A Geometric Approach to Linear Cryptanalysis

Tim Beyne

imec-COSIC, ESAT, KU Leuven, Belgium
name.lastname@esat.kuleuven.be

Abstract. A new interpretation of linear cryptanalysis is proposed. This 'geometric approach' unifies all common variants of linear cryptanalysis, reveals links between various properties, and suggests additional generalizations. For example, new insights into invariants corresponding to non-real eigenvalues of correlation matrices and a generalization of the link between zero-correlation and integral attacks are obtained. Geometric intuition leads to a fixed-key motivation for the piling-up principle, which is illustrated by explaining and generalizing previous results relating invariants and linear approximations. Rank-one approximations are proposed to analyze cell-oriented ciphers, and used to resolve an open problem posed by Beierle, Canteaut and Leander at FSE 2019. In particular, it is shown how such approximations can be analyzed automatically using Riemannian optimization.

Keywords: Linear cryptanalysis · Nonlinear cryptanalysis · Piling-up lemma · Correlation matrices · Block cipher invariants

1 Introduction

At EUROCRYPT 1993, Matsui [31] introduced linear cryptanalysis as a new known-plaintext attack on the block cipher DES. Linear cryptanalysis is based on probabilistic linear relations or *linear approximations*, a concept introduced by Tardy-Corfdir and Gilbert [36].

The success of Matsui's attack led to the development of a myriad of extensions and variants of linear approximations, and to more advanced techniques for their analysis [16, 32]. Despite significant advances, many questions related to linear cryptanalysis and its theoretical foundations remain unresolved.

Kaliski and Robshaw [25] suggested using multiple linear approximations. Hermelin, Cho and Nyberg [23] proposed the related multidimensional linear attack. Both extensions are widely used. Generalizations of linear cryptanalysis to groups other than $\mathbb{F}_2^n$ were proposed by Granboulan, Levieil and Piret [20] and Baignères, Stern and Vaudenay [3]. The use of nonlinear approximations is another natural extension, and has been attempted by Knudsen and Robshaw [26], Harpes, Kramer and Massey [21] with I/O sums, Harpes and Massey [22] with partitioning attacks and recently by Beierle, Canteaut and Leander [4].

All of the above techniques rely on heuristic methods to glue together several approximations over multiple rounds of a cipher. These methods will be collectively referred to as the *piling-up principle*. This principle has traditionally been justified using independence or Markov chain assumptions [2, 31, 42], which can be related to earlier work on Markov ciphers in the context of differential cryptanalysis [28]. However, such assumptions are hard to reconcile with the key-dependence of approximations and the increased importance of cryptographic permutations. In fact, key-dependence is one of the fundamental difficulties of nonlinear cryptanalysis. Alternatively, the correlation matrix framework of Daemen *et al.* [16] is more suitable for the fixed-key setting. It motivates the piling-up principle using the dominant trail hypothesis. Beierle *et al.* [4] extend this approach by applying linear cryptanalysis to a nonlinearly transformed variant of the cipher.

In a different direction, Rijmen and Bogdanov [13] introduced zero-correlation linear cryptanalysis to exploit unbiased linear approximations. The construction of zero-correlation approximations relies on the miss-in-the-middle technique as opposed to the piling-up principle. At ASIACRYPT 2012, Bogdanov *et al.* [12] established a link between multidimensional zero-correlation approximations and integral distinguishers [27].

Finally, several lightweight block ciphers have been found vulnerable to weak-key attacks based on invariant subspaces [30] and nonlinear invariants [39]. These attacks have led to renewed interest in linear cryptanalysis and its generalizations. Abdelraheem *et al.* [1] found links between invariant subspaces and linear cryptanalysis. Moreover, nonlinear invariants provide one of the most compelling examples of nonlinearity in cryptanalysis, with applications including the analysis of SCREAM, iSCREAM, Midori-64 and MANTIS [6, 39]. At ASIACRYPT 2018, it was shown that invariant subspaces and nonlinear invariants can be described as eigenvectors of correlation matrices [6]. Furthermore, one of the invariants discovered in [6] corresponds to a perfect linear approximation. These results established a strong link between nonlinear invariants and linear cryptanalysis, but a true statistical generalization of the nonlinear invariant attack was left open. Lastly, Beierle *et al.* [4] extended the links discovered by Abdelraheem *et al.* to some classes of nonlinear invariants.

Contribution. A conceptually new way of thinking about linear cryptanalysis is introduced. It provides an alternative viewpoint for the foundations of linear cryptanalysis and has a number of concrete benefits. Firstly, it results in a systematized and unified description of the above-mentioned variants of linear cryptanalysis. Secondly, it leads to generalizations of the connections between these attacks, such as the link between integral and zero-correlation cryptanalysis and the links between invariants and linear approximations. Some of these results are illustrated in Table 1, and are discussed in more detail below. Thirdly, it suggests a general form of the piling-up principle. Finally, to illustrate the relevance for the working cryptanalyst, the approach is used to solve a problem posed by Beierle *et al.* [4].

Section 3 introduces a correspondence between cryptanalytic properties and vector spaces of complex-valued functions on the domain of a primitive. This results in a uniform description of the properties (sets, linear and nonlinear

Boolean functions, ...) that are used in different variants of linear cryptanalysis. The correspondence generalizes the idea introduced in [6] that invariant subspaces and nonlinear invariants can be represented by complex vectors, which led to their characterization as eigenvectors of correlation matrices.

Table 1: Approximations for a function F from the geometric viewpoint. Here, U and V are vector spaces (of dimension d) of functions. The notation follows Sections 3 to 5.

	Zero-correlation $C^{\mathsf{F}}U \perp V$ $\overset{\text{Thm. 4.2}}{\rightleftarrows}$	Perfect $C^{\mathsf{F}}U \subseteq V$ $\xrightarrow{\text{Sect. 5.3}}$	General $\langle V, U\rangle_{\mathsf{F}}$
$d = 1$	Linear zero-correlation [13] Nonlinear zero-correlation (Ex. 4.3)	Invariant subspaces [30] Nonlinear invariants [39] Eigenvectors of C^{F} [6]	Linear cryptanalysis [31] Abelian groups [3] I/O sums [21] Beierle *et al.* [4] Rank-one (Section 6)
$d \geq 1$	Multidimensional zero-correlation [12]	Integral attacks [27] General invariants (Def. 4.3, Ex. 4.2)	Multiple linear [9, 25] Multidim. linear [23] Partitioning [22] Projection, χ^2 [2, 41, 42]

Definition 4.1 characterizes an approximation of a cipher as a pair of vector spaces (U, V), corresponding to input and output properties as sketched above. This results in a systematization of many variants of linear cryptanalysis, as summarized in Table 1. It will be shown that the type and quality of approximations is related to the geometric properties of the spaces U and V. Section 4.1 illustrates how this results in new insight into block cipher invariants and gives a realistic example of invariants related to non-real eigenvalues of correlation matrices, a problem that was left open at ASIACRYPT 2018 [6]. Theorem 4.2 generalizes the links between zero-correlation and integral attacks discovered by Bogdanov *et al.* [12]. For general approximations, *principal correlations* are introduced as a natural extension of the correlation of a linear approximation and it is shown how they relate to the complexity of optimal distinguishers discussed by Baignères, Junod and Vaudenay [2].

A general piling-up principle is given in Theorem 5.1. Its motivation is the result of geometric intuition. This avoids independence and Markov chain assumptions and simplifies working with fixed keys. Furthermore, the result evades the issues that are encountered when the dominant-trail approach of Daemen *et al.* is extended to the nonlinear case. Theorem 5.1 allows for much greater flexibility than previous formulations of the piling-up principle. In particular, it becomes possible to build trails that combine diverse cryptanalytic properties. This is illustrated in Section 5.3 by strengthening the links between linear approximations and invariants, extending previous work by Abdelraheem *et al.* [1] and Beierle *et al.* [4].

Finally, Section 6 introduces rank-one approximations to analyze cell-oriented ciphers. A tool to find optimal rank-one trails is introduced, and its application to searching for invariants is discussed. Perhaps surprisingly, the tool is based on numerical optimization on a Riemannian manifold. This is enabled by the generality of Sections 3 to 5, which relaxes the search space by introducing many new types of approximations. The tool is provided as supplementary material. Rank-one approximations and the aforementioned tool are used in Section 7.3 to resolve a problem introduced by Beierle *et al.* [4], who describe it as “a major open problem”. It is representative of other concrete problems, and its solution relies on the general techniques that are introduced in Sections 3 to 5.

2 Functions on Abelian Groups

The goal of this section is to introduce several concepts that will be used to develop a general theory of linear cryptanalysis in Sections 3 to 5. These concepts provide the setting for the proposed geometric approach. It is assumed that the reader is familiar with finite Abelian groups and linear algebra in finite-dimensional inner product spaces.

It will be shown in Section 3 that many cryptanalytic properties can be described by complex-valued functions on the domain of a primitive. Section 2.1 discusses preliminaries related to the set of such functions. Section 2.2 introduces the Fourier transformation on finite Abelian groups. This will be an important tool to simplify the effect of constant (including key) additions. Finally, Section 2.3 discusses the geometry of subspaces of an inner product space.

2.1 Inner Product Space of Functions

Let G be a finite Abelian group, for example the domain of a block cipher. In fact, all of the properties in this section are valid for any *set* G. However, the results in Section 2.2 will require the assumption that G is a finite Abelian group. The $\mathbb{C}$-vector space of all functions from G to $\mathbb{C}$, with the usual pointwise addition and scalar multiplication, will be denoted by $\mathbb{C}G$. The standard inner product between two functions $f, g \in \mathbb{C}G$ is defined by

$$\langle f, g \rangle = \sum_{x \in G} \overline{f(x)} g(x),$$

where $\overline{f(x)}$ denotes the complex-conjugate of $f(x)$. Hence, the vector space $\mathbb{C}G$ is a finite-dimensional inner product space. One also has a norm $\|f\|_2 = \sqrt{\langle f, f \rangle}$, which carries the geometric interpretation of length. The modulus of the inner product between two normalized vectors can be interpreted as the cosine of the smallest angle enclosed by them – although for non-real vectors, several definitions of angles are plausible. The theory developed in Sections 4 and 5 will draw on these geometric concepts for intuition.

The functions δ_x, which are equal to one at $x \in G$ and zero everywhere else, clearly form an orthonormal basis for $\mathbb{C}G$. This basis will be referred to as the *standard basis*. It follows that $\mathbb{C}G$ is isomorphic to $\mathbb{C}^{|G|}$ as an inner product space.

Example 2.1. The indicator function $\mathbb{1}_S : G \to \mathbb{C}$ of a set $S \subseteq G$ is defined by $\mathbb{1}_S(x) = 1$ if $x \in S$ and zero elsewhere. The coordinates of $\mathbb{1}_S$ in the standard basis are given by $\langle \delta_x, \mathbb{1}_S \rangle = \mathbb{1}_S(x)$ for $x \in G$. Given a second set $T \subseteq G$, it holds that $\langle \mathbb{1}_S, \mathbb{1}_T \rangle = |S \cap T|$. One reason to consider indicator functions such as $\mathbb{1}_S$ and $\mathbb{1}_T$ as complex-valued rather than real-valued functions is that $\mathbb{C}$ is algebraically closed. This will be convenient in Section 2.2 below, where the Fourier transformation of such functions is introduced.

Recall that the tensor product of $\mathbb{C}$-vector spaces $V_1, \dots, V_n$ is another $\mathbb{C}$-vector space $V_1 \otimes \cdots \otimes V_n$ of dimension $\prod_{i=1}^n \dim V_i$ together with a multilinear map $\otimes : \prod_{i=1}^n V_i \to \bigotimes_{i=1}^n V_i$, which has the universal property that it uniquely linearizes arbitrary multilinear maps. Specifically, for any $T : \prod_{i=1}^n V_i \to W$ linear in each variable (multilinear), there exists a unique *linear* map $L : \bigotimes_{i=1}^n V_i \to W$ such that $T(v_1, \dots, v_n) = L(v_1 \otimes \cdots \otimes v_n)$.

For the purposes of this paper, readers who are not familiar with tensor products may take the following characterization as a definition. Let $G = A \oplus B$ be a direct sum of Abelian groups A and B. That is, the group G consists of all pairs (a, b) with $a \in A$ and $b \in B$. The tensor product of $\mathbb{C}A$ and $\mathbb{C}B$ can then be characterized by $\mathbb{C}A \otimes \mathbb{C}B \cong \mathbb{C}G$. Indeed, the linear map defined by $\delta_{(a,b)} \mapsto \delta_a \otimes \delta_b$ for all $a \in A$ and $b \in B$ is an isomorphism. In this paper, $\mathbb{C}G$ and $\mathbb{C}A \otimes \mathbb{C}B$ will always be identified through this isomorphism. Hence, for $f \in \mathbb{C}A$ and $g \in \mathbb{C}B$, it can be said that $f \otimes g \in \mathbb{C}G$ with $(f \otimes g)(a, b) = f(a)\, g(b)$.

A rank-one vector $v \in \bigotimes_{i=1}^n V_i$ is a vector of the form $v = v_1 \otimes \cdots \otimes v_n$. Given bases for $V_1, \dots, V_n$, the set of all their tensor products is a basis of rank-one vectors for $\bigotimes_{i=1}^n V_i$. More generally, for any vector v there exists an integer $r \geq 0$ such that

$$v = \sum_{i=1}^{r} \lambda_i \bigotimes\nolimits_{j=1}^{n} v_{i,j},$$

for some vectors $v_{i,j} \in V_j$ and scalars $\lambda_i \in \mathbb{C}$. The smallest r for which such a decomposition exists is called the tensor rank of v.

Example 2.2. Let $G = \mathbb{F}_2^2 = \mathbb{F}_2 \oplus \mathbb{F}_2$. The vector $\delta_{(0,0)} = \delta_0 \otimes \delta_0$ in $\mathbb{C}\mathbb{F}_2 \otimes \mathbb{C}\mathbb{F}_2$ has tensor rank one. Furthermore, it is easy to check that the vector $\delta_{(0,0)} + \delta_{(1,1)} = \delta_0 \otimes \delta_0 + \delta_1 \otimes \delta_1$ has rank two. However, the vector $\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)}$ only has rank one because it is equal to $(\delta_0 + \delta_1) \otimes (\delta_0 + \delta_1)$.

2.2 Fourier Analysis

Given a function $f \in \mathbb{C}G$ and a constant $t \in G$, one can define a new function by $x \mapsto f(x + t)$. The effect of translations on the coordinates of functions in the standard basis of $\mathbb{C}G$ is inconvenient: the basis vectors are shuffled around by the permutation $\delta_x \mapsto \delta_{x+t}$, which corresponds to multiplication by a Toeplitz matrix. It would be more convenient if the effect of translation would be a simple scaling of the coordinates, *i.e.* multiplication by a diagonal matrix. This can be achieved by working with respect to a different basis.

To achieve the goal of diagonalization, the new basis vectors should be eigenvectors of the set of translation operations. This is achieved for any homomorphism $\chi : G \to \mathbb{C}^\times$ from G to the multiplicative group of complex numbers $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$, since $\chi(x+t) = \chi(t)\chi(x)$ for any $x, t \in G$. This leads to the following definition.

Definition 2.1 (Group characters [37]). *Let G be a finite Abelian group. A (complex) character of G is a group homomorphism $G \to \mathbb{C}^\times$. The (Pontryagin) dual of G is the group $\widehat{G}$ of all characters of G with respect to the pointwise product.*

It is not hard to see that $\widehat{G}$ is indeed an Abelian group. For example, the inverse of $\chi \in \widehat{G}$ is the character $x \mapsto \chi(-x)$. That is, $\chi(-x) = \overline{\chi(x)}$.

Example 2.3. The dual of the additive group $\mathbb{F}_2$ is $\widehat{\mathbb{F}}_2 = \{x \mapsto 1, x \mapsto (-1)^x\}$. Indeed, these are the only two group homomorphisms $\mathbb{F}_2 \to \mathbb{C}^\times$. ▷

The functions in the dual group $\widehat{G}$ form a basis for $\mathbb{C}G$ that behaves well with respect to translation. Further properties of the dual group are given in Theorem 2.1 below. In particular, property (2) shows that the basis of characters is orthogonal.

Theorem 2.1 (Properties of dual groups [37]). *If G is a finite Abelian group with dual $\widehat{G}$, then*

(1) The dual group $\widehat{G}$ is isomorphic to G.

