A Fusion of Maximum Likelihood and Structural Steganalysis

Andrew D. Ker

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

Abstract. This paper draws together two methodologies for the detection of bit replacement steganography: the principle of maximum likelihood, which is statistically well-founded but has lead to weak detectors in practice, and so-called structural detection, which is sensitive but lacks optimality and can suffer from complicated exposition. The key novelty is to extend structural analysis to include a hypothetical "pre-cover", from which the cover object is imagined to derive. Here, maximum likelihood detection is presented for three structural detectors. Although the algebraic derivation is long, and maximizing the likelihood function difficult in practice, conceptually the new detectors are reasonably simple. Experiments show that the new detectors are the best performers yet, very significantly so in the detection of replacement of multiple bit planes.

1 Introduction

There is no doubt that replacement of low-order bits, whether in digital images, movies, or audio, is an insecure method of embedding and there is a large body of literature on the detection of this steganographic method. Broadly, detectors fall into three categories: methods that target the *structure* of bit replacement[1], those that apply statistical techniques to derive *maximum likelihood* (ML) detectors based on features such as histogram or co-occurrence matrix, and *blind* classifier-based methods that pick ad-hoc features and train on cover and stego images.

Each class has its advantages. The structural detectors, which include (in increasing order of complexity) [1–7], are easily the most sensitive. But their exposition can become complex, and the methods themselves are often based on dubious statistical principles (for example, assuming that all observations of random variables equal the expectation). Statistical rigour is at the heart of maximum likelihood detectors [8–11] but their performance is weak. The blind classifier-based methods can detect a range of embedding methods (not limited to bit replacement) but have neither good detection power nor statistical rigour and we shall not consider this last class here.

This paper combines the theoretical rigour of ML detection with the sensitivity of structural steganalysis. Not only does this produce superior detectors, it

---

[1] An even cover sample can be incremented or unchanged, but never decremented, when the least significant bit is replaced, conversely for odd samples; similar structure occurs in replacement of other bits.

also presents a framework which avoids some of the difficulties that plague structural steganalysis using large groups of pixels [6]. Indeed, with hindsight, this is arguably the mathematical setting within which structural steganalysis should always have been presented. We will confine our analysis to digital grayscale images with spatial-domain embedding, but the principles described here should apply in other domains too.

The contents of the paper is as follows. Section 2 is a brief survey of ML and structural steganalysis methods for detection of bit replacement. Section 3 describes ML detection of least significant bit (LSB) replacement using an extended structural analysis, deriving a likelihood function and explaining a maximization procedure (at this stage only a simple, slow, maximization is used), with experimental results following. Sections 4 and 5 repeat the same process and experiments for the detection of embedding in two lowest bit planes and then a more complex structural analysis of embedding in just the LSB plane. Finally, Sect. 6 draws conclusions. In order to include multiple detectors the later sections contain little detail; this paper aims to present the principles only, and we refer to an extended version of this paper [12] for fuller derivations, explicit likelihood functions, and wider experimental results.

2 ML Steganalysis and Structural Steganalysis

The starting point for ML steganalysis is the Neyman-Pearson Lemma, which states that the optimal discriminator between two point hypotheses is given by the likelihood ratio. Further, good performance for composite hypotheses is given by the generalized likelihood ratio (GLR) test and for parameter estimators by the method of maximum likelihood (although the optimality of these extensions is not universal). One of the earliest uses of ML in steganalysis is found in [8], which derives the effect of LSB replacement on the probability mass function (PMF) of a signal; if the PMF of the cover source is known, it is possible to create a GLR test for the presence of data hidden by LSB overwriting and a ML estimate for the size of payload.

In practice the PMF of the cover is not known: it must be estimated by either filtering the observed PMF [8] or postulating an "ideal" cover PMF ([9] uses this latter approach, in a transform domain). The detectors are weak for three reasons: estimation of the cover PMF is subject to inaccuracy; considering only the PMF discards Markovicity in the cover source; and the detectors are unable to exploit the aforementioned structure of LSB replacement. Improved detectors based on maximum likelihood principles can be found in [11], which models the source as a Markov chain, and [10] which also tries to exploit some of the structure of LSB replacement via the so-called "stair-step" effect. Nonetheless, even the improved detectors remain unable to detect even moderate payloads[2].

In contrast, the most successful class of steganalyzers for bit replacement (including LSB embedding as well as replacement of multiple bit planes) are

---

[2] We should be clear that some of the ML detectors cited are not specialised towards LSB replacement, so weak performance is expected.

those which focus on the structure of the embedding process. Their general strategy, described in [5], is as follows. First, define a feature set of stego images which depends on the size of payload p, a vector $\mathbf{S}(p)$; second, derive how $\mathbf{S}(p)$ depends on p and $\mathbf{S}(0)$ and then invert so that, given a stego image, we can hypothesise a value for p and compute what this would imply for $\mathbf{S}(0)$. Third, express a model for covers in terms of $\mathbf{S}(0)$. Then we create an estimator for the payload size p as whichever implies a value of $\mathbf{S}(0)$ closest to the model.

The structure of LSB replacement in individual samples is trivial and does not give a detector, but extending to pairs or larger groups of pixels produces extremely sensitive detectors. The earliest detector used the structure implicitly [1]; analysis of the effect of bit replacement on specific structural features was first given in [2] and extended in [4,5]. Further application, to large groups of pixels and replacement of two bit planes, can be found in [6,7]. A key novelty is the principle of *least-squares* [3], which defines the closest cover fit to be when the features' sum-square deviation (from an ideal) is minimized.

Although very effective, the structural detectors have drawbacks. The analysis of the effect of bit replacement on groups of three or more pixels can be handled in an elegant manner [5], but specification of the cover model for such groups can become desperately complicated, as demonstrated in [6]. And the statistical methodology is poor: it is not truly the case that stego features $\mathbf{S}(p)$ depend deterministically on p and $\mathbf{S}(0)$ (it also depends on the payload content, usually assumed random) and it is necessary to take expectations. Neither is the cover model exact. Moreover a least-squares cover fit, while plausible, cannot be shown optimal (indeed we shall later be able to see that it is not).

3   Maximum Likelihood and the "Couples" Structure

We now present a fusion of the sensitive features used by the structural detectors with the principles of maximum likelihood. In this section we consider the features used by the detectors known as *Sample Pairs* (SPA) [2,3] or *Couples* [5] (other possibilities will be examined in later sections).

