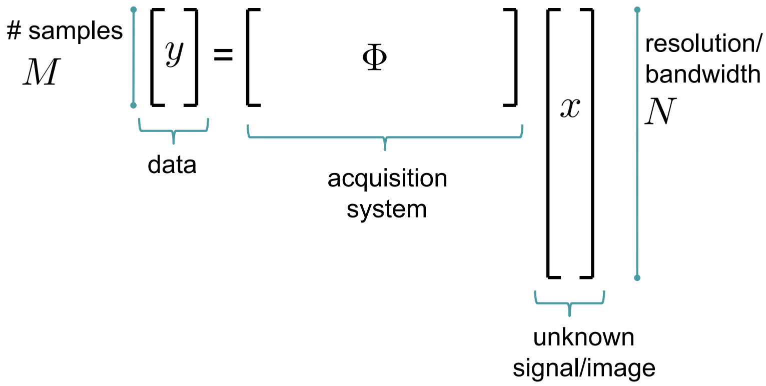
ℓ_1 minimization

We will now focus on underdetermined systems of equations:



Suppose we observe $y = \Phi x_0$, and given y we attempt to estimate x_0 by applying the pseudo-inverse of Φ to y — we are implicitly solving the program

$$\min_x \|x\|_2 \quad \text{subject to} \quad \Phi x = y.$$

Of course, we will recover x_0 exactly only under very special circumstances, namely that x_0 is in Range(Φ^*); if $x_0 = \Phi^*\alpha$ for some $\alpha \in \mathbb{R}^M$, then

$$\begin{aligned}\Phi^*(\Phi\Phi^*)^{-1}y &= \Phi^*(\Phi\Phi^*)^{-1}\Phi x_0 \\ &= \Phi^*(\Phi\Phi^*)^{-1}\Phi\Phi^*\alpha \\ &= \Phi^*\alpha \\ &= x_0.\end{aligned}$$

So if x_0 lies in a particular M-dimensional subspace of $\mathbb{R}^N$, it can be recovered from observations through the $M \times N$ matrix Φ.

Yesterday, we hinted that a different variational framework, one based on ℓ_1 minimization instead of ℓ_2 minimization, would allow us to recover sparse vectors. Given y, we solve

$$\min_x \ \|x\|_1 \quad \text{subject to} \quad \Phi x = y. \tag{1}$$

It is our goal today to formalize the conditions under which this recovery procedure is effective.

Necessary and sufficient conditions for ℓ_1 recovery

Let x_0 be a vector supported on a set $\Gamma \subset \{1, 2, \ldots, N\}$, and let $y = \Phi x_0$. It is clear that x_0 is a solution to (1) if and only if

$$\|x_0 + h\|_1 \geq \|x_0\|_1 \qquad \forall h \text{ with } \Phi h = 0. \tag{2}$$

It is always true that

$$\begin{aligned} \|x_0 + h\|_1 - \|x_0\|_1 &= \sum_{\gamma \in \Gamma} (|x_0[\gamma] + h[\gamma]| - |x_0[\gamma]|) + \sum_{\gamma \in \Gamma^c} |h[\gamma]| \\ &\geq \sum_{\gamma \in \Gamma} \operatorname{sgn}(x_0[\gamma]) h[\gamma] + \sum_{\gamma \in \Gamma^c} |h[\gamma]| \end{aligned}$$

since

$$|a + b| - |a| \geq \operatorname{sgn}(a) b \quad \text{for all } a, b \in \mathbb{R}.$$

Thus (2) holds when

$$-\sum_{\gamma \in \Gamma} \operatorname{sgn}(x_0[\gamma]) h[\gamma] \leq \sum_{\gamma \in \Gamma^c} |h[\gamma]| \quad \text{for all} \quad h \in \operatorname{Null}(\Phi). \tag{3}$$

This condition is also necessary, since if there is an $h \in \operatorname{Null}(\Phi)$ with

$$\sum_{\gamma \in \Gamma^c} |h[\gamma]| < -\sum_{\gamma \in \Gamma} \operatorname{sgn}(x_0[\gamma]) h[\gamma],$$

