Why Sparse?

Thongchai Dumrongpokaphan, Olga Kosheleva, Vladik Kreinovich, and Aleksandra Belina

Abstract In many situations, a solution to a practical problem is *sparse*, i.e., corresponds to the case when most of the parameters describing the solution are zeros, and only a few attain non-zero values. This surprising empirical phenomenon helps solve the corresponding problems – but it remains unclear why this phenomenon happens. In this paper, we provide a possible theoretical explanation for this mysterious phenomenon.

1 Formulation of the Problem

Need to reconstruct a function. In many practical situations, we are interested in a function: e.g., we want to reconstruct a signal $s(t)$ based on the noisy measurements, or we want to reconstruct the original image $I(x,y)$ from the observed noisy one.

General functions can be described via an appropriate basis. Many algorithms for determining a function are based on the fact that every function – under certain restrictions like continuity – can be represented as an infinite sum

$$f(x) = \sum_{i=1}^{\infty} c_i \cdot e_i(x),$$

Thongchai Dumrongpokaphan
Department of Mathematics, Faculty of Science, Chiang Mai University, Thailand
e-mail: tcd43@hotmail.com

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, El Paso, TX 79968, USA
e-mail: olgak@utep.edu, vladik@utep.edu

Aleksandra Belina
Department of Building Structures, Silesian University of Technology, Gliwice, Poland
e-mail: aleksandra.belina@polsl.pl

where:

- the functions $e_1(x)$, $e_2(x)$, ..., are fixed (the set of these functions is known as a *basis*), and
- different functions $f(x)$ correspond to different values of the coefficients c_1, c_2, ...

For example:

- smooth functions can be represented by Taylor series, with $e_1(x) = 1$, $e_2(x) = x$, $e_3(x) = x^2$, ...,
- general functions on a given interval can be represented as Fourier series, with $e_1(x) = \sin(\omega \cdot x)$, $e_2(x) = \cos(\omega \cdot x)$, $e_3(x) = \sin(2\omega \cdot x)$, $e_4(x) = \cos(2\omega \cdot x)$, ...

In all these cases, the fact that the function is a limit of a convergent sum means that the size of the terms $c_n \cdot e_n(x)$ tends to 0 as n increases.

In practice, it is sufficient to determine a finite number of coefficients. To represent arbitrary functions exactly, we need infinitely many coefficients. However, in most practical problems, it is sufficient to represent the functions with some accuracy. For such a representation, we can safely ignore small terms corresponding to large values n. Thus, in practical problems, it is sufficient to use only a fixed number of terms in the corresponding representation, i.e., to consider approximations of the type

$$f(x) \approx \sum_{i=1}^{k} c_i \cdot e_i(x).$$

Sparsity: a mysterious empirical fact. Somewhat surprisingly, in many practical situations, the desired reconstructed function, in an appropriate basis, is *sparse*, in the sense that most coefficients c_i are equal to 0, and only a few are non-zeros.

This sparsity helps design more efficient algorithms for reconstructing the desired function (see, e.g., [2, 3, 4, 5, 6, 8, 9, 10, 11, 15, 16, 17, 20, 21]), but why this happens in the real world remains largely a mystery. To the best of our knowledge, the only theoretical explanation so far is an explanation based on formalizing an intuitive idea that all values be small [7]. The problem with this explanation is that it is somewhat subjective. It is desirable to have an objective – i.e., expert-independent – explanation.

What we do in this paper. In this paper, we provide a possible objective theoretical explanation for this mysterious empirical phenomenon.

2 Main Idea

Informal reformulation of the problem. Measurement uncertainty means that, based on the measurement results, we cannot uniquely determine the desired function $f(x)$. In other words, there exist several different functions which are all con-

sistent with all the measurement results. Out of all these functions, we would like to select (prefer) one which is, in some reasonable sense, the most appropriate.

How to formalize this description. According to decision theory (see, e.g., [12, 14, 13, 18, 19]), preferences of a rational decision maker (for whom preferences are transitive and antisymmetric) can be described by a real-valued function called *utility*, so that:

- between several alternative,
- the decision maker always selects the one which has the largest value of the utility.

Thus, to describe user's preferences, we need to know his/her utility function.

In our case, different alternatives are different functions $f(x)$, i.e., equivalently, different values of the coefficients $c_1,\ldots,c_k$. Thus, to describe the user's preferences, we need to know how the user's utility u depends on the values $c_1,\ldots,c_k$, i.e., we need to know the dependence $u(c_1,\ldots,c_k)$.

Let us analyze what are the reasonable properties of this dependence.