(2) For all $\chi, \psi \in \widehat{G}$, it holds that $\langle \chi, \psi \rangle = |G|\, \delta_\chi(\psi)$.

(3) If $G = H_1 \oplus H_2$ with $\oplus$ the internal direct sum, then $\widehat{G} = \widehat{H}_1 \oplus \widehat{H}_2$.

By Theorem 2.1 (1), $\widehat{G}$ can be identified with G. In general, this identification is not unique. However, there is a *functorial* isomorphism between the double dual of G and G itself, which identifies $g \in G$ with the evaluation map $\chi \mapsto \chi(g)$ in the dual of $\widehat{G}$ [37]. This result justifies the term 'dual group'. In order to avoid arbitrary choices, isomorphisms between $\widehat{G}$ and G will be avoided throughout this paper. This makes no difference in specific calculations, but it is theoretically more elegant.

Example 2.4. Since the additive group $\mathbb{F}_2^n$ is the direct sum of n copies of $\mathbb{F}_2$, it follows from Theorem 2.1 (3) that the dual group is essentially the direct sum of n copies of $\widehat{\mathbb{F}}_2$. Specifically, $\widehat{\mathbb{F}}_2^n = \{x \mapsto \prod_{i=1}^n (-1)^{u_i x_i} = (-1)^{u^\top x} \mid u \in \mathbb{F}_2^n\}$. Note that identifying $\widehat{\mathbb{F}}_2^n$ and $\mathbb{F}_2^n$ requires choosing a basis for $\mathbb{F}_2^n$. ▷

The Fourier transformation $\mathcal{F}$ is essentially a change of basis from the standard basis to the character basis. However, in order to avoid identifying $\widehat{G}$ and G, we shall define $\mathcal{F}$ as a transformation from $\mathbb{C}G$ to $\mathbb{C}\widehat{G}$. With this definition, the Fourier transformation maps a character $\chi \in \widehat{G} \subset \mathbb{C}G$ directly to a multiple of the standard basis vector $\delta_\chi \in \mathbb{C}\widehat{G}$. Since group characters are orthogonal, Definition 2.2 achieves the desired basis transformation.

Definition 2.2 (Fourier transformation [37]). *Let $f : G \to \mathbb{C}$ be a function. The Fourier transformation of f is the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by*

$$\widehat{f}(\chi) = \langle \chi, f \rangle = \sum_{x \in G} \overline{\chi(x)} f(x).$$

The Fourier transformation is the map $\mathcal{F} : \mathbb{C}G \to \mathbb{C}\widehat{G}$ such that $\mathcal{F}f = \widehat{f}$.

The transformation $\mathcal{F}$ is a vector space isomorphism. In fact, since $\mathbb{C}G$ and $\mathbb{C}\widehat{G}$ are algebras with either the pointwise product or convolution, $\mathcal{F}$ is an isomorphism of algebras which swaps the pointwise product and convolution. This is by construction, since the set of convolution operators is generated by translations.

The vector space $\mathbb{C}\widehat{G}$ is also an inner product space. In fact, due to the orthogonality of characters, the inner product between $f_1, f_2 \in \mathbb{C}G$ coincides with the inner product of their Fourier transforms up to a constant factor:

$$\langle \widehat{f}_1, \widehat{f}_2 \rangle = \sum_{\chi \in \widehat{G}} \overline{\widehat{f}_1(\chi)} \widehat{f}_2(\chi) = |G| \, \langle f_1, f_2 \rangle.$$

In other words, $\mathcal{F}/\sqrt{|G|}$ is a unitary map and $\mathcal{F}^{-1} = \mathcal{F}^*/|G|$ with $\mathcal{F}^*$ the adjoint (conjugate transpose) of $\mathcal{F}$.

To end this section, consider the case $G = A \oplus B$. As mentioned above, one has $\mathbb{C}G = \mathbb{C}A \otimes \mathbb{C}B$ (technically up to isomorphism). By Theorem 2.1 (3), the dual group satisfies $\widehat{G} = \widehat{A} \oplus \widehat{B}$. Hence, one also has $\mathbb{C}\widehat{G} = \mathbb{C}\widehat{A} \otimes \mathbb{C}\widehat{B}$. Consequently, the Fourier transformation on $\mathbb{C}G$ is given by $\mathcal{F}_A \otimes \mathcal{F}_B$. Equivalently, the matrix representation of $\mathcal{F}$ in the standard basis is the Kronecker product of the matrix representations of $\mathcal{F}_A$ and $\mathcal{F}_B$ in the standard basis.

2.3 Subspaces of $\mathbb{C}G$ and $\mathbb{C}\widehat{G}$

Sections 3 and 4 will demonstrate that subspaces of $\mathbb{C}G$ and $\mathbb{C}\widehat{G}$ are often more interesting for cryptanalysis than individual functions. For this reason, it will be convenient to extend the inner product notation $\langle \cdot, \cdot \rangle$ to subspaces of $\mathbb{C}G$. For subspaces $U \subseteq \mathbb{C}G$ and $V \subseteq \mathbb{C}G$, define the linear map $\langle V, U \rangle : U \to V$ by $\langle V, U \rangle = \pi_V \, \iota_U$, where $\iota_U : U \to \mathbb{C}G$ is the inclusion map and $\pi_V : \mathbb{C}G \to V$ is the orthogonal projection on V. A similar definition can be given for subspaces of $\mathbb{C}\widehat{G}$. Note that $\langle V, U \rangle = \langle U, V \rangle^*$ since projection and inclusion are adjoint.

Example 2.5. Let U and V be one-dimensional subspaces of $\mathbb{C}G$ spanned by unit-norm vectors u and v respectively. By definition, $\iota_U(\lambda u) = \lambda u$ and $\pi_V(x) = v\langle v, x \rangle$. Consequently, $\langle V, U \rangle : U \to V$ is the map $\lambda u \mapsto \langle v, u \rangle \lambda v$. The matrix representation of this map is thus simply the 1×1 matrix containing the inner product $\langle v, u \rangle$. ▷

The transformation $\langle V, U \rangle$ comes with a geometric interpretation, which will be important in Sections 4 and 5. Due to standard properties of orthogonal projection, $\langle V, U \rangle$ maps any $u \in U$ to the nearest vector $v \in V$. In addition, no other vector in V of the same length makes a smaller angle to u than v. This suggests that $\langle V, U \rangle$ encodes all information about the 'angles' between U and V. This can be made precise using the notion of principal angles between subspaces, which is due to Jordan [24]. The characterization below follows Björck and Golub [10].

Definition 2.3 (Principal angles). *Let U and V be subspaces of an inner product space over $\mathbb{C}$ of finite dimension and let $d = \min\{\dim U, \dim V\}$. The principal angles $0 \leq \theta_1 \leq \ldots \leq \theta_d \leq \pi/2$ between U and V are recursively defined by (for $i = 1, 2, \ldots, d$)*

$$\cos\theta_i = \frac{\langle u_i, v_i \rangle}{\|u_i\|_2 \|v_i\|_2} = \max_{\substack{u \in U_i \setminus \{0\} \\ v \in V_i \setminus \{0\}}} \frac{|\langle u, v \rangle|}{\|u\|_2 \|v\|_2}\,,$$

where $u_i \in U_i$ and $v_i \in V_i$ are nonzero vectors for which the maximum on the right is achieved with $\langle u_i, v_i \rangle$ a non-negative real number, $U_i = U \cap \{u_1, \ldots, u_{i-1}\}^\perp$ and $V_i = V \cap \{v_1, \ldots, v_{i-1}\}^\perp$.

The cosines of the principal angles are precisely the singular values of $\langle V, U \rangle$, and the singular vectors are the directions along which these angles are to be measured. This follows directly from the variational characterization of singular values. Further details may be found in [10].

3 Cryptanalytic Properties

Many cryptanalytic techniques rely only on partial information about the inputs and outputs of a primitive, such as membership of a set or the value taken by a Boolean function. Below, the structure of the inputs (or outputs) will be informally referred to as cryptanalytic input (or output) properties.

One of the obstacles to a more general approach to linear cryptanalysis and its variants, is the fact that different cryptanalytic properties are often described by disparate mathematical objects (such as sets, linear or nonlinear functions, ...). In a few cases, overcoming this difficulty has resulted in new or generalized results. Examples include the projection function approach of Wagner [42] and Baignères *et al.* [2], which enables unifying the data-complexity analysis of several attacks, and the observation that both invariant subspaces and nonlinear invariants correspond to eigenvectors of correlation matrices [6].

Section 3.1 introduces a general correspondence between cryptanalytic properties and subspaces of the inner product space $\mathbb{C}G$. It works for all properties relevant to linear cryptanalysis and its variants, and in particular generalizes both examples just mentioned above. Section 3.2 describes how properties change when a function is applied to the state. This leads to a more general perspective on correlation matrices.

3.1 Correspondence Between Properties and Subspaces

The purpose of this section is to show that the cryptanalytic properties used in linear cryptanalysis and its variants are naturally described by functions $G \to \mathbb{C}$, *i.e.* functions in the inner product space $\mathbb{C}G$ from Section 2.1. This will be motivated from two viewpoints, which are dual to one another. Specifically, the following two perspectives will be advanced:

(i) Cryptanalytic properties correspond to functions in $\mathbb{C}G$.

(ii) Cryptanalytic properties corrsepond to linear functions $\mathbb{C}G \to \mathbb{C}$.

From viewpoint (i), a cryptanalytic property characterizes the state of a collection of inputs or outputs. For instance, probability distributions on G can be represented by functions $G \to [0,1] \subset \mathbb{C}$. Similarly, any subset S of G has an indicator function $\mathbb{1}_S \in \mathbb{C}G$. It will be shown below that the general idea of associating not just positive numbers, but also arbitrary complex-valued weights, to the elements of G is necessary to describe other types of properties.

According to (ii), properties describe a measurement or observation of the state of a collection of inputs or outputs. Importantly, only *linear* functions of the state vector are considered in the present framework. The set of linear functions $\mathbb{C}G \to \mathbb{C}$ is itself a vector space $\mathbb{C}G^*$, *i.e.* the dual vector space of $\mathbb{C}G$. However, the explicit choice of the inner product in Section 2.1 identifies $\mathbb{C}G$ and $\mathbb{C}G^*$. Indeed, $f \in \mathbb{C}G$ corresponds to the function $g \mapsto \langle f, g\rangle$ in $\mathbb{C}G^*$. This correspondence will be used throughout this paper, and both (i) and (ii) will be represened by elements of $\mathbb{C}G$. For example, for a subset S, the indicator function $\mathbb{1}_S$ is dual to the function $f \mapsto \langle \mathbb{1}_S, f\rangle = \sum_{x\in S} f(x)$.

More generally, consider a subspace V of $\mathbb{C}G$. Any function in V can then be interpreted according to either (i) or (ii). The assumption that the property must correspond to a *subspace* of $\mathbb{C}G$ implies that it is possible to make arbitrary linear combinations of these functions.

Representing properties as subspaces of $\mathbb{C}G$ comes with a geometric interpretation. Specifically, the inner product yields the observed outcome when pair of properties with interpretations (i) and (ii) are combined. This aspect will be discussed in detail in Section 4. The remainder of this section is intended as a dictionary between conventional cryptanalytic properties and their corresponding subspaces.

A short summary for $G = \mathbb{F}_2^n$ is given in Table 2. The table includes both the subspaces of $\mathbb{C}G$ and their Fourier transforms, which are subspaces of $\mathbb{C}\widehat{G}$. Importantly, there are other useful subspaces which do not correspond to any of the constructions discussed below. One example will be discussed in Section 6.

Probability distributions. Several properties correspond to subspaces spanned by one or more probability distributions. Subspaces and sets are one example, since any set corresponds to the uniform distribution on that set (equivalently, its indicator function). Affine spaces are an important example and are used in the invariant subspace attack of Leander *et al.* [30].

Table 2: Commonly used cryptanalytic properties and their corresponding subspaces. The characters of $\mathbb{F}_2^n$ are denoted by $\chi_u(x) = (-1)^{u^\top x}$, where $u \in \mathbb{F}_2^n$.

Property	Basis for subspace		Applications
	$V \subseteq \mathbb{C}G$	$\mathcal{F}(V) \subseteq \mathbb{C}\widehat{G}$	
Affine space $a + U \subseteq \mathbb{F}_2^n$	$\{\mathbb{1}_{a+U}\}$	$\{\chi_a \, \mathbb{1}_{U^\perp}\}$	Invariant subspaces
Affine spaces $a_1 + U_1, \ldots \subseteq \mathbb{F}_2^n$	$\{\mathbb{1}_{a_1+U_1}, \ldots\}$	$\{\chi_{a_1} \, \mathbb{1}_{U_1^\perp}, \ldots\}$	Integral cryptanalysis
Probability distribution $p : \mathbb{F}_2^n \to [0, 1]$	$\{p\}$	$\{\,\widehat{p}\,\}$	Statistical saturation
Linear *Mask* $u \in \mathbb{F}_2^n$	$\{\chi_u\}$	$\{\delta_{\chi_u}\}$	Linear approximations
Multidimensional linear *Subspace* $U \subseteq \mathbb{F}_2^n$	$\{\chi_u \mid u \in U\}$	$\{\delta_{\chi_u} \mid u \in U\}$	Multidimensional linear approximations
Multiple linear *Subset* $U \subseteq \mathbb{F}_2^n$	$\{\chi_u \mid u \in U\}$	$\{\delta_{\chi_u} \mid u \in U\}$	Multiple linear approximations
Nonlinear *Function* $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2$	$\{(-1)^{\mathsf{F}}\}$	$\{\mathcal{F}[(-1)^{\mathsf{F}}]\}$	Nonlinear invariants I/O sums
Projection *Function* $\mathsf{F} : \mathbb{F}_2^n \to X$	$\{\delta_x \circ \mathsf{F} \mid x \in X\}$	$\{\widehat{\delta_x \circ \mathsf{F}} \mid x \in X\}$	Partitioning attacks χ^2 distinguishers

Integral and division properties [17,38] are also examples[1], but their analysis is not the main focus of this paper. In this case, the corresponding vector space could be spanned by the indicator function of a set which is balanced on certain bits. However, the intermediate and output properties typically correspond to higher-dimensional vector spaces because they express several possible sets in which the state could be contained. Equivalently, following (ii), one observes the marginal (but *not necessarily* joint) distribution of several state bits.

Not many variants of linear cryptanalysis are directly based on non-uniform probability distributions. The statistical saturation attack of Collard and Standaert [15], in its original form, may be considered an example. In this attack, one estimates the key-dependent probability distribution of the state of a block cipher when some of the plaintext bits are constant and the others are uniform random. However, depending on how the estimated distribution is used, it may be more appropriate to approach this attack using the projection functions discussed below.

Projection functions. Let $\mathsf{F} : G \to H$ be a function between finite Abelian groups G and H, with H typically much smaller than G. In fact, H need not be a group for the construction below to work, but this will be assumed for simplicity. Such functions play an important role in Wagner's framework of 'commutative diagram cryptanalysis', where they are called *projections* [42]. Baignères *et al.* [2] analyze the statistical properties of distinguishers based on balanced projections,

[1] The present framework only describes zero-sum properties.

such as χ^2-attacks [41], partitioning cryptanalysis [22] and multidimensional linear attacks [23].

From the viewpoint of (ii), a projection property gives access to the evaluation of F on the state. Equivalently, the property allows observing any linear combination of the functions $\delta_h \circ \mathsf{F}$, where $\{\delta_h \mid h \in H\}$ is the standard basis of $\mathbb{C}H$. More generally, any function on H can be 'pulled back' to G along the projection function F and the projection property corresponds to the vector space of all such functions. This leads to Definition 3.1 below.

Definition 3.1 (Pullback). *Let $\mathsf{F} : G \to H$ be a function. The pullback operator along F is the linear operator $T^{\mathsf{F}^*} : \mathbb{C}H \to \mathbb{C}G$ defined by $f \mapsto f \circ \mathsf{F}$. The pullback space of $\mathbb{C}H$ along F is the image of T^{F^*}:*

$$\operatorname{im} T^{\mathsf{F}^*} = \{f \circ \mathsf{F} \mid f \in \mathbb{C}H\} \subseteq \mathbb{C}G\,.$$

Similarly, the Fourier transformation $\mathcal{F}(\operatorname{im} T^{\mathsf{F}^*})$ of $\operatorname{im} T^{\mathsf{F}^*}$ will be called the pullback of $\mathbb{C}H$ to $\mathbb{C}\widehat{G}$ along F.

