# *Exact signal reconstruction from highly incomplete nonlinear system*

https://doi.org/10.31713/MCIT.2025.063

Cherhykalo Denys
Faculty of Computer Science and Cybernetics
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
denischergicalo@gmail.com

*Abstract*— **L1 regularization has long been used to obtain sparse representations of data, but is it possible to extend the convergence conditions from the linear to the nonlinear case? L1 regularization does not guarantee accurate reconstruction of a sparse state from incomplete data, even under linear constraints. However, once the amount of data exceeds a certain threshold, the probability of reconstruction failure with random matrices will be negligible. Developing a theory to determine the probability of success in the nonlinear case seems useful. This requires defining various classes of random nonlinear functions and examining how their convergence conditions vary. In particular, linear processes are often affected by nonlinear distortions, which are commonly modeled as noise. Identifying the physics and chemical kinetics of these distortions and incorporating them into the model will help improve the probability of successful reconstruction and enable the identification of factors that influence these distortions. This is especially important in biology, where understanding not only the balance of internal processes but also the factors influencing the stability of the system is crucial.**

**Keywords — L1 regularization, random matrices, sparse reconstruction, nonlinear inverse problems.**

## I. Introduction

There are many studies related to L1 regularization for nonlinear cases[9][10][11]. In general, the optimum is determined by the Karush–Kuhn–Tucker conditions. But in this article we will try to describe a basis within which all solutions satisfy nonlinear conditions and within which we will obtain a problem without constraints.

It should also be clarified that there are already problems with some types of nonlinearity that can be reduced to linear ones, which in turn will link the convergence conditions with the linear case[2], although this was done for a specific case.

As in this article, we use singular value decomposition. This decomposition makes it easiest to construct the necessary basis, and this basis is also used in the algorithm itself.

In this article, we also leave room for discussion on how exactly to achieve the optimum in this new basis, there are already articles with practical applications for this[3][4][5][6][7][8].

We will consider those cases when a nonlinear operator can be translated by simple transformations into a nearly linearized form, so that in a given space it looks like this:

$$A(x) = Bx + \varepsilon F(x), B \in R^{n\times m}, m < n,$$

$$x \in U \subset R^n.$$

And also for the given output y there must exist x such that:

$$A(\dot{x}) = y$$

$$|\{i: \dot{x}_i \neq 0\}| = d \ll M$$

We also assume that:

$\|J(F,x)\|_{L_1} \leq 1, x \in U, J(F,x)$ - Jacobian of F.

And the matrix $B$ has independent random coefficients with uniform distribution in the interval $[-\mu, \mu], \mu > 0$.

Without loss of generality, we will assume that only the first d elements are non-zero:

$$\dot{x}_i = 0, i > d$$

Let us consider the singular value decomposition of matrix B:

$$B = U\Sigma V$$

To begin, let's find at least one solution that satisfies our conditions, not necessarily sparse:

$$A(\bar{x}) = B\bar{x} + \varepsilon F(\bar{x}), B \in R^{n\times m}, m < n$$

Let's consider:

$$x = \bar{x} + dx$$

$$A(x) = B(\bar{x} + dx) + \varepsilon F(\bar{x} + dx) = y = B\bar{x} + \varepsilon F(\bar{x})$$

Thus we have:

$$Bdx + \varepsilon\big(F(\bar{x} + dx) - F(\bar{x})\big) = 0$$

Let's replace dx with $V^T b$ and multiply by U^T on the left:

$$U^T U\Sigma V V^T b + \varepsilon U^T\big(F(\bar{x} + V^T b) - F(\bar{x})\big) = 0$$

We reduce unitary matrices and obtain:

$$\Sigma b + \varepsilon U^T\big(F(\bar{x} + V^T b) - F(\bar{x})\big) = 0$$

Using the auxiliary function $G$ we have:

$$G(\alpha) = G_{\bar{x}}(\alpha) = U^T\big(F(\bar{x} + V^T\alpha) - F(\bar{x})\big)$$

$$\Sigma b + \varepsilon G(b) = 0$$

To begin with, let's consider a linear branch, that is, $\varepsilon = 0$:

$$\Sigma b = 0$$

## II. Linear Case

Terence Tao and others[1] proved for the linear case that for:

$$d \le m\frac{\alpha(r)}{log(n)}$$

the recovery probability for l1 minimization is:

$$p_{rec} = 1 - O(n^{-r}),$$

where $\alpha(r)$ some function of m close to:

$$\alpha(r) \approx \frac{1}{29.6(r+1)}$$

Similar probability estimates to this article can be obtained by following these steps:

$$rank(\Sigma) = m \text{ thus:}$$
$$b_i = 0, i \le m$$
$$b_i \in R, i > m$$
$$b = I_{n-m}^n c, c \in R^{n-m}$$

