Xiang-Gen Xia

Undesired Cross Terms

This lecture note briefly describes two examples in mathematics on how and why cross terms are disliked: orthogonal space-time block code (OSTBC) from orthogonal designs and free probability theory. The first example shows that the elimination of cross terms can significantly simplify the optimization that has many practical applications in, for instance, multiple antenna systems in wireless communications. The second example gives an intuitive and/or fundamental understanding of free probability theory, which has important applications in random matrix theory.

In signal processing, optimization is always an important—if not the most important—task that has applications in nearly all areas. Current active applications include data analytics, artificial intelligence, and wireless communications. Any methods that help reduce the complexity in optimization are critically important. The elimination of cross terms is a way to simplify the optimization and therefore, it has critical applications in many areas, including signal processing.

Large-size random matrices play an important role in not only data analytics but also current and future wireless communications, known as *massive multiple-input, multiple-output* (*MIMO*) *systems*. Free probability theory is an important tool used to study large-size random matrices. As explained later in this article, the viewpoint of vanishing cross terms provides an intuitive and/or fundamental understanding of free probability theory.

Digital Object Identifier 10.1109/MSP.2020.3018085
Date of current version: 28 October 2020

Prerequisites

This lecture note requires some basic knowledge of probability theory, such as moments, distribution functions, and probability density functions. It also requires some basic understanding of algebra and digital communications theory.

Problem statement and solution

Crossings have made this world rich and nonsimple. People often hear the word *multidisciplinary*, in particular, in the research community, which means that different disciplines come together to share their ideas to spark new ideas. Today, multidisciplinary in most, if not all, research funding agencies means creative and big. Not many people have the ongoing patience for small items or small ideas anymore. These kinds of crossings are for creative ideas.

It is known that the crossings of the two simplest digits, 0 and 1, have created all the digital products in use today, including computers, cellphones, and so forth. This is now making people all over the world crazy about creating everything by machine or crossings between 0 and 1, called *artificial intelligence*. These kinds of crossings are for creative products.

Although crossings bring many joys and surprises, mathematically, mathematicians do not like crossings much and sometimes even dislike them. This is mainly because crossings may also bring a lot of inconvenience and annoyance to mathematicians. Crossings cause cross terms as follows. Let x_1 and x_2 be two variables. Then, some of their cross terms are $p_1(x_1)p_2(x_2)$, where $p_1(x_1)$ and $p_2(x_2)$ are two nonconstant polynomials of x_1 and x_2, respectively, while a general cross term of these two variables is a multivariate polynomial of the two variables. The simplest cross term of x_1 and x_2 is x_1x_2. When the number of variables is not two, their cross terms become more complicated. For example, for three variables, x_1, x_2, and x_3, one of their simplest cross terms is $x_1x_2 + x_2x_3 + x_1x_3$.

The nth power of the sum of several variables can be expanded using the cross terms of these variables, but its expansion becomes tedious when n is not small. However, if all of the cross terms of the variables are gone, that is, the variables do not cross, then the nth power of their sum would be the sum of their nth powers: $(x_1 + \cdots + x_p)^n = x_1^n + \cdots + x_p^n$.

 1053-5888/20©2020IEEE

What a simple formula this would be!

The aforementioned formula is, if correct, not only simple for the expansion, but also very useful in many applications. For example, say you want to optimize an objective function with the p variables mentioned in the previous section. If you can be sure that the p variables in this objective function have no cross terms, that is, you can decompose this objective function into a sum of p independent subobjective functions, then the joint optimization of the objective function of the p variables can be converted into separate optimizations of the p subobjective functions, and each of the p subobjective functions has only one variable. In this way, the joint optimization of p variables becomes p many individual optimization problems of a single variable. This greatly reduces the optimization complexity and will change an infeasible solution to a feasible solution in many cases.

Orthogonal space-time block codes from orthogonal designs

An example pertains to the topic of multiantenna space-time modulation and demodulation in wireless communications around the year 2000. If m transmit antennas transmit p variables (or symbols), these p variables must be solved at the receiver. In general, their optimal solution is to jointly search for the values of the p variables. When p is not small, the complexity may be prohibitively high. However, if an appropriate space-time coding is performed at the transmitter, then at the receiver, the cross terms of the p variables can be eliminated, that is, there is no cross term. At this time, as explained previously, their demodulation is equivalent to the demodulation of each of the p variables separately, and the demodulation complexity is thus only linear in terms of the number p of the variables.

The first such design was given by Siavash Alamouti at AT&T for two transmit antennas in 1998. It is named the *Alamouti code* in the research community. In the case of an arbitrary number of transmit antennas, such a code is called an OSTBC [1], which means that all of the variables will have no cross terms at the receiver after some simple operations without losing any information. It was later discovered [1] that the design of OSTBC, in a general case, can be traced back to the oldest mathematics—namely, numbers—and was a core problem in mathematics. They are equivalent to the generalizations of the complex number field, called *number domains*, and further generalizations are norm identities and compositions of quadratic forms [2].

