

An Implementation of Sparse Recovery via l_1 Minimization

William Hilton

*Department of Mathematics
University of Central Florida*

MOTIVATION

- ▶ The problem of reconstructing, detecting, classifying, or estimating a signal from incoherent measurements is encountered in a variety of areas, e.g. optics, medical imaging, and data acquisition.
- ▶ Specific motivations include: reducing radiation exposure in MRIs; reducing cost of photon-detecting equipment; reducing data bandwidth, etc.

NARRATIVE

- ▶ Let α be some N -dimensional data.
- ▶ Let Ψ map α to some superposition, denoted by P .
- ▶ The support of P in Ψ is $K \leq N$ dimensions.
- ▶ On Ψ , take M *exact* measurements, m , of P via, say, Φ .
- ▶ Solve a minimization problem and then reconstruct P as $\hat{P}$.
- ▶ Track $\|P - \hat{P}\|$ as M and K vary.

PRINCIPAL THEORY

Let $\alpha \in \mathbb{R}^{21}$, $\Psi : [-1, 1] \subset \mathbb{R}^{21} \rightarrow [-1, 1] \subset \mathbb{R}^{21}$ of Vandermonde-type, $\Psi\alpha$ S -sparse, $\Phi : \mathbb{R}^M \rightarrow \mathbb{R}^N$, and define

$$P := \Psi\alpha \tag{1}$$

$$m := \Phi P \tag{2}$$

$$\hat{\alpha} := \min \|\alpha_0\|_1 \in \mathbb{R}^N \text{ s.t. } \Phi\Psi\alpha_0 = m \tag{3}$$

$$\hat{P} := \Psi\hat{\alpha} \tag{4}$$

$$K \sim S \tag{5}$$

then

$$M \geq K \log(N) \implies \|P - \hat{P}\| = 0 \tag{6}$$

as long as (2) is “uniformly uncertain” or “incoherent”, i.e.

$$\frac{4M}{5N} \|\Psi\alpha\|_2^2 \leq \|\Phi\Psi\alpha\|_2^2 \leq \frac{5M}{4N} \|\Psi\alpha\|_2^2 \tag{7}$$

APPROACH

- ▶ Let entries in α , Ψ , and Φ vary randomly (save for the K -sparse control of α and the Vandermonde structure of Ψ), all restricted to $[-1, 1]$ for computational preconditioning.
- ▶ Note: Results can be generalized to the desired interval.
- ▶ Next, compute $\|P - \hat{P}\|$ as $K \leq 21$ and $M \leq 3 \cdot 21$ vary using $l - 1$ magic and Matlab's `fmincon` for Eqn. (3)
- ▶ Repeat 25 times to assess variational tendencies

ISSUE

- ▶ Even the 'preconditioned' Vandermonde operator, Ψ , and the uniformly uncertain measurement operator, Φ , always enjoy a very large condition number and their composition even more so.
- ▶ Although control of the condition number of Φ preserves its uniform uncertainty, control of the sample space generating Ψ has no observable effect on the conditioning of Ψ . **Moreover, directly controlling the condition number of Ψ always results in a false sample space (dim. > 21).**
- ▶ Hint: $\text{cond}(\Psi)$ is always near 10^{21} . Hmmm...

OBSERVATION

- ▶ **Condition number of Vandermonde matrices is BOUNDED BELOW by an exponential for real entries.** Only under very strict conditions (namely when sample space is equidistant points on unit circle in $\mathbb{C}$) can Vandermonde matrix be well-conditioned – hence not a random sample space.
- ▶ Proof: Google “Arxiv Victor Pan How Bad Are Vandermonde Matrices”

RESOLUTION

- ▶ Use matrix-free conjugate gradients version of minimization algorithms to avoid condition number issue and proceed as planned.
- ▶ Consider the fixed, well-conditioned case separately, where we force $\text{cond}(\Phi) \equiv 1$ and $\text{cond}(\Psi) \equiv 1$ by exploiting singular value decomposition for the former and unit circle sample space $x_k = \cos(k\frac{2\pi}{21}) + i\sin(k\frac{2\pi}{21})$, $k = 0, \dots, 20$ for latter.

