Estimating the Information Theoretic Optimal Stego Noise

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England
adk@comlab.ox.ac.uk

Abstract. We recently developed a new benchmark for steganography, underpinned by the *square root law of capacity*, called *Steganographic Fisher Information* (SFI). It is related to the multiplicative constant for the square root capacity rate and represents a truly information theoretic measure of asymptotic evidence. Given a very large corpus of covers from which the joint histograms can be estimated, an estimator for SFI was derived in [1], and certain aspects of embedding and detection were compared using this benchmark.

In this paper we concentrate on the evidence presented by various spatial-domain embedding operations. We extend the technology of [1] in two ways, to convex combinations of arbitrary so-called *independent embedding functions*. We then apply the new techniques to estimate, in genuine sets of cover images, the spatial-domain stego noise shape which optimally trades evidence – in terms of asymptotic KL divergence – for capacity. The results suggest that smallest embedding changes are optimal for cover images not exhibiting much noise, and also for cover images with significant saturation, but in noisy images it is superior to embed with more stego noise in fewer locations.

1 Introduction

A particular challenge, for the design of better steganographic embedding algorithms, is the lack of universal benchmarks. When a new method is proposed, just about the best that can be done is to test it against leading steganalysis algorithms, and if their detection accuracy is diminished then the steganography method is considered an advance. In practice, new embedding methods usually turn out to be easily broken by a modified detector. The root of the problem is that the metric was really one for novelty, not security.

Information theoretic models of stego systems [2] provide the foundation for an alternative: the Kullback-Leibler (KL) divergence between cover and stego distributions can bound secure embedding capacity, but such distributions are arguably incognisable [3] and certainly infeasible to estimate in full. However, in [1] we argued that the *asymptotic* KL divergence is sufficient, and that this is determined by so-called *Steganographic Fisher Information* (SFI). Furthermore, SFI can indeed be estimated for small groups of pixels, and it was argued that this is highly relevant for practical steganalysis which almost inevitably takes

A.T.S. Ho et al. (Eds.): IWDW 2009, LNCS 5703, pp. 184–198, 2009.

evidence from small groups. Some experimental results, in [1], used the estimator to compare a few simple embedding functions' security, but mainly focused on lessons for steganalysis.

This is a sequel to [1], using SFI to evaluate spatial-domain embedding functions. We extend the SFI estimator to remove some of the limitations in [1] and to *convex combinations* of different embedding functions. Then we apply the new estimator to find the *optimal* combination of certain simple spatial-domain embedding functions. We may have confidence in the true optimality of these combinations because the metric has well-founded information theoretic roots. Our results are not surprising – in noisy covers it is better to embed with larger stego noise – but allow, for the first time, calculation of an optimized embedding function for real-world cover sources.

This paper contains: (Sect. 2) a brief recapitulation of the argument and results of [1], and an explanation of why slightly different notation must be adopted for the present work; (Sect. 3) an extension of the estimator of [1], both to arbitrary embedding functions and to convex combinations thereof; (Sect. 4) some experiments using the SFI estimate to choose optimal combinations of embedding functions, thereby deriving optimally-shaped stego noise for simple variable-base (mod k)-matching embedding; (Sect. 5) a conclusion.

Some notational conventions: random variables and distributions will be denoted by upper-case letters, and realizations of random variables the corresponding lower case. Vectors of either random variables or realizations will be boldface $\boldsymbol{x} = (x_1, \ldots x_n)$, with n implicit. All logs will be to natural base.

2 Steganographic Fisher Information

We model stego objects as random variables with distribution $P(\lambda)$, where λ indicates the payload size (how the size is measured is important and we will return to this momentarily), so that $P(0)$ is the distribution of covers. KL divergence cannot be increased by processing, and thus we reach the well-known limit on the accuracy of *any* detector for the presence of steganography, in terms of $D_{\mathrm{KL}}(P(0) \,\|\, P(\lambda))$ [2]. This justifies using KL divergence as a measure of *evidence*. In [4] it is argued that we should focus on *asymptotic* capacity, as relative payload size tends to zero, because repeated communication must reduce the embedding rate or face eventual certain detection. So in order to make an asymptotic judgement about secure capacity it is sufficient to consider the asymptotic behaviour of $D_{\mathrm{KL}}(P(0) \,\|\, P(\lambda))$ as $\lambda \to 0$, and usually (see [5]), this is locally quadratic in λ, i.e.

$$D_{\mathrm{KL}}(P(0) \,\|\, P(\lambda)) \sim \tfrac{1}{2} I \lambda^2 + O(\lambda^3).$$

I is called, in this setting, *Steganographic Fisher's Information*. Unlike most other benchmarks, SFI is a single figure which can be used to compare the asymptotic performance of embedding methods or, by considering SFI of projections of the stego object space, the evidence available to various feature sets. It also seems to be easier to estimate SFI than KL divergence directly.

We suppose that stego objects are made up of n *locations* – pixel values, transform coefficients, or suchlike – and that embedding alters some locations. How λ measures payload size is critical to the interpretation of SFI. If we define λ to be the relative number of embedding changes – the proportion of cover locations changed by embedding – then we call it *SFI with respect to change rate* and write I_c. But this does not correctly take account of the cover size n, nor does it correctly compare embedding methods with different *embedding efficiency* – usually defined as the average number of covert payload bits conveyed per embedding change [6], and denoted e – so [1] defined *SFI with respect to payload rate*

$$I_p = \frac{I_c}{ne^2}.$$

It was I_c which was directly estimated in [1], and converted to I_p as above for proper comparison of embedding and detection methods. I_p has the following interpretation, to connect it with the square root law of steganographic capacity [7,8]: if one embeds a small m bit payload into a cover with n locations, using an embedding method with SFI I_p, one expects to produce a KL divergence of approximately $I_p(m^2/n)$ nats of evidence.

