Linking Intrinsic Difficulty and Regret to Properties of Multivariate Gaussians in Image Steganalysis

Antoine Mallet, Rémi Cogranne, Patrick Bas

Linking Intrinsic Difficulty and Regret to Properties of Multivariate Gaussians in Image Steganalysis

Antoine Mallet**
antoine.mallet@utt.fr
LIST3N - Univ. of Technology of Troyes
Troyes, Grand Est, France

Rémi Cogranne
remi.cogranne@utt.fr
Univ. of Technology of Troyes
Troyes, Grand Est, France

Patrick Bas
Patrick.Bas@centralelille.fr
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL
Villeneuve d'Ascq, Hauts de France
France

ABSTRACT

This paper deals with the Cover-Source Mismatch (CSM) problem faced in operational steganalysis. Based on a multivariate Gaussian model of the distribution of the noise contained in natural images, it provides proxies for the two important empirical measures of CSM: intrinsic difficulty and regret. The former can be modeled with the determinant of the covariance matrix of the noise present in an image. The latter can be predicted with a modified Kullback-Leibler divergence between the distribution of the noises of images coming from different cover-sources. We first recall the reasoning behind the multivariate Gaussian model of the noise, and detail how to compute the statistic of the distribution of the noise. Then, our proposed models are compared to empirical data with a specifically designed cover-source generation process. For both quantities, very high correlation coefficients between the model and the observations are obtained. Finally, realistic cover-sources are used to further illustrate the relevance of our model.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; • **Applied computing** → **Investigation techniques**; • **Security and privacy**;

KEYWORDS

forensics, steganalysis, cover-source mismatch, image processing pipeline, fingerprint

ACM Reference Format:
Antoine Mallet, Rémi Cogranne, and Patrick Bas. 2024. Linking Intrinsic Difficulty and Regret to Properties of Multivariate Gaussians in Image Steganalysis. In *Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security (IHMMSec '24), June 24–26, 2024, Baiona, Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3658664.3659643

*This work has been funded by the EU's Horizon 2020 program under grant agreement No. 101021687 (UNCOVER project), and the French ANR PACeS project No. ANR-21-CE39- 0002.

IHMMSec '24, June 24–26, 2024, Baiona, Spain

ACM ISBN 979-8-4007-0637-0/24/06
https://doi.org/10.1145/3658664.3659643

1 INTRODUCTION

In the field of multimedia security, steganography is the art of secret communication. It uses cover objects from an innocent medium, such as images, videos, or text files, and embeds data to form so-called stego objects. On the other side of the coin, steganalysis develops tools to distinguish between cover and stego objects, generally based on supervised Machine Learning (ML) models.

In an operational context, however, the steganalyst might face images that widely differ from the ones used to design his tools. This mismatch between the training and testing samples is a major problem in steganalysis, known as the *Cover-Source Mismatch* (CSM) problem. The CSM is the fact that samples from different cover-sources follow different distributions. Its effect on steganalysis is that, when training a detector on some cover-sources, its decision rule might become irrelevant on some other ones, to the point where it can get completely blind.

The CSM problem in steganalysis is well-known and studied. Most of the research is focusing on mitigating it [3, 14, 25], while some study the causes of CSM [1, 10]. Providing a model for the cover-source is very difficult in practice since it can be possibly defined by infinite and undefined processings, especially ones that are under proprietary software, and/or using non-standard and new algorithms. This is partially why, despite the growing interest in the issue, there is barely any attention given to providing an actual model of the cover-source.

On the other hand, modeling the CSM, which essentially consists of describing the discrepancy between cover-sources, also remains a largely open problem. From the point of view of the steganalyst, this discrepancy should relate to the empirically observed drop in the accuracy of detectors. However, it is clear that the embedding scheme and the design of the detector also impact this drop in accuracy. Therefore, the measure of the CSM problem, which is the measure of the cost of testing and training on different sources, necessarily also captures the impacts of the embedding and the model.

1.1 Related work

Although there are over 100 papers dealing with the CSM problem in steganalysis [17], there is currently no proper model of a cover-source. This task is indeed extremely challenging since the causes of CSM are very broad. CSM has been a known problem in steganalysis for 20 years now; but the awareness of the community truly starts with the BOSS contest in 2010. The following decade mostly bears studies trying to mitigate the impact of CSM, along with a rough understanding of its causes. The first deep dive into the causes of

CSM is rather recent [10]. It comes in the aftermath of the ALASKA competition, which further raised the issue of generalization of detectors in operational steganalysis.

Since then, there has been more focus on properly identifying the causes of CSM, and quantifying their impact. Some papers even give a metric or pseudo-metric for specific causes of CSM, showing how it can relate to empirical measures: for JPEG quantization tables [21, 24] or texture complexity [13] for instance. These can be considered partial models of the CSM, focusing on one cause.

The current most advanced work is probably [2], exploring a geometrical approach of the CSM. Stating that the cover sources define manifolds of lower dimensions, in which samples coming from a given cover-source live, one can predict the impact of CSM as the angle between the manifolds of two cover-sources. Defined as a dot product, this signed measure of an angle can actually reflect the asymmetry of the regret, which can be high from one source to another but low the other way around [1, 4, 10].

1.2 Contributions & paper's outline

The present paper provides another approach to the CSM and an assessment of its impact on steganalysis. The contributions are listed as the following:

(1) A discussion on what is the cover source mismatch, clearly identifying causes and consequences, and definitions of the intrinsic difficulty and regret that take into account the role of the detector in measuring the CSM problem.
(2) Evidences of the limits of measuring the CSM through its impact on steganalysis, i.e. via the regret.
(3) A statistical approach to the question of the CSM model, providing a way to predict the intrinsic difficulty and the source inconsistency, both necessary to compute the regret.

To answer these questions, the paper is organized as follows: Sec. 2 recalls the considered stochastic model of the noise in natural images and explains how it can be leveraged as a fingerprint of the cover-source. Sec. 3 provides rationales and detailed explanations of how this fingerprint can be used to estimate the empirical impact of the CSM, namely the source's intrinsic difficulty and the regret. Sec. 4 details the settings of the proposed experiments. Results in a controlled environment are discussed in Sec. 5, and in a realistic setup in Sec. 6. Finally, Sec. 7 concludes the present paper.

