

8-2010

Why Ellipsoid Constraints, Ellipsoid Clusters, and Riemannian Space-Time: Dvoretzky's Theorem Revisited

Karen Villaverde

Olga Kosheleva

The University of Texas at El Paso, olgak@utep.edu

Martine Ceberio

The University of Texas at El Paso, mceberio@utep.edu

Follow this and additional works at: https://scholarworks.utep.edu/cs_techrep



Part of the [Computer Engineering Commons](#)

Comments:

Technical Report: UTEP-CS-10-19

Recommended Citation

Villaverde, Karen; Kosheleva, Olga; and Ceberio, Martine, "Why Ellipsoid Constraints, Ellipsoid Clusters, and Riemannian Space-Time: Dvoretzky's Theorem Revisited" (2010). *Departmental Technical Reports (CS)*. 19.

https://scholarworks.utep.edu/cs_techrep/19

This Article is brought to you for free and open access by the Computer Science at ScholarWorks@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.

Why Ellipsoid Constraints, Ellipsoid Clusters, and Riemannian Space-Time: Dvoretzky's Theorem Revisited

Karen Villaverde¹, Olga Kosheleva², and Martine Ceberio²

¹ Department of Computer Science, New Mexico State University,
Las Cruces, NM 88003, USA, kvillave@cs.nmsu.edu

² University of Texas at El Paso, El Paso, TX 79968, USA,
{olgak,mceberio}@utep.edu

Abstract. In many practical applications, we encounter ellipsoid constraints, ellipsoid-shaped clusters, etc. A usual justification for this ellipsoid shape comes from the fact that many real-life quantities are normally distributed, and for a multi-variate normal distribution, a natural confidence set (containing the vast majority of the objects) is an ellipsoid. However, ellipsoids appear more frequently than normal distributions (which occur in about half of the cases). In this paper, we provide a new justification for ellipsoids based on a known mathematical result – Dvoretzky's Theorem.

Keywords: ellipsoids; constraints; clusters; tensors; space-time physics; Dvoretzky's theorem

1 Formulation of the Problem

Ellipsoids are ubiquitous. In many practical applications, we encounter ellipsoid constraints, ellipsoid-shaped clusters, etc. (see, e.g., [2]), i.e., sets in an n -dimensional space described by the formula

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^n a_i \cdot x_i \leq a_0. \quad (1)$$

Reformulation in terms of tensors. The above formula (1) shows that to describe an ellipsoid, we need to have a vector (= tensor of order 1) a_i and a tensor a_{ij} of order 2.

A usual probabilistic explanation of the ellipsoid shape. A usual justification for this ellipsoid shape comes from the fact that many real-life quantities are normally distributed, and for a multi-variate normal distribution, a natural confidence set (containing the vast majority of the objects) is an ellipsoid.

Indeed, it is known that uncertainty can be often described by the Gaussian (= normal) distribution, with the probability density

$$\rho(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right). \quad (2)$$

This possibility comes from the Central Limit Theorem (see, e.g., [12]), according to which the sum $x = \sum_{i=1}^N x_i$ of a large number N of independent small random variables x_i has an approximately Gaussian distribution. (To be more precise, the theorem says that in the limit $N \rightarrow \infty$, the distribution of the sum tends to the Gaussian distribution.)

In practice, often, the measurement error is caused by a joint effect of a large number of small independent factors, so it makes sense to conclude that the distribution is approximately Gaussian. This theoretical conclusion has been experimentally confirmed on the example of many actual measuring instruments; see, e.g., [9].

The above result is about the 1-D distribution: of a random number. For the multi-D case – of a random vector $x = (x_1, \dots, x_n)$ – a similar result also leads to multi-D Gaussian distribution, with an expression

$$\rho(x) = \text{const} \cdot \exp\left(-\sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot (x_i - a_i) \cdot (x_j - a_j)\right). \quad (3)$$

This probability density $\rho(x)$ is everywhere positive; thus, in principle, an arbitrary tuple Δx is possible. In practical statistics, however, tuples with very low probability density $\rho(\Delta x)$ are considered impossible. For example, in 1-dimensional case, we have a “three sigma” rule: values for which $|\Delta x_1| > 3\sigma_1$ are considered to be almost impossible. In the multi-dimensional case, it is natural to choose some threshold $t > 0$, and consider only tuples for which $\rho(\Delta x) \geq t$ as possible ones. This formula is equivalent to $\ln(\rho(x)) \geq \ln(t)$. For Gaussian distribution, this equality takes the form $\sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot (x_i - a_i) \cdot (x_j - a_j) \leq -\ln(t)$, i.e., the form of an ellipsoid.

