FROM CH. III.3. *ON THE CHOICE OF A KERNEL FUNCTION IN SYMMETRIC SPACES* OF THE BOOK "*PATTERN RECOGNITION: THE METHOD OF POTENTIAL FUNCTIONS*".†

M.A. AIZERMAN, E.M. BRAVERMAN, AND L.I. ROZONOER

The book *"Pattern recognition: the method of potential functions"* was published in Soviet Union more than 45 years ago. It was never translated to English and many ideas of the book remained unknown outside of Russia. We present here a sligtly abridged version of Chapte III.3 of the book. Technical details of some proofs are omitted and replaced by a short sketch of the main steps of the proof. An interested reader can either fill those details or consult the original Russian edition.

The chapter was translated by Benjamin Rozonoer. The translation was edited by Maxim Braverman.

1. **Symmetric space.** In this section we consider a metric space of special type, called *symmetric spaces*. An m-dimensional cube, which often appears in problems of machine learning, is an important particular instance of such spaces. For symmetric spaces it is possible to give more precise choice of a system of functions $\phi_i(x)$ and the kernel function $K(x,y)$, and also to justify the usefulness of the choice of a kernel function as a function of distance.

The definition of a symmetric space is based on the concept of a isometric transformation of a metric space into itself. A transformation of a metric space into itself is called *isometric* if it preserves the distance between any pair of points. If A is an isometric transformation, such that point x transforms to point Ax as a result of the transformation, then, in accordance with the definition, for any pair of points $x, y \in X$,

$$\rho(Ax, Ay) \ = \ \rho(x,y). \tag{32}$$

It is clear that the successive application of several isometric transformations is an isometric transformation and that the inverse transformation exists and is also isometric. thus the set of isometric transformations forms a group, where, as usual, the product of two elements (transformations) means successive application of these transformations; the identical transformation can be considered as the groups unit, and the inverse transformation plays the role of the inverse element.

†Translated to English by Benjamin Rozonoer.

A metric space X is called *symmetric* if, for any two pairs of points x', y' and x'', y'', located at the same distance $\rho(x', y') = \rho(x'', y'')$, there exists an isometric transformation A of the space X into itself, such that $x'' = Ax'$ and $y'' = Ay'$.

An example of a symmetric space is the above-mentioned set of vertices of an m-dimensional cube, containing $N = 2m$ points, if we use Hammings definition of distance. In this case the isometric transformations are rotations and reflections of the cube (c.f. article [5] below for more details).

Another example of a symmetric space is the set, consisting of N points, evenly distributed along a circle in such a way that the lengths of the shortest arcs between any pair of neighboring points is the same. In this space we set the distance between any two points to be the length of the shortest arc which connects them. Then the isometric transformations are rotations and reflections relative to the corresponding diameters.

From now on, when talking about symmetric space, we will assume that it contains only a finite number of points.

2. **Quadratic quality functionals on symmetric spaces.** Next we define functionals of type $\mathscr{F}\{f(x)\}$, which evaluates the "quality of the function $f(x)$, i. e. its "smoothness, simplicity etc. From the point of view of intuitive notions about the quality of functions, it is natural to demand for the functional $\mathscr{F}\{f(x)\}$ to possess the following properties:

$$\mathscr{F}\{\lambda f(x)\} = \mathscr{F}\{f(x)\}; \tag{33}$$

$$\mathscr{F}\{Af(x)\} = \mathscr{F}\{f(x)\}, \tag{34}$$

where λ is any non-zero constant, and A is any isometric transformation of the symmetric space X, on which f(x) is defined. Really, the multiplication of a function by a non-zero constant does not change its "spectrum composition", which is what defines the quality of the function. The second requirement is justified by the fact that the function $f(Ax)$ is simply "shift" of the function $f(x)$.

In this chapter we consider quality functionals of the form

$$\mathscr{F}\{f(x)\} \ := \ \frac{\sum_{x,y\in X} \ L(x,y)\, f(x)\, f(y)}{\|f\|^2}, \tag{35}$$

where

$$\|f\| \ := \ \sqrt{\sum_{x\in X} |f(x)|^2}.$$

Without loss of generality we can assume that the kernel $L(x, y)$ is symmetric

$$L(x,y) \ = \ L(y,x).$$

The specific form of the functional is determined by the choice of the kernel $L(x, y)$. The requirement (33) is satisfied automatically due to the appearance of the quantity $\|f\|^2$

in the denominator. The requirement (34) considerably limits the potential form of the kernel $L(x, y)$. Namely, the following theorem holds.

Theorem I. *Let X be a symmetric space. Then, the functional* (35) *satisfies* (34) *for every function f(x) and every isometric transformation A if and only if the kernel L(x, y) is a* function of distance between points x and y*:*

$$L(x, y) \ = \ L\big(\rho(x, y)\big).$$

We precede the proof of theorem II with the following

Lemma I. *For a function $\psi(x, y)$ defined on a finite symmetric space X to be a function of distance $\rho(x, y)$, it is both necessary and sufficient that for any isometric transformation A*

$$\psi(Ax, Ay) = \psi(x, y).$$

Proof of Lemma I. The necessity of the lemmas condition immediately follows from the definition an isometry since

$$\psi\big(\rho(Ax, Ay)\big) \ = \ \psi\big(\rho(x, y)\big).$$

Let us prove the sufficiency of the of the lemmas conditions. Let x, y and x', y' be two pares of points . We must prove, that if $\rho(x, y) = \rho(x', y')$, then it follows from the assumptions of the lemma that
$psi(x, y) = \psi(x', y')$. Since the space X is symmetric there exists an isometry A such that

$$x' = Ax, \quad y' = Ay.$$

Hence, $\psi(x', y') = \psi(Ax, Ay)$. But by the assumptions of the lemma $\psi(Ax, Ay) = \psi(x, y)$. Thus $\psi(x', y') = \psi(x, y)$. The lemma is proven. □

Proof of Theorem I. We rewrite the condition (34) as

$$\frac{\sum_{x,y\in X} \ L(x, y)\, f(Ax)\, f(Ay)}{\|f(Ax)\|^2} \ = \ \frac{\sum_{x,y\in X} \ L(x, y)\, f(x)\, f(y)}{\|f\|^2}. \tag{36}$$

The denominators in both parts of these expressions are equal. Hence, after the change of variables $u = Ax, \ v = Ay$ the equality (36) is equivalent to

$$\sum_{u,v\in X} L(A^{-1}u, A^{-1}v) f(u) f(v) \ = \ \sum_{u,v\in X} L(u, v)\, f(u) f(v). \tag{37}$$