First reasonable property: coefficients c_i are independent. In most practical situations, coefficients c_i are independent in the following sense: for each of these coefficients, there are some preferred values, so that if we have two tuples with the same values of all other coefficients and different values $c_i \neq c_i'$, then, if select c_i in one such case, we should select c_i and not c_i' in all such cases, irrespective of what are the other values $c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_k$.

For example, for quadratic Taylor series $f(x) = c_1 + c_2\cdot x + c_3\cdot x^2$, if we consider a linear dependence more reasonable, then between two functions differing only by their coefficients $c_3 \neq c_3'$, we should select a one for which the value of $|c_3|$ is the smallest – irrespective of the values $c_1 = c_1'$ and $c_2 = c_2'$.

Similarly, for Fourier series, if we believe that nonlinear effects – leading to double frequencies – are small, then between the two functions differing only by the double frequency terms c_3 and c_4, we should prefer functions for which these coefficients are smaller – irrespective of the values of c_1 and c_2.

It is known (see, e.g., [14]) that under this independence assumption, the utility function has either the form $u(c_1,\ldots,c_k) = \sum\limits_{i=1}^{k} u_i(c_i)$ or the form $u(c_1,\ldots,c_k) = \prod\limits_{i=1}^{k} U_i(c_i)$ for some functions $u_i(c_i)$ or $U_i(c_i)$.

Maximizing the product $\prod\limits_{i=1}^{k} U_i(c_i)$ is equivalent to maximizing its logarithm $\sum\limits_{i=1}^{k} u_i(c_i)$, where we denoted $u_i(c_i) \stackrel{\text{def}}{=} \ln(U_i(c_i))$. Thus, without losing generality, we can assume that we select alternatives for which the sum

$$\sum_{i=1}^{k} u_i(c_i) \tag{1}$$

attains the smallest possible value – among all the combinations $(c_1,\dots,c_k)$ for which the function $f(x)=\sum\limits_{i=1}^{k} c_i\cdot e_i(x)$ is consistent with all the measurement results.

Thus, we arrive at the following definition.

Definition 1.

- *By a* criterion *for selecting coefficients, we mean a tuple*

$$u=(u_1(c_1),\dots,u_k(c_k))$$

of k smooth functions $u_i(c_i)$, $1\le i\le k$.
- *Let u be a criterion for selecting coefficients. We say that a tuple $c=(c_1,\dots,c_k)$ is u-*better *than a tuple $c'=(c'_1,\dots,c'_k)$ (and denote it $c>c'$) if*

$$\sum_{i=1}^{k} u_i(c_i)>\sum_{i=1}^{k} u_i(c'_i).$$

- *We say that a tuple $c=(c_1,\dots,c_k)$ is* of the same *u-*quality *as a tuple $c'=(c'_1,\dots,c'_k)$ (and denote it $c\equiv c'$) if $\sum\limits_{i=1}^{k} u_i(c_i)=\sum\limits_{i=1}^{k} u_i(c'_i)$.*

Second reasonable property: scale-invariance. Numerical values of a physical quantity depend on our choice of a measurement unit, and this choice is rather arbitrary. For example, if we originally measured the signal in Volts, and then decided to switch to milliVolts, the signal remains the same but all its numerical values $s(t)$ gets multiplied by a 1000: $s(t)\to 1000\cdot s(t)$. In general, if we change the original measuring unit to a new one which is λ times smaller, all the values of the corresponding function $f(x)$ get multiplies by λ: $f(x)\cdot f_1(x)=\lambda\cdot f(x)$.

From the fact that $f(x)=\sum\limits_{i=1}^{k} c_i\cdot e_i(x)$, we conclude that

$$f_1(x)=\lambda\cdot f(x)=\lambda\cdot\left(\sum_{i=1}^{k} c_i\cdot e_i(x)\right)=\sum_{i=1}^{k}(\lambda\cdot c_i)\cdot e_i(x).$$

Thus, in terms of the coefficients c_i, multiplying all the values $f(x)$ by a constant λ is equivalent to multiplying all the coefficients c_i by the same coefficient λ:

$$c_i\to c'_i=\lambda\cdot c_i.$$

It is reasonable to require that the relative quality of two different functions – i.e., equivalently, of two different tuples $(c_1,\dots,c_k)$ – should not change if we simply multiply all the coefficients by the same positive number λ.

Up to now, we only consider functions $f(x)$ – like images – which are described by non-negative functions.