Let V be the vector space corresponding to the projection property defined by F, *i.e.* the pullback of $\mathbb{C}H$ along F. It was already mentioned above that $\{\delta_h \circ \mathsf{F} \mid h \in H\}$ is a basis for V. However, it is often more convenient to use the basis of functions $\chi \circ \mathsf{F}$ where $\chi \in \widehat{H}$. This choice behaves particularly well for homomorphisms $\mathsf{F} : G \to H$ when working with the Fourier transformation of V, since $\widehat{\chi \circ \mathsf{F}} = \delta_{\chi \circ \mathsf{F}}$ in that case.

The following example describes the vector space corresponding to a Boolean projection function in more detail. Such properties are closely related to classical linear cryptanalysis, and more generally the I/O-sums of Harpes *et al.* [21] and the nonlinear approximations considered by Beierle *et al.* [4]. However, as discussed below, there is subtle difference.

Example 3.1. Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function. Denote the characters of $\mathbb{F}_2^n$ by $\chi_u(x) = (-1)^{u^\top x}$. The pullback space V of $\mathbb{C}\mathbb{F}_2$ along F is equal to

$$V = \operatorname{span}\{\delta_0 \circ \mathsf{F}, \delta_1 \circ \mathsf{F}\} = \operatorname{span}\{\mathbb{1}, (-1)^{\mathsf{F}}\}\,,$$

with $\mathbb{1} = \chi_0$ the trivial character of $\mathbb{F}_2^n$. Hence, the Fourier transformation of V is given by

$$\mathcal{F}(V) = \operatorname{span}\{\delta_{\mathbb{1}}, \mathcal{F}[(-1)^{\mathsf{F}}]\}\,.$$

The function $\mathcal{F}[(-1)^{\mathsf{F}}]$ is often called the Walsh-Hadamard transform of F. If F is a linear function, then $\mathsf{F}(x) = u^\top x$ for some $u \in \mathbb{F}_2^n$. Hence, $(-1)^{\mathsf{F}} = \chi_u$ and consequently $\mathcal{F}(V) = \operatorname{span}\{\delta_{\mathbb{1}}, \delta_{\chi_u}\}$. ▷

Example 3.1 suggests that ordinary linear properties correspond to a vector space $V = \operatorname{span}\{\delta_{\mathbb{1}}, \delta_\chi\}$, where χ is a character of the additive group $\mathbb{F}_2^n$. Table 2 instead lists the one-dimensional space $\operatorname{span}\{\delta_\chi\} \subset V$. For the analysis of permutations, there is no significant difference since $\delta_{\mathbb{1}}$ corresponds to a trivial invariant for any permutation (its domain). However, for general functions, the vector space V represents a strictly stronger property.

In general, many commonly used cryptographic properties correspond to subspaces of pullback spaces. This difference is not easily expressed in the formalism of Baignères *et al.* [2] and Wagner [42]. The next paragraph discusses several important examples.

Subspaces of pullbacks. Example 3.1 generalizes to other finite Abelian groups. Let $\mathsf{F} : G \to H$ be a homomorphism. Since $\chi \circ \mathsf{F} \in \widehat{G}$ for any character χ of H, the pullback V of $\mathbb{C}H$ to $\mathbb{C}\widehat{G}$ is spanned by the functions $\delta_{\chi\circ\mathsf{F}}$ with $\chi \in \widehat{H}$. Hence, $\dim V = |H|$. However, the dimension could be reduced by one for permutations. This is essentially the generalization of linear cryptanalysis proposed by Granboulan *et al.* [20, §3]. However, it is also reasonable to consider only one of the functions $\delta_{\chi\circ\mathsf{F}}$. Since this results in one-dimensional subspaces and is closer to the spirit of ordinary linear cryptanalysis. This is essentially the generalization of linear cryptanalysis proposed by Baignères *et al.* [3]. The approaces of Baignères *et al.* and its multidimensional generalization were recently used in the cryptanalysis of FF3.1 [8].

The difference between *multiple* and *multidimensional* linear cryptanalysis is of the same nature. For multiple linear properties, one uses a subspace spanned by one or more standard basis vectors δ_χ. In multidimensional linear cryptanalysis, the considered characters form a subgroup of $\widehat{G}$ and consequently the subspace is the pullback of a homomorphism to some subgroup of G.

3.2 Transformations on $\mathbb{C}G$ and $\mathbb{C}\widehat{G}$

This section investigates how properties, *i.e.* subspaces of $\mathbb{C}G$, change when a function $\mathsf{F} : G \to H$ is applied to the state of the primitive under analysis.

Definition 3.2 (Transition matrix). *Let $\mathsf{F} : G \to H$ be a function. Define $T^{\mathsf{F}} : \mathbb{C}G \to \mathbb{C}H$ as the unique linear operator defined by $\delta_x \mapsto \delta_{\mathsf{F}(x)}$ for all $x \in G$. The transition matrix of F is the coordinate representation of T^{F} with respect to the standard bases of $\mathbb{C}G$ and $\mathbb{C}H$.*

Definition 3.2 only specifies the action of T^{F} on the standard basis of $\mathbb{C}G$, but this uniquely defines T^{F} on all of $\mathbb{C}G$. The choice of the notations T^{F^*} and T^{F} for pullback (Definition 3.1) and transition (Definition 3.2) operators respectively is not arbitrary: these operators are indeed represented by conjugate-transposed matrices. In fact, T^{F} could also be called the *pushforward* operator.

Note that the notation T^{F} will be overloaded, referring to both the operator and its standard matrix representation. The coordinates of the matrix T^{F} will be indexed by elements of G and H rather than by integers, since this avoids choosing an arbitrary ordering of the standard basis. In particular,

$$T^{\mathsf{F}}_{y,x} = \langle \delta_y, T^{\mathsf{F}} \delta_x \rangle = \langle \delta_y, \delta_{\mathsf{F}(x)} \rangle = \delta_y(\mathsf{F}(x)).$$

An analog of Definition 3.2 for $\mathbb{C}\widehat{G}$ is given in Definition 3.3. It generalizes the definition of correlation matrices given in [6] to arbitrary finite Abelian groups. The term *correlation matrix* is due to Daemen *et al.* [16], who defined these matrices in terms of their coordinates.

Definition 3.3 (Correlation matrix). *Let $\mathsf{F} : G \to H$ be a function between finite Abelian groups G and H. Define $C^{\mathsf{F}} : \mathbb{C}\widehat{G} \to \mathbb{C}\widehat{H}$ as the Fourier transformation of T^{F}. That is, $C^{\mathsf{F}} = \mathcal{F}_H\, T^{\mathsf{F}}\, \mathcal{F}_G^{-1}$, with $\mathcal{F}_H$ and $\mathcal{F}_G$ the Fourier transformation on $\mathbb{C}H$ and $\mathbb{C}G$ respectively. The correlation matrix of F is the coordinate representation of C^{F} with respect to the standard bases of $\mathbb{C}\widehat{G}$ and $\mathbb{C}\widehat{H}$.*

$$\begin{tikzcd}
\mathbb{C}G \arrow[r, "T^{\mathsf{F}}"] \arrow[d, "\mathcal{F}_G"'] & \mathbb{C}H \arrow[d, "\mathcal{F}_H"] \\
\mathbb{C}\widehat{G} \arrow[r, "C^{\mathsf{F}}"] & \mathbb{C}\widehat{H}
\end{tikzcd}$$

The notation C^{F} will refer to both the linear operator and its standard matrix representation. Contrary to [6, 16], the coordinates will be indexed by elements of $\widehat{G}$ in order to avoid arbitrary choices. Since $T^{\mathsf{F}}_{y,x} = \delta_y(\mathsf{F}(x))$, the coordinates are given by

$$C^{\mathsf{F}}_{\chi,\psi} = \langle \delta_\chi, C^{\mathsf{F}}\delta_\psi \rangle = \frac{1}{|G|}\langle \chi, T^{\mathsf{F}}\psi \rangle = \frac{1}{|G|}\sum_{x \in G} \overline{\chi(\mathsf{F}(x))}\psi(x).$$

For $G = \mathbb{F}_2^n$ and $H = \mathbb{F}_2^m$, and after identifying these groups with their dual, the expression above coincides with the original definition of correlation matrices given by Daemen *et al.* [16].

The following two theorems list the main properties of transition and correlation matrices that will be used throughout this paper. The last two properties in Theorem 3.1 also apply to correlation matrices. For (2), this follows from the fact that $\mathcal{F}_{G_1 \oplus G_2}$ is essentially the same as $\mathcal{F}_{G_1} \otimes \mathcal{F}_{G_2}$.

Theorem 3.1 (Properties of transition matrices). *Let $\mathsf{F} : G \to H$ be a function. The transition matrix of T^{F} of F has the following properties:*

(1) If F is a bijection, then T^{F} is a permutation matrix.

(2) If $\mathsf{F} = (\mathsf{F}_1, \ldots, \mathsf{F}_n)$ with $\mathsf{F}_i : G_i \to H_i$, then $T^{\mathsf{F}} = \bigotimes_{i=1}^n T^{\mathsf{F}_i}$.

(3) If $\mathsf{F} = \mathsf{F}_2 \circ \mathsf{F}_1$, then $T^{\mathsf{F}} = T^{\mathsf{F}_2}T^{\mathsf{F}_1}$.

Proof. The first two claims directly follow from $T^{\mathsf{F}}_{y,x} = \delta_y(\mathsf{F}(x))$. The third property is an immediate consequence of Definition 3.2. □

Theorem 3.2 (Properties of correlation matrices). *Let $\mathsf{F} : G \to H$ be a function between finite Abelian groups G and H. The correlation matrix C^{F} of F has the following properties:*

(1) If F is a bijection, then C^{F} is a unitary matrix.

(2) If F is a group homomorphism, then $C^{\mathsf{F}}_{\chi,\psi} = \delta_{\chi\circ\mathsf{F}}(\psi)$.

(3) If $G = H$ and $\mathsf{F}(x) = x - t$ for some constant $t \in G$, then C^{F} is a diagonal matrix with $C^{\mathsf{F}}_{\chi,\chi} = \chi(t)$.

Proof. By Theorem 3.1 (1), if F is a permutation, then T^{F} is a permutation matrix and thus unitary. Furthermore, since $|G| = |H|$, both $\mathcal{F}^*_H/\sqrt{|G|}$ and $\mathcal{F}_G/\sqrt{|G|}$ are unitary matrices. Property (1) follows since the product of unitary matrices is unitary and $C^{\mathsf{F}} = \mathcal{F}_H T^{\mathsf{F}} \mathcal{F}_G^{-1}$.

For (2), note that if F is a group homomorphism, then so is $\chi \circ \mathsf{F} : G \to \mathbb{C}^\times$. Hence, by the orthogonality of group characters, $C^{\mathsf{F}}_{\chi,\psi} = \delta_{\chi\circ\mathsf{F}}(\psi)$. As discussed in Section 2.2, property (3) holds by construction of the Fourier transformation. Indeed, note that the action of F corresponds to a translation by t. □

4 Approximations

An approximation of a function $\mathsf{F} : G \to H$ is essentially a pair consisting of an input and an output property. By the correspondence in Section 3, these properties can be represented by subspaces U and V. As discussed in Section 3, $u \in U$ represents a state and $v \in V$ corresponds to a linear measurement function or observation. The inner product $\langle v, T^{\mathsf{F}} u\rangle$ gives the outcome of such an observation. This leads to Definition 4.1 below, where the *approximation map* represents all such inner products without relying on the choice of a specific basis. Given orthonormal bases $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$ for U and V respectively, the coordinates of the matrix representing the approximation map are given by the inner products $\langle v_i, T^{\mathsf{F}} u_i\rangle$.

Definition 4.1 (Approximation). *Let G and H be finite Abelian groups. An approximation of a function $\mathsf{F} : G \to H$ is a pair (U, V) of subspaces $U \subseteq \mathbb{C}\widehat{G}$ and $V \subseteq \mathbb{C}\widehat{H}$. The approximation map of (U, V) is a linear transformation $\langle V, U\rangle_{\mathsf{F}} : U \to V$ defined by $\langle V, U\rangle_{\mathsf{F}} = \pi_V\, C^{\mathsf{F}}\, \iota_U$, with $\iota_U : U \to \mathbb{C}\widehat{G}$ the inclusion map and $\pi_V : \mathbb{C}\widehat{H} \to V$ the orthogonal projection on V.*

$$\begin{tikzcd}
\mathbb{C}\widehat{G} \arrow[r, "C^{\mathsf{F}}"] & \mathbb{C}\widehat{H} \arrow[d, "\pi_V"] \\
U \arrow[u, "\iota_U"] \arrow[r, "\langle V, U\rangle_{\mathsf{F}}"] & V
\end{tikzcd}$$

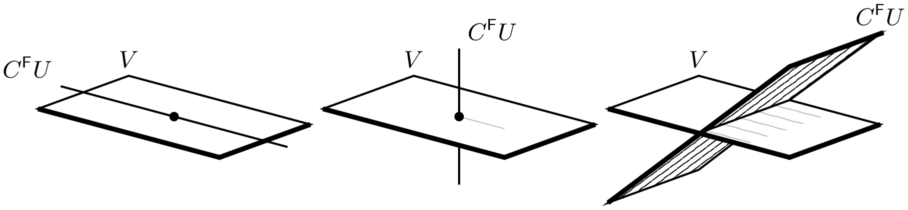
Definition 4.1 refers to subspaces of $\mathbb{C}\widehat{G}$ and $\mathbb{C}\widehat{H}$. An equivalent definition could be given for the subspaces $\mathcal{F}^*_G(U) \subseteq \mathbb{C}G$ and $\mathcal{F}^*_H(V) \subseteq \mathbb{C}H$, taking into account that C^{F} should be replaced by T^{F}. The same remark applies to all definitions in this section and Section 5.

Note that the notation $\langle V, U\rangle_{\mathsf{F}}$ is intentionally similar to the 'inner product of subspaces' notation $\langle V, U\rangle$ from Section 2.3. It will be shown in Theorem 4.1 that the maps $\langle V, U\rangle_{\mathsf{F}}$ and $\langle V, C^{\mathsf{F}} U\rangle$ are indeed closely related and encode the same geometric information.

Example 4.1. Consider a linear approximation for a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. As listed in Table 2, linear properties correspond to one-dimensional spaces $U = \operatorname{span}\{\delta_{\chi_u}\}$ and $V = \operatorname{span}\{\delta_{\chi_v}\}$ with masks $u \in \mathbb{F}_2^n$ and $v \in \mathbb{F}_2^m$. As in Example 2.5, one has the inclusion map $\iota_U(x) = x$ and the orthogonal projection $\pi_V(x) = \langle \delta_{\chi_v}, x\rangle \delta_{\chi_v}$. Hence, $\langle V, U\rangle_{\mathsf{F}}$ is given by $\lambda\delta_{\chi_u} \mapsto \langle \delta_{\chi_v}, C^{\mathsf{F}}\delta_{\chi_u}\rangle\lambda\delta_{\chi_v} = C^{\mathsf{F}}_{\chi_v,\chi_u}\lambda\delta_{\chi_v}$. The same result holds for any pair of finite Abelian groups. ▷

The main purpose of this section is to show that Definition 4.1 indeed encompasses all variants of linear cryptanalysis mentioned in Section 1, and leads to new insights for several of them.

As illustrated in Figure 1, two geometrically intuitive edge cases of Definition 4.1 can be identified: parallel or orthogonal spaces V and $C^{\mathsf{F}}U$. Approximations in the former category will be called 'perfect'. This includes the important case of invariants. The latter case corresponds to a broad generalization of zero-correlation linear approximations. In the remaining cases, the vector spaces V and $C^{\mathsf{F}}U$ are neither completely parallel nor fully orthogonal. All three cases are discussed in detail in Sections 4.1 to 4.3.



(a) Perfect. (b) Zero-correlation. (c) General case.

Fig. 1: Geometric interpretation of Definition 4.1.

The geometric intuitions illustrated in Figure 1 can be quantified using the concept of principal angles that was introduced in Section 2.3. This leads to the following definition of 'principal correlations'. For linear approximations, the unique principal correlation coincides with the ordinary absolute correlation. Further aspects of principal correlations, such as their relation to the 'capacity' in multiple linear cryptanalysis, are discussed in Section 4.3.