We assume a single-channel cover image[3] which consists of N pixels with intensities $s_1, s_2, \ldots, s_N$ in the range $0 \ldots 2M+1$ (typically $M = 127$). A *sample pair* is a pair of pixels (j, k) for some $j \neq k$. Let $\mathcal{P}$ be the set of all pairs that represent adjacent pixels and define some subsets of $\mathcal{P}$, called the *trace subsets*:

$$\mathcal{D}_m = \{(j, k) \in \mathcal{P} \mid s_k = s_j + m\}$$
$$\mathcal{E}_m = \{(j, k) \in \mathcal{P} \mid s_k = s_j + m, \text{ with } s_j \text{ even}\}$$
$$\mathcal{O}_m = \{(j, k) \in \mathcal{P} \mid s_k = s_j + m, \text{ with } s_j \text{ odd}\}$$

for $-2M + 1 \leq m \leq 2M + 1$. Analysis of the movement of sample pairs among these subsets, when payload is embedded by LSB replacement, is the key to structural steganalysis. The cover model in [2,3] is $|\mathcal{E}_m| = |\mathcal{O}_m|$, although only odd indices can be used for steganalysis.

---

[3] Colour images are usually separated into colour channels, whose signals are either treated separately or concatenated.
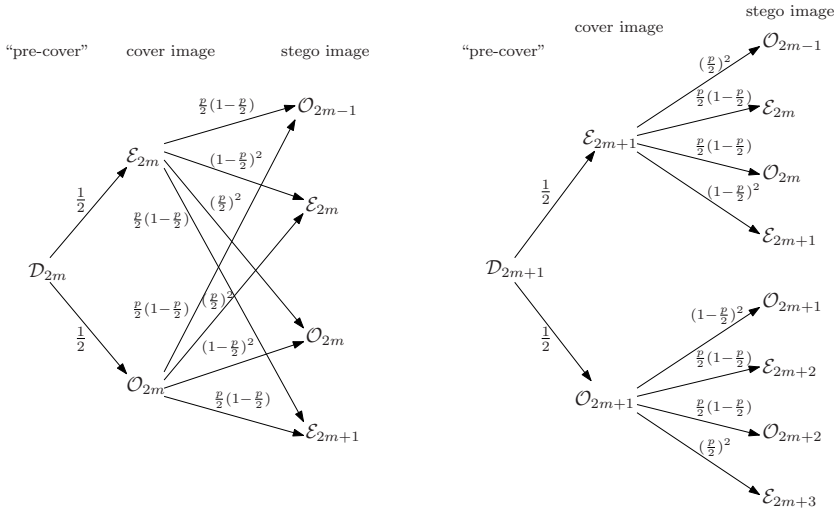
Fig. 1. The evolution of the *Couples* trace subsets, from "pre-cover" to stego-image, when proportion $\frac{p}{2}$ least significant bits are flipped

3.1 Derivation of Likelihood Function

We present here a new analysis of trace subsets. To model trace subsets in cover images, we use the technique of [13], which proposed a model as part of the analysis of errors in least-squares steganalysis. Imagine that, for each m, the set $\mathcal{D}_m$ is first determined (in what we will call a "pre-cover") and then partitioned binomially to make the cover image, with each sample pair in $\mathcal{D}_m$ placed independently and equiprobably into $\mathcal{E}_m$ or $\mathcal{O}_m$. Of course this bears no relationship to the physical processes which create, for example, an image in the CCD of a digital camera – there is no "pre-cover" except in our imagination – but it is a natural explanation of the cover property $|\mathcal{E}_m| = |\mathcal{O}_m|$ and it is demonstrated, in [13], very accurate in most circumstances[4].

Second, we model how LSB replacement converts trace subsets in the cover to trace subsets in a stego image, depending on the size of payload. We suppose that the LSB of each sample is flipped, independently, with probability $\frac{p}{2}$ – this corresponds to embedding a payload (uncorrelated with the cover) of length pN by randomized LSB replacement, or perhaps the embedding of a longer payload when LSB flipping is used as part of a source coding scheme [14] (of course the estimator will be truly for twice the number of flipped pixels, regardless of the connection this has with payload size; it could be nothing else). Under this

---

[4] In [13] the limitations of this model are explored; it is not very accurate for m near zero and there is observed a negative correlation between the sizes of $\mathcal{E}_1$ and $\mathcal{E}_{-1}$; in fact it is possible to explain these features in terms of over-saturated images, but that is not within the scope of this paper. We content ourselves with using the model as is, in the knowledge that it is not perfect.

embedding operation each quartet of trace subsets $\mathcal{O}_{2m-1}$, $\mathcal{E}_{2m}$, $\mathcal{O}_{2m}$, $\mathcal{E}_{2m+1}$ undergoes permutation, with probabilities determined by p. We will not repeat that analysis, since it can be found in much of the other steganography literature including [2, 3, 5].

The overall evolution of sample pairs, from "pre-cover", through cover, to stego image, with the probability of each transition, is shown in Fig. 1. Note that transitions originating from $\mathcal{D}_m$ are of different structure depending on whether m is even or odd, because $\mathcal{E}_{2m}$ and $\mathcal{O}_{2m}$ can be interchanged by embedding whereas $\mathcal{E}_{2m+1}$ and $\mathcal{O}_{2m+1}$ cannot. Figure 1 is the foundation from which we will derive a likelihood for an observed stego image, given the sizes of the $\mathcal{D}_m$ in the pre-cover and p.

Consider for example the trace subset $\mathcal{E}_{2m+1}$ in the stego image. A sample pair has arrived there by one of four possible paths:

(i) Beginning in $\mathcal{D}_{2m}$ in the pre-cover, it was placed into $\mathcal{E}_{2m}$ in the cover and moved to $\mathcal{E}_{2m+1}$ under embedding; the probability of a pair beginning in $\mathcal{D}_{2m}$ making these transitions is $\frac{1}{2}\frac{p}{2}(1 - \frac{p}{2})$.

(ii) Beginning in $\mathcal{D}_{2m}$, it was placed into $\mathcal{O}_{2m}$ and moved to $\mathcal{E}_{2m+1}$ under embedding; again the probability of a pair in $\mathcal{D}_{2m}$ making such transitions is $\frac{1}{2}\frac{p}{2}(1 - \frac{p}{2})$.

(iii) Beginning in $\mathcal{D}_{2m+1}$, it was placed into $\mathcal{E}_{2m+1}$ and did not move under embedding; the probability of a pair in $\mathcal{D}_{2m+1}$ doing so is $\frac{1}{2}(1 - \frac{p}{2})^2$.

(iv) Beginning in $\mathcal{D}_{2m-1}$, it was placed into $\mathcal{O}_{2m-1}$ and moved to $\mathcal{E}_{2m+1}$ under embedding; the probability of a pair in $\mathcal{D}_{2m+1}$ doing so is $\frac{1}{2}(\frac{p}{2})^2$.