2 GAUSSIAN MODEL OF THE NOISE

Whereas [2] develops a geometrical approach to model the CSM, We explore here a statistical approach, based on estimating the statistics of the multivariate model of the noise in images, leveraged by [6, 23] in Natural Steganography.

2.1 Rationale

The correlation of neighboring pixels is a fingerprint of the Image Processing Pipeline (IPP), as illustrated in Fig. 1. A natural image (as opposed to forged or generated images), goes through an acquisition step and a list of processings. Right after acquisition, pixels bear spatially independent noise. Many steps of the IPP, however, modify pixel values according to one anothers' – e.g. during demosaicking, missing color values are interpolated using neighboring RAW pixel values. We can hypothesize that different IPPs will correlate the

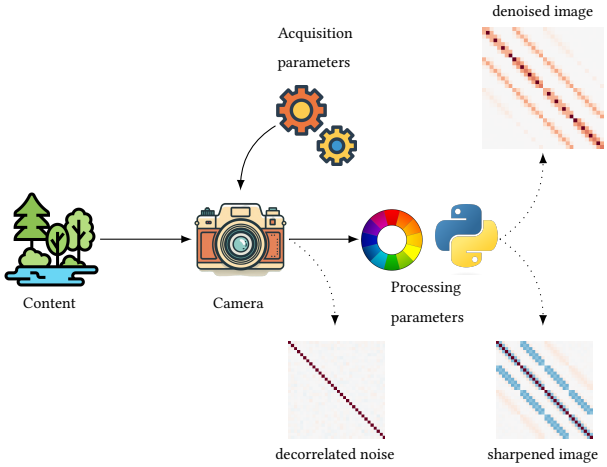


Figure 1: Schematic reasoning for the multivariate distribution of the noise in natural images. Below are correlations of neighboring pixels in the RAW (just after acquisition) and developed domains. Red cells indicate a positive correlation, and blue cells a negative correlation.

pixels differently. For example, we can see that the correlated noises after sharpening and after denoising are very different, with each line of the matrices indicating the correlation of one pixel of a 8×4 neighborhood with every pixel of this neighborhood, and where a red (resp. blue) cell indicates a positive (resp. negative) correlation. In particular, note that, as compared to the noises obtained in developed images, the noise in the RAW domain is decorrelated.

Being able to compare fingerprints of different pipelines, in particular being able to quantify their difference would be a good estimation of the difference between cover-sources, and the impact of their mismatch in steganalysis.

2.2 Formalization

In a RAW image, there exists a noise stemming from the stochastic nature of the acquisition process [7]. We can write the value of a pixel in the RAW domain x_i as its "true" value, to which an univariate heteroscedastic noise is added:

$$x_i = z_i + e_i, \tag{1}$$

with

$$e_i \sim \mathcal{N}\left(0, a \times z_i + b\right), \tag{2}$$

and where a and b are called the heteroscedastic parameters of the distribution. These, in the most usual setup where the steganalyst does not have access to the RAW image, are hard to estimate [9].

Furthermore, the noise in a developed image can be considered to follow a multivariate Gaussian distribution under two conditions on the IPP [8]. First, it needs to be linear, i.e. there exists a linear transformation that can transform the RAW image into the developed image. Rather than considering the whole image, one can assume that this transformation can be defined on a smaller portion of the image, such that the $\sqrt{n} \times \sqrt{n}$ section of the RAW image will be transformed into the $\sqrt{m} \times \sqrt{m}$ developed image, with m and

n perfect squares. Formally, we can define $H_k \in \mathcal{M}^{m \times n}$ the set of 2-D matrices of size $m \times n$, that transforms the k-th vectorized RAW image block $X_k \in \mathbb{N}^n$ into the corresponding vectorized developped image block $Y_k \in \mathbb{N}^m$:

$$Y_k = H_k X_k, \tag{3}$$

In this paper, we consider developed image blocks of size 8×8. To account for the pixels on the edge of the block, that are correlated with pixels outside the block, we consider an extra outer layer of raw pixels. Therefore, $m = 8^2$ and $n = 10^2$. Secondly, it needs to be stationary, i.e. this linear application is the same for every block in the image:

$$H_k = \mathbf{H}. \tag{4}$$

If both conditions are verified, then the noise in the developed domain follows a multivariate Gaussian distribution:

$$y_k \sim \mathcal{N}(\mathbf{H}z_k, \Sigma), \tag{5}$$

where Σ is the covariance matrix. Let us now define $\bar{X}$ the matrix of size $n \times m_x$, with m_x the number of vectorized image block in the RAW image X, and similarly with $\bar{Y}$. Then, given a linear and stationary IPP, the transformation matrix $\mathbf{H}$ can be computed using the least-square methods:

$$\mathbf{H} = \bar{Y}\bar{X}^T(\bar{X}\bar{X}^T)^{-1}. \tag{6}$$

When dealing with synthetic data, we can get rid of the heteroscedasticity by generating an artificial homoscedastic noise by giving the same mean and variance to every pixel of a RAW image, and developing it. We finally get Σ as:

$$\Sigma = \mathbf{H}\mathbf{H}^T. \tag{7}$$

3 CSM MODEL

Adopting a formalism inspired by [22], let us define the following supervised steganalysis detector:

$$\begin{aligned} f(x|\theta_{p,\gamma}) : X &\to \{cover, stego\} \\ x &\mapsto y \end{aligned} \tag{8}$$

where p and γ are the parameters characterizing the IPP and the steganography respectively, and $\theta_{p,\gamma}$ the learnt parameters w.r.t. them. We introduce in the rest of the section the intrinsic diffculty and the regret, and our models for both. We assume that the steganography is fixed, thus that no stego-scheme mismatch will occur. Therefore, we will drop the γ parameter wherever possible.

