# EFFECT OF COVER QUANTIZATION ON STEGANOGRAPHIC FISHER INFORMATION

*Jessica Fridrich*

Department of ECE, SUNY Binghamton, NY, USA
fridrich@binghamton.edu

## ABSTRACT

This article presents an extension of the square root law of imperfect steganography to consider the effects of quantization on the steganographic Fisher information. We make the assumption that the cover elements are quantized i.i.d. samples drawn from an underlying continuous-valued 'precover' distribution. In the fine quantization limit, the Fisher information exhibits power scaling with an exponent determined jointly by the smoothness of the precover distribution and the properties of the embedding function. This extension is relevant for understanding the effects of pixel color depth and JPEG quality factor on secure payload of imperfect steganography realized using a mutually independent embedding operation.

## 1. INTRODUCTION

The square-root law (SRL) of steganography [2, 3, 6] is an asymptotic scaling result concerning the size of the secure payload for stegosystems that are imperfect, such as all stegosystems designed for empirical (and thus fundamentally incognizable) cover sources, example of which are digital media files. The law manifests when the Warden is granted the full knowledge of the cover source and the embedding method while the sender uses mutually independent embedding.[1] When the Warden needs to *learn* the cover source, the law may manifest in a different form depending on the knowledge available to the Warden [5]. The SRL becomes readily apparent when expanding the KL divergence between the cover and stego distributions at $\beta = 0$ (where $\beta$ is the change rate) using Taylor series – the leading quadratic term is $\frac{1}{2}N\beta^2 I(0)$, where $N$ is the cover length and $I(0)$ is the steganographic Fisher information [1, 4].

Digital media are typically obtained by quantizing an analog (precover) signal, such as photon counts or transform coefficients. The quantization affects the statistical properties of covers and thus the asymptotic scaling laws through changes in $I(0)$. In particular, this paper shows that $I(0) \propto \triangle^s$, where $\triangle$ is the quantization step and $s \geq 0$ is an exponent determined jointly by the embedding operation and the smoothness of the precover distribution. This generalizes the SRL as one can state that constant statistical detectability is obtained when $\frac{1}{2}N\beta^2\triangle^s = const.$ Unlike the SRL, which is quite robust and observable in practice even when the source is empirical and the Warden uses classifiers instead or optimal detectors, the scaling due to quantization sensitively depends on the precover distribution. This limits the current contribution to a theoretical treatment that cannot readily quantify the scaling in empirical cover sources with empirical measures of security (e.g., relating classifier error to color bit depth or JPEG quality factor).

The paper starts in the next section with a formalization of basic concepts. The main results are established in Section 3 for smooth precover distributions as well as distributions with a singularity. The theory is applied to four common embedding operations and two precover models in Section 4. Section 5 discusses the limitations of observing the scaling in practice when using empirical sources and detectors. A summary appears in Section 6.

An expanded journal version of this paper, that contains additional results as well as the proofs of all theorems, has been submitted to IEEE Transactions on Information Forensic and Security in June 2012.

## 2. PRELIMINARIES

Throughout the paper, $a_{ij}$, $i, j \in \mathbb{Z}$, denotes a (potentially infinite) two-dimensional array with elements $a_{ij}$. Calligraphic font is reserved for sets, while random variables are always represented with capital letters. For

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9950-12-1-0124. 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. The author thanks Jan Kodovský and Tomáš Filler for useful discussions.

[1] When both the sender and the Warden are fully informed, the sender can communicate with perfect security with a positive rate [8].

real functions $g, h$, we define $g(x) = \Theta(h(x))$ at $x = a$ if $G_1 h(x) \le g(x) \le G_2 h(x)$ for $G_1, G_2 > 0$ on some neighborhood of $a \in \{\mathbb{R}, -\infty, \infty\}$. We write $g \approx h$ whenever $\lim_{x\to a} g(x)/h(x)$ exists and is positive. We also use the standard Landau big-O and little-o notation.

Given a countable set of scalar bin centroids, $\mathcal{M} = (m_j)$, $m_j < m_{j+1}$, a scalar quantizer is a mapping $Q_{\mathcal{M}} : \mathbb{R} \to \mathcal{M}$, defined as $Q_{\mathcal{M}}(x) = \arg\min_{m_j\in\mathcal{M}} |x - m_j|$. Even though the main result of this paper could be established for a more general class of quantizers, for simplicity, we will assume that $Q_{\mathcal{M}}$ is uniform, $m_j = j\triangle$, $j \in \mathbb{Z}$, where $\triangle > 0$ is the bin width. A uniform quantizer with bin width $\triangle$ will be denoted $Q_\triangle$.

