Optimally Weighted Least-Squares Steganalysis

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England

ABSTRACT

Quantitative steganalysis aims to estimate the amount of payload in a stego object, and such estimators seem to arise naturally in steganalysis of Least Significant Bit (LSB) replacement in digital images. However, as with all steganalysis, the estimators are subject to errors, and their magnitude seems heavily dependent on properties of the cover. In very recent work we have given the first *derivation* of estimation error, for a certain method of steganalysis (the Least-Squares variant of Sample Pairs Analysis) of LSB replacement steganography in digital images. In this paper we make use of our theoretical results to find an improved estimator and detector. We also extend the theoretical analysis to another (more accurate) steganalysis estimator (Triples Analysis) and hence derive an improved version of that estimator too. Experimental results show that the new steganalyzers have improved accuracy, particularly in the difficult case of never-compressed covers.

Keywords: Steganography, Structural Steganalysis, Error Distribution

1. INTRODUCTION

Many steganalysis methods do more than simply diagnose the presence or absence of hidden data: they form an estimate for the size of embedded payload. Such *quantitative* steganalysis seems to present itself naturally in certain frameworks, including the leading class[1] of detectors for embedding by bit replacement in images. But the estimates are subject to errors, whose nature, for some particular cases, have been investigated empirically.[2–4] The empirical results show that the magnitude of the errors varies hugely, and is influenced by properties of the cover object used for embedding.

Suppose, then, that a steganalyst has developed a quantitative method. In a particular instance, how accurate is the payload estimate, and how much confidence should the steganalyst have in the result? This question goes to the heart of steganalysis, and it seems clear that a measure of steganalysis confidence could enhance the reliability of detectors. In very recent work[5] we addressed this question head-on and, for one particular payload estimator – the *Least-Squares Method* (LSM) variant[6] of the *Sample Pairs Analysis* detector[7*] – applied to one particular embedding method (simple LSB replacement in images), were able to derive the distribution of the estimator in the restricted case when no payload was in fact embedded. It was suggested that the theoretical results could lead to development of better detection methods, and here we will follow up that comment by constructing improved steganalysis estimators.

The aim of this paper is twofold: to use the results of Ref. 5 to improve the steganalyzer (by reducing its bias and variance), and to extend the results to a newer, related, estimator called *Triples*.[1] Benchmarking of the various steganalyzers is an important part of this work and we will include a substantial suite of experiments to demonstrate the extent of the improvement. It will be seen that the improved estimators and detectors are generally more accurate, and they are particularly successful in the most difficult case (never-compressed bitmap covers) for which other work on LSB replacement steganalysis[1,8] struggled to make much headway.

The rest of this paper is presented as follows. In Sect. 2 we sketch the Couples/LSM estimator and then summarise the theoretical results of Ref. 5, which predict its error distribution in the case when no payload is present. In Sect. 3 we suggest a modified estimator which introduces weighting into the least-squares computation and derive optimal weights according to the theory of Sect. 2. New theory is sketched out in Sect. 4, which extends the results of Ref. 5 to the Triples/LSM estimator, and a similar optimal weighting is derived. Section 5 includes a comprehensive experimental survey to demonstrate the extent to which the new weighted estimators, and other estimators arising naturally out of the error-derivation work, are superior. Finally we draw conclusions in Sect. 6.

Further author information: E-mail: adk@comlab.ox.ac.uk, Telephone: +44 1865 283530, Fax: +44 1865 276790

*In Ref. 5 and this paper we use the almost-equivalent detector we have called *Couples*, which avoids some special-case behaviour without affecting performance. See Ref. 1.

2. ERROR DISTRIBUTION OF COUPLES/LSM STEGANALYSIS

We present only an outline of the Couples/LSM estimator, including the principles that drive it but not going so far as to repeat an explicit formula for the estimator itself (a clear exposition can be found in Ref. 9). We will include just enough of the theoretical analysis of Ref. 5 for our subsequent application and extension.

2.1. The Couples/LSM Estimator

Suppose that a digital image consists of a series of N samples with values $s_1, s_2, \ldots, s_N$ in the range $0 \ldots 2M+1$ (typically $M = 127$). A *sample pair* is a pair of sample locations (j, k) for some $1 \leq j \neq k \leq N$. Let $\mathcal{P}$ be a set of sample pairs; we will use the set of all pairs that come from horizontally adjacent pixels. We then count how many sample pairs, in a fixed cover image, lie in certain *trace subsets*:

$$\begin{aligned} e_m &= |\{(j,k) \in \mathcal{P} \mid s_k = s_j + m, \text{ with } s_j \text{ even}\}| \\ o_m &= |\{(j,k) \in \mathcal{P} \mid s_k = s_j + m, \text{ with } s_j \text{ odd}\}| \\ d_m &= |\{(j,k) \in \mathcal{P} \mid s_k = s_j + m\}| \end{aligned}$$

for $-2M+1 \leq m \leq 2M+1$. We also count these quantities in the same image after a payload has been embedded by LSB replacement, and call the counts e'_m, o'_m, and d'_m respectively. The key to *structural steganalysis*[1] is to relate the cover and stego counts via the size of embedded payload.

We suppose that embedding a payload flips the least significant bit of each sample in each pair, independently, with probability $\frac{p}{2}$, for example when a payload of length pN is embedded using the standard form of LSB replacement that spreads the payload pseudorandomly throughout the cover. Structural detectors consider the effect of least significant bit flipping on pairs in e_m and o_m, deriving:

$$\begin{pmatrix} e'_{2m} \\ o'_{2m-1} \\ e'_{2m+1} \\ o'_{2m} \end{pmatrix} \approx \begin{pmatrix} (1-\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2}) & (\frac{p}{2})^2 \\ \frac{p}{2}(1-\frac{p}{2}) & (1-\frac{p}{2})^2 & (\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2}) \\ \frac{p}{2}(1-\frac{p}{2}) & (\frac{p}{2})^2 & (1-\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2}) \\ (\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2}) & (1-\frac{p}{2})^2 \end{pmatrix} \begin{pmatrix} e_{2m} \\ o_{2m-1} \\ e_{2m+1} \\ o_{2m} \end{pmatrix}. \tag{1}$$

We will not repeat the derivation because it can be found in many places in the literature, including Refs. 1 and 9 or (using the original definition of trace subsets, with a slightly more complex derivation) Refs. 6 and 7. The equation is approximate because the counts e'_m, o'_m depend on the content of the payload; if the payload is random or randomly placed then the equation refers to their expectations and the Law of Large Numbers tells us that it is exact asymptotically as the size of cover and payload tends (in fixed ratio) to infinity.

Inverting (1) is possible as long as $p \neq 1$: the inverse system is

$$\begin{pmatrix} e_{2m} \\ o_{2m-1} \\ e_{2m+1} \\ o_{2m} \end{pmatrix} \approx \frac{1}{(1-p)^2} \begin{pmatrix} (1-\frac{p}{2})^2 & -\frac{p}{2}(1-\frac{p}{2}) & -\frac{p}{2}(1-\frac{p}{2}) & (\frac{p}{2})^2 \\ -\frac{p}{2}(1-\frac{p}{2}) & (1-\frac{p}{2})^2 & (\frac{p}{2})^2 & -\frac{p}{2}(1-\frac{p}{2}) \\ -\frac{p}{2}(1-\frac{p}{2}) & (\frac{p}{2})^2 & (1-\frac{p}{2})^2 & -\frac{p}{2}(1-\frac{p}{2}) \\ (\frac{p}{2})^2 & -\frac{p}{2}(1-\frac{p}{2}) & -\frac{p}{2}(1-\frac{p}{2}) & (1-\frac{p}{2})^2 \end{pmatrix} \begin{pmatrix} e'_{2m} \\ o'_{2m-1} \\ e'_{2m+1} \\ o'_{2m} \end{pmatrix}. \tag{2}$$

Equation (2) holds for each m.

To construct the detector we need an assumption about cover objects. In LSB replacement detectors based on the structure of sample pairs this is

$$e_{2m+1} \approx o_{2m+1} \text{ for each } m; \tag{3}$$

such approximate equalities have been called *symmetries*[8] and they are well-justified for continuous-tone natural images. Putting together (3) with the relevant elements of (2) gives the following approximate equation for p which involves only observations of the stego image:

$$0 \approx e_{2m+1} - o_{2m+1} \approx \frac{1}{(1-p)^2}(s'_m + t'_m p + u'_m p^2) \tag{4}$$

where

$$\begin{aligned} s'_m &= e'_{2m+1} - o'_{2m+1} \\ t'_m &= \tfrac{1}{2}(d'_{2m+2} - d'_{2m}) - (e'_{2m+1} - o'_{2m+1}) \\ u'_m &= \tfrac{1}{4}(d'_{2m} - d'_{2m+2} + o'_{2m-1} - e'_{2m+3} + e'_{2m+1} - o'_{2m+1}) \end{aligned}$$

There are many such equations, for one each $-M \leq m \leq M$. The principle of least-squares steganalysis, introduced in Ref. 6, is to find the value of $p = \hat{p}$ which implies that all the approximately-zero quantities (4) are as close as possible to zero by minimizing their sum-square. The calculations for finding such an estimate $\hat{p}$ (*mutatis mutandis* to account for our modified definition of trace subsets, and different notation) can be found in Ref. 6 and we will not need to repeat it here. We will refer to this estimator as *Couples/LSM*, to emphasise its dependence on finding a least-square cover model fit, and its use of pairs of pixels in the analysis of bit replacement structure.

