Synchronization Minimizing Statistical Detectability for Side-Informed JPEG Steganography

Quentin Giboulot, Patrick Bas, Rémi Cogranne

Synchronization Minimizing Statistical Detectability for Side-Informed JPEG Steganography

Quentin Giboulot
Troyes University of Technology,
ROSAS dept., LM2S Lab.
Email: quentin.giboulot@utt.fr

Patrick Bas
CNRS and École Centrale de Lille,
CRIStAL Lab.
Email: Patrick.Bas@centralelille.fr

Rémi Cogranne
Troyes University of Technology,
ROSAS dept., LM2S Lab.
Email: remi.cogranne@utt.fr

***Abstract*—Current schemes in steganography relying on synchronization are all based on a general heuristic to take into account interactions between embedding changes. However these approaches, while often competitive, lack a clear model for the relationship between pixels/DCT coefficient and the distortion function, and, as such, do not give any guarantees in terms of detectabilty. To solve this problem, we herein propose a synchronized side-informed scheme in the JPEG domain based on minimizing statistical detectability which achieves state-of-the-art performances. This is done by exploiting a statistical model that takes into account correlations between DCT coefficients and adding an optimal steganographic-signal with covariance which is a scaled version of the cover noise covariance. This method allows a clear understanding of the reasons why, depending on the processing pipeline, synchronization using both intra and inter-block dependencies allows such gains in performance.**

I. Introduction

Imperfect steganography has historically always been grounded on the assumption of independence between samples (pixels or DCT). Indeed, most of the current-art steganographic scheme assumes that the modification of one cover element does not affect the detectability over neighboring elements and hence does not change the probability of modifying another. Recently however, several strategies have been proposed to take into account interactions between embedding changes. In fact there is a striking asymmetry between steganography and steganalysis on this assumption of independent cover samples. On the one hand, steganography largely relies on this assumption, in large part because of practical aspects such as the use of the STC that use an addition distortion function. On the other hand, steganalysis does exploit more and more accurately the correlation between neighboring samples in order to detect minor changes due to data hiding.

Roughly speaking, there have been two types of approaches that have been proposed in order to exploit sample correlation in the design of steganography. The most general of method uses the so-called Gibbs construction [1] to allow the use of any non-additive distortion function which can be expressed as a sum of locally supported potential. Despite its generality, the success of this approach has been quite mild as it did not shed any light on the link between a non-additive distortion function and empirical detectability. Without such a link, a number of heuristic schemes have been proposed [2]–[4]. While those approaches allow improving substantially the performance of steganography with respect to the empirical detectors, these heuristic approaches do not give any guarantees of performance as their distortion functions are not based on any statistical model. Yet, there now exist a few approaches, both in the spatial [5], [6] and JPEG domain [7]–[9] for designing schemes using the framework initially proposed in [10], [11] based on hypothesis testing and minimizing the power of the most powerful detector. This framework has the advantage of giving guarantees of performance in terms of detectability as long as the cover and stego under scrutiny follow the model assumed by the steganographic scheme .

A scheme using synchronization under this framework has yet to be proposed. This is the main contribution of this work.

In this paper, we design a scheme, focusing on the JPEG domain though the method can also be used in the spatial domain, which minimizes statistical detectability through the use of synchronization. More specifically, we extend our previous work on Gaussian embedding [9] to take into account correlations between DCT coefficients. In the first section, we present the model of the cover and stego model and derive the performance of the optimal detector as well as the optimal form of the stego signal. In the second section we design a multivariate version of Gaussian embedding which fully takes the covariance of DCT blocks into account using the results of Section II. We finally present the results of our method in the last section for different processing pipelines and quality factors.

II. Cover and stego models

In this section, we present the model that will be used in the design of our synchronized embedding scheme. The model we consider here is applicable both to the spatial and JPEG domain. Following the same motivations as in [9], [12]–[14], [23], we model the precover as an image consisting of N independent $M \times M$ blocks P_i following a multivariate Gaussian. Though the signal will eventually be embedded in the discretized domain, we model a steganographer trying to minimize the detectability in the continuous domain with a payload constraint in the discrete domain. To do so, we model a "pre-stego" $\mathbf{Q}$ as the image to which was added

signal consisting of N independent $M \times M$ blocks following a multivariate Gaussian with zero-mean. Formally, let :