An $n$-element precover source will be represented using a random variable $Z \triangleq (Z_1, \dots, Z_n)$ where $Z_k$ are mutually independent and identically distributed (i.i.d.) continuous-valued random variables $Z_k \sim f(x)$, where $f(x)$ is a probability density function. For convenience, we will assume that $f$ is even and that its domain can be continuously extended to $\mathbb{R}$. A cover source $X$ corresponding to a precover source $Z$ is obtained by applying $Q_\triangle$ to each element of $Z$, $X = (X_1, \dots, X_n) = (Q_\triangle(Z_1), \dots, Q_\triangle(Z_n))$ with $X_k \sim p$, a probability mass function on $\mathcal{M}$:

$$p_j = \Pr(X_k = m_j) = \int\limits_{m_j-\triangle/2}^{m_j+\triangle/2} f(x)\mathrm{d}x. \qquad (1)$$

Since the specific details of the embedding (and extraction) algorithms are not important for our study, we only model the probabilistic impact of embedding. In particular, we narrow our study to the so-called mutually independent embedding that modifies every cover element $X_k$ independently to a corresponding element of the stego object $Y_k$ with probability

$$\Pr(Y_k = m_j | X_k = m_i) \triangleq b_{ij}(\beta) = \begin{cases} 1 + \beta c_{ii} & \text{if } i = j \\ \beta c_{ij} & \text{otherwise,} \end{cases} \qquad (2)$$

for some constants $c_{ij} \ge 0$ for $i \ne j$. Since $\sum_j b_{ij} = 1$, we must have $c_{ii} = -\sum_{j\ne i} c_{ij}$ for each $i$. The scalar parameter $\beta \ge 0$ will typically be the change rate. We restrict ourselves to embedding with changes from a limited range: $c_{ij} = 0$ whenever $|i - j| > L$ for some fixed positive integer $L$. This restriction is quite reasonable since most practical steganographic schemes modify cover elements to a few neighboring values, such as by ±1. Having said this, it is possible (and perhaps also meaningful) to study embedding operations whose range $L$ increases with $\triangle \to 0$.

Given the embedding operation (2), the stego object is an i.i.d. sequence of random variables $Y \triangleq (Y_1, \dots, Y_n)$ with $Y_k \sim q(\beta)$. For example, ignoring the boundary issues, for LSB matching (LSBM), LSB replacement (LSBR), the F5 embedding operation [9] (F5), and symmetrized Jsteg [7] (symJsteg) with an embedding change rate $\beta$, we respectively have:

$$q_j(\beta) = (1-\beta)p_j + \frac{1}{2}\beta(p_{j+1} + p_{j-1}) \quad \text{for all } j, \qquad (3)$$

$$q_j(\beta) = (1-\beta)p_j + \beta p_{j+(-1)^j} \quad \text{for all } j, \qquad (4)$$

$$q_j(\beta) = \begin{cases} (1-\beta)p_j + \beta p_{j+\mathrm{sign}(j)} & \text{for } j \ne 0, \\ p_0 + \beta(p_1 + p_{-1}) & \text{for } j = 0, \end{cases} \qquad (5)$$

$$q_j(\beta) = \begin{cases} (1-\beta)p_j + \beta p_{j+\mathrm{sign}(j)} & \text{for } j \text{ odd}, \\ (1-\beta)p_j + \beta p_{j-\mathrm{sign}(j)} & \text{for } j \text{ even}, \\ p_j & \text{for } j = 0. \end{cases} \qquad (6)$$

Assumption (2) leads to the following relationship between $p_j$ and $q_j(\beta)$:

$$q_j(\beta) = \Pr(Y_k = m_j) = \sum_i b_{ij} p_i = p_j + \beta \sum_i c_{ij} p_i. \qquad (7)$$

By expanding the KL divergence between $p$ and $q(\beta)$ at $\beta = 0$, the following standard result is obtained:

$$D_{\mathrm{KL}}(p||q) = \sum_j p_j \log \frac{p_j}{q_j(\beta)} = \frac{1}{2}\beta^2 I_\triangle(0) + O(\beta^3), \qquad (8)$$

where, using (7),

$$I_\triangle(0) = \sum_j \frac{1}{p_j} \left( \left. \frac{\mathrm{d}q_j(\beta)}{\mathrm{d}\beta} \right|_{\beta=0} \right)^2 = \sum_j \frac{1}{p_j} \left( \sum_i c_{ij} p_i \right)^2. \qquad (9)$$

is the steganographic Fisher information (FI), which encapsulates the effect of embedding. Note that in our framework $I_\triangle(0)$ depends on both $f$ and $\triangle$ through (1). It is precisely this relationship that is of interest in this paper.