Here, we find it more convenient to use a different parameterization for λ. Let us measure payload as the relative number of payload locations used for embedding, whether changed or not. In the case of embedding one bit per symbol, this is exactly the relative payload size, but if embedding k-ary symbols in m locations the total payload transmitted is $m \log_2 k$ bits. This measure is convenient because different embedding methods have different probabilities of changing a location, which is otherwise an algebraic nuisance. We call the SFI thus derived *SFI with respect to location rate* and denote it I_l; it is I_l which will be estimated in Sect. 3. Then I_p can be recovered as

$$I_p = \frac{I_l}{ne'^2}, \tag{1}$$

where e' denotes the number of covert bits transmitted per location used. We will later consider (mod k)-matching embedding, for which $e' = \log_2 k$.

2.1 Estimating SFI

The dimensionality of the space of digital images is outrageously large, so it is not possible to estimate true SFI for entire images. In [1] we advocated the following lower-dimensional model: imagine that an image is made up of many independent pixel *groups*, where the groups are of fixed size such as 1×2 pixels, 2×2, 3×3, etc. Thus we reduce each image to its histogram of groups: in the case of 1×1 groups this is the standard histogram, in the case of 1×2 groups it is the *co-occurrence matrix*, and so on. We argued that, although this certainly destroys information, it is a fact that most leading steganalysis methods do exactly the same: they base their decision on information extracted from histograms, adjacency histograms, or (in the case of JPEG images) 8×8

blocks. (This is no surprise because models of digital media are usually local.) So, if we do likewise, computing SFI for small pixel groups gives us asymptotic bounds on the performance of these steganalysis methods. Indeed, the main focus of [1] was the comparison of evidence in different pixel groups.

Having reduced an image to independently-considered groups of pixels, we obtained the SFI as a function of the group frequencies and embedding function, via a Taylor expansion of KL divergence in change rate, but only for a particular type of embedding operation which changes cover samples to one of a fixed number of alternatives, equiprobably: this is suitable for LSB embedding, but not for more complex examples such as the convex combinations we explore later in this paper. An estimator for SFI was obtained by plugging the empirical group histogram, obtained from a corpus of genuine covers, into the SFI formula. This estimator has limitations – we need a very large corpus from which to estimate the histogram, particularly for larger groups of pixels where the histogram itself has very many bins – but does converge in probability to the true value as the corpus size tends to infinity.

We performed some experiments, mostly with just one corpus of covers, to compare the SFI found in different types of pixel groups in grayscale images. Some brief experiments compared the relative security of LSB replacement and 2LSB replacement (where each pixel carries two bits of payload, at the cost of higher embedding noise), motivated by an observation in [9] that 2LSB embedding was, on a per-payload basis, slightly less sensitively detected by structural detectors. Our experiments in a set of very well-regulated cover images contradicted this hypothesis, but brief experiments on noisier image sets were consistent with it. This raises the questions addressed in this paper: given the options of embedding more payload per change with greater stego noise, or less payload with lower stego noise, which is better? And what of intermediate options?

3 Extending the SFI Estimator to Arbitrary Embedding

With weaker assumptions, but using similar techniques as in [1], we will compute I_l by expanding the KL divergence in location rate. Our model is that the cover is made up of a fixed-length sequence of symbols $(X_1, \ldots, X_n)$, each drawn from finite alphabet $\mathcal{X}$ (with arbitrary distribution: the components X_i need not be independent). The corresponding stego object is denoted $(Y_1, \ldots, Y_n)$. We are concerned with *independent embedding*, where the embedding function chooses whether to locate a payload in each cover symbol independently with probability λ, and if location X_i is chosen then it is altered randomly according to a matrix $B = (b_{ij})$, so that $P(Y_i{=}y \,|\, X_i{=}x) = b_{xy}$ in the chosen locations (otherwise $Y_i = X_i$). For this to be well-defined, B must be stochastic: $\sum_j b_{ij} = 1$ for all i. Most non-adaptive steganography methods are accurately described by this model, including bit replacement, (mod k)-matching, and additive noise.

We also assume that the distribution of cover sequences $P(\boldsymbol{X}{=}\boldsymbol{x})$ is such that $P(\boldsymbol{X}{=}\boldsymbol{x}) = 0 \iff P(\boldsymbol{Y}{=}\boldsymbol{x}) = 0$. This ensures that the KL divergence between cover and stego sequences is finite. And we assume that the embedding is not

perfect: for at least some $\boldsymbol{x}$, $P(\boldsymbol{X}{=}\boldsymbol{x}) \neq P(\boldsymbol{Y}{=}\boldsymbol{x})$, otherwise SFI is zero and the square root law of capacity does not apply.