$$P_i \sim \mathcal{N}(\mu_i, \Sigma_i), \tag{1}$$

$$Q_i \sim \mathcal{N}(\mu_i, \Sigma_i + \boldsymbol{\epsilon}_i). \tag{2}$$

Following the methodology proposed in [5], we design a steganographic scheme which generates a signal that minimizes the power of the most powerful (MP) detector. To cast the steganography problem into the continuous domain, we will work under the assumption that the Warden knows $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_N)$ as well as the model parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)$ and $\boldsymbol{\Sigma} = (\Sigma_1, \Sigma_2, \ldots, \Sigma_N)$ and analyses the image blocks before rounding $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$. The Warden's goal is to decide between the two hypotheses $\forall i \in \{1, 2, \ldots, N\}$:

$$\begin{cases} \mathcal{H}_0 & = \{z_i \sim \mathcal{N}(\mu_i, \Sigma_i)\}, \\ \mathcal{H}_1 & = \{z_i \sim \mathcal{N}(\mu_i, \Sigma_i + \boldsymbol{\epsilon}_i)\}. \end{cases} \tag{3}$$

Let the pdf of the noise distribution under $\mathcal{H}_0$, $p_{\Sigma_i}(x)$, and $q_{\Sigma_i, \boldsymbol{\epsilon}_i}(x)$ under $\mathcal{H}_1$ as

$$p_{\Sigma_i}(x) = \frac{\exp\left((x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right)}{\sqrt{2\pi|\Sigma_i|}} \tag{4}$$

$$q_{\Sigma_i, \boldsymbol{\epsilon}_i}(x) = \frac{\exp\left((x-\mu_i)^T (\Sigma_i + \boldsymbol{\epsilon}_i)^{-1} (x-\mu_i)\right)}{\sqrt{2\pi|\Sigma_i + \boldsymbol{\epsilon}_i|}}. \tag{5}$$

We can then use the Neyman-Pearson criterion of optimality. In this case the Warden constructs a test $\delta : \mathbb{R} \to \{\mathcal{H}_0, \mathcal{H}_1\}$ which maximizes the power of the test $P_D \triangleq \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_1)$ under a given false-alarm probability $P_{FA} \triangleq \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_0)$.

Under these assumptions, the problem of the Warden (3) is reduced to a choice between two simple hypotheses for which the Neyman-Pearson Lemma states that the most-powerful test is the likelihood ratio test (LRT), defined, in our case as follows:

$$\Lambda_i(z, \Sigma_i, \boldsymbol{\epsilon}_i) = \ln\left(\frac{p_{\Sigma_i}(z)}{q_{\Sigma_i, \boldsymbol{\epsilon}_i}(z)}\right), \tag{6}$$

$$\Lambda(\mathbf{z}, \boldsymbol{\Sigma}, \boldsymbol{\epsilon}) = \sum_{i=0}^{N} \Lambda_i(z_i, \Sigma_i, \boldsymbol{\epsilon}_i) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \tau, \tag{7}$$

since we assume independence between image blocks.

We now give, without proof due to space constraints, the asymptotic power of the LRT.

Let $\mathbf{A}_i = \Sigma_i^{-1} - (\Sigma_i + \boldsymbol{\epsilon}_i)^{-1}$ and let the j-th eigenvalue of $\mathbf{A}_i \Sigma_i$ be written as k_j^0 and the j-th eigenvalue of $\mathbf{A}_i (\Sigma_i + \boldsymbol{\epsilon}_i)$ as k_j^1. The first two moments of the LRT under each hypothesis are given by:

As the number N of independent blocks of cover elements $N \to \infty$, Linderberg's central limit theorem implies that:

$$\Lambda(\mathbf{z}, \boldsymbol{\Sigma}, \boldsymbol{\epsilon}) \rightsquigarrow \begin{cases} \mathcal{N}(\mathbb{E}_{\mathcal{H}_0}, \mathrm{Var}_{\mathcal{H}_0}), & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\mathbb{E}_{\mathcal{H}_1}, \mathrm{Var}_{\mathcal{H}_1}), & \text{under } \mathcal{H}_1 \end{cases} \tag{8}$$

	$\mathbb{E}_{\mathcal{H}_i}[\Lambda]$	$Var_{\mathcal{H}_i}[\Lambda]$
0	$-\sum_{i=1}^{N} D_{KL}\left(p_{\Sigma_i^c} \| q_{\Sigma_i^s}\right)$	$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}(k_{ij}^0)^2$
1	$\sum_{i=1}^{N} D_{KL}\left(q_{\Sigma_i^s} \| p_{\Sigma_i^c}\right)$	$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}(k_{ij}^1)^2$

where $\rightsquigarrow$ denotes convergence in distribution.