Although some limited experiments in Ref. 6 support a claim that the least-squares method improves the accuracy of payload estimation (over standard SPA,[7] which effectively sums the equations (4)), we will see in Sect. 5 that this is not universally (or even often) the case. Regardless of its merits as an improvement on standard SPA, the Couples/LSM estimator has the advantage that its error distribution is amenable to the analysis of Ref. 5, which we now sketch.

2.2. Derivation of Error Distribution

Reference 4 points out that there are generally two approximations made in derivation of a quantitative steganalysis estimator, leading to two sources of error. Assumptions about cover images (here (3)) cause error which is dependent only on the cover, called *between-image* error. Assumptions about the payload and the embedding process (here the approximation (1)) cause error which may be influenced by the cover but depend primarily on the payload, called *within-image* error. Within-image error is closely linked with payload size and for small payloads is practically zero; it is demonstrated in Ref. 4 that, for some quantitative steganalysis methods including Couples/LSM, only for very large payloads does within-image error become a serious consideration.

In Ref. 5 we are only able to consider between-image error, deriving the distribution of the estimator in images that have no payload. It was noted that this is at least sufficient for the computation of a *p-value* for the presence of payload, and it does address the more significant source of error. Nonetheless, it would be valuable in future work to extend the analysis to both sources of error and images with any payload size.

When there is no payload, all estimation error is due to (3). We need to quantify deviations from exact equality, and Ref. 5 proposes a simple model in which each sample pair (j, k) with $s_k = s_j + m$ (i.e. that counts towards d_m) has s_j even, or odd, independently and equiprobably. That is, we assume that the d_m are not random but the e_m (and hence o_m) become binomial random variables. For large enough d_m this model implies

$$e_m - o_m \dot{\sim} \mathrm{N}(0, d_m). \tag{5}$$

In Ref. 5 this model is investigated and found good at least for never-compressed images if $|m| > 3$ and marginal for $|m| = 3$, but not accurate for $|m| < 3$ or for JPEG-compressed covers. It was suggested that the Couples/LSM detector could be modified to remove any dependence on this model for $|m| \leq 3$, and we will return to this possibility later. The theoretical results should *not* be assumed accurate for covers that have been subject to JPEG compression, a process which certainly damages their parity structure.

We will identify the relevant deviations from (3) by writing $\varepsilon_m = e_{2m+1} - o_{2m+1}$ and, with a caveat regarding $m \in \{-2, -1, 0, 1\}$, we have that $\varepsilon_m \sim \mathrm{N}(0, d_{2m+1})$ are independent random variables.

We now return to the Couples/LSM estimator. When there is no payload, each $e'_m = e_m$ and $o'_m = o_m$ (these are exact equalities of course) so the formula for $\hat{p}$ becomes

$$\hat{p} = \operatorname*{argmin}_p \sum_{m=-M}^{M} \left(\frac{1}{(1-p)^2}(s'_m + t'_m p + u'_m p^2) \right)^2 \tag{6}$$

where

$$\begin{array}{rcl} s'_m &=& e_{2m+1} - o_{2m+1} \\ t'_m &=& \frac{1}{2}(d_{2m+2} - d_{2m}) - (e_{2m+1} - o_{2m+1}) \\ u'_m &=& \frac{1}{4}(d_{2m} - d_{2m+2} + o_{2m-1} - e_{2m+3} + e_{2m+1} - o_{2m+1}) \end{array}$$

We write $\boldsymbol{s'}$ (respectively $\boldsymbol{t'}$, $\boldsymbol{u'}$, $\boldsymbol{\varepsilon}$) for vectors whose entries are each s'_m (t'_m, u'_m, ε_m) for $-M \le m \le M$. Let us decompose $\boldsymbol{s'}$, etc, each into two components $\boldsymbol{s'} = \boldsymbol{s} + \bar{\boldsymbol{s}}$, etc, thus separating out the supposedly nonrandom d_m from the influence of the random deviations ε_m. We have

$$\begin{array}{rclcrcl} \boldsymbol{s} &=& \boldsymbol{0} & & \bar{\boldsymbol{s}} &=& \boldsymbol{\varepsilon} \\ t_m &=& \frac{1}{2}(d_{2m} - d_{2m+2}) & & \bar{\boldsymbol{t}} &=& -\boldsymbol{\varepsilon} \\ u_m &=& \frac{1}{4}(d_{2m} - d_{2m+2}) + \frac{1}{8}(d_{2m-1} - d_{2m+3}) & & \bar{u}_m &=& \varepsilon_m - \frac{1}{8}(\varepsilon_{m-1} + \varepsilon_{m+1}) \end{array} \tag{7}$$

Finally, we note that (6) has a natural geometric interpretation:

$$\hat{p} = \operatorname*{argmin}_p \left\| \frac{(\boldsymbol{s} + \bar{\boldsymbol{s}}) + (\boldsymbol{t} + \bar{\boldsymbol{t}})p + (\boldsymbol{u} + \bar{\boldsymbol{u}})p^2}{(1-p)^2} \right\| \tag{8}$$

where $\| \cdot \|$ represents the L^2-norm, so that $\hat{p}$ is the parameter p that places the path $\boldsymbol{r'} = \frac{(\boldsymbol{s}+\bar{\boldsymbol{s}})+(\boldsymbol{t}+\bar{\boldsymbol{t}})p+(\boldsymbol{u}+\bar{\boldsymbol{u}})p^2}{(1-p)^2}$ closest to the origin. Note that this path can be considered a randomly perturbed version of $\boldsymbol{r} = \frac{\boldsymbol{s}+\boldsymbol{t}p+\boldsymbol{u}p^2}{(1-p)^2}$ and the latter passes exactly through the origin at $p = 0$. Reference 5 includes a geometric argument (which we will not repeat here, partly because we are going to prove something more general in Sect. 4) to show that the value of $\hat{p}$ can be approximated by

$$\hat{p} \approx -\frac{\bar{\boldsymbol{s}}.\boldsymbol{t}}{\boldsymbol{t}.\boldsymbol{t}} + 2\frac{\big((\bar{\boldsymbol{t}} + 2\bar{\boldsymbol{s}}).\boldsymbol{t}\big)(\bar{\boldsymbol{s}}.\boldsymbol{t})}{(\boldsymbol{t}.\boldsymbol{t})^2} - \frac{\bar{\boldsymbol{s}}.(\bar{\boldsymbol{t}} + 2\bar{\boldsymbol{s}})}{\boldsymbol{t}.\boldsymbol{t}}$$

(“.” represents the scalar product) which, in our case, gives

$$\hat{p} \approx -\frac{\boldsymbol{\varepsilon}.\boldsymbol{t}}{\boldsymbol{t}.\boldsymbol{t}} + 2\frac{(\boldsymbol{\varepsilon}.\boldsymbol{t})^2}{(\boldsymbol{t}.\boldsymbol{t})^2} - \frac{\boldsymbol{\varepsilon}.\boldsymbol{\varepsilon}}{\boldsymbol{t}.\boldsymbol{t}}. \tag{9}$$

The first term of (9) is a linear combination of Gaussian random variables with mean zero, and hence is itself Gaussian with mean zero. The second and third terms, it is argued in Ref. 5, contribute little to the shape of the distribution and their primary significance is to shift the mean. Using $\mathrm{E}[\varepsilon_m] = 0$, $\mathrm{E}[\varepsilon_m^2] = \mathrm{Var}[\varepsilon_m] = d_{2m+1}$, and independence of the ε_m we therefore derive the following approximation to the distribution of the steganalysis estimator when the true value of p is zero:

$$\begin{aligned} \hat{p} \approx \mathrm{N}\big(\mu(\boldsymbol{d}), v(\boldsymbol{d})\big), \text{ where } v(\boldsymbol{d}) &= \frac{4\sum_m (d_{2m+2} - d_{2m})^2 d_{2m+1}}{\big(\sum_m (d_{2m+2} - d_{2m})^2\big)^2} \\ \mu(\boldsymbol{d}) &= 2v(\boldsymbol{d}) - \frac{4\sum_m d_{2m+1}}{\sum_m (d_{2m+2} - d_{2m})^2}. \end{aligned} \tag{10}$$

We emphasise that this distribution is image-specific: it does not tell us the distribution of the estimates as we vary over different images because each image can have a different value for $\mu(\boldsymbol{d})$ and $v(\boldsymbol{d})$. For this reason we refer to $\mu(\boldsymbol{d})$ as *image-specific bias* and $v(\boldsymbol{d})$ as *image-specific variance*.