3.1 Model of the intrinsic difficulty

Definition 1. *The **intrinsic difficulty** of a detector on a cover-source is the error of the detector tested on images coming from the same cover-source, assuming that the steganographic embedding is also the same:*

$$I_D(f|p) = \mathbb{E}_{(x,y)\sim\mathbb{P}((x,y)|p,\gamma)}(f(x|\theta_{p,\gamma}) \neq y). \tag{9}$$

The determinant of a matrix is generally understood as a measure of the volume of the parallelepiped defined by the rows of the matrix. Alternatively, it can be defined as an overall correlation factor. $\det(I) = 1$, where I is the identity matrix, and $\det(C) = 0$ implies that C contains linearly dependant (i.e. perfectly correlated) vectors. We hypothesize that the determinant of the covariance matrix derived from Eq. (7), noted $|\Sigma|$, is a good indicator of the intrinsic difficulty of a source. It is a measure of the "available free space" for the steganographer to modify pixel values without deviating from the distribution, and becoming detectable. In other words, the more pixels are correlated, the less freedom for steganographic modifications under compliance with these correlations.

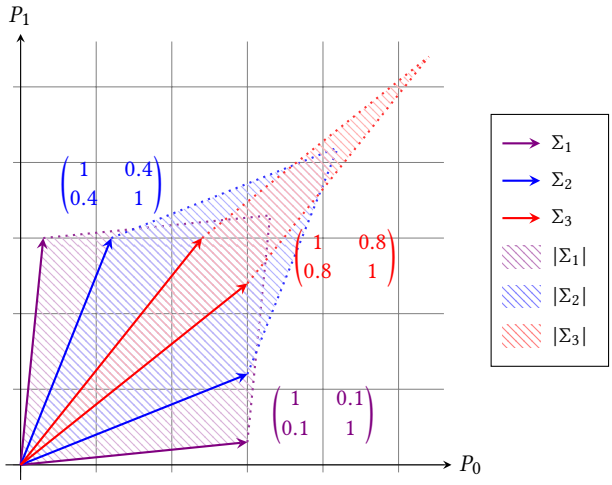


Figure 2: Correlation matrices visualized for 2-pixel samples. As we can see, the stronger each pair is correlated, the smallest the determinant of their correlation matrix: $|\Sigma_1| = 0.99$, $|\Sigma_2| = 0.84$ and $|\Sigma_3| = 0.36$.

To illustrate this statement, let us look at Fig. 2, depicting three covariance matrices. One can see that the more correlated pixel values are, the smaller the determinant (i.e. the volume) becomes. Accordingly, Fig. 3 shows the distribution of cover and stego samples of 2-pixel images, drawn from 2-D multivariate gaussian distributions with different covariance matrices. The stego samples are generated by randomly adding +1 or −1 to one pixel of each image. As one can see, looking from Fig. 3a to Fig. 3d, the more correlated pixels get, i.e. the smaller the covariance matrix's determinant gets, hence the more steganography should become detectable.

3.2 Model of the regret

The intrinsic difficulty is the measure of the performance of a detector in a clairvoyant scenario. When the detector faces samples from another cover-source, it will likely show some inconsistency.

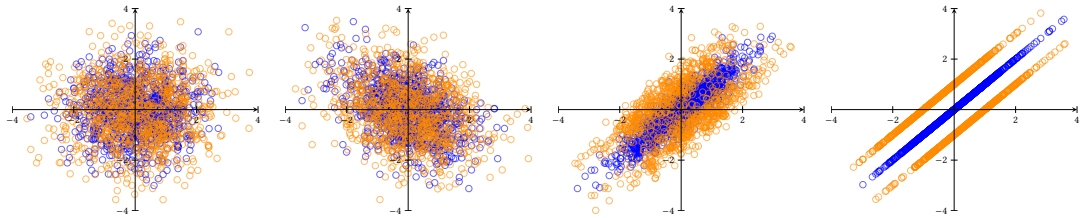
Definition 2. *The **source inconsistency** of a detector between two cover-sources is the detection accuracy, of testing the detector on a cover-source, given that it was trained on the other:*

$$S_I(f|p_1, p_2) = \mathbb{E}_{(x,y)\sim\mathbb{P}((x,y)|p_2,\gamma)}(f(x|\theta_{p_1,\gamma}) \neq y). \tag{10}$$

Then, the regret is just the cost of this inconsistency in regard of the intrinsic difficulty.

Definition 3. *The **regret** of a detector between two cover-sources is the difference between the source inconsistency and the intrinsic difficulty of source which the detector is tested on:*

$$r^f_{p_1,p_2} = S_I(f|p_1, p_2) - I_D(f|p_2). \tag{11}$$



(a) $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (b) $\Sigma = \begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}$ (c) $\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ (d) $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

Figure 3: Scatter plot of the values of 1000 synthetic 2-pixel images, drawn from multivariate normal distributions, of the specified covariance matrices. in blue are the cover images, in orange their stego counterparts, generated by randomly adding -1 or $+1$ to one of the pixels. The more correlated the pixels are, the easier the steganography is detected. The determinant of the covariance matrix Σ can be a good indication of the detectability of steganography.

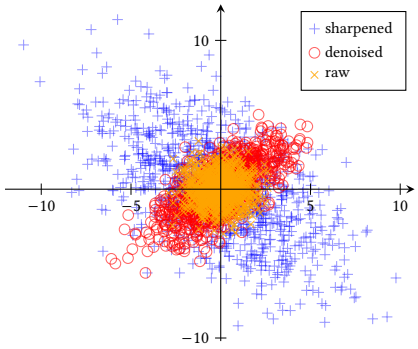


Figure 4: Effect of different processings to the covariance of the pixel noise in an image. The RAW image has decorrelated noise (orange). Denoising tend to create positive correlation between the pixels (red), whereas sharpening creates negative correlations.

In this paper, we investigate whether or not the discrepancy between distributions of the noise coming from different IPPs, characterized by their covariance matrices, can be used as an accurate measure in order to predict the regret. This idea follows the observation that different IPPs will produce different statistics of the noise, as illustrated in Fig. 4. It shows joint distributions of 2 neighboring pixels in images that are either decorrelated in the RAW domain (orange), positively correlated when processed with a denosing operation (red) or negatively correlated with a sharpening one (blue).