Definition 4.2 (Principal correlations). *Let (U, V) be an approximation for a function $\mathsf{F} : G \to H$ between finite Abelian groups G and H. Let $d = \min\{\dim U, \dim V\}$. The principal correlations of the approximation (U, V) are the d largest singular values of the approximation map $\langle V, U\rangle_{\mathsf{F}}$.*

The geometric interpretation of the principal correlations is due to the following result, which relates them to the principal angles between the subspaces $C^{\mathsf{F}}U$ and V.

Theorem 4.1. *Let (U, V) be an approximation for a function $\mathsf{F} : G \to H$ between finite Abelian groups G and H. Let $d = \min\{\dim U, \dim V\}$. If F is* *injective, then the principal correlations of the approximation (U, V) are equal to the cosines of the d smallest principal angles between the subspaces $C^{\mathsf{F}}U$ and V.*

Proof. By Theorem 3.2 (1), C^{F} is a unitary matrix if F is a permutation. More generally, $[C^{\mathsf{F}}]^* C^{\mathsf{F}}$ is a nonzero multiple of the identity map if F is an injection. That is, C^{F} preserves the inner product up to multiplication by a constant. To prove this, show that the result holds for T^{F} (by direct calculation) and then apply the same argument as in the proof of Theorem 3.2 (1).

If C^{F} preserves the inner product up to multiplication by a nonzero constant, then $u_{i+1} \perp u_i$ implies $C^{\mathsf{F}}u_{i+1} \perp C^{\mathsf{F}}u_i$. Hence, the result follows from the fact that the variational characterization of singular values is then equivalent to the definition of principal angles (Definition 2.3). □

4.1 Invariants and Perfect Approximations

If the subspaces U and V are aligned as in Figure 1a, the approximation (U, V) will be called perfect. More formally, (U, V) is perfect if $C^{\mathsf{F}}U \subseteq V$. Alternatively, an approximation *over a permutation* F is perfect if its principal correlations are equal to one.

Integral and division properties are of this type, but these traditionally 'algebraic' properties are not the main focus of this work. However, the case $U = V$ is of particular interest since it leads to a class of approximations that will be called *invariants*, and which includes the invariant subspaces of Leander *et al.* [30] and the nonlinear invariants of Todo *et al.* [39].

Definition 4.3 (Invariant). *Let $\mathsf{F} : G \to G$ be a function. An approximation (V, V) such that $C^{\mathsf{F}}V \subseteq V$ will be called an invariant for F.*

If F is a permutation, all principal correlations of an invariant (V, V) are equal to one. For general functions, this is not necessarily true. For example, if two distinct input distributions result in the same output distribution, it is natural to consider the difference of their probability mass functions as invariant.

Since transition matrices and correlation matrices of permutations have finite multiplicative order, they are diagonalizable. Consequently, by a standard linear algebra result for algebraically closed fields, any invariant V splits into one-dimensional invariant subspaces spanned by the eigenvectors of C^{F}. Hence, Definition 4.3 reduces to the characterization of invariants introduced in [6, Definition 2]

Despite the fact that the eigenvectors of C^{F} determine all possible invariants, the more general characterization of invariants in Definition 4.3 sometimes leads to additional insight. This will be illustrated using the following example, which involves eigenvectors whose corresponding eigenvalue is imaginary – thereby addressing a problem left as future work by [6].

Example 4.2. Consider the following 4-bit S-box, defined in cycle notation:

$$\mathsf{S} = (\mathtt{0\ 7\ b\ 3\ \ d\ 5\ 9\ 6\ \ 8\ 2\ 1\ e\ \ a\ f\ c\ 4}).$$

Further details about this S-box, including a lookup table representation, are given as supplementary material in Appendix A. From a cryptanalytic perspective, the properties of S are seemingly excellent: the linear and differential properties are optimal, and it does not have any fixed points since it is a cyclic permutation. The last property implies that all eigenspaces of C^{S} are one-dimensional, see for instance [6, §4.2]. An immediate consequence of this is that S does not have any nontrivial invariant subspaces.

Denote the ring of integers modulo four by $\mathbb{Z}_4$ and let $f : \mathbb{F}_2^4 \to \mathbb{Z}_4$ be the function defined by

$$\begin{aligned} f(\{\texttt{0}, \texttt{d}, \texttt{8}, \texttt{a}\}) &= 0, & f(\{\texttt{b}, \texttt{9}, \texttt{1}, \texttt{c}\}) &= 2, \\ f(\{\texttt{7}, \texttt{5}, \texttt{2}, \texttt{f}\}) &= 1, & f(\{\texttt{3}, \texttt{6}, \texttt{e}, \texttt{4}\}) &= 3. \end{aligned}$$

By inspection of the cycle structure of S, one can see that $f(\mathsf{S}(x)) = f(x) + 1$ for all $x \in \mathbb{F}_2^4$. This property is reminiscent of nonlinear invariants, and in fact yields a nonlinear invariant for S when reduced modulo two. Nevertheless, the property is more powerful than a nonlinear invariant since its defining function takes values in $\mathbb{Z}_4$ rather than $\mathbb{F}_2$. In fact, the use of $\mathbb{Z}_4$-approximations has been previously suggested by Parker and Raddum [33]. Properties such as f are to nonlinear invariants as nonlinear invariants are to invariant sets: just as a nonlinear invariant can be interpreted as a pair of sets that are potentially swapped by S, f can be interpreted as a pair of nonlinear invariants that are swapped by S.

To obtain a subspace V of $\mathbb{C}\widehat{\mathbb{F}_2^4}$ from f, the pullback construction from Section 3.1 can be applied. Since $\mathbb{Z}_4$ is cyclic of order four, one can deduce from Theorem 2.1 that $\widehat{\mathbb{Z}}_4 = \{x \mapsto \zeta_4^{kx}\ \ k \in \mathbb{Z}_4\}$ with ζ_4 a primitive fourth root of unity such as $\sqrt{-1}$. Hence, using the basis of functions $\widehat{\chi \circ f}$ where $\chi \in \widehat{\mathbb{Z}}_4$, yields

$$\begin{aligned} V &= \operatorname{span}\{\widehat{\zeta_4^{0}}\ , \widehat{\zeta_4^{f}}, \widehat{\zeta_4^{2f}}, \widehat{\zeta_4^{3f}}\} \\ &= \operatorname{span}\{(1,\ 0,0,0,0,\ \ 0,0,\ 0,\ \ 0,0,0,\ 0,0,\ \ \ 0,0,0)^\top, \\ &\qquad (0,\overline{\zeta_8},0,0,0,2\zeta_8,0,\overline{\zeta_8},\ \ 0,0,0,\overline{\zeta_8},0,-\overline{\zeta_8},0,0)^\top/\sqrt{8}, \\ &\qquad (0,\ 0,1,0,1,\ \ 0,0,\ 0,-1,0,0,\ 0,0,\ \ \ 0,1,0)^\top/2, \\ &\qquad (0,\zeta_8,0,0,0,2\overline{\zeta_8},0,\zeta_8,\ \ 0,0,0,\zeta_8,0,-\zeta_8,0,0)^\top/\sqrt{8}\}. \end{aligned}$$

The choice of $\widehat{\chi \circ f}$ (up to a scalar multiple) as a basis is not arbitrary: since $\chi(f(\mathsf{S}(x))) = \chi(1)\chi(f(x))$, it ensures that each basis vector is an eigenvector of C^{S}. Consequently, it is immediately clear that V is indeed an invariant. Note that the first vector listed above is the trivial eigenvector with eigenvalue one. The second and fourth vectors are complex-conjugate eigenvectors corresponding to the conjugate eigenvalues ζ_4 and $\overline{\zeta_4}$. Finally, the third vector is an eigenvector with eigenvalue $\zeta_4^2 = -1$. It corresponds to the nonlinear invariant obtained by reduction modulo two that was mentioned above.

For the purpose of obtaining an interesting example, the S-box S was carefully chosen. In particular, by taking appropriate linear combinations of the two complex-conjugate eigenvectors above, one can see that V is spanned by four real vectors $v_1, \ldots, v_4$ such that $v_1^{\otimes 16}, \ldots, v_4^{\otimes 16}$ are all eigenvectors of C^{L}, where L is the linear layer of Midori-64. Furthermore, these vectors are invariant under the round-constant and key-additions for 2^{32} weak keys. In fact, $v_3^{\otimes 16}$ is itself a nonlinear invariant for the same number of weak keys, but it has been shown that there exists a stronger four-dimensional invariant.

Moreover, there is a larger set of 2^{96} weak keys for which $v_1^{\otimes 16}$ and $v_2^{\otimes 16}$ are still invariants for the whole cipher. This is due to the fact that Midori-64 alternates round keys, and because $C^{\mathsf{S}} v_2 = -v_4$ and $C^{\mathsf{S}} v_4 = v_2$. However, neither v_2 nor v_4 corresponds to a nonlinear invariant for S. One can think of the invariant obtained here as a 'remnant' of the stronger – yet valid for fewer keys – invariant described above. Appendix A contains additional details regarding the preceding claims. ▷

The invariant obtained at the end of the previous example could also have been identified by searching for nonlinear invariants for two rounds which are not necessarily an invariant for the S-box layer, as in [6, Algorithm 1]. However, there is a subtle difference. The examples given in [6, §5.3] correspond to eigenvectors of $[C^{\mathsf{S}}]^2$ with eigenvalue $+1$, whereas the example above is based on an eigenvector of $[C^{\mathsf{S}}]^2$ with eigenvalue -1. This manifests itself when comparing the behaviour for an all-zero key: in the former case, an invariant set for two rounds of Midori-64 is obtained; in the above example this is only achieved for four rounds. However, both examples result in a nonlinear invariant for two rounds of Midori-64.

In general, a one-dimensional periodically repeating perfect approximation for a function F must be an eigenvector of $[C^{\mathsf{F}}]^l$ with eigenvalue one for some positive integer l. These eigenvectors are linear combinations of the eigenvectors of C^{F} with eigenvalues of order divisible by l.

4.2 Zero-Correlation Approximations

Zero-correlation linear approximations were introduced by Bogdanov and Rijmen [13]. They correspond to linear approximations $(\operatorname{span}\{\delta_\psi\}, \operatorname{span}\{\delta_\chi\})$ such that $C^{\mathsf{F}}_{\chi,\psi} = 0$. That is, δ_χ is orthogonal to $C^{\mathsf{F}}\delta_\psi$. This corresponds to the geometric situation sketched in Figure 1b, motivating the following definition.

Definition 4.4 (Zero-correlation approximation). *Let $\mathsf{F} : G \to H$ be a function. An approximation (U, V) such that $V \perp C^{\mathsf{F}} U$ will be called a zero-correlation approximation for F. Equivalently, all principal correlations of a zero-correlation approximation (U, V) are zero.*

Zero-correlation and perfect approximations are closely related, despite being opposite extremes. In fact, this is clear from a geometrical point of view, see for instance Figures 1a and 1b.

Theorem 4.2. *If (U, V) is a zero-correlation approximation, then $(U, V^{\perp})$ is a perfect approximation and conversely.*

Proof. Since (U, V) is a zero-correlation approximation, any $v \in C^{\mathsf{F}} U$ is orthogonal to V. Hence, $C^{\mathsf{F}} U \subseteq V^{\perp}$. The proof of the converse result is analogous. □

The statement and proof of Theorem 4.2 are deceptively simple, but the result is powerful. Indeed, it generalizes the well-known correspondence between multidimensional linear zero-correlation approximations and integral properties, first noted by Bogdanov *et al.* at ASIACRYPT 2012 [12][2] and discussed futher by Sun *et al.* [35].

Definition 4.4 leads to a useful generalization of the miss-in-the-middle approach that is commonly used to find zero-correlation linear approximations. Suppose $\mathsf{F} = \mathsf{F}_2 \circ \mathsf{F}_1$. Let (U_1, V_1) and (U_2, V_2) be approximations such that

$$C^{\mathsf{F}_1} U_1 \subseteq V_1 \perp V_2 \supseteq [C^{\mathsf{F}_2}]^* U_2.$$

It then follows that (U_1, U_2) is a zero-correlation approximation for $\mathsf{F}_2 \circ \mathsf{F}_1$. Recall from Theorem 3.2 (1) that if F_2 is invertible, then $[C^{\mathsf{F}_2}]^* = C^{\mathsf{F}_2^{-1}}$.

Example 4.3. The key-recovery attacks on Midori-64 and MANTIS from ASIACRYPT 2018 [6] are based on a one-dimensional *nonlinear* zero-correlation approximation, and this property was obtained by connecting an ordinary integral property with a nonlinear invariant using the miss-in-the-middle approach discussed above. For completeness, a fully worked out version of this approximation is provided as supplementary material in Appendix B. ▷

The zero-correlation approximation in Example 4.3 can still be explained by mismatching activity patterns in the middle. The benefit of the geometric approach here is mainly that it clarifies that the combination of integral properties with invariants is a natural example of a more general principle, rather than just a 'trick'. However, in some cases, a more refined and possibly key-dependent analysis is necessary to establish the orthogonality of the subspaces V_1 and V_2. Such an example will be encountered in Section 7.3.

4.3 General Approximations

It follows from Example 4.1 that the unique principal correlation for an ordinary linear approximation equals the absolute value of the (conventional) correlation of the linear approximation. For a fixed advantage, the data-complexity of a linear distinguisher is inversely proportional to the square of the correlation.

More generally, Baignères *et al.* [2] discuss the optimal data-complexity of distinguishers for a permutation $\mathsf{F} : G_1 \to G_2$ based on balanced projections $\mathsf{P}_1 : G_1 \to H_1$ and $\mathsf{P}_2 : G_2 \to H_2$. As discussed in Section 3.1, these projections correspond to subspaces $U = \operatorname{span}\{\delta_x \circ \mathsf{P}_1 \mid x \in H_1\} \subseteq \mathbb{C}G_1$ and $V = \operatorname{span}\{\delta_x \circ$

[2] For the case of multidimensional zero-correlation approximations with 'coupled masks', apply Theorem 4.2 to the function $x \mapsto (x, \mathsf{F}(x))$ to obtain their result.

$\mathsf{P}_2 \mid x \in H_2\} \subseteq \mathbb{C}G_2$ by the pullback construction. The approximation map $\langle V, U\rangle_{\mathsf{F}}$ can be represented by a matrix M with coordinates

$$\begin{aligned} M_{y,x} &= \frac{\langle \delta_y \circ \mathsf{P}_2, T^{\mathsf{F}}[\delta_x \circ \mathsf{P}_1]\rangle}{\|\delta_y \circ \mathsf{P}_2\|_2\, \|\delta_x \circ \mathsf{P}_1\|_2} \\ &= \sqrt{\frac{|G_1|}{|G_2|}\frac{\Pr\left[\mathsf{P}_1(\boldsymbol{z}_1) = x\right]}{\Pr\left[\mathsf{P}_2(\boldsymbol{z}_2) = y\right]}}\; \Pr\left[\mathsf{P}_2(\mathsf{F}(\boldsymbol{z}_1)) = y \,|\, \mathsf{P}_1(\boldsymbol{z}_1) = x\right], \end{aligned}$$

where $\boldsymbol{z}_1$ is uniform random on G_1 and $\boldsymbol{z}_2$ is uniform random on G_2. Since the approximations considered by Baignères *et al.* are balanced, $\Pr\left[\mathsf{P}_1(\boldsymbol{z}_1) = x\right] = |H_1|/|G_1|$ and $\Pr\left[\mathsf{P}_2(\boldsymbol{z}_2) = y\right] = |H_2|/|G_2|$, so the prefactor simplifies to $\sqrt{|H_1|}/\sqrt{|H_2|}$. Recall that the Frobenius norm $\|\cdot\|_F$ of a linear operator is the square root of the sum of its squared singular values. Equivalently, its square equals the sum of all squared coordinates of an arbitrary matrix representation with respect to an orthonormal basis. It follows that the Frobenius norm of $\langle V, U\rangle_{\mathsf{F}}$ is given by

$$\|\langle U, V\rangle_{\mathsf{F}}\|_F^2 = \frac{|H_1|}{|H_2|} \sum_{\substack{x \in H_1 \\ y \in H_2}} \Pr\left[\mathsf{P}_2(\mathsf{F}(\boldsymbol{z}_1)) = y \,|\, \mathsf{P}_1(\boldsymbol{z}_1) = x\right]^2.$$

In particular, $\|\langle U, V\rangle_{\mathsf{F}}\|_F^2 - 1$ is equal to the *squared Euclidean imbalance* as defined by Baignères *et al.* [2, Definition 7]. The term -1 is due to the trivial invariant corresponding to the uniform distribution. If this is omitted, one obtains that the data-complexity of an optimal distinguisher is inversely proportional to the sum of the squared principal correlations. This generalizes to multiple linear distinguishers (which are not necessarily of projection type), in which case the squared Frobenius norm corresponds to the fixed-key capacity.

