Fisher Information Determines Capacity of ε-Secure Steganography

Tomáš Filler and Jessica Fridrich

Department of ECE, SUNY Binghamton, NY, USA
{tomas.filler,fridrich}@binghamton.edu

Abstract. Most practical stegosystems for digital media work by applying a mutually independent embedding operation to each element of the cover. For such stegosystems, the Fisher information w.r.t. the change rate is a perfect security descriptor equivalent to KL divergence between cover and stego images. Under the assumption of Markov covers, we derive a closed-form expression for the Fisher information and show how it can be used for comparing stegosystems and optimizing their performance. In particular, using an analytic cover model fit to experimental data obtained from a large number of natural images, we prove that the ± 1 embedding operation is asymptotically optimal among all mutually independent embedding operations that modify cover elements by at most 1.

1 Introduction

The key concept in essentially all communication schemes is the channel capacity defined as the amount of information, or largest payload, that can be safely transmitted over the channel. So far, the capacity of steganographic channels was studied mainly for the case of perfectly secure stegosystems, for which the number of bits that can be safely transmitted in an n-element cover (the *steganographic capacity*) scales linearly w.r.t. n. In this sense, the communication rate (payload per cover element)[1] is *non-vanishing* [1,2,3]. A crucial assumption in these works is the full knowledge of the cover source or the detector. In practice, when dealing with empirical cover sources, such as digital media files, it is unlikely that the communicating parties (Alice and Bob) will have the same knowledge as the Warden. In fact, history teaches us that no matter how sophisticated Alice and Bob are in creating their steganographic scheme that embeds in empirical covers, it is relatively easy for the Warden to identify statistics violated by the embedding and thus mount an attack. Consequently, practical stegosystems are likely to exhibit positive KL divergence between cover and stego objects in some appropriate cover model. We call such systems *imperfect*.

For imperfect stegosystems, the communication rate is not a good descriptor of the channel because it approaches zero with increasing n. Alice, however, still

[1] In this paper, we measure "capacity" as the total number of bits and instead use the term "communication rate" for capacity expressed per cover element.

S. Katzenbeisser and A.-R. Sadeghi (Eds.): IH 2009, LNCS 5806, pp. 31–47, 2009.

needs to know what level of risk she is exposing herself to when sending a message to Bob. It is critical for her to know how much information she can send using her stegosystem in an n-element cover while keeping the KL divergence between cover and stego objects below some chosen ε. It was recently shown that under fairly general assumptions, the amount of information that she can hide scales as $r\sqrt{n}$ [4], with r constant. This *Square Root Law of imperfect steganography* (SRL) was experimentally verified for various embedding algorithms in both spatial and DCT domains [5]. The SRL was also proved by Ker [6] for the case of batch steganography.

In this paper, we propose to use the proportionality constant r from the SRL as a more refined measure of steganographic capacity of imperfect stegosystems. By the form of the law, the constant r, for which we coin the term *the root rate,* essentially expresses the capacity per square root of cover size. We derive a closed form expression for the root rate under the assumption that covers form a Markov chain and embedding is realized by applying a sequence of independent embedding operations to individual cover elements. The root rate depends on the Fisher information rate w.r.t. the the change rate, which was shown to be a perfect security descriptor equivalent to the KL divergence between distributions of cover and stego objects [7]. Expressing the Fisher information rate analytically as a quadratic form allows us to evaluate, compare, and optimize security of stegosystems. To this end, we derive an analytic cover model from a large database of natural images represented in the spatial domain and show that the ± 1 embedding operation is asymptotically optimal among all mutually independent embedding operations that modify cover elements by at most 1. Finally, using the Fisher information rate, we compare security of several practical stegosystems, including LSB embedding and ± 1 embedding. Our findings appear to be consistent with results previously obtained experimentally using steganalyzers and are in good agreement with the recent experimental study reported in [8].

This paper is structured as follows. In the next section, we introduce notation and formulate our assumptions. In Section 3, we introduce the concept of the root rate as a measure of steganographic capacity of imperfect stegosystems. At the same time, we derive a closed form expression for the Fisher information rate on which the root rate depends. Section 4 contains the theoretical foundation for comparing stegosystems and for maximizing the root rate with respect to the embedding operation for a fixed cover source. In Section 5, we present comparison of several known embedding operations for three spatial domain analytic cover models derived from databases of raw, JPEG, and scanned images. Also, we prove that ternary ± 1 embedding has the highest root rate among all stegosystems that modify cover elements by at most 1. The paper is concluded in Section 6.

2 Assumptions

The results reached in this paper will be derived from three basic assumptions. The first assumption concerns the impact of embedding. We postulate that the

stego object is obtained by applying a mutually independent embedding operation to each cover element. This type of embedding can be found in majority of practical embedding methods (see, e.g., [9] and the references therein). The second assumption is our model of covers. We require the individual cover elements to form a first-order stationary Markov chain because this model is analytically tractable while allowing study of more realistic cover sources with memory. Finally, the third assumption essentially states that the resulting stegosystem is imperfect.

Throughout the paper, we use $\mathbb{A} = (a_{ij})$ to denote a matrix with elements a_{ij}, calligraphic font ($\mathcal{X}$) to denote sets, and capital letters (X, Y) to denote random variables, both vector and scalar. If y is a vector with components $y = (y_1, \ldots, y_n)$, y_k^l denotes the subsequence $y_k^l = (y_k, \ldots, y_l)$. If $Y = (Y_1, \ldots, Y_n)$ is a random vector with underlying probability distribution P, then $P(Y_k^l = y_k^l)$ or simply $P(y_k^l)$ denotes the marginal probability $P(Y_k = y_k, Y_{k+1} = y_{k+1}, \ldots, Y_l = y_l)$.

An n-element cover source will be represented using a random variable $X_1^n \triangleq (X_1, \ldots, X_n)$ distributed according to some general distribution $P^{(n)}$ over $\mathcal{X}^n$, $\mathcal{X} \triangleq \{1, \ldots, N\}$. A specific cover object is a realization of X_1^n and will be denoted with the corresponding lower case letter $x_1^n \triangleq (x_1, \ldots, x_n) \in \mathcal{X}^n$. A stegosystem is a triple $S_n = (X_1^n, Emb^{(n)}, Ext^{(n)})$ consisting of the random variable describing the cover source, embedding mapping $Emb^{(n)}$, and extraction mapping $Ext^{(n)}$. The embedding mapping $Emb^{(n)}$ applied to X_1^n induces another random variable $Y_1^n \triangleq (Y_1, \ldots, Y_n)$ with probability distribution $Q_\beta^{(n)}$ over $\mathcal{X}^n$. Specific realizations of Y_1^n are called stego objects and will be denoted $y_1^n \triangleq (y_1, \ldots, y_n)$. Here, $\beta \geq 0$ is a scalar parameter of embedding whose meaning will be explained shortly.