There exist many distances between statistical distributions, such as the Mahalanobis distance or the total variation distance. But note that the regret is generally asymmetrical. For its link with the Neyman-Pearson's lemma, we therefore suggest using the well-known Kullback-Leibler divergence, which is a measure of the cost (in bits) of encoding samples from a distribution P for a code optimized for another distribution Q, rather than using a code optimized for P. In the general case, it is quantified as the expectation of the log-likelihood ratio of the probability distributions of P and Q. In the case of multivariate Gaussian distributions, however, we can leverage the following expression

Table 1: Kullback-Leibler divergences obtained on the example distributions of Fig. 4. The columns give us the so-called "reference" distributions, and the rows give the "observations".

$D_{KL}(P\|\|Q)$		Q Shar	Q Raw	Q Den
P	Shar	0	7.67	6.90
P	Raw	1.34	0	0.41
P	Den	0.96	1.19	0

of the Kullback-Leibler divergence:

$$D_{KL}\left(\mathcal{N}_1||\mathcal{N}_2\right) = \frac{1}{2}(\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) + \ln\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) - n). \tag{12}$$

As shown in Sec. 4, we use synthetic developed noise for our statistical model. This noise has the same mean value on the whole "image". Therefore, the last term of Eq. (12) is canceled out:

$$D_{KL}\left(\mathcal{N}_1||\mathcal{N}_2\right) = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma_2^{-1}\Sigma_1\right) + \ln\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right)\right). \tag{13}$$

To complete our example, we can apply the D_{KL} to the joint distributions shown in Fig. 4. After estimating the covariance matrices of the obtained joint distributions, we get the divergences shown in table 1. We can see that the cost is greater when the sharpened distribution is taken as the observation with the denoised or raw ones as references. The asymmetry is clearly visible, and can also be linked to the asymmetry of the regret between "difficult" noisy sources and "easier" denoised sources [2].

4 EXPERIMENTAL SETUP

4.1 Choice of processing pipeline

In order to test our hypotheses, we need to create a setup where the gaussianity of the distribution of the noise is safe.

In mainstream pipelines, however, it is likely to find one or more non-linear processings: gamma correction is non-linear, as well as most denoising operations; demosaicking can be non-stationary,…

Demosaicking: We chose the bilinear demosaicking [16] algorithm, as implemented in the *color-demosaicing* python library.

Histogram stretch: To ensure having a broad range of pixel values, we modify the demosaicked image I_{d} such that the darkest (resp. brightest) pixel value equals 0 (resp. 255). We therefore compute I_{hs} by performing the following histogram stretch operation:

$$I_{\mathrm{hs}} = (I_{\mathrm{d}} - \min(I_{\mathrm{d}})) \times \frac{\max(I_{\mathrm{hs}})}{\max(I_{\mathrm{d}})}. \tag{14}$$

Grayscale conversion: We then convert our rgb images obtained via demosaicking to single-channel grayscale images, using the conversion proposed in the *opencv* library:

$$I_{\mathrm{gs}} = 0.299 I^R_{\mathrm{hs}} + 0.587 I^G_{\mathrm{wb}} + 0.114 I^B_{\mathrm{hs}}. \tag{15}$$

Post-processing: We chose a variant of the famous unsharp masking method:

$$I_{\mathrm{pp}} = I_{\mathrm{gs}} + p \times (\mathcal{L} \circledast I_{\mathrm{gs}}), \tag{16}$$

where:

$$\mathcal{L} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \tag{17}$$

and where p is a parameter used to set the strength of the operation. Note that for values of p close to zero, $p < 0$ corresponds to denoising, and $p > 0$ to sharpening.

4.2 Construction of the dataset

We define 21 sources, by using values of the strength parameter p defined in Eq. (16) in $\{-1, -0.9, -0.8, \ldots, 0.9, 1\}$ For our experiments, we use two kinds of images. First, we use synthetic images, made out of noise, to get statistical models as precise as possible. These synthetic RAW images are created by setting all of their values to an univariate heteroscedastic Gaussian distribution, with $a = 1$ and $b = 0$:

$$x \sim \mathcal{N}(\mu, \mu), \tag{18}$$

and with $\mu = 2^{13}$. These images will be used to compute the covariances matrices, their determinant and the Kulback-Leibler divergence between cover-sources.

Second, we use regular images to measure empirically the CSM effect in a steganalysis task. They are developed using the same pipeline, with an additional center cropping before compression, such that they are all of size 512×512. Each cover-source contains 15 000 images, randomly chosen in a pool of 50 000 coming from the ALASKA#2 dataset [5] These images are used to create the empirical measures of the CSM, via steganalysis.

4.3 Steganalysis

We perform steganography with J-UNIWARD [12]. On the other side, We use DCTR features [11] to do steganalysis, as they show good results and are rather fast to compute. We use them alongside several simple detectors: with a linear classifier as proposed in [2], and with SVMs with linear and Gaussian kernels. Each detector is trained on 10,000 images, randomly selected as cover or stegos, and tested on the remaining 5,000 images, also randomly from either class.

Table 2: Detailed processings used to define the realistic cover-sources in the experiments of Sec. 6.

Realistic cover-sources			
USM	DPD	DPD-USM	USM-DPD
50	30	90-50	350-30
150	50	90-150	350-50
250	70	90-250	350-70
350	90	90-350	350-90

4.4 Case of a realistic setup

To conduct the last experiments of this paper, we designed another more realistic setup, where the choice of cover-sources include real processings, performed with the *Rawtherapee* software. All the cover-sources apply the same bilinear demosaicking and white balance operation. They are defined by the following post-processings. The processings are the directional-pyramid denoising (DPD) [18] and the unsharp masking (USM) [20]. For each, 4 levels of intensities are chosen, giving 8 1-processing cover-sources. 8 additional 2-processings cover-sources are generated by first applying the DPD (resp. USM) with the strongest intensity followed by one of the 4 intensities of USM (resp. USM). Details on the value of the intensities can be found in table 2. Finally, images are converted to grayscale, center cropped and JPEG compressed with a quality factor of 100.