We begin with

$$P(Y{=}y \,|\, X{=}x) = (1-\lambda)\delta_{xy} + \lambda b_{xy}$$

from which we derive

$$\begin{aligned} P(\boldsymbol{Y}=\boldsymbol{y}) &= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P(\boldsymbol{Y}{=}\boldsymbol{y} \,|\, \boldsymbol{X}{=}\boldsymbol{x}) P(\boldsymbol{X}{=}\boldsymbol{x}) \\ &= (1-\lambda)^n P(\boldsymbol{X}{=}\boldsymbol{y}) + \lambda(1{-}\lambda)^{n-1} A(\boldsymbol{y}) + \lambda^2(1{-}\lambda)^{n-2} B(\boldsymbol{y}) + O(\lambda^3) \\ &= P(\boldsymbol{X}{=}\boldsymbol{y}) + \lambda\big[-nP(\boldsymbol{X}{=}\boldsymbol{y}) + A(\boldsymbol{y})\big] \\ &\quad + \lambda^2\big[\tfrac{n(n-1)}{2} P(\boldsymbol{X}{=}\boldsymbol{y}) - (n{-}1)A(\boldsymbol{y}) + B(\boldsymbol{y})\big] + O(\lambda^3) \end{aligned}$$

where

$$A(\boldsymbol{y}) = \sum_{i=1}^{n} \sum_{u\in\mathcal{X}} P(\boldsymbol{X}{=}\boldsymbol{y}[u/y_i]) b_{uy_i}, \tag{2}$$

$$B(\boldsymbol{y}) = \sum_{\substack{i,j=1 \\ i<j}}^{n} \sum_{u,v\in\mathcal{X}} P(\boldsymbol{X}{=}\boldsymbol{y}[u/y_i, v/y_j]) b_{uy_i} b_{vy_j},$$

and $\boldsymbol{y}[u/y_i]$ denotes the sequence $(y_1, \ldots, y_{i-1}, u, y_{i+1}, \ldots, y_n)$, $\boldsymbol{y}[u/y_i, v/y_j]$ analogously. $A(\boldsymbol{y})$, respectively $B(\boldsymbol{y})$, represents the probability of observing $\boldsymbol{y}$ in a stego object given exactly one, respectively two, locations used. Now, using $\log(1+z) = z - \frac{z^2}{2} + O(z^3)$, we can expand the KL divergence:

$$\begin{aligned} D_{\mathrm{KL}}(\boldsymbol{X} \,\|\, \boldsymbol{Y}) &= -\sum_{\boldsymbol{y}\in\mathcal{X}^n} P(\boldsymbol{X}{=}\boldsymbol{y}) \log\Big(\tfrac{P(\boldsymbol{Y}=\boldsymbol{y})}{P(\boldsymbol{X}=\boldsymbol{y})}\Big) \\ &= \lambda\Big[n\textstyle\sum P(\boldsymbol{X}{=}\boldsymbol{y}) - \sum A(\boldsymbol{y})\Big] \\ &\quad +\lambda^2\Big[\tfrac{n}{2}\textstyle\sum P(\boldsymbol{X}{=}\boldsymbol{y}) - \sum A(\boldsymbol{y}) - \sum B(\boldsymbol{y}) + \tfrac{1}{2}\sum \tfrac{A(\boldsymbol{y})^2}{P(\boldsymbol{X}=\boldsymbol{y})}\Big] + O(\lambda^3) \\ &= \tfrac{\lambda^2}{2}\Big[\textstyle\sum \tfrac{A(\boldsymbol{y})^2}{P(\boldsymbol{X}=\boldsymbol{y})} - n^2\Big] + O(\lambda^3). \end{aligned} \tag{3}$$

For the final step, we use $\sum_{\boldsymbol{y}} P(\boldsymbol{X}{=}\boldsymbol{y}) = 1$, and

$$\begin{aligned} \sum_{\boldsymbol{y}} A(\boldsymbol{y}) &= \sum_{\boldsymbol{y}} \sum_{i=1}^{n} \sum_{u} P(\boldsymbol{X}{=}\boldsymbol{y}[u/y_i]) b_{uy_i} \\ &= \sum_{i=1}^{n} \sum_{u} \sum_{\substack{\boldsymbol{y} \\ \text{except } y_i}} P(\boldsymbol{X}{=}\boldsymbol{y}[u/y_i]) \sum_{y_i} b_{uy_i} \\ &= \sum_{i=1}^{n} \sum_{\boldsymbol{y}} P(\boldsymbol{X}{=}\boldsymbol{y}) = n, \end{aligned}$$

similarly $\sum_{\boldsymbol{y}} B(\boldsymbol{y}) = \frac{n(n-1)}{2}$.

Thus, once we know the embedding efficiency per location for our embedding function e', we combine (3) and (1) to compute the SFI with respect to payload rate

$$I_p = \frac{\sum_{\boldsymbol{y}\in\mathcal{X}^n} \frac{A(\boldsymbol{y})^2}{P(\boldsymbol{X}=\boldsymbol{y})} - n^2}{n{e'_i}^2} \tag{4}$$

as a function of the true frequencies of each symbol group in $\mathcal{X}^n$ (note that $A(\boldsymbol{y})$ is a linear combination of such frequencies). As in [1], we can estimate this quantity from a large corpus of cover objects, simply by plugging the empirical frequencies into (4). We must omit any terms where $P(\boldsymbol{X}=\boldsymbol{y}) = 0$, i.e. groups of n which never occur in the corpus, but if sufficiently large then this should happen never or rarely, and the missing terms should be negligible.