One can easily establish the power function and the false alarm probability of the LR test using the limiting distribution of the log-likelihood ratio (11). To this end we will first establish the threshold τ that guarantees that the test (7) satisfies a given constraint on the false alarm rate:

$$P_{FA}(\tau) = \mathbb{P}(\Lambda(\mathbf{z}, \boldsymbol{\Sigma}, \boldsymbol{\epsilon}) > \tau | \mathcal{H}_0) = Q\left(\frac{\tau - \mathbb{E}_{\mathcal{H}_0}}{\sqrt{\mathrm{Var}_{\mathcal{H}_0}}}\right),$$

$$\Leftrightarrow \qquad \tau = Q^{-1}(P_{FA})\sqrt{\mathrm{Var}_{\mathcal{H}_0}} - \mathbb{E}_{\mathcal{H}_0}. \tag{9}$$

where Q is the tail distribution function of the standard normal distribution.

Similarly the power function is given

$$P_D(\tau) = \mathbb{P}(\Lambda(\mathbf{z}, \boldsymbol{\Sigma}, \boldsymbol{\epsilon}) > \tau | \mathcal{H}_1) = Q\left(\frac{\tau - \mathbb{E}_{\mathcal{H}_1}}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}}}\right). \tag{10}$$

Replacing in (10) the expression of the threshold given in (9) eventually yields to:

$$P_D = \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_1) \tag{11}$$

$$= Q\left(\frac{Q^{-1}(P_{FA})\sqrt{\mathrm{Var}_{\mathcal{H}_0}} + \mathbb{E}_{\mathcal{H}_1} - \mathbb{E}_{\mathcal{H}_0}}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}}}\right), \tag{12}$$

We show in the appendix that the "optimal" covariance stego signal, that minimizes the power function, has the following form:

$$\boldsymbol{\epsilon}_i = \alpha \Sigma_i, \tag{13}$$

with $\alpha \in \mathbf{R}^+$.

III. Multivariate Gaussian embedding

In this section we design a synchronized embedding scheme, which we will name Multivariate Gaussian embedding (MGE), which is able to take into account correlations between neighboring pixels/DCT coefficients by leveraging the model described in the preceding section. Due to space considerations, we will here only consider the case of a Payload-Limited Sender (PLS).

In our case, the payload-limited sender wants to minimize the power of the LRT (11) under a given payload constraint:

$$\begin{cases} \min\limits_{\boldsymbol{\epsilon}_i} & P_D(\boldsymbol{\epsilon}_i) \\ R & = \sum\limits_{i=0}^{n}\sum\limits_{j\in\mathbb{Z}} \beta_i^{(j)} \log\left(\beta_i^{(j)}\right) \end{cases} \tag{14}$$

where $\beta_i^{(j)}$ is the probability of adding $+j$ to the i^{th} coefficient, n is the total number of elements (DCT coefficients) and R is the desired payload in bits. Also note that we minimize

the power of the detector in the continuous domain while the payload constraint is expressed in the quantized domain.

Assuming each $M \times M$ block of DCT coefficient independent and using the results of the appendix; $\boldsymbol{\epsilon}_i$ has the form given in Eq (13). To minimize P_D under the payload constraint we therefore only have to perform a binary search on α until the constraint is met.

The problem is now reduced to computing the probabilities of adding $+j$ to the i^{th} coefficient $\beta_i^{(j)}$ efficiently. To do so we compute the Gaussian parameters $\bar{\mu}_i$ and $\bar{\epsilon}_i$ of each element of the stego-signal conditioned on the realizations of previous embedding. This can be done efficiently using the Cholesky decomposition of Σ_i :

$$\Sigma_i = \mathbf{L}_i \mathbf{L}_i^T. \tag{15}$$

Using the Cholesky decomposition, we can first sample the stego-signal as realizations of a standard normal distribution $\mathbf{x}$. We then correlate the samples using $\mathbf{L}$:

$$\mathbf{x}_i' = \mathbf{L}_i \mathbf{x}_i, \tag{16}$$

finally scaling them accordingly to meet the constraint.