The specific details of the embedding (and extraction) mappings are immaterial for our study. We only need to postulate the *probabilistic impact* of embedding.

Assumption 1. [Mutually independent embedding] *The embedding algorithm modifies every cover element X_k independently to a corresponding element of the stego object Y_k with probability*

$$Q_\beta(Y_k = j | X_k = i) \triangleq b_{ij}(\beta) = \begin{cases} 1 + \beta c_{ii} & \text{if } i = j \\ \beta c_{ij} & \text{otherwise,} \end{cases} \tag{1}$$

for some constants $c_{ij} \geq 0$ for $i \neq j$. Note that because $\sum_{j=1}^N b_{ij} = 1$, we must have $c_{ii} = -\sum_{j \neq i} c_{ij}$ for each $i \in \mathcal{X}$. Also note that we can find sufficiently small β_0 such that $b_{ii}(\beta) > 0$ for $\beta \in [0, \beta_0]$ and all $i \in \mathcal{X}$. The embedding and extraction mappings also impose a bound on the range of β, $\beta \in [0, \beta_{MAX}]$.

The matrix $\mathbb{C} \triangleq (c_{ij})$ reflects the inner workings of the embedding algorithm, while the parameter β captures the *extent* of embedding changes. Due to the fact that $Pr(Y_k \neq X_k) = -\beta c_{ii}$, we can think of β as a parameter controlling the relative number of changes or the change rate. Because the matrix $\mathbb{B}_\beta \triangleq (b_{ij}(\beta))$ does not depend on $k \in \{1, \ldots, n\}$ or the history of embedding changes, one can say that the stego object is obtained from the cover by applying to

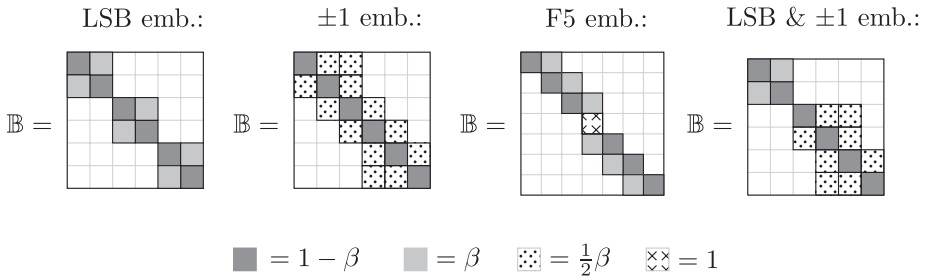

Fig. 1. Examples of several embedding methods in the form of a functional matrix $\mathbb{B}$. The last matrix represents an embedding method that uses LSB embedding to embed in the first two cover values and ternary ± 1 embedding in the last four values.

each cover element a Mutually Independent embedding operation (we speak of *MI embedding*). The independence of embedding modifications implies that the conditional probability of the stego object given the cover object can be factorized, i.e., $Q_\beta^{(n)}(Y_1^n|X_1^n) = \prod_{i=1}^n Q_\beta(Y_i|X_i)$. For simplicity, we omit the index β from the functional matrix $\mathbb{B}_\beta$.

Many embedding algorithms across different domains use MI embedding. Representative examples are LSB embedding, ± 1 embedding, stochastic modulation, Jsteg, MMx, and various versions of the F5 algorithm [9]. Examples of matrix $\mathbb{B}$ for four selected embedding methods are shown in Figure 1. The last matrix $\mathbb{B}$ in this figure represents a practical method that merges ternary ± 1 embedding with LSB embedding.

Next, we formulate our assumption about the cover source.

Assumption 2. [Markov cover source]. *We assume that the cover source X_1^n is a first-order stationary Markov Chain (MC) over $\mathcal{X}$, to which we will often refer as just Markov chain for brevity. This source is completely described by its stochastic transition probability matrix $\mathbb{A} \triangleq (a_{ij}) \in \mathbb{R}^{N\times N}$, $a_{ij} = Pr(X_k = j|X_{k-1} = i)$, and by the initial distribution $Pr(X_1)$. The probability distribution induced by the MC source generating n-element cover objects satisfies $P^{(n)}(X_1^n = x_1^n) = P^{(n-1)}(X_1^{n-1} = x_1^{n-1})a_{x_{n-1}x_n}$, where $P^{(1)}(X_1)$ is the initial distribution. We further assume that the transition probability matrix of the cover source satisfies $a_{ij} \geq \delta > 0$, for some δ and thus the MC is irreducible. The stationary distribution of the MC source is a vector $\pi \triangleq (\pi_1, \ldots, \pi_N)$ satisfying $\pi\mathbb{A} = \pi$. In this paper, we will always assume that the initial distribution $P^{(1)}(X_1) = \pi$, which implies $P^{(n)}(X_k) = \pi$ for every n and k. This assumption simplifies the analysis without loss of generality because the marginal probabilities $P^{(n)}(X_k)$ converge to π with exponential rate w.r.t. k (see Doob [10], equation (2.2) on page 173). In other words, MCs "forget" their initial distribution with exponential rate.*

Under the above assumption and the class of MI embedding, the source of stego images no longer exhibits the Markov property and forms a Hidden Markov Chain[2] (HMC) instead [12]. The HMC model is described by its hidden states (cover elements) and output transition probabilities (MI embedding). Hidden states are described by the cover MC, while the output probability transition matrix $\mathbb{B}$ is taken from the definition of MI embedding.

Unless stated otherwise, in the rest of this paper $Q_\beta^{(n)}$ denotes the probability measure induced by the HMC source embedded with parameter β into n-element MC cover objects. By the stationarity of the MC source, the marginal probabilities $P^{(n)}(X_k^{k+1}) = P^{(2)}(X_1^2)$ and $Q_\beta^{(n)}(Y_k^{k+1}) = Q_\beta^{(n)}(Y_1^2)$ for all k. Sometimes, we will omit the number of elements, n, and denote as P and Q_β the probability distributions over cover and stego images, respectively.

The third assumption we formulate concerns the entire stegosystem S_n. In this work, we only deal with imperfect stegosystems.