Considering digital images, for $n \geq 4$ it can be challenging even to compute the empirical histogram, because there are potentially 256^n histogram bins and our image corpus will consist of at least 10^{10} groups of pixels, so computer memory is soon exhausted. We solved this problem in [1], using red-black trees to create overlapping histogram chunks, shuffle-sorting the chunks, making a second pass through the histogram to adjoin the value of $A(\boldsymbol{x})$ to each $P(\boldsymbol{X}=\boldsymbol{x})$, and finally summing the ratio $A(\boldsymbol{x})^2/P(\boldsymbol{X}=\boldsymbol{x})$. We will not go into the detail here because the same techniques can be used, although the second stage is somewhat slower because $A(\boldsymbol{x})$ depends on potentially the entire histogram, but for the experiments we report here on (mod k)-matching it is still the case that only portions local to $\boldsymbol{x}$ need be examined. With our available computing resources (a cluster of 20 dual-core machines) it is feasible to estimate I_p for pixel groups of size up to about 9, but our image libraries are only large enough adequately to sample the histograms for $n \leq 6$.

3.1 Convex Combinations of Embedding Functions

As well as extending the estimator to arbitrary independent embedding, we will consider the combination of embedding functions. Suppose that the steganographer and recipient share k different embedding options each of which matches the hypotheses of the previous section. Let us denote the change probabilities for embedding method i by the matrix B_i, and the embedding efficiency per location as e'_i. They can construct a hybrid embedding method which, on a per-symbol basis, picks embedding method i with fixed probability π_i such that $\sum_i \pi_i = 1$ (the correspondence between symbols and embedding functions can be generated from their shared secret key). This convex combination has overall embedding efficiency per location $\sum \pi_i e'_i$ and its change matrix is $B = \sum_i \pi_i B_i$, and this allows us to vary continuously between the different options. In particular, we can vary the tradeoff between higher stego noise and higher embedding rates. Here, we examine the SFI of such a combination, and later will demonstrate that combinations can indeed provide better transmission rates, at comparable levels of risk, than any of the individual options alone.

Recall that SFI is defined in terms of $A(\boldsymbol{y})$, the probability of observing $\boldsymbol{y}$ in a stego group with exactly one embedding location used. Observe in (2) that $A(\boldsymbol{y})$

is a linear function of B. Therefore if embedding method i has corresponding function $A_i(\boldsymbol{y})$, for the convex combination we have $A(\boldsymbol{y}) = \sum_i \pi_i A_i(\boldsymbol{y})$.

Therefore, the SFI with respect to payload rate is given by

$$I_p = \frac{\displaystyle\sum_{\boldsymbol{y}\in\mathcal{X}^n} \frac{(\sum_i \pi_i A_i(\boldsymbol{y}))^2}{P(\boldsymbol{X}=\boldsymbol{y})} - n^2}{n(\sum_i \pi_i e_i')^2} = \frac{\sum_{i,j} c_{ij}\pi_i\pi_j - n^2}{n(\sum_i \pi_i e_i')^2} \tag{5}$$

where

$$c_{ij} = \sum_{\boldsymbol{y}\in\mathcal{X}^n} \frac{A_i(\boldsymbol{y})A_j(\boldsymbol{y})}{P(\boldsymbol{X}=\boldsymbol{y})}. \tag{6}$$

The optimal convex combination is the probability vector π which minimizes (5): lower SFI means lower KL divergence – less accurate detection – or alternatively a greater secure capacity for equivalent risk. SFI is inversely proportional to the square of the "root rate", the asymptotic constant in secure capacity $r\sqrt{n}$ where n denotes cover size [7,8].

Equation (5) is a ratio between two quadratic forms and there does not seem to be an easy analytic form for the minimum, but the optimization can be performed very efficiently by numerical methods because all c_{ij} and e_i' must be positive, and I_l must also positive, so both $\sum_{i,j} c_{ij}\pi_i\pi_j$ and $(\sum_i \pi_i e_i')^{-1}$ are positive and convex in $\boldsymbol{\pi}$. So (5) can be written as the product of positive convex functions, and therefore is a convex function. Thus, given c_{ij} and e_i', numerical optimization of (5), subject to $\sum \pi_i = 1$, can be performed using standard convex programming methods.

4 Results

We now apply the extended estimator to find the optimal convex combinations of some simple embedding functions. Of course, the results depend on the cover source: there is no universally-optimal embedding function, and we expect different results for different sources. We will restrict our attention to spatial-domain **(mod k)-matching** embedding in grayscale digital images: each selected pixel conveys one k-ary symbol ($\log_2 k$ bits) of information in its remainder (mod k), and the embedding function alters the cover pixel to the nearest value with the correct remainder. We consider only odd $k = 2j + 1$, so that the embedding is symmetric. Most of the time, this results in additive noise uniformly distributed from the range $-j, \ldots, j$, but for pixels near to saturation at 0 or 255 the absolute value of the noise could reach $2j$. (Although it was LSB and 2LSB embedding which was briefly considered in [1], here we have excluded bit replacement and (mod $2k$)-matching embedding because it has been demonstrated, time and again, that asymmetrical embedding causes additional weaknesses [9,10].)

The case of (mod 3)-matching is also sometimes known as ± 1 embedding, and $k = 5, 7$ can be called $\pm 2, \pm 3$ embedding, respectively. However, we eschew

this terminology for two reasons. First, ± 1 more accurately describes the effect of LSB *replacement* while (mod 3)-matching can cause stego noise of ± 2 when applied to saturated pixels. Second, there is some confusion in the literature as to exactly what shape stego noise ± 2 denotes: uniform distortions of ± 2, or including ± 1 noise as well, or some other shape? Our preferred terminology is *ternary embedding* for (mod 3)-matching, *quinary embedding* for (mod 5)-matching, *septenary embedding* for (mod 7)-matching, and so on.