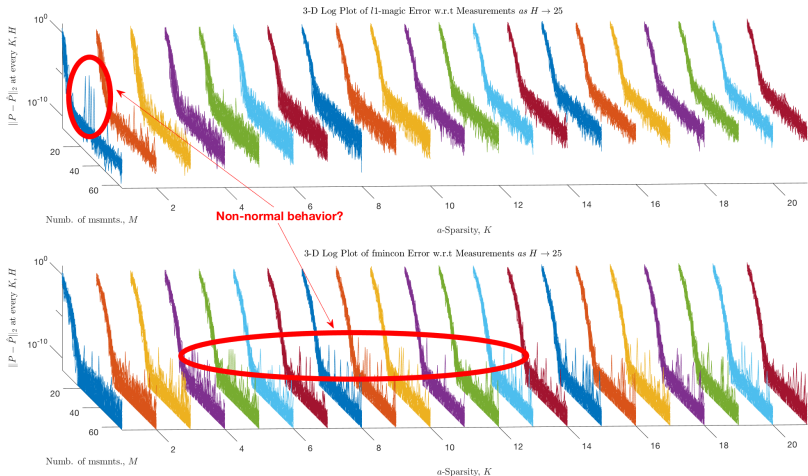
INITIAL NOTES

Default Stopping Criteria

	Error tolerance	Max. iterations
Least squares guess	10^{-8}	500
$l - 1$ magic	10^{-8}	500
fmincon	10^{-6}	3000

- ▶ Take error tolerance to be sufficient condition for algorithm's success
- ▶ Compute least squares/minimum energy guess for algorithm initialization
- ▶ Visual inspection of data suggests behavior may not be statistically normal. See *Figure 1*.

FIG. 1



RESOLUTION

Measure the standard deviation at every (M, K) with respect to H . See *Figure 2* and *Figure 3*.

REMARKS

- ▶ Error measures are reasonably normal over space of hyperiterations (i.e. ≤ 2)
- ▶ `fmincon` enjoys a tighter error distribution, likely due to default iterative stopping criterion
- ▶ Therefore, we may proceed by considering the mean errors at every (M, K) as reasonably representative

RESULTS

For $l - 1$ magic, successful (M, K) pairs occur about when

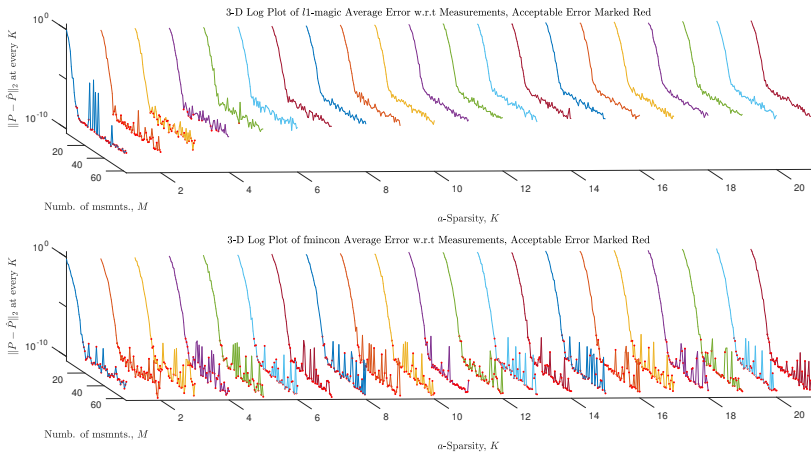
$$\frac{M}{K \log(21)} \geq 2.2 \quad (8)$$

excluding all $(M = 1, K)$ and all $(M, K > 5)$. For `fmincon`, successful (M, K) almost always occur when

$$M \geq 17 \quad (9)$$

and admits no reasonable inequality bound like $l - 1$ magic, with similar $M = 1$ behavior. See *Figure 4*, *Figure 5*, and *Figure 6*.

FIG. 4



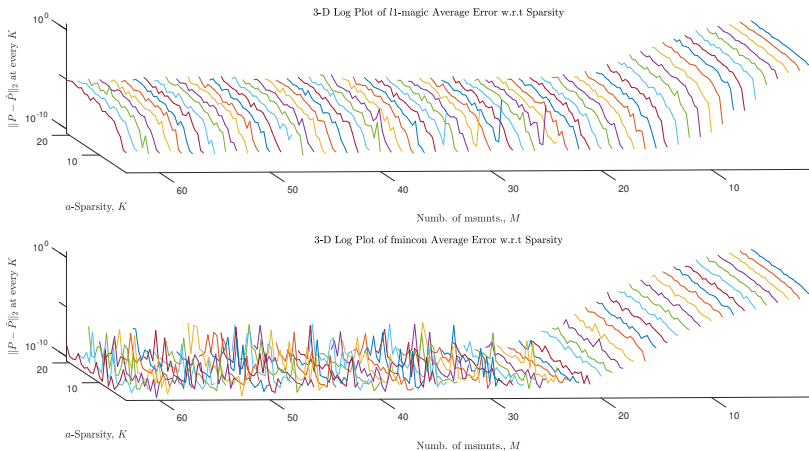
[illegible]

[illegible][illegible]

RESULTS (CONT.)

For $l - 1$ magic, error as K varies is relatively smooth (i.e., predicts) nearby error at different M . This does not hold for `fmincon`. See *Figure 7*.