Similar calculations can be carried out for all trace subsets $\mathcal{E}_m$ and $\mathcal{O}_m$: each can have arisen from one of three of the $\mathcal{D}_m$ in the pre-cover. We display these probabilities in Tab. 1, with the columns corresponding to the source $\mathcal{D}_m$ and the rows to the $\mathcal{E}_m$ and $\mathcal{O}_m$ in the stego image. Therefore the full table, for 8 bit images, has 511 columns (trace subsets $\mathcal{D}_m$ exist for $-255 \leq m \leq 255$) and 1020 rows ($\mathcal{O}_{-255}, \mathcal{E}_{-254}, , \mathcal{O}_{-254}, ..., \mathcal{E}_{254}, \mathcal{O}_{254}, \mathcal{E}_{255}$) and only a portion is displayed, from which the rest can be constructed. We have made the convenient abbreviation $\pi_i = \frac{1}{2}(\frac{p}{2})^i(1 - \frac{p}{2})^{2-i}$ for $i = 0, 1, 2$. Observe that fifth displayed row corresponds to our analysis, above, of trace subsets finishing in $\mathcal{E}_{2m+1}$.

Now suppose that an image is observed, with unknown payload. Let us form a vector of the observed sizes of trace subsets in the stego image, interleaving as $\boldsymbol{A}' = (O'_{-255}, E'_{-254}, , O'_{-254}, ..., E'_{255})^T$, to correspond with the row labels of the complete table excerpted in Tab. 1. We write $\boldsymbol{p}_m$ for the column of Tab. 1 headed by the pre-cover trace subset $\mathcal{D}_m$ and suppose that the size of this subset was d_m. Because each sample pair beginning in the pre-cover must end up in some trace subset of the stego image, and under the assumption of random uncorrelated payload all destinations are independent, we have that the number of pairs in each trace subset originating from $\mathcal{D}_m$ takes a multinomial distribution. Summing up over all subsets $\mathcal{D}_m$ in the pre-cover, we thus deduce that $\boldsymbol{A}' = \sum_m \boldsymbol{A}'_m$, where each $\boldsymbol{A}'_m$ has a multinomial distribution with size parameter d_m and probability vector $\boldsymbol{p}_m$.

Table 1. Table showing the probability that, given a random payload of pN, each sample pair from a trace subset $\mathcal{D}_m$ in the pre-cover is observed in each trace subset $\mathcal{E}_m$ or $\mathcal{O}_m$ in a stego image

	$\mathcal{D}_{2m-1}$	$\mathcal{D}_{2m}$	$\mathcal{D}_{2m+1}$	$\mathcal{D}_{2m+2}$
$\mathcal{E}_{2m-1}$	π_0	0	0	0
$\mathcal{O}_{2m-1}$	π_0	$2\pi_1$	π_2	0
$\mathcal{E}_{2m}$	π_1	$\pi_0+\pi_2$	π_1	0
$\mathcal{O}_{2m}$	π_1	$\pi_0+\pi_2$	π_1	0
$\mathcal{E}_{2m+1}$	π_2	$2\pi_1$	π_0	0
$\mathcal{O}_{2m+1}$	0	0	π_0	$2\pi_1$
$\mathcal{E}_{2m+2}$	0	0	π_1	$\pi_0+\pi_2$
$\mathcal{O}_{2m+2}$	0	0	π_1	$\pi_0+\pi_2$
$\mathcal{E}_{2m+3}$	0	0	π_2	$2\pi_1$
$\mathcal{O}_{2m+3}$	0	0	0	0

In order to find the likelihood for this sum of distributions we make the standard multivariate Gaussian approximation to the multinomial distribution (generally accurate as long as the size of the original trace subset is not very small):

$$\boldsymbol{A} \stackrel{.}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu} = \sum_m d_m \boldsymbol{p}_m$, $\boldsymbol{\Sigma} = \sum_m d_m (\boldsymbol{\Delta}_{\boldsymbol{p}_m} - \boldsymbol{p}_m \boldsymbol{p}_m^T)$, and $\boldsymbol{\Delta}_{\boldsymbol{v}}$ represents a diagonal matrix with $\boldsymbol{v}$ on the diagonal. This allows us to compute the (log-)likelihood that a given image with trace subsets $\boldsymbol{a}$ arose from a pre-cover with specific trace subset sizes $\boldsymbol{d}$ and a particular proportionate payload p:

$$l(\boldsymbol{a}; p, \boldsymbol{d}) = -\frac{L}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu})$$

where L is the length of the vector of observations $\boldsymbol{a}$.

Although we omit the intermediate calculations, it is worthwhile to see the full form of the mean vector $\boldsymbol{\mu} = (\dots, \mathrm{E}[O'_{2m-1}], \mathrm{E}[E'_{2m}], \mathrm{E}[O'_{2m}], \mathrm{E}[E'_{2m+1}], \dots)^T$ and covariance matrix $\boldsymbol{\Sigma}$ whose entries are $\mathrm{Cov}[E'_m, O'_m]$. Extracting $\boldsymbol{p}_m$ from Tab. 1 and using $\boldsymbol{\mu} = \sum_m d_m \boldsymbol{p}_m$, we derive

$$\begin{aligned}
\mathrm{E}[O'_{2m-1}] &= \pi_0 d_{2m-1} + 2\pi_1 d_{2m} + \pi_2 d_{2m+1} \\
\mathrm{E}[E'_{2m}] &= \pi_1(d_{2m-1} + d_{2m+1}) + (\pi_0 + \pi_2)d_{2m} \\
\mathrm{E}[O'_{2m}] &= \pi_1(d_{2m-1} + d_{2m+1}) + (\pi_0 + \pi_2)d_{2m} \\
\mathrm{E}[E'_{2m+1}] &= \pi_2 d_{2m-1} + 2\pi_1 d_{2m} + \pi_0 d_{2m+1}
\end{aligned}$$

for each m. From $\boldsymbol{\Sigma} = \sum_m d_m (\boldsymbol{\Delta}_{\boldsymbol{p}_m} - \boldsymbol{p}_m \boldsymbol{p}_m^T)$ we can compute all covariances; they are displayed in the following table. We make a further abbreviation, $\sigma = \pi_0 + \pi_2$, and elide the covariances of O'_{2m} (which are almost identical to those of E'_{2m}) in order to fit the table onto the page.