The accuracy of the approximate distribution is verified in Ref. 5. It turns out to be highly accurate only if the components $m = -1, 0$ (and sometimes also $m = -2, 1$) are excluded from the calculation of the Couples/LSM estimator because, for these values of m, Eq. (5) is inaccurate. It would be a shame to have to make this exclusion, because it forces us to ignore pixels in the stego image which are close in value to adjacent pixels, and this can be quite a large proportion (in Ref. 5 it is stated that, in typical cover images, excluding the components $m = -2, -1, 0, 1$ means ignoring on average half of the pixels in each cover). In most cases we will

not exclude these components, so the theoretical error prediction will be imperfect, but it is adequate to obtain better estimators.

Finally, it can be demonstrated that the value of $\mu(\boldsymbol{d})$ is usually quite close to zero, in comparison with the standard deviation, in digital images. Therefore our estimate of the true value of p (zero) by $\hat{p}$ is almost unbiased. But in the novel estimators described in this paper we will sometimes have to take account of the bias.

3. OPTIMAL WEIGHTING

We now make use of the theory to describe a modification to the estimator, in which a *weighted* sum of the squared deviations (4) is minimized. We propose to find the value $\hat{p}_w$ of p which minimizes

$$\sum_m w_m \Big(\tfrac{s'_m + t'_m p + u'_m p^2}{(1-p)^2} \Big)^2$$

where the w_m are the weighting components. The analogy to (8) is

$$\hat{p}_w = \operatorname*{argmin}_p \left\| \frac{\boldsymbol{s'_w} + \boldsymbol{t'_w} p + \boldsymbol{u'_w} p^2}{(1-p)^2} \right\|$$

where $\boldsymbol{s'_w}$ has components $\sqrt{w_m} s'_m$, and similarly for $\boldsymbol{t'_w}$ and $\boldsymbol{u'_w}$. Using the results of the previous section, we now have

$$\begin{aligned} \hat{p}_w \approx \mathrm{N}\big(\mu(\boldsymbol{d},\boldsymbol{w}), v(\boldsymbol{d},\boldsymbol{w})\big), \text{ where } v(\boldsymbol{d},\boldsymbol{w}) &= \frac{4\sum_m w_m^2 (d_{2m+2} - d_{2m})^2 d_{2m+1}}{\big(\sum_m w_m (d_{2m+2} - d_{2m})^2\big)^2} \\ \mu(\boldsymbol{d},\boldsymbol{w}) &= 2v(\boldsymbol{d},\boldsymbol{w}) - \frac{4\sum_m w_m d_{2m+1}}{\sum_m w_m (d_{2m+2} - d_{2m})^2}. \end{aligned} \tag{11}$$

We seek a weight vector $\boldsymbol{w}$ to minimize the variance of the estimator. It is derived from the following simple result:

LEMMA 3.1. *If all $a_i \geq 0$ and all $b_i > 0$ then the quantity*

$$v = \frac{\sum w_i^2 a_i b_i}{\big(\sum w_i a_i\big)^2}$$

is minimized when all $w_i \propto b_i^{-1}$.

Proof. We compute

$$\frac{\partial v}{\partial w_j} = \frac{2 w_j a_j b_j \sum w_i a_i - 2 a_j \sum w_i^2 a_i b_i}{\big(\sum w_i a_i\big)^3}$$

which is zero if

$$w_j b_j = \frac{\sum w_i^2 a_i b_i}{\sum w_i a_i},$$

which is constant in j. (One may proceed to compute the Hessian, to check that the stationary point is indeed a maximum, but we omit to do so here. An alternative proof is possible using Lagrange multipliers.) □

This immediately implies that we should take

$$w_m = \frac{1}{d_{2m+1}}. \tag{12}$$

in order to minimize $v(\boldsymbol{d},\boldsymbol{w})$.

There are two caveats to the optimality of the weighting. First, we have already said that the cover model (5) is not accurate for m close to zero. Either we must exclude some components from the calculation of the

estimate (effectively forcing $w_m = 0$ for a few values of m close to zero), or accept that applying the theory will predict suboptimal weightings for these few components. In Sect. 5 we will see that, slightly suboptimal or not, we can use the weightings (12) for all components and still see a performance gain.

Second, note that the theory outlined in the previous section is *only* accurate for the distribution of $\hat{p}$ when $p = 0$ (we have not yet been able fully to generalise the theory to work for all p) and by continuity we expect that the weights are close to optimal when p is small. But our detector may not be optimally-weighted when the true value of p is large. An additional problem here is that $\boldsymbol{d}$ is a property of the *cover* image, and of course when payload is embedded we do not know the cover image: one solution is to estimate the values of d_m from d'_m given an initial estimate of p, but we will not pursue that idea in this paper. Instead, we will simply use the observed $\boldsymbol{d}$ as an estimate of the same quantity in the cover, and again this will only be a good estimate as long as the true value of p is small. Since LSM estimators seem always to suffer from poor performance for large values of p anyway (for one explanation why, see Ref. 9), this will not concern us.

Despite these reservations, we expect that $\hat{p}_w$, which we call the *Couples Weighted Least-Squares Method* (or Couples/WLSM) estimator, will be a more accurate estimator of p than the unweighted version. We will see, in Sect. 5, that this is true to the extent that the estimator variance is reduced. But there is a price to pay when introducing weighting: it turns out that the weighted estimator bias $\mu(\boldsymbol{d}, \boldsymbol{w})$ is often substantially larger than the unweighted bias $\mu(\boldsymbol{d})$ (one cannot minimize both bias and variance with the same weights). However there is a solution to this conundrum, because we can introduce a *bias corrected* estimator $\hat{p}_w - \mu(\boldsymbol{d}, \boldsymbol{w})$. This solution is not perfect because the same caveats apply to the theoretically-predicted bias as do to the variance: a) the theory is only correct when the true value of p is zero, and b) the flaw in the cover deviation model (5) will require us to modify the detector, weakening its power, if we want the bias computation to be exact. It turns out that bias correction is necessary if the weighted estimator is to be used, and that it works well for reasonably small true values of p even given the caveats.

Finally, there is one further way in which we could aim to improve the reliability of the Couples/WLSM method. Although weighting reduces the image-specific variance, some images still have a high value of $v(\boldsymbol{d}, \boldsymbol{w})$ and these are the prime candidates to give outliers (large errors) in their estimate. In particular, when we consider the problem of discrimination between covers and stego objects (as opposed to payload estimation), we would like to lend much less significance to images with a high value of $v(\boldsymbol{d}, \boldsymbol{w})$. A solution is to introduce the *standardized* statistic $(\hat{p}_w - \mu(\boldsymbol{d}, \boldsymbol{w}))/\sqrt{v(\boldsymbol{d}, \boldsymbol{w})}$ as a discriminator between stego and cover images (it is, of course, no longer a payload estimate). The aim of the statistic is to separate cover and stego objects with a very low rate of false positives, because outliers in the null distribution are suppressed. We will see that, in some particular circumstances, the standardized detector does succeed in this aim.

4. EXTENSION TO LEAST-SQUARES TRIPLES STEGANALYSIS

Having applied the theory of Ref. 5, outlined in Sect. 2, to produce improved estimators, we now give new theory extending the error distribution derivation to another estimator. The estimator in question is that called *Triples* in Ref. 1, but here we will call it *Triples/LSM* since it does indeed use a least-squares cover-fitting approach. We will first outline the detector; as previously, we will not go so far as to include the final formula for the estimate since it is not necessary to the error analysis and can be found in full detail in Ref. 1. Then we derive an optimally-weighted version. The mathematics is not fundamentally different to that in Sects. 2 and 3, but the algebra is rather more involved.

4.1. The Triples/LSM Estimator

The Triples/LSM estimator analyzes the structure of bit replacement in triplets of pixels. Its derivation can be presented in very similar way to the Couples/LSM estimator. Again we suppose that the image has N samples $s_1, s_2, \ldots, s_N$ in the range $0 \ldots 2M + 1$ and we consider triplets of distinct sample locations (j, k, l). Let $\mathcal{T}$ be a set of triplets: we will use all groups of three adjacent pixels in horizontal rows. As before, we count trace subsets in a fixed cover image, classifying triplets according to successive pixel value differences and the parity of the first:

$$\begin{aligned} e_{m,n} &= |\{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n,\ s_j \text{ even}\}| \\ o_{m,n} &= |\{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n,\ s_j \text{ odd}\}| \\ d_{m,n} &= |\{(j,k,l) \in \mathcal{T} \mid s_k = s_j + m, s_l = s_k + n\}| \end{aligned}$$

Similarly we count $e'_{m,n}$, etc, the number in each type of trace subset in the stego image. Potentially we could have $-M \le m, n \le M$. When both m and n are large the total difference between the first and last pixel sample in the triplet would have to exceed the possible dynamic range of the pixels, but this presents no problem: the number of such triplets is zero, and this allows us to be free with the range of m and n where it simplifies calculations.