Assumption 3. [FI condition]. *We assume the stegosystem* $S_n = (X_1^n, Emb^{(n)}, Ext^{(n)})$ *to be imperfect, meaning that it is not perfectly secure in the sense of Cachin [13], i.e., the KL divergence* $D_{KL}(P^{(n)}||Q_\beta^{(n)}) > 0$ *for* $\beta > 0$. *For our special case of Markov cover sources* X_1^n *and MI embedding* $Emb^{(n)}$, *this assumption can be equivalently stated in two different forms:*

1. *The pair* $(P^{(2)}, Q_\beta^{(2)})$ does not *satisfy the so called* Fisher Information condition,

$$\forall y_1^2 \in \mathcal{X}^2 \quad \Big(P^{(2)}\big(X_1^2 = y_1^2\big) > 0\Big) \;\Rightarrow\; \Big(\frac{d}{d\beta} Q_\beta^{(2)}\big(y_1^2\big)\big|_{\beta=0} = 0\Big). \tag{2}$$

2. *There exists a pair of states* (i, j) *such that*

$$P\big(X_1^2 = (i,j)\big) \neq Q_\beta\big(Y_1^2 = (i,j)\big) \text{ for all } \beta{>}0. \tag{3}$$

For proof of these statements, see [7, Cor. 7].

Finally, we would like to stress that Assumptions 1–3 are not overly restrictive and will likely be satisfied for all practical steganographic schemes in some appropriate representation of the cover. For example, a stegosystem that preserves the Markov model is likely to be detectable by computing higher-order dependencies among pixels. Thus, the stegosystem will become imperfect when representing the cover as pairs or groups of pixels/coefficients or some other quantities computed from the cover.

3 Capacity of Imperfect Stegosystems

In this section, we introduce the concept of root rate as a measure of capacity of imperfect stegosystems. We start by explaining the relationship between

[2] In contrast to [11], we opted not to approximate stego objects by Markov chain as it is not entirely clear what consequences this simplifying step has.

steganographic capacity of stegosystems satisfying Assumptions 1–3 and the Fisher information w.r.t. the parameter β

$$I_n(0) = E_P\left[\left(\frac{d}{d\beta}\ln Q_\beta^{(n)}(y_1^n)\Big|_{\beta=0}\right)^2\right]. \tag{4}$$

Then in Section 3.2, we derive its closed form expression and write it in terms of the expected relative payload α instead of parameter β as this form is more informative for the steganographer.

3.1 Fisher Information in Steganography

Fisher information is a fundamental quantity that frequently appears in theoretical steganography and in general in signal detection and estimation. For example, the Cramer-Rao lower bound states that the reciprocal of Fisher information, $1/I_n(\beta)$, is the lower bound on the variance of unbiased estimators of β (quantitative steganalyzers). Fisher information also appears in the leading term of Taylor expansion of the KL divergence $d_n(\beta) \triangleq D_{KL}(P^{(n)}||Q_\beta^{(n)}) = \beta^2 I_n(0)/(2\ln 2) + O(\beta^3)$, where

$$D_{KL}\left(P^{(n)}||Q_\beta^{(n)}\right) \triangleq \sum_{x_1^n\in\mathcal{X}^n} P^{(n)}(x_1^n)\log_2\frac{P^{(n)}(x_1^n)}{Q_\beta^{(n)}(x_1^n)}.$$

From here, we see that zero KL divergence implies zero Fisher information. Although the opposite is not true in general, it holds for all stegosystems with MI embedding and arbitrary cover model [7]. For such stegosystems, Fisher information $I_n(0)$ represents a perfect security descriptor equivalent to the KL divergence. Fisher information was also proposed for benchmarking steganalyzers [14].

The relationship between the Fisher information rate and steganographic capacity of stegosystems satisfying Assumptions 1–3 was established in [4]. It was essentially shown that such stegosystems are subject to the Square Root Law, which means that payloads that grow faster than $\sqrt{n}$, i.e., $\lim_{n\to\infty}\beta(n)n/\sqrt{n} = \infty$, can be detected arbitrarily accurately, whereas payloads that grow slower than $\sqrt{n}$, i.e., $\beta(n)n/\sqrt{n} \le K < \infty$, lead to ε-secure stegosystems, $d_n(\beta) < \varepsilon$.[3] This result tells us that the payload that can be securely transmitted over the steganographic channel scales as $r\sqrt{n}$. Consequently, the sequence of embedding parameters $\beta(n)$ must approach zero for ε-secure systems and thus the communication rate tends to zero. Due to this fact, it makes sense to evaluate steganographic capacity in the limit of $\beta(n) \to 0$.

[3] Here, we assumed that there exists a linear relationship between $\beta(n)$ and the relative payload $\alpha(n)$ (e.g., the stegosystem does not employ matrix embedding). Indeed, application of matrix embedding does not invalidate our arguments as $\alpha(n)$ differs from $\beta(n)$ only by a multiplicative factor bounded by $\log n$.

3.2 Root Rate

The problem of steganalysis can be formulated as the following hypothesis testing problem

$$\begin{aligned} &\mathrm{H}_0 : \beta = 0 \\ &\mathrm{H}_1 : \beta > 0. \end{aligned} \tag{5}$$

We show that for small (and known) β and large n, the likelihood ratio test with test statistic

$$\frac{1}{\sqrt{n}} T^{(n)}_{\beta_0}(X) = \frac{1}{\sqrt{n}} \ln \big(Q^{(n)}_{\beta_0}(X) / P^{(n)}(X) \big), \tag{6}$$

is a mean-shifted Gauss-Gauss problem.[4] This property, usually called the Local Asymptotic Normality (LAN) of the detector, allows us to quantify and correctly compare security of embedding algorithms operating on the same MC cover model for small values of β.

In this case, the detector performance can be completely described by the deflection coefficient d^2, which parametrizes the ROC curve as it binds the probability of detection, P_D, as a function of the false alarm probability, P_{FA},

$$P_D = Q\big(Q^{-1}(P_{FA}) - \sqrt{d^2} \big).$$

Here, $Q(x) = 1 - \Phi(x)$ and $\Phi(x)$ is the cdf of a standard normal variable $N(0,1)$. Large value of the deflection coefficient implies better detection or weaker steganography.

First, we state the LAN property for the HMC model w.r.t. the embedding parameter β and then extend this result with respect to the relative payload α.