	O'_{2m-1}	E'_{2m}	O'_{2m}	E'_{2m+1}
O'_{2m-3}	$-\pi_0\pi_2 d_{2m-1}$	$-\pi_1\pi_2 d_{2m-1}$	$\leftarrow$	$-\pi_2^2 d_{2m-1}$
E'_{2m-2}	$-\pi_0\pi_1 d_{2m-1}$	$-\pi_1^2 d_{2m-1}$	$\leftarrow$	$-\pi_1\pi_2 d_{2m-1}$
O'_{2m-2}	$-\pi_0\pi_1 d_{2m-1}$	$-\pi_1^2 d_{2m-1}$	$\leftarrow$	$-\pi_1\pi_2 d_{2m-1}$
E'_{2m-1}	$-\pi_0^2 d_{2m-1}$	$-\pi_0\pi_1 d_{2m-1}$	$\leftarrow$	$-\pi_0\pi_2 d_{2m-1}$
O'_{2m-1}	$\pi_0(1-\pi_0)d_{2m-1}$ $+2\pi_1(1-2\pi_1)d_{2m}$ $+\pi_2(1-\pi_2)d_{2m+1}$	$-\pi_0\pi_1 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_1\pi_2 d_{2m+1}$	$\leftarrow$	$-\pi_0\pi_2 d_{2m-1}$ $-4\pi_1^2 d_{2m}$ $-\pi_0\pi_2 d_{2m+1}$
E'_{2m}	$-\pi_0\pi_1 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_1\pi_2 d_{2m+1}$	$\pi_1(1-\pi_1)d_{2m-1}$ $+\sigma(1-\sigma)d_{2m}$ $+\pi_1(1-\pi_1)d_{2m+1}$	$\swarrow$	$-\pi_1\pi_2 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_0\pi_1 d_{2m+1}$
O'_{2m}	$-\pi_0\pi_1 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_1\pi_2 d_{2m+1}$	$-\pi_1^2 d_{2m-1}$ $-\sigma^2 d_{2m}$ $-\pi_1^2 d_{2m+1}$	$\searrow$	$-\pi_1\pi_2 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_0\pi_1 d_{2m+1}$
E'_{2m+1}	$-\pi_0\pi_2 d_{2m-1}$ $-4\pi_1^2 d_{2m}$ $-\pi_0\pi_2 d_{2m+1}$	$-\pi_1\pi_2 d_{2m-1}$ $-2\pi_1\sigma d_{2m}$ $-\pi_0\pi_1 d_{2m+1}$	$\leftarrow$	$\pi_2(1-\pi_2)d_{2m-1}$ $+2\pi_1(1-2\pi_1)d_{2m}$ $+\pi_0(1-\pi_0)d_{2m+1}$
O'_{2m+1}	$-\pi_0\pi_2 d_{2m+1}$	$-\pi_0\pi_1 d_{2m+1}$	$\leftarrow$	$-\pi_0^2 d_{2m+1}$
E'_{2m+2}	$-\pi_1\pi_2 d_{2m+1}$	$-\pi_1^2 d_{2m+1}$	$\leftarrow$	$-\pi_0\pi_1 d_{2m+1}$
O'_{2m+2}	$-\pi_1\pi_2 d_{2m+1}$	$-\pi_1^2 d_{2m+1}$	$\leftarrow$	$-\pi_0\pi_1 d_{2m+1}$
E'_{2m+3}	$-\pi_2^2 d_{2m+1}$	$-\pi_1\pi_2 d_{2m+1}$	$\leftarrow$	$-\pi_0\pi_2 d_{2m+1}$

These hold for all m, and other covariances are zero. Although fairly appalling to look at, the matrix can quickly be computed from p and d.

It is worthwhile to contrast the principle of structural ML estimation with the standard structural estimators. Consider, for example, [2]: the cover model used there (with a minor variation) is that $|\mathcal{E}_{2m+1}| = |\mathcal{O}_{2m+1}|$ for each m. In Fig. 1 we can see that this is true *in expectation*, but a more sophisticated model can quantify deviations from exact equality. Similarly, the standard structural analysis of embedding is the same as the transitions from cover to stego object in Fig. 1, but it assumes that the observed stego image trace subset sizes are exactly their expectation, not allowing for random variation. The ML method allows us, in effect, to take account of the fact that such approximations are more accurate when the numbers involved are larger. Finally, the principle of least-squares estimation [3, 7] is optimal if the random variables whose least squares are minimized are Gaussian and independent. Under our model they are Gaussian, but considering the covariance matrix $\mathbf{\Sigma}$, one can show that the relevant quantities are not independent, and hence least-squares steganalysis is, at best, an approximation to optimality.

3.2   Implementing a ML Estimator

In principle, the rest is simple: given an image we observe the trace subsets a and find the value of p (along with, as nuisance parameters, all the d) which maximizes the log-likelihood. In practice there are some difficulties.

First, how to find the maximum? There is no apparent closed form for it so we must proceed numerically. The function is differentiable, but computing the derivative (particularly with respect to p) seems extremely difficult and we have not yet completed the calculation. This prevents us from using standard iterative/scoring methods to locate the maximum. So that we can test the accuracy of a structural ML detector, without being unduly distracted by implementation issues, we have settled on a temporary, computationally expensive, solution: the maximization method of Nelder and Mead [15] which walks a simplex through the surface to be optimized and does not require derivatives. Of course, such a method converges only slowly to a maximum. We must also find a point from which to start the optimization: we used one of the standard methods for estimating LSB replacement payload [2] as an initial value of p, and created initial values for $\boldsymbol{d}$ by inverting the second half of the transition diagram in Fig. 1 (such an inversion is already part of the standard method for structural steganalysis of LSB replacement).

Unfortunately, maximization of the full likelihood function is computationally unfeasible: it is a 512-dimensional problem, which is simply too complex to optimize in any reasonable amount of time. But we can cheaply reduce the dimensionality by discarding the parameters d_m for $|m|$ large: very few adjacent pairs of pixels have a large difference, and discarding a minority of pairs does not reduce the evidential base too much. We found that considering only $\mathcal{D}_{-11}$ to $\mathcal{D}_{11}$ still gave excellent performance (in our test set of cover images, 75.1% of sample pairs are found in this range) and reduces the dimensionality of the maximization to 24.

We can further ameliorate the computational cost by evaluating the log-likelihood efficiently: because the matrix in Tab. 1 has many zeros (recall that each trace subset in the stego image can come from only one of three subsets in the pre-cover) the covariance matrix $\boldsymbol{\Sigma}$ is zero except near to the diagonal and there exist efficient methods to compute $|\boldsymbol{\Sigma}|$ and $(\boldsymbol{a}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{a}-\boldsymbol{\mu})$ exploiting the *Cholesky decomposition* of $\boldsymbol{\Sigma}$.

A final improvement is to observe that $\mathcal{E}_{2m}$ and $\mathcal{O}_{2m}$ are always treated equally in Tab. 1, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. This means that it is safe to combine such subsets, observing and computing likelihood of only the sum $E'_{2m}+O'_{2m}$, but still separating E'_{2m+1} and O'_{2m+1}. This leaves the dimensionality of the maximization unchanged, but reduces the size of the quadratic form which must be computed for each evaluation of the likelihood, from $L = 44$ to $L = 33$.

Nonetheless the ML estimator is still quite slow due to the inefficient maximization method. Using a moderately-optimized implementation in C, and a computer with a 64-bit processor running at 2Ghz, we timed the standard *SPA* estimator [2], a least-squares variant *Couples/LSM* [3], and our new *Couples/ML* estimator, on various image sizes. The results are displayed in Tab. 2; observe that the size of the images is almost irrelevant to the ML estimator which spends almost all of its time maximizing the likelihood function, whereas the standard methods spend most of their time simply counting the trace subsets and hence their time complexity is roughly linear in the image size.

