DIMACS Technical Report 2000-19
August 2000

Bounds on packings of spheres in the Grassmann manifolds

by

A. Barg [1] D. Nogin [2]

[1] Bell Labs, Lucent Technologies, 600 Mountain Ave., Rm. 2C-375, Murray Hill, NJ 07974.
[2] IPPI RAN, Bol'shoj Karetnyj 19, Moscow 101447, Russia. Research done while visiting DIMACS, Bell Labs, November 1999. Supported in part by the Russian Foundation for Basic Research, project no. 99-01-01204.

ABSTRACT

We derive the Varshamov–Gilbert and Hamming bounds for packings of spheres (codes) in the Grassmann manifolds over $\mathbb{R}$ and $\mathbb{C}$. The distance between two k-planes is defined as $\rho(p,q) = (\sin^2\theta_1 + \cdots + \sin^2\theta_k)^{1/2}$, where $\theta_i, 1 \leq i \leq k$, are the principal angles between p and q.

1. Introduction. Let X be a Riemannian homogeneous space with metric ρ and (normalized) invariant measure dh. A *code* C is a finite subset of X. Let $\delta = \delta(C)$ be the minimum distance between distinct points in C. One of the main problems of coding theory is establishing the maximum size of a code with a given distance δ. The most well-known examples are the sphere $S^{n-1}(\mathbb{R})$ and, in the discrete case, the Hamming space H_q^n. One of the possible generalizations of the former is studying codes in the Grassmann manifold $G_{k,n}(L)$, where $L = \mathbb{R}$ or $\mathbb{C}$. It is a homogeneous space of the group $O(n)$ or $U(n)$, respectively. For instance, $G_{k,n}(\mathbb{R}) \cong O(n)/O(k) \times O(n-k)$. Presently this space is in the focus of attention for a number of reasons. From a purely geometric point of view, packings in $G_{k,n}$ form a natural generalization of spherical codes. However, their study seems to have been first addressed only a few years ago [2], [13], motivated in part by a group-theoretic application that connects them with the theory of quantum codes. Apart from this, codes in $G_{k,n}(\mathbb{C})$ arise naturally in the area of multiple-antenna transmission (especially, in the case of low noise), and there are a few papers that underline this connection [1], [15]. Therefore, it seems timely to address basic coding-theoretic questions such as the sphere-packing bounds on the size of codes. This is the aim of the present paper.

Let $B(\delta)$ be the metric ball of radius δ in X. Obviously, if M is any number such that

$$Mh(B(\delta)) < 1, \tag{1}$$

there exists a code in X of size $M+1$ and distance δ. Indeed, as long as (1) holds true, it is always possible to pick a point $x \in X$ so that the centers of already chosen balls together with x form a code with distance at least δ. This principle is called the Varshamov–Gilbert lower bound. On the other hand, obviously for any code C,

$$|C|h(B(\delta/2)) \le 1.$$

This is the Hamming upper bound on the size of codes.

To define the distance in $G_{k,n}$, we have to introduce principal angles between planes p and q. Let $a \in p$ and $b \in q$ be two vectors and $\theta = \arccos|\langle a, b\rangle|$ the angle between them. As a varies over p and b varies over q, θ has k stationary points,

$$0 \le \theta_k \le \cdots \le \theta_1 \le \pi/2,$$

corresponding to some pairs of vectors $(a_i, b_i), 1 \le i \le k$. The sets of vectors (a_i) and (b_i) form orthogonal bases in their respective planes, and if $k \le n/2$, then a_i is orthogonal to b_j for any $i \ne j$. For a proof see, for instance, [10]. Note also that if A_p is a generator matrix of a plane p, i.e., the matrix whose rows form a basis of p, and A_q is the same for q, then for the eigenvalues $\lambda_1, \ldots, \lambda_k$ of $A_q A_p^* A_p A_q^*$ we have $\lambda_i = \cos^2\theta_i$. In other words, this means that we can change the basis and apply a suitable unitary rotation so that p is generated by

$$\begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ & & \ldots & & \ldots & & \ldots & \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \end{bmatrix} \tag{2}$$

and q by

$$\begin{bmatrix} \cos^2\theta_1 & 0 & \dots & 0 & \sin^2\theta_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \cos^2\theta_2 & \dots & 0 & 0 & \sin^2\theta_2 & \dots & 0 & 0 & \dots & 0 \\ & & \dots & & & & \dots & & & \dots & \\ 0 & 0 & \dots & \cos^2\theta_k & 0 & 0 & \dots & \sin^2\theta_k & 0 & \dots & 0 \end{bmatrix}. \tag{3}$$