Our experiments will involve four sets of cover images, chosen for different levels of noise, to test the hypothesis that greater stego noise is optimal for noisy covers.

Set A: 2121 grayscale images taken with a single digital camera, all sized approximately 4.7 Mpixels. The histograms were computed from the images in each of four orientations, to boost the evidential base, so that a total of just over $4 \cdot 10^{10}$ pixel groups were used to estimate the joint histograms. The images had never been subject to JPEG compression, but as part of their conversion from RAW format were substantially denoised; also, images with significant areas of saturation were removed. This set of images is extremely well-behaved and it is the main corpus used for the results in [1].

Set B: 1040 grayscale images taken with a mixture of digital cameras, all sized approximately 1.5 Mpixels, for a total of over $6 \cdot 10^9$ pixel groups. Again, the images were never JPEG compressed and had been denoised in conversion from RAW format, but the denoising is not as aggressive as in set A.

Set C: 3200 grayscale images taken with the same camera as set A (in fact, the parent RAW files for set A are a subset of these), for a total of about $1.5 \cdot 10^{10}$ pixel groups. In conversion from RAW format, all optional denoising was disabled, so these images are visibly more noisy than those of sets A or B. Note that, unlike in set A, images with saturated areas (typically over-exposed highlights) have not been excluded.

Set D: 10000 grayscale decompressed JPEG images from a photo library CD, all sized about 900×600. Like set A, the images were re-used in each of four orientations, for a total of about $2 \cdot 10^{10}$ pixel groups. These images are certainly noisy, but feature quantization noise rather than sensor noise.

In each case we use the technology of [1] to estimate the histograms of individual pixels, and pixel groups of shapes 1×2, 1×3, 1×4, 2×2 (we will not follow [1] to even larger group sizes, to be sure that the histograms are not undersampled). For ternary, quinary, and (sometimes) septenary embedding, the coefficients c_{ij} (6) are computed and (5) minimized numerically to find the optimal combination.

4.1 Combination of Ternary and Quinary Embedding

We begin by considering combinations of ternary and quinary embedding: this allows stego noise up to level ± 2 (except at saturated cover locations). The embedding matrices are

$$B_1 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \ddots & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$$B_2 = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & \ddots & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

and we consider the mixture $B = \pi_1 B_1 + \pi_2 B_2$, where $\pi_1 + \pi_2 = 1$, i.e. the embedding will use proportion π_1 ternary symbols and π_2 quinary symbols. (The embedder and recipient may need to transcode the payload via a variable-base format, but we will not concern ourselves with the technicalities of doing so.)

Figure 1 shows the results of SFI estimation, considering both 1×2 and 1×4 blocks (most other shapes were similar; we shall see shortly that 1×1 groups are anomalous). For each image set we first plot the SFI with respect to location rate I_l, as a function of π_2: for most image sets, the graphs are rising, indicating that the greater the proportion of quinary symbols, the greater the evidence of payload. This is no surprise, because quinary embedding causes greater stego noise. (However, the slight decrease for small values of π_2 in set B, and the significant U-shape in set D, means that it can be less suspicious to embed with more noise, a paradox which probably deserves further study.)

The second column in Fig. 1 shows how the embedding efficiency e', measured in bits per payload location, increases as we increase the proportion of quinary symbols (which carry more bits each). This linear function is the same for all image sets, of course. The final column computes the trade-off between these functions, showing SFI per payload: for a given (small) payload, this indicates how much evidence is available to the opponent for each combination of ternary and quinary embedding. For cover image set A, which contains little noise, pure ternary embedding is best. For set B, which is more noisy, the optimum is about 1/3 ternary and 2/3 quinary embedding: the exact minimum does depend on whether we look at 1×2 or 1×4 pixel groups. For set C, which is very noisy, we might have expected an even higher proportion of quinary embedding, but in fact observe the opposite: we attribute this to the saturated areas in some of the images, because the embedding noise is exaggerated for saturated pixels. For set D, which has quantization noise, the optimum is almost pure quinary embedding for 1×2 pixel groups, but nearer to an even mixture for 1×4 groups.