For example, for the Alamouti code of two transmit antennas, it sends two symbols with carried information in two time slots. When both symbols are real valued, the Alamouti code is equivalent to a complex number. However, in wireless communications, it is not energy efficient to transmit only real-valued symbols. When the two symbols are complex valued, the Alamouti code is equivalent to a quaternion number. Along this direction, we obtained some results in the 2000s on general orthogonal designs for complex variables of an arbitrary size, which include some rate upper bounds and an inductive design for all the numbers of transmit antennas. Interested readers are referred to my homepage [3] on the University of Delaware's website.

Free probability theory

Another example where mathematicians do not like cross terms is the free probability theory that appeared in the 1980s. The theory of probability, as we all know, is for real variables, and it can be generalized to complex variables. The multiplication of these variables is commutative, for example, $x_1x_2x_1x_2 = x_1^2x_2^2$. If we take its expectation, we have $E(x_1x_2x_1x_2) = E(x_1^2x_2^2) \geq 0$, which is not equal to 0, unless x_1x_2 is 0. This means that the cross terms of x_1 and x_2 exist, even after taking the expectation. In the 1980s, Dan-Virgil Voiculescu invented the theory of free probability to study a von Neumann conjecture of 1967. The random variables he studied were not real nor complex valued, but their multiplications are not commutative, such as random matrices. In fact, abstract noncommutative elements can often be represented by matrices in mathematics. Then, the troubling problem is how to define distribution functions for these random variables. One knows that random variables in the conventional probability theory often have moments that are based on the expectation E, and the moments of a random variable form an infinite series. In many cases, this series of moments is equivalent to the characteristic function of the random variable, and the characteristic function and the distribution function of a random variable are Fourier transform pairs. Therefore, in many cases, the moment series of a random variable uniquely determines the distribution function of the random variable.

Note that the relationship between a series of moments and a characteristic function is similar to the Laurent expansion of a complex-variable function. When a moment series uniquely determines a characteristic function is related to when a complex-valued function exists for Laurent expansion, which, in turn, is related to complex analysis. From this observation, Voiculescu used moments to describe noncommutative random variables, which leads to free probability theory, where expectation E is replaced by a linear functional E on a set (or an algebra) of noncommutative elements (or random variables) with $E(e) = 1$, with e as the identity element.

The most important theorem in undergraduate probability theory is the central limit theorem. It is a theorem about where the sum of independent random variables goes to when the number of random variables in the sum goes to infinity. The theorem says that, as long as these independent random variables are properly normalized, their sum converges to a Gaussian random variable. In free probability theory, Voiculescu also cared about the limit of the sum of multiple noncommutative random variables. However, for noncommutative random variables, we only deal with their moments, not the traditional distribution functions. For multiple random variables, the high-order moments of their sum are the expectations of high-order multivariate polynomials of the random variables. After the expansion, it is very complicated and difficult to continue any fruitful investigation, particularly when these random variables are not commutative. To deal with this problem, Voiculescu observed that, if the cross terms of these multiple random variables disappear or most of them disappear after the expectation, the summation would become much simpler to study, similar to what was discussed in the previous section. Therefore, Voiculescu defines that if these random variables have such properties, i.e., the expectations of their cross terms or most of their cross terms are 0, when the expectations of these random variables by themselves are 0, they are called *free random variables*, or *free*.

Freeness definition

Random variables $x_1, \cdots, x_p$ are called *free random variables* or *free*, if for any m polynomials $p_1(x), \cdots, p_m(x), m \geq 2$,

$$E(p_1(x_{i_1})p_2(x_{i_2})\cdots p_m(x_{i_m})) = 0$$

when $E(p_k(x_{i_k})) = 0$ for all $k, 1 \leq k \leq m$, and any two neighboring indices i_l and i_{l+1} are not equal, i.e., $1 \leq i_1 \neq i_2 \neq \cdots \neq i_m \leq p$.

For example, random variables x_1 and x_2 are free random variables (or free), one condition is as follows. By taking $p_1(x) = p_2(x) = p_3(x) = p_4(x) = x$ and $m = 4$ in the aforementioned definition, $E(x_1x_2x_1x_2) = 0$ when $E(x_1) = E(x_2) = 0$, where $i_1 = i_3 = 1$, $i_2 = i_4 = 2$. This is different from the conventional commutative random variables mentioned earlier. Two nonzero, independent conventional real-valued random variables x_1 and x_2 of 0 mean can never be free because $E(x_1x_2x_1x_2) = E(x_1^2x_2^2) = E(x_1^2)E(x_1^2) > 0$.