Thus, the problem is reduced to the usual non-smooth optimization:

$$F(c) = \|V^T I_{n-m}^n c + \bar{x}\|_{l_1} = \|Kc + \bar{x}\|_{l_1}$$

The optimum point is determined by the presence of a zero vector in the subgradient:

$$0 \in \partial F(c) = S(Kc + \bar{x})K,$$
$$S(x) = \{s: s_i = sign(x_i) if x_i \ne 0, |s_i| \le 1 if x_i = 0\}$$

That is, for $\dot{x}$ to be an optimum, the following conditions must be met:

$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d})K^T = 0, \|l\|_\infty \le 1$$
$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d}) = r_1 I_n^m V, \|l\|_\infty \le 1, r_1 \in R^m$$
$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d}) = r_2 \Sigma V, \|l\|_\infty \le 1, r_2 \in R^m$$
$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d}) = rU\Sigma V, \|l\|_\infty \le 1, r \in R^m$$
$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d}) = rB, \|l\|_\infty \le 1$$
$$\big(sign(\dot{x}_1), \dots, sign(\dot{x}_d)\big) = rB_1, \|rB_2\|_\infty \le 1$$

$$\beta = M[\xi_1^2 + \dots + \xi_n^2] = n\frac{\mu^2}{3}, \xi_i = U([-\mu, \mu])$$

Where $M[\cdot]$ - mathematical expectation for random matrices with a fixed $\dot{x}$.

$$\dot{r} = \beta^{-1}\big(sign(\dot{x}_1), \dots, sign(\dot{x}_d)\big)B_1^T$$
$$M[T(B_1)] = M[E - \beta^{-1}B_1^T B_1] = 0$$
$$M[B_1^T B_2] = 0$$
$$M[\dot{r}B_1] = \big(sign(\dot{x}_1), \dots, sign(\dot{x}_d)\big)$$
$$M[\dot{r}B_2] = 0$$
$$r = \big(sign(\dot{x}_1), \dots, sign(\dot{x}_d)\big)\big(E - T(B_1)\big)^{-1}\beta^{-1}B_1$$

We substitute this $r(B_1, \dot{x})$ and we obtain sufficient stability conditions for $\dot{x}$:

$$r(B_1, \dot{x})B_1 = \big(sign(\dot{x}_1), \dots, sign(\dot{x}_d)\big)$$
$$\|r(B_1, \dot{x})B_2\|_\infty < 1$$

To do this, it is sufficient to prove:

$$\|\dot{r}B_2\|_\infty + \|(r(B_1, \dot{x}) - \dot{r})B_2\|_\infty < 1$$

Because of independence of the matrices $B_1, B_2$ we have $\|\dot{r}B_2\|_\infty$ is close to zero.

Because of $T(B_1)$ is close to zero we have$\|(r(B_1, \dot{x}) - \dot{r})B_2\|_\infty$ is close to zero.

Therefore, although this is a sufficient but not a necessary condition, it describes a large number of cases of convergence

## III. Nonlinear Case

For the nonlinear case, we consider:

$$b = (e_1, \dots, e_m, c_1, \dots, c_{n-m}), b = I_m^n e + I_{n-m}^n c$$
$$e \in R^m, c \in R^{n-m}$$

Substituting into the expression we will have:

$$\Sigma(I_m^n e + I_{n-m}^n c) + \varepsilon G(I_m^n e + I_{n-m}^n c) = 0$$
$$\Sigma I_m^n e + \varepsilon G(I_m^n e + I_{n-m}^n c) = 0$$

Solving this equation for $e$ we get $e(c)$:

$$\Sigma I_m^n e(\mathrm{c}) + \varepsilon G(I_m^n e(\mathrm{c}) + I_{n-m}^n c) = 0$$