We will illustrate the theoretical results on two continuous densities commonly used for modeling the distribution of digital media elements: the generalized Gaussian (GGD) and generalized Cauchy distribution (GCD):

$$f_{\mathrm{GG}}(x) = \frac{\alpha}{2b\Gamma(1/\alpha)} \exp\left( -\frac{|x-\mu|^\alpha}{b} \right), \qquad (10)$$

$$f_{\mathrm{GC}}(x) = \frac{\tau - 1}{2b} \left( 1 + \frac{|x-\mu|}{b} \right)^{-\tau}, \qquad (11)$$

with $b > 0$, $\alpha > 0$, $\tau > 1$. Note that besides the case when $\alpha \ge 2$, $\alpha \in \mathbb{Z}$, the GGD has a singularity at $x = \mu$ as its derivatives become unbounded there, starting with the $\lceil\alpha\rceil$th derivative. In contrast, all one-sided derivatives of the GCD are bounded but do not exist at $x = \mu$.

## 3. SCALING DUE TO QUANTIZATION

In this section, we analyze the effects of cover quantization on the FI. First, we state a general result for smooth

**Table 1**. Leading order $k^\star$ for common embedding operations. For F5, the numbers in brackets are at $j = 0$.

| Embedding | $k^\star$ | $w_{k^\star}$ |
|---|---|---|
| LSBM | 2 | 1 |
| LSBR | 1 | 1 |
| F5 | 1 (0) | 1 (2) |
| symJsteg | 1 | 1 |
| $c_{j-m,j} = c_{j+m,j}$ | 2 | $2\sum_{m=1}^{L} m^2 c_{j+m,j}$ |

precover densities and then for densities with singularities to cover the GGD as well as the GCD.

By introducing $F_\triangle(x) = \int_{x-\triangle/2}^{x+\triangle/2} f(t)\mathrm{d}t$, for $\triangle > 0$ and $x \in \mathbb{R}$, $p_j = F_\triangle(j\triangle)$, and (9) becomes:

$$I_\triangle(0) = \sum_j \frac{\left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2}{F_\triangle(j\triangle)}. \tag{12}$$

The sum in the numerator is a discrete filter with (a generally non stationary) kernel $c_{\cdot j}$ applied to $F_\triangle$ sampled at $j\triangle$. It is shown below that the scaling w.r.t. $\triangle$ depends on the leading exponent of $\sum_i c_{ij} F_\triangle(i\triangle) \propto \triangle^k$, which in turn depends on the smoothness of $f$ as well as the embedding operation.

### 3.1. $f$ differentiable

From the first mean value theorem for integration:

$$F_\triangle(j\triangle) = \triangle f(u_j) \tag{13}$$

for some $u_j \in (j\triangle - \frac{\triangle}{2}, j\triangle + \frac{\triangle}{2})$. For $i \neq j$, $|i - j| \leq L$, we expand $F_\triangle(i\triangle) = F_\triangle(j\triangle + (i - j)\triangle)$ at $j\triangle$ using Taylor series with Lagrange remainder. Assuming $F_\triangle(x)$ is $k \geq 0$ times continuously differentiable

$$F_\triangle(i\triangle) = \sum_{l=0}^{k-1} \frac{F_\triangle^{(l)}(j\triangle)}{l!} \triangle^l (i-j)^l + \frac{F_\triangle^{(k)}(\xi_{ij})}{k!} \triangle^k (i-j)^k, \tag{14}$$

where $\xi_{ij} \in (j\triangle, i\triangle)$ or $(i\triangle, j\triangle)$, depending on whether $i > j$ or $j > i$. Therefore,

$$\sum_i c_{ij} F_\triangle(i\triangle) = \sum_{l=0}^{k-1} \frac{F_\triangle^{(l)}(j\triangle)}{l!} \triangle^l w_{jl} \tag{15}$$