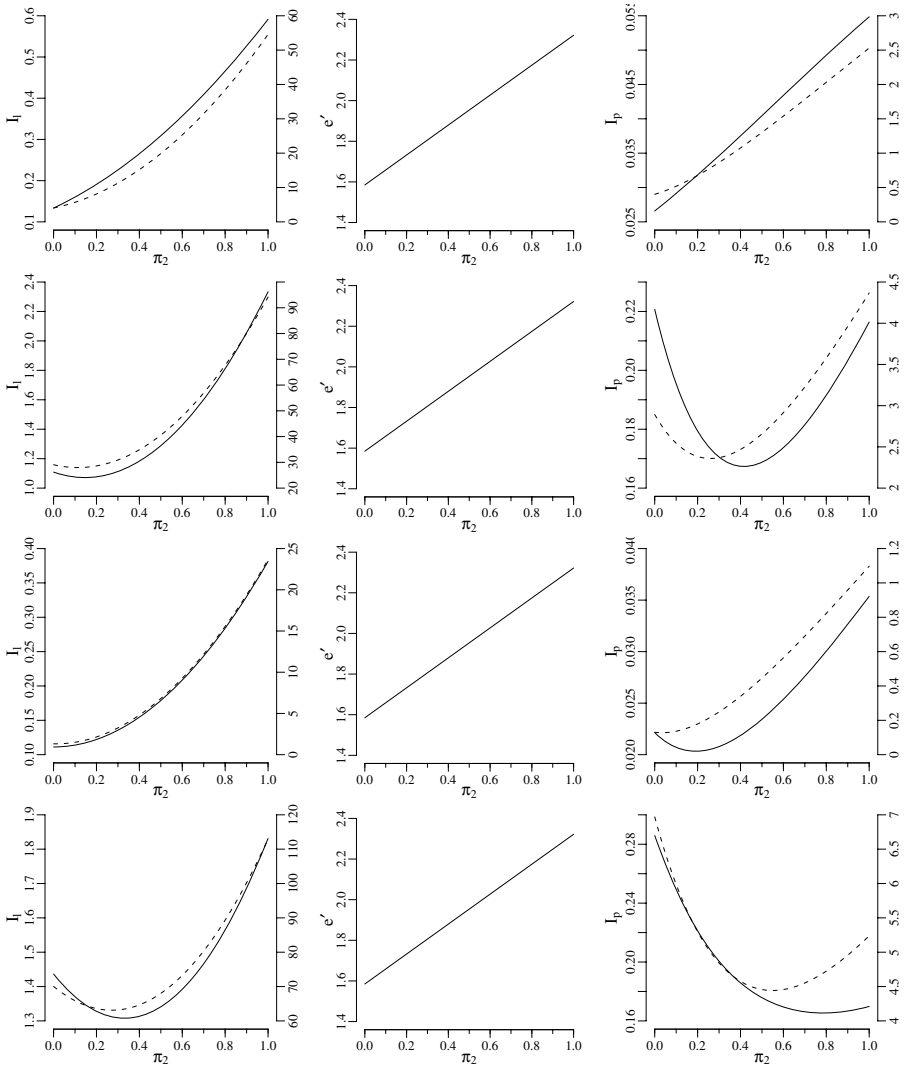

Fig. 1. SFI for varying combinations of ternary and quinary embedding. In each case the x-axis represents the proportion of quinary locations, so that leftmost points correspond to entirely ternary and rightmost entirely quinary embedding. Left graphs, SFI with respect to location rate I_l; middle, embedding efficiency e'; right, SFI with respect to payload I_p. For SFI measures, the solid line is derived from pixel pairs and is denoted on the left axis, the dotted line from 1×4 groups on the right axis. From top to bottom, image sets A to D.

4.2 Optimal Embedding Noise Up to ±3

We can extend the analysis further, but we will go only as far as mixtures of ternary, quinary, and septenary embedding. Such a mixture is specified by

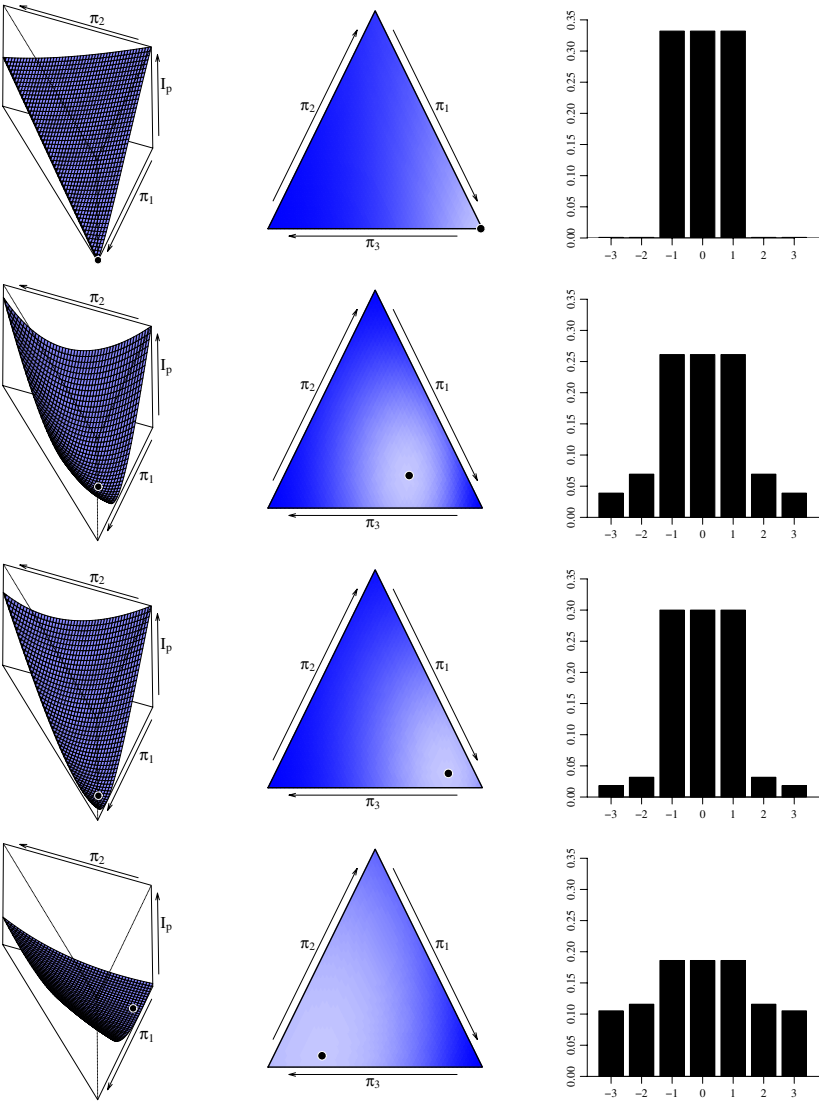

Fig. 2. Combinations of ternary (proportion π_1), quinary (proportion π_2), and septenary (proportion π_3) embedding. Left, a three-dimensional depiction of the surface I_p, as it depends on π_1 and π_2. Centre, the same information in two dimensions, where lighter shading indicates lower SFI. In both cases the location of the minimum is marked. Right, the shape of optimal stego noise, at the SFI minimum. In all cases 1×2 pixel groups have been used to estimate SFI. From top to bottom, image sets A to D.