Theorem 1. [LAN of the LLRT]. *Under Assumptions 1–3, the likelihood ratio* (6) *satisfies the local asymptotic normality (LAN), i.e., under both hypotheses and for values of β up to order β^2*

$$\sqrt{n}\big(T^{(n)}_\beta / n + \beta^2 I/2 \big) \xrightarrow{d} N(0, \beta^2 I) \text{ under } \mathrm{H}_0 \tag{7}$$

$$\sqrt{n}\big(T^{(n)}_\beta / n - \beta^2 I/2 \big) \xrightarrow{d} N(0, \beta^2 I) \text{ under } \mathrm{H}_1, \tag{8}$$

where I is the Fisher information rate, $I = \lim_{n\to\infty} \frac{1}{n} I_n(0)$, and $\xrightarrow{d}$ is the convergence in distribution. The detection performance is thus completely described by the deflection coefficient

$$d^2 = \frac{(\sqrt{n}\beta^2 I/2 + \sqrt{n}\beta^2 I/2)^2}{\beta^2 I} = n\beta^2 I.$$

[4] In hypothesis testing, the problem of testing $N(\mu_0, \sigma^2)$ vs. $N(\mu_1, \sigma^2)$ is called the mean-shifted Gauss-Gauss problem and its detection performance is completely described by the deflection coefficient $d^2 = (\mu_0 - \mu_1)^2/\sigma^2$ [15, Chapter 3].

Proof. Due to limited space, we only provide a brief outline of the proof. The Gaussianity of the test statistic follows from the Central Limit Theorem (CLT) due to the fact that the test statistic is close to being i.i.d. Formal proof of this uses exponential forgetting of the prediction filter [16, Lemma 9] and follows similar steps as the proof of the CLT for Markov chains [10]. The mean and variance of the likelihood ratio (6) is obtained by expanding (6) in Taylor series w.r.t. β and realizing that the leading term is the quadratic term containing the Fisher information rate.

We now reformulate the conclusion of the theorem in terms of the payload rather than the parameter β. Matrix embedding (syndrome coding) employed by the stegosystem may introduce a non-linear relationship $\beta = f(\alpha)$ between both quantities. In general, the payload embedded at each cover element may depend on its state $i \in \mathcal{X}$ (e.g., see the last two matrices in Figure 1). Thus, the expected value of the relative payload that can be embedded in each cover is $\alpha(\beta) = \sum_{i\in\mathcal{X}} \pi_i \alpha_i(\beta)$, where $\alpha_i(\beta)$ stands for the number of bits that can be embedded into state $i \in \mathcal{X}$ and π_i is the stationary distribution of the MC. The value of β for which α is maximal will be denoted as β_{MAX}

$$\beta_{MAX} = \arg\max_{\beta} \alpha(\beta).$$

For example, for ternary ± 1 embedding $\beta_{MAX} = 2/3$ and $\alpha_i(\beta_{MAX}) = \log_2 3$, while for binary ± 1 embedding $\beta_{MAX} = 1/2$ and $\alpha_i(\beta_{MAX}) = 1$ (see Figure 1 for the corresponding matrices). Notice that the matrix $\mathbb{C}$ is the same for both embedding methods. The only formal difference is the range of the parameter β. We also remark that unless all α_i are the same, the maximal payload will depend on the distribution of individual states π_i.

To simplify our arguments, we assume a linear relationship between β and α (e.g., we do not consider in this paper the effects of matrix embedding). Therefore, we can write

$$\beta = f(\alpha) = \frac{\beta_{MAX}}{\alpha_{MAX}}\alpha, \tag{9}$$

where $\alpha \in [0, \alpha_{MAX}]$ and $\alpha_{MAX} = \alpha(\beta_{MAX})$ denotes the average number of bits that can be embedded into cover element while embedding with $\beta = \beta_{MAX}$ (maximum change rate).

From (9), the deflection coefficient can be expressed in terms of the relative payload α by substituting $\beta = f(\alpha)$ from (9) into Q_β

$$d^2 = n\alpha^2 \left(\frac{\beta_{MAX}}{\alpha_{MAX}}\right)^2 I. \tag{10}$$

In practice, Alice can control statistical detectability by bounding $d^2 < \varepsilon$ for some fixed ε, obtaining thus an upper bound on the total number of bits (payload) αn that can be safely embedded (this requires rearranging the terms in (10))

$$\alpha n \leq \frac{\alpha_{MAX}}{\beta_{MAX}} \sqrt{\frac{\varepsilon}{I}} n. \tag{11}$$

In analogy to the communication rate, it is natural to define *the root rate*

$$r \triangleq \frac{\alpha_{MAX}}{\sqrt{I}\beta_{MAX}} \tag{12}$$

as the quantity that measures steganographic security of imperfect stegosystems in bits per square root of cover size per square root of KL divergence. We use the root rate for comparing stegosystems with a MC cover model.

In the next theorem, proved in the appendix, we establish the existence of the main component of the root rate, the Fisher information rate I, and express it in a closed form.

Theorem 2. [Fisher information rate]. *Let $\mathbb{A} = (a_{ij})$ define the MC cover model and $\mathbb{B}$, defined by matrix $\mathbb{C} = (c_{ij})$, capture the embedding algorithm. Then, the normalized Fisher information $I_n(0)/n$ approaches a finite limit I as $n \to \infty$. This limit can be written as $I = \mathbf{c}^T \mathbb{F} \mathbf{c}$, where $\mathbf{c}$ is obtained by arranging $\mathbb{C}$ into a column vector of size N^2 with elements c_{ij}.*[5] *The matrix $\mathbb{F}$ of size $N^2 \times N^2$ is defined only in terms of matrix $\mathbb{A}$ and does not depend on the embedding algorithm. The elements of matrix $\mathbb{F}$ are*

$$f_{(i,j),(k,l)} = [j = l]V(i,j,k) - U(i,j,k,l), \tag{13}$$

where by the Iverson notation $[j = l]$ is one if $j = l$ and zero otherwise and

$$V(i,j,k) = \left(\sum_{z \in \mathcal{X}} \pi_z a_{zi} \frac{a_{zk}}{a_{zj}}\right)\left(\sum_{z \in \mathcal{X}} a_{iz} \frac{a_{kz}}{a_{jz}}\right)$$

$$U(i,j,k,l) = \pi_i \left(a_{ik} - a_{il}\frac{a_{jk}}{a_{jl}}\right) + \pi_k \left(a_{ki} - a_{kj}\frac{a_{li}}{a_{lj}}\right).$$

Moreover, $|I_n(0)/n - I| \leq C/n$ for some constant C. This constant depends only on the elements of matrix $\mathbb{A}$ and not on the embedding algorithm. The quadratic form $I(\mathbf{c}) = \mathbf{c}^T \mathbb{F} \mathbf{c}$ is semidefinite, in general.