Table 2. Images processed per second, for three detectors and three image sizes

Detector	Image size		
	0.5Mpixel	1.0Mpixel	1.5Mpixel
SPA	36.1	21.0	15.11
Couples/LSM	36.7	21.5	15.26
Couples/ML	0.401	0.392	0.387

There still remains a difficulty with implementing the estimator. As p tends to zero, the covariance matrix $\boldsymbol{\Sigma}$ becomes singular. Thus the likelihood function can grow without bound as p approaches zero, if the $\boldsymbol{d}$ are chosen in a certain way, and the likelihood optimizer can find its way to an incorrect "solution". This is a standard problem and can be seen as either overfitting – some of the observed trace subsets can match their expectations with arbitrarily good accuracy, at the expense of others – or a breakdown in the multivariate Gaussian approximation to the multinomial when the probabilities involved are small. The common solution is to place a prior distribution on the difficult parameter (here p) and convert the ML estimator to a maximum a posteriori (MAP) estimator. If the prior is Gaussian this amounts to placing a quadratic penalty on the log-likelihood function.

We already needed to find another estimate $\hat{p}$ from which to begin the optimization, and found that using a Gaussian prior with mean $\hat{p}$ and variance $\hat{p}^2/100$ (knowing that we want to keep the estimator away from $p = 0$ without fixing it too closely to the less-accurate prior estimate) was an effective solution.

3.3   Experimental Results

We now benchmark the MAP estimator *Couples/ML* against the best of the other structural detectors based on pairs of pixels: *SPA* [2], *Couples/LSM* (in the form described in [7]) and a new weighted least-squares version *Couples/WLSM* [16]. We report only results which use a set of 3000 never-compressed grayscale covers, which have been downsampled to approximately 0.3Mpixels from scanned high-resolution originals. LSB replacement steganography was simulated with embedding rates $0, 0.05, \ldots 1$, and accuracy of the resulting estimates compared.

Making a comparison of estimator accuracy is not quite straightforward because the shape of error distribution is not necessarily the same for all estimators. Therefore it can be misleading to compare, for example, only mean-square error (such a measure will heavily penalize estimators with a few large outliers) or only interquartile range (which is completely insensitive to outliers). Furthermore, there is both estimator bias (least-squares estimators consistently underestimate large payloads) and spread to consider.

We will display a number of measures in order to give a balanced assessment: mean-square error encompasses both bias and spread, sample mean estimates
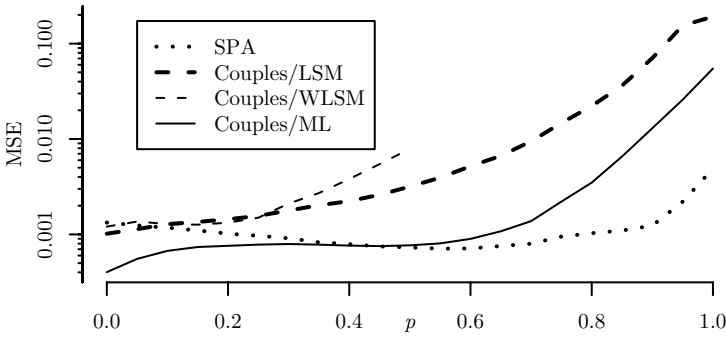
	Detector	$p = 0$	$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 1.0$
Bias	SPA	0.059	0.018	0.012	0.127	−6.298
	Couples/LSM	−0.061	−0.780	−1.864	−6.380	−36.597
	Couples/WLSM	0.094	−1.499	−5.974		
	Couples/ML	0.824	−0.490	−0.466	−2.349	−20.315
IQR	SPA	2.699	2.576	2.618	3.383	4.037
	Couples/LSM	2.782	2.814	3.265	6.881	27.738
	Couples/WLSM	2.527	2.792	5.780		
	Couples/ML	1.034	2.190	2.652	4.860	6.673
MAE	SPA	2.076	1.909	1.822	2.245	6.305
	Couples/LSM	2.086	2.323	2.939	6.823	36.675
	Couples/WLSM	2.035	2.382	6.120		
	Couples/ML	0.826	1.725	1.843	3.525	21.039

Fig. 2. Above, chart displaying how the estimator mean-square error (MSE; log axis) of the structural detectors depends on the proportionate payload p. Below, table displaying other measures of estimator accuracy: sample bias, interquartile range (IQR), mean absolute error (MAE), all displayed $\times 10^2$, for five payload sizes.

bias, interquartile range is a measure of spread without regard to outliers, and mean absolute error is a combination measure which is robust but not insensitive to outliers. The first of these is charted, and the others tabulated, in Fig. 2.

Observe that the *Couples/ML* estimator is uniformly superior to both *Couples/LSM* and *Couples/WLSM* (the latter is not benchmarked for proportionate payloads greater than 50% because its "optimal" weighting depends on the true payload being small). It is also superior to the standard *SPA* method for payloads $p < 0.5$, and particularly for very small payloads. Of course, in some applications the increased computational costs of the maximization will outweigh the accuracy advantages. The slight positive bias at zero payloads, for the ML method, is because negative estimates can never happen.

It is curious that one of the very first payload estimators, which claims no optimality, should appear the best performer for large payloads. In fact this can

be explained by reference to a weakness in the cover model used here, which the standard SPA method coincidentally does exactly the right thing to avoid, but such an explanation is beyond the scope of this work.

4   Maximum Likelihood and the "2Couples" Structure

We now outline the extension of ML structural analysis to embedding in the two lowest bit planes; this was called "2LSB" embedding in [7], where a least-squares structural steganalyzer was presented. Again we consider pairs of pixels evolving from a "pre-cover" through a cover to a stego-image. The trace subsets, which we re-use from [7], are

$$\mathcal{D}_m = \{(j,k) \in \mathcal{P} \mid s_k = s_j + m\}$$
$$\mathcal{B}_m^i = \{(j,k) \in \mathcal{P} \mid s_k = s_j + m, \text{ with } s_j \equiv i \ (\text{mod } 4)\}$$

for $i = 0, 1, 2, 3$. They serve the same purpose as in the Couples case, exposing exactly the structure of replacement of two bit planes. Traditional least-squares steganalysis assumes that the sizes of $\mathcal{B}_m^0$, $\mathcal{B}_m^1$, $\mathcal{B}_m^2$ and $\mathcal{B}_m^3$ are all equal in cover images, for each m, although in [7] it was noted that not all such equalities are useful for steganalysis – a difficulty that the ML framework is able to avoid.