respective probabilities π_1, π_2, π_3. The matrix for septenary embedding, B_3, is analogous to B_1 and B_2 in Subsect. 4.1 and the same procedure, albeit more computationally expensive, can be used to determine the coefficients c_{ij} for $1 \leq i, j \leq 3$. With two degrees of freedom, the result can be visualised as

Table 1. The optimal mixture of ternary (π_1), quinary (π_2), and septenary (π_3) embedding, for each image set and considering joint histograms from five different pixel group shapes. The final column indicates the relative SFI for the optimum, compared with pure ternary embedding.

Image Set	Group Size	π_1	π_2	π_3	$\frac{I_p(\pi_1,\pi_2,\pi_3)}{I_p(1,0,0)}$
A	1×1	1.000	0.000	0.000	1.000
A	1×2	1.000	0.000	0.000	1.000
A	1×3	1.000	0.000	0.000	1.000
A	1×4	1.000	0.000	0.000	1.000
A	2×2	1.000	0.000	0.000	1.000
B	1×1	0.130	0.115	0.755	0.346
B	1×2	0.576	0.152	0.272	0.684
B	1×3	0.684	0.099	0.216	0.707
B	1×4	0.718	0.119	0.163	0.749
B	2×2	0.707	0.157	0.136	0.792
C	1×1	0.434	0.071	0.495	0.576
C	1×2	0.805	0.067	0.128	0.883
C	1×3	0.921	0.041	0.038	0.944
C	1×4	0.961	0.024	0.015	0.971
C	2×2	1.000	0.000	0.000	1.000
D	1×1	0.000	0.987	0.013	0.181
D	1×2	0.210	0.053	0.736	0.444
D	1×3	0.187	0.346	0.467	0.437
D	1×4	0.408	0.228	0.364	0.522
D	2×2	0.487	0.291	0.222	0.594

either a three-dimensional surface, or a two-dimensional "heatmap"; both types of graphic are displayed in Fig. 2, for SFI in pixel pairs.

In the first row, corresponding to image set A, the I_p surface slants sharply down towards the point where $\pi_1 = 1$: pure ternary embedding is clearly optimal. For image set B the surface is curved, and at the optimum a majority of ternary embedding is mixed with smaller amounts of both quinary and septenary symbols. Set C is similar but with a lower proportion of quinary and septenary symbols, despite the extra noise in the covers: again, we attribute this to saturation. Finally, for set D the mixture features a majority of septenary embedding: these covers are so noisy that, had we extended our analysis to nonary embedding and beyond, it is likely that we would have seen even larger stego noise in the mixture as well. (Of course, there exist other detectors for steganography in previously JPEG-compressed images, which make use of the 8×8 JPEG block structure and can be extremely sensitive [11]. Such detectors are not accounted for in our analysis, which only covers smaller pixel groups.)

To examine other pixel groups, we show how the location of the minimum depends on the group size and shape in Tab. 1. Although there is certainly

variation with pixel group size, most of the results are broadly similar as the group size changes. Further examination (not included here) shows that the surface slopes rather gently near the optimum so that the optimum for, say, 1×2 pixel groups is quite close to optimal for the others. The notable exceptions are the results for 1×1 groups, which have markedly different optima in all cases except set A. There is no contradiction here, and it underlines an important lesson: optimizing embedding to best preserve image histograms is far from optimal when inter-pixel dependencies are considered. This is a familiar pattern from steganalysis literature.

How much difference does it make, to use the optimal combination of embedding functions instead of, say, pure ternary embedding? The final column of Tab. 1 shows the ratio between the SFI I_p of optimally-mixed and pure ternary embedding. For example, looking at set B, we see that the SFI is about 30% lower with a suitable mixture of embedding functions, and this means that a payload of about $(1/0.7)^{1/2} \approx 1.2$ times as large can be carried with equivalent asymptotic KL divergence.

5 Conclusions

Steganographic Fisher Information can be estimated from a large corpus of covers, and we have demonstrated that the technology of SFI estimation can be used to examine convex combinations of embedding functions. It is then simple to find the optimal embedding function combination for a given cover source, though of course the results vary depending on the nature of the cover objects. Optimal SFI is a true information theoretic optimality, indicating lowest asymptotic KL divergence and therefore best security again detection. Except in the image set subject to heavy denoising, combinations of ternary, quinary, and septenary embedding outperform any single embedding method.

Of course, true optimality happens if the embedding method is perfect (preserves the distribution of covers exactly), in which case the SFI is zero and secret payload can be conveyed at a linear, not square root, rate. But constructing such an embedding is difficult and requires perfect knowledge of the cover source, whereas a pseudorandom combination between ternary, quinary, etc, embedding is very simple to implement at both embedder and receiver (though we have not considered the difficulty of transcoding the payload into variable-base). We could take this work further, into quasi-adaptive embedding where the rows of the matrix B are not regular, and find the optimal matrix, but again this asks a lot of the sender and recipient. For the same reasons, we have assumed that the embedder does make use of source coding [6], which usually requires solving systems of linear equations. We must acknowledge that the presence of source coding can complicate the analysis, and may lead to different conclusions.