By the assumptions of Theorem I this equality hold for any function f. Form this we immediately conclude that the function $L(A^{-1}u, A^{-1}v)$ and $L(u, v)$ coincide:

$$L(A^{-1}u, A^{-1}v) \ = \ L(u, v). \tag{38}$$

Since (38) holds for every isometry A, the theorem follows now from Lemma I. □

Theorem I allows us to rewrite expression (35) in the form

$$\mathscr{F}\{f(x)\} \;:=\; \frac{\sum_{x,y\in X} L\big(\rho(x,y)\big)\, f(x)\, f(y)}{\|f\|^2}, \tag{39}$$

Using Theorem I one can show that the functional (39) is uniquely determined by the functional

$$\widetilde{\mathscr{F}}\{f(x)\} \;:=\; \frac{\sum_{x,y\in X} L\big(\rho(x,y)\big)\, \big[f(x)-f(y)\big]^2}{\|f\|^2}, \tag{40}$$

by the formula

$$\widetilde{\mathscr{F}}\{f\} \;=\; 2\,\big(C - \mathscr{F}\{f\}\big), \tag{41}$$

where C is a constant depending only on the kernel $L(\rho)$.

To define this constant we consider a function $S(\rho)$, whose value equals the number of points of symmetric space X that lie on the sphere $Sp_x(\rho)$ of radius ρ with center at an arbitrary point x. Since the space X is symmetric, $S(\rho)$ does not depend on x. The expression for constant C has the form

$$C \;=\; \sum_{\rho} L(\rho)\, S(\rho). \tag{42}$$

Indeed, opening the parentheses in formula (40), we obtain

$$\widetilde{\mathscr{F}}\{f(x)\} \;:=\; 2\,\frac{\sum_{x,y\in X} L\big(\rho(x,y)f(y)^2}{\|f\|^2} \;-\; 2\mathscr{F}\{f(x)\}. \tag{43}$$

Let us first sum over x in (43), summing successively over spheres with radii $\rho = 0, \rho_1, \rho_2, \ldots$ with center in some fixed point y. Since the space X is symmetric, the number of points $S(\rho)$ on a sphere of radius ρ does not depend onthe choice of the center y. For this reason

$$\sum_{x\in X} L\big(\rho(x,y)\big) \;=\; \sum_{\rho} L(\rho)\, S(\rho) \;=\; C$$

is independent of y. Summing now over y in (43) and using the definition of $\|f\|$ we obtain (41).

The notation of the quality functional in form (40) is convenient in the sense that it quite clearly reflects ones intuitive perceptions of a functions quality, as it directly includes the difference of the functions values at points x and y, located at a distance $\rho(x,y)$. In particular, if the kernel $L\big(\rho(x,y)\big)$ is non-negative, then the functional takes the minimal (zero) value on constant functions. Meanwhile, since with the “worsening of function $f(x)$, the differences $[f(x) - f(y)]^2$ increase, broadly speaking, we can assume, that with the growth of the value of functional (40) the function “worsens. Conversely, in accordance with formula (41) with a positive kernel $L\big(\rho(x,y)\big)$, it is the decrease of functional (39) that corresponds to the “worsening of the function.

3. **The assignment of classes for functions of equal quality.** In this section we will need some knowledge from the theory of representation of groups.

Let G be a group and let $\mathscr{L}$ be a linear space. A representation of G in $\mathscr{L}$ is an assignment of a linear transformation $T(A) : \mathscr{L} \to \mathscr{L}$ to each element $A \in G$ such that the product of elements A and B of a group G corresponds to the product of operators, i.e.

$$T(A \cdot B) = T(A) \circ T(B).$$

The *dimension* of a representation T is defined to be dimension of $\mathscr{L}$.

In this work $\mathscr{L}$ is the linear space of real-valued functions on X. If, as we assume here, the space X consists of a finite number N of points, then the dimension of space $\mathscr{L}$ is N.

Let us A be an isometric transformation of the space X into itself. To every function $f(x) \in \mathscr{L}$ we will assign a corresponding function $g(x) = f(A^{-1}x) \in \mathscr{L}$. This correspondence between functions $f(x)$ and $g(x)$ specifies the operator $T(A)$:

$$g = T(A)f.$$

The operator $T(A)$ depends, of course, on what kind of isometric transformation A is considered, but no matter how A is chosen, the corresponding operator $T(A)$ is linear. In fact, for any functions $f_1(x)$ and $f_2(x)$ and numbers λ_1, λ_2 we have

$$T(A)[\lambda_1 f_1(x) + \lambda_2 f_2(x)] \;=\; \lambda_1 f_1(Ax) + \lambda_2 f_2(Ax) \;=\; \lambda_1 T(A) f_1 + \lambda_2 T(A) f_2.$$

The correspondence $A \mapsto T(A)$ between the isometric transformations of space X and linear operators on space $\mathscr{L}$ of functions on X, is precisely the representation of group G of isometric transformations that interests us.

Since the dimension of $\mathscr{L}$ is finite, a choice of a basis $\phi_1(x), \dots, \phi_N(x)$ in $\mathscr{L}$ allows us to represent the operator $T(A)$ by a square matrix $\|T_{ik}(A)\|$. Thus the representation $T(A)$ is given by the correspondence between isometric transformations and square $(N \times N)$-matrices.

If $\phi_1(x), \dots, \phi_N(x)$ is an orthonormal bases then the matrix $\|T_{ik}(A)\|$ is orthogonal for any isometry A.

A representation T is called *irreducible*, if there are no non-trivial subspaces of L which are invariant for all operators $T(A)$. Otherwise the representation is called *reducible*.

It is a classical result of the representation theory of finite groups that a finite-dimensional linear space $\mathscr{L}$ on which a finite group G acts decomposes into a direct sum of irreducible representations. In particular,

$$\mathscr{L} \;=\; \mathscr{L}_0 \oplus \mathscr{L}_1 \oplus \cdots \oplus \mathscr{L}_m. \tag{44}$$

Here each $\mathscr{L}_i$ is an invariant subspace of $T(A)$ for all isometries $A : X \to X$. For $s = 0, \dots, m$, we set $N_s := \dim \mathscr{L}_s$ and we denote by $T^s(A)$ the restriction of $T(A)$ to $\mathscr{L}_s$. Then each T^s is an irreducible representation of the group of isometries.