Definition. Let $p, q \in G_{k,n}(L)$. The distance between p and q is defined as follows:

$$\rho(p,q) = \sqrt{\sin^2\theta_1 + \cdots + \sin^2\theta_k}.$$

This is not the only possible definition of the distance; the reasons for taking this one are discussed at length in [2]. One of them is that under this definition $G_{k,n}(\mathbb{R})$ affords an isometric embedding into a sphere of radius $r = \sqrt{k(n-k)/k}$ in $\mathbb{R}^{(n-1)(n+2)/2}$. It is realized as follows. The matrix $\Pi_p = A^tA$ is an orthogonal projection from L^n on p. As shown in [2], for any p the Euclidean norm of the matrix $\Pi_p - (k/n)I_n$ equals r, so p is represented as a point on the sphere. Moreover, under this embedding $\rho^2(p,q) = (1/2)\|\Pi_p - \Pi_q\|^2$. Observe that the standard Plükker embedding maps $G_{k,n}$ into a space of a much higher dimension, namely $P(\bigwedge^k(L^n))$, and is not isometric.

Likewise, in the complex case the Hermitian matrix $\Pi_p = A^*A$ is an orthogonal projector on p. We again have $\rho^2(p,q) = \frac{1}{2}\|\Pi_p - \Pi_q\|^2$, $\operatorname{tr}\Pi_p = \frac{k}{n}$, and $\|\Pi_p - \frac{k}{n}I_k\|^2 = k(n-k)/n$. A Hermitian matrix with fixed trace can be represented by a point on the sphere in $\mathbb{R}^{n^2-1}$.

The existence of these embeddings implies that upper bounds on the size of codes on the sphere apply to codes in $G_{k,n}(L)$ with distance function ρ. In particular, the well-known Rankin bounds in the real case imply the following:

$$\delta(C) \le \begin{cases} \frac{k(n-k)}{n}\frac{|C|}{|C|-1}, & \text{if } |C| \le n(n+1)/2 \\ \frac{k(n-k)}{n} & \text{if } |C| > n(n+1)/2. \end{cases}$$

This inequality is an analog of the Plotkin bound of coding theory. Interestingly, there exist sequences of codes in $G_{k,n}(\mathbb{R})$ that meet these bounds [13].

We are interested in codes in $G_{k,n}$ whose size grows exponentially with n. Our goal is to derive an expression for the Varshamov–Gilbert and Hamming bounds on codes. The answer can be written in a compact form only in the asymptotic setting. We assume that $n \to \infty$ and k is a fixed constant. Further the quantity $R = (1/n)\ln|C|$ is called the rate of the code. Note that the case $G_{1,n}(\mathbb{R})$ with the (Riemannian) metric $\tilde\rho(p,q) = \theta$ was treated by Shannon [12]. He proved that there exist sequences of codes with distance θ and

$$R \gtrsim -\ln\sin\theta + o(1).$$

Note that in the case $k = 1$ codes optimal for the metric $\tilde\rho$ will be also optimal for the metric ρ.

1.1. *Sphere packing in $G_{k,n}$.* Our main result, proved in Sect. 3, is as follows: Let $B(\delta)$ be the ball of radius δ in $G_{k,n}(L)$. Then

$$h(B(\delta)) = \Big(\frac{\delta}{\sqrt{k}}\Big)^{\beta nk+o(n)} \qquad (n \to \infty,\ k = \text{const}), \tag{4}$$

where $\beta = 1, 2$ for $L = \mathbb{R}, \mathbb{C}$, respectively. With the above, this implies the following theorem.

Theorem 1 *Let $n \to \infty$, $k = const$. There exist sequences of codes in $G_{k,n}$ with asymptotic rate*

$$R \gtrsim -\beta k \ln(\delta/\sqrt{k}). \tag{5}$$

For any sequence of codes,

$$R \lesssim -\beta k \ln(\delta/2\sqrt{k}).$$

1.2. *Distance distribution.* Apart from the minimum distance, an important parameter of codes is their distance distribution, i.e., the average number of neighbors of a code point at a given distance. Since t in principle can be any number between δ and $\sqrt{k}$, it is convenient to consider the distance density of C defined as follows:

$$a(t) = \frac{1}{|C|}\big|\{(p,q) : t - (1/n) \le \rho(p,q) \le t\}\big|.$$

Suppose C is chosen in X with uniform probability distribution. Then

$$\frac{\mathsf{E}a(t)}{|C|} = h(B(t)) - h(B(t - n^{-1})),$$

where E denotes the mathematical expectation. From (4), the right-hand side equals $h(B(t))(1 - o(1))$. By standard methods of coding theory it is possible to show that among sequences of codes that meet the bound (5), i.e., for which the distance $\delta(C) \gtrsim \sqrt{k}e^{-R/\beta k}$, there exist codes, called below *uniform*, whose distance density is bounded above as

$$a(t) \le p(n)|C|h(B(t))(1 - o(1)) \quad (e^{-R/\beta k} \le t/\sqrt{k} \le 1),$$

where $p(n)$ is some function of polynomial growth. In other words, the logarithm of the average number of neighbors for uniform codes is bounded above as

$$\ln a(t) \le n(R + \beta k \ln(\delta/\sqrt{k}) + o(1)).$$

Uniform codes in other spaces of interest to coding and information theory (the binary Hamming space H_q^n and the sphere $S^{n-1}(\mathbb{R})$) have a number of interesting properties. The most important of them is related to the use of codes for transmission of information over noisy channels. In this situation uniform codes account for the best known exponential upper bounds on the probability of incorrect recovery of the code vector transmitted from the noisy version of this vector received from the channel [12].

2. Invariant densities in $G_{k,n}$. To prove (4), we need explicit volume forms on $G_{k,n}(L)$. A general construction of invariant measures in homogeneous spaces with applications to classical groups and related manifolds is given, for instance, in [11]. A combinatorial approach is presented in [8]. Necessary background material can be looked up in any textbook on geometry, for instance, [14], [16].

We begin with the real case. Density for the submanifold of critical angles was calculated several times in statistics (see [7]). Let $a_1, \dots, a_k$ and $b_1, \dots, b_{n-k}$ be orthonormal column vectors that span a plane p and its orthogonal complement. The invariant measure on $G_{k,n}(\mathbb{R})$ is (locally) given by the $k(n-k)$-form

$$v_k^n = \bigwedge_{j=1}^{n-k} \bigwedge_{i=1}^{k} b_j^t da_i,$$

where ^{t} means transposition.

To isolate the part of this form that corresponds to the density on principal angles, we also introduce the Stiefel manifold $V_{k,n}$, i.e., the manifold of orthonormal k-frames in R^n. In particular, $V_{k,k}$ is the orthogonal group $O(k)$. It is proved in [7] that the open part of $G_{k,n}$ decomposes into a direct product of the simplex

$$\Theta = \{(\theta_1, \dots, \theta_k) : \pi/2 > \theta_1 > \cdots > \theta_k > 0\}$$

and two manifolds, $\tilde{V}_{k,k}$ and $V_{k,n-k}$, where $\tilde{V}$ is the submanifold of the Stiefel manifold specified by those frames in which in each vector the first coordinate is positive. Based on this, it is possible to write v_k^n as a product of three independent densities and compute the marginal distribution on Θ [7]. This gives the answer in the real case (see ω_k below). In the complex case it is easier, though not so intuitive, to rely upon the distribution of eigenvalues of random Gaussian unitary matrices [9, 5]. We can assume that q is a fixed plane with generator matrix $J = [I_k \, 0]$ and p is uniformly distributed on $G_{k,n}(L)$. Then we are interested in the distribution of the eigenvalues $\lambda_i = \cos^2\theta_i$ of the matrix $JA_p^*A_pJ^*$. It can be shown that this distribution is related to the distribution of eigenvalues of Wishart matrices, i.e., matrices of the form G^*G where G is a $k \times n$ matrix with Gaussian $\mathcal{N}(0,1)$ elements [5, p. 202]. The final answer has the form

$$\omega_k = K(k,n) \prod_{i=1}^{k} (\sin\theta_i)^{\beta(n-2k)} \prod_{1 \le i < j \le k} (\sin^2\theta_i - \sin^2\theta_j)^{\beta} d\theta_1 \wedge \cdots \wedge d\theta_k,$$

where the constant $K(k,n)$ is chosen to normalize the measure of Θ. In the real case its value is obvious from the above geometric considerations:

$$K(k,n) = \frac{\operatorname{vol}(\tilde{V}_{k,k})\operatorname{vol}(V_{k,n-k})}{\operatorname{vol}(G_{k,n})}.$$