5 Trails

Most cryptographic primitives F do not allow for a direct computation of the approximation map $\langle V, U\rangle_{\mathsf{F}}$, even when U and V are low-dimensional. Indeed, if F is devoid of structure, one is forced to estimate the approximation map empirically. Consequently, finding good approximations of the general type discussed in Section 4.3 is nontrivial.

However, cryptographic primitives are often a composition of highly structured round functions. That is, $\mathsf{F} = \mathsf{F}_r \circ \mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1$. By exploiting the structure of the functions F_i, one can often find approximations (V_i, V_{i+1}) such that $\langle V_{i+1}, V_i\rangle_{\mathsf{F}_i}$ can be efficiently computed. This is for instance the case for linear cryptanalysis, and Section 6 will introduce rank-one approximations as another example for cell-oriented ciphers. The remaining task is to combine or 'pile-up' the individual approximations (V_i, V_{i+1}) for F_i in order to obtain an approximation (V_1, V_{r+1}) for F. The purpose of the piling-up principle, which will be discussed in Section 5.1, is to obtain an estimate of the approximation map $\langle V_{r+1}, V_1\rangle_{\mathsf{F}}$.

Definition 5.1 (Trail). *Let $G_1, G_2, \ldots, G_{r+1}$ be finite Abelian groups. A trail of vector spaces for a function $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ with $\mathsf{F}_i : G_i \to G_{i+1}$ is a tuple $(V_1, V_2, \ldots, V_{r+1})$ of subspaces $V_1 \subseteq \mathbb{C}\widehat{G}_1, \ldots, V_{r+1} \subseteq \mathbb{C}\widehat{G}_{r+1}$.*

Similarly to ordinary linear trails, Definition 5.1 defines a sequence of compatible intermediate approximations. In particular, if all vector spaces V_i are spanned by a standard basis vector $\delta_{\chi_i} \in \mathbb{C}\widehat{G}_i$, one obtains ordinary linear trails as defined by Matsui [31] and generalized to other groups by Baignères *et al.* [3]. Note that the compatibility requirement does not exclude taking one or more of the functions F_i as the identity map.

5.1 Piling-up Principle

As discussed in Section 1, methods for piling-up the approximations within a trail are often motivated by Markov chain assumptions, or a dominant trail hypothesis. Unfortunately, when the former assumption fails, it is often hard to understand why or how to resolve the problem. The latter approach has been mostly limited to the case of simple linear cryptanalysis.

Theorem 5.1 below provides an alternative motivation for the piling-up principle. The premise is that each approximation in a trail corresponds to a transformation of its input space, followed by an orthogonal projection on the input space of the next approximation. Each of these successive projections introduces an error, but orthogonal projection is optimal in the sense that it keeps the inner product between the state and its approximation maximal and the norm of the error minimal (see Section 2.3).

Theorem 5.1 (Piling-up principle). *Let $(V_1, V_2, \ldots, V_{r+1})$ be a trail for a function $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$. The approximation map of the approximation (V_{r+1}, V_1) for F can be written as*

$$\langle V_{r+1}, V_1 \rangle_{\mathsf{F}} = \langle V_{r+1}, V_r \rangle_{\mathsf{F}_r} \cdots \langle V_2, V_1 \rangle_{\mathsf{F}_1} + E\,,$$

where the error term E is the transformation given by

$$E = \sum_{i=1}^{r-1} \langle V_{r+1}, V_{i+1} \rangle_{\mathsf{F}_r \circ \cdots \circ \mathsf{F}_{i+1}} \langle V_{i+1}^{\perp}, V_i \rangle_{\mathsf{F}_i} \cdots \langle V_2, V_1 \rangle_{\mathsf{F}_1}\,.$$

Proof. The proof follows the above intuition of successive orthogonal projection, but keeps track of the error term. Recall from Definition 4.1 that $\langle V, U \rangle_{\mathsf{F}} = \pi_V C^{\mathsf{F}} \iota_U$ where π_V is the orthogonal projector on V and ι_U the inclusion map. Since $\pi_V + \pi_{V^\perp}$ is equal to the identity map, one has the following decomposition for $i = 1, \ldots, r-1$:

$$\begin{aligned}
&\langle V_{r+1}, V_i \rangle_{\mathsf{F}_r \circ \cdots \circ \mathsf{F}_i} \\
&= \pi_{V_{r+1}} C^{\mathsf{F}_r \circ \cdots \circ \mathsf{F}_{i+1}} (\pi_{V_{i+1}} + \pi_{V_{i+1}^{\perp}}) C^{\mathsf{F}_i} \iota_{V_i} \\
&= \langle V_{r+1}, V_{i+1} \rangle_{\mathsf{F}_r \circ \cdots \circ \mathsf{F}_{i+1}} \langle V_{i+1}, V_i \rangle_{\mathsf{F}_i} + \langle V_{r+1}, V_{i+1} \rangle_{\mathsf{F}_r \circ \cdots \circ \mathsf{F}_{i+1}} \langle V_{i+1}^{\perp}, V_i \rangle_{\mathsf{F}_i}\,.
\end{aligned}$$

The result follows by successively decomposing the factor $\langle V_{r+1}, V_{i+1}\rangle_{\mathsf{F}_r\circ\cdots\circ\mathsf{F}_{i+1}}$ using the same expression. □

Theorem 5.1 generalizes the piling-up principle as used in many variants of linear cryptanalysis. This will be demonstrated in Section 5.2. Furthermore, allowing arbitrary subspaces V_i increases flexibility. Even if the spaces V_1 and V_{r+1} correspond to a specific type of property, the intermediate vector spaces can represent seemingly unrelated properties. This will be illustrated in Section 5.3, and again in Section 6. In addition, since the formulation of Theorem 5.1 is basis-free, the choice of basis for these spaces can be arbitrary[3]. This may have computational benefits.

5.2 Discussion of Theorem 5.1

In the one-dimensional case with V_i spanned by δ_{χ_i}, Theorem 5.1 reduces to

$$C^{\mathsf{F}}_{\chi_{r+1},\chi_1} = \textstyle\prod_{i=1}^{r} C^{\mathsf{F}_i}_{\chi_{i+1},\chi_i} + e,$$

where the error term e can be written as a sum over all other linear trails. This is the fixed-key piling-up principle as stated in [16, §6.1] for $\mathbb{F}_2^n$. It also implies the piling-up lemma as stated by Matsui [31] and generalized by Baignères *et al.* [3] to other groups (after taking the variance with respect to independent round keys). The composition result of Beierle *et al.* [4, Theorem 3] for one-dimensional nonlinear approximations is another special case.

A few examples of the higher-dimensional case can be found in the literature. Consider the case where all spaces V_i are pullbacks of $\mathbb{C}H_i$ along balanced projection functions $\mathsf{P}_i : G_i \to H_i$, as in Baignères *et al.* [2] and Wagner [42]. Like all results in this paper, Theorem 5.1 is basis-free and also applies to the spaces $U_i = \mathcal{F}^{-1}(V_i) \subseteq \mathbb{C}G$ provided that one replaces C^{F_i} by T^{F_i}. As shown in Section 4.3, relative to the bases $\{\delta_x \circ \mathsf{P}_i / \|\delta_x \circ \mathsf{P}_i\|_2 \mid x \in H_i\}$ for U_i, the map $\langle U_{i+1}, U_i\rangle_{\mathsf{F}_i}$ can be represented by a matrix M with coordinates

$$M_{y,x} = \sqrt{\frac{|H_i\quad|}{|H_{i+1}|}}\ \Pr\left[\mathsf{P}_{i+1}(\mathsf{F}(\boldsymbol{z})) = y \,|\, \mathsf{P}_i(\boldsymbol{z}) = x\right],$$

where $\boldsymbol{z}$ is uniform random on $|G_i|$. That is, there exist diagonal matrices D_i and D_{i+1} such that $D_{i+1} M D_i^{-1}$ is the transition matrix considered in [2, 42]. These works follow the Markov chain assumption, which leads to using the product of round transition matrices as an approximation for the true transition matrix. The factors D_i and D_{i+1} indeed cancel out, so that Theorem 5.1 yields the same result up to initial and final multiplication by D_{r+1} and D_1^{-1} respectively.

In the case of multiple linear cryptanalysis [9, 25], it is common practice to combine many individual linear trails by adding their correlations. Alternatively,

[3] If B_i is a matrix whose columns form a basis for V_i, then the matrix-representation of $\langle V_{i+1}, V_i\rangle_{\mathsf{F}_i}$ with respect to these bases is $(B_{i+1}^* B_{i+1})^{-1} B_{i+1}^* C^{\mathsf{F}_i} B_i (B_i^* B_i)^{-1}$. Note the normalization factors for non-orthonormal bases.

the squared correlations are added in order to estimate the variance of the correlation under the assumption of independent round keys. However, in general, strong approximations can often be found by taking into account the correlations between all pairs of approximations. Theorem 5.1 reflects this because, for multiple linear approximations, the coordinate representation of $\langle V_{i+1}, V_i \rangle_{\mathsf{F}_i}$ in the standard basis is a submatrix of the correlation matrix C^{F_i}. This approach has been (sometimes implicitly) used in several works, notably in analyses of Present [14], Puffin [29] and Spongent [11]. Note that this is often combined with key-averaging, but a careful analysis of the key-dependency would be both feasible and preferable in many cases.

5.3 Clustering and Linear Approximations from Invariants

A minimal condition for the applicability of the piling-up approximation is that one chooses the best trail from within a predetermined class of candidates, where the principal correlations can be used as a measure of quality. Indeed, by decomposing the error term in Theorem 5.1, one can see that it can be large if other trails result in better or comparable approximations.

However, it is also possible that the class of candidate trails is too limited to obtain a good estimate for $\langle V_{r+1}, V_1 \rangle_{\mathsf{F}}$. In the context of linear cryptanalysis, this phenomenon has been called *clustering* by Daemen and Rijmen [18]. In some cases, clustering can be explained by broadening the set of candidate trails. At ASIACRYPT 2018, an example of a perfect linear approximation over full Midori-64 (with modified round constants) was presented [6]. However, full-round Midori-64 does not admit any high-correlation linear *trails*. This observation can be thought of as an extreme case of a more general phenomenon. At CRYPTO 2012, Abdelraheem *et al.* [1] showed that invariant subspaces give rise to linear approximations with higher-than-expected correlation. The same observation was later generalized to plateaued nonlinear invariants by Beierle *et al.* [4]. Plateaued nonlinear invariants are characterized by a flat Walsh-Hadamard transform, taking only two values up to sign. The results of Beierle *et al.* [4] can be summarized and generalized as follows.

Theorem 5.2. *Let* $\mathsf{F} : G \to G$ *be a function on a finite Abelian group* G*. Let* $v \in \mathbb{C}\widehat{G}$ *be any function such that* $|v(\chi)| = 1/\sqrt{|\mathrm{supp}\ v|}$ *for all* $\chi \in \mathrm{supp}\ v$ *and zero elsewhere. If* $\mathrm{span}\{v\}$ *is an invariant of* F *in the sense of Definition 4.3, then there exist characters* $\chi, \psi \in \mathrm{supp}\ v$ *such that* $|C^{\mathsf{F}}_{\chi,\psi}| \geq 1/|\mathrm{supp}\ v|$.

Proof. By Definition 4.3, it holds that (the sum is over $\chi, \psi \in \mathrm{supp}\ v$)

$$1 = |\langle v, C^{\mathsf{F}} v \rangle| = \left| \sum_{\chi,\psi} \overline{v(\chi)} v(\psi) C^{\mathsf{F}}_{\chi,\psi} \right| \leq |\mathrm{supp}\ v| \max_{\chi,\psi} |C^{\mathsf{F}}_{\chi,\psi}|.$$

It follows that $|C^{\mathsf{F}}_{\chi,\psi}| \geq 1/|\mathrm{supp}\ v|$ for at least one pair (χ, ψ). □

Note that the same result is spread over two theorems in previous work [4, Theorem 4 and 5]: one for invariant subspaces, and one for plateaued nonlinear invariants. This illustrates the convenience of the general definitions in Section 4. To apply the results to the case of invariant subspaces, one only needs to know that the Fourier transformation of the indicator function of a subgroup $H \subseteq G$ is flat with support size $|G|/|H|$. This follows from the Poisson-summation formula [37, Theorem 1]. See also the first entry of Table 2 for $G = \mathbb{F}_2^n$.

Theorem 5.2 and the results above illustrate that a strong approximation using one kind of property tends to result in unexpectedly good approximations using other properties. This can be understood using Theorem 5.1. For example, let $\operatorname{span}\{v\}$ with $\|v\|_2 = 1$ be any one-dimensional invariant for C^{F}. Consider an ordinary linear approximation, *i.e.* a pair $(\operatorname{span}\{\delta_\psi\}, \operatorname{span}\{\delta_\chi\})$ where ψ, χ are characters. Assuming $\delta_\psi \not\perp v$ and $\delta_\chi \not\perp v$, the correlation of the linear approximation over F can be estimated using the following trail:

$$\delta_\psi \xrightarrow[\langle v, \delta_\psi \rangle]{I} v \xrightarrow[1]{C^{\mathsf{F}}} v \xrightarrow[\langle \delta_\chi, v \rangle]{I} \delta_\chi.$$

Theorem 5.1 yields the estimate $|\langle v, \delta_\psi \rangle \langle \delta_\chi, v \rangle| = |v(\psi) v(\chi)|$ for the absolute correlation. If v is flat as in Theorem 5.2, then the piling-up approximation suggests that all approximations with $\psi, \chi \in \operatorname{supp} v$ will have a correlation of roughly $1/|\operatorname{supp} v|$. In fact, this resolves a problem of Beierle *et al.*, who note that “our arguments are non-constructive and therefore, we are not able to identify those highly-biased linear approximations” [4, §1]. In fact, it is easy to identify the highly-biased approximations in practice: generically, *any* approximation with $\psi, \chi \in \operatorname{supp} v$ will do.

6 Rank-One Approximations

It is often convenient to represent the domain of a cipher as an array of m cells of n-bit vectors, because most of the operations in the cipher act on the cells in an independent way. In fact, in ciphers such as the AES, only the linear layer results in diffusion between cells. That is, let $G = (\mathbb{F}_2^n)^m$. Recall from Section 2 that $\mathbb{C}(\mathbb{F}_2^n)^m \cong [\mathbb{C}\mathbb{F}_2^n]^{\otimes m}$ and similarly for the dual group. For example, the probability distribution of a state with independent cells having distributions $p_1, \ldots, p_m$, is represented by the rank-one tensor $p_1 \otimes \cdots \otimes p_m \in [\mathbb{C}\mathbb{F}_2^n]^{\otimes m}$ (see Section 2.1 for definitions).

A rank-one approximation (U, V) is any approximation such that U and V are spanned by a rank-one tensor. No further conditions are imposed on U and V. An important class of rank-one approximations is obtained from balanced Boolean functions $f : (\mathbb{F}_2^n)^m \to \mathbb{F}_2$ such that $f(x_1, \ldots, x_m) = \sum_{i=1}^m f_i(x_i)$. As shown in Table 2, the corresponding vector space for such a property is spanned by the function $(-1)^f = \bigotimes_{i=1}^m (-1)^{f_i}$. Equivalently, the Fourier transformation of the corresponding vector space is spanned by

$$\mathcal{F}[(-1)^f] = \bigotimes\nolimits_{i=1}^m \mathcal{F}[(-1)^{f_i}],$$

where $\mathcal{F}[(-1)^{f_i}]$ is precisely the Walsh-Hadamard transform of f_i. The invariants discussed in [6] and the nonlinear approximations considered by Beierle *et al.* [4] are of this type.