We note the the the representation T of the group of isometries of X in the space $\mathscr{L}$ of linear functions on X is always reducible. Indeed, the space $\mathscr{L}_0 \subset \mathscr{L}$ of constant functions is a non-trivial invariant subspace. Thus $\mathscr{L}$ is a direct sum of non-trivial irreducible representations. Every function $f(x) \in \mathscr{L}$ can be projected onto each subspace $\mathscr{L}_s$. We denote by f^s this projection. Then f can be written as a direct sum of its "irreducible" components

$$f(x) \;=\; \sum_{s=0}^{m} c_s \phi^s(x), \tag{45}$$

where $c_s := \|f^s\|$ and $\phi^s := f^s/\|f^s\|$ is the normalized projection onto $\mathscr{L}_s$.

The decomposition of the space $\mathscr{L}$ into orthogonal subspaces $\mathscr{L}_s$ has a direct relation to the question that interests us concerning the evaluation of the quality of functions. And indeed, the following theorem holds.

Theorem II. *The functional* (39) *takes the same value* $\mathscr{F}_s$ *on all functions* $f \in \mathscr{L}_s$ *(*$s = 0, \dots, m$*). The value* $\mathscr{F}_s$ *depends only on* s *and the choice of kernel* $L\big(\rho(x,y)\big)$.

For the proof of Theorem II we will need the following lemma.

Lemma II. *The representations* $T_s(A)$ *and* $T_q(A)$ *are not equivalent when* $s \neq q$.

Proof of Lemma II. Let us assume the contrary, that the representations $T_s(A)$ and $T_q(A)$ are equivalent. Then one can choose the orthonormal bases $\phi_1, \dots, \phi_l$ and $\psi_1, \dots, \psi_l$ of the subspaces $\mathscr{L}_s$ and $\mathscr{L}_q$ respectively ($N_s = N_q = l$) such that the matrices $\|T^s_{ik}(A)\|$ and $\|T^q_{ik}(A)\|$ are orthogonal and coincide for all isometries A.

Consider the function

$$\Phi(x,y) \;:=\; \sum_{i=1}^{l} \phi_i(x)\psi_i(y). \tag{46}$$

One immediately checks that for any isometric operator A the following identity holds:

$$\Phi(Ax, Ay) \;=\; \Phi(x,y). \tag{47}$$

By Lemma I, any function satisfying (47) is a function of distance. Thus

$$\sum_{i=1}^{l} \phi_i(x)\psi_i(y) \;=\; \Phi\big(\rho(x,y)\big). \tag{48}$$

Since $\Phi\big(\rho(x,y)\big) = \Phi\big(\rho(y,x)\big)$ it follows from (48) that

$$\sum_{i=1}^{l} \phi_i(x)\psi_i(y) \;=\; \sum_{i=1}^{l} \phi_i(y)\psi_i(x).$$

This equality is contradictory, since setting choosing $y = y^*$ such that at least one of the values $\phi_i(y^*),\ \psi_i(y^*)$ $(i = 1,\dots,l)$ is different from zero, we come to the conclusion, that the system of orthonormal functions $\phi_1,\dots,\phi_l;\psi_1,\dots,\psi_l$ is linearly dependent. The obtained contradiction refutes the assumption about the equivalence of the representations of $T_s(A)$ and $T_q(A)$. Lemma II is proven. □

To each function $L\big(\rho(x,y)\big)$ we associate a linear operator $\widehat{L} : \mathscr{L} \to \mathscr{L}$, defined by the formula

$$\widehat{L}f \;:=\; \sum_{y\in X} L\big(\rho(x,y)\big)\cdot f(y). \tag{49}$$

One readily checks that $\widehat{L}$ commutes with all the operators $T(A)$:

$$T(A)\widehat{L} = \widehat{L}T(A). \tag{50}$$

Lemma III. *For each $s = 0,\dots,m$, all functions $f \in \mathscr{L}_s$ are eigenfunctions of $\widehat{L}$ with the same eigenvalue, i.e. there exists $\mu_s \in \mathbb{R}$ such that*

$$\widehat{L}f \;=\; \mu_s f, \quad \textit{for all} \;\; f \in \mathscr{L}_s.$$

Proof of Lemma III. Consider the set $L'_s := \widehat{L}(\mathscr{L}_s) \subset \mathscr{L}$. We will show that this subspace is invariant relative to each of the operators of the representation T(A), i.e.

$$T(A)f' \in \mathscr{L}_s, \qquad \text{for all} \;\; f' \in \mathscr{L}_s. \tag{51}$$

Indeed, by (50), we have $T(A)f' = T(A)\widehat{L}f = \widehat{L}T(A)f$, and since $T(A)f \in \mathscr{L}_s$, we conclude that $T(A)f' \in \mathscr{L}_s$.

Let us now examine two possible cases:

a) subspace $\mathscr{L}_s$ is a trivial subspace, containing only the null vector $f'(x) = 0$;

b) in subspace $\mathscr{L}'_s$ there exists at least one nonzero vector.

In case a) the lemmas assertion is obvious, since $\widehat{L}f = 0$ for all $f \in \mathscr{L}_s$ and, hence, all functions in $\mathscr{L}_s$ are eigenfunctions of $\widehat{L}$ with eigenvalue 0.

Therefore from now on we only consider the case b). Representation $T^s(A)$ of G on L_s induces a representation $T'^s(A)$ of G on $\mathscr{L}'_s = \widehat{L}(\mathscr{L}_s)$. Since T^s is an irreducible representation, so is T'^s. Hence, $\mathscr{L}'_s$ coincides with one of L_q in the decomposition (44). We will now show that $\mathscr{L}'_s$ coincides with $\mathscr{L}_s$. Indeed, $\widehat{L} : \mathscr{L}_s \to \mathscr{L}'_s = \mathscr{L}_q$. By Shur's lemma this map is either 0 or an isomorphism of representations. As we assumed that $\mathscr{L}'_s \neq 0$ this map must be an isomorphism. The assertion follows now from Lemma II.