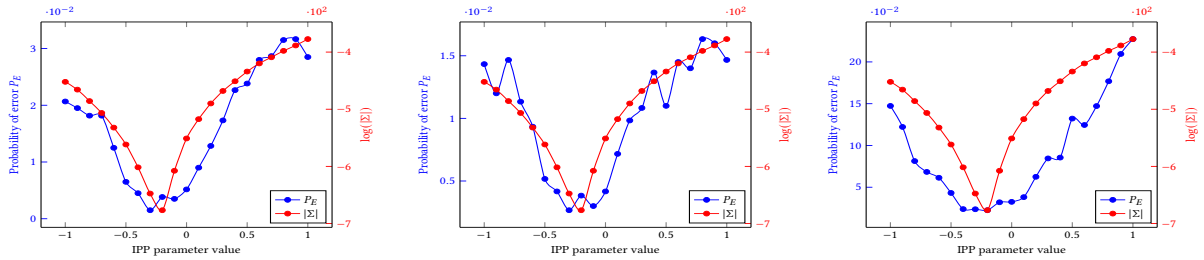
The steganalysis scheme is also more realistically applied, with a lower (but still relatively high) payload $\rho = 0.5$ bpnzAC.

5 EXPERIMENTAL RESULTS

5.1 Relation between intrinsic difficulty and determinant of the covariance matrix

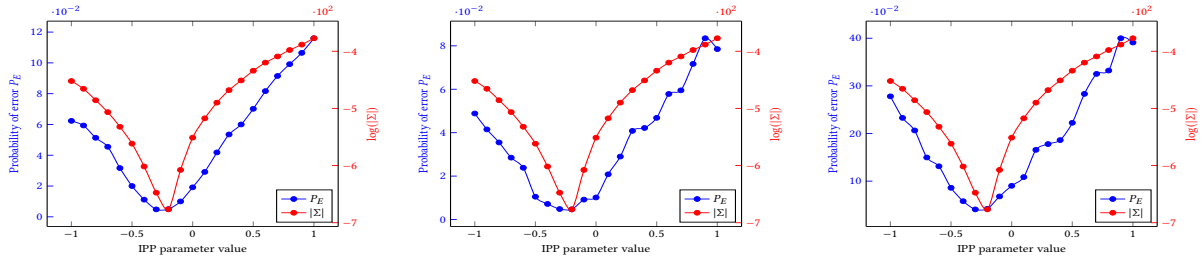
On one side, for each source, we perform steganalysis using the three detectors described in Sec. 4.3, at both payloads $\beta = \{1, 1.5\}$ bpnzAC. On the other, we compute the determinant of the covariance matrix, using the methodology presented in Sec. 2.2, on synthetic images, developed on each source. Due to the curse of dimensionality, the volume described in 64-D is extremely small. Hence, we use the logarithm of the determinant as a comparison. We end up with one determinant curve, and 6 intrinsic difficulties curves. Fig. 5 shows the curves obtained at payload $\beta = 1.5$ bpnzAC, and Fig. 6 the ones obtained at payload $\beta = 1$ bpnzAC. Although they are of different scales, the shape of the determinant's curve fits the intrinsic difficulty curves very well. We can express it in terms of different indicators. We suggest the Pearson correlation coefficient ($\mathrm{P_{CC}}$) [19] (see Sec. VII), the Spearman rank-correlation coefficient ($\mathrm{S_{CC}}$) [26] (see Sec. XIV.7), and Kendall's τ [15]. Results are shown in Table 3. Note that, as the logarithm is strictly monotonous, the rank-correlation measures are not affected by it.

Results show, contrary to our initial expectations, that applying a strong "denoising" using the proposed post-processing actually increases the value of the determinant. This is because this operation acts as as "noise remover" rather than a proper denoiser. It appears that under a certain threshold, it starts removing noise that is not there, hence ends up adding noise. This is visually confirmed



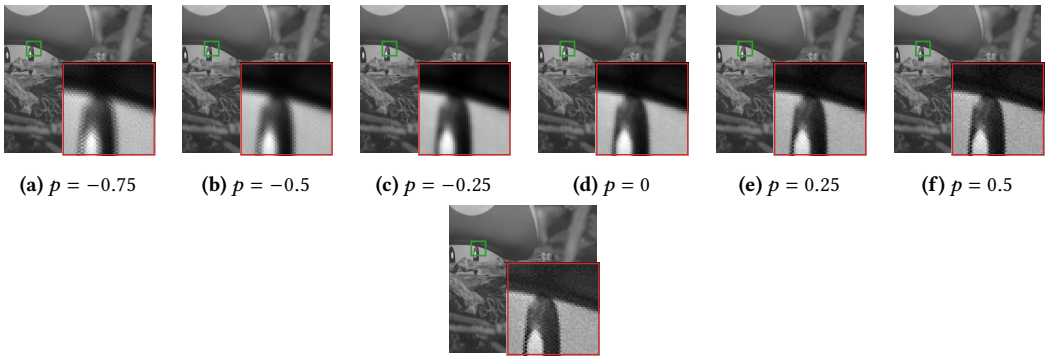
(a) Logistic Regression **(b) Linear SVM** **(c) Gaussian SVM**

Figure 5: Determinant and intrinsic difficulty of different detectors, for J-UNIWARD with payload $\beta = 1.5$ bpnzAC.



(a) Logistic Regression **(b) Linear SVM** **(c) Gaussian SVM**

Figure 6: Determinant and intrinsic difficulty of different detectors, for J-UNIWARD with payload $\beta = 1$ bpnzAC.



(a) $p = -0.75$ **(b)** $p = -0.5$ **(c)** $p = -0.25$ **(d)** $p = 0$ **(e)** $p = 0.25$ **(f)** $p = 0.5$

(g) $p = 0.75$

Figure 7: Effect of the processing with different parameter values. Positive values correspond to "adding" noise, and negative values to "removing" noise. However, we see that for $p < -0.25$, the image seems to get noisier again.

in Fig. 7, where values of $p < -0.25$ seem to enhance contrasts again, which corresponds to the minimal value of the determinant obtained at $p = -0.2$.