6.1 Theoretical Analysis of Rank-One Trails

By Theorem 3.1 (2), the correlation matrix of a layer of m identical S-boxes S is equal to $(C^{\mathsf{S}})^{\otimes m}$. Indeed, correlation matrices are themselves tensors and the tensor rank (not to be confused with matrix rank) of $(C^{\mathsf{S}})^{\otimes m}$ is one. This expresses the fact that the S-box layer preserves independence of cells. A similar result holds for the key-addition step. Whereas the S-box layer preserves the rank-one structure of approximations, the linear layer tends to increase the rank. In fact, it is reasonable to interpret the rank as a measure of diffusion between the state cells. The correlation matrix of any function $\mathsf{F} : (\mathbb{F}_2^n)^m \to (\mathbb{F}_2^n)^m$ is itself a tensor and can be decomposed as

$$C^{\mathsf{F}} = \sum_{i=1}^{r} \lambda_i \bigotimes\nolimits_{j=1}^{m} C_{i,j},$$

where $C_{i,j}$ are $2^n \times 2^n$ matrices and r is the tensor rank of C^{F}.

Lemma 6.1. *Let $\mathsf{F} : (\mathbb{F}_2^n)^m \to (\mathbb{F}_2^n)^m$ be a function such that $\mathsf{F} = (\mathsf{G}, \mathsf{G}, \ldots, \mathsf{G})$ for some $\mathsf{G} : \mathbb{F}_2^n \to \mathbb{F}_2^n$. If $C^{\mathsf{G}} = \sum_{i=1}^{r} \lambda_i \bigotimes_{j=1}^{n} C_{i,j}$, then*

$$C^{\mathsf{F}} = \sum_{i_1,\ldots,i_m \in [r]} \left(\textstyle\prod_{k=1}^{m} \lambda_{i_k}\right) \bigotimes\nolimits_{k=1}^{m} \bigotimes\nolimits_{j=1}^{n} C_{i_k,j},$$

where $[r] = \{1, \ldots, r\}$. In particular, the tensor rank of C^{F} is at most r^m.

Proof. By Theorem 3.1 (2), it holds that $C^{\mathsf{F}} = (C^{\mathsf{G}})^{\otimes m}$. The result follows by expanding this expression using the multilinearity of tensor products. □

Lemma 6.1 can be used to obtain a decomposition of the correlation matrix of the `MixColumn` map of Midori-64 and MANTIS into 2^8 rank-one terms. This map $\mathsf{M} : (\mathbb{F}_2^4)^4 \to (\mathbb{F}_2^4)^4$ can be represented by the following matrix over $\mathbb{F}_{2^4}$:

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Up to a reordering of the input bits, one can think of M as a map $\widetilde{\mathsf{M}} = (\mathsf{L}, \mathsf{L}, \mathsf{L}, \mathsf{L})$ where L corresponds to the same matrix as above, but over $\mathbb{F}_2$. Specifically, $\widetilde{\mathsf{M}} = \sigma \mathsf{M} \sigma$ where $\sigma : (\mathbb{F}_2^4)^4 \to (\mathbb{F}_2^4)^4$ is the bit permutation defined by $\sigma_i(x_1, \ldots, x_4) = (x_{1,i}, \ldots, x_{4,i})$. Since C^{L} is a 16×16 matrix, one can check that

$$C^{\mathsf{L}} = \frac{1}{2}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes 4} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{\otimes 4} + \begin{pmatrix} 1 & 0 \\ 0 & \text{-}1 \end{pmatrix}^{\otimes 4} - \begin{pmatrix} 0 & 1 \\ \text{-}1 & 0 \end{pmatrix}^{\otimes 4}\right].$$

To see this, it is helpful to observe that C^{L} is symmetric as a tensor. Since $\widetilde{\mathsf{M}} = \sigma \mathsf{M} \sigma$ where σ is a linear map corresponding to a reordering of bits, it follows from Theorem 3.2 (2) and Lemma 6.1 that

$$C^{\mathsf{M}} = 2^{-4} \sum_{i_1,i_2,i_3,i_4 \in [4]^4} \left(\prod_{j=1}^4 \lambda_{i_j}\right) \left[\bigotimes_{j=1}^4 C_{i_j}\right]^{\otimes 4}.$$

with $\lambda_1 = \lambda_2 = \lambda_3 = 1$ and $\lambda_4 = -1$ and

$$C_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad C_3 = \begin{pmatrix} 1 & 0 \\ 0 & \text{-}1 \end{pmatrix}, \quad C_4 = \begin{pmatrix} 0 & 1 \\ \text{-}1 & 0 \end{pmatrix}.$$

Hence, the tensor rank of C^{M} is at most 2^8. This is significantly lower than the worst-case of 2^{16}. Practically speaking, this enables a detailed analysis of rank-one approximations for Midori-64 in Section 7.3. In fact, one can show that this decomposition is minimal *i.e.* the rank of C^{M} is equal to 2^8.

Lemma 6.2 (Lemma 3.5 in [19]). *Let $V_1, \ldots, V_d$ be finite-dimensional vector spaces over $\mathbb{C}$. If $x_{i,1}, \ldots, x_{i,r} \in V_i$ are linearly independent for $i = 1, \ldots, d$, then the vector $\sum_{i=1}^r x_{1,i} \otimes x_{2,i} \otimes \cdots \otimes x_{d,i}$ in $\bigotimes_{i=1}^r V_i$ has tensor rank r.*

To see why Lemma 6.2 implies the result, let V_i be the vector space of 16×16 matrices over $\mathbb{C}$. This is an inner product space under the Frobenius inner product $\mathrm{Tr}\,(A^* B)$ between matrices A and B. It is easy to check that the matrices C_i defined above are mutually orthogonal with respect to this inner product. This implies the mutual orthogonality of the matrices $\left[\bigotimes_{j=1}^4 C_{i_j}\right]^{\otimes 4}$. The result follows by the linear independence of orthogonal vectors.

6.2 Automated Analysis of Rank-One Trails

Let $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ be a permutation on $(\mathbb{F}_2^n)^m$. By Theorem 5.1, an optimal rank-one trail for F can be found by solving the following optimization problem:

$$\begin{aligned} &\text{maximize} \sum_{i=1}^{r} \log_2 \left|\left\langle \textstyle\bigotimes_{j=1}^m v_{i+1,j},\, C^{\mathsf{F}_i} \bigotimes_{j=1}^m v_{i,j} \right\rangle\right| \\ &\text{subject to } \|v_{i,j}\|_2 = 1 \text{ for } i = 1, \ldots, r+1,\ j = 1, \ldots, m \\ &\qquad\qquad v_{i,j}(\mathbb{1}) = 0 \text{ for } (i,j) \in A \text{ and } v_{i,j} = \delta_{\mathbb{1}} \text{ otherwise,} \end{aligned}$$

where the last condition ensures that the vectors $v_{i,j}$ are active and balanced, *i.e.* orthogonal to $\delta_{\mathbb{1}}$, on a predetermined pattern of cells A. Clearly, at least one cell must be active to obtain a nontrivial result. In practice, it is better to take the logarithm of the objective function in order to avoid vanishing gradients.

Restricting to real-valued $v_{i,j}$, the above is an optimization problem over the product of several copies of the $(2^n - 1)$-dimensional unit sphere. This domain is a Riemannian manifold, and common iterative numerical optimization techniques such as steepest descent and conjugate gradient have been generalized to this setting [34]. This is the basic approach behind the automated method proposed in this section. The source code of the tool is provided as supplementary material and relies on the Pymanopt library [40].

The power of this method lies in the fact that it enables iterative convergence to an optimal trail. This is made possible because the general nature of rank-one approximations results in a relaxed, continuous optimization problem rather than a discrete one. Although it is sometimes necessary to ensure that the outermost vectors of the trail correspond to (for example) a Boolean function, there is no reason to impose the same condition on vectors which are internal to the trail.

Example 6.1. The tool can be applied to find rank-one invariants of arbitrary functions with a limited number of input and output bits, which is a difficult problem in general [6]. For example, Figure 2 shows the iterative convergence towards an invariant of the Midori-64 linear layer. This process takes about a second on an ordinary computer. By optimizing over the ellipsoid of unit-norm vectors in the eigenspaces $E_\lambda(C^{\mathsf{S}})$ of the correlation matrix C^{S}, joint invariants for the linear and S-box layer can be found. Instructions to reproduce this example are included as supplementary material in Appendix D. The tool also implements a barrier method to find *all* rank-one invariants for a given linear layer. ▷

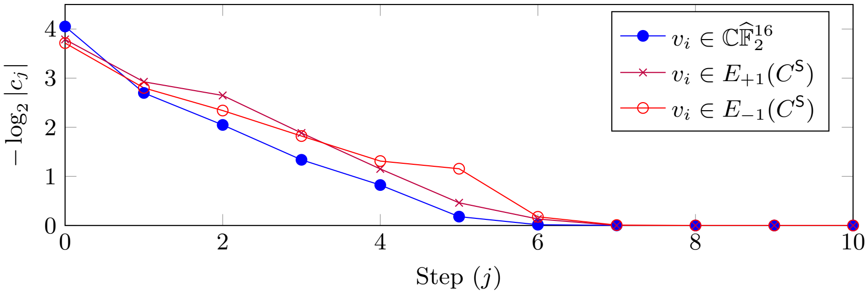


Fig. 2: Correlation c_j at each step of the optimization process for finding invariants of the form $v_1 \otimes v_2 \otimes v_3 \otimes v_4$ with $v_i(\mathbb{1}) = 0$ for the Midori-64 linear layer.

A number of challenges remain for larger problems. These include addressing key-dependence, which is simplified due to the use of the Fourier transform, and convergence issues. For completeness, Appendix D summarizes the (somewhat technical) steps that were taken to address these challenges.

7 Open Problem of Beierle *et al.*

This section explains observations of Beierle *et al.* [4] regarding a nonlinear approximation for two rounds of Midori-64. More broadly, the results in this section lead to a deeper understanding of many nonlinear approximations of the Midori-64 round function.

7.1 Problem Statement

Beierle *et al.* [4, Section 4.4] consider a nonlinear approximation over two rounds of Midori-64, restricted to a single column of the state. Denote this function by F. Its correlation matrix is equal to

$$C^{\mathsf{F}} = C^{\mathsf{M}}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{K}_2}C^{\mathsf{M}}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{K}_1},$$

where K_1 and K_2 are key-addition maps, S is the S-box and M the matrix defined in Section 6.1. Recall from Section 1 that Beierle *et al.* [4] describe nonlinear approximations using linear properties of a nonlinearly transformed representation of the cipher. The details of their approach will not be discussed here; the geometric framework developed in Sections 4 and 5 will be used instead. The nonlinear functions considered by Beierle *et al.* are of the form $\sum_{i=1}^{4} f_i(x)$ with $f_i : \mathbb{F}_2^4 \to \mathbb{F}_2$ and consequently, as discussed in Section 6 on page 24, correspond to approximations spanned by rank-one vectors. Specifically, the pair of nonlinear functions considered in [4, Section 4.4] corresponds to a one-dimensional approximation $(\operatorname{span}\{u \otimes v^{\otimes 3}\}, \operatorname{span}\{u \otimes v^{\otimes 3}\})$ for F with

$$\begin{aligned} u &= 1/4 \cdot (0, 1, 0, -1, 0, 1, 0, -1, 0, -1, 0, 1, 0, -1, 0, -3)^\top \\ v &= 1/2 \cdot (0, 0, 0, -1, 0, 0, 0, -1, 0, \quad 0, 0, 1, 0, \quad 0, 0, -1)^\top. \end{aligned}$$

The coordinates above are given for the character basis δ_{χ_w} with lexicographic ordering of w. Note that v is an eigenvector of C^{S}. Beierle *et al.* estimate the correlation of the above approximation by (from the perspective of this paper) the following one-round trail, which has absolute correlation at least 9/32:

$$u \otimes v \otimes v \otimes v \xrightarrow[\pm 1 \text{ or } \pm 1/2]{[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{K}_i}} u \otimes v \otimes v \otimes v \xrightarrow[9/16]{C^{\mathsf{M}}} u \otimes v \otimes v \otimes v. \qquad (1)$$

The computation of the correlation over C^{M} was done by a direct evaluation of the inner product $\langle u \otimes v^{\otimes 3}, C^{\mathsf{M}}\, u \otimes v^{\otimes 3}\rangle$. This trail was believed to hold whenever $K_i \in \mathbb{F}_2^4 \times \mathcal{K}^3$ for $i = 1, 2$, with $\mathcal{K} = \{(0, 0, x, y) \mid x, y \in \mathbb{F}_2\}$. The weak key set $\mathcal{K}$ ensures the invariance of the tensor product factor v under key addition. Based on the above, one estimates an absolute correlation of at least $(9/32)^2$ over F. However, Beierle *et al.* experimentally observe that this estimate is not accurate:

(i) When $K_2 \in (\mathbb{F}_2^4 \setminus \mathcal{K}) \times \mathcal{K}^3$, the correlation is found to equal zero.

(ii) For other keys, the correlation takes on various values, but is always significantly larger than the estimated minimum of 81/1024. Specifically, for $K_1, K_2 \in \mathcal{K}^4$, the correlation ranges from 35/64 to 40/64 = 5/8. For other keys, it lies between 39/256 and 65/256.

In their conclusion, the authors remark that understanding this phenomenon is "a major open problem". Sections 7.2 and 7.3 completely explains the above observations using the methods developed in Sections 4 and 5.

7.2 Optimal Rank-One Trail

As shown in Section 6.1, the effect of the linear layer is nontrivial and this makes finding an optimal rank-one trail difficult. Hence, a simple explanation for observation (ii) could be that the trail (1) proposed by Beierle *et al.* is not a good guess. Using the tool from Section 6.2, it is easy to find the optimal rank-one trail – ignoring the effect of key-addition for now. Running the tool (the configuration is given in Appendix C) yields the following trail with absolute correlation at most 9/16:

$$u \otimes v^{\otimes 3} \xrightarrow[\pm 3/4 \text{ or } \pm 1/4]{[C^{\mathsf{S}}]^{\otimes 4} C^{\mathsf{K}_1}} v^{\otimes 4} \xrightarrow[1]{C^{\mathsf{M}}} v^{\otimes 4} \xrightarrow[\pm 1]{[C^{\mathsf{S}}]^{\otimes 4} C^{\mathsf{K}_2}} v^{\otimes 4} \xrightarrow[3/4]{C^{\mathsf{M}}} u \otimes v^{\otimes 3}.$$

A short calculation shows that the third step requires $K_2 \in \mathcal{K}$, otherwise the trail has correlation zero. Furthermore, the correlation 3/4 in the first step occurs if and only if $K_1 \in \mathcal{K}^4$. In hindsight, one might have guessed the above trail without detailed analysis: the choice of $v^{\otimes 4}$ as an intermediate step is natural, since $v^{\otimes 4}$ is an invariant for the round function. This is an instance of the general phenomenon discussed in the last paragraph of Section 5.3.

7.3 Theoretical Analysis of the Problem

The correlations predicted by the rank-one trail obtained in Section 7.2 are within 10 to 30% of the observed correlations reported by Beierle *et al.* [4, Tables 1–4]. However, the trail does not yet explain the zero-correlation approximation. In this section, the results from Section 6.1 will be used to find a *minimal and complete* set of rank-one trails for the approximation.

The propagation of $u \otimes v^{\otimes 3}$ under the Midori-64 round function will first be analyzed. For the zero-correlation case, the miss-in-the-middle strategy from Section 4.2 will be used. It will then be shown that a relatively short formula for the exact key-dependent correlation of the approximation can be computed.

Let $K_1 = (k_1, k_2, \ldots, k_{16}) \in \mathbb{F}_2^{16}$ and $K_2 = (k'_1, k'_2, \ldots, k'_{16}) \in \mathbb{F}_2^{16}$. The results in Section 6.1 can be used to compute the image of $u \otimes v^{\otimes 3}$ under one round:

$$C^{\mathsf{M}} [C^{\mathsf{S}}]^{\otimes 4} C^{\mathsf{K}_1} \, u \otimes v^{\otimes 3} = -\nu \, C^{\mathsf{M}} (C^{\mathsf{S}} C^{k_1 \| \cdots \| k_4} u) \otimes v^{\otimes 3} = \nu \, v \otimes \left(\textstyle\sum_{i=1}^{16} c_i \, v_i^{\otimes 3}\right),$$

where $\nu = -\prod_{i=2}^{4} (-1)^{k_{4i-1} + k_{4i}}$. The coefficients c_i and the vectors v_i are listed in Table 4 in Appendix C. Note that, because C^{M} has rank 2^8, one initially obtains 2^8 terms. However, this can be reduced to 16 by grouping terms appropriately. This can be done manually by exploiting the structure of the rank-decomposition, but Sage code to automate this is also provided as supplementary material. Since the vectors v_i are mutually orthogonal and this is preserved when multiplied with (the same) orthogonal matrices, Lemma 6.2 implies that the above decomposition is minimal. Interestingly, not all of the vectors v_i correspond to Boolean functions or probability distributions.