By inspecting the proof of the theorem, the matrix $\mathbb{F}$ can be seen as the Fisher information rate matrix w.r.t. the parameters $\{b_{ij} | 1 \leq i, j \leq N\}$. It describes the natural sensitivity of the cover source to MI embedding. The quadratic form then combines these sensitivities with coefficients given by the specific embedding method and allows us to decompose the intrinsic detectability caused by the cover source from the detectability caused by the embedding algorithm.

Corollary 1. *For the special case when the MC degenerates to an i.i.d. cover source with distribution $P = \pi$, the Fisher information rate simplifies to*

$$I = \sum_{i,j,k \in \mathcal{X}} c_{ij} \frac{\pi_i \pi_k}{\pi_j} c_{kj}.$$

[5] The order of elements in $\mathbb{C}$ is immaterial as far as the same ordering is used for pairs (i,j) and (k,l) in matrix $\mathbb{F}$.

4 Maximizing the Root Rate

In the previous section, we established that the steganographic capacity of imperfect stegosystems should be measured as the root rate (12) defined as the payload per square root of the cover size and per square root of KL divergence. The most important component of the root rate is the stegosystem's Fisher information rate, for which an analytic form was derived in Theorem 2. The steganographer is interested in designing stegosystems (finding $\mathbb{C}$) with the highest possible root rate. This can be achieved by minimizing the Fisher information rate or by embedding symbols from a larger alphabet, i.e., increasing the ratio α_{MAX}/β_{MAX}. In this section, we describe two general strategies for maximizing the root rate that are applicable to practical stegosystems. In Section 5, we draw conclusions from experiments when these strategies are applied to real cover sources formed by digital images.

Before proceedings with further arguments, we point out that the highest root rate is obviously obtained when the Fisher information rate is zero, $I = 0$. This can happen for non-trivial embedding ($\mathbb{C} \neq 0$) in certain sources because the Fisher information rate is a semidefinite quadratic form. Such stegosystems, however, would be perfectly secure and thus by Assumption 3 are excluded from our consideration.[6]

The number of bits, α_i, that can be embedded at each state $i \in \mathcal{X}$ is bounded by the entropy of the ith row of $\mathbb{B} = \mathbb{I} + \beta\mathbb{C}$, $H(\mathbb{B}_{i\bullet})$. Thus, in the most general setting, we wish to maximize the root rate

$$\frac{\sum \pi_i H\left(\mathbb{B}_{i\bullet}(\beta_{MAX})\right)}{\beta_{MAX}} \frac{1}{\sqrt{I}}$$

w.r.t. matrix $\mathbb{C}$. The nonlinear objective function makes the analysis rather complicated and the result may depend on the distribution of individual states π. Moreover, even if we knew the optimal solution, care needs to be taken in interpreting such results, because a practical algorithm allowing us to communicate the entropy of the additive noise may not be available. We are only aware of a few practical embedding algorithms that communicate the maximal amount of information (LSB embedding with binary symbols and ± 1 embedding with ternary symbols). In practice, stochastic modulation [17] can be used in some cases to embed information by adding noise with a specific pmf (matrix $\mathbb{C}$), but the specific algorithms described in [17] are suboptimal.

In the rest of this section, we present two different approaches how to optimize the embedding algorithm under different settings that are practically realizable.

4.1 Optimization by Convex Combination of Known Methods

One simple and practical approach to optimize the embedding method is obtained by combining existing stegosystems $S^{(1)}$ and $S^{(2)}$. Suppose Alice and

[6] An example of such a stegosystem is LSB embedding in i.i.d. covers with $\pi_{2i} = \pi_{2i+1}$ for all i.

Bob embed a portion of the message into λn elements, $0 < \lambda < 1$, using $S^{(1)}$ and use the remaining $(1-\lambda)n$ elements to embed the rest of the message using $S^{(2)}$. If both parties select the elements pseudo-randomly based on a stego key, the impact on a single cover element follows a distribution obtained as a convex combination of the noise pmfs of both methods. Note that the methods are allowed to embed a different number of bits per cover element since Bob knows which symbol to extract from each part of the stego object. Let $S^{(i)}$ represent the ith embedding method with matrix $\mathbb{C}^{(i)}$, or its vector representation $\mathbf{c}^{(i)}$, with ratio $\rho^{(i)} = \alpha^{(i)}_{MAX}/\beta^{(i)}_{MAX}$ for $i \in \{1,2\}$. The root rate $r(\lambda)$ of the method obtained by the above approach (convex embedding) with parameter λ can be written as

$$\begin{aligned} r(\lambda) &= \frac{\lambda\rho^{(1)} + (1-\lambda)\rho^{(2)}}{\sqrt{(\lambda\mathbf{c}^{(1)} + (1-\lambda)\mathbf{c}^{(2)})^T \mathbb{F}(\lambda\mathbf{c}^{(1)} + (1-\lambda)\mathbf{c}^{(2)})}} \\ &= \frac{\lambda\rho^{(1)} + (1-\lambda)\rho^{(2)}}{\sqrt{\lambda^2 I^{(1)} + (1-\lambda)^2 I^{(2)} + 2\lambda(1-\lambda)I^{(1,2)}}}, \end{aligned} \tag{14}$$

where $I^{(i)}$ is the Fisher information rate of $S^{(i)}$ and $I^{(1,2)} = \left(\mathbf{c}^{(1)}\right)^T \mathbb{F}\mathbf{c}^{(2)}$. Here, we used the symmetry of $\mathbb{F}$ to write $I^{(1,2)} = I^{(2,1)}$.

4.2 Minimizing the Fisher Information Rate

In an alternative setup, we deal with the problem of optimizing the shape of the additive noise pmf under the assumption that the number of bits, α_i, embedded at each state $i \in \mathcal{X}$ is constant. For example, we may wish to determine the optimal pmf that would allow us to communicate 1 bit per element ($\alpha_i = 1$, $\forall i \in \mathcal{X}$) by changing each cover element by at most 1. In this problem, the ratio α_{MAX}/β_{MAX}, as well as the cover model (matrix $\mathbb{A}$), are fixed and known. The task is to minimize the Fisher information rate I.