As before, we must model differences between the sizes of $\mathcal{B}_m^i$ in covers. We will imagine that the "pre-cover" has fixed trace subsets $\mathcal{D}_m$ for each m, and that every pair in $\mathcal{D}_m$ moves to one of $\mathcal{B}_m^i$ in the cover object, independently and equiprobably. Again, this corresponds with traditional structural steganalysis in expectation. Next, embedding a payload of length $2pN$ (or randomizing the two least bits of each pixel with probability p) causes the 16 trace subsets $\mathcal{B}_{4m}^0$, $\mathcal{B}_{4m+1}^0$, $\mathcal{B}_{4m+2}^0$, $\mathcal{B}_{4m+3}^0$ $\mathcal{B}_{4m-1}^1$, $\mathcal{B}_{4m}^1$, $\mathcal{B}_{4m+1}^1$, $\mathcal{B}_{4m+2}^1$ $\mathcal{B}_{4m-2}^2$, $\mathcal{B}_{4m-1}^2$, $\mathcal{B}_{4m}^2$, $\mathcal{B}_{4m+1}^2$ $\mathcal{B}_{4m-3}^3$, $\mathcal{B}_{4m-2}^3$, $\mathcal{B}_{4m-1}^3$, $\mathcal{B}_{4m}^3$ to be permuted with certain probabilities, which are described in [7] and will not be repeated here.

The added complexity means that we cannot fit in a diagram analogous to Fig. 1: a pair in $\mathcal{D}_m$ in the pre-cover can move to any one of 32 trace subsets in the stego image. We will skip directly to the analogue of Tab. 1, which we display in Tab. 3 using the abbreviation $\pi_i = \frac{1}{4}(\frac{p}{4})^i(1 - \frac{3p}{4})^{2-i}$.

It would be horrendous to display the mean and covariance matrices here, which are much more complex than those in Sect. 3, and we omit all calculations. But the principle of ML estimation can now proceed as in the previous section: concatenate the observed trace subset sizes into a vector $\boldsymbol{a}$, and maximize the log-likelihood function

$$l(\boldsymbol{a}; p, \boldsymbol{d}) = -\frac{L}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{a} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{a} - \boldsymbol{\mu})$$

where $\boldsymbol{\mu} = \sum_m d_m \boldsymbol{p}_m$, $\boldsymbol{\Sigma} = \sum_m d_m (\boldsymbol{\Delta}_{\boldsymbol{p}_m} - \boldsymbol{p}_m \boldsymbol{p}_m^T)$, and $\boldsymbol{p}_m$ are the columns of Tab. 3. We must convert to a MAP estimator to avoid overfitting (the same prior is appropriate) and again we find it necessary to reduce the dimensionality of the optimization by considering only $\mathcal{D}_{-11}$ to $\mathcal{D}_{11}$, for a 24 dimensional surface.

Table 3. Table showing the probability that, given a random payload of $2pN$ embedded by 2LSB replacement, each sample pair from a trace subset $\mathcal{D}_m$ in the pre-cover is observed in each trace subset $\mathcal{B}^i_m$ in a stego image

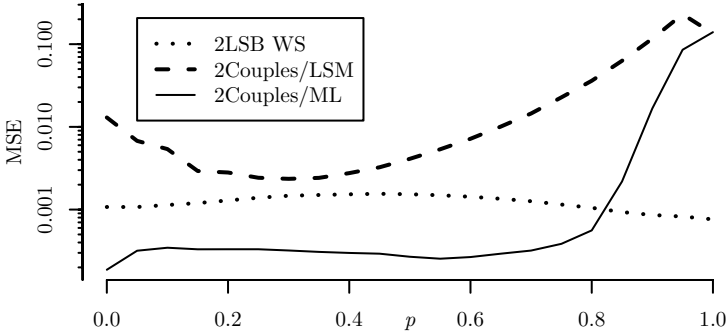|	$\mathcal{D}_{4m-3}$	$\mathcal{D}_{4m-2}$	$\mathcal{D}_{4m-1}$	$\mathcal{D}_{4m}$	$\mathcal{D}_{4m+1}$	$\mathcal{D}_{4m+2}$	$\mathcal{D}_{4m+3}$
$\mathcal{B}^3_{4m-3}$	π_0	$2\pi_1$	$2\pi_1+\pi_2$	$2\pi_1+2\pi_2$	$3\pi_2$	$2\pi_2$	π_2
$\mathcal{B}^3_{4m-2}$	π_1	$\pi_0+\pi_2$	$2\pi_1+\pi_2$	$2\pi_1+2\pi_2$	$\pi_1+2\pi_2$	$2\pi_2$	π_2
$\mathcal{B}^3_{4m-1}$	π_1	$\pi_1+\pi_2$	$\pi_0+2\pi_2$	$2\pi_1+2\pi_2$	$\pi_1+2\pi_2$	$\pi_1+\pi_2$	π_2
$\mathcal{B}^3_{4m}$	π_1	$\pi_1+\pi_2$	$\pi_1+2\pi_2$	$\pi_0+3\pi_2$	$\pi_1+2\pi_2$	$\pi_1+\pi_2$	π_1
$\mathcal{B}^2_{4m-2}$	π_1	$\pi_0+\pi_2$	$2\pi_1+\pi_2$	$2\pi_1+2\pi_2$	$\pi_1+2\pi_2$	$2\pi_2$	π_2
$\mathcal{B}^2_{4m-1}$	π_2	$2\pi_1$	$\pi_0+2\pi_2$	$2\pi_1+2\pi_2$	$2\pi_1+\pi_2$	$2\pi_2$	π_2
$\mathcal{B}^2_{4m}$	π_2	$\pi_1+\pi_2$	$2\pi_1+\pi_2$	$\pi_0+3\pi_2$	$2\pi_1+\pi_2$	$\pi_1+\pi_2$	π_2
$\mathcal{B}^2_{4m+1}$	π_2	$\pi_1+\pi_2$	$\pi_1+2\pi_2$	$2\pi_1+2\pi_2$	$\pi_0+2\pi_2$	$\pi_1+\pi_2$	π_1
$\mathcal{B}^1_{4m-1}$	π_1	$\pi_1+\pi_2$	$\pi_0+2\pi_2$	$2\pi_1+2\pi_2$	$\pi_1+2\pi_2$	$\pi_1+\pi_2$	π_2
$\mathcal{B}^1_{4m}$	π_2	$\pi_1+\pi_2$	$2\pi_1+\pi_2$	$\pi_0+3\pi_2$	$2\pi_1+\pi_2$	$\pi_1+\pi_2$	π_2
$\mathcal{B}^1_{4m+1}$	π_2	$2\pi_2$	$2\pi_1+\pi_2$	$2\pi_1+2\pi_2$	$\pi_0+2\pi_2$	$2\pi_1$	π_2
$\mathcal{B}^1_{4m+2}$	π_2	$2\pi_2$	$\pi_1+2\pi_2$	$2\pi_1+2\pi_2$	$2\pi_1+\pi_2$	$\pi_0+\pi_2$	π_1
$\mathcal{B}^0_{4m}$	π_1	$\pi_1+\pi_2$	$\pi_1+2\pi_2$	$\pi_0+3\pi_2$	$\pi_1+2\pi_2$	$\pi_1+\pi_2$	π_1
$\mathcal{B}^0_{4m+1}$	π_2	$\pi_1+\pi_2$	$\pi_1+2\pi_2$	$2\pi_1+2\pi_2$	$\pi_0+2\pi_2$	$\pi_1+\pi_2$	π_1
$\mathcal{B}^0_{4m+2}$	π_2	$2\pi_2$	$\pi_1+2\pi_2$	$2\pi_1+2\pi_2$	$2\pi_1+\pi_2$	$\pi_0+\pi_2$	π_1
$\mathcal{B}^0_{4m+3}$	π_2	$2\pi_2$	$3\pi_2$	$2\pi_1+2\pi_2$	$2\pi_1+\pi_2$	$2\pi_1$	π_0

But there are twice as many trace subsets to count and there is less symmetry than before, so each evaluation of the likelihood involves a quadratic form of dimension 80; accordingly the estimator is even slower than that of the previous section (about ten times slower, in our poorly-optimized implementation).