A similar computation can be performed for the inverse of the second round. Specifically, recalling that S and M are involutions,

$$C^{\mathsf{K}_2} [C^{\mathsf{S}}]^{\otimes 4} C^{\mathsf{M}} \, u \otimes v^{\otimes 3} = \nu' \, C^{k'_1 \| \cdots \| k'_4} v \otimes \left(\textstyle\sum_{i=1}^{8} c'_i \bigotimes_{j=1}^{3} (C^{k'_{4j} \| \cdots \| k'_{4j+4}} v'_i)\right).$$

The coefficients c'_i and the vectors v'_i are listed in Table 5 in Appendix C and $\nu' = (-1)^{k'_3+k'_4+1}$. The minimality of the above decomposition can again be established using Lemma 6.2.

Zero-correlation approximation. Let $U = \operatorname{span}\{v\} \otimes (\mathbb{C}\widehat{\mathbb{F}_2^4})^{\otimes 3}$ and $V = \operatorname{span}\{C^{k'_1\|\cdots\|k'_4}\, v\} \otimes (\mathbb{C}\widehat{\mathbb{F}_2^4})^{\otimes 3}$. The decompositions above clearly imply the following inclusions:

$$C^{\mathsf{M}}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{K}_1}\, u \otimes v^{\otimes 3} \in U \quad \text{and} \quad C^{\mathsf{K}_2}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{M}}\, u \otimes v^{\otimes 3} \in V.$$

Consequently, if $U \perp V$, the general miss-in-the-middle principle discussed in Section 4.2 implies that the approximation has correlation zero. This happens whenever $\langle v, C^{k'_1\|\cdots\|k'_4}\, v\rangle = 0$. That is,

$$\begin{aligned}
&\big\langle (0,0,0,1,0,0,0,1,0,0,0,-1,0,0,0,1)^\top, \\
&\quad (0,0,0,1,0,0,0,(-1)^{k'_2},0,0,0,(-1)^{1+k'_1},0,0,0,(-1)^{k'_1+k'_2})^\top \big\rangle \\
&\quad = 1 + (-1)^{k'_1} + (-1)^{k'_2} + (-1)^{k'_1+k'_2},
\end{aligned}$$

which equals zero unless $k'_1 = k'_2 = 0$. This explains the condition $K_2 \in (\mathbb{F}_2^4 \setminus \mathcal{K}) \times \mathcal{K}^3$ observed by Beierle *et al.* [4].

Refining the correlation estimate. Now assume $K_2 \in \mathcal{K}^4$, so that the correlation is nonzero. A closer inspection of the vectors v_i and v'_j reveals that $|\langle v_i, C^{k'_{4j}\|\cdots\|k'_{4j+4}} v'_j\rangle| \leq 1/2$ unless $i = 3$ and $j = 1$. That is, when the inner product $\langle C^{\mathsf{K}_2}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{M}}\, u \otimes v^{\otimes 3}, C^{\mathsf{K}_2}[C^{\mathsf{S}}]^{\otimes 4}C^{\mathsf{M}}\, u \otimes v^{\otimes 3}\rangle$ is expanded using the decomposition above, the term corresponding to $c_3c'_1$ has a weight of one whereas all other terms have weight at most 2^{-3}. Since $v_3 = v'_1 = v$, this term corresponds to the trail from Section 7.2.

The correlation estimate can be improved by including additional trails. In principle, all 128 terms in the expanded inner product between the forward and backward expressions can be computed. The supplementary material contains a Sage script that computes a short formula for the exact key-dependent correlation of the approximation, which is also listed in Appendix C.

In fact, due to the low rank of C^{M}, the same technique can be used to analyze *any* rank-one approximation of F. This includes all linear approximations. In general, the minimal number of rank-one trails can be higher or lower than 16×8 (depending on the choice of the input and output property).

8 Conclusion

A conceptually new 'geometric' approach to linear cryptanalysis has been developed, thereby addressing several of the issues that were discussed in the introduction.

Linear approximations were generalized to pairs of subspaces of inner product spaces. These subspaces can be related to sets, probability distributions, Boolean functions, pullbacks of projections and their subspaces, and many other unexplored properties. This viewpoint helps to clarify the links between different variants of linear cryptanalysis, such as integral and zero-correlation approximations and invariants and ordinary linear approximations. The geometric properties of approximations determine their type and quality. A piling-up principle for general approximations was derived from geometric principles, giving a more transparent motivation for commonly used heuristics in the fixed-key setting. Rank-one approximations were introduced and used to resolve a concrete open problem posed by Beierle *et al.* [4]. In addition, it was shown how such approximations can be found using iterative optimization methods.

The focus of this paper has been on developing general tools, rather than their application to specific situations. Most potential applications are consequently left as future work. In addition, the algorithmic and statistical aspects of using general approximations to set up distinguishers and key-recovery attacks were not discussed in this work.

Acknowledgments. This work is based on my master's thesis "Linear Cryptanalysis in the Weak-Key Model" [7]. I'm grateful to Vincent Rijmen for supervising this thesis, and for comments on a draft of this paper. I also thank Gregor Leander and Christof Beierle for interesting discussions about this work at Ruhr-University Bochum. The author is supported by a PhD Fellowship from the Research Foundation – Flanders (FWO).

References

1. Abdelraheem, M.A., Ågren, M., Beelen, P., Leander, G.: On the distribution of linear biases: Three instructive examples. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO 2012. LNCS, vol. 7417, pp. 50–67. Springer, Heidelberg (Aug 2012). https://doi.org/10.1007/978-3-642-32009-5_4
2. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P.J. (ed.) ASIACRYPT 2004. LNCS, vol. 3329, pp. 432–450. Springer, Heidelberg (Dec 2004). https://doi.org/10.1007/978-3-540-30539-2_31
3. Baignères, T., Stern, J., Vaudenay, S.: Linear cryptanalysis of non binary ciphers. In: Adams, C.M., Miri, A., Wiener, M.J. (eds.) SAC 2007. LNCS, vol. 4876, pp. 184–211. Springer, Heidelberg (Aug 2007). https://doi.org/10.1007/978-3-540-77360-3_13
4. Beierle, C., Canteaut, A., Leander, G.: Nonlinear approximations in cryptanalysis revisited. IACR Trans. Symm. Cryptol. **2018**(4), 80–101 (2018). https://doi.org/10.13154/tosc.v2018.i4.80-101
5. Beierle, C., Jean, J., Kölbl, S., Leander, G., Moradi, A., Peyrin, T., Sasaki, Y., Sasdrich, P., Sim, S.M.: The SKINNY family of block ciphers and its low-latency variant MANTIS. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016, Part II. LNCS, vol. 9815, pp. 123–153. Springer, Heidelberg (Aug 2016). https://doi.org/10.1007/978-3-662-53008-5_5

6. Beyne, T.: Block cipher invariants as eigenvectors of correlation matrices. In: Peyrin, T., Galbraith, S. (eds.) ASIACRYPT 2018, Part I. LNCS, vol. 11272, pp. 3–31. Springer, Heidelberg (Dec 2018). https://doi.org/10.1007/978-3-030-03326-2_1
7. Beyne, T.: Linear Cryptanalysis in the Weak Key Model. Master's thesis, KU Leuven (2019), https://homes.esat.kuleuven.be/~tbeyne/masterthesis/thesis.pdf
8. Beyne, T.: Linear cryptanalysis of FF3-1 and FEA. In: Malkin, T., Peikert, C. (eds.) Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12825, pp. 41–69. Springer (2021). https://doi.org/10.1007/978-3-030-84242-0_3, https://doi.org/10.1007/978-3-030-84242-0_3
9. Biryukov, A., De Cannière, C., Quisquater, M.: On multiple linear approximations. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 1–22. Springer, Heidelberg (Aug 2004). https://doi.org/10.1007/978-3-540-28628-8_1
10. Björck, Å., Golub, G.H.: Numerical methods for computing angles between linear subspaces. Mathematics of computation **27**(123), 579–594 (1973)
11. Bogdanov, A., Knežević, M., Leander, G., Toz, D., Varici, K., Verbauwhede, I.: Spongent: A lightweight hash function. In: Preneel, B., Takagi, T. (eds.) CHES 2011. LNCS, vol. 6917, pp. 312–325. Springer, Heidelberg (Sep / Oct 2011). https://doi.org/10.1007/978-3-642-23951-9_21
12. Bogdanov, A., Leander, G., Nyberg, K., Wang, M.: Integral and multidimensional linear distinguishers with correlation zero. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 244–261. Springer, Heidelberg (Dec 2012). https://doi.org/10.1007/978-3-642-34961-4_16
13. Bogdanov, A., Rijmen, V.: Linear hulls with correlation zero and linear cryptanalysis of block ciphers. Designs, Codes and Cryptography **70**(3), 369–383 (Mar 2014). https://doi.org/10.1007/s10623-012-9697-z
14. Cho, J.Y.: Linear cryptanalysis of reduced-round PRESENT. In: Pieprzyk, J. (ed.) CT-RSA 2010. LNCS, vol. 5985, pp. 302–317. Springer, Heidelberg (Mar 2010). https://doi.org/10.1007/978-3-642-11925-5_21
15. Collard, B., Standaert, F.X.: A statistical saturation attack against the block cipher PRESENT. In: Fischlin, M. (ed.) CT-RSA 2009. LNCS, vol. 5473, pp. 195–210. Springer, Heidelberg (Apr 2009). https://doi.org/10.1007/978-3-642-00862-7_13
16. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In: Preneel, B. (ed.) FSE'94. LNCS, vol. 1008, pp. 275–285. Springer, Heidelberg (Dec 1995). https://doi.org/10.1007/3-540-60590-8_21
17. Daemen, J., Knudsen, L.R., Rijmen, V.: The block cipher Square. In: Biham, E. (ed.) FSE'97. LNCS, vol. 1267, pp. 149–165. Springer, Heidelberg (Jan 1997). https://doi.org/10.1007/BFb0052343
18. Daemen, J., Rijmen, V.: The wide trail design strategy. In: Honary, B. (ed.) 8th IMA International Conference on Cryptography and Coding. LNCS, vol. 2260, pp. 222–238. Springer, Heidelberg (Dec 2001)
19. De Silva, V., Lim, L.H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM Journal on Matrix Analysis and Applications **30**(3), 1084–1127 (2008)
20. Granboulan, L., Levieil, É., Piret, G.: Pseudorandom permutation families over Abelian groups. In: Robshaw, M.J.B. (ed.) FSE 2006. LNCS, vol. 4047, pp. 57–77. Springer, Heidelberg (Mar 2006). https://doi.org/10.1007/11799313_5

21. Harpes, C., Kramer, G.G., Massey, J.L.: A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma. In: Guillou, L.C., Quisquater, J.J. (eds.) EUROCRYPT'95. LNCS, vol. 921, pp. 24–38. Springer, Heidelberg (May 1995). https://doi.org/10.1007/3-540-49264-X_3
22. Harpes, C., Massey, J.L.: Partitioning cryptanalysis. In: Biham, E. (ed.) FSE'97. LNCS, vol. 1267, pp. 13–27. Springer, Heidelberg (Jan 1997). https://doi.org/10.1007/BFb0052331
23. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional linear cryptanalysis of reduced round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) ACISP 08. LNCS, vol. 5107, pp. 203–215. Springer, Heidelberg (Jul 2008)
24. Jordan, C.: Essai sur la géométrie à n dimensions. Bulletin de la Société mathématique de France **3**, 103–174 (1875)
25. Kaliski Jr., B.S., Robshaw, M.J.B.: Linear cryptanalysis using multiple approximations. In: Desmedt, Y. (ed.) CRYPTO'94. LNCS, vol. 839, pp. 26–39. Springer, Heidelberg (Aug 1994). https://doi.org/10.1007/3-540-48658-5_4
26. Knudsen, L.R., Robshaw, M.J.B.: Non-linear approximations in linear cryptanalysis. In: Maurer, U.M. (ed.) EUROCRYPT'96. LNCS, vol. 1070, pp. 224–236. Springer, Heidelberg (May 1996). https://doi.org/10.1007/3-540-68339-9_20
27. Knudsen, L.R., Wagner, D.: Integral cryptanalysis. In: Daemen, J., Rijmen, V. (eds.) FSE 2002. LNCS, vol. 2365, pp. 112–127. Springer, Heidelberg (Feb 2002). https://doi.org/10.1007/3-540-45661-9_9
28. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) EUROCRYPT'91. LNCS, vol. 547, pp. 17–38. Springer, Heidelberg (Apr 1991). https://doi.org/10.1007/3-540-46416-6_2
29. Leander, G.: On linear hulls, statistical saturation attacks, PRESENT and a cryptanalysis of PUFFIN. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 303–322. Springer, Heidelberg (May 2011). https://doi.org/10.1007/978-3-642-20465-4_18
30. Leander, G., Abdelraheem, M.A., AlKhzaimi, H., Zenner, E.: A cryptanalysis of PRINTcipher: The invariant subspace attack. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 206–221. Springer, Heidelberg (Aug 2011). https://doi.org/10.1007/978-3-642-22792-9_12
31. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Helleseth, T. (ed.) EUROCRYPT'93. LNCS, vol. 765, pp. 386–397. Springer, Heidelberg (May 1994). https://doi.org/10.1007/3-540-48285-7_33
32. Nyberg, K.: Linear approximation of block ciphers (rump session). In: Santis, A.D. (ed.) EUROCRYPT'94. LNCS, vol. 950, pp. 439–444. Springer, Heidelberg (May 1995). https://doi.org/10.1007/BFb0053460
33. Parker, M., Raddum, H.: $\mathbb{Z}_4$-linear cryptanalysis. Tech. rep., NESSIE Internal Report: NES/DOC/UIB/WP5/018/1 (06 2020)
34. Smith, S.T.: Optimization techniques on riemannian manifolds. Fields institute communications **3**(3), 113–135 (1994)
35. Sun, B., Liu, Z., Rijmen, V., Li, R., Cheng, L., Wang, Q., AlKhzaimi, H., Li, C.: Links among impossible differential, integral and zero correlation linear cryptanalysis. In: Gennaro, R., Robshaw, M.J.B. (eds.) CRYPTO 2015, Part I. LNCS, vol. 9215, pp. 95–115. Springer, Heidelberg (Aug 2015). https://doi.org/10.1007/978-3-662-47989-6_5
36. Tardy-Corfdir, A., Gilbert, H.: A known plaintext attack of FEAL-4 and FEAL-6. In: Feigenbaum, J. (ed.) CRYPTO'91. LNCS, vol. 576, pp. 172–181. Springer, Heidelberg (Aug 1992). https://doi.org/10.1007/3-540-46766-1_12

37. Terras, A.: Fourier analysis on finite groups and applications. Cambridge University Press (1999)
38. Todo, Y.: Structural evaluation by generalized integral property. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015, Part I. LNCS, vol. 9056, pp. 287–314. Springer, Heidelberg (Apr 2015). https://doi.org/10.1007/978-3-662-46800-5_12
39. Todo, Y., Leander, G., Sasaki, Y.: Nonlinear invariant attack - practical attack on full SCREAM, iSCREAM, and Midori64. In: Cheon, J.H., Takagi, T. (eds.) ASIACRYPT 2016, Part II. LNCS, vol. 10032, pp. 3–33. Springer, Heidelberg (Dec 2016). https://doi.org/10.1007/978-3-662-53890-6_1
40. Townsend, J., Koep, N., Weichwald, S.: Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. Journal of Machine Learning Research **17**(137), 1–5 (2016)
41. Vaudenay, S.: An experiment on DES statistical cryptanalysis. In: Gong, L., Stern, J. (eds.) ACM CCS 96. pp. 139–147. ACM Press (Mar 1996). https://doi.org/10.1145/238168.238206
42. Wagner, D.: Towards a unifying view of block cipher cryptanalysis. In: Roy, B.K., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 16–33. Springer, Heidelberg (Feb 2004). https://doi.org/10.1007/978-3-540-25937-4_2

A Supplementary Material for Example 4.2

A lookup table representation of the S-box in shown in Table 3.