Additionally, although the correlation coefficients are high – validating our approach, we still see substantial variations between payloads. With the linear SVM, at $\beta = 1$ bpnzAC, $S_{CC} = 0.991$ but only equals 0.915 at $\beta = 1.5$ bpnzAC. A similar comment can be made for the choice of model, where $P_{CC} = 0.889$ for the Gaussian SVM at $\beta = 1.5$ and $P_{CC} = 0.950$ for the logistic regressor at the same payload.

5.2 Relation between regret and statistical divergence

To compare the regret between two sources and the D_{KL} between the Gaussian models of their noise, we first build the regret matrix and the D_{KL} matrix. The former is obtained by computing the regret between every pair of training and testing sets:

$$R_p^f = (r_{p_i,p_j}^f) \tag{19}$$

The latter is obtained in the same manner, computing the D_{KL} between all sources. Before comparing both matrices, one should address the scaling problem. Indeed, the D_{KL} is defined in $[0, +\infty[$, but the regret is bounded in $[0, 0.5]$. To ensure that both quantities

Table 3: Correlation measures between the determinant of the covariance matrix $|\Sigma|$ and the intrinsic difficulties obtained for the different detectors and payloads.

Detector	β (bpnzAC)	P_{CC}	S_{CC}	Kendall's τ
LogReg	1.5	0.950	0.978	0.902
	1.0	0.953	0.993	0.952
SVM lin	1.5	0.925	0.915	0.775
	1.0	0.940	0.991	0.943
SVM rbf	1.5	0.889	0.974	0.895
	1.0	0.932	0.978	0.905

are defined on the same set, we suggest using an activation function, such as the *sigmoid*:

$$\begin{aligned} \text{sig}[x] : \mathbb{R} &\rightarrow [0,1] \\ x &\mapsto \frac{1}{1+e^{-x}}. \end{aligned} \tag{20}$$

We note D^*_{KL} the following transformed divergence:

$$D^*_{KL}(\mathcal{N}_{p_i}||\mathcal{N}_{p_j}) = \frac{1}{2} \times \text{sig}[D_{KL}(\mathcal{N}_{p_i}||\mathcal{N}_{p_j})]. \tag{21}$$

To mitigate the diverging nature of the D_{KL} we also considered the following transformation, noted D^{**}_{KL}:

$$D^{**}_{KL}(\mathcal{N}_{p_i}||\mathcal{N}_{p_j}) = \frac{1}{2} \times \log[D_{KL}(\mathcal{N}_{p_i}||\mathcal{N}_{p_j}) + 1], \tag{22}$$

where +1 accounts for the fact that the D_{KL} between identical cover-sources is zero. Note that the rank-correlation coefficients are not impacted by the transformations of Eq. (21) and Eq. (22), as both are striclty monotonous.

The regret matrix obtained with the LogReg detector at payload $\beta = 1.5$ bpnzAC and the D_{KL} matrix are shown in Fig. 8. Similar asymmetries are clearly visible in both matrices. On the leftmost part of the plot (up until $p = -0.6$), however, we also observe clear differences. We report results in table 4 using the three correlation measures presented in Sec. 5.1 and for the three versions of the D_{KL} of Eqs. (13), (21) and (22). As expected after visual investigation, the correlations are not as good as the ones obtained in Sec.5.1, although they still confirm that there exists a strong link between our statistical model and the regret. The differences from one detector to another, and from one payload to another, also highlight their non-negligible impact on the regret. This observation further validates that the CSM and its impact on steganalysis are two different (even if highly correlated) phenomena.

6 STATISTICAL MODEL AGAINST REALISTIC SOURCES

In this last section, we investigate the relevance of our models of the intrinsic difficulty and regret when dealing with empirical measures conducted on the realistic cover-sources described in Sec. 4.4. We also provide a visual exploration of the results based on scatter plots, to illustrate the joint distribution of the statistical and empirical measures. For the sake of focusing on the impact of the choice of sources, we perform the comparison for the logistic regressor only, with a payload of $\beta = 0.5$ bpnzAC.

Table 4: Correlation measures between the D_{KL} and the regret matrices obtained for the different detectors & payloads. For Pearson's correlation, the results are given for the 3 versions of the D_{KL}. The the two rank-correlations, the results are the same, thus only reported once.

Detector	β	P_{CC}			S_{CC}	Kendall's τ
		D_{KL}	D^*_{KL}	D^{**}_{KL}		
LogReg	1.5	0.367	0.386	0.684	**0.713**	0.505
	1.0	0.379	0.375	**0.708**	0.678	0.463
SVM lin	1.5	0.337	0.383	0.658	**0.709**	0.494
	1.0	0.359	0.374	**0.693**	0.671	0.455
SVM rbf	1.5	0.414	0.348	**0.737**	0.713	0.513
	1.0	0.443	0.341	**0.763**	0.719	0.532

Table 5: Correlation measures between the median log-determinant of the covariance matrix of realistic cover-sources and the intrinsic difficulty obtained with the logistic regressor at payload $\beta = 0.5$ bpnzAC.

Detector	β (bpnzAC)	P_{CC}	S_{CC}	Kendall's τ
LogReg	0.5	0.324	0.914	0.778

6.1 Case of the intrinsic difficulty

The intrinsic difficulties of the realistic sources are obtained following the procedure shown in Sec. 5.1. On the other hand, since we are dealing with realistic cover-sources processed through *Rawtherapee*, developing Gaussian homoscedastic noise is not easy. Rather, for each cover-source, we take 100 true developed images and their RAW counterpart, and, for each pair, compute the transition matrix $\mathbf{H}$ as in Eq. 6 before compute the associated covariance matrix as in Eq. 7. We then select the median value of the determinant out of the distributions of each cover-source as the candidate determinant. The joint distribution of the log-determinant and the intrinsic difficulty is shown in Fig. 9. A strong positive correlation is clearly visible. Furthermore, we obtain a distribution that is very consistent with our understanding of the effect of the cover-source on the intrinsic difficulty. Indeed, the more the source is denoised, the lower the intrinsic difficulty (in blue and red) becomes. Oppositely, the more sharpened they are, the more difficult they become (in orange and green).