We benchmark this MAP estimator for 2LSB replacement, which we call *2Couples/ML*, against the least-squares version *2Couples/LSM* [7]. We also include another detector of multiple bit plane replacement, which is a modification of the method known as *WS* and found in [17] (we will call it *2LSB WS*). Results for the corresponding experiments to those in Sect. 3 are displayed in Fig. 3.

Observe that the new estimator is many times more accurate than the others, except for very large payloads. It remains more accurate than the LSM detector for all embedding rates, but above proportionate payloads of about 80% the WS-based detector is the only one not to suffer from weak performance (this seems to be a general feature of WS detectors, and their area of strength). Clearly the method of maximum likelihood has made a vast improvement to the reliability of payload estimation, particularly in the difficult case of small payloads.

5 Maximum Likelihood and the "Triples" Structure

Finally, we sketch how the ML technique can be applied to a more intricate structural analysis of (single plane) LSB replacement, based on triplets of pixels.

Fig. 3. Above, chart displaying how the MSE (log axis) of the *2Couples/LSM* structural estimator [7], *2LSB WS* [17], and the proposed *2Couples/ML* estimator, depends on the proportionate payload p. Below, table displaying other measures of estimator accuracy: sample bias, IQR, and MAE, all displayed $\times 10^2$, for five payload sizes.

	Detector	$p = 0$	$p = 0.25$	$p = 0.5$	$p = 0.75$	$p = 1.0$
	2LSB WS	0.148	1.612	2.033	1.535	0.093
Bias	2Couples/LSM	0.806	−0.990	−2.765	−10.243	−20.521
	2Couples/ML	0.600	−0.239	−0.094	−0.374	4.806
	2LSB WS	3.466	3.912	4.129	3.658	2.991
IQR	2Couples/LSM	1.905	1.954	2.849	9.270	41.608
	2Couples/ML	0.774	1.488	1.599	2.192	41.863
	2LSB WS	2.341	2.828	3.050	2.647	2.028
MAE	2Couples/LSM	2.624	1.921	3.154	10.371	20.521
	2Couples/ML	0.601	1.134	1.116	1.456	28.516

Let $\mathcal{T}$ be the set of horizontal groups of three adjacent pixels with notation analogous to Sects. 3 and 4, and define trace subsets as follows:

$$\mathcal{D}_{m,n} = \{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n\}$$
$$\mathcal{E}_{m,n} = \{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n, \text{ with } s_j \text{ even}\}$$
$$\mathcal{O}_{m,n} = \{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n, \text{ with } s_j \text{ odd}\}$$

These capture the third-order effects of replacing LSBs; in covers we expect that $\mathcal{E}_{m,n}$ and $\mathcal{O}_{m,n}$ will be of approximately equal size, and the corresponding least-squares structural payload estimators [5] are generally (although not uniformly) more accurate than those based on Couples structure.

For an ML detector we again imagine that $\mathcal{E}_{m,n}$ and $\mathcal{O}_{m,n}$ derive binomially from $\mathcal{D}_{m,n}$. This time it is the eight trace subsets $\mathcal{O}_{2m-1,2n}$, $\mathcal{O}_{2m-1,2n+1}$, $\mathcal{E}_{2m,2n}$, $\mathcal{E}_{2m,2n+1}$, $\mathcal{O}_{2m,2n-1}$, $\mathcal{O}_{2m,2n}$, $\mathcal{E}_{2m+1,2n-1}$, $\mathcal{E}_{2m+1,2n}$ which are permuted by LSB operations – see [5] for the transition probabilities – and the analogy to Tab. 1 is displayed in part in Tab. 4. All instances of the displayed transitions make up

Table 4. Extract from table showing probabilities of transition from pre-cover trace subsets to stego object trace subset, when the Triples structure is analyzed

	$\mathcal{D}_{2m-1,\,2n}$	$\mathcal{D}_{2m-1\atop 2n+1}$	$\mathcal{D}_{2m\atop 2n-1}$	$\mathcal{D}_{2m\atop 2n}$	$\mathcal{D}_{2n\atop 2n+1}$	$\mathcal{D}_{2m+1\atop 2n-1}$	$\mathcal{D}_{2m+1\atop 2n}$
$\mathcal{O}_{2m-1,2n}$	π_0	π_1	π_1	$\pi_1+\pi_2$	π_2	π_2	π_3
$\mathcal{O}_{2m-1,2n+1}$	π_1	π_0	π_2	$\pi_1+\pi_2$	π_1	π_3	π_2
$\mathcal{E}_{2m,2n}$	π_1	π_2	π_2	$\pi_0+\pi_3$	π_1	π_1	π_2
$\mathcal{E}_{2m,2n+1}$	π_2	π_1	π_3	$\pi_1+\pi_2$	π_0	π_2	π_1
$\mathcal{O}_{2m,2n-1}$	π_1	π_2	π_0	$\pi_1+\pi_2$	π_3	π_1	π_2
$\mathcal{O}_{2m,2n}$	π_2	π_1	π_1	$\pi_0+\pi_3$	π_2	π_2	π_1
$\mathcal{E}_{2m+1,2n-1}$	π_2	π_3	π_1	$\pi_1+\pi_2$	π_2	π_0	π_1
$\mathcal{E}_{2m+1,2n}$	π_3	π_2	π_2	$\pi_1+\pi_2$	π_1	π_1	π_0

the complete table and the abbreviation is $\pi_i = \frac{1}{2}(\frac{p}{2})^i(1-\frac{p}{2})^{3-i}$. Then a MAP estimator can be derived in the same way as for Couples and 2Couples structure.

The optimization problem for this likelihood is much harder than those occurring in Sects. 3 and 4, because the number of nuisance parameters (the sizes of $\mathcal{D}_{m,n}$ in the pre-cover) is squared by the finer categorization of pixel differences. We must radically reduce the number of subsets under consideration in order to achieve a tractable optimization. For now we consider only $|m| \leq 3$ and $|n| \leq 3$, but only 36.1% of sample triplets in our test covers are found in this range so we are ignoring quite a large proportion of the evidence. Even so, this leads to 48 dimensional optimization problem, and each likelihood evaluation involves a quadratic form of length 72, so this *Triples/ML* estimator is much slower to evaluate than *Couples/ML* of Sect. 3 (over a minute per image).