$$+ \frac{\triangle^k}{k!} \sum_i c_{ij} (i-j)^k F_\triangle^{(k)}(\xi_{ij}) \tag{16}$$

$$= \sum_{l=0}^{k-1} \frac{f^{(l)}(\phi_{jl})}{l!} \triangle^{l+1} w_{jl} \tag{17}$$

$$+ \frac{\triangle^{k+1}}{k!} \sum_i c_{ij} (i-j)^k f^{(k)}(\tilde{\phi}_{ij}). \tag{18}$$

where $\phi_{jl} \in (j\triangle - \frac{\triangle}{2}, j\triangle + \frac{\triangle}{2})$, $\tilde{\phi}_{ij} \in (\xi_{ij} - \frac{\triangle}{2}, \xi_{ij} + \frac{\triangle}{2})$ and the weights

$$w_{jl} = \sum_i c_{ij} (i-j)^l, \tag{19}$$

depend only on the embedding operation but not on the density $f$. Equation (18) follows from $F_\triangle^{(l)}(x) = f^{(l-1)}(x + \frac{\triangle}{2}) - f^{(l-1)}(x - \frac{\triangle}{2}) = \triangle f^{(l)}(\phi)$ for $1 \leq l \leq k+1$ and some $\phi \in (x - \frac{\triangle}{2}, x + \frac{\triangle}{2})$.

For LSB replacement and symmetrized Jsteg, $w_{j0} = \sum_i c_{ij} = 0$ for all $j$. The condition $w_{j0} = 0$ for all $j$ is equivalent with $b_{ij}$ being doubly stochastic (the sum of rows and columns is equal to 1). For LSB matching, $w_{j0} \neq 0$ only at the boundary of the dynamic range of cover elements. Since the boundary is usually sparsely populated for typical digital media distributions, the boundary bins cannot be used to make any statistically reliable conclusions and thus should not affect the asymptotic w.r.t. $\triangle$. This is why we made the dynamic range unbounded when defining $f$ in Section 2. Under these circumstances, the only embedding operation with non-zero $w_{j0}$ is F5 where $w_{00} = 2$ because $c_{-1,0} = c_{1,0} = 1$, $c_{00} = 0$ (see (5)). As will be seen below, bins $j$ for which $w_{j0} \neq 0$ have an effect on the scaling of the FI w.r.t. $\triangle$.

**Example 1.** Hypothetically, one could construct embedding operations with $w_{j0} \neq 0$ for all $j$, such as a scheme that does not embed in bins $3j$ and always changes $3j - 1$ and $3j + 1$ into $3j$.

Note that $w_{j1} = 0$ for embedding operations that are "symmetrical" in the sense that they modify each value of the cover by $\pm(i - j)$ with equal probabilities, e.g., for LSBM. For LSBR, symJsteg, and F5 (at $j \neq 0$), $|w_{j1}| = 1$. Finally, in general $w_{j2} > 0$ for all embedding operations except when the embedding does not embed in bin $j$ ($c_{ij} = 0$ whenever $i \neq j$). In particular, for all four embedding operations $w_{j2} = 1$ for all $j$ with the exception of F5, where $w_{02} = 2$, and symJsteg where $w_{0k} = 0$ for all $k \geq 0$.

**Definition 1.** We define the leading order $k^\star$ of the sum (15) at $x = j\triangle$ as the largest $k$ for which $w_{jl} = 0$ for all $l < k$. Furthermore, if $|w_{jk^\star}| = w_{k^\star}$ for all $j$, such that $j\triangle \in \mathcal{I} \subset \mathbb{R}$, we say that the embedding operation is bin-invariant on $\mathcal{I}$. If $w_{jl} = 0$ for all $l \geq 0$, we set $k^\star = \infty$ (as is the case of symJsteg for $j = 0$). Note that in general $k^\star \in \{0, 1, 2, \infty\}$.

Table 1 summarizes the leading order for the most common embedding operations used in steganography. The leading order for F5 is 1 for $j \neq 0$ and it is equal to 0 for $j = 0$. LSBR and LSBM are bin-invariant on $\mathbb{R}$ while F5 and symJsteg are bin-invariant on $\mathbb{R} - \{0\}$ as they both apply a different embedding rule at $j = 0$.

The leading order determines the scaling of the FI through the following theorem.