In some situations – e.g., if we process signals – the values $f(x)$ can be both positive and negative. The selection of the sign is usually also arbitrary: e.g.:

- we consider the current positive if all electrons move in one direction, but
- we could as well call this direction negative.

So, it is reasonable to require that nothing should change if we simply change the sign of all the values $f(x)$ – or, equivalently, that we change the signs of all the coefficients c_i.

Together with invariance with respect to multiplying by any positive number, we can now conclude that the user's preference is invariant with respect to multiplying by any real number.

Thus, we arrive at the following definition.

Definition 2. *We say that a criterion u is* scale-invariant *if for every $\lambda \neq 0$, the following two conditions are satisfied:*

- *if a tuple $c = (c_1, \ldots, c_k)$ is u-better than a tuple $c' = (c'_1, \ldots, c'_k)$, then the tuple $\lambda \cdot c \stackrel{\text{def}}{=} (\lambda \cdot c_1, \ldots, \lambda \cdot c_k)$ is u-better than $\lambda \cdot c' = (\lambda \cdot c'_1, \ldots, \lambda \cdot c'_k)$;*
- *if a tuple $c = (c_1, \ldots, c_k)$ is of the same u-quality as a tuple $c' = (c'_1, \ldots, c'_k)$, then the tuple $\lambda \cdot c \stackrel{\text{def}}{=} (\lambda \cdot c_1, \ldots, \lambda \cdot c_k)$ has the same u-quality as the tuple*

$$\lambda \cdot c' = (\lambda \cdot c'_1, \ldots, \lambda \cdot c'_k).$$

3 Main Result: Formulation and Discussion

Proposition. *Every scale-invariant criterion is equivalent to optimizing the sum $\sum\limits_{i=1}^{k} a_i \cdot |c_i|^p$ for some constants $p, a_1, \ldots, a_k$.*

Discussion. By replacing c_i with $c'_i = |a_i|^{1/p} \cdot c_i$ and $e_i(x)$ with $e'_i(x) = e_i \cdot |a_i|^{-1/p}$, we conclude that the optimized sum has a simplified form $\sum\limits_{i=1}^{k} |c'_i|^p$, where $f(x) = \sum\limits_{i=1}^{k} c'_i \cdot e'_i(x)$.

So, in general, we optimize the sum of the p-th powers:

- for $p = 2$, we get the usual least squares method of minimizing $\sum c_i^2$;
- for $p = 1$, we get a robust ℓ^1-*method* of minimizing the sum $\sum |c_i|$;
- for $p \to \infty$, since optimizing $\sum |c_i|^p$ is equivalent to maximizing $\|c\|_p \stackrel{\text{def}}{=} (\sum |c_i|^p)^{1/p}$, we minimize the limit $\lim\limits_{p \to \infty} \|c\|_p = \max |c_i|$, i.e., we minimize the largest coefficient;
- finally, when $p \to 0$, $|c_i| \to |c_i|^0 = 1$ when $c_i \neq 0$ and $|c_i|^p = 0 \to 0$ if $c_i = 0$; thus, when p tends to 0, the sum $\sum |c_i|^p$ tends to the number of non-zero coefficients c_i.

In the last case, minimizing the sum becomes minimizing the number of non-zero elements – which is exactly what sparsity is about.

Thus, *we have the desired explanation of why sparsity naturally appears in many practical problems.*

4 Proof

1°. If we subtract the same constant from all the values of the objective function, the relative quality of different tuples does not change. In particular, if instead of the original functions $u_i(c)$, we consider new functions $\widetilde{u}_i(c) \stackrel{\text{def}}{=} u_i(c) - u_i(0)$ for which $u_i'(0) = 0$, the new sum $\sum\limits_i \widetilde{u}_i(c_i)$ differs from the old sum by a constant $\sum\limits_i u_i(0)$.

Thus, without losing generality, we can safely assume that $u_i(0) = 0$ for all i.

2°. Let us first prove that the functions $u_i(c_i)$ do not change value is we simply change the sign of the coefficient, i.e., that $u_i(-c_i) = u_i(c_i)$ for all c_i.

Indeed, let us consider the tuple $c = (0, \ldots, 0, c_i, 0, \ldots, 0)$ in which only the i-th element is different from 0.

If c is better than $-c$, i.e., if $c > -c$, then, due to invariance under multiplying by -1, we conclude that $-c > c$, i.e., that $-c$ is better than c – a contradiction.

Similarly, if $-c$ is better than c, i.e., if $-c > c$, then, due to invariance under multiplying by -1, we conclude that $c > -c$, i.e., that c is better than $-c$: also a contradiction.