Volumes of the manifolds involved are well known [11], and we get

$$K(k,n) = \prod_{i=1}^{k} \frac{O(k-i+1)^2 O(n-k-i+1)}{2O(n-i+1)},$$

where $O(k) = 2\pi^{k/2}/\Gamma(k/2)$ is the area of the unit sphere in $\mathbb{R}^k$. Note that $K(k,n)$ grows polynomially in n for fixed k, so in our context its exact form is not essential. This also holds true for $\mathbb{C}$, namely, we have [6]

$$K(k,n) = \prod_{i=1}^{k} \frac{(n-i)!}{(i!)^2(n-k-i)!}.$$

So the volume of the ball of radius δ is given by $h(B(\delta)) = K(k,n)J_k(\delta)$, where

$$J_k(\delta) = \int\limits_{\substack{(\theta_1,\dots,\theta_k)\in\Theta \\ \sin^2\theta_1+\cdots+\sin^2\theta_k\le\delta^2}} \frac{\omega_k}{K(k,n)}. \tag{6}$$

3. Asymptotics. We would like to compute the logarithmic asymptotics of $h(B(\delta))$. Let us compute the behavior of J_k. We have

$$J_k = \int\!\!\cdots\!\!\int\limits_{\substack{0<x_k<\cdots<x_1<1 \\ x_1^2+\cdots+x_k^2\le\delta^2}} \frac{(x_1x_2\dots x_k)^{\beta(n-2k)}\prod_{i<j}(x_i^2-x_j^2)^\beta}{\sqrt{(1-x_1^2)(1-x_2)^2\dots(1-x_k)^2}}dx_1\dots dx_k.$$

Note that for $\beta = 2$ (but not for 1), J_k is symmetric in $x_1,\dots,x_k$, so we can divide out $k!$ and remove the ordering condition. To treat both cases simultaneously, we compute the asymptotics of the integral by the Laplace method [3, 4]. In the multidimensional case the corresponding theorem has the following form [4, p. 131]:

Let Ω be a connected open domain in $\mathbb{R}^k$, $\partial\Omega$ be its boundary, $[\Omega] = \Omega\cup\partial\Omega$ and let $f(x), S(x) \in C([\Omega])$ be real functions. Further suppose that the maximum $\max_{x\in[\Omega]} S(x)$ is attained only at the point $x^0 \in \partial\Omega$, and at this point
(a) $\partial S(x^0)/\partial\boldsymbol{n} \neq 0$, *where $\partial/\partial\boldsymbol{n}$ is differentiation along the interior normal to $\partial\Omega$ at x^0;*
(b) *the matrix $H = \left(\frac{\partial^2 S(x^0)}{\partial\xi_i\partial\xi_j}\right)_{i,j=1}^{k-1}$ is negative definite, where $(\xi_1,\dots,\xi_{k-1})$ is an orthonormal basis in the tangent space $T\,\partial\Omega_{x^0}$ to $\partial\Omega$ at x^0;*
(c) *$f, S, \partial\Omega \in C^\infty$ in the neighborhood of x^0.*

Then

$$\int_\Omega f(x)e^{\lambda S(x)}dx = \lambda^{-\frac{k+1}{2}} e^{\lambda S(x^0)} \sum_{m=0}^{\infty} a_m\lambda^{-m}, \quad (\lambda\to\infty)$$

for some constants a_m. Moreover, a_0 is proportional to $f(x_0)$.

Let us use this result in our problem. We have $x \in \mathbb{R}^k$, $\lambda = n$,

$$S(x) = \beta \ln(x_1 \dots x_k) = \beta \sum_i \ln x_i,$$

$$f(x) = \Big(\prod_{i=1}^{k} x_i^{-2\beta k} (1 - x_i^2)^{-1/2} \Big) \prod_{i<j} (x_i^2 - x_j^2),$$

$$[\Omega] = \Big[(x_1, \dots, x_k) : 0 \le x_k \le \dots \le x_1 \le 1, \ \sum_i x_i^2 \le \delta^2 \Big].$$

As is easily checked, the maximum of $S(x)$ over $[\Omega]$ is attained at $x^0 = k^{-(1/2)}(\delta, \delta, \dots, \delta)$. To satisfy the conditions of the theorem, we have to adjust the integration domain in several ways.

(i) Observe that $S(x)$ has discontinuities at the hyperplanes $x_i = 0$ and $f(x)$ has discontinuities at the hyperplanes $x_i = 1$. Therefore, let us shift the domain $[\Omega]$ from these hyperplanes.

(ii) At the point x^0 the boundary $\partial\Omega$ is not differentiable. Therefore, let us extend the domain by including small sectors.