The parameters are then obtained by :

$$\bar{\epsilon}_i = \sqrt{\alpha} \mathbf{L}_{ii}, \tag{17}$$

$$\bar{\mu}_i = \sqrt{\alpha} \left(\mathbf{x}_i' - \bar{\epsilon}_i \mathbf{x}_i \right). \tag{18}$$

The $\beta_i^{(j)}$ are then obtained by :

$$\begin{aligned} \beta_i^{(j)} = & \boldsymbol{\Phi} \left(\frac{j - r_i - \bar{\mu}_i + 0.5}{\bar{\epsilon}_i} \right) \\ & - \boldsymbol{\Phi} \left(\frac{j - r_i - \bar{\mu}_i - 0.5}{\bar{\epsilon}_i} \right), \end{aligned} \tag{19}$$

where $\boldsymbol{\Phi}(\cdot)$ represents the cumulative distribution function of the standard normal distribution and $r_i = x_i - [x_i]$ denotes the rounding error of i-th DCT coefficient. In practice, the alphabet size of the embedding scheme is finite; j is thus constrained to a finite range and the $\beta_i^{(j)}$ must be normalized accordingly.

IV. Multivariate Embedding in Practice

In the preceding section we showed how to simulate MGE for a given covariance matrix of the cover noise. In this section we explain how we estimate the covariance matrix in practice and how the block size is chosen. We also give the rationale for using this scheme with Syndrome-Trellis-Codes [15] (STC).

A. Estimation of the Covariance Matrix

Before estimating the covariance matrix, the steganographer has to decide what block size $M \times M$ to use. This block size determines what dependencies are taken into account by the embedding scheme. If a block size of 8×8 is chosen, only intra-block dependencies are taken into account while a higher block-size will allow capturing inter-block dependencies. This choice is then a trade-off between computational tractability of the covariance matrix and security of the embedding scheme. In practice we decided to fix the block size as 24×24 in order to capture inter-block dependencies in every direction (horizontal, vertical and diagonal blocks) while still allowing fast computation of the covariance matrix.

The covariance matrix itself is estimated using the method described in [9, Section II]. The $\mathbf{H}$ matrix modeling the processing pipeline is estimated using a simple least square regression. To that end, we use a synthetic constant RAW image to which centered Gaussian noise with constant variance is added. This image is then processed using the relevant processing pipeline. When using both intra-block and inter-block dependencies, the RAW and developed images are then reshaped as arrays of $(24k + 2) \times (24k + 2)$ and 24×24 blocks respectively where k is the resize factor of the image from the RAW domain to the JPEG domain (which is equal to 1 if no resizing was performed). When using only intra-block dependencies, we use $(8k + 2) \times (8k + 2)$ and 8×8 respectively. Denoting the blocks in the RAW domain and in the developed domain as $\mathbf{X}^{RAW}$ and $\mathbf{X}^{dev}$, respectively, $\mathbf{H}$ is obtained by solving:

$$\mathbf{X}^{dev} = \mathbf{H} \mathbf{X}^{RAW}. \tag{20}$$

B. Embedding using STC

We now outline an implementation of MGE using STCs. When using STCs, we need to compute the embedding probabilities $\beta_i^{(j)}$ using the methodology given in the preceding section and convert them to costs using:

$$\beta_i^{(j)} = \frac{e^{-\lambda \rho_i^{(j)}}}{1 + \sum_{j \in \mathcal{A} \setminus \{0\}} e^{-\lambda \rho_i^{(j)}}}. \tag{21}$$

However, the probability of embedding in the k-th coefficient of a given block depends on the actual embedding performed by the STC on all the preceding coefficients in that block. Such an implementation will thus require to be performed iteratively.