We must relate the $e'_{m,n}$ and $o'_{m,n}$ to $e_{m,n}$ and $o_{m,n}$ via the payload size. Again assuming that embedding flips the LSB of each sample in each pair, independently, with probability $\frac{p}{2}$ (e.g. when the payload is pN) we can follow the effect of embedding on trace subsets and derive a system of linear equations. The calculations can be seen in Ref. 1 and the conclusion is:

$$\begin{pmatrix} e'_{2m,2n} \\ o'_{2m-1,2n} \\ e'_{2m+1,2n-1} \\ o'_{2m,2n-1} \\ e'_{2m,2n+1} \\ o'_{2m-1,2n+1} \\ e'_{2m+1,2n} \\ o'_{2m,2n} \end{pmatrix} \approx \begin{pmatrix} (1-\frac{p}{2})^3 & \frac{p}{2}(1-\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2})^2 & (\frac{p}{2})^2(1-\frac{p}{2}) & \cdots \\ \frac{p}{2}(1-\frac{p}{2})^2 & (1-\frac{p}{2})^3 & (\frac{p}{2})^2(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2})^2 & \\ \frac{p}{2}(1-\frac{p}{2})^2 & (\frac{p}{2})^2(1-\frac{p}{2}) & (1-\frac{p}{2})^3 & \frac{p}{2}(1-\frac{p}{2})^2 & \\ (\frac{p}{2})^2(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2})^2 & \frac{p}{2}(1-\frac{p}{2})^2 & (1-\frac{p}{2})^3 & \\ \frac{p}{2}(1-\frac{p}{2})^2 & (\frac{p}{2})^2(1-\frac{p}{2}) & (\frac{p}{2})^2(1-\frac{p}{2}) & (\frac{p}{2})^3 & \\ (\frac{p}{2})^2(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2})^2 & (\frac{p}{2})^3 & (\frac{p}{2})^2(1-\frac{p}{2}) & \\ (\frac{p}{2})^2(1-\frac{p}{2}) & (\frac{p}{2})^3 & \frac{p}{2}(1-\frac{p}{2})^2 & (\frac{p}{2})^2(1-\frac{p}{2}) & \cdots \\ (\frac{p}{2})^3 & (\frac{p}{2})^2(1-\frac{p}{2}) & (\frac{p}{2})^2(1-\frac{p}{2}) & \frac{p}{2}(1-\frac{p}{2})^2 & \end{pmatrix} \begin{pmatrix} e_{2m,2n} \\ o_{2m-1,2n} \\ e_{2m+1,2n-1} \\ o_{2m,2n-1} \\ e_{2m,2n+1} \\ o_{2m-1,2n+1} \\ e_{2m+1,2n} \\ o_{2m,2n} \end{pmatrix}.$$

For economy of space we have displayed only half of the matrix, which is rotationally symmetric. We will not display the full inverse, for the same reason, instead picking out the two components that matter to the Triples/LSM detector:

$$\begin{aligned} e_{2m+1,2n+1} &\approx \frac{1}{(1-p)^3}\Big((1-\tfrac{p}{2})^3(e'_{2m+1,2n+1}) - \tfrac{p}{2}(1-\tfrac{p}{2})^2(e'_{2m+1,2n+2} + e'_{2m,2n+2} + o'_{2m,2n+1}) \\ &\qquad + (\tfrac{p}{2})^2(1-\tfrac{p}{2})(e'_{2m,2n+3} + o'_{2m-1,2n+2} + o'_{2m,2n+2}) - (\tfrac{p}{2})^3(o'_{2m-1,2n+3})\Big) \\ o_{2m+1,2n+1} &\approx \frac{1}{(1-p)^3}\Big((1-\tfrac{p}{2})^3(o'_{2m+1,2n+1}) - \tfrac{p}{2}(1-\tfrac{p}{2})^2(o'_{2m+2,2n} + o'_{2m+1,2n} + e'_{2m+2,2n+1}) \\ &\qquad + (\tfrac{p}{2})^2(1-\tfrac{p}{2})(o'_{2m+2,2n-1} + e'_{2m+2,2n} + e'_{2m+3,2n}) - (\tfrac{p}{2})^3(e'_{2m+3,2n-1})\Big) \end{aligned}$$

Again we need a property of cover images, from which to derive an estimate for p. This can be a little more complicated than the Couples case because there are a number of plausible symmetries, but it is sufficient to pick just one class of symmetry which looks very similar to (3):

$$e_{2m+1,2n+1} \approx o_{2m+1,2n+1} \text{ for each } m \text{ and } n. \tag{13}$$

Other symmetries, including $e_{2m+1,2n} \approx o_{2m+1,2n}$ and $e_{2m+1,2n+1} \approx e_{2n+1,2m+1}$ are possible but their inclusion makes almost no difference to the performance of the estimator, and complicates the analysis greatly, so we will not include them. For a thorough investigation of the difficulties presented by multiple symmetries, see Ref. 8.

We are now in a position to state an equation estimating the payload p, analogous to (4). There is one equation for each m and n:

$$0 \approx e_{2m+1,2n+1} - o_{2m+1,2n+1} \approx \frac{1}{(1-p)^3}(s'_{m,n} + t'_{m,n}p + u'_{m,n}p^2 + v'_{m,n}p^3) \tag{14}$$

where

$$\begin{aligned} s'_{m,n} &= e'_{2m+1,2n+1} - o'_{2m+1,2n+1} \\ t'_{m,n} &= \tfrac{1}{2}(o'_{2m+2,2n} + o'_{2m+1,2n} + e'_{2m+2,2n+1} - e'_{2m+1,2n+2} - e'_{2m,2n+2} - o'_{2m,2n+1}) - \tfrac{3}{2}(e'_{2m+1} - o'_{2m+1}) \end{aligned}$$

Since we will not need to use $u'_{m,n}$ or $v'_{m,n}$ in any of the subsequent analysis (although of course they are needed in order to compute the estimate) we will not display their long formulae.

The traditional Triples/LSM estimator finds an estimate $\hat{p}$ to minimize the sum-square of all of the approximately zero quantities (14) and, as shown in Ref. 1 (verified again here in Sect. 5), it is almost always superior to both the Couples/LSM estimator and other traditional steganalysis methods.

4.2. Weighted Least-Squares and Derivation of Error Distribution

We now derive the error distribution of the Triples/LSM estimator, again only for the case when truly no data is embedded. Our presentation will differ from that of Sect. 2, though, because we will immediately depart from Ref. 1 by introducing weighting into the minimized sum-square:

$$\hat{p}_w = \operatorname*{argmin}_p \sum_{m,n=-M}^{M} w_{m,n}\, a_{m,n}^2$$

where $a_{m,n}$ represents the right of (14). Thus we will avoid having to repeat ourselves by giving calculations for both unweighted and weighted estimators.

Once again we need to model deviations from exact equality in (13) and we will use the same idea: we assume that the $d_{m,n}$ are nonrandom and that $e_{m,n}$ (which then determines $o_{m,n}$) are independent binomial random variables; this leads to

$$e_{m,n} - o_{m,n} \dot{\sim} \mathrm{N}(0, d_{m,n}). \tag{15}$$

Once again there are some situations in which this model is not accurate. Recall that (5) was not necessarily a good model for values of m very close to zero; the same problem occurs with (15) but it is harder to identify the components where there are problems. There seem to be a few more cases where the distribution appears to have heavier tails than Gaussian. Furthermore, the assumption that all deviations (15) are independent is also less accurate than for the Couples case.

We will not allow ourselves to be sidetracked into investigation of this problem. Instead, we will simply admit, as in the Couples case, that the model (15) is imperfect, and the consequent weighted least-squares detector will not be quite optimally weighted. We will use it anyway in the hope that our weighting is better than none at all; the experimental results of Sect. 5 will show that we have created an improved estimator in spite of the dubious accuracy of (15) in a few cases.