The only remaining case is $c \equiv -c$, which means that

$$u(c) = \sum_j u_j(c_j) = u_i(c_i) = u(-c) = \sum_j u_j(-c_j) = u_i(-c_i).$$

Thus, we have $u_i(-c_i) = u_i(c_i)$ for all i and c_i, i.e., equivalently, $u_i(c_i) = u_i(|c_i|)$. So, it is sufficient to determine the values of the functions $u_i(c_i)$ for positive values $c_i > 0$.

2°. Let us now consider the case when two values c_i and c_j differ from 0, and all others are equal to 0. For such tuples, the objective function has the form

$$u_i(c_i) + u_j(c_j).$$

For such functions, scale-invariance means, in particular, that if

$$u_i(c_i) + u_j(c_j) = u_i(c_i') + u_j(c_j'),$$

then for every $\lambda > 0$, we have

$$u_i(\lambda \cdot c_i) + u_j(\lambda \cdot c_j) = u_i(\lambda \cdot c_i') + u_j(\lambda \cdot c_j').$$

3°. Let us consider the case when:

- c_i' is close to c_i, i.e., when $c_i' = c_i + \Delta c$ for a small value Δc, and
- c_j' is close to c_j, i.e., $c_j' = c_j + k \cdot \Delta c + o(\Delta c)$ for an appropriate k.

Substituting these values c_i' and c_j' into the above equality, we get

$$u_i(c_i) + u_j(c_j) = u_i(c_i + \Delta c) + u_j(c_j + k \cdot \Delta c).$$

Here,

$$u_i(c_i + \Delta c) = u_i(c_i) + u_i'(c_i) \cdot \Delta c + o(\Delta c),$$

where f', as usual, denotes the derivative of a function f.

Similarly,

$$u_j(c_j + k \cdot \Delta c) = u_j(c_j) + u_j'(c_j) \cdot k \cdot \Delta c + o(\Delta c),$$

so the above equality implies that

$$u_i'(c_i) \cdot \Delta c + u_j'(c_j) \cdot k \cdot \Delta c + o(\Delta c) = 0.$$

Diving both sides by Δc and taking $\Delta c \to 0$, we get

$$u_i'(c_i) + u_j'(c_j) \cdot k = 0,$$

hence

$$k = -\frac{u_i'(c_i)}{u_j'(c_j)}.$$

The condition

$$u_i(\lambda \cdot c_i) + u_j(\lambda \cdot c_j) = u_i(\lambda \cdot c_i') + u_j(\lambda \cdot c' j2)$$

similarly takes the form

$$u_i'(\lambda \cdot c_i) + u_j'(\lambda \cdot c_j) \cdot k = 0,$$

i.e.,

$$u_i'(\lambda \cdot c_i) - u_j'(\lambda \cdot c_j) \cdot \frac{u_i'(c_i)}{u_j'(c_j)} = 0.$$

Thus,

$$u_i'(\lambda \cdot c_i) = u_j'(\lambda \cdot c_j) \cdot \frac{u_i'(c_i)}{u_j'(c_j)}.$$

By moving all the terms related to c_1 to the left-hand side and all other terms to the right-hand side, we get

$$\frac{u_i'(\lambda \cdot c_i)}{u_i'(c_i)} = \frac{u_j'(\lambda \cdot c_j)}{u_j'(c_j)}$$

for all λ, c_i, and c_j.

This means that the ratio $\frac{u_i'(\lambda\cdot c_i)}{u_i'(c_i)} = \frac{u_j'(\lambda\cdot c_j)}{u_j'(c_j)}$ does not depend on c_i or c_j, it only depends on λ:

$$\frac{u_i'(\lambda\cdot c_i)}{u_i'(c_i)} = F(\lambda)$$

for some function $F(\lambda)$.

For $\lambda = \lambda_1\cdot\lambda_2$, we have

$$F(\lambda) = \frac{u_i'(\lambda\cdot c_i)}{u_i'(c_i)} = \frac{u_i'(\lambda_1\cdot\lambda_2\cdot c_i)}{u_i'(c_i)} =$$

$$\frac{u_i'(\lambda_1\cdot(\lambda_2\cdot c_i))}{u_i'(\lambda_2\cdot c_i)}\cdot\frac{u_i'(\lambda_2\cdot c_i)}{u_i'(c_i)} = F(\lambda_1)\cdot F(\lambda_2),$$

i.e.,

$$F(\lambda_1\cdot\lambda_2) = F(\lambda_1)\cdot F(\lambda_2).$$

It is known (see, e.g., [1]) that every continuous function satisfying this property has the form $F(\lambda) = \lambda^q$ for some real number q.