This paper has a number of limitations. First, our model for covers is of independent groups of pixels. We have argued that, although the model is certainly not accurate for digital images, it mirrors the practice of steganalysis methods which inevitably base their decisions on joint histograms of pixel groups (although the group size might be larger than we are able to examine here), and

therefore SFI is properly connected with the security against such detectors. One difficulty in selecting an optimal embedding combination is that the optimum depends on the size of the pixel groups examined: there is no easy solution to this conundrum, but it makes sense to base decisions on the largest possible group size, since the evidence in large groups subsumes that in small groups. Thankfully, roughly similar results seem to appear in most pixel group shapes with the notable exception of 1×1 groups. We re-iterate this observation: selecting an embedding method to preserve, as best as possible, the histogram of image pixels is a poor strategy. This lesson has been observed a number of times in the literature, with steganography methods touted as "perfect" because of histogram preservation soon falling to steganalysis which considers pairs of pixels or other higher-order information. Nonetheless, a number of authors continue to advance ad hoc embedding methods to preserve cover histograms.

We note that we make the implicit assumption, when using SFI as a benchmark, that the enemy steganalyst has complete knowledge of both the cover source and the chosen embedding function. This is in keeping with Kerckhoffs' Principle but could be argued too pessimistic. However, any other scenario is difficult to examine using KL divergence.

Our experiments were carried out using four sets of cover images, which happened to be conveniently available to the author. In some respects the choice was unfortunate, because they differ in both noise levels and saturation, and there appears to be some interplay between these factors regarding the optimal embedding function. In future work we could examine systematically the effects of noise, saturation, prior JPEG compression, or other macroscopic properties, in isolation, though the computational demands may be considerable.

We should contrast SFI, as an information theoretic measure of asymptotic evidence, with Maximum Mean Discrepancy (MMD), applied to information hiding in [12]. MMD is now quite well-studied though its application in information hiding is still in infancy, and there are efficient estimators allowing MMD to be computed for large-dimensional feature sets. However, although there is some connection between MMD and the performance of kernelized support vector machines, it is not a truly entropic measure and we know no analogue of the connection between KL divergence and maximum hypothesis test performance. Nonetheless, it would be interesting to derive an estimator for asymptotic MMD, and repeat these experiments with that metric to see whether similar results arise.

We may also contrast the SFI estimator here and in [1] with an independent approach to the same problem by Filler & Fridrich [13]. Their estimator differs significantly, modelling the images as a Markov chain with a parameterised transition matrix. They also examine convex combinations, but only of LSB replacement and ternary embedding. Hopefully there will be a confluence of ideas in the area of Fisher Information estimation, which only recently emerged as the true asymptotic benchmark for steganography [4].

Acknowledgements

The author is a Royal Society University Research Fellow. Thanks are due to Tomáš Filler and Rainer Böhme, who both suggested that SFI can be used to optimize embedding functions.

References

1. Ker, A.: Estimating Steganographic Fisher Information in real images. In: Proc. 11th Information Hiding Workshop (to appear, 2009)
2. Cachin, C.: An information-theoretic model for steganography. Information and Computation 192(1), 41–56 (2004)
3. Böhme, R.: Improved Statistical Steganalysis using Models of Heterogeneous Cover Signals. PhD thesis, Technische Universität Dresden (2008)
4. Ker, A.: The ultimate steganalysis benchmark? In: Proc. 9th ACM Workshop on Multimedia and Security, pp. 141–148 (2007)
5. Kullback, S.: Information Theory and Statistics. Dover, New York (1968)
6. Fridrich, J., Soukal, D.: Matrix embedding for large payloads. IEEE Transactions on Information Forensics and Security 1(3), 390–394 (2006)
7. Ker, A., Pevný, T., Kodovský, J., Fridrich, J.: The square root law of steganographic capacity. In: Proc. 10th ACM Workshop on Multimedia and Security, pp. 107–116 (2008)
8. Filler, T., Ker, A., Fridrich, J.: The square root law of steganographic capacity for Markov covers. In: Proc. SPIE. Media Forensics and Security XI, vol. 7254, pp. 801–811 (2009)
9. Ker, A.: Steganalysis of embedding in two least significant bits. IEEE Transactions on Information Forensics and Security 2(1), 46–54 (2007)
10. Ker, A.: A general framework for the structural steganalysis of LSB replacement. In: Barni, M., Herrera-Joancomartí, J., Katzenbeisser, S., Pérez-González, F. (eds.) IH 2005. LNCS, vol. 3727, pp. 296–311. Springer, Heidelberg (2005)
11. Böhme, R.: Weighted stego-image steganalysis for JPEG covers. In: Solanki, K., Sullivan, K., Madhow, U. (eds.) IH 2008. LNCS, vol. 5284, pp. 178–194. Springer, Heidelberg (2008)
12. Pevný, T., Fridrich, J.: Benchmarking for steganography. In: Solanki, K., Sullivan, K., Madhow, U. (eds.) IH 2008. LNCS, vol. 5284, pp. 251–267. Springer, Heidelberg (2008)
13. Filler, T., Fridrich, J.: Fisher Information determines capacity of ϵ-secure steganography. In: Proc. 11th Information Hiding Workshop (to appear, 2009)