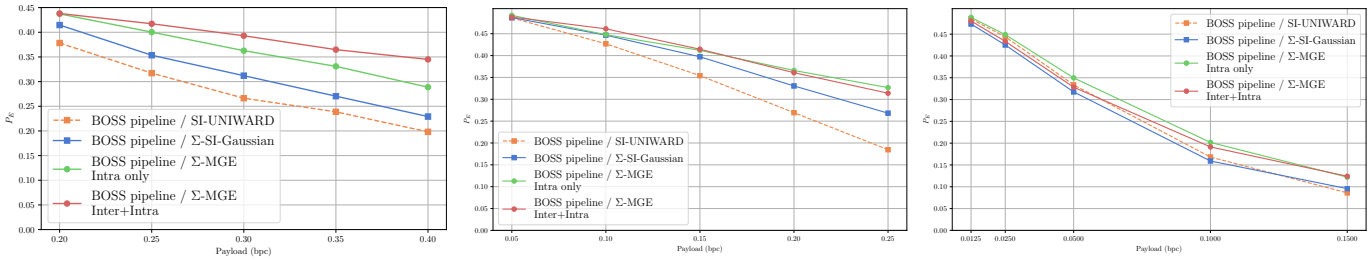
In the first iteration we compute the embedding probabilities in the first coefficient of each block. Here, we necessarily have $\bar{\mu}_i = 0$ for all coefficients, $\beta_i^{(j)}$'s can thus directly be computed and converted to costs. Once the payload has been embedded, the $\mathbf{x}_i$ corresponding to each of these coefficient must be computed. To do so we can sample $\mathbf{x}_i'$ using rejection sampling until the rounded value of $\mathbf{x}_i'$ matches the actual embedding. We then compute $\mathbf{x}_i$ using Eq. (16).

The k-th iteration is done in exactly the same manner, for the k-th coefficient of each block, except, this time, we compute $\bar{\mu}_i$ using the previously computed $\mathbf{x}_i$ and Eq (18).

The only caveat with this approach is that α must be fixed before the embedding and be identical for each lattice. However, since the actual entropy depends on the $\beta_i^{(j)}$, hence on the actual embedding performed, it is theoretically not possible to compute the minimal α that would allow to reach the payload size. In practice, this is not such a problem as we observed during simulations that the realizations of the $\mathbf{x}_i$ play a very small role on the entropy; it will usually make it change by 1 or 2 nats at most if at all. A good rule of thumb would

TABLE I: Names and operations of the processing pipelines used in the experiments. Gamma correction is never performed except when explicitly stated. The operations are performed in the order they are presented in the table

Pipeline name	**Demosaicking**	**White Balance**	**RGB to grey**	**Downsampling method**
Linear Pipeline	Bilinear	No	Yes	Edge crop, 256×256
BOSS Pipeline	PPG	Yes, Camera	Yes	Resize from 768×768 (Edge crop) to 256×256, Lanczos kernel



(a) QF100 (b) QF95 (c) QF75

Fig. 1: Comparison of the proposed embedding schemes with prior art using the P_E as a function of the payload size for BossBase developed with the BOSS pipeline.

then be to find the optimal α for a slightly higher payload and to use it to ensure the message will fit on the first try.

V. Results

In this section, we study the performance of our different extensions of Gaussian Embedding in the JPEG domain. In order to have access to a precise estimation of the covariance matrix, we use the estimation method described in our previous work [9, Section II]. Consequently we use the knowledge of the RAW file and of the processing pipeline. We use the BOSS [16] RAW dataset excluding the M9 camera because of its peculiar distribution of its photonic (see [17], Fig. 2) which would lead to imprecise estimation of the covariance matrix. From this dataset comprising 7642 RAW images taken with 6 different cameras we produce two new datasets using two different processing pipeline : a linear processing pipeline and a processing pipeline close to the original BOSSBase [16]. The details are exposed in Table I.

The different embedding schemes used as well as their parameters are described in Table II. Steganalysis was performed with DCTR [18] and the Low-complexity Linear Classifier (LCLC) [19]. Eventually, the empirical security of the schemes is evaluated using the minimal probability of error under equal priors: $P_E = \min_{P_{FA}} \frac{P_{MD}+PFA}{2}$. Results are given in Figure 1 and 2.

TABLE II: Nomenclature of the embedding schemes

Name	**Meaning**
Σ-SI-Gaussian	Minimizes the power of the MP detector in the continuous domain supposing the DCT coefficient to be independent as described in [9].
Σ-MGE Intra Only	Minimizes the power of the MP detector in the continuous domain supposing 8×8 DCT block to be independent.
Σ-MGE Intra+Inter	Minimizes the power of the MP detector in the continuous domain supposing 24×24 DCT block to be independent.
SI-UNIWARD	Side informed distortion based schemes as described in [20].