Problem. While the probabilistic explanation is convincing, it does not cover all the cases. Indeed, according to [9], normal distributions occur in approximately half of the cases, while in many practical applications, ellipsoids appear more frequently.

How can we explain this ubiquity of ellipsoids?

Taylor expansion: a possible explanation. Another possible explanation comes from the fact that the function $g(x_1, \dots, x_n)$ describing a general constraint $g(x_1, \dots, x_n) \leq 0$ is usually smooth; thus, it can be usually expanded in Taylor series. In this expansion, terms of higher order become smaller and smaller, so we can usually safely keep only a few first terms in this expansion. In particular,

if we only keep linear and quadratic terms, we get an expression (1) – i.e., an ellipsoid.

This argument is reasonable, but it does not explain why in most cases, the first two terms are sufficient and not, e.g., the first three – which would lead to more complex shapes of constraints and clusters (and the use of tensors of higher order).

Comment. An alternative explanation comes from the fact that ellipsoids are known to be the *optimal* approximation sets for different problems with respect to several reasonable optimality criteria; see, e.g., [5, 6]. However, they are optimal only if we consider approximating families of sets characterized by the smallest possible number of parameters.

2 New Explanation Based on Dvoretzky's Theorem

What is Dvoretzky's Theorem. In this paper, we propose a new explanation of the ubiquity of ellipsoids. This explanation is based on a mathematical result called Dvoretzky's theorem.

The original version of this theorem [3] answered a question raised in 1956 by Alexander Grothendieck, one of the most important mathematicians of the 20 century. A. Dvoretzky proved that Grothendieck's hypothesis is indeed true, and that in general, convex sets in large dimensions have sections whose shape is close to ellipsoidal – the larger the dimension, the close this shape to the shape of an ellipsoid.

In 1971, V. L. Milman [7] strengthened this result by proving that not only *there exists* an almost ellipsoidal shape, but also that *almost all* low-dimensional sections of a convex set have an almost ellipsoidal shape. (Strictly speaking, he proved that for every $\varepsilon > 0$, the probability to get a shape which is more than ε -different from ellipsoidal goes to 0 as the dimension of the convex set increases.)

How Dvoretzky's theorem explains the ubiquity of ellipsoid clusters and ellipsoid constraints. In clustering, one of the main problems is that usually, we only measure a few quantities, not enough to easily classify objects. For example, in military applications, the need to classify sonar records into submarine sounds, whale sounds, and noise comes from the fact that we only have a weak (partially observed) signal. Based on a high-quality low-noise recording, it is relatively easy to distinguish between sounds produced by submarines and sounds produced by whales.

Theoretically, each real-life object can be characterized by a point (vector) containing the results of measuring all possible quantities characterizing this object. In this theoretical description, objects are represented by points in a (very) high-dimensional space, and natural classes of objects are sets in this high-dimensional space.

However, in the real world, we only observe a few of these quantities. Thus, what we observe is a lower-dimensional section of a high-dimensional set – and

we know that, according to Dvoretzky's theorem, this section is almost always almost ellipsoidal.

A similar argument can be made about constraints. The actual physical constraints depend not only on the observed quantities x_1, \dots, x_n , they also depend on other quantities whose values we do not measure in our experiments. For example, to avoid unnecessary side effects, it is usually recommended that the amount x_1 of a medicine that a doctor prescribes to a patient must lie within bounds depending on the patient's body weight x_2 . In other words, we have a constraint of the type $x_1 \leq k \cdot x_2$, where the constant k depends on the specific medicine. However, the actual effect of the medicine depends not only on the body weight, it depends on many other characteristics of a patient – such as physical fitness, general allergic reactions – characteristics that usually, we do not measure. Similarly, in recipes for cooking, the amount of salt x_1 is usually listed depending on the amount of, say, meat x_2 used in the cooking. However, in reality, it should depend also on the parameters that a usual cook does not measure exactly – such as the humidity in the air, etc. (That is why, in contrast to typical US cookbooks that list the exact amounts of all the ingredients, in Mexican, Russian, and French cookbooks these amounts are only approximately listed – so that a skilled cook can take into account other parameters that are difficult to measure :-)

In general, a physical constraint actually has a form $g(x_1, \dots, x_n, x_{n+1}, \dots, x_N) \leq 0$, where x_{n+1}, \dots, x_N are quantities that we do not measure in this particular experiment. Thus, the corresponding n -dimensional constraint set $\{x = (x_1, \dots, x_n) : g(x_1, \dots, x_n) \leq 0\}$ is a *section* of the actual (unknown) multi-dimensional constraint set $\{x = (x_1, \dots, x_n, \dots, x_N) : g(x_1, \dots, x_n, \dots, x_N) \leq 0\}$ – and we already know that in almost all cases, such sections are almost ellipsoidal.