Thus $\widehat{L} : \mathscr{L}_s \to \mathscr{L}_s = \mathscr{L}'_s$. Since the representation $\mathscr{L}_s$ is irreducible it follows form the Shur's lemma that the restriction of $\widehat{L}$ to $\mathscr{L}_s$ is a non-zero constant μ_s. Thus $\widehat{L}f = \mu_s f$ for all $s \in \mathscr{L}_s$. The lemma is proven. □

Proof of Theorem II. Let μ_s be as in Lemma III. Notice that the functional $\mathscr{F}$ can be written as

$$\mathscr{F}\{f\} \;=\; \frac{\langle f, \widehat{L}f\rangle}{\langle f, f\rangle}.$$

Then for every $f \in \mathscr{L}_s$ we have

$$\mathscr{F}\{f\} \;=\; \frac{\langle f, \mu_s f\rangle}{\langle f, f\rangle} \;= \mu_s \;\equiv\; \mathscr{F}_s.$$

□

From theorem II follows a simple formula for the value of the functional (35), if function f(x) is defined by the decomposition (45). Namely,

$$\mathscr{F}\{f\} \;=\; \frac{\sum_{s=0}^{m} c_s^2\, \mathscr{F}_s}{\sum_{s=0}^{m} c_s^2}. \tag{52}$$

From (41) it follows that a similar formula holds for the functional $\widetilde{\mathscr{F}}$:

$$\widetilde{\mathscr{F}}\{f\} \;=\; \frac{\sum_{s=0}^{m} c_s^2\, \widetilde{\mathscr{F}_s}}{\sum_{s=0}^{m} c_s^2}. \tag{53}$$

From (52) we see that the choice of the kernel $L\big(\rho(x, y)\big)$ in functional (39) is reflected only in values of $\mathscr{F}_s$, evaluating the "quality of the irreducible component $\mathscr{L}_s$. Thus, for evaluating the complexity of an arbitrary function, one can define the numbers $\mathscr{F}_s$ instead of defining the kernel $L\big(\rho(x, y)\big)$, see Section 4 below.

Theorem II allows us to identify the classes of equivalent (in terms of their quality) functions without specifying the concrete functional of quality. No matter how the intuitive conceptions concerning a functions quality may be formalized, the ordering of functions based on their quality is connected with the introduction of order relations between two functions f and g: $f \preceq g$. This notation is read as: *"function f is not worse than function g*. The relation $\preceq$ is transitive.

We will say: *"a function f is equivalent to a function* g and denote $f \sim g$, if $f \preceq g$ and $g \preceq f$. We now postulate the following properties of the introduced order relations:

$$f(Ax) \sim f(x); \tag{54}$$

$$\text{if } f \preceq g, \text{ then } \;\lambda f(x) + \mu g(x) \preceq g(x). \tag{55}$$

Condition (54), essentially, coincides with condition (34), while condition (55) is stronger than (33). Postulating condition (33) allows us to establish the equivalence of functions belonging to the same irreducible component, without using a quality functional.

Theorem III. *If two non-zero functions $f(x)$ and $g(x)$ belong to the same irreducible component $\mathscr{L}_s$ then $f \sim g$.*

Proof of Theorem III. Let $f, g \in \mathscr{L}_s$ be non-zero functions. Since $\mathscr{L}_s$ is an irreducible representation of G, g belongs to the linear span on functions $\{T(A)f : A \in G\}$. In other words, $g(x)$ can be written as

$$g(x) \;=\; \sum_{i=1}^{N_G} \lambda_i f(A_i x), \tag{56}$$

where N_G is the rank of the group G. It follows from this equation, (54) and (55) that $g(x) \preceq f(x)$. Similarly, $f(x) \preceq g(x)$ and, hence, $f \sim g$. □

Thus the equivalence classes of functions of "the same quality" are determined without specifying the quality functional. However, to compare two functions, which have nonzero projections on at least two different layers, we must use functional (39), i.e. to specify the kernel function $L(\rho)$. This immediately leads to the establishment of concrete "weights" $\mathscr{F}_s$, attributed to the irreducible components, and it is necessary that these weights correspond to intuitive conceptions about the complexity of functions from these components. One has to remember this when choosing the kernel $L(\rho)$. An example of such a choice will be given in the beginning of Section 5.

4. **The power series expansion of a function of distance.** It was shown above, that the kernel of the quality functional (39) must be chosen to be a function of distance $\rho(x, y)$. In this section we examine the properties of a distance functions on symmetric spaces, related to their expansion into a series of some system of functions, also depending on the distance. The choice of this system of function is closely related to the decomposition of $\mathscr{L}$ into irreducible components, discussed in the previous section.

In each component $\mathscr{L}_s$ we choose an orthonormal basis ϕ_j^s $(j = 1, \dots, N_s,\ s = 0, \dots, m)$. Clearly, $N = \sum N_s$ and the collection ϕ_j^s is a basis of $\mathscr{L}$.

For each $s - 0, \dots, m$, define the function

$$K_s(x, y) \;:=\; \sum_{j=1}^{N_s} \phi_j^s(x) \phi_j^s(y). \tag{57}$$

Theorem IV. *The functions K_s are independent of the choice of the basis ϕ_j^s and are functions of distance*

$$K_s(x, y) = K_s\big(\rho(x, y)\big).$$

The system of functions $K_s(x,y)$ ($s = 0, \dots, m$) is a complete system of function in the space of functions of ρ.

Sketch of the proof of Theorem IV. One readily checks that the operator $\widehat{K}_s : \mathscr{L} \to \mathscr{L}$

$$\widehat{K}_s f(x) \ := \ \sum_{y \in X} K_s(x,y) f(y)$$

is the orthogonal projection onto $\mathscr{L}_s$. All the assertion of the theorem follow directly from this fact. □

It follows from this theorem that any function of distance is a linear combination of the functions K_s. In case of the kernel function (39) one readily sees that

$$L\big(\rho(x,y)\big) \ = \ \sum_{s=0}^{m} \mathscr{F}_s K_s(x,y), \tag{58}$$

where $\mathscr{F}_s$ are the weights defined in Theorem II.

We will now list some useful properties and relations connected with the functions $K_s(\rho)$.

First property. For all ρ

$$|K_s(\rho)| \le K_s(0) \ = \ \frac{N_s}{N} > 0. \tag{59}$$

Proof of the first property.

$$N K_s(0) = \sum_{x \in X} K_s(0) = \sum_{x \in X} \sum_{j=1}^{N_s} \phi_j^s(x)^2 = \sum_{j=1}^{N_s} \sum_{x \in X} \phi_j^s(x)^2 = \sum_{j=1}^{N_s} \|\phi_j^s\|^2 \ = N_s.$$

Hence, $K_s(0) = \frac{N_s}{N}$. This proves the second equality of (59). The first inequality of (59) follows from the Cauchy inequality. □

Second property. The functions K_s are orthogonal with weight $S(\rho)$:

$$\sum_{\rho} S(\rho) K_s(\rho) K_q(\rho) \ = \ \delta_{sq} K_s(0), \tag{60}$$

where $S(\rho)$ is the number of points in the sphere of radius ρ around any point $x \in X$ and δ_{sq} is the Kronecker symbol.