When no payload is embedded the equation for the weighted estimate becomes

$$\hat{p}_w = \operatorname*{argmin}_p \sum_{m,n=-M}^{M} \left(\frac{1}{(1-p)^3} (s'_{m,n} + t'_{m,n} p + u'_{m,n} p^2 + v'_{m,n} p^3) \right)^2 \tag{16}$$

where

$$s'_{m,n} = \sqrt{w_{m,n}} (e_{2m+1,2n+1} - o_{2m+1,2n+1})$$

$$\begin{aligned} t'_{m,n} = {} & \frac{\sqrt{w_{m,n}}}{2} \left(o_{2m+2,2n} + o_{2m+1,2n} + e_{2m+2,2n+1} - e_{2m+1,2n+2} - e_{2m,2n+2} - o_{2m,2n+1} \right) \\ & - \frac{3\sqrt{w_{m,n}}}{2} \left(e_{2m+1,2n+1} - o_{2m+1,2n+1} \right) \end{aligned}$$

and we need not consider $u'_{m,n}$ or $v'_{m,n}$. We will write $\boldsymbol{s'}$ for the vector of all $s'_{m,n}$ (in some order, it does not matter which) and similarly $\boldsymbol{t'}$, $\boldsymbol{u'}$, $\boldsymbol{v'}$. Analogously to (7) we separate $\boldsymbol{s'}$ into a nonrandom part $\boldsymbol{s}$, and a part $\bar{\boldsymbol{s}}$ that depends on the deviations from (13). Similarly for $\boldsymbol{t'}$ (but $\boldsymbol{u'}$ and $\boldsymbol{v'}$ we need not display here).

$$\begin{aligned} \boldsymbol{s} &= \boldsymbol{0} & \qquad \bar{s}_{m,n} &= \varepsilon_{m,n} \\ t_{m,n} &= \tfrac{\sqrt{w_{m,n}}}{4} \left(d_{2m+2,2n} + d_{2m+1,2n} + d_{2m+2,2n+1} - d_{2m+1,2n+2} - d_{2m,2n+2} - d_{2m,2n+1} \right) & \qquad \bar{t}_{m,n} &= \delta_{m,n} - \tfrac{3}{2}\varepsilon_{m,n} \end{aligned}$$

where $\varepsilon_{m,n} = \sqrt{w_{m,n}} (e_{2m+1,2n+1} - o_{2m+1,2n+1}) \sim \mathrm{N}(0, w_{m,n} d_{2m+1,2n+1})$ and $\delta_{m,n}$ is some Gaussian random variable with mean zero and is independent of all $\varepsilon_{m,n}$ (the latter because $\delta_{m,n}$ consists of linear combinations of $e_{j,k} - o_{j,k}$ never including both j and k odd).

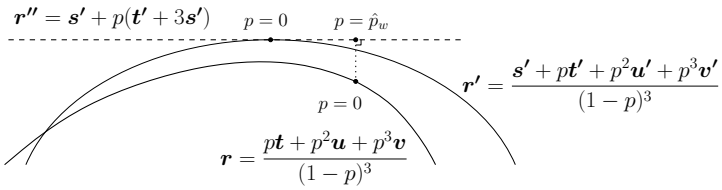


Figure 1. The effect of a small perturbation on a path of the form $\boldsymbol{r} = \frac{\boldsymbol{s}+\boldsymbol{t}p+\boldsymbol{u}p^2+\boldsymbol{v}p^3}{(1-p)^3}$ with $\boldsymbol{s} = \boldsymbol{0}$.

Just as in the Couples/LSM case, we give (16) a geometric interpretation:

$$\hat{p}_w = \operatorname*{argmin}_p \left\| \frac{(\boldsymbol{s}+\bar{\boldsymbol{s}}) + (\boldsymbol{t}+\bar{\boldsymbol{t}})p + (\boldsymbol{u}+\bar{\boldsymbol{u}})p^2 + (\boldsymbol{v}+\bar{\boldsymbol{v}})p^3}{(1-p)^3} \right\|$$

so that $\hat{p}_w$ is the parameter p which places the path $\boldsymbol{r'} = \frac{(\boldsymbol{s}+\bar{\boldsymbol{s}})+(\boldsymbol{t}+\bar{\boldsymbol{t}})p+(\boldsymbol{u}+\bar{\boldsymbol{u}})p^2+(\boldsymbol{v}+\bar{\boldsymbol{v}})p^3}{(1-p)^3}$ (let us call this path P') closest to the origin. This path can be considered a randomly perturbed version of P, given by $\boldsymbol{r} = \frac{\boldsymbol{s}+\boldsymbol{t}p+\boldsymbol{u}p^2+\boldsymbol{v}p^3}{(1-p)^3}$ and the latter passes exactly through the origin at $p = 0$ because $\boldsymbol{s} = \boldsymbol{0}$.

The geometric approximation, which extends that in Ref. 5, is to replace P' by its tangent at $p = 0$, which passes through $\boldsymbol{s'}$ and has direction vector $\frac{\mathrm{d}\boldsymbol{r'}}{\mathrm{d}p}|_{p=0} = \boldsymbol{t'} + 3\boldsymbol{s'}$. It is easy to say when a straight line passes closest to the origin: its position must be orthogonal to its direction vector (see Fig. 1), so we have $(\boldsymbol{s'} + \hat{p}_w(\boldsymbol{t'} + 3\boldsymbol{s'})).(\boldsymbol{t'} + 3\boldsymbol{s'}) = 0$, which occurs when

$$\hat{p}_w = -\frac{\boldsymbol{s'}.(\boldsymbol{t'} + 3\boldsymbol{s'})}{(\boldsymbol{t'} + 3\boldsymbol{s'}).(\boldsymbol{t'} + 3\boldsymbol{s'})}.$$

Note that the approximation of P' by a straight line has removed any dependence on $\boldsymbol{u'}$ and $\boldsymbol{v'}$, which is why they could be disregarded.

Now writing $\boldsymbol{s'} = \boldsymbol{s} + \bar{\boldsymbol{s}}$, $\boldsymbol{t'} = \boldsymbol{t} + \bar{\boldsymbol{t}}$, and using $\boldsymbol{s} = \boldsymbol{0}$, we have

$$\hat{p}_w = -\frac{\bar{\boldsymbol{s}}.\boldsymbol{t} + \bar{\boldsymbol{s}}.(\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}})}{\boldsymbol{t}.\boldsymbol{t} + 2\boldsymbol{t}.(\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}}) + (\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}}).(\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}})}$$

Expanding in the perturbations $\bar{\boldsymbol{s}}$ and $\bar{\boldsymbol{t}}$, disregarding terms with more than a square perturbation in magnitude, we have

$$\hat{p}_w \approx -\frac{\bar{\boldsymbol{s}}.\boldsymbol{t}}{\boldsymbol{t}.\boldsymbol{t}} + 2\frac{\big((\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}}).\boldsymbol{t}\big)(\bar{\boldsymbol{s}}.\boldsymbol{t})}{(\boldsymbol{t}.\boldsymbol{t})^2} - \frac{\bar{\boldsymbol{s}}.(\bar{\boldsymbol{t}} + 3\bar{\boldsymbol{s}})}{\boldsymbol{t}.\boldsymbol{t}}$$

which, in our case, gives

$$\hat{p}_w \approx -\frac{\boldsymbol{\varepsilon}.\boldsymbol{t}}{\boldsymbol{t}.\boldsymbol{t}} + 3\frac{(\boldsymbol{\varepsilon}.\boldsymbol{t})^2}{(\boldsymbol{t}.\boldsymbol{t})^2} + 2\frac{(\boldsymbol{\varepsilon}.\boldsymbol{t})(\boldsymbol{\delta}.\boldsymbol{t})}{(\boldsymbol{t}.\boldsymbol{t})^2} - \frac{\boldsymbol{\varepsilon}.\boldsymbol{\delta}}{\boldsymbol{t}.\boldsymbol{t}} - \frac{3\boldsymbol{\varepsilon}.\boldsymbol{\varepsilon}}{2\boldsymbol{t}.\boldsymbol{t}}. \tag{17}$$

As in the Couples/LSM case, only the first term – which is a linear combination of Gaussian random variables with mean 0 – has a significant contribution to distributional shape (the others have variance one power of N smaller) and therefore we can approximate the distribution of $\hat{p}_w$ by a Gaussian with mean equal to the mean of (17) and variance equal to the variance of $\frac{\boldsymbol{\varepsilon}.\boldsymbol{t}}{\boldsymbol{t}.\boldsymbol{t}}$.