Because of its complexity we have not yet conducted as full benchmarks as for *Couples/ML* and *2Couples/ML*. Some reduced experiments are reported in Fig. 4. In order to make a fair comparison with least-squares Triples [5] (denoted *Triples/LSM*) we weakened the standard estimator by similarly limiting it to trace subsets with maximum absolute index of 3. The experimental results indicate that the *Triples/ML* estimator is more accurate only for small payloads, with the critical payload being somewhere between 25% and 50%. It is not clear that the performance gain is worth the hugely inflated computational costs.

6   Conclusions and Directions for Further Research

Bringing together the statistical foundations of the maximum likelihood method with the sensitivity of structural steganalysis has been fruitful in terms of new and more accurate payload estimators for bit replacement steganography. At the moment there is a large computational cost associated with maximizing the likelihood function, but it is likely to be significantly reduced when the function is differentiated. There is no fundamental bar to this: the problem is only one

	Detector	$p = 0$	$p = 0.1$	$p = 0.25$	$p = 0.5$
Bias	Triples/LSM$^-$	-0.128	-0.478	-0.857	-0.326
	Triples/ML	0.785	-0.695	-0.828	-1.031
IQR	Triples/LSM$^-$	3.127	3.262	3.495	1.914
	Triples/ML	0.957	1.998	2.537	2.631
MAE	Triples/LSM$^-$	2.362	2.584	2.899	1.522
	Triples/ML	0.785	1.750	1.990	2.027
MSE	Triples/LSM$^-$	0.153	0.182	0.237	0.076
	Triples/ML	0.042	0.074	0.099	0.095

Fig. 4. Comparison of estimator accuracy for weakened *Triples/LSM* (denoted *Triples/LSM$^-$*) and *Triples/ML* methods, with four payload sizes

of algebraic complexity. Another possibility for reducing the complexity is to formulate a parametric model for the pixel difference parameters *d*. However, no such model will be perfect and inaccuracies modelling *d* could damage the accuracy of the ML estimator.

Regardless of computational cost, there is a certain elegance to the combined method, particularly in that it avoids the potentially extremely complex question of which equalities between trace subsets should be included in the cover model. We believe that the ML foundation provides the "ultimate" structural steganalyzers and that, once the efficiency is improved and one hole in the cover model patched, these detectors will be the last and best bit replacement detectors.

The next application of a likelihood calculation might be the development of a likelihood ratio test for the presence of hidden data (it may also be possible to develop tests for which type of bit replacement – LSB, 2LSB, etc – have been used). In some preliminary experiments we did implement such a detector but its performance was disappointing – as a classifier for the presence or absence of hidden data it was better to use simply the payload estimate. Very likely this is because of the necessity of imposing a prior distribution, to avoid overfitting, so that the usual generalized likelihood ratio statistic is not available.

We should note, though, that there is no guarantee of complete optimality with maximum likelihood estimators, let alone the MAP estimator we have to use in practice: their unbiased and minimum variance properties only apply asymptotically. Indeed for a single image we have only one observation of the trace subset sizes which make up the multinomial likelihood, although it can be seen as many observations: the type of each sample pair or triplet. Finally, the trace subsets used here have been demonstrated highly sensitive but this does not completely preclude the possibility that better feature sets exist.

Even so, the experimental results included in this paper demonstrate that the fusion of ML methods with structural analysis has practical value as well as a well-principled derivation.

Acknowledgements

References

1. Fridrich, J., Goljan, M., Du, R.: Detecting LSB steganography in color and grayscale images. IEEE Multimedia 8, 22–28 (2001)
2. Dumitrescu, S., Wu, X., Wang, Z.: Detection of LSB steganography via sample pair analysis. IEEE Transactions on Signal Processing 51, 1995–2007 (2003)
3. Lu, P., Luo, X., Tang, Q., Shen, L.: An improved sample pairs method for detection of LSB embedding. In: Fridrich, J. (ed.) IH 2004. LNCS, vol. 3200, pp. 116–127. Springer, Heidelberg (2004)
4. Dumitrescu, S., Wu, X.: A new framework of LSB steganalysis of digital media. IEEE Transactions on Signal Processing 53, 3936–3947 (2005)
5. 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)
6. Ker, A.: Fourth-order structural steganalysis and analysis of cover assumptions. In: Security, Steganography and Watermarking of Multimedia Contents VIII. Proc. SPIE, vol. 6072, pp. 25–38 (2006)
7. Ker, A.: Steganalysis of embedding in two least significant bits. IEEE Transactions on Information Forensics and Security 2, 46–54 (2007)
8. Dabeer, O., Sullivan, K., Madhow, U., Chandrasekaran, S., Manjunath, B.: Detection of hiding in the least significant bit. IEEE Transactions on Signal Processing 52, 3046–3058 (2004)
9. Hogan, M., Hurley, N., Silvestre, G., Balado, F., Whelan, K.: ML detection of steganography. In: Security, Steganography and Watermarking of Multimedia Contents VII. Proc. SPIE, vol. 5681, pp. 16–27 (2005)
10. Draper, S., Ishwar, P., Molnar, D., Prabhakaran, V., Ramchandran, K., Schonberg, D., Wagner, D.: An analysis of empirical PMF based tests for least significant bit image steganography. In: Barni, M., Herrera-Joancomartí, J., Katzenbeisser, S., Pérez-González, F. (eds.) IH 2005. LNCS, vol. 3727, pp. 327–341. Springer, Heidelberg (2005)
11. Sullivan, K., Madhow, U., Chandrasekaran, S., Manjunath, B.: Steganalysis for Markov cover data with applications to images. IEEE Transactions on Information Forensics and Security 1, 275–287 (2006)
12. Ker, A.: A fusion of maximum likelihood and structural steganalysis, extended technical report in preparation (2007)
13. Ker, A.: Derivation of error distribution in least-squares steganalysis. IEEE Transactions on Information Forensics and Security 2, 140–148 (2007)
14. Fridrich, J., Soukal, D.: Matrix embedding for large payloads. In: Security, Steganography and Watermarking of Multimedia Contents VIII. Proc. SPIE, vol. 6072 (2006)
15. Nelder, J., Mead, R.: A simplex algorithm for function minimization. Computer Journal 7, 308–313 (1965)
16. Ker, A.: Optimally weighted least-squares steganalysis. In: Security, Steganography and Watermarking of Multimedia Contents IX. Proc. SPIE, vol. 6505 (2007)
17. Yu, X., Tan, T., Wang, Y.: Extended optimization method of LSB steganalysis. In: Proc. IEEE International Conference on Image Processing, vol. 2, pp. 1102–1105 (2005)