**Theorem 1.** Let the embedding operation be bin-invariant with leading order $k^\star$ everywhere and $f(x)$ $k^\star+1$ times continuously differentiable. Assuming $\exists M > 0$, such that 1) $|f^{(k)}(x)|$ is monotone decreasing for $x > M$ and $0 \le k \le k^\star$, 2) $\exists \delta_0 > 0$ such that $\int_M^\infty \frac{(f^{(k^\star)}(x))^2}{f(x+\delta)}\mathrm{d}x$ is convergent in Riemann sense $\forall \delta \in (0, \delta_0]$,

$$\lim_{\triangle\to 0^+} \frac{I_\triangle(0)}{\triangle^{2k^\star}} = \frac{w_{k^\star}^2}{k^\star!^2} \int\limits_{-\infty}^{\infty} \frac{(f^{(k^\star)}(x))^2}{f(x)}\mathrm{d}x. \qquad (20)$$

**Discussion:** The theorem assumptions are easily established for both the GGD and GCD for all $k \ge 0$, $k \in \mathbb{Z}$. Assumption 2 essentially requests that $f(x)$ does not fall to zero too quickly. An example of a p.d.f. that satisfies 2) for $\delta = 0$ but not for any $\delta > 0$ is the double exponential, $\exp(-\exp(x^2))$.

The theorem is proved by writing $I_\triangle(0)$ as a sum of three *partial sums*, $I_\triangle(0) = s_{-\infty,-M} + s_{-M,M} + s_{M,\infty}$, where

$$s_{m_1,m_2}(\triangle) = \sum_{m_1 \le j\triangle \le m_2} \frac{\left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2}{F_\triangle(j\triangle)}. \qquad (21)$$

After substituting (18) and (13) into the numerator and denominator, respectively, the infinite partial sums are shown to be $\triangle^{2k^\star} \times o(M)$ due to the restrictions at $\infty$ imposed on $f$.

The result can be easily extended to non-even densities by imposing assumptions 1) and 2) at both $\pm\infty$. It can also be generalized to embedding operations that are not bin-invariant. Quite often, special embedding rules are adopted at $j = 0$. For example, the F5 embedding operation has $w_{00} \ne 0$ in which case, $I_\triangle(0) \propto \triangle^1$ since at most $2L+1$ bins (a number which does not depend on $\triangle$) are affected. In general, modifying the embedding rule so that the leading order is $k' < k^\star$ for finitely many bins $j$ changes the scaling from $\triangle^{2k^\star}$ to $\triangle^{2k'+1}$. Finally, note that $I_\triangle(0) \propto \triangle^0$ for the operation from Example 1 since $k^\star = 0$ for all $j$.

**Table 2**. Scaling of the sum $s_{(L+2)\triangle,\epsilon}$ near the singularity for the GGD.

| $k^\star$ | $s_{(L+2)\triangle,\epsilon}$ | $\alpha$ |
|---|---|---|
| 1 | $\triangle^{1+2\alpha}$ | $\alpha \le 1/2$ |
| | $\triangle^2$ | $\alpha > 1/2$ |
| 2 | $\triangle^{1+2\alpha}$ | $\alpha \le 3/2$ |
| | $\triangle^4$ | $\alpha > 3/2$ |

**Table 3**. Scaling of Fisher information $I_\triangle(0)$ w.r.t. the quantization step $\triangle$ for four embedding operations for the GGD and GCD; $\alpha > 0$ is the shape parameter of GGD. The scaling is invariant to the shape parameter $\tau$ of the GCD.

| Embedding | GGD | GCD |
|---|---|---|
| LSBM | $\triangle^{\min\{4,1+2\alpha\}}$ | $\triangle^3$ |
| LSBR | $\triangle^{\min\{2,1+2\alpha\}}$ | $\triangle^2$ |
| F5 | $\triangle$ | $\triangle$ |
| symJsteg | $\triangle^{\min\{2,1+2\alpha\}}$ | $\triangle^2$ |

### 3.2. $f$ with singularity

We restrict our study to precover densities with a sole singularity at $x = 0$. Since the analysis of partial sums on unbounded intervals and on closed intervals not containing $x = 0$ is covered by Theorem 1, we divide the treatment to 1) *an immediate neighborhood of the singularity* (partial sums on $[(L+2)\triangle, \epsilon]$ for a fixed $\epsilon > 0$) and 2) *at singularity* (the remaining $2L+3$ terms $j$ for $|j| \le L+1$). If the singularity is such that $f$ is not differentiable at $x = 0$ but the one-sided derivatives exist and are bounded, the scaling of $s_{(L+2)\triangle,\epsilon}$ is again given by Theorem 1 (e.g., for the GCD). The case when the derivatives are unbounded at zero is covered by the following theorem.