The means of the first, third, and fourth terms of (17) are zero, because all components of $\boldsymbol{\varepsilon}$ are independent of all components of $\boldsymbol{\delta}$ and have mean zero. For the other terms we use $\mathrm{E}[\varepsilon_{m,n}^2] = \mathrm{Var}[\varepsilon_{m,n}] = w_{m,n}d_{2m+1,2n+1}$ to derive

$$\begin{aligned} \hat{p}_w &\approx \mathrm{N}\big(\mu(\boldsymbol{d},\boldsymbol{w}), v(\boldsymbol{d},\boldsymbol{w})\big), \text{ where } v(\boldsymbol{d},\boldsymbol{w}) = \frac{16\sum_{m,n} w_{m,n}^2 \tilde{d}_{m,n}^2 d_{2m+1,2n+1}}{\big(\sum_{m,n} w_{m,n}\tilde{d}_{m,n}^2\big)^2} \\ \mu(\boldsymbol{d},\boldsymbol{w}) &= 3v(\boldsymbol{d},\boldsymbol{w}) - \frac{24\sum_{m,n} w_{m,n} d_{2m+1,2n+1}}{\sum_{m,n} w_{m,n}\tilde{d}_{m,n}^2}. \end{aligned} \tag{18}$$

where $\tilde{d}_{m,n} = d_{2m+2,2n} + d_{2m+1,2n} + d_{2m+2,2n+1} - d_{2m+1,2n+2} - d_{2m,2n+2} - d_{2m,2n+1}$.

We have not expended a lot of effort testing the accuracy of this approximate distribution, as we did for the Couples/LSM work in Ref. 5. We expect that it will not be highly accurate unless some components are excluded from the estimator because of the imperfections in the model (15). Instead we will be content to demonstrate that we can use it to derive improved estimators.

Now we are in a position to derive "optimal" weightings by minimizing $v(\boldsymbol{d}, \boldsymbol{w})$. Lemma 3.1 immediately gives

$$w_{m,n} = \frac{1}{d_{2m+1,2n+1}}$$

and we call the detector thus weighted the *Triples/WLSM* estimator. As in the case of Couples/WLSM, we must be prepared to make a bias correction to the estimator because weighting to reduce the variance will generally increase the bias.

5. EXPERIMENTAL RESULTS

We now measure the improvement that weighting the least-squares estimators brings, as well as testing some other modified detectors suggested by the error models in Sects. 2–4. Each batch of tests involves a large set of cover images, into which payloads of different lengths are repeatedly embedded to test the steganalysis methods' ability a) to estimate the payload size accurately, and b) to classify cover and stego objects correctly. It requires a large amount of computation, which we perform using the distributed steganalysis project outlined in Ref. 2.

Most of our experiments will be performed on cover images derived from one parent set: 3000 never-compressed bitmaps downloaded from `http://photogallery.nrcs.usda.gov`. Originally very high resolution colour images apparently scanned from film, we reduced them in size to 640×416 pixels. In order to compare steganalysis performance between different types of cover we repeated the tests on the colour images, then using the same images reduced to grayscale, and then again with the same images subject to moderate JPEG compression ("quality factor" 80) prior to embedding. Using the same images, but with added JPEG compression and/or conversion to grayscale, ensures that any differences in performance are due to the image type and not content. In the final experiment we also use another, independent, set of cover images: 1000 larger never-compressed bitmaps taken directly from a variety of digital cameras as raw bitmaps and subject to no postprocessing.

Potentially there are a very large number of steganalyzers to test. We wish to include one or two traditional LSB steganalysis methods: the earliest "second-generation" payload estimator called *RS*[10] and the original *SPA*[7] method. We can then combine any of the following options for the LSM estimators: Couples- or Triples-based structure; LSM or new WLSM; plain estimate, bias-corrected (subtract the theoretically-predicted bias, assuming no payload embedded), or standardized (subtract the theoretically-predicted bias and divide by standard deviation); excluding components for which the model of cover deviations is imperfect, or not.

We have indeed tested almost every possible combination, but we cannot display all the results here because of the space it would require. Instead we whittle down the possibilities by ruling out certain combinations that are never good performers, without displaying results to back up these particular claims. First, standardized estimates (which in any case cannot estimate payload) are without value unless we do exclude components for which the cover model is imperfect (it seems that dividing by an incorrect standard deviation simply messes up the results); on the other hand, excluding components only weakens the detectors if we are not standardizing the estimate. Additionally, we were not able to form an estimator which excluded the problem components in the Triples detector (they are hard to identify). Second, correcting the bias is essential if we are aiming for quantitative steganalysis, but if we are only using the output of the detector as a classifier between the cases of zero and nonzero payload then subtracting bias is not necessary (in fact it very slightly weakens the results). Finally, the WLSM steganalyzers uniformly outperform the LSM steganalyzers, although we will certainly want to include some of the experiments that verify this.

Therefore we will display results from the following algorithms as estimators of payload:

(a) The Couples method (practically equivalent to SPA[7]);

(b) Couples/LSM (practically equivalent to SPA/LSM[6]);

(c) Couples/WLSM with bias correction;

(d) Triples/LSM[1];

(e) Triples/WLSM with bias correction.

When testing reliability as a discriminator between cover and stego objects, we will display:

(a) Standard RS[10] and Couples;

(b) Couples/LSM;

(c) Couples/WLSM *without* bias correction;

(d) Standardized Couples/WLSM with components $m = -2, -1, 0, 1$ excluded;

(e) Triples/LSM;

(f) Triples/WLSM *without* bias correction.

We should clarify some of the minor parameters used in these tests. For each detector we used groups of pixels (pairs or triplets or, in the case of RS, quadruplets) in horizontal rows. When testing on colour images we pooled the trace subset counts (and their analogy in the RS method) between the three colour channels. The RS "mask" used was $[0, 1, 1, 0]$ (see Ref. 10 for details).

Also, it is wise to alter the "optimal" weightings by suppressing any components that involve very few pixels, because they would be assigned an overly-large weight. This is natural, not least because the Gaussian approximation used to derive (5) and (15) is not accurate for very small numbers. Without much tuning we decided to exclude all components when d_{2m+1} (or, in the case of Triples, $d_{2m+1,2n+1}$) was lower than 10, zeroing out that particular weight. This amounts to excluding a few parts of the image where pixel difference is unusually high.

Finally, there is the possibility that each of these estimators can "fail": in the case of RS or Couples there is a quadratic equation to solve and it might have no roots; in the case of the LSM detectors it can rarely happen that the estimate of p is wildly outside the plausible region of $[0, 1]$. For the former the only options are to return an estimate of 1 (on the grounds that failures are more common for high embedding rates) or to give no estimate at all; for the latter either give no estimate or fall back to an estimate produced by a simpler estimator. We elected to exclude images for which the RS or Couples algorithm fails to have a root, but in fact no such images occurred in our tests because we only considered relatively low embedding rates. For the Couples and Triples LSM and WLSM estimators we caused them to fall back to the standard Couples estimate in the (very rare) cases where the algorithm gave an estimate of p outside the range $[-0.2, 1.2]$.

5.1. Reduction in Image-Specific Variance

Our first set of experiments determine the extent to which the theory predicts that optimal weighting will decrease image-specific variance. We compared $v(\boldsymbol{d})$ from (10) and $v(\boldsymbol{d}, \boldsymbol{w})$ from (11), with $\boldsymbol{w}$ as in (12), in each image of the various sets of covers. Those in the set of 3000 never-compressed grayscale covers showed reductions in theoretically-predicted variance of between 0.19% and 45.28%, with a mean reduction of 15.19%. Very similar numbers (0.14% to 38.96%, mean 14.07%) were observed in colour covers. Lower numbers (mean around 8%) were observed in previously JPEG-compressed sets, but we do not expect that the theoretically-derived variance is correct for JPEG covers because it relies on a cover model which does not extend to such images.

The analogous calculations for standard and weighted Triples showed a greater improvement, with mean reductions in image-specific variance of around 25%. We found it quite surprising that the unweighted estimators were as close to optimality as these figures suggest, but at least the theory does predict that the weighted estimators should be an improvement on the traditional unweighted methods.

On the other hand, we have already mentioned that weighting tends to increase the image-specific bias. Comparing $\mu(\boldsymbol{d})$ and $\mu(\boldsymbol{d}, \boldsymbol{w})$, we observed that in every case the bias was increased, by a factor of between 1.35 and 372 (mean factor 18.3) in grayscale never-compressed covers and similar results in the other cover sets. Comparing weighted and unweighted Triples bias (18) showed a rather smaller increase: in a few cases the weighted Triples bias is smaller than for the unweighted estimator, but generally it was larger by a factor of

around 3. These figures indicate how important it is to introduce bias correction to weighted estimators. But recall that the theoretically-predicted bias is for true values of p close to zero, so we should expect that the weighted and bias-corrected estimators will perform poorly for large values of p.

5.2. Improved Estimator Accuracy

We now measure the accuracy of the estimators on our sets of cover images. Inaccuracy takes two forms: *bias* in the estimation, and *spread* of the estimator value. We measure these quantities by computing the median observed error, and either sample interquartile range or sample standard deviation, when a certain payload is embedded into each image in each cover set. The argument for choosing median, rather than mean, as a measure of location is presented in Ref. 4, where it is noted that the tails of the steganalysis error distributions are heavy enough to cast doubt on the convergence of even quite low-order sample moments. We find it useful to continue to compute sample standard deviation, however, because it allows us to measure how well the steganalysis estimators suppress outliers.