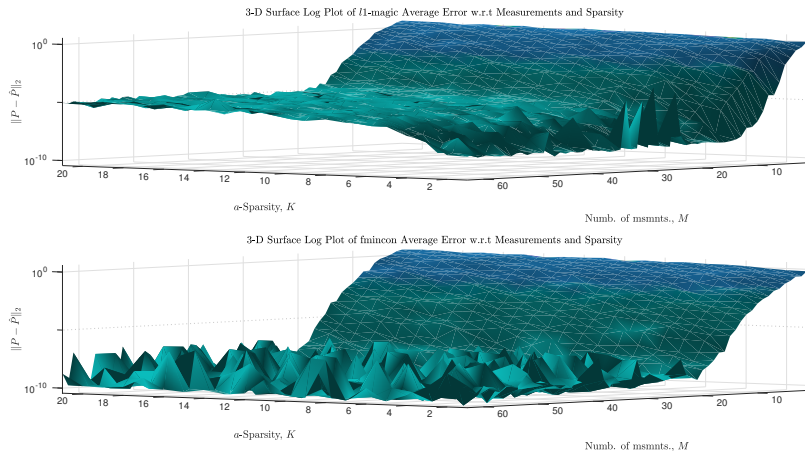
FIG. 7



RESULTS (CONT.)

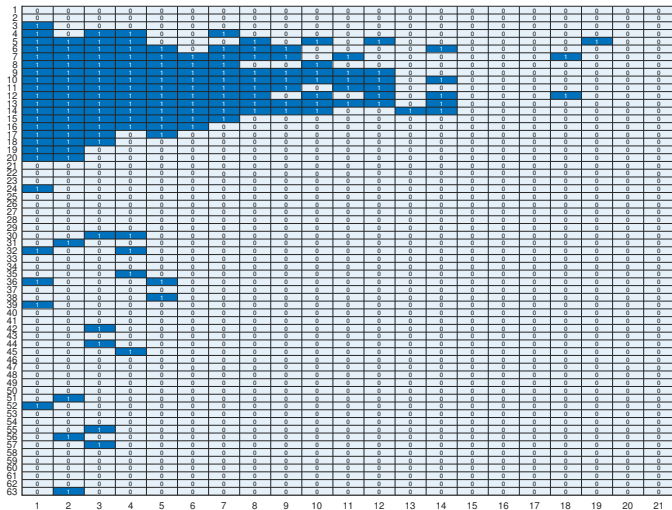
- ▶ On the whole, `fmincon` is superior to $l - 1$ magic using default options for both. Notably, total computation time for **`fmincon` exceeded five hours, while $l - 1$ magic required only about twenty-five minutes.** See Figure 8.
- ▶ Caveat: Don't be deceived by log plots! Error is indeed small for large M .

FIG. 8



RESULTS (CONT.)

- In *Figure 9*, observe the very notable exceptions where the particularly expensive `fmincon` function loses to $l - 1$ magic.

FIG. 9: (M, K) WHERE FMINCON LOSES (=1)

SUMMARY OF FINDINGS

	K small	K large
M small	Quad. 1	Quad. 4
M large	Quad. 2	Quad. 3

- **Quad. 1:** Most surprising finding. Cheap conjugate gradients method ($l - 1$ magic) is superior to very expensive fmincon function. Suggests problems in application require fine-tuned numerical schemes.

SUMMARY OF FINDINGS

	K small	K large
M small	Quad. 1	Quad. 4
M large	Quad. 2	Quad. 3

- **Quad. 2:** Here, $l - 1$ magic partly validates theory by relating sparsity of S to K (via coeff. ≈ 2.2) while `fmincon` cannot resolve even when given ample resources.

SUMMARY OF FINDINGS

	K small	K large
M small	Quad. 1	Quad. 4
M large	Quad. 2	Quad. 3

- ▶ **Quad. 3:** “Inverse” of Quad 2. `fmincon` partly validates theory by picking up where $l - 1$ magic fails. $l - 1$ magic cannot resolve.

SUMMARY OF FINDINGS

	K small	K large
M small	Quad. 1	Quad. 4
M large	Quad. 2	Quad. 3

- **Quad. 4:** Most expected finding. Both algorithms unstable (via std.dev. measure), problem essentially ill-posed.