Using this property one can compute the coefficients μ_s by the formula

$$\mu_s = \frac{1}{K_s(0)} \sum_{\rho} L(\rho) K_s(\rho) S(\rho). \tag{61}$$

Sketch of the proof of the second property. Since K_s is the kernel of the orthogonal projection onto $\mathscr{L}_s$, K_s and K_q are orthogonal as functions on $X \times X$. Thus

$$\delta_{sq} K_s(x,x) = \sum_{x,y \in X} K_s(x,y) K_q(x,y) = \sum_{\rho} \sum_{\rho(x,u)=\rho} K_s(x,y) K_q(x,y) = \sum_{\rho} S(\rho) K_s(\rho) K_q(\rho)$$

□

Third property. The functions K_s satisfy the second orthogonality relation:

$$\sum_{s=0}^{m} \frac{K_s(\rho) K_s(\kappa)}{K_s(0)} \;=\; \frac{1}{S(\kappa)} \delta_{\rho\kappa}. \tag{62}$$

Proof of the third property. Fix κ and consider $\delta_{\rho\kappa}$ as a of ρ. Since $K_s(\rho)$ is a complete system of functions we can write

$$\delta_{\rho\kappa} \;=\; \sum_{s=0}^{m} \lambda_s(\kappa) K_s(\rho). \tag{63}$$

To compute the coefficients $\lambda_s(\kappa)$ we multiply (63) by $K_q(\rho) S(\rho)$ and sum over ρ. Then using (60) we have

$$\lambda_q(\kappa) K_q(0) = \sum_{\rho} \delta_{\rho\kappa} K_q(\rho) S(\rho) = K_q(\kappa) S(\kappa).$$

Hence,

$$\lambda_q(\kappa) \;=\; \frac{K_q(\kappa) S(\kappa)}{K_q(0)}. \tag{64}$$

Substituting this expression to (63) we obtain (62). □

From Theorem IV and the second property we obtain the following important result about symmetric spaces: *The number $m+1$ of irreducible components in decomposition (44) is equal to the number of different distances between the points of X.*

Notice now that the "complexity" of a function $K_s(\rho)$ also can be evaluated by the quality functional $\mathscr{F}_s$ or $\widetilde{\mathscr{F}}_s$. To make it precise, let us fix a point $x^* \in X$ and consider the function

$$g_s(y) := K_s\big(\rho(x^*, y)\big).$$

By (57) the function $g_s(y) \in \mathscr{L}_s$. It follows that $\mathscr{F}(g_s) = \mathscr{F}_s$ and $\widetilde{\mathscr{F}}\{g_s\} = \widetilde{\mathscr{F}}_s$. In this sense the "complexity" of K_s is the same as the "complexity" of the functions in $\mathscr{L}_s$. Using this result we can compute the "complexity" of any function $L(\rho)$ through its decomposition into a linear combination of K_s. Indeed if

$$L(\rho) = \sum_{s=0}^{m} \mu_s K_s(\rho),$$

then, using (52), we obtain

$$\mathscr{F}\{L(\rho(x^*,y))\} \;=\; \frac{\sum_{s=0}^{m}\mu_s^2 K_s(0)\mathscr{F}_s}{\sum_{s=0}^{m}\mu_s^2 K_s(0)}.$$

5. **The potential function in a symmetric space.** To explain how the above facts are used for the choice of a potential function in the method of potential functions, we need to specify a concrete form of the kernel function $L(\rho)$ in (39) and (40).

Let us define the kernel by the formula

$$L(\rho) = \frac{1}{4S(\rho_1)}\delta_{\rho\rho_1}, \tag{65}$$

where ρ_1 is the smallest distance between non-equal points of X. Then the value of the functional (40) is proportional to the sum of the squares of the differences of the values of f in the neighboring points. Thus the value of $\mathscr{F}\{f\}$ is bigger for the functions which we intuitively consider "worst".

One easily checks the following properties of the functional (40) with kernel (65):

1) $0 \le \widetilde{\mathscr{F}}\{f\} \le 1$;
2) if $f(x) \ge 0$ then $0 \le \widetilde{\mathscr{F}}\{f\} \le \frac{1}{2}$,

and as for any functional of this type

3) $\widetilde{\mathscr{F}}(const) = 0$.

For the kernel (65) the weights $\mathscr{F}_s$ and $\widetilde{\mathscr{F}_s}$ are expressed in terms of the functions K_s as follows

$$\mathscr{F}_s = \frac{1}{4}\,\frac{K_s(\rho_1)}{K_s(0)}, \tag{66}$$

$$\widetilde{\mathscr{F}_s} = \frac{1}{2}\left(1 - \frac{K_s(\rho_1)}{K_s(0)}\right). \tag{67}$$

These formulas follow easily from (62).

From now on we enumerate the irreducible components in (44) by weights $\widetilde{\mathscr{F}_s}$ so that

$$0 = \widetilde{\mathscr{F}_0} < \widetilde{\mathscr{F}_1} < \cdots < \widetilde{\mathscr{F}_m}. \tag{68}$$

We now turn to the question of choosing the potential function $K(x,y)$ in the *method of potential functions*. In practice whenever this method is used the function $K(x,y)$ is chosen as a function of distance: $K(x,y) = K(\rho(x,y))$. This choice for function $K(x,y)$ is justified by the following reasons. In Section 4 we introduced the system of functions $K_s(\rho)$ and proved its completeness. For this reason, no matter how the function $K(x,y) =$

$K(\rho(x,y))$ is chosen, it can be expressed as a sum

$$K\big(\rho(x,y)\big) = \sum_{s=0}^{m} \mu_s K_s\big(\rho(x,y)\big). \tag{69}$$

The method potential functions discussed in the previous sections of this book implies that the coefficients $\mu_s \geq 0$, which, in view of (59), that

$$K(0) > 0, \tag{70}$$

and for all ρ

$$|K(\rho)| < K(0). \tag{71}$$

In order to determine which further restrictions are reasonable, when choosing the potential function $K(\rho(x,y))$, i. e. for the assignment of non-negative numbers μ_s in decomposition (69), we will examine the machine realization of the method of potential functions (cf. 3 ch. II of this book). On every n-th step the machine realization comes down to computing the sum

$$f^n(x) = \sum_{i=0}^{n-1} r^i K(x, x^{i+1}). \tag{72}$$

If we introduce the function

$$\pi^n(x) = \sum_{i=0}^{n-1} r^i \delta_{x,x^{i+1}},$$

then

$$f^n(x) = \sum_{y\in X} K\big(\rho(x,y)\big)\pi^n(y). \tag{73}$$

Function $\pi(x)$ is equal to zero everywhere, except at points x_i, used in the process of learning. The problem of learning makes sense only when the number of points used in the process of learning is much less than the total number of points in space X. For this reason the function $\pi(x)$ is different from zero only in points that are separated from each other and it intuitively becomes clear that it is very "tattered, "plateresque". This is evident also from the value of the functional $\widetilde{\mathscr{F}}\big\{\pi^n(x)\big\}$. Really, it is easy to calculate the value of this functional with the assumption that among the chosen points x_i there are no adjacent points (i. e. $\rho(x_i,x_j) > \rho_1$ when $i \neq j$). This value is equal to