Again, we can quantify the positive correlation, with the same three tools as in Sec. 5.1. Results are reported in table 5. They show that, despite the weak correlation captured by Pearson's correlation coefficient, the rank correlation coefficients still indicate a very strong link between the two quantities.

6.2 Case of the regret

We construct the regret matrix in the same fashion as in Sec. 5.2. As mentioned in Sec. 6.1, we can not directly leverage developed

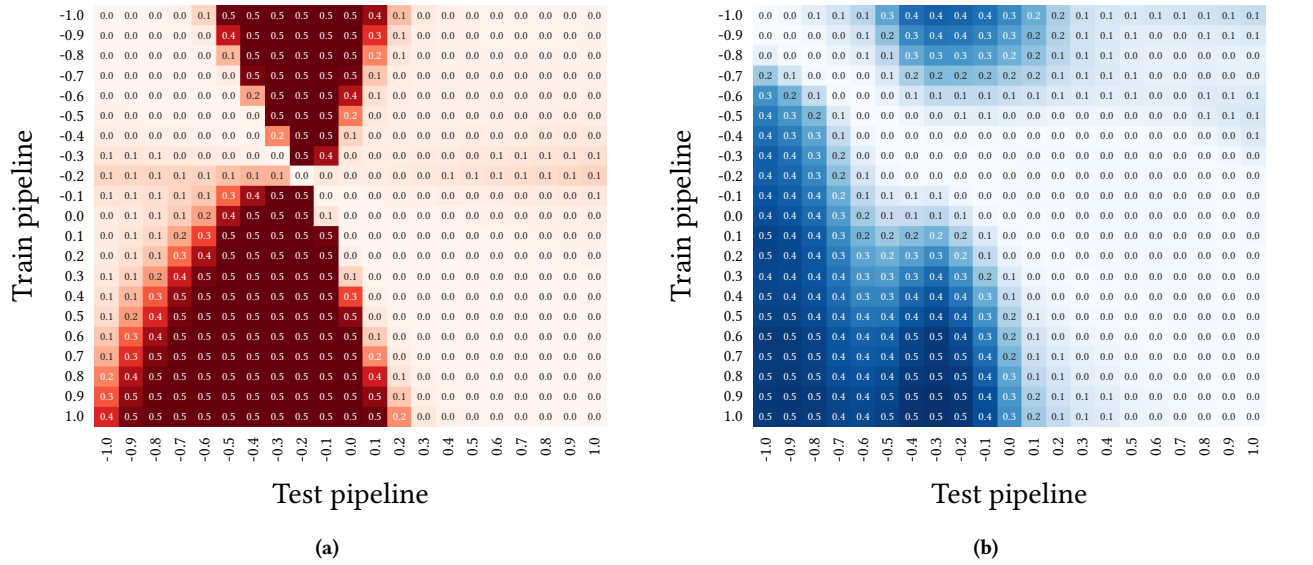


Figure 8: Modified Kullback-Leibler divergences' matrix (8a) and regret matrix (8b) of the LogReg detector at payload $\beta = 1.5$ bpnzAC.

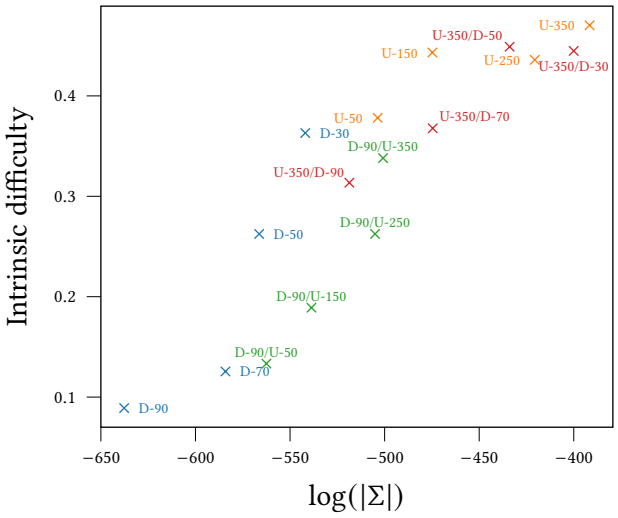


Figure 9: Scatter plot showing the joint distribution of the log determinant and the intrinsic difficulties.

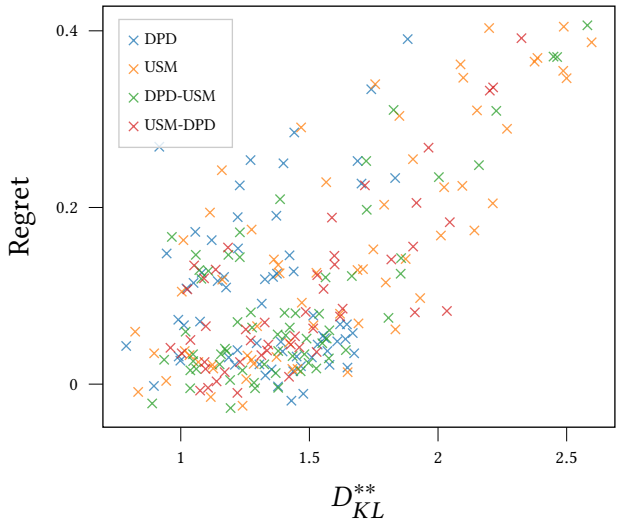


Figure 10: Joint distribution of the regret and modified Kullback-Leibler divergences between the realistic cover-sources.

noise; this time, we use an averaged covariance matrix over 100 RAW-developed pairs, estimated using the methodology presented in Sec. 2.2, as the parameter of the Gaussian distribution of the noise. Then, the D_{KL} matrix can be computed. We can finally plot the joint distribution of the regrets and the divergence, which is shown in Fig. 10 for the D_{KL}^{**}. This time, although a positive correlation is still visible, it appears less convincing.