Auxiliary result: why Riemannian space-time? A similar argument can explain why, contrary to physicists' expectations, experiments seem to confirm the Riemannian models of space-time. Before we provide this explanation, let us briefly explain what is the Riemannian model and why physicists expected it to be experimentally disproved.

Before Einstein's General Relativity theory, it was assumed that space is Euclidean, i.e., that, in appropriate coordinates, the distance $d(x, x + \Delta x)$ between two close points can be described as $d^2(x, x + \Delta x) = \sum_{i=1}^n (\Delta x_i)^2$. In general (not necessarily orthonormal) coordinates, this distance takes a more general form $\sum_{i=1}^n \sum_{j=1}^n g_{ij} \cdot \Delta x_i \cdot \Delta x_j$. Einstein suggested that the space-time is *locally* Euclidean, so that in the small vicinity of each point, there are coordinates in which the distance is Euclidean – but there are no coordinates in which the distance formula is Euclidean at all the points [8]. Such spaces are known as *Riemannian*.

Einstein himself experimented with extending his theory from the usual (observed) $(3 + 1)$ -dimensional space-time to space-times of higher dimension [4].

It later turned out that higher dimensions are needed to make quantum field theory consistent; see, e.g., [11].

A local Euclidean metric can be characterized by the fact that in this metric, the unit ball is an ellipsoid. In principle, there are other metrics (e.g., l^p -metric for which $d^p(x, x + \Delta x) = \sum_{i=1}^n |\Delta x_i|^p$) with different convex bodies for unit balls. The corresponding generalization of Riemannian space-time is called a *Finsler space* [1, 10].

One of the main ideas of quantum physics is that in contrast to classical physics, where, e.g., some trajectories are allowed and some are not, in quantum physics, all trajectories are allowed – just the probability of very non-standard probabilities is small. Similarly, metrics should not be limited to Riemannian metrics, Finsler metrics should also be possible – with some probability. However, while experiments confirm non-standard trajectories of quantum particles and non-standard behavior of quantum fields, surprisingly, all experimental data so far confirms Riemannian metric.

Dvoretzky's theorem explains this phenomenon: indeed, the actual space is multi-dimensional, so we only observe a section of the corresponding convex unit ball, and such a section is close to an ellipsoid.

References

1. Chern, S.-S., Shen, Z.: Riemann-Finsler Geometry. World Scientific, Singapore (2005)
2. Chernousko, F.L.: State Estimation for Dynamic Systems. CRC Press, Boca Raton, Florida (1994)
3. Dvoretzky, A.: Some results on convex bodies and Banach spaces, In: Proceedings of the 1960 International Symposium on Linear Spaces, pp. 123–160. Jerusalem, Jerusalem Academic Press, Jerusalem, and Pergamon Press, Oxford (1961)
4. Einstein, A., Bergmann, P.: On the generalization of Kaluza's theory of electricity. Ann. Phys. 39, 683–701 (1938)
5. Finkelstein, A., Koshelova, O., Kreinovich, V.: Astrogeometry, error estimation, and other applications of set-valued analysis. ACM SIGNUM Newsletter 31(4), 3–25 (1996)
6. Li, S., Ogura, Y., Kreinovich, V.: Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables. Kluwer Academic Publishers, Dordrecht (2002)
7. Milman, V.D.: A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies, Functional Analysis and Its Applications 5(4), 28–37 (1971), in Russian
8. Misner, C.W., Thorne, K.S., Wheeler, J.A.: Gravitation. W.H. Freeman, New York (1973)
9. Novitskii, P.V., Zograph, I.A.: Estimating the Measurement Errors. Energoatomizdat, Leningrad (1991) in Russian
10. Pavlov, D.G., Atanasiu, G., Balan, V. (eds.) Space-Time Structure. Algebra and Geometry. Russian Hypercomplex Society, Lilia Print, Moscow (2007)
11. Polchinski, J.: String Theory, Vols. 1, 2, Cambridge University Press (1998)
12. Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. Springer Verlag, New York (2005)