$$\widetilde{\mathscr{F}}\big\{\pi^n(x)\big\} \ = \ \frac{1}{2}$$

and depend neither on the number n of featured points (as long as there are no adjacent ones among them), nor on the values r^i (i. e. on the concrete learning algorithm). The value of the functional, equal to 1/2, corresponds to the very "tattered function (to which attests, for example, property 2) of functional $\mathscr{F}$). As for function $f^n(x)$, it should be smooth enough, since for large enough values of n (but still much less than than the total number of points in space X) it must approximate the function $f^*(x)$ that is being restored, which is assumed to be "smooth, not "tattered, i. e. possessing a high quality (c.f. section 1.1 of this chapter). To such functions $f^n(x)$ must correspond the small value of the functional $\widetilde{\mathscr{F}}\{\pi^n(x)\}$.

In formula (73) the function $K(\rho(x,y))$ can be viewed as the kernel of the linear "integral" operator $\widehat{K}$, which transforms a function $\pi^n(x)$ into a function $f^n(x)$. It follows from the above discussion that the operator $\widehat{K}$ must map the a function of bad quality (with a large value of $\widetilde{\mathscr{F}}$) to functions of good quality (with a small value of $\widetilde{\mathscr{F}}$). Therefore it makes sense to introduce the following definition: operator $\widehat{K}$ with kernel $K(\rho(x,y))$ is called *bettering* (resp. *worsening*), if $\widetilde{\mathscr{F}}\{\widehat{K}f\} \leq \widetilde{\mathscr{F}}\{f\}$ (resp. $\widetilde{\mathscr{F}}\{\widehat{K}f\} \geq \widetilde{\mathscr{F}}\{f\}$ for any function $f(x)$.

Assume that the irreducible components $\mathscr{L}_s$, and hence also the coefficients μ_s, are numbered in accordance with (68). Then the following theorem holds.

Theorem V. *Assume that the function $K(\rho)$, corresponding to operator $\widehat{K}$, is given by* (69)*. Then the operator $\widehat{K}$ is bettering (worsening) only if and only if the sequence $|\mu_s|$, $s = 0, 1, \ldots, m$ is non-increasing (non-decreasing).*

Proof of Theorem V. It is enough to proof the assertion of theorem V for bettering operators.
a) Proof of necessity. Suppose the requirement of the theorem are not satisfied, i. e. $|\mu_s| > |\mu_j|$ for some $k > j$. Consider the function

$$f(x) = \frac{1}{\sqrt{2}}\phi_j(x) + \frac{1}{\sqrt{2}}\phi_k(x),$$

where ϕ_i and ϕ_k are any functions from $\mathscr{L}_j$ and $\mathscr{L}_k$, respectively. By (53)

$$\widetilde{\mathscr{F}}(f) = \frac{1}{2}\left(\widetilde{\mathscr{F}_i} + \widetilde{\mathscr{F}_j}\right).$$

Since

$$\widehat{K}f = \frac{\mu_j}{\sqrt{2}}\phi_j + \frac{\mu_k}{\sqrt{2}}\phi_k(x),$$

we also have

$$\widetilde{\mathscr{F}}\{\widehat{K}f\} = \frac{\mu_j^2\widetilde{\mathscr{F}}_j + \mu_k^2\widetilde{\mathscr{F}}_k}{\mu_j^2 + \mu_k^2} = \widetilde{\mathscr{F}_j} + \frac{\mu_k^2}{\mu_j^2 + \mu_k^2}(\widetilde{\mathscr{F}_k} - \widetilde{\mathscr{F}_j}).$$

But, by our assumption, $\mu_k^2 > \mu_j^2$. Hence

$$\frac{\mu_k^2}{\mu_j^2 + \mu_k^2} > \frac{1}{2}.$$

Besides this, since the irreducible components are ordered in accordance with the values of the functional, $\widetilde{\mathscr{F}}_k - \widetilde{\mathscr{F}}_j > 0$. Thus

$$\widetilde{\mathscr{F}}\{\widehat{K}f\} > \widetilde{\mathscr{F}}_j + \frac{1}{2}(\widetilde{\mathscr{F}}_k - \widetilde{\mathscr{F}}_j) = \frac{1}{2}(\widetilde{\mathscr{F}}_k + \widetilde{\mathscr{F}}_j) = \widetilde{\mathscr{F}}\{f\}.$$

Hence, if the condition of the theorem is not satisfied, then the operator K is not a bettering operator.

b) Proof of sufficiency. Let the conditions of the theorem be satisfied. Consider an arbitrary function $f(x) = \sum_{s=0}^m c_s \phi^s(x)$ and the number

$$\Delta := \widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}\{\widehat{K}f\}.$$

We need to show that $\Delta > 0$. Indeed, using (53) we obtain

$$\Delta = \widetilde{\mathscr{F}}\{f\} - \frac{\sum_{s=0}^m \mu_s^2 c_s^2 \widetilde{\mathscr{F}}_s}{\sum_{s=0}^m \mu_s^2 c_s^2} = \frac{\sum_{s=0}^m \mu_s^2 c_s^2 \Big[\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big]}{\sum_{s=0}^m \mu_s^2 c_s^2}.$$

Since the sequence $\widetilde{\mathscr{F}}_s$ is increasing and $0 = \widetilde{\mathscr{F}}_0 \le \widetilde{\mathscr{F}}\{f\} \le \widetilde{\mathscr{F}}_m$, there exists k such that

$$\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_{k-1} \ge 0, \quad \widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_k \le 0.$$

Hence, we can write

$$\Delta = \frac{1}{\sum_{s=0}^m \mu_s^2 c_s^2} \left[\sum_{s=0}^{k-1} \mu_s^2 c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) - \sum_{s=k}^{m} \mu_s^2 c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) \right], \tag{74}$$

where the summands in both sums are non-negative.