We formulate our optimization problem by restricting the form of the matrix $\mathbb{C} = (c_{ij})$, or its vector representation $\mathbf{c} = (c_{ij}) \in R^{N^2 \times 1}$, to the following linear parametric form

$$\mathbf{c} = \mathbb{D}v + e, \tag{15}$$

where $\mathbb{D} = (d_{ij})$ is a full-rank real matrix of size $N^2 \times k$, e is a real column vector of size N^2, and $v = (v_1, \ldots, v_k)^T$ is a k-dimensional column vector. We assume $v \in \mathcal{V}$, where $\mathcal{V}$ is bounded by a set of linear inequalities[7] and the constraint $\sum_j c_{ij} = 0$ for all $i \in \{1, \ldots, N\}$. In other words, we decompose the matrix $\mathbb{C}$ into k real parameters v_i, $i \in \{1, \ldots, k\}$. The following example shows one such representation for a stegosystem whose embedding changes are at most 1.

Example 1. [Tridiagonal embedding]. We set $c_{ii} = -1$, $c_{i,i-1} = v_{i-1}$, and $c_{i,i+1} = 1 - v_{i-1}$ for $i \in \{2, \ldots, N-1\}$ (and suitably defined at the boundaries). This allows us to model ± 1 embedding, LSB embedding, and all possible MI embedding methods that modify every element by at most 1. By setting $c_{ii} = -1$ for

[7] E.g., we must have $\mathbb{B} \geq 0$.

all i, we constrain ourselves to stegosystems that embed the same payload into every state $i \in \mathcal{X}$ for all $\beta \geq 0$. This model has $k = N - 2$ parameters and the set $\mathcal{V}$ is formed by $v_j \in [0, 1]$, $j \in \{1, \ldots, k\}$.

Our task is to minimize the Fisher information rate for embedding methods given by (15). The function $I(v) = (\mathbb{D}v + e)^T \mathbb{F}(\mathbb{D}v + e)$ can attain its minimum either at a point with a zero gradient[8] (a critical point) or on the boundary of $\mathcal{V}$. We now derive a set of linear equations for the set of all possible critical points. This approach will be used in Section 5 to prove that ternary ± 1 embedding is asymptotically optimal within the class of tridiagonal embedding in spatial domain.

For our parametrization, the gradient w.r.t. every parameter v_j can be expressed as

$$\frac{\partial}{\partial v_j} I(v) = \frac{\partial}{\partial v_j} (\mathbb{D}v + e)^T \mathbb{F}(\mathbb{D}v + e) = 2(\mathbb{D}_{\bullet j})^T \mathbb{F}(\mathbb{D}v + e),$$

where $\mathbb{D}_{\bullet j}$ is the jth column of matrix $\mathbb{D}$. Because every possible candidate v_0 for the optimal parameters must satisfy $(\partial/\partial v_j) I(v)|_{v=v_0} = 0$ for every $j \in \{1, \ldots, k\}$, all critical points are solutions of the following linear system

$$\mathbb{D}^T \mathbb{F} \mathbb{D} v = -\mathbb{D}^T \mathbb{F} e. \tag{16}$$

If this system has a unique solution $v_0 \in \mathcal{V}$, then v_0 corresponds to matrix $\mathbb{C}$ achieving the global minimum of the Fisher information rate, which corresponds to the best MI embedding method w.r.t. $\mathcal{V}$ and a given MC cover source.

5 Experiments

In the previous section, we outlined two strategies for maximizing the root rate for practical stegosystems. This section presents specific results when these strategies are applied to stegosystems operating on 8-bit gray-scale images represented in the spatial domain. Although images are two dimensional objects with spatial dependencies in both directions, we represent them in a row-wise fashion as a first-order Markov Chain over $\mathcal{X} = \{0, \ldots, 255\}$. The MC model represents the first and simplest step of capturing pixel dependencies while still retaining the important advantage of being analytically tractable. Then, we adopt a parametric model for the transition probability matrix of this Markov cover source and show that it is a good fit for the empirical transition probability matrix $\mathbb{A}$ estimated from a large number of natural images. We use the analytic model to evaluate the root rate (12) of several stegosystems obtained by a convex combinations of known methods. Finally, we show that the optimal embedding algorithm that modifies cover elements by at most 1 is very close to ± 1 embedding.

In principle, in practice we could calculate the Fisher information rate using equation (13) with an empirical matrix $\mathbb{A}$ estimated from a large number of

[8] Note that the semidefiniteness of $\mathbb{F}$ guarantees that the extremum must be a minimum.

images. However, this approach may give misleading results because (13) is quite sensitive to small perturbations of a_{ij} with a small value (observe that $I = +\infty$ if $a_{ij} = 0$). We do not expect this to be an issue in practice since rare transitions between distant states are probable but content dependent, which makes them difficult to be utilized for steganalysis. Because small values of a_{ij} can not be accurately estimated in practice, we represent the matrix $\mathbb{A}$ with the following parametric model

$$a_{ij} = \frac{1}{Z_i} e^{-(|i-j|/\tau)^\gamma}, \tag{17}$$

where $Z_i = \sum_{j=1}^{256} e^{-(|i-j|/\tau)^\gamma}$ is the normalization constant. The parameter γ controls the shape of the distribution, whereas τ controls its "width." The model parameters were found in the logarithmic domain using the least square fit between (17) and its empirical estimate. To validate this model, we carried out the least square fit separately for three image databases: never compressed images taken by several digital cameras[9] (CAMRAW), digital scans[10] (NRCS), and decompressed JPEG images[11] (NRCS-JPEG). Figure 2 shows the comparison between the empirical matrix $\mathbb{A}$ estimated from the CAMRAW database and the corresponding fit. Although this model cannot capture some important macroscopic properties of natural images, such as pixel saturations, it remains analytically tractable and is valid for many natural images.

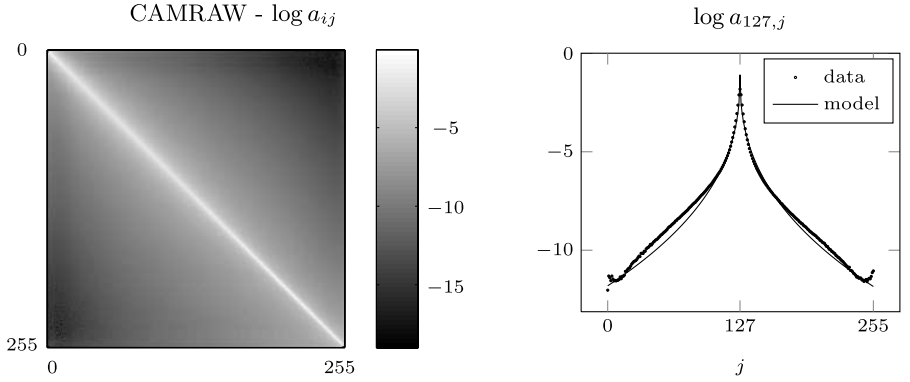

Fig. 2. Left: plot of the empirical matrix $\mathbb{A}$ estimated from CAMRAW database in log domain. Right: comparison of the 128th row of matrix $\mathbb{A}$ estimated from the same database with the analytic model (17).