**Theorem 2.** Assuming the embedding operation is bin-invariant on $\mathbb{R} - \{0\}$ with leading order $k^\star \ge 1$ and $f(x)$ has a singularity at $x = 0$ such that on some neighborhood of zero, $f^{(k^\star)}(x) = g(x)|x|^{-n}$, $n > 0$ for a continuous $g(x)$ with $g(0) \ne 0$, for all sufficiently small $\epsilon > 0$,

$$s_{(L+2)\triangle,\epsilon} = \begin{cases} \Theta(\triangle^{2k^\star+1-2n}) & \text{when } n \ge 1/2, \\ \Theta(\triangle^{2k^\star}) & \text{when } n < 1/2. \end{cases} \qquad (22)$$

**Discussion:** The proof starts with rewriting (18) using the fact that $\sum_i \gamma_i f(x_i) = f(\overline{x}) \sum_i \gamma_i$, $\overline{x} \in [\min_i x_i, \max_i x_i]$, for any $f$ continuous when all $\gamma_i$ are of the same sign. Then, thanks to the assumption on the $k^\star$th derivative, the sum over $j$ can be squeezed between two integrals that are shown to exhibit the same scaling.

Since for the GGD (10), $f'(x) \approx |x|^{\alpha-1}$ and $f''(x) \approx |x|^{\alpha-2}$ at $x = 0$, Theorem 2 gives the scaling of the partial sum summarized in Table (2).

Instead of providing a general result at the singularity, we show how the scaling of the remaining $2L+3$ terms of the partial sum for $|j| \le L+1$ can be obtained for a specific embedding algorithm. The assumptions of Theorem 2 imply the following form of the density

$$f(x) = a_0 - a_1|x|^\lambda + o(|x|^\lambda) \text{ at } x = 0, \qquad (23)$$

for some $\lambda > 0$, $a_0 > 0$, and $|a_1| > 0$. Thus, the scaling of the terms $\frac{1}{p_j}\left(\sum_i c_{ij} p_i\right)^2$ in (9), $|j| \le L+1$, can be obtained simply by determining $\lambda$ and evaluating the integrals (1). Carrying out these steps allows us to obtain in the next section a complete scaling result for the four embedding operations listed in Section (2) and two precover distributions.

## 4. GG AND GC PRECOVERS

**GGD:** By combining the scaling in the immediate neighborhood of the singularity (Table 2) with the fact that $f_{\mathrm{GG}}(x) \approx \exp(-|x|^\alpha/b) = 1-|x|^\alpha/b+o(|x|^\alpha) \implies \lambda = \alpha$ in (23), we obtain the result shown in the second column of Table 3 graphically rendered in Figure 1. The results are verified using simulations obtained by evaluating $p_j$ (1) in the sum (9) by numerical integration.

**GCD:** The derivative does not exist at zero but the one-sided ones do and are bounded. The scaling will thus be determined solely by the $2L+3$ terms at the singularity. The expansion of (11) at zero is

$$f_{\mathrm{GC}}(x) = \frac{\tau-1}{b}\left(1-\frac{\tau}{b}|x|+\frac{\tau(\tau+1)}{b^2}a^2x^2+\cdots\right), \tag{24}$$

which means that $\lambda = 1$ independently of the shape parameter $\tau$. Since for LSBM all bins with the exception of $-1, 0, 1$ scale as $\triangle^4$ ($k^\star = 2$, see Section 3), the resulting scaling for the GCD is $\triangle^{1+2\lambda} = \triangle^3$. For LSBR and symJsteg, the scaling at smooth points is $\triangle^2$ (since $k^\star = 1$) while the scaling at bin 0 is $\triangle^3$, giving the final scaling $\triangle^2$ for both algorithms for all $\tau > 0$. Finally, for the F5 operation, the scaling is determined by the zero bin, which scales as $\triangle^1$ for all $\tau$. The scaling for the GCD is summarized in the third column of Table 3.

## 5. SCALING IN PRACTICE

The original SRL is quite robust in the sense that even though it has been established only for artificial cover sources, such as Markov chains, it manifests quite robustly for empirical cover sources despite the fact that practitioners build detectors using machine learning rather than as likelihood ratio tests and that pixels (transform coefficients) are quite heterogeneous and non stationary.