By our assumptions $\mu_s^2 \ge \mu_k^2$ for $s \le k-1$ and $\mu_s^2 \le \mu_k^2$ for $s \ge k$. Hence,

$$\sum_{s=0}^{k-1} \mu_s^2 c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) \ge \mu_k^2 \sum_{s=0}^{k-1} c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big)$$
$$\sum_{s=k}^{m} \mu_s^2 c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) \le \mu_k^2 \sum_{s=k}^{m} c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big).$$

Substituting these inequalities into (74) we obtain

$$\Delta \ge \frac{\mu_k^2}{\sum_{s=0}^m c_s^2} \left[\sum_{s=0}^{k-1} c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) - \sum_{s=k}^{m} c_s^2 \Big(\widetilde{\mathscr{F}}\{f\} - \widetilde{\mathscr{F}}_s \Big) \right]$$
$$= \mu_k^2 \frac{\sum_{s=0}^m c_s^2}{\sum_{s=0}^m \mu_s^2 c_s^2} \left(\widetilde{\mathscr{F}}\{f\} - \frac{\sum_{s=k}^m c_s^2 \widetilde{\mathscr{F}}_s}{\sum_{s=k}^m c_s^2} \right). \tag{75}$$

By (53) the last quantity in parentheses is equal to zero, and therefore $\Delta \ge 0$. This proves the sufficiency of the theorem. Theorem V is proven completely. □

Theorem V establishes those additional restrictions on the choice of the potential function, which was discussed above. Namely, in connection with the fact that the operator $\widehat{K}$ must be bettering, the coefficients μ_s in decomposition (69) must not only be nonnegative, but also nonincreasing: $\mu_0 \geq \mu_1 \geq \cdots \geq \mu_m$.

In the conclusion of this point we will summarize those reasons which must be considered when choosing a potential function $K(x, y)$ in symmetric spaces:

(i) It is expedient to choose the potential function $K(x, y)$ as a function $K(\rho(x, y))$, depending only on the distance $\rho(x, y)$.
(ii) This function can be defined by the decomposition (69) into a series in the system of functions $K_s(\rho)$. The system of functions $K_s(\rho)$ is uniquely defined for a given space X.
(iii) The coefficients μ_s in decomposition (69) must be positive.
(iv) If we order the coefficients μ_s in accordance with (68), then the sequence $\mu_0, \mu_1, \ldots, \mu_m$ must be monotonically decreasing.

In the cases when the potential function $K(\rho)$ is initially defined in closed form (for example, with expressions $\widehat{K}(\rho) = e^{-\alpha\rho^2}$, $K(\rho) = 1/(1 + \alpha\rho^2)$ and so forth) to check properties (iii) and (iv) one can calculate the coefficients μ_s, making use of formula (61). Functions $K(\rho)$, for which properties (70) and (71) are not satisfied, are certainly not suitable as potential functions.

The practice of applying potential functions shows that the results of using the method depend little on how the coefficients μ_s are chosen, if the satisfy the above restrictions.

6. **On the choice of a potential function in the space of vertices of an m-dimensional cube.** Out of the various symmetric spaces which we encounter in practice, the most significant is the space of vertices of an m-dimensional cube. We encounter such a space, for example, in the recognition of black-and-white images. In the present paragraph we will show how the theory outlined above is applied in this concrete space. It will be convenient for us to consider that the coordinates $x_1, \ldots, x_m$ of the cubes vertices taking values ± 1, i. e. that the center of the cube is located at the origin, and the edge has euclidean length 2. As the distance between the vertices $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ we use the usual *Hamming distance*

$$\rho(x, y) = \frac{1}{4}\sum_{i=1}^{m}(x_i - y_i)^2 = \frac{1}{2}\left(m - \sum_{i=1}^{m} x_i y_i\right), \tag{76}$$

equal to the number of distinct places in the codes of the vertices. The space of vertices of an m-dimensional cube with metric (76) is from here on called a *Hamming space*. This space contains $N = 2^m$ points. A Hamming space is a symmetric space, as one can easily check.

Let us now find out, in what way the linear space $\mathscr{L}$ of functions on the Hamming space decomposes into irreducible components L_s.

Consider the system of functions in $\mathscr{L}$, consisting of the constant function $\phi_0 = 1/2^{m/2}$ and functions:

$$\left.\begin{array}{cccc} \frac{1}{2^{m/2}}x_1, & \frac{1}{2^{m/2}}x_2, \dots, & \frac{1}{2^{m/2}}x_i, \dots, & \frac{1}{2^{m/2}}x_m; \\ \frac{1}{2^{m/2}}x_1x_2, & \frac{1}{2^{m/2}}x_1x_3, \dots, & \frac{1}{2^{m/2}}x_ix_j, \dots, & \frac{1}{2^{m/2}}x_{m-1}x_m; \\ \dots & \dots\dots & \dots\dots & \dots \\ & & \frac{1}{2^{m/2}}x_1x_2\cdots x_m & \end{array}\right\} \qquad (77)$$

so that the functions, written in the s-th line, have the form

$$\phi_{i_1 i_2 \dots i_s} = \frac{1}{2^{m/2}} x_{i_1} x_{i_1} \cdots x_{i_s},$$

and the indices $i_1, i_2, \dots, i_s$ take all possible values from 1 to m that satisfy the condition $i_1 < i_2 < \cdots < i_s$. Thus the s-th line contains exactly $\binom{m}{s}$ functions, and their total number (including the constant function) equals 2^m, i. e. equals the number N of points of the Hamming space. Bearing in mind, that the above system of functions is orthonormal, we conclude, that it is the orthonormal basis in linear space $\mathscr{L}$ of functions on the Hamming space.

One can easily check that the functions written in the s-th line (77) belong to one irreducible component $\mathscr{L}_s$ and make up its basis. It follows that the number of irreducible components (including the component $\mathscr{L}_0$ consistent of constant functions) is equal to $m + 1$, i.e. in accordance with the general theory, to the number of distinct distances in the examined symmetric space. The dimension of the s-th component $\mathscr{L}_s$ equals $N_s = \binom{m}{s}$.

Next we calculate the functions of $K_s(\rho)$ for a Hamming space.

For the component $\mathscr{L}_0$

$$K_0(\rho) = \frac{1}{2^{m/2}} \cdot \frac{1}{2^{m/2}} = \frac{1}{2^m}.$$

For component $\mathscr{L}_s$ we have

$$K_s\big(\rho(x, y)\big) = \frac{1}{2^m} \sum_{i_1 < \cdots < i_s} x_{i_1} \cdots x_{i_s} \cdot y_{i_1} \cdots y_{i_s}.$$

Set $z_i := x_i y_i$ $(i = 1, \dots, m)$. Then

$$K_s\big(\rho(x, y)\big) = \frac{1}{2^m} \sum_{i_1 < \cdots < i_s} z_{i_1} \cdots z_{i_s}. \qquad (78)$$

If the distance between x and y equals ρ, then in accordance with (76) there will be ρ negative (-1) and $m - \rho$ positive (+1) values among z_i. The summands in (78) can be broken into groups with j positive and $s - j$ negative z_i's. In both cases $0 \le j \le \min\{s, \rho\}$.