Let's take the Jacobian of the expression:

$$\Sigma I_m^n J(e, \mathrm{c}) + \varepsilon J\big(G, b(c)\big)(I_m^n J(e, \mathrm{c}) + I_{n-m}^n) = 0$$
$$\big(\Sigma I_m^n + \varepsilon J\big(G, b(c)\big)I_m^n\big)J(e, \mathrm{c}) = -\varepsilon J\big(G, b(c)\big)I_{n-m}^n$$

From which we obtain an expression for the $J(e, \mathrm{c})$:

$$-\varepsilon\big(\Sigma I_m^n + \varepsilon J\big(G, b(c)\big)I_m^n\big)^{-1} J\big(G, b(c)\big)I_{n-m}^n$$

where:

$$b(\mathrm{c}) = I_m^n e(\mathrm{c}) + I_{n-m}^n c$$

Our problem with nonlinear constraints has been reduced to the following:

$$F(c) = \|V^T b(c) + \bar{x}\|_{l_1} \to min$$

The optimum point is determined by the presence of a zero vector in the Clarke subdifferential:

$$0 \in \partial F(c) = S(Kc + \bar{x})V^T J(b, c)$$
$$S(x) = \{s: s_i = sign(x_i) if x_i \ne 0, |s_i| \le 1 if x_i = 0\}$$

To obtain the stability conditions, we introduce the substitution:

$$\dot{B} = Bx + \varepsilon J(F, \dot{x}), \dot{F}(x) = F(x) - J(F, \dot{x})x$$

We have a problem where at $J(\dot{F}, \dot{x}) = 0$:

$$A(x) = \dot{B}x + \varepsilon\dot{F}(x), B \in R^{n \times m}, m < n$$

Then, since:

$$\dot{B} = \dot{U}\dot{\Sigma}\dot{V}, \dot{K} = \dot{V}I_{n-m}^n$$
$$\dot{G}_{\dot{x}}(\alpha) = \dot{U}^T\left(\dot{F}(\dot{x} + \dot{V}^T\alpha) - \dot{F}(\dot{x})\right)$$

For $J(\dot{e}, \dot{\mathrm{c}})$ we have:

$$-\varepsilon\left(\dot{\Sigma}I_m^n + \varepsilon J\left(\dot{G}, \dot{b}(\dot{c})\right)I_m^n\right)^{-1} J\left(\dot{G}, \dot{b}(\dot{c})\right)I_{n-m}^n$$

Therefore, if:

$$(sign(\dot{x}_1), \dots, sign(\dot{x}_d), l_1, \dots, l_{n-d})\dot{K} = 0$$
$$\|l\|_\infty < 1$$
$$\dot{K} = \left(E - \varepsilon(\dot{\Sigma}I_m^n)^{-1}\dot{U}^T J(F, \dot{x})\right)K$$

Which doesn't spoil the conditions too much, especially if there is:

$$\big(sign(\dot{x}_1), \dots, sign(\dot{x}_d), \dot{l}_1, \dots, \dot{l}_{n-d}\big)K = 0$$
$$\left\|\dot{l}\right\|_\infty < 1 - C\varepsilon$$

Where the constant C, as can be seen, depends on the distribution of singular values $\dot{\Sigma}I_m$for a given class of random matrices.

## IV. Implementation of the algorithm

The author of the article believes that nonlinear reconstruction algorithms, especially in the class described in the introduction, benefit from the transition to the basis that was considered.

We will demonstrate a potential implementation using a simple algorithm:

$$\bar{g}_0 = g_0 \in \partial F(c_0)$$
$$\bar{g}_{k+1} = (1 - \beta)g_k + \beta\bar{g}_k, g_k \in \partial F(c_k)$$
$$\mathrm{c}_{k+1} = c_k - \lambda_k\frac{\bar{g_{k+1}}}{\|\bar{g_{k+1}}\|_\infty + \delta}$$

$$\lambda_k = \frac{\lambda_0}{1 + k^{\frac{2}{3}}}$$

where $\delta$ - computational error.

To apply any subradient algorithm we need to calculate the subgradient, and for the subradient we need to calculate the function $vJ(b,c), v \in S(Kc + \bar{x})V^T$.

