$\mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1\right) \leq \mathrm{P}\left(\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1 \,|\, \mathcal{E}\right) + \mathrm{P}\left(\mathcal{E}^c\right) \tag{135}$$

$$\leq 2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right). \tag{136}$$

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

$$\mathrm{P}\left(\bigcup_{i \in \mathcal{S}^c} \left\{\left|\mathbf{A}_i^T \vec{\boldsymbol{\alpha}}_{\text{cert}}\right| \geq 1\right\}\right) \leq n\left(2\exp\left(-\frac{m}{8s}\right) + \exp\left(-C'\frac{m}{s}\right)\right). \tag{137}$$

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

$$m \geq Cs\log n \tag{138}$$

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

# 4 Matrix completion

## 4.1 Missing data

At first glance, the problem of completing a matrix such as this one

$$\begin{bmatrix} 1 & ? & 5 \\ ? & 3 & 2 \end{bmatrix} \tag{139}$$

seems completely ill posed. We can fill in the missing entries arbitrarily! In more mathematical terms, the completion problem is equivalent to an underdetermined system of equations

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_{11} \\ M_{21} \\ M_{12} \\ M_{22} \\ M_{13} \\ M_{23} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 2 \end{bmatrix}. \tag{140}$$

In order to solve the problem, we need to make an assumption on the structure of the matrix that we aim to complete. In compressed sensing we make the assumption that the original signal is sparse. In the case of matrix completion, we make the assumption that the original matrix is low rank. This implies that there exists a high correlation between the entries of the matrix, which may make it possible to infer the missing entries from the observations. As a very simple example, consider the following matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 & ? & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ ? & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{141}$$

Setting the missing entries to 1 yields a rank 1 matrix, whereas setting them to any other number yields a rank 2 or rank 3 matrix.

The low-rank assumption implies that if the matrix has dimensions $m \times n$ then it can be factorized into two matrices that have dimensions $m \times r$ and $r \times n$. This factorization allows to encode the matrix using $r\,(m+n)$ parameters. If the number of observed entries is larger than $r\,(m+n)$ parameters then it may be possible to recover the missing entries. However, this is not enough to ensure that the problem is well posed.

## 4.2 When is matrix completion well posed?

Whether low-rank matrix completion is well posed obviously depend on the subset of entries that are observed. For example, completion is impossible unless we observe at least one entry in each row and column. To see why let us consider a rank 1 matrix for which we do not observe the second row,

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ ? & ? & ? & ? \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ ? \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}. \tag{142}$$

If we set the missing row to equal the same value, we obtain a rank-1 matrix consistent with the measurements. In this case, the problem is not well posed.

In general, we need samples that are distributed across the whole matrix. This may be achieved by sampling entries uniformly at random. Although this model does not completely describe matrix completion problems in practice (some users tend to rate more movies, some movies are very popular and are rated by many people), making the assumption that the revealed entries are random simplifies theoretical analysis and avoids dealing with adversarial cases designed to make deterministic patterns fail.

We now turn to the question of what matrices can be completed from a subset of entries samples uniformly at random. Intuitively, matrix completion can be achieved when the information contained in the entries of the matrix is *spread out* across multiple entries. If the information is very localized then it will be impossible to reconstruct the missing entries. Consider a simple example where the matrix is sparse

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{143}$$

If we don't observe the nonzero entry, we will naturally assume that it was equal to zero.

The problem is not restricted to sparse matrices. In the following matrix the last row does not seem to be correlated to the rest of the rows,

$$M := \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ -3 & 3 & -3 & 3 \end{bmatrix}. \tag{144}$$

This is revealed by the singular-value decomposition of the matrix, which decomposes it into two rank-1 matrices.

$$M = U\,SV^T \tag{145}$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \end{bmatrix} \tag{146}$$

$$= 8 \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix} + 6 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} -0.5 & 0.5 & -0.5 & 0.5 \end{bmatrix} \tag{147}$$

$$= \sigma_1 U_1 V_1^T + \sigma_2 U_2 V_2^T. \tag{148}$$

The first rank-1 component of this decomposition has information that is very spread out,

$$\sigma_1 U_1 V_1^T = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{149}$$

The reason is that most of the entries of $V_1$ are nonzero and have the same magnitude, so that each entry of $U_1$ affects every single entry of the corresponding row. If one of those entries is missing, we can still recover the information from the other entries.