CONCLUSION

Checklist for measuring:

- Measure randomly in as many dimensions as possible while avoiding noise
- Hope that your target data is sparse in these dimensions
- Custom-tailor your minimization algorithm to behave when given your measurements/sparsity
- Reconstruct and enjoy

FIG. 10: PRIMAL DUAL METHOD'S NOISE

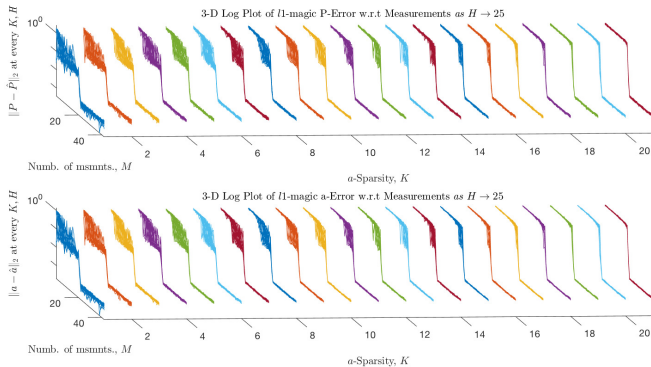


FIG. 11: PRIMAL DUAL METHOD'S AVG. SUCCESS INDICATIONS

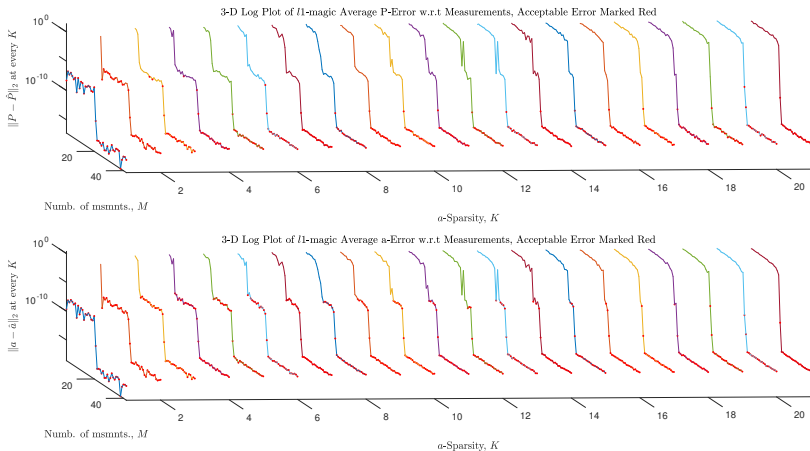


FIG. 12: PRIMAL DUAL METHOD'S SMOOTHNESS W.R.T. SPARSITY

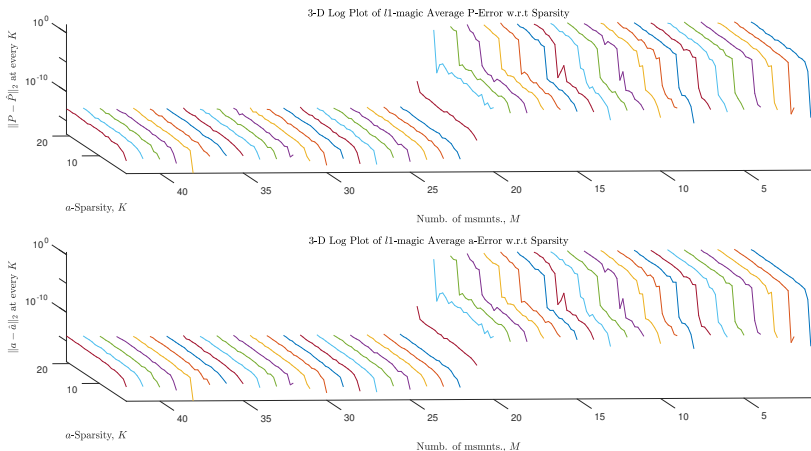
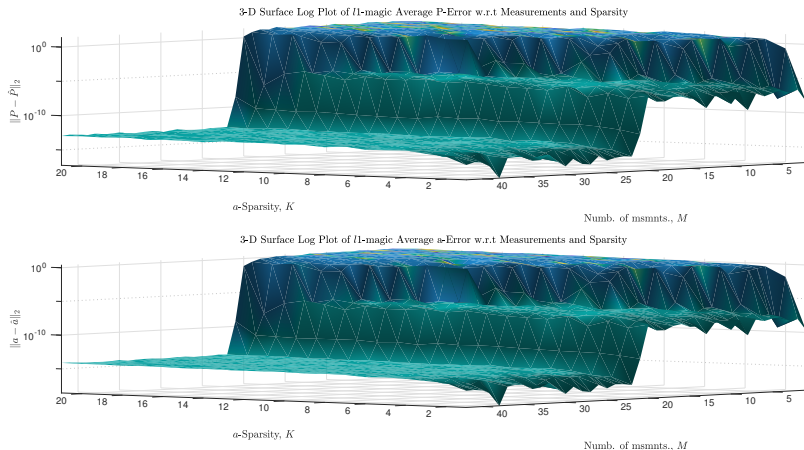


FIG. 13: PRIMAL DUAL METHOD'S ERROR COMPARISON



CONCLUSION

- Fixed sample space results indicate that, if possible, one should measure with respect to an optimal sample space.

END

Thank you for listening.