The scaling of the FI w.r.t. the quantization step undoubtedly manifests in practice as well. However, the specific scaling strongly depends on the precover distribution. For example, while both the GGD and GCD are commonly used for modeling transform coefficients, both models lead to quite different scaling (see Table 3). Even though one could conceivably conduct an experiment with a database of images compressed with a range of JPEG quality factors and perform steganalysis using classifiers implemented using machine learning in some feature space, relating the results of these experiments with the theoretical results of this paper derived for i.i.d. sources is likely not possible. This is mainly because DCT coefficients form a rather complicated mixture – each DCT mode has a different distribution that additionally depends on each image.

There is an additional trouble posed by the limited number of samples typically available in practice. When sampling the cover distribution, one must avoid under- as well as over-sampling. Oversampling (too small $\triangle$) will produce a noisy estimate of the cover distribution, $p_j$, which will completely change the scaling. On the contrary, $\triangle$ that is too large will not yet exhibit the limiting behavior when $\triangle \to 0$.

Next, we provide a crude qualitative analysis of the impact of over-sampling. Given a cover object with $N$ elements from a finite range $\mathcal{R}$, the normalized population of the $j$th bin, $P_j$, is a random variable whose binomial distribution will be approximated with a Gaussian $P_j \sim \mathrm{N}(p_j, p_j(1-p_j)/N)$. If $P_j$ were independent, the expected value of $\left(\sum_i c_{ij}P_i\right)^2 = \left(\sum_i c_{ij}p_i + \sum_i c_{ij}(P_i-p_i)\right)^2$ would be $\mu_j^2 = \left(\sum_i c_{ij}p_i\right)^2 + \sum_i c_{ij}^2 p_i(1-p_i)/N$. Since in practice, there will be finitely many bins, $j\triangle \in \mathcal{R} \subset [-R, R]$, where $R$ is the dynamic range of $X$,

$$E\left[\sum_{j\in\mathcal{R}} \frac{\left(\sum_i c_{ij}P_i\right)^2}{P_j}\right] \doteq \sum_{j\in\mathcal{R}} \frac{\mu_j^2}{p_j} = I_\triangle(0) \tag{25}$$

$$+\sum_{j\in\mathcal{R}} \frac{\sum_i c_{ij}^2 p_i(1-p_i)}{Np_j} \propto \frac{1}{N\triangle}. \tag{26}$$

because $I(0) \propto \triangle^s$ with $s \geq 0$ and $|\mathcal{R}|$ is inversely proportional to the quantization step $\triangle$.

To assess how the scaling might manifest in practice, we generated $N = 10^4 - 10^8$ i.i.d. samples from a GGD with $\alpha = 1.5$ and $b = 1$ and computed $I_\triangle(0)$ for a range of $\triangle$. The result shown in Figure 2 confirms the crude analysis of over-sampling with the scaling exponent $s \doteq -1$ for small $\triangle$. The scaling matches the theoretical result, $s \doteq 1+2\alpha = 4$, for midrange values and breaks up when $\triangle$ becomes too large. The range of proper scaling gets smaller with decreasing $N$.

## 6. CONCLUSION

The square root law of imperfect steganography connects statistical detectability with the cover size and the change rate in the asymptotic limit of large covers and small change rates. The current paper extends this law to the case when the cover object is obtained by quantizing a precover that follows a continuous-valued distribution. In particular, constant statistical detectability is obtained when $N\beta^2\triangle^s = \mathit{const.}$, where $N$ is the cover

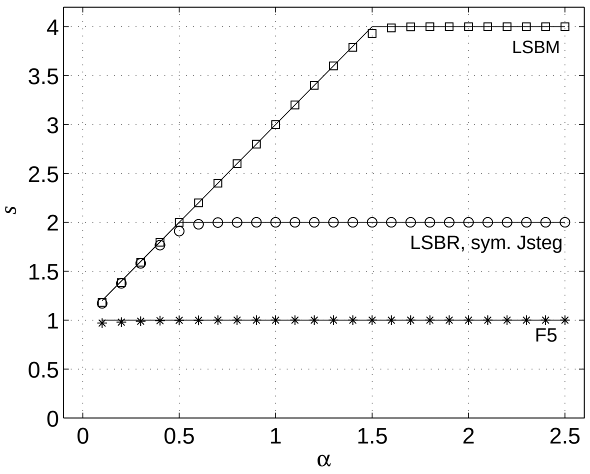

**Fig. 1**. Scaling exponent $s$ versus the parameter $\alpha$ for the generalized Gaussian precover model and four embedding operations. Solid lines show the theoretical result; the markers are from numerical simulations.