The condition $\frac{u_i'(\lambda\cdot c_i)}{u_i'(c_i)} = F(\lambda)$ now takes the form

$$u_i'(\lambda\cdot c_i) = u_i'(c_i)\cdot F(\lambda) = u_i'(c_i)\cdot\lambda^p.$$

In particular, for $c_i = 1$, we get

$$u_i'(\lambda) = A_i\cdot\lambda^q,$$

where $A_i \stackrel{\text{def}}{=} u_i'(1)$. In other words, $u_i'(c_i) = A_i\cdot c_i^q$.

We have an expression for the derivative $u_i'(c_i)$ of the desired function $u_i(c_i)$. To get $u_i(c_i)$, we therefore need to integrate this derivative. For this integration, we have two different formulas: for $q = -1$ and for all other q.

Let us show that the value $q = -1$ is impossible. Indeed, if $q = -1$, we get $u_i(c_i) = A_i\cdot\ln(c_i) + \text{const}$, which contradicts to the above requirement that $u_i(0) = 0$.

Thus, we have $q \neq -1$. Therefore, integration leads to

$$u_i(c_i) = \frac{A_i}{q+1}\cdot c_i^{q+1} + \text{const}.$$

The condition $u_i(0) = 0$ now implies that $u_i(c_i) = \frac{A_i}{q+1}\cdot c_i^{q+1}$ for $c_i \geq 0$.

Since, according to Part 2 of this proof, we have $u_i(c_i) = u_i(|c_i|)$, we thus get $u_i(c_i) = \frac{A_i}{q+1}\cdot|c_i|^{q+1}$ for all c_i. Therefore,

$$u(c) = \sum_{i=1}^{k} u_i(c_i) = \sum_{i=1}^{k} \frac{A_i}{q+1} \cdot |c_i|^{q+1}.$$

This is exactly the desired form, with $a_i = \dfrac{A_i}{q+1}$ and $p = q+1$. The proposition is proven.

Acknowledgments

This work is supported by Chiang Mai University, Thailand, and by the US National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

1. J. Aczel, *Lectures on Functional Equations and Their Applications*, Dover, New York, 2006.
2. B. Amizic, L. Spinoulas, R. Molina, and A. K. Katsaggelos, "Compressive blind image deconvolution", *IEEE Transactions on Image Processing*, 2013, Vol. 22, No. 10, pp. 3994–4006.
3. E. J. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Comm. Pure Appl. Math.*, 2006, Vol. 59, pp. 1207–1223.
4. E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 2, pp. 489–509.
5. E. J. Candès and T. Tao, "Decoding by linear programming", *IEEE Transactions on Information Theory*, 2005, Vol. 51, No. 12, pp. 4203–4215.
6. E. J. Candès and M. B. Wakin, "An Introduction to compressive sampling", *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 21–30.
7. F. Cervantes, B. Usevitch, L. Valera, and V. Kreinovich, "Why sparse? fuzzy techniques explain empirical efficiency of sparsity-based data- and image-processing algorithms", *Proceedings of the 2016 World Conference on Soft Computing*, Berkeley, California, May 22–25, 2016, pp. 165–169.
8. D. L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2005, Vol. 52, No. 4, pp. 1289–1306.
9. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 83–91.
10. T. Edeler, K. Ohliger, S. Hussmann, and A. Mertins, "Super-resolution model for a compressed-sensing measurement setup", *IEEE Transactions on Instrumentation and Measurement*, 2012, Vol. 61, No. 5, pp. 1140–1148.
11. M. Elad, *Sparse and Redundant Representations*, Springer Verlag, 2010.
12. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
13. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
14. P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
15. J. Ma and F.-X. Le Dimet, "Deblurring from highly incomplete measurements for remote sensing", *IEEE Transactions on Geosciences Remote Sensing*, 2009, Vol. 47, No. 3, pp. 792–802.

16. L. McMackin, M. A. Herman, B. Chatterjee, and M. Weldon, "A high-resolution swir camera via compressed sensing", *Proceedings of SPIE*, 2012, Vol. 8353, No. 1, p. 8353-03.
17. B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 1995, Vol. 24, pp. 227–234.
18. H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrows 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
19. H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.
20. Y. Tsaig and D. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 4, pp. 1289–1306.
21. L. Xiao, J. Shao, L. Huang, and Z. Wei, "Compounded regularization and fast algorithm for compressive sensing deconvolution", *Proceedings of the 6th International Conference on Image Graphics*, 2011, pp. 616–621.