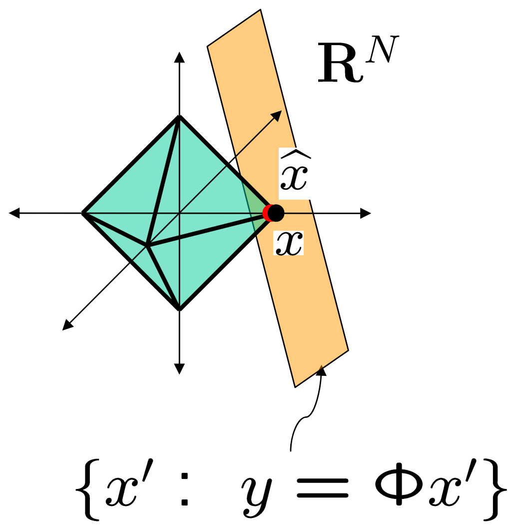
then the same is true for ϵh for all $\epsilon >$, and $\Phi(x_0 + \epsilon h) = y$, and

$$\begin{aligned} \|x_0 + \epsilon h\|_1 &= \sum_{\gamma\in\Gamma} |x_0[\gamma] + \epsilon h[\gamma]| + \sum_{\gamma\in\Gamma^c} |\epsilon h[\gamma]| \\ &< \sum_{\gamma\in\Gamma} |x_0[\gamma] + \epsilon h[\gamma]| - \epsilon \,\mathrm{sgn}(x_0[\gamma]) h[\gamma] \\ &\le \sum_{\gamma\in\Gamma} |x_0[\gamma]| \quad \text{for some small enough } \epsilon > 0 \\ &= \|x_0\|_1 \end{aligned}$$

It is interesting to note that given the observation matrix Φ, our ability to recover a vector x_0 is determined only by

1. the set Γ on which x_0 is supported (the locations of its "active elements"), and
2. the signs of the elements on this set.

The magnitudes of the entries are not involved at all. Geometrically, the support set coupled with the sign sequence specifies the *facet* of the ℓ_1 ball on which x_0 lives:



Duality and optimality

We can get a more workable sufficient condition for exact recovery of x_0 by looking at the dual of the convex program

$$\min_x \ \|x\|_1 \quad \text{subject to} \quad \Phi x = y. \tag{4}$$

Let's start by considering a general optimization program with linear equality constraints:

$$\min_x \ f(x) \quad \text{subject to} \quad \Phi x = y. \tag{5}$$

The *Lagrangian* for this problem is

$$L(x, v) = f(x) + v^*(\Phi x - y)$$

If f is differentiable, then x is a solution to (5) if it is feasible, $\Phi x = y$, and there exists a $v \in \mathbb{R}^M$ such that

$$\nabla_x L(f, v) = \nabla_x f(x) + \Phi^* v = 0.$$

We are interested in the particular functional $f(x) = \|x\|_1$, which is not differentiable but is convex. Fortunately, there is an easy modification to the condition above in this case — x is a solution to (5) if it is feasible and there exists a $v \in \mathbb{R}^M$ such that

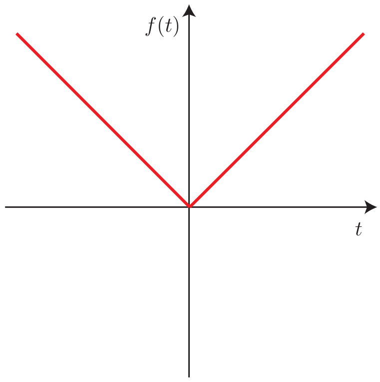
$$\Phi^* v \text{ is a "subgradient" of } f \text{ at } x.$$

A vector $u \in \mathbb{R}^N$ is a subgradient of a function f at x_0 if it is normal to a "supporting hyperplane"; that is if

$$f(x) \geq f(x_0) + u^*(x - x_0).$$

If f is differentiable at x_0, then $\nabla f(x_0)$ is the only subgradient.

To make this concept more concrete, consider the (very relevant) one dimensional example $f(t) = |t|$.



In this case, the set of subgradients is simply the derivative (± 1) for t away from the origin, and all slopes between -1 and $+1$ at the origin:

$$\{\text{subgradients}\} = \begin{cases} \text{sgn}(t) & t \neq 0 \\ [-1, 1] & t = 0 \end{cases}.$$

In $\mathbb{R}^N$, for $f(x) = \|x\|_1$, the subgradients of f at x_0 are the collection of vectors u such that

$$\begin{aligned} u[\gamma] &= \text{sgn}(x_0[\gamma]), & \gamma \in \Gamma \\ |u[\gamma]| &\leq 1, & \gamma \in \Gamma^c, \end{aligned}$$

where Γ is the support set of x_0.