For instance, for $k = 2$ the domain $[\Omega]$ is formed by the intersection of the sector of radius δ and angle $\pi/4$ in the first quadrant and the strip $0 \le x_1 \le 1$. The discussed extension amounts to increasing the angle to $\pi/4 + \varphi$ for some small finite φ. Note that the maximum of $S(x)$ over the extended domain does not shift from x^0. In the general case, consider the adjusted domain

$$\Omega_\varepsilon = \Big[(x_1, \dots, x_k) : \varepsilon \le x_i \le 1 - \varepsilon, \ x_{i+1} \le x_i(1 + \varepsilon), \ \sum_i x_i^2 \le \delta^2 \Big].$$

Note that to Ω_ε corresponds an adjusted domain Θ_ε in the coordinates $\theta_1, \dots, \theta_k$, where $\mathrm{vol}(\Theta \setminus \Theta_\varepsilon) + \mathrm{vol}(\Theta_\varepsilon \setminus \Theta) = o(\varepsilon)$. Since the integrand in (6) is bounded by 1, this implies that J_k differs from the corresponding integral over Ω_ε by $o(\varepsilon)$ uniformly in n.

(iii) Finally, the case $\delta = \sqrt{k}$ also has to be excluded since then $f(x)$ has a singularity at x^0. In this case the ball exhausts the entire space X except for a set of measure 0, and the bounds are extended by continuity.

We need to verify conditions (a) and (b) of the theorem. Note that in the basis $(x_1, \dots, x_k)$ the Hessian $H(x) = (\frac{\partial^2 S(x)}{\partial x_i \partial x_j})$ has the form $-\mathrm{diag}(x_1^{-2}, \dots, x_k^{-2})$, and so

$$H(x^0) = -\frac{\delta^2}{k} I_k.$$

This quadratic form is negative definite. Clearly, H remains negative definite under the restriction to the subspace spanned by $(\xi_1 \dots, \xi_{k-1})$. To verify (a), note that the interior normal to the sphere $\partial\Omega$ has the form $\boldsymbol{n} = (-1, \dots, -1)$. The level surfaces of $S(x)$ are given by $x_1 \dots x_k = \mathrm{const}$, which are convex hypesurfaces. Moreover, they are homothetic to each other with respect to the origin, therefore $S(x)$ strictly increases along the diagonal.

Therefore, let us apply the theorem to J_k. We have $f(x^0) = 0$; thus

$$J_k = n^{-\frac{k+1}{2}} \Big(\frac{\delta}{\sqrt{k}}\Big)^{\beta kn} O(n^{-1}).$$

This concludes the proof of (4).

References

[1] D. Agrawal, T. Richardson, R. Urbanke, *Multiple-antenna signal constellations for fading channels*, IEEE Trans. Inform. Theory, submitted.

[2] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, *Packing lines, planes, etc.: Packings in Grassmannian spaces*, Experimental Mathematics **5** (1996), no. 2, 139–159.

[3] N. G. de Bruijn, *Asymptotic methods in analysis*, North-Holland, Amsterdam, 1958.

[4] M. V. Fedoryuk, *Asimptotika: integraly i ryady* [*Asymptotics: Iintegrals and series*], Nauka, Moscow (In Russian), 1987.

[5] V. L. Girko, *An introduction to statistical analysis of random arrays*, VSP Press, Utrecht, 1998.

[6] B. Hassibi, B. Hochwald, and T. Marzetta, *Space-time autocoding*, IEEE Trans. Inform. Theory, submitted.

[7] A. T. James, *Normal multivariate analysis and the orthogonal group*, Ann. Math. Statistics **25** (1954), 40–75.

[8] D. A. Klain and G-C. Rota, *Introduction to geometric probability*, Cambr. Univ. Press, Cambridge, 1997.

[9] M. L. Mehta, *Random matrices*, Boston: Academic Press, 1991.

[10] S. N. Roy, *A note on critical angles between two flats in hyperspace with certain statistical applications*, Sankhyā **8** (1947), 177–194.

[11] L. A. Santaló, *Integral geometry and geometric probability*, Addison-Wesley, Reading, MA, 1976.

[12] C. E. Shannon, *Probability of error for optimal codes in a Gaussian channel*, Bell Syst. Techn. Journ. **38** (1959), no. 3, 611–656.

[13] P. W. Shor and N. J. A. Sloane, *A family of optimal packings in Grassmannian manifolds*, J. Algebraic Combinatorics **7** (1998), 157–163.

[14] M. Spivak, *Calculus on manifolds*, Benjamin Inc., New York, 1965.

[15] L. Zheng and D. N. C. Tse, *Packing spheres in the Grassmann manifold: A geometric approach to the non-coherent multi-antenna channel*, IEEE Trans. Inform. Theory, submitted.

[16] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, New York: Springer-Verlag, 1983.