size, $\beta$ the change rate, and $s$ the scaling exponent w.r.t. the quantization step $\triangle$ that tends to zero. The scaling exponent is determined jointly by the embedding operation and the smoothness of the precover distribution. In general, $s$ is larger for smoother distributions and for embedding operations that act as low-pass filters of the first-order statistic of cover samples.

This work reveals a connection between statistical detectability and the complex interplay between the precover distribution and the embedding operation. This is relevant for understanding the effect of color bit depth and/or the JPEG quality factor on security. However, unlike the original square root law, which is quite robust, the scaling w.r.t. $\triangle$ strongly depends on the precover distribution, which prevents theoretical quantification of scaling in empirical sources analyzed with empirical detectors.

## 7. REFERENCES

[1] T. Filler and J. Fridrich. Fisher information determines capacity of $\epsilon$-secure steganography. In *Information Hiding, 11th International Workshop*, volume 5806 of *Springer LNCS*, pages 31–47, 2009.

[2] T. Filler, A. D. Ker, and J. Fridrich. The square root law of steganographic capacity for Markov covers. In *Proc. SPIE, Electronic Imaging, Security and Forensics of Multimedia XI*, volume 7254, pages 08 1–08 11, 2009.

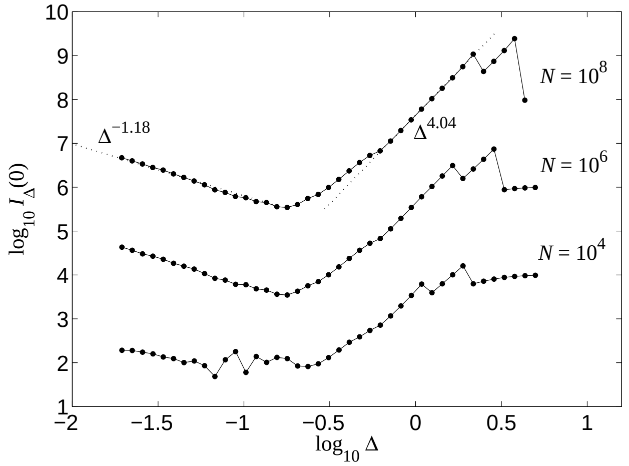

**Fig. 2**. Fisher information vs. $\triangle$ for $N = 10^8, 10^6, 10^4$ samples from a GGD with $\alpha = 1.5$, $b = 1$, $\mu = 0$. Note the region of over-sampling for small $\triangle$ where $I_\triangle(0) \propto \triangle^{-1}$ and under-sampling for large $\triangle$. The range of $\triangle$ where $s \doteq \max\{4, 1 + 2\alpha\} = 4$, as predicted by the theory, becomes smaller with decreasing $N$. Dotted lines are linear fits in their corresponding ranges of $\triangle$.

[3] A. D. Ker. A capacity result for batch steganography. *IEEE Signal Processing Letters*, 14(8):525–528, 2007.

[4] A. D. Ker. Estimating steganographic Fisher information in real images. In *Information Hiding, 11th International Workshop*, volume 5806 of *Springer LNCS*, pages 73–88, 2009.

[5] A. D. Ker. The square root law in stegosystems with imperfect information. In *Information Hiding, 12th International Conference*, volume 6387 of *Springer LNCS*, pages 145–160, 2010.

[6] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich. The square root law of steganographic capacity. In *Proc. of the 10th ACM MM&Sec Workshop*, pages 107–116, 2008.

[7] J. Kodovský and J. Fridrich. Quantitative structural steganalysis of Jsteg. *IEEE TIFS*, 5(4):681–693, 2010.

[8] Y. Wang and P. Moulin. Perfectly secure steganography: Capacity, error exponents, and code constructions. *IEEE Transactions on Information Theory, Special Issue on Security*, 55(6):2706–2722, 2008.

[9] A. Westfeld. High capacity despite better steganalysis (F5 – a steganographic algorithm). In *Information Hiding, 4th International Workshop*, volume 2137 of *Springer LNCS*, pages 289–302, 2001.