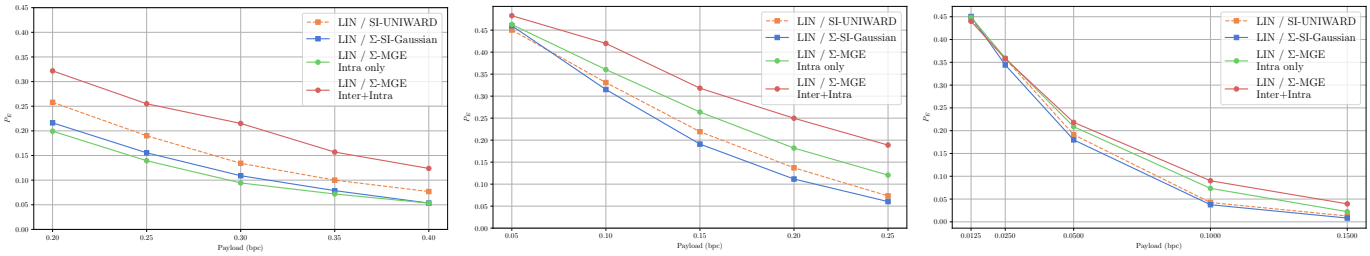
First of all, one can observe that performance improves drastically when using the intra AND inter block dependencies especially for higher quality factors – ie. QF100 and QF95. The average absolute gain in P_E with respect to SI-UNIWARD ranges from 6% to 9% respectively for the linear pipeline and from 11% to 6% respectively for the BOSS pipeline. This gain is only of 2% on average for QF75.

However, we also observe an average relative gain of 1.5 which is consistent across every pipeline and quality factor, except at the two lowest payloads for the BOSS pipeline at QF75 where all the schemes perform approximately the same.

Interestingly, at QF100 with the linear pipeline, using only intra-block dependencies does not bring any gain in performance compared to the scheme where the DCT coefficients are deemed independent. However, the performance of the method using only the intra-block tends to match with the one using both inter and intra-block dependencies as the QF gets lower. This phenomenon was already observed with Natural Steganography [22] where not taking into account inter-block dependencies would lead to a useless scheme at QF100, yet the scheme would still perform acceptably for lower QF.

This phenomenon can be explained by the fact that the inter-block covariances tend to be smaller by one or two order of magnitude than the intra-block covariances. Higher quantization steps and rounding will thus tend to reinforce this fact, making the inter-block dependencies, and even eventually the intra-block dependencies negligible.

The fact that this phenomenon is much less pronounced for the Bosslike pipeline has to do with the fact that the demosaicking algorithm, PPG, works independently on 8×8 blocks, thus not creating inter-block dependencies in the JPEG domain by itself. The only inter-block dependencies which are introduced are thus due to the filtering used for the resizing. However since we use a resize factor of 3, most of these dependencies are lost during the decimation process



(a) QF100 (b) QF95 (c) QF75

Fig. 2: Comparison of the proposed embedding schemes with prior art using the P_E as a function of the payload size for BossBase developed with the Linear pipeline.

of the resizing process. The gain in using the inter-block dependencies hence disappears far quicker with respect to the QF than for the linear pipeline.

VI. Conclusion

In this paper, we presented a novel method able to take correlations between cover elements into account. In particular, we showed that, under some assumptions on the precover, the optimal covariance of the prestego signal is simply a scaling of the precover covariance. We then studied the performance of this scheme and observed that using both intra and inter-block dependencies consistently gave the best performance. As future work, we will expand this scheme to forego the assumption of independence between blocks in the model. We will also propose a method to be able to use this scheme when the RAW file is unavailable to the steganographer.

Appendix

A. Representation of the Optimal Stego Covariance

In this appendix we show that the optimal covariance of the stego signal when the cover is corrupted by a non-stationary multivariate Gaussian noise is a scaling of the covariance of the cover noise.

This proof relies on three steps. In the first part we show that if the cover element are independent Gaussians random variables, then then the Gaussian stego signal that minimizes the DKL, is also composed of independent element. Following this, we show that one optimal detector in the general case of a non-stationary multivariate signal can be expressed in the domain where the noise is independent. From this, we use the result derived in [9] that if the stego signal and the cover element are independent Gaussians, the variance of the optimal stego signal is a scaled version of the variance of the corresponding cover element. We then conclude on the form of the optimal covariance of the stego signal in the original domain.

Let $\mathbf{P}$ be a precover image with N elements corrupted by a multivariate Gaussian noise:

$$\mathbf{P} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_c) \tag{22}$$

where Σ_c is a $N \times N$ covariance matrix. Similarly let the pre-stego $\mathbf{Q}$ follow a centered multivariate Gaussian noise :

$$\mathbf{Q} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma_s) \tag{23}$$

The steganographer wants to minimize the power of the optimal detector given a payload constraint R.