The experiments were repeated with five estimators and embedding rates $p = 0, 0.01, 0.02, \ldots, 1$ in order to examine how the true payload effects the estimators' performance. Results, broken down by cover image set (colour or grayscale, never-compressed or previously JPEG-compressed) are shown in Figure 2. We already know[1, 4, 9] that LSM-based detectors have poorer performance as p grows, and also that bias correction depends on the true payload being close to zero, so we have opted to show only the performance up to $p = 0.25$.

From these tables and charts we conclude as follows. Generally the bias is much less significant than the estimator spread. In all cases except for colour JPEG covers, the Couples/LSM estimator is *worse* than the plain Couples estimator; such performance was not visible in Ref. 6 (it could be seen in the tables in Refs. 1 and 8 but it was not commented on). In fact the plain Couples estimator collapses when presented with colour JPEGs. The Triples/LSM estimator is better than plain Couples (non LSM) when used on colour bitmap or JPEG covers (as tested in Ref. 1) but barely so on grayscale covers and only for very small values of p. Thankfully the WLSM methods (with bias correction) improve on their LSM counterparts, usually reducing estimator spread by around 10-30%, so that Triples/WLSM is always the best estimator for small enough values of p. We can also see that the LSM and WLSM estimators start to suffer from a negative bias as p grows, and their estimator spread also increases. We can conclude that the Triples/WLSM estimator is the best of all the least-squares estimators, and the overall best choice for p below about 0.2. Above that level it would be best to return to the simplest structural estimator, plain Couples.

This accords with our expectations (we already knew that LSM algorithms are weak for large values of p). Of course, estimation of small payloads is the interesting case, when it is hard to distinguish them from zero payloads.

5.3. Improved Discrimination Between Cover and Stego Images

Although accurate estimation of payload is useful, the primary task for a steganalyst is to decide whether any payload is present. Since the LSM estimators have best performance for smallest payloads, it seems likely that they will be able to discriminate between covers and stego objects with the most sensitivity.

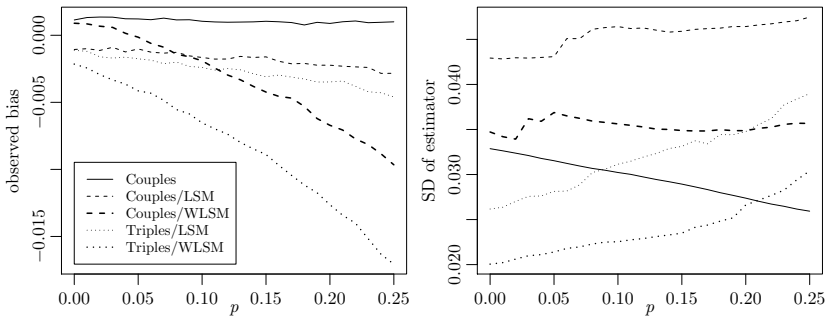
In general, accuracy of discrimination is measured by false positive and false negative detections and often displayed as a Receiver Operating Characteristic (ROC) curve; to characterise the behaviour of the discriminator one should display ROCs for every embedding rate, which is unmanageable. We must therefore reduce the dimensionality of the data in some way, in order to obtain a good performance metric, but there are no perfect ways to do this. To summarise concisely the steganalysis methods' ability to detect a payload, we have made an arbitrary but plausible definition of "reliable" detection to mean the ability to separate covers from stego objects with 5% false positives and 50% false negatives (the same was used in Refs. 1 and 11). We then determine experimentally the lowest embedding rate at which this reliable detection is achieved, by each discriminator and in each cover set separately. Table 1 shows the results.

We can see again that the traditional detectors collapse when presented with JPEG-compressed colour covers, and that in other situations the Couples/LSM method has fairly poor performance as a discriminator. These parallel the results for estimation. The unweighted Triples detector is the most sensitive of all the unweighted

Zero-payload profile

Detector	Bias	IQR (SD)
Couples	0.114	1.96 (3.29)
Couples/LSM	-0.106	2.87 (4.29)
Couples/WLSM	0.089	2.53 (3.47)
Triples/LSM	-0.112	2.18 (2.62)
Triples/WLSM	-0.215	1.69 (2.01)

all figures $\times 10^{-2}$

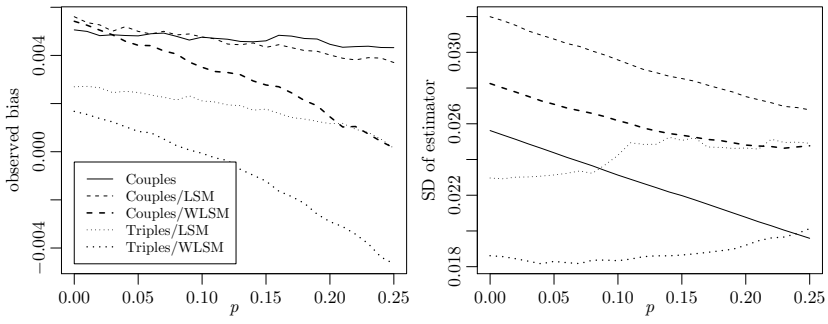


(a) Grayscale, never-compressed covers.

Zero-payload profile

Detector	Bias	IQR (SD)
Couples	0.506	2.08 (2.56)
Couples/LSM	0.561	2.71 (3.20)
Couples/WLSM	0.542	2.41 (2.83)
Triples/LSM	0.270	1.79 (2.30)
Triples/WLSM	0.168	1.42 (1.86)

all figures $\times 10^{-2}$

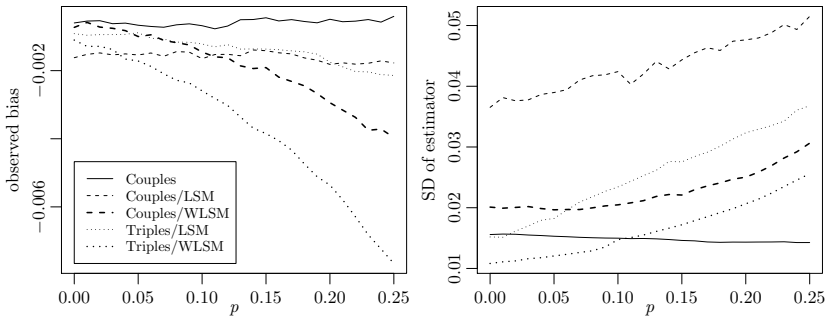


(b) Colour, never-compressed covers.

Zero-payload profile

Detector	Bias	IQR (SD)
Couples	-0.060	1.17 (1.56)
Couples/LSM	-0.162	1.47 (3.65)
Couples/WLSM	-0.073	1.30 (2.01)
Triples/LSM	-0.915	0.93 (1.52)
Triples/WLSM	-0.110	0.88 (1.08)

all figures $\times 10^{-2}$

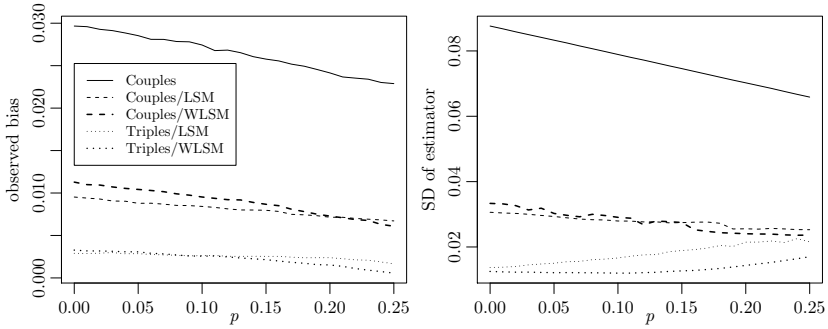


(c) Grayscale, previously-compressed covers.

Zero-payload profile

Detector	Bias	IQR (SD)
Couples	2.967	6.91 (8.76)
Couples/LSM	0.953	2.39 (3.06)
Couples/WLSM	1.128	2.57 (3.33)
Triples/LSM	0.290	0.97 (1.37)
Triples/WLSM	0.327	1.00 (1.25)

all figures $\times 10^{-2}$



(d) Colour, previously-compressed covers.

Figure 2. Estimator error (observed bias and spread) as computed in four sets of 3000 images, for five different estimators. The tables show the bias, interquartile range, and standard deviation of the observed estimators when no payload is embedded, and the charts show how the performance is affected by the payload size, up to proportionate payload $p = 0.25$. The WLSM estimators include bias correction.