From this freeness definition of random variables, one can see that Voiculescu wanted to make as many cross terms as possible disappear, while one can also see that not all cross terms disappear [4]. For more details about free random variables, we refer the reader to the work of Mingo and Speicher [4] or to my simplified introductory article [5].

In fact, this freeness of noncommutative random variables corresponds to (but are not identical to, as seen earlier) the independence of commutative random variables. Consider the likelihood function $\log f(x_1, \cdots, x_p)$ of p conventional random variables $x_1, \cdots, x_p$, with their individual probability density functions (pdfs) $f_i(x_i)$ and joint pdf $f(x_1, \cdots, x_p)$. Then, these p random variables are independent if and only if $\log f(x_1, \cdots, x_p) = \log f_1(x_1) + \cdots + \log f_p(x_p)$.

This means that the cross terms of the random variables also disappear in their joint likelihood function. Let us also use moments to see the conventional independence of random variables. For any m polynomials, $p_1(x), \cdots, p_m(x), 2 \leq m \leq p$, such that $E(p_i(x_i)) = 0$, then, if $x_1 \cdots, x_p$ are independent, we have

$$\begin{aligned} &E(p_1(x_1)p_2(x_2)\cdots p_m(x_m)) \\ &= E(p_1(x_1))E(p_2(x_2))\cdots E(p_m(x_m)) \\ &= 0 \end{aligned} \tag{1}$$

which implies that the cross terms vanish. Thus, from the vanishing cross-term viewpoint, the freeness and the independence of random variables correspond to each other. Note that a key difference between the independence of commutative random variables and the freeness of noncommutative random variables exists in the fact that for the conventional commutative random variables, their product, as a crossing of these variables, can be sorted into a product of single-variable polynomials as shown in (1), which does not apply to noncommutative random variables.

Using the aforementioned definition, Voiculescu also came up with a theorem called the *free central limit theorem* for free random variables similar to the central limit theorem for independent random variables. He said that the sum of multiple free random variables after a proper normalization converges to a random variable with semicircle distribution. This semicircle distribution is similar to the Gaussian distribution in the conventional probability theory. In fact, Eugene Wigner first found that the distribution of eigenvalues of random matrices with independent and identically distributed Gaussian components goes to a random variable with semicircle distribution when the dimension of the matrices tends to infinity—this is called *Wigner's semicircle law*.

Voiculescu's free probability theory was naturally applied to large-size random matrices and has become an active research topic in recent years. As the size/dimension of matrices becomes larger and larger, if all the elements in the matrices are independent of each other and have 0 mean, then their cross terms are getting closer to 0 mean. So, the expectations of the cross terms of the matrices are getting closer to 0 as well. In other words, large-size random matrices are approximately free. With this in mind, free probability theory can be applied to study large-size random matrices with applications in, for example, big data and massive MIMO systems. One such application is the asymptotic eigenvalue analysis of large-size random matrices, which may be used to determine the performance limit of a massive MIMO communications system. For more details, see [4] and [5].

What we have learned

In this note, it was briefly explained that cross terms are sometimes troublesome for mathematicians. Although only two examples were mentioned, namely, orthogonal designs and free probability theory, where mathematicians do not like cross terms, there should be many other examples. After saying so, cross terms exist more often, as in the expansion of the nth power of a sum of multiple variables, as mentioned in the beginning of this note, or for nonindependent conventional random variables. No matter whether mathematicians (or we) like them or not, we have to deal with them.

Author

Xiang-Gen Xia (xxia@ee.udel.edu) received his Ph.D. degree in electrical engineering from the University of Southern California in 1992 and is currently the Charles Black Evans Professor of Electrical and Computer Engineering at the University of Delaware. He is the author of *Modulated Coding for*

Intersymbol Interference Channels. He received the National Science Foundation CAREER Award in 1997, the Office of Naval Research Young Investigator Award in 1998, and the Information Theory Outstanding Overseas Chinese Scientist Award in 2019. His research interests include multiple-input, multiple-output systems and radar signal processing. He is a Fellow of IEEE.

References

[1] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999. doi: 10.1109/18.771146.

[2] D. B. Shapiro, *Compositions of Quadratic Forms*. New York: De Gruyter, 2000.

[3] X.-G. Xia, "Mathematics and electrical engineering," Nankai Univ., Tianjin, China, June 28, 2018. [Online]. Available: https://www.eecis.udel.edu/~xxia/Math_EE.pdf

[4] J. A. Mingo and R. Speicher, *Free Probability and Random Matrices*. New York: Springer-Verlag, 2017.

[5] X.-G. Xia, "A simple introduction to free probability theory and its application to random matrices," *Math. Res. Rev.*, vol. 1, no. 2, Art no. 22, pp. 1–22, Nov. 2019. [Online]. Available: https://www.prior-sci-pub.com/mrr2019_iss2_art22.pdf