1) Independant stego signal for independant Gaussian cover noise: We here show that if Σ_c is diagonal, which we will write as $\Sigma_c = \mathrm{diag}(\sigma_c^2)$ then one of the optimal covariance for the prestego is also diagonal. In this subsection only, we will write Σ_s to talk about the non diagonal stego covariance matrix, while we will re-write $\mathrm{diag}(\sigma_s^2)$ if we talk a bout the diagonal one.

As $N \to \infty$, it is sufficient to show that the KL-divergence between the cover and the stego is greater if Σ_s is not diagonal. We will assume here than N is large enough for the asymptotic regime to hold.

Now, using the chain rule of KL-divergence and the fact that the cover elements are diagonal:

$$D_{KL}(P||Q) = \sum_{i=1}^{N} D_{KL}\left(p(z_i)||q(z_i|z_1 \dots z_{i-1})\right). \tag{24}$$

If $\Sigma_s = \mathrm{diag}(\sigma_s^2)$ then:

$$D_{KL}(P||Q) = \sum_{i=1}^{N} \log\left(\frac{\sigma_{s,i}}{\sigma_{c,i}}\right) + \sum_{i=1}^{N} \frac{\sigma_{c,i}^2}{2\sigma_{s,i}^2} - \frac{N}{2}. \tag{25}$$

If Σ_s is not diagonal, then:

$$D_{KL}(P||Q) = \sum_{i=1}^{N} \log\left(\frac{\bar{\sigma}_{s,i}}{\sigma_{c,i}}\right) + \sum_{i=1}^{N} \frac{\sigma_{c,i}^2}{2\bar{\sigma}_{s,i}^2} + \sum_{i=1}^{N} \frac{\bar{\mu}_{s,i}^2}{2\bar{\sigma}_{s,i}^2} - \frac{N}{2}, \tag{26}$$

where $\bar{\mu}_{s,i}$ and $\bar{\sigma}_{s,i}^2$ is the mean and variance of the i-th conditioned Gaussian in Eq (24).

Now we use the payload constraint to observe that:

$$\log(|\Sigma_s|) = \log\left(|\mathrm{diag}(\sigma_s^2)|\right) \tag{27}$$

$$= \sum_{i=1}^{N} \log\left(2\pi e \sigma_{s,i}^2\right) \tag{28}$$

$$= \sum_{i=1}^{N} \log\left(2\pi e \bar{\sigma}_{s,i}^2\right). \tag{29}$$

Now, it is easy to show, using the technique of Lagrange multipliers in the same way as in [9], that there is a unique solution which minimizes the first and the second term of the LHS of Eq (25) and Eq (26) for a given constraint. Since the constraint is the same for both cases we can conclude that in order to minimize the first two terms of the LHS for both cases, we necessarily have $\mathrm{diag}(\sigma_{s,i}^2) = \mathrm{diag}(\bar{\sigma}_{s,i}^2) \ \ \forall i \in \{1 \dots N\}$.

However, since $\bar{\mu}_{s,i}^2 \geq 0$, the minimum D_{KL} when Σ_s is not diagonal is necessarily greater than the D_{KL} when it is.

2) Optimal detector: Let the eigendecomposition of Σ_c be written as :

$$\Sigma_c = \mathbf{U}\mathbf{K}\mathbf{U}^{-1}, \tag{30}$$

and

$$\mathbf{z}_u = \mathbf{U}\mathbf{z}. \tag{31}$$

Since Σ_c is symmetric positive definite, it follows that $\mathbf{U}$ is a full-rank non-singular matrix. Hence, expressing the samples in this new basis has no impact on the power of the LRT. An equivalent expression of the optimal test expressed in Eq (7) is thus (see [21, Chapter 3 Section 3] for a derivation):

$$\mathbf{z}_u^T \left(\mathbf{K}^{-1} - \left(\mathbf{U}\Sigma_s\mathbf{U}^T \right)^{-1} \right) \mathbf{z}_u \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \tau_u, \tag{32}$$