Table 1. The lowest embedding rate (secret bits per cover pixel, which equals proportion of maximum payload) for which "reliable" discrimination from $p = 0$ is achieved. Here, "reliable" is taken to mean a false positive rate of 5% and a false negative rate of 50%. Results computed for seven methods, four sets of 3000 cover images, and accurate to 0.001. WLSM discriminators do not include bias correction.

	Never-Compressed Bitmaps		Previously JPEG Compressed	
	Grayscale	Colour	Grayscale	Colour
RS	0.036	0.054	0.023	0.283
Couples	0.036	0.052	0.018	0.220
Couples/LSM	0.045	0.060	0.022	0.068
Couples/WLSM	0.040	0.056	0.016	0.063
Standardized Couples/WLSM	0.060	0.037	0.080	0.156
Triples/LSM	0.033	0.041	0.014	0.026
Triples/WLSM	**0.028**	**0.035**	**0.011**	**0.025**

detectors. It is clear that weighting the LSM calculation always leads to improved discrimination, and the Triples/WLSM detector is the best performing. The improvement due to weighting is modest but at its greatest in the case of uncompressed covers (this should not surprise us, since the weighting factors are derived from a cover model which is not accurate for JPEG images).

The standardized version of Couples/WLSM is a poor performer with respect to this metric (except, curiously, for the case of colour never-compressed covers where its performance is second-best of all the discriminators). Again, this should not surprise us: the intended application of the standardized statistic is to high-reliability steganalysis, i.e. a very low rate of false positives. So to measure the detectors' ability to discriminate with very high reliability, we repeated the same experiments with alternative critical values of 0.1% false positives and 50% false negatives; these results are displayed in Table 2.

It is notable that the payloads necessary for detection at these levels are much higher than if we allow a rate of 5% false positives. This time we see the advantage conferred by the standardized Couples/WLSM discriminator. In the case of grayscale never-compressed covers it is comfortably the most sensitive detector. In the case of colour never-compressed covers it ties the Triples/WLSM estimator. But it remains a weak performer on previously JPEG-compressed covers. Again no surprise, because the cover model that the theory is based on is not accurate for JPEGs.

With so many experiments coming from a single parent set of 3000 covers, we thought it was important to perform one final additional set of experiments. For these we used a completely independent set of cover images: 1000 never-compressed bitmap images taken directly from a variety of digital cameras in raw format, all sized 1504×1000 (so substantially larger images than the 3000 used in the previous experiments). These images were not used for any tuning or selection of detectors, and indeed these experiments were conducted at the very end of the preparation of this paper. We advocate that to perform a set of experiments on an independent set of covers, on which the detectors have not been tuned, is good experimental practice. It verifies that any improvements in performance are not due to the detectors becoming too adapted to a particular cover set.

We repeated the experiment of finding the minimum payload at which 5% false positives and 50% false negatives could be observed (with only 1000 image it is not sensible to test lower false positive rates, as they depend on extreme order statistics). Table 3 shows the results. The first thing to note is that, although these images are much noisier than the original set of 3000 covers, their larger size means that substantially more sensitive detection is possible (dependence on cover size, if all other features remain the same, can be deduced from the theoretical distributions (10) and (18)). These results confirm those seen in Table 1: the Couples/LSM

Table 2. The lowest embedding rate (secret bits per cover pixel) for which "highly reliable" discrimination from $p = 0$ is achieved. "Highly reliable" is taken to mean a false positive rate of 0.1% and a false negative rate of 50%. Results computed for seven methods, four sets of 3000 cover images, and accurate to 0.001. WLSM discriminators do not include bias correction.

	Never-Compressed Bitmaps		Previously JPEG Compressed	
	Grayscale	Colour	Grayscale	Colour
RS	0.159	0.179	0.100	0.811
Couples	0.195	0.185	0.109	0.627
Couples/LSM	0.255	0.180	0.099	0.227
Couples/WLSM	0.249	0.168	0.065	0.176
Standardized Couples/WLSM	**0.117**	**0.110**	0.158	0.433
Triples/LSM	0.232	0.119	0.072	**0.070**
Triples/WLSM	0.216	**0.110**	**0.058**	**0.070**

method is not very good, and weighting only brings its performance back into line with traditional non-LSM steganalysis methods, but Triples/WLSM is comfortably the most sensitive detector of payload.

6. CONCLUSIONS AND DIRECTIONS FOR FURTHER WORK

A theoretical model of estimation error is very valuable. Even when, as here and in Ref. 5, we only know the error distribution in the case of zero payload, the theory has an immediate application in the development of better steganalysis. Here we have shown that weighted least-squares steganalysis outperforms traditional unweighted least-squares steganalysis, particularly in the difficult case of never-compressed covers. We have also extended the theory of between-image error to Triples/LSM steganalysis, and the new Triples/WLSM estimator and discriminator are the most accurate detectors of LSB replacement steganography yet known.

The most obvious direction for further work is to refine the models for cover deviation, (5) and (15). Table 2 shows the potential that standardized outputs have in the pursuit of very low false positive rates, but at present the standardized discriminators only work for Couples/WLSM and require some components to be excluded. Perhaps better models can be derived from those for pixel difference in natural images, but the great advantage of the models leading to (5) and (15) is that they are not parametric – it might be difficult to estimate the parameters needed for more complex models, from a single image. If the cover models can be improved to work for all components as well as they work for most then we confidently expect to produce even better steganalysis.

Another way to improve the theoretical results is to extend them to work with images containing a payload. We can see, in Fig. 2, that bias correction with the improperly-computed bias is causing poor performance for moderate payloads. But it was noted in Ref. 5 that analysis of error distributions in the case of general payloads has certain technical difficulties.

More generally, we hope to use knowledge of factors influencing the errors further to refine the structural detectors. At present, structural steganalysis has an "all or nothing" character: a cover model which (until Ref. 5) was assumed to hold precisely, and also an application of the Law of Large Numbers (suppressed here only because we assume no payload) to assume that the realisation of the random variables e'_m, o'_m equal their expectations. We are moving towards a more subtle approach, in which deviations from these assumptions are permitted and can be quantified, which may eventually lead to a maximum likelihood estimator.

Table 3. The lowest embedding rate (secret bits per cover pixel) for which "reliable" discrimination from $p = 0$ is achieved, again taken to mean a false positive rate of 5% and a false negative rate of 50%. These results are from a set of 1000 large (1.5 megapixel) never-compressed images taken directly from digital cameras with no postprocessing. Results accurate to 0.0001. WLSM discriminators do not include bias correction.

	Never-Compressed Bitmaps	
	Grayscale	Colour
RS	0.0073	0.0082
Couples	0.0068	0.0072
Couples/LSM	0.0107	0.0087
Couples/WLSM	0.0079	0.0076
Standardized Couples/WLSM	0.0120	0.0083
Triples/LSM	0.0051	0.0034
Triples/WLSM	**0.0040**	**0.0024**

ACKNOWLEDGMENTS

The author is a Royal Society University Research Fellow. The author is grateful to Jessica Fridrich and Tomáš Pevný, who supplied the set of 1000 images from digital cameras used in the final series of experiments.

REFERENCES

1. A. Ker, "A general framework for the structural steganalysis of LSB replacement," in *Proc. 7th Information Hiding Workshop*, *Springer LNCS* **3727**, pp. 296–311, 2005.
2. A. Ker, "Quantitive evaluation of Pairs and RS steganalysis," in *Security, Steganography, and Watermarking of Multimedia Contents VI*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **5306**, pp. 83–97, 2004.
3. R. Böhme, "Assessment of steganalytic methods using multiple regression models," in *Proc. 7th Information Hiding Workshop*, *Springer LNCS* **3727**, pp. 278–295, 2005.
4. R. Böhme and A. Ker, "A two-factor error model for quantitative steganalysis," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6072**, pp. 59–74, 2006.
5. A. Ker, "Derivation of error distribution in least-squares steganalysis." To appear in *IEEE Transactions on Information Forensics and Security*, 2007.
6. P. Lu, X. Luo, Q. Tang, and L. Shen, "An improved sample pairs method for detection of LSB embedding," in *Proc. 6th Information Hiding Workshop*, *Springer LNCS* **3200**, pp. 116–127, 2004.
7. S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," *IEEE Transactions on Signal Processing* **51**(7), pp. 1995–2007, 2003.
8. A. Ker, "Fourth-order structural steganalysis and analysis of cover assumptions," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6072**, pp. 25–38, 2006.
9. A. Ker, "Steganalysis of embedding in two least significant bits." To appear in *IEEE Transactions on Information Forensics and Security*, 2007.
10. J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," *Proc. ACM Workshop on Multimedia and Security*, pp. 27–30, 2001.
11. A. Ker, "Improved detection of LSB steganography in grayscale images," in *Proc. 6th Information Hiding Workshop*, *Springer LNCS* **3200**, pp. 97–115, 2004.