First we need to calculate the function $e(c_k)$:

Algorithm 1

For each $e(c_k)$ we generate a sequence $e_i$:

$$e_0 = \begin{cases} e(c_{k-1}) if k > 0 \\ 0 if k = 0 \end{cases}$$

$$e_{i+1} = -\varepsilon(\Sigma I_m^n)^{-1} G(I_m^n e_i + I_{n-m}^n c_k)$$

while($\|e_{i+1} - e_i\|_{l_1} > 2\delta$).

For convergence, it is sufficient that:

$$T_1(\alpha) = -\varepsilon(\Sigma I_m^n)^{-1} G(I_m^n \alpha + I_{n-m}^n c_k)$$

was a non-stretching operator.

Now we can already define the iterations for the calculation $g = vJ(b, c_k)$:

Algorithm 2

$$g = -\varepsilon v\big(\Sigma I_m^n + \varepsilon J\big(G, b(c_k)\big) I_m^n\big)^{-1} J\big(G, b(c_k)\big) I_{n-m}^n$$

$$g = -\varepsilon w(\Sigma I_m^n)^{-1} J\big(G, b(c_k)\big) I_{n-m}^n$$

$$w = v\big(E_m + \varepsilon(\Sigma I_m^n)^{-1} J\big(G, b(c_k)\big) I_m^n\big)^{-1}$$

To find $w(c_k)$, we apply the algorithm:

$$w_0 = \begin{cases} w(c_{k-1}) if k > 0 \\ v if k = 0 \end{cases}$$

$$w_{i+1} = v - \varepsilon w_i (\Sigma I_m^n)^{-1} J\big(G, b(c)\big) I_m^n$$

while($\|w_{i+1} - w_i\|_{l_1} > 2\delta$)

For convergence, it is sufficient that:

$$T_2(\alpha) = v - \varepsilon\alpha(\Sigma I_m^n)^{-1} J\big(G, b(c)\big) I_m^n$$

was a non-stretching operator.

## V. Acknowledgments

The work was carried out with the financial support of the National Academy of Sciences of Ukraine (project "Нові субградієнтні та екстраградієнтні методи для негладких задач регресії", 0124U002162).

## References

[1] Candès, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52(2), 489-509.

[2] Wang, Y., Hua, Y., Candes, E. J., & Pilanci, M. (2025). Overparameterized relu neural networks learn the simplest model: Neural isometry and phase transitions. IEEE Transactions on Information Theory.

[3] Parvaresh, F., Vikalo, H., Misra, S., & Hassibi, B. (2008). Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays. IEEE Journal of Selected Topics in Signal Processing, 2(3), 275-285.

[4] Yi, J., Jiang, H., Wang, X., & Tan, Y. (2024). A comprehensive review on sparse representation and compressed perception in optical image reconstruction. Archives of Computational Methods in Engineering, 31(5), 3197-3209.

[5] Marques, E. C., Maciel, N., Naviner, L., Cai, H., & Yang, J. (2018). A review of sparse recovery algorithms. IEEE access, 7, 1300-1322.

[6] Ji, S., Xue, Y., & Carin, L. (2008). Bayesian compressive sensing. IEEE Transactions on signal processing, 56(6), 2346-2356.

[7] Gorodnitsky, I. F., & Rao, B. D. (2002). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Transactions on signal processing, 45(3), 600-616.

[8] Daubechies, I., DeVore, R., Fornasier, M., & Güntürk, C. S. (2010). Iteratively reweighted least squares minimization for sparse recovery. Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 63(1), 1-38.

[9] Fu, H., Liu, H., Han, B., Yang, Y., & Hu, Y. (2017). A proximal iteratively regularized Gauss–Newton method for nonlinear inverse problems. Journal of Inverse and Ill-posed Problems, 25(3), 341-356.

[10] Zhong, M., Wang, W., & Tong, S. (2022). An asymptotical regularization with convex constraints for inverse problems. Inverse Problems, 38(4), 045007.

[11] Wang, J., Han, B., & Wang, W. (2019). Elastic-net regularization for nonlinear electrical impedance tomography with a splitting approach. Applicable Analysis, 98(12), 2201-2217.