In contrast, the information in the second rank-1 component is very localized, due to the fact that the corresponding left singular vector is very sparse,

$$\sigma_2 U_2 V_2^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -3 & 3 & -3 & 3 \end{bmatrix}. \tag{150}$$

Each entry of the right singular vector only affects one entry of the component. If we don't observe that entry then it will be impossible to recover.

This simple example shows that sparse singular vectors are problematic for matrix completion. In order to quantify to what extent the information is spread out across the low-rank matrix we define a coherence measure that depends on the singular vectors.

**Definition 4.1** (Coherence)**.** *Let $U S V^T$ be the singular-value decomposition of an $n \times n$ matrix $M$ with rank $r$. The coherence $\mu$ of $M$ is a constant such that*

$$\max_{1 \leq j \leq n} \sum_{i=1}^{r} U_{ij}^2 \leq \frac{n\mu}{r} \tag{151}$$

$$\max_{1 \leq j \leq n} \sum_{i=1}^{r} V_{ij}^2 \leq \frac{n\mu}{r}. \tag{152}$$

This condition was first introduced in [5]. Its exact formulation is not too important. The point is that matrix completion from uniform samples only makes sense for matrices which are incoherent, and therefore do not have spiky singular values. There is a direct analogy with the super-resolution problem, where sparsity is not a strong enough constraint to make the problem well posed and the class of signals of interest has to be further restricted to signals with supports that satisfy a minimum separation.

## 4.3 Dual certificate for matrix completion

As discussed in the notes on convex optimization, nuclear-norm regularization is an effective method for matrix completion. In this section we derive a dual certificate that can be used to provide a theoretical analysis of the technique. The following corollary to Theorem 2.15 derives the dual of the nuclear-norm minimization problem.

**Corollary 4.2.** *Let $\Omega$ be a subset of $m$ entries of a $n_1 \times n_2$ matrix, and let $\vec{y} \in \mathbb{R}^m$. The dual of the optimization problem*

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} ||X||_* \quad \text{such that } X_\Omega = \vec{y} \tag{153}$$

*is*

$$\max_{\vec{\alpha} \in \mathbb{R}^m} \langle \vec{\alpha}, \vec{y} \rangle \qquad \text{subject to } ||M_\Omega(\vec{\alpha})||_\infty \leq 1, \tag{154}$$

*where for any $\vec{b} \in \mathbb{R}^m$ $M_\Omega(\vec{b})$ is a $n_1 \times n_2$ matrix containing $\vec{b}$ in the entries indexed by $\Omega$ and zeros elsewhere.*

*Proof.* To apply Theorem 2.15 we need to derive the adjoint of the linear operator that extracts the entries indexed by $\Omega$ from a $n_1 \times n_2$ matrix. The adjoint equals $M_\Omega$ since for any matrix $A \in \mathbb{R}^{n_1 \times n_2}$, and any $\vec{b} \in \mathbb{R}^k$,

$$\langle A_\Omega, \vec{b} \rangle = \langle A, M_\Omega(\vec{b}) \rangle. \tag{155}$$

In addition, we need the dual norm of the nuclear norm, which is the operator norm since for any matrix $A$

$$\begin{aligned} ||A||_d &:= \max_{||B||_* \leq 1} \langle A, B \rangle & (156)\\ &= ||A|| \max_{||B||_* \leq 1} \left\langle \frac{A}{||A||}, B \right\rangle & (157)\\ &= ||A|| \,, & (158) \end{aligned}$$

by definition of the nuclear norm. The result then follows automatically from the theorem. □