Every such summand equals $(-1)^j$. Their number equals $\binom{\rho}{j}\binom{m-\rho}{s-j}$. Now, summing over j, we will get

$$K_s(\rho) = \frac{1}{2^m} \sum_{j=0}^{\min\{s,\rho\}} \binom{\rho}{j}\binom{m-\rho}{s-j}(-1)^j. \tag{79}$$

With this formula one can calculate, in particular,

$$K_1(\rho) = \frac{1}{2^m}(m-2\rho),$$
$$K_2(\rho) = \frac{1}{2^m}\frac{(m-2\rho)^2-m}{2},$$

etc. Also

$$K_s(1) = \frac{1}{2^m} \sum_{j=0}^{1} \binom{\rho}{j}\binom{m-\rho}{s-j}(-1)^j = \left(\binom{m-1}{s} - \binom{m-1}{s-1}\right)\frac{1}{2^m}$$
$$K_s(0) = \frac{1}{2^m}\binom{m}{s}.$$

Thus

$$\frac{K_s(1)}{K_s(0)} = 1 - \frac{2s}{m}.$$

and, consequently, by (67) the value $\widetilde{\mathscr{F}}_s$ of the quality functional $\widetilde{\mathscr{F}}$ with kernel (65) on functions $f \in \mathscr{L}_s$ equals

$$\widetilde{\mathscr{F}}_s = \frac{s}{m}. \tag{80}$$

Formula (80) shows that the values of the functional increase with the growth of s. This is in full compliance with our intuitive understanding of the change of complexity of functions (77) when moving from the top to the bottom lines in (80). Indeed, one can show that each function, in the s-th line, possesses the following property: among the m vertices of the cube that are located at the minimal distance $\rho = 1$ from any fixed vertex x^*, there are exactly s vertices in which the values of the function differ by a sign from its value in x^*; in the remaining $m - s$ adjacent vertices the values of the function coincide with the value in x^*. The modules of the values of all the functions in (80) are the same in all vertices and equal $1/2^{m/2}$.

In accordance with the note at the end of section 4, the values (80) of the quality functional characterize the complexity of the functions $K_s(\rho)$. With the growth of the number s the function $K_s(\rho)$ becomes more complex. In this case function $K_s(\rho)$ is a polynomial of order s in ρ, and, correspondingly, with the growth of s grows the number of its sign changes, extrema and other intuitive indicators of complexity.

Let us now turn to the question of the decomposition a function of distance (in particular, a potential function), defined in a Hamming space into a linear combination of functions $K_s(\rho)$. Here, of course, we can use formula (61), where the function $S(rho)$ (the number of points on a sphere of radius ρ) in this case, as can be easily shown, has the form

$$S(\rho) = \binom{m}{\rho}.$$

However, the practical use of formula (61) leads to difficult calculations. In a number of cases it is possible to calculate (precisely or approximately) the decomposition coefficients μ_s without resorting to a direct calculation by formula (61), but using the following identity:

$$\left(\frac{1-u}{2}\right)^{\rho} \cdot \left(\frac{1+u}{2}\right)^{m-\rho} = \sum_{s=0}^{m} K_s(\rho) u^s, \tag{81}$$

whose proof is left to the reader (see page 134 of the Russian edition for a detailed proof).

As an example of a precise calculation of coefficients μ_s we use this identity for the decomposition of function $K(\rho) = e - \alpha\rho$ (where α is some constant). With this goal we choose the value of u in (81) so that

$$u = \frac{e^\alpha - 1}{e^\alpha + 1} \quad \Leftrightarrow \quad \ln \frac{1+u}{1-u} = \alpha.$$

Substituting this value of U in (81) we obtain

$$e^{-\alpha\rho} = \sum_{s=0}^{m} (1 - e^{-\alpha})^s (1 + e^{-\alpha})^{m-s} K_s(\rho).$$

Thus for $K(\rho) = e^{-\alpha\rho}$ the decomposition coefficients μ_s are given by

$$\mu_s = (1 - e^{-\alpha})^s (1 + e^{-\alpha})^{m-s}. \tag{82}$$

This, in particular, shows that (in the Hamming space) the function $e - \alpha\rho$ can be used as a potential function.

We will now show, how formula (81) can be used for an asymptotic (for $m \to \infty$) evaluation of the decomposition coefficients of one quite broad class of functions of distance. Namely, we consider functions of the form

$$K(\rho) = f(\rho/m),$$

where $f(z)$ is a sufficiently smooth function, defined on the segment $0 \leq z \leq 1$. To calculate the decomposition coefficients μ_s of function $K(\rho)$ we will multiply both sides of formula (81) by

$$K(\rho)S(\rho) = f(\rho/m) \binom{m}{\rho}.$$

Summing over ρ between 0 and m, recalling (61) and setting $z = (1-u)/2$, we get:

$$\sum_{\rho=0}^{m} \binom{m}{\rho} z^{\rho}(1-z)^{m-\rho} f(\rho/m) \;=\; \sum_{s=0}^{m} (-2)^s \left(z - \frac{1}{2}\right)^s \binom{m}{s} \frac{\mu_s}{2^m}. \tag{83}$$

The left hand side of this expression is a polynomial of S.N. Bernstein of function $f(z)$. It is known that this polynomial approximates function $f(z)$ for $m \to \infty$ uniformly on the segment $0 \le z \le 1$ and therefore

$$f(z) \sim \sum_{s=0}^{m} (-2)^s \left(z - \frac{1}{2}\right)^s \binom{m}{s} \frac{\mu_s}{2^m}, \tag{84}$$

and the error diminishes with the growth of m uniformly in z, for example, as $1\sqrt{m}$, if we demand only the continuity of $f(z)$, and as $1/m$, if $f(z)$ is twice differentiable. Suppose now that $f(z)$ has a Taylor series expansion in a neighborhood of $z = 1/2$. Then, comparing this Taylor series with the right hand side of (83), we get the expression for the coefficients μ_s:

$$2^{-m} \binom{m}{s} \mu_s \sim \frac{1}{s!(-2)^s} f^{(s)}\left(\frac{1}{2}\right), \tag{85}$$

which is asymptotically accurate for $m \to \infty$. However, for a finite m it makes sense to use formula (85) only for relatively small values of s, since with the growth of s the right side of (85) becomes comparable with the error of this formula. We can easily assess the errors for the use of formula (85), if we use the well-known results concerning the evaluation of approximation precision by the polynomials of S.N. Bernstein.

Formula (85) allows us to check, whether or not the function $K(\rho) = f(\rho/m)$ can serve as a potential function in a Hamming space.