Given an $M \times N$ matrix Φ and a vector $x_0 \in \mathbb{R}^N$ supported on Γ, set $y = \Phi x_0$. Then x_0 is a solution to (4) if there exits a $v \in \mathbb{R}^M$ such that

$$\begin{aligned} (\Phi^* v)[\gamma] &= \operatorname{sgn}(x_0[\gamma]), & \gamma \in \Gamma \\ |(\Phi^* v)[\gamma]| &\leq 1, & \gamma \in \Gamma^c. \end{aligned}$$

In addition, if Φ is injective on the set of all vectors supported on Γ and we can find a v such that the second condition is

$$|(\Phi^* v)[\gamma]| < 1, \qquad \gamma \in \Gamma^c$$

then x_0 is the unique solution.

Choosing a particular dual vector

Our recovery conditions boxed above simply ask that we be able to find one such v (there could be many). Many of the results in the fields of sparse recovery and compressed sensing have narrowed down the condition down by simply testing a prescribed vector. Let

$$\Phi_\Gamma = \text{ the } M \times |\Gamma| \text{ submatrix containing the columns of } \Phi \text{ indexed by } \Gamma$$

and let

$$z_0 \in \mathbb{R}^{|\Gamma|} \text{ contain the signs of } x_0 \text{ on } \Gamma.$$

We set

$$u_0 = \Phi^* \Phi_\Gamma (\Phi_\Gamma^* \Phi_\Gamma)^{-1} z_0. \tag{6}$$

Now sufficient conditions for x_0 to be the unique minimizer of (4) are

1. $\Phi_\Gamma^*\Phi_\Gamma$ is invertible. If this is true, then the expression above for u_0 is well-behaved, and b construction we will have $u_0[\gamma] = \operatorname{sgn}(x_0[\gamma])$;
2. If 1) holds, then we need $|u_0[\gamma]| < 1$.

Note on complex vectors: Almost everything we have said so far about ℓ_1 minimization extends in a straightforward manner to complex-valued vectors. First, it is worth mentioning that if $x \in \mathbb{C}^N$, then

$$\|x\|_1 = \sum_{n=1}^{N} |x[n]| = \sum_{n=1}^{N} \sqrt{\operatorname{Re}\{x[n]\}^2 + \operatorname{Im}\{x[n]\}^2}.$$

In this case, the ℓ_1 minimization program can no longer be recast as a linear program, but rather is what is called a "sum of norms" program (which is a particular type of "second order cone program"). This type of problem, however, is not too much more difficult to solve from a practical perspective.

The sufficient conditions for recovery are the same, but now we take $\operatorname{sgn}(z)$ to be the *phase* of the complex number z. That is, if $z = A\mathrm{e}^{\mathrm{j}\theta}$, then $\operatorname{sgn}(z) = \theta$.

Example: Spikes and Sines Dictionary

Let us return to the Φ created by concatenating the "spikes" orthobasis and the Fourier orthobasis:

$$\Phi = \begin{bmatrix} I & F \end{bmatrix}.$$

Here the matrix is $n \times 2n$. Suppose that

$$\alpha_0 \in \mathbb{C}^{2n} \text{ is supported on a set } \Gamma = T \cup \Omega$$

where as before T denotes the "spike" support and Ω the "sine" support.

Yesterday, in establishing the Donoho-Starck uncertainty principle, we saw that we could write

$$\Phi_\Gamma^* \Phi_\Gamma = I + G, \quad \text{where} \quad \|G\| \le \frac{|T|\,|\Omega|}{n}.$$