As in the case of compressed sensing, we can exploit duality to show that a matrix can be successfully completed. Let $X_{\text{true}}$ be a matrix such that $(X_{\text{true}})_\Omega = \vec{y}$. If there exists a dual feasible variable $\vec{\alpha}$ such that

$$||X_{\text{true}}||_* = \langle \vec{\alpha}, \vec{y} \rangle, \tag{159}$$

then $X_{\text{true}}$ is a solution to the primal problem by weak duality. Since

$$\begin{aligned} \langle \vec{\alpha}, \vec{y} \rangle &= \langle M_\Omega(\vec{\alpha}), M_\Omega(\vec{y}) \rangle & (160)\\ &= \langle M_\Omega(\vec{\alpha}), X_{\text{true}} \rangle, & (161) \end{aligned}$$

$G := M_\Omega(\vec{\alpha})$ must be of the form $UV^T + W$ where $USV^T$ is the SVD of $X_{\text{true}}$ and $W$ is such that $||W|| \leq 1$, $U^T W = 0$ and $W V = 0$ (the argument is analogous to the one in Lemma 2.18). Such an object is a subgradient of the nuclear norm at $X_{\text{true}}$ by Theorem 3.20 in the notes on convex optimization. As a result, for any $X$ such that $X_\Omega = (X_{\text{true}})_\Omega$

$$\begin{aligned} ||X||_* &\geq ||X_{\text{true}}||_* + \langle X - X_{\text{true}}, G \rangle & (162)\\ &= ||X_{\text{true}}||_* \,, & (163) \end{aligned}$$

because $X - X_{\text{true}}$ is zero on $\Omega$. Under a certain constraint on the sampling pattern and the slightly stricter condition $||W|| < 1$, a variation of this argument establishes that if $G$ exists then $X_{\text{true}}$ is the unique solution to problem (153). We omit the proof, which can be found in [5]. In order to show that matrix completion via nuclear-norm minimization succeeds, we need to show that such a dual certificate exists with high probability. For this we will need the matrix to be incoherent, since otherwise $UV^T$ may have large entries which are not in $\Omega$. This would make it very challenging to construct $G$ in a way that $UV^T = G - W$ for a matrix $W$ with bounded operator norm. The first guarantees for matrix completion were obtained by constructing such a certificate in [5] and [7]. Subsequently, the results were improved in [8], where it is shown that an *approximate* dual certificate also allows to establish exact recovery, and simplifies the proofs significantly.

## 4.4 Completing a rank-1 matrix

Let

$$X_{\text{true}} := \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & b & b \end{bmatrix} \tag{164}$$

$$= \frac{1}{\sqrt{3}} \begin{bmatrix} a & b & b \\ a & b & b \\ a & b & b \end{bmatrix}, \quad a \in (0,1), \; b := \sqrt{\frac{1-a^2}{2}}. \tag{165}$$

The matrix is normalized so that the SVD $USV^T$ of $X_{\text{true}}$ is given by

$$U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad S = 1, \qquad V = \begin{bmatrix} a \\ b \\ b \end{bmatrix}. \tag{166}$$

The value of $a$ controls the *spikiness* of the right singular vector of the matrix. In this section we use this example to illustrate the use of dual certificates in matrix completion. Our goal is to determine for what values of $a$ the matrix is recoverable from a subset of its entries via nuclear-norm minimization.

We consider measurements $\vec{y}$ corresponding to the set of indices $\Omega$ such that

$$M_\Omega(\vec{y}) := \frac{1}{\sqrt{3}} \begin{bmatrix} 0 & b & b \\ a & 0 & b \\ a & b & 0 \end{bmatrix}, \tag{167}$$

i.e. the main diagonal is missing. To show that $X_{\text{true}}$ is recovered by nuclear-norm minimization, we need to build a dual certificate $G = UV^T + W$ supported on $\Omega$. Equivalently, we need to build $W$ such that $UV^T + W$ is zero on $\Omega^c$ which fixes those entries