The left part of Figure 3 shows the root rate (14), $r(\lambda)$, for a convex combination of LSB and ± 1 embedding, $\lambda \in [0, 1]$, for different image sources. The higher

[9] Expanded version of CAMERA_RAW database from [18] with 4547 8-bit images.
[10] Contains 2375 raw scans of negatives coming from the USDA Natural Resources Conservation Service (`http://photogallery.nrcs.usda.gov`).
[11] Images from NRCS database compressed with JPEG quality factor 70.

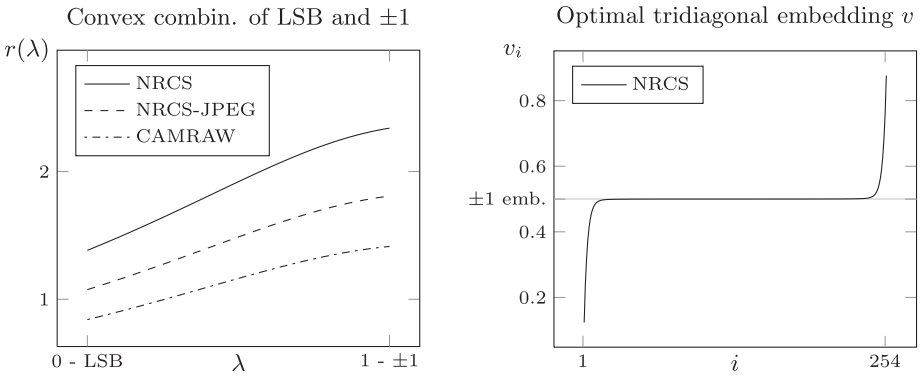

Fig. 3. Left: the root rate $r(\lambda) = \alpha_{MAX}/(\beta_{MAX}\sqrt{I})$ of a convex combination of LSB and ± 1 embedding for different image sources. Right: optimal parameters $v = (v_1, \ldots, v_{254})$ of MI embedding (15) minimizing the Fisher information rate while modifying cover elements by at most 1. The difference between ± 1 embedding and optimal MI embedding is due to boundary effects that vanish as $N \to \infty$.

the root rate $r(\lambda)$, the better the stegosystem. The results are consistent with the thesis that ± 1 embedding is less detectable than LSB embedding. Similarly, the capacity of stegosystems with covers from NRCS (scans) is believed to be higher than the capacity of stegosystem with decompressed JPEGs or images from digital cameras. This fact is in agreement with our result obtained for all values of the convex combination of LSB and ± 1 embedding and we attribute it to the fact that scans contain a higher level of noise that masks embedding changes. In contradiction with our expectations, decompressed JPEGs from NRCS-JPEG have a higher root rate than raw images from digital cameras (CAMRAW). This phenomenon is probably caused by the simplicity of the MC model, which fails to capture JPEG artifacts because they span across larger distances than neighboring pixels.

We now use the methodology described in Section 4.2 and maximize the root rate with respect to stegosystems that modify each cover element by at most 1. We do so for the cover model fit obtained from the NRCS database. Assuming the embedding operation is binary, it can embed one bit per cover element. Thus, it is sufficient to find the MI embedding that attains the minimum Fisher information rate. We use the parametrization from Example 1 and solve the system of equations (16). This system has only one solution $v = (v_1, \ldots, v_{254}) \in \mathcal{V} = [0, 1]^{254}$ and thus it represents MI embedding with minimum Fisher information rate. This solution is shown in the right part of Figure 3 along with the representation of the ± 1 embedding operation. The optimal MI embedding differs from ± 1 embedding only at the boundary of the dynamic range. This is due to the finite number of states in the MC model. We experimentally verified that the

relative number of states with $|v_i - 0.5| \geq \delta$ tends to zero for a range of $\delta > 0$ as $N \to \infty$ for fixed parameters of the analytic model.[12] Thus, the boundary effect is negligible for large N. This suggests that the loss in capacity when using ± 1 embedding algorithm is negligible for large N or, in other words, ± 1 embedding is asymptotically optimal.

6 Conclusion

In sharp contrast with the well established fact that the steganographic capacity of perfectly secure stegosystems increases linearly with the number of cover elements, n, a recently derived result states that steganographic capacity of a quite wide class of imperfect stegosystems is only proportional to $\sqrt{n}$. The communication rate of imperfect stegosystems is thus non-informative because it tends to zero with n. Instead, an appropriate measure of capacity is the constant of proportionality in front of $\sqrt{n}$, for which we coin the term the *root rate* whose unit is bit per square root of cover size per square root of KL divergence. The root rate is shown to be inversely proportional to the square root of the Fisher information rate of the stegosystem. Adopting a Markov model for the cover source, we derive an analytic formula for the root rate with Fisher information rate expressible as a quadratic form defined by the cover transition probability matrix evaluated at a vector fully determined by the embedding operation. This analytic form is important as it enables us to compare the capacity of imperfect stegosystems as well as optimize their embedding operation (maximize the root rate). We fit a parametric model through the empirical transition probability matrix for neighboring pixels of real images and use this model to compute and compare the root rate of known steganographic schemes and their convex combinations. In agreement with results previously established experimentally using blind steganalyzers, our analysis indicates that ternary ± 1 embedding is more secure than LSB embedding and it is also optimal among all embedding methods that modify pixels by at most 1. Furthermore, by analyzing image databases of raw images from different sources, we established that the root rate is larger for images with higher noise level as is to be expected. Among the surprising results of our effort, we point out the fact that the root rate for ± 1 embedding is only about twice larger than for LSB embedding, which contrasts with the fact that current best steganalyzers for LSB embedding are markedly more accurate than the best steganalyzers of ± 1 embedding. This hints at the existence of significantly more accurate detectors of ± 1 embedding that are yet to be found.

The results presented here offer several interesting research directions worth pursuing in the future. In particular, we may attempt to determine the embedding operation that maximizes the root rate for a given Markov cover source for a wider class of matrices $\mathbb{C}$. Additionally, we intend to extend our results to JPEG images represented in the DCT domain using appropriate analytic models.