Table 3: An S-box with optimal linear and differential properties.

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathsf{S}(x)$	7	e	1	d	0	9	8	b	2	6	f	3	4	5	a	c

A real basis for the invariant subspace V for S from Example 4.2 is given by

$$\begin{aligned} V = \operatorname{span}\bigl\{&(1,0,0,0,0,0,0,0, \quad 0,0,0,0,0, \quad 0,0,0)^\top, \\ &(0,0,0,0,0,1,0,0, \quad 0,0,0,0,0, \quad 0,0,0)^\top, \\ &(0,0,1,0,1,0,0,0,-1,0,0,0,0, \quad 0,1,0)^\top/2, \\ &(0,1,0,0,0,0,0,1, \quad 0,0,0,1,0,-1,0,0)^\top/2\bigr\}. \end{aligned}$$

In Example 4.2, the four vectors above are denoted by $v_1, \ldots, v_4$. It follows from [6, Theorem 9] that $v_1^{\otimes 16}, \ldots, v_4^{\otimes 16}$ are all eigenvectors of C^{L}, where L is the linear layer of Midori-64. Supporting source code can be found at `https://github.com/TimBeyne/Geometric-approach` in the script '`invariant_example`'.

B Supplementary Material for Example 4.3

It was mentioned in Example 4.3 that the attacks on MANTIS and Midori-64 from [6] are based on nonlinear zero-correlation approximations, which are constructed by connecting a nonlinear invariant and an integral property using the miss-in-the-middle technique that was described in Section 4.2. Below, this is fully worked out for the case of MANTIS. The Midori-64 example is similar, but it uses a more advanced integral property.

Recall that the MANTIS [5] state can be represented as an array of 16 four-bit cells. Hence, it is natural to represent the ambient space by $\bigotimes_{i=1}^{16} \mathbb{C}\widehat{\mathbb{F}}_2^4$.

Let F_1 denote the first two rounds of the cipher, and F_2 the remaining rounds. The invariant for F_2 is one-dimensional and is extended to a perfect approximation $(\operatorname{span}\{u\}, \operatorname{span}\{v\})$ with $C^{\mathsf{F}_2} v = \lambda u$. More specifically, the vectors v and u have tensor rank one. That is,

$$v = \bigotimes_{i=1}^{16} v_i\,,$$

$$u = \bigotimes_{i=1}^{16} u_i\,.$$

Let V_2 be the subspace spanned by all $\bigotimes_{i=1}^{16} \delta_{\chi_{w_i}}$ where $\chi_{w_i}(x) = (-1)^{w_i^\top x}$ such that $w_i \neq 0$ for $i = 1, \ldots 16$. Since $\langle \delta_{\chi_0}, v_i \rangle = 0$ for $i = 1, \ldots 16$, it follows that

$v \in V_2$. Hence, if $U_2 = \operatorname{span}\{u\}$, then

$$V_2 \supset [C^{\mathsf{F}_2}]^* U_2\,.$$

Since v is known, it is technically not necessary to use the above inclusion in the large space V_2. However, the inclusion corresponds to the intuitive property that every cell in the middle state must be active.

Let U_1 be the subspace $\operatorname{span}\{\delta_{\mathbb{1}} \otimes f\}$ for $f = \bigotimes_{i=1}^{15} \widehat{\delta}_{c_i}$ with $c_i \in \mathbb{F}_2^4$ arbitrary. The function $\delta_{\mathbb{1}} \otimes f$ represents a state with one uniform random cell, and all other cells constants. By standard methods of integral cryptanalysis, it is easy to show that F_1 maps such this state to a state in which several cell are individually uniform random after two rounds. This is just the 2-round integral property shown in [6, Figure 9]. Hence, $C^{\mathsf{F}_1} U_1$ is contained in the subspace V_1 spanned by all $\bigotimes_{i=1}^{16} \delta_{\chi_{w_i}}$ with $w_i = 0$ for at least one i:

$$V_1 \supset C^{\mathsf{F}_1} U_1\,.$$

It is clear that $V_1 \perp V_2$. By the miss-in-the-middle approach, it can be concluded that (U_1, U_2) is a zero-correlation approximation for MANTIS-4. This is the property used to set up a key-recovery attack in [6]. Note, however, that (U_1, U_2) is not a standard linear zero-correlation approximation: the input space U_1 corresponds to a specific set and U_2 corresponds to a nonlinear Boolean function.

By Theorem 4.2, the above also implies that $(U_1, U_2^\perp)$ is a perfect approximation – in this case it can be interpreted as a nonlinear integral property. Indeed, U_1 corresponds to the uniform distribution on the first cell and $U_2^\perp$ is the span of all distributions which are balanced on the nonlinear function with Walsh-Hadamard transformation u. The individual integral properties corresponding to one-dimensional subspaces of $U_2^\perp$ spanned by set indicators, are rather weak for the simple reason that they are large.

C Supplementary Material for Section 7.3

For $K_1, K_2 \in \mathcal{K}^4$, the exact absolute correlation of the approximation is given by the following expression:

$$\begin{aligned}
&\big|303/1024\,(-1)^{k_1+k_2+k_3+s_3+s_4} + 189/2048\,[1+(-1)^{k_1}](-1)^{k_3+s_3+s_4}\\
&+ 47/512\,(-1)^{k_2+k_3+s_3+s_4} + 69/2048\,(-1)^{k_1+k_2+k_3+s_4} + 17/2048\,(-1)^{k_1+k_2+s_4}\\
&- 1/128\,(-1)^{k_3+s_4} + 7/1024\,(-1)^{k_2+s_3+s_4} + 5/1024\,(-1)^{s_4}\\
&+ 9/2048\,[1-(-1)^{k_3}]\,(-1)^{k_1+k_2+k_3+s_3}\\
&+ 7/2048\,[1+(-1)^{s_3}+(-1)^{k_1}-(-1)^{s_3+s_4}]\,(-1)^{k_1+s_3+s_4}\\
&- 3/1024\,(-1)^{k_2+k_3+s_3} + 3/2048\,[1+(-1)^{k_3}]\,(-1)^{k_3+s_3}\big|,
\end{aligned}$$

with $s_3 = \sum_{i=2}^{4} k'_{4i-1}$ and $s_4 = \sum_{i=2}^{4} k'_{4i}$. This can be verified using the Sage script `midori_rankone.sage` available at

https://github.com/TimBeyne/Geometric-approach.

Table 4: Vectors v_i and corresponding coefficients in the forward decomposition. The notation $\kappa_i = (-1)^{k_i}$ is used.

i	$2\,v_i^\top$	$\kappa_4\,c_i$
1	$(0,0,0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1)$	$1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-\kappa_1\kappa_3-\kappa_1+2\kappa_2)$
2	$(0,0,0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1)$	$-1/32\,(\kappa_1\kappa_3-2\kappa_2\kappa_3+\kappa_1+2\kappa_2+\kappa_3+1)$
3	$(0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1)$	$-1/16\,\kappa_3\,(3\kappa_1\kappa_2+\kappa_1+\kappa_2+1)$
4	$(0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1)$	$1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2+2\kappa_1-\kappa_3+1)$
5	$(0,0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0)$	$1/32\,(2\kappa_1\kappa_2+\kappa_1\kappa_3-\kappa_1-\kappa_3-1)$
6	$(0,0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2+\kappa_1\kappa_3-2\kappa_2\kappa_3-\kappa_1)$
7	$(0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-2\kappa_2\kappa_3-2\kappa_2+\kappa_3-1)$
8	$(0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0)$	$-1/16\,(\kappa_2-1)$
9	$(0,\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0,0)$	$1/32\,\kappa_1\,(3\kappa_2\kappa_3+\kappa_2-\kappa_3+1)$
10	$(0,\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0)$	$-1/32\,(2\kappa_1\kappa_2+\kappa_1\kappa_3+\kappa_1-\kappa_3+1)$
11	$(0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0,0)$	$1/16\,\kappa_3\,(3\kappa_1\kappa_2-1)$
12	$(0,-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0)$	$1/32\,(3\kappa_1\kappa_2\kappa_3+\kappa_1\kappa_2-2\kappa_1\kappa_3+\kappa_3+1)$
13	$(\ 1,0,0,0,-1,0,0,0,\ 1,0,0,0,\ 1,0,0,0)$	$-1/32\,(\kappa_1\kappa_3-\kappa_1-\kappa_3+1)$
14	$(\ 1,0,0,0,-1,0,0,0,-1,0,0,0,-1,0,0,0)$	$-1/32\,\kappa_1\,(3\kappa_2\kappa_3-\kappa_2-\kappa_3-1)$
15	$(-1,0,0,0,-1,0,0,0,\ 1,0,0,0,-1,0,0,0)$	$-1/32\,(3\kappa_1\kappa_2\kappa_3-\kappa_1\kappa_2-\kappa_3-1)$
16	$(-1,0,0,0,-1,0,0,0,-1,0,0,0,\ 1,0,0,0)$	$1/16\,(\kappa_1-1)$

Table 5: Vectors v_i' and corresponding coefficients in the backward decomposition. The notation $\kappa_i = (-1)^{k'_{4j+i}}$ is used.

i	$4\,(C^{k'_{4j}\Vert\cdots\Vert k'_{4j+4}}v_i')^\top$	c_i'
1	$2(0,\ 0,\ 0,\ -1,\ 0,\ 0,\ 0,\ -1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ -1)$	$-3/8\kappa_3\kappa_4$
2	$2(0,\ 0,-\kappa_3,-\kappa_2\kappa_3,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,-\kappa_3,\kappa_3\kappa_4)$	$1/8$
3	$(0,\ 0,\ 2\kappa_3,\ 0,-2,\ 0,\ 0,\ 0,-1,\ \kappa_4,\ \kappa_3,\ \kappa_3\kappa_4,\ 1,\ \kappa_4,-\kappa_3,\kappa_3\kappa_4)$	$1/8$
4	$(0,\ 0,\ 0,\ 0,\ 0,-2\kappa_4,\ 2\kappa_3,\ 0,\ 1,-\kappa_4,-\kappa_3,\ \kappa_3\kappa_4,\ 1,\ \kappa_4,\ \kappa_3,\kappa_3\kappa_4)$	$1/8$
5	$(0,\ \kappa_4,\ 0,-\kappa_3\kappa_4,-1,\ 2\kappa_4,\ \kappa_3,2\kappa_3\kappa_4,\ 0,-\kappa_4,\ 0,\ \kappa_3\kappa_4,-1,\ 0,\ \kappa_4,\ 0)$	$-1/8$
6	$(0,-\kappa_4,\ 0,-\kappa_3\kappa_4,\ 1,\ 0,\ \kappa_3,\ 0,-1,\ 2\kappa_4,\ \kappa_3,\ 0,\ 0,\ -\kappa_4,\ 2\kappa_3,\kappa_3\kappa_4)$	$1/8$
7	$(0,-\kappa_4,\ 0,\ \kappa_3\kappa_4,\ 1,\ 0,-\kappa_3,\ 0,\ 1,\ 0,\ \kappa_3,2\kappa_3\kappa_4,-2,\ \kappa_4,\ 0,\kappa_3\kappa_4)$	$1/8$
8	$(0,\ \kappa_4,\ 0,\ \kappa_3\kappa_4,\ 1,\ 0,\ \kappa_3,\ 0,\ 0,-\kappa_4,\ 2\kappa_3,\ \kappa_3\kappa_4,\ 1,-2\kappa_4,-\kappa_3,\ 0)$	$-1/8$

D Supplementary Material for Section 6.2

In all of the examples in this paper, the optimization method used is the standard Riemannian variant of the conjugate gradient algorithm. For further details such as the line search method, the reader is referred to the PYMANOPT source code (no custom optimizations were introduced).

A first issue mentioned in Section 6.2 is the potential key-dependence of the correlation. In many cases, it is possible to fix the key and analyze the key-dependence afterwards (using the Fourier transformation simplifies this process). The disadvantage of this approach is that it does not ensure that the approximation will hold for many keys. This can be resolved by optimizing over ellipsoids $B\,\mathbb{S}^{2^n}$ where B is a matrix whose columns are an orthonormal basis for an invariant subspace of several key-addition correlation matrices.

Another issue is that the optimization problem may have many local optima. Lack of global convergence is mainly an issue when the number of variables in the problem is large – for the examples discussed in this paper, this issue was not encountered. For large problems, several restarts may be necessary to find a globally optimal solution. The tool automates this process, but restarting necessarily slows down convergence. For this reason, it is advisable to predetermine the activity pattern and to enforce symmetries wherever possible.

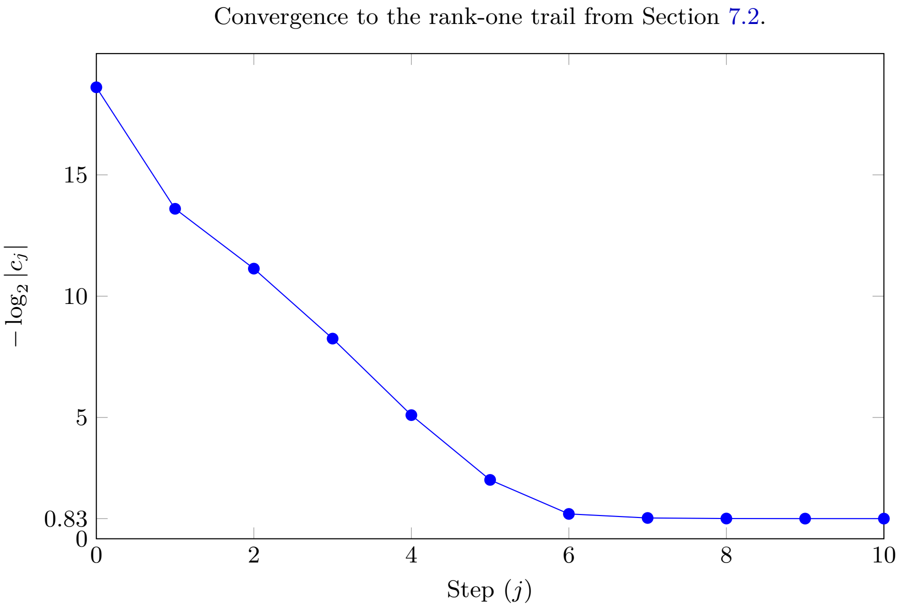
The source code of the tool be accessed at `https://github.com/TimBeyne/Geometric-approach`. To run it, the following steps are required (on a Linux or Unix-type system):

- Install PYMANOPT using `sage -pip install pymanopt`.
- Install ADEPT for automatic differentiation from `http://www.met.reading.ac.uk/clouds/adept/`.
- Compile the file `cost.cpp` into a shared object. Use `g++ -O3 -fPIC -shared cost.cpp -ladept -o cost.so` for gcc.
- Use SAGE to execute `find_invariants.sage` or `trail_midori.sage`.

D.1 Recovering the Rank-One Trail from Section 7.2

The supplementary file '`trail_midori.sage`' contains a script that uses the Riemannian optimization tool to automatically recover the rank-one trail from Section 7.3. Except for balancedness, no constraints were placed on the approximation.

Figure 3 illustrates that the tool quickly converges to the optimal trail. The starting point was random, and the curve corresponds to a single run. The run-time for this experiment was negligible (one to two seconds on a personal computer). The initial guess had a correlation of around 2^{-16}, as expected for a random 2-round trail. The figure illustrates that the tool iteratively steps to an optimal solution. In fact, the correlation increases exponentially with each step.



Convergence to the rank-one trail from Section 7.2.

Fig. 3: Correlation at each step of the optimization process for the rank-one trail from Section 7.3. Note that the final correlation equals $9/16 \approx 2^{-0.83}$.

D.2 Finding Invariants

The file '`find_invariants.sage`' provides a generic tool for finding rank-one invariants (and more generally high-correlation approximations) over a linear layer. It can be used to find a single invariant, or any number of them (if they exist). The latter functionality is implemented using a barrier method, but this is not necessarily the best way.

Figure 2 illustrates the convergence behaviour for finding one nontrivial invariant $v_1 \otimes v_2 \otimes v_3 \otimes v_4$ of the Midori-64 linear layer C^{M}. Three scenarios for $v_i \perp \delta_{\mathbb{1}}$ are shown: v_i arbitrary, v_i an eigenvector of C^{S} with eigenvalue $+1$, and v_i an eigenvector of C^{S} with eigenvalue -1. In all three cases, fast convergence to an invariant is observed.