$$W_{\Omega^c} = -(UV^T)_{\Omega^c}. \tag{168}$$

$W$ is consequently of the form,

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & w_3 & w_5 \\ w_1 & -b & w_6 \\ w_2 & w_4 & -b \end{bmatrix}, \tag{169}$$

for some value of $w_1, \ldots, w_6$. To ensure $U^T W = 0$ and $W V = 0$, these numbers must satisfy the following system of equations,

$$w_1 + w_2 = a \tag{170}$$
$$w_3 + w_4 = b \tag{171}$$
$$w_5 + w_6 = b \tag{172}$$
$$w_3 + w_5 = \frac{a^2}{b} \tag{173}$$
$$a w_1 + b w_6 = b^2 \tag{174}$$
$$a w_2 + b w_4 = b^2. \tag{175}$$

The equations are dependent, with rank 5. We fix $w_1 := w$ and solve them to obtain

$$W = \frac{1}{\sqrt{3}} \begin{bmatrix} -a & a - \frac{wb}{a} & \frac{wb}{a} \\ w & -b & b - w \\ \frac{a^2}{b} - w & b - \frac{a^2}{b} + w & -b \end{bmatrix}, \tag{176}$$

This is a valid dual certificate as long as $||W|| < 1$. In order to determine for what values of $a$ nuclear-norm minimization achieves exact recovery we evaluate the largest singular value of $W$ for a range of values of $a$ and $w$. Figure 9 shows the results: as long as $a \leq 0.81$ then we can find values of $w$ for which $||W|| < 1$. We can confirm numerically that for $a = 0.82$ (which implies $b = 0.4047$), the solution is not $X_{\text{true}}$ but rather

$$X^* := \begin{bmatrix} 0.8095 & 0.82 & 0.82 \\ 0.4047 & 0.4047 & 0.4047 \\ 0.4047 & 0.4047 & 0.4047 \end{bmatrix}, \tag{177}$$

where $||X^*||_* = 1.7320 < 1.7321 = ||X_{\text{true}}||_*$.

# 5 Proofs

## 5.1 Proof of Lemma 3.3

By symmetry of the Gaussian probability density function, we just need to bound the probability that $u > t$. Applying Markov's inequality (Theorem 2.9 in the notes on randomization) we have

$$\mathrm{P}\left(\mathbf{u} \geq t\right) = \mathrm{P}\left(\exp\left(\mathbf{u}t\right) \geq \exp\left(t^2\right)\right) \tag{178}$$
$$\leq \mathrm{E}\left(\exp\left(\mathbf{u}t - t^2\right)\right) \tag{179}$$
$$= \exp\left(-\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2}\right) \mathrm{d}x \tag{180}$$
$$= \exp\left(-\frac{t^2}{2}\right). \tag{181}$$

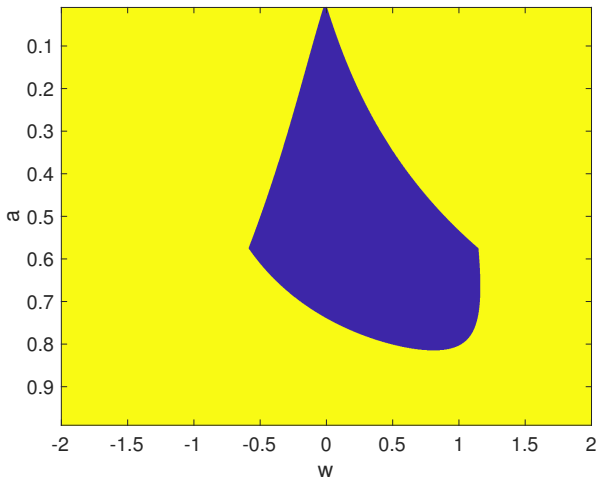


**Figure 9:** The blue entries indicate values of $w$ and $a$ for which $||W|| < 1$ in the example of Section 4.3. This reveals the values of $a$ for which nuclear-norm minimization achieves exact recovery.

# References


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

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

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

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

[5] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[6] 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.

[7] E. J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5):2053–2080, 2010.

[8] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.