[12] We believe the same to be true for all $\delta > 0$.

Acknowledgements

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9550-08-1-0084. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of AFOSR or the U.S. Government.

References

1. Wang, Y., Moulin, P.: Perfectly secure steganography: Capacity, error exponents, and code constructions. IEEE Transactions on Information Theory, Special Issue on Security (June 2008)
2. Comesana, P., Pérez-Gonzáles, F.: On the capacity of stegosystems. In: Dittmann, J., Fridrich, J. (eds.) Proceedings of the 9th ACM Multimedia & Security Workshop, Dallas, TX, September 20-21, pp. 3–14 (2007)
3. Harmsen, J.J., Pearlman, W.A.: Capacity of steganographic channels. Submitted to IEEE Transactions on Information Theory (2008), `http://arxiv.org/abs/0810.4171`
4. Filler, T., Fridrich, J., Ker, A.D.: The square root law of steganographic capacity for Markov covers. In: Delp, E.J., Wong, P.W., Memon, N., Dittmann, J. (eds.) Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia XI, San Jose, CA, January 18-21 (2009)
5. Ker, A.D., Pevný, T., Kodovský, J., Fridrich, J.: The square root law of steganographic capacity. In: Ker, A., Dittmann, J., Fridrich, J. (eds.) Proceedings of the 10th ACM Multimedia & Security Workshop, Oxford, UK, September 22-23, pp. 107–116 (2008)
6. Ker, A.D.: A capacity result for batch steganography. IEEE Signal Processing Letters 14(8), 525–528 (2007)
7. Filler, T., Fridrich, J.: Complete characterization of perfectly secure stego-systems with mutually independent embedding operation. In: Proceedings IEEE, International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, April 19-24 (2009)
8. Ker, A.D.: Estimating steganographic Fisher information in real images. In: Information Hiding, 11th International Workshop, Darmstadt, Germany. LNCS. Springer, Heidelberg (June 7-10, 2009)
9. Kodovský, J., Fridrich, J., Pevný, T.: Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In: Dittmann, J., Fridrich, J. (eds.) Proceedings of the 9th ACM Multimedia & Security Workshop, Dallas, TX, September 20-21, pp. 3–14 (2007)
10. Doob, J.L.: Stochastic processes, 1st edn. Wiley, New York (1953)
11. Sullivan, K., Madhow, U., Manjunath, B., Chandrasekaran, S.: Steganalysis for Markov cover data with applications to images. IEEE Transactions on Information Forensics and Security 1(2), 275–287 (2006)
12. Sidorov, M.: Hidden Markov models and steganalysis. In: Dittmann, J., Fridrich, J. (eds.) Proceedings of the 6th ACM Multimedia & Security Workshop, Magdeburg, Germany, September 20-21, pp. 63–67 (2004)

13. Cachin, C.: An information-theoretic model for steganography. In: Aucsmith, D. (ed.) IH 1998. LNCS, vol. 1525, pp. 306–318. Springer, Heidelberg (1998)
14. Ker, A.D.: The ultimate steganalysis benchmark? In: Dittmann, J., Fridrich, J. (eds.) Proceedings of the 9th ACM Multimedia & Security Workshop, Dallas, TX, September 20-21, pp. 141–148 (2007)
15. Kay, S.M.: Fundamentals of Statistical Signal Processing, Detection Theory, vol. II. Prentice-Hall, Englewood Cliffs (1998)
16. Filler, T.: Important properties of normalized KL-divergence under HMC model. Technical report, DDE Lab, SUNY Binghamton (2008), `http://dde.binghamton.edu/filler/kl-divergence-hmc.pdf`
17. Fridrich, J., Goljan, M.: Digital image steganography using stochastic modulation. In: Delp, E.J., Wong, P.W. (eds.) Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents V, Santa Clara, CA, January 21-24, vol. 5020, pp. 191–202 (2003)
18. Goljan, M., Fridrich, J., Holotyak, T.: New blind steganalysis and its implications. In: Delp, E.J., Wong, P.W. (eds.) Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII, San Jose, CA, January 16-19, vol. 6072, pp. 1–13 (2006)
19. Filler, T.: Fisher information determines capacity of ε-secure steganography - proofs. Technical report, SUNY Binghamton (2009), `http://dde.binghamton.edu/filler/pdf/Fill09ihwproofs.pdf`

Appendix

Proof of Theorem 2: Here, we only present the main idea of the proof, leaving all technical details to the report [19]. The decomposition of the sequence $I_n(0)/n$ to a quadratic form and its properties can be obtained directly from the definition of Fisher information

$$\frac{1}{n}I_n(0) = \frac{\ln 2}{n}\frac{\partial^2}{\partial\beta^2}d_n(\beta)\Big|_{\beta=0} =$$
$$= -\sum_{(i,j)}\sum_{(k,l)}\frac{\ln 2}{n\ln 2}E_P\Big[\underbrace{\Big(\frac{\partial^2}{\partial b_{ij}b_{kl}}\ln Q_\beta(Y_1^n)\Big|_{\mathbb{B}=\mathbb{I}}\Big)}_{\triangleq g(Y_1^n,i,j,k,l)}\Big]\underbrace{\Big(\frac{\partial b_{ij}}{\partial\beta}\Big|_{\beta=0}\Big)}_{=c_{ij}}\underbrace{\Big(\frac{\partial b_{kl}}{\partial\beta}\Big|_{\beta=0}\Big)}_{=c_{kl}}.$$

The derivatives of the log-likelihood are evaluated at $\mathbb{B}=\mathbb{I}$ because $\mathbb{B}(\beta) = \mathbb{I}+\beta\mathbb{C}$ and $\beta = 0$. By using $Q_\beta(y_1^n) = \sum_{x_1^n\in\mathcal{X}^n}P(x_1^n)Q_\beta(y_1^n|x_1^n)$, the random variable $g(Y_1^n,i,j,k,l)$ does not depend on the embedding method. This is because the derivatives are evaluated at $\mathbb{B}=\mathbb{I}$ and thus only contain the elements of the cover source transition matrix $\mathbb{A}$. The proof of the convergence of $-\frac{1}{n}E_P[g(Y_1^n,i,j,k,l)]$ to $f_{(i,j),(k,l)}$ and its closed form is more involved and is presented in the report [19]. The semidefinitness of the quadratic form follows from semidefiniteness of the Fisher information matrix $\mathbb{F}$. It is not positively definite because for an i.i.d. cover source all rows of matrix $\mathbb{F}$ coincide and are thus linearly dependent.