Thus, if

$$|T|\,|\Omega| \le \delta n,$$

for some $0 < \delta < 1$, we have

$$\|(\Phi_\Gamma^* \Phi_\Gamma)^{-1}\| \le \frac{1}{1-\delta}.$$

So our first sufficient condition holds when $|T|\,|\Omega| < n$, which is implied by

$$|\Gamma| = |T| + |\Omega| < 2\sqrt{n}.$$

For the second condition, construct u_0 as in (6), where now z_0 are the phases of α_0 on Γ. Then for a particular entry $\gamma \in \Gamma^c$,

$$\begin{aligned} u_0[\gamma] &= \Phi_\gamma^* \Phi_\Gamma (\Phi_\Gamma^* \Phi_\Gamma)^{-1} z_0 \\ &= \langle w_\gamma, z_0 \rangle \end{aligned}$$

where Φ_γ is the column of Φ corresponding to γ, and w_γ is the vector

$$w_\gamma = (\Phi_\Gamma^* \Phi_\Gamma)^{-1} \Phi_\Gamma^* \Phi_\gamma.$$

Using Cauchy-Schwarz,

$$\begin{aligned} |u_0[\gamma]| &\leq \|w_\gamma\| \, \|z_0\| \\ &\leq \|(\Phi_\Gamma^* \Phi_\Gamma)^{-1}\| \cdot \|\Phi_\Gamma^* \Phi_\gamma\|_2 \cdot \sqrt{|T| + |\Omega|}. \end{aligned}$$

The entries in the vector $\Phi_\Gamma^* \Phi_\gamma$ are either equal to zero or $1/\sqrt{n}$, and so

$$|u_0[\gamma]| \ \leq \ \frac{1}{1-\delta} \cdot \sqrt{\frac{|T| + |\Omega|}{n}} \cdot \sqrt{|T| + |\Omega|}.$$

When can we guarantee that the expression on the right less than 1? We know that we will at least need

$$|\Gamma| = |T| + |\Omega| \leq \sqrt{n}.$$

In this case, $|T| \cdot |\Omega| \leq n/4$, so we can take $\delta = 1/4$, and so if

$$|\Gamma| = |T| + |\Omega| \leq \frac{3\sqrt{n}}{4}$$

then

$$\max_{\gamma \in \Gamma^c} |u_0[\gamma]| < 1$$

and α_0 will be the minimizer to (4).

Example: Dictionaries with small coherence

Now suppose that Φ is an $M \times N$ matrix with normalized columns,

$$\|\Phi_\gamma\|_2 = 1, \quad \gamma = 1, 2, \ldots, N$$

and *coherence*

$$\mu = \max_{\substack{1 \le \gamma_1, \gamma_2 \le N \\ \gamma_1 \ne \gamma_2}} |\langle \Phi_{\gamma_1}, \Phi_{\gamma_2} \rangle|.$$

The quantity μ is essentially a measure of how closely aligned any two columns of Φ are.

Let Γ be a fixed subset of $\{1, 2, \ldots, N\}$, and x_0 be a vector supported on Γ with sign sequence z_0. For the first optimality condition note that we can write $\Phi_\Gamma^* \Phi_\Gamma$ as

$$\Phi_\Gamma^* \Phi_\Gamma = I + G$$

where each entry of the $|\Gamma| \times |\Gamma|$ matrix G is less than or equal to μ. We have

$$\begin{aligned} \|G\| &\le \|G\|_F \\ &= \sqrt{\sum_{j,k} |G_{j,k}|^2} \\ &\le \mu |\Gamma|, \end{aligned}$$

and so we can guarantee $\Phi_\Gamma^* \Phi_\Gamma$ is invertible when

$$|\Gamma| \ \le \ \frac{1}{\mu}.$$

To check the second condition, we have for $\gamma \in \Gamma^c$

$$\begin{aligned} |u_0[\gamma]| &\le \|(\Phi_\Gamma^* \Phi_\Gamma)^{-1}\| \, \|\Phi_\Gamma^* \Phi_\gamma\|_2 \, \|z_0\|_2 \\ &\le \frac{1}{1 - \mu|\Gamma|} \|\Phi_\Gamma^* \Phi_\gamma\|_2 \sqrt{|\Gamma|}. \end{aligned}$$

Since $\gamma \notin \Gamma$, all the entries of $\Phi_\Gamma^* \Phi_\gamma$ will have magnitude less than μ, and so $\|\Phi_\Gamma^* \Phi_\gamma\|_2 \leq \mu\sqrt{|\Gamma|}$, and so

$$|u_0[\gamma]| \leq \frac{\mu|\Gamma|}{1 - \mu|\Gamma|}.$$

Thus

$$|\Gamma| \; < \; \frac{1}{2\mu}$$

ensures that x_0 will be recoverable from observations $y = \Phi x_0$.