Let $\mathbf{K}_\epsilon = \mathbf{U}\Sigma_s\mathbf{U}^T - \mathbf{K}$ be the covariance of the stego signal in this new basis. Since $\mathbf{K}$ is diagonal, we have seen that the optimal $\mathbf{K}_\epsilon$ must also be diagonal. More precisely, using the results of our preceding work [9] we know that the optimal $\mathbf{K}_\epsilon$ is given by scaling of each variance by the Lagrange multiplier λ:

$$\mathbf{K}_\epsilon = \sqrt{\frac{\lambda}{2}}\mathbf{K} \triangleq \alpha\mathbf{K}, \tag{33}$$

Going back to the original basis of the samples, the expression of the optimal stego covariance matrix is directly given by:

$$\boldsymbol{\epsilon} = \mathbf{U}^{-1}\alpha\mathbf{K}\mathbf{U} = \alpha\Sigma_c. \tag{34}$$

The optimal covariance of the stego signal is thus a scaling of the original covariance matrix of the cover noise. Furthermore note that the performance of the LRT is obtained in a similar way as for the independent case and thus its power can be quantified by:

$$\varrho = \sqrt{\sum_{i=1}^{N} \frac{k_{\epsilon i}^2}{2k_i^2}}, \tag{35}$$

where k_i^2 and $k_{\epsilon i}^2$ are the eigenvalues of Σ_c and $\Sigma_s - \Sigma_c$ respectively.

REFERENCES

[1] T. Filler and J. Fridrich "Gibbs Construction in Steganography". *IEEE* Transactions on Information Forensics and Security, vol. 5, no. 4, pp. 705–720, dec 2010.

[2] L. Bin and W. Ming, L. Xiaolong, T. Shunquan and H. Jiwu, "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 1905–1917, sep 2015.

[3] T. Denemark, and J. Fridrich, "Improving steganographic security by synchronizing the selection channel," in - *IHMMSec2015*.

[4] X. Hu, J. Ni, W. Su and J. Huang "Model based image steganography using asymmetric embedding," *Electronic Imaging*, vol. 27, no. 4, 2018.

[5] V. Sedighi, R. C ogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, Feb 2016.

[6] M. Sharifzadeh, M. Aloraini, and D. Schonfeld, "Quantized Gaussian embedding steganography," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

[7] R. Cogranne, Q. Giboulot and P. Bas "Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images," in *ACM IH&MMSEC*, pp. 161–167, Denver, USA, Jun 2020

[8] T. Denemark and J. Fridrich, "Model based steganography with precover," *Electronic Imaging*, vol. 2017, no. 7, pp. 56–66, Jan 2017.

[9] Q. Giboulot, R. Cogranne, and P. Bas, "JPEG steganography with side information from the processing pipeline," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelone, Spain, May 2020.

[10] J. Fridrich and J. Kodovskỳ, "Multivariate Gaussian model for designing additive distortion for steganography," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2949–2953, 2013.

[11] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-Adaptive Pentary Steganography Using the Multivariate Generalized Gaussian Cover Model," *SPIE Electronic Imaging*, ser. Media Watermarking, Security, and Forensics, Feb. 2015.

[12] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, oct 2008.

[13] T.H. Thai, R. Cogranne, and F. Retraint, "Camera model identification based on the heteroscedastic noise model," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250–263, Jan 2014.

[14] T.H. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients: Application in the steganalysis of jsteg algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980–1993, May 2014.

[15] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, sep 2011.

[16] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system — the ins and outs of organizing BOSS," in *Information Hiding, 13th International Workshop*, Prague, Czech Republic, May 18–20, 2011, pp. 59–70, LNCS vol.6958.

[17] T. Denemark, P. Bas, and J. Fridrich, "Natural Steganography in JPEG Compressed Images," in *Electronic Imaging*, *Proc. IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics* San Francisco, United States, Jan.2018.

[18] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, Feb 2015.

[19] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevny, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?," in *International Workshop on Information Forensics and Security (WIFS)*. Nov 2015, IEEE.

[20] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, jan 2014.

[21] H. Van Trees "Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory." John Wiley and Sons, 2004.

[22] T. Taburet, P. Bas, J. Fridrich, and W. Sawaya, "Computing dependencies between DCT coefficients for natural steganography in JPEG domain," in - *IHMMSec2019*.

[23] T. Taburet, P. Bas, W. Sawaya and J. Fridrich "Natural Steganography in JPEG Domain With a Linear Development Pipeline." in *IEEE Transactions on Information Forensics and Security, Vol 16, 2021, p 173-186*.