To evaluate the relevance of our approach, we compare the correlation coefficients obtained with the three correlation coefficients, and with the three versions of the D_{KL} for Pearson's correlation in table 6. The results still indicate that there is a relation between the Kullback-Leibler divergence and the empirical regret. Although they are still promising results, they highlight the current limitations of our approach to model the regret, especially in a realistic setup.

7 CONCLUSION

In this paper, we proposed a model for the two important practical quantities of the Cover-Source Mismatch problem in steganalysis. First, we showed that the determinant of the covariance matrix of

Table 6: Correlation measures between the D_{KL} and its two modified versions with the regret matrices obtained for the realistic cover-sources with the logistic regressor at payload $\beta = 0.5$ bpnzAC.

Detector	β	D_{KL}	P_{CC} D^{*}_{KL}	D^{**}_{KL}	S_{CC}	Kendall's τ
LogReg	0.5	0.664	0.382	**0.665**	0.550	0.399

the developed noise in an image was a promising approximation of the intrinsic difficulty of a source. The high correlations between the determinant and the empirical measure of difficulty validate our approach. Furthermore, we showed that, while the choice of detector and the payload both impacted the results, the correlations were maintained.

Second, we showed that the Kullback-Leibler divergence between the multivariate Gaussian models of the processing pipelines is a promising approximation of the regret between the cover-sources.

These encouraging results are the first step towards practical mitigation strategies of the CSM. In particular, they could be used in constrained environments, where the labels are very costly since predicting the intrinsic difficulty as well as the generalization ability of a source can help design efficient training sets in holistic approaches.

REFERENCES

[1] Rony Abecidan, Vincent Itier, Jérémie Boulanger, Patrick Bas, and Tomáš Pevný. Using Set Covering to Generate Databases for Holistic Steganalysis. In *WIFS*, pages 1–6. IEEE, 2022.

[2] Rony Abecidan, Vincent Itier, Jérémie Boulanger, Patrick Bas, and Tomáš Pevný. Leveraging data geometry to mitigate csm in steganalysis. In *2023 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, 2023.

[3] Mauro Barni, Giacomo Cancelli, and Annalisa Esposito. Forensics-Aided Steganalysis of Heterogeneous Images. In *ICASSP*, pages 1690–1693. IEEE, 2010.

[4] Dirk Borghys, Patrick Bas, and Helena Bruyninckx. Facing the Cover-Source Mismatch on JPHide using Training-Set Design. In *IH&MMSec*, pages 17–22. ACM, 2018.

[5] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. Alaska# 2: Challenging academic research on steganalysis with realistic images. In *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–5. IEEE, 2020.

[6] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. Efficient steganography in jpeg images by minimizing performance of optimal detector. *IEEE Transactions on Information Forensics and Security*, 17:1328–1343, 2021.

[7] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing*, 17(10):1737–1754, 2008.

[8] Quentin Giboulot. *Statistical Steganography based on a Sensor Noise Model using the Processing Pipeline*. PhD thesis, Université de Technologie Troyes, 2022.

[9] Quentin Giboulot, Rémi Cogranne, and Patrick Bas. Detectability-based jpeg steganography modeling the processing pipeline: The noise-content trade-off. *IEEE Transactions on Information Forensics and Security*, 16:2202–2217, 2021.

[10] Quentin Giboulot, Rémi Cogranne, Dirk Borghys, and Patrick Bas. Effects and Solutions of Cover-Source Mismatch in Image Steganalysis. *SPIC*, 86:115888, 2020.

[11] Vojtěch Holub and Jessica Fridrich. Low-complexity features for jpeg steganalysis using undecimated dct. *IEEE Transactions on Information forensics and security*, 10(2):219–228, 2014.

[12] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security*, 2014:1–13, 2014.

[13] Donghui Hu, Zhongjin Ma, Yuqi Fan, Shuli Zheng, Dengpan Ye, and Lina Wang. Study on the interaction between the cover source mismatch and texture complexity in steganalysis. *Multimedia Tools and Applications*, 78:7643–7666, 2019.

[14] Tomer Itzhaki, Yassine Yousfi, and Jessica Fridrich. Data Augmentation for JPEG Steganalysis. In *WIFS*, pages 1–6. IEEE, 2021.

[15] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[16] O. Losson, L. Macaire, and Y. Yang. Comparison of color demosaicing methods. In *Advances in Imaging and Electron Physics*, volume 162, pages 173–265. EUSIPCO, 2010.

[17] Antoine Mallet, Martin Beneš, and Rémi Cogranne. Cover-source Mismatch in Steganalysis: Systematic Review. In *JIS*. eurasip, 2024.

[18] Truong T Nguyen and Soontorn Oraintara. The shiftable complex directional pyramid—part ii: Implementation and applications. *IEEE Transactions on Signal Processing*, 56(10):4661–4672, 2008.

[19] K Pearson. Notes on regression and inheritance in the case of two parents proceedings of the royal society of london, 58, 240-242. *K Pearson*, 1895.

[20] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing*, 9(3):505–510, 2000.

[21] Elena Rodríguez-Lois, David Vázquez-Padín, Fernando Pérez-González, and Pedro Comesana-Alfaro. A Critical Look into Quantization Table Generalization Capabilities of CNN-based Double JPEG Compression Detection. In *EUSIPCO*, pages 1022–1026. IEEE, 2022.

[22] Dominik Šepák, Lukáš Adam, and Tomáš Pevnỳ. Formalizing cover-source mismatch as a robust optimization. In *EUSIPCO: European Signal Processing Conference, Belgrade, Serbia*, 2022.

[23] Théo Taburet, Patrick Bas, Wadih Sawaya, and Jessica Fridrich. Natural steganography in jpeg domain with a linear development pipeline. *IEEE Transactions on Information Forensics and Security*, 16:173–186, 2020.

[24] Yassine Yousfi and Jessica Fridrich. Jpeg steganalysis detectors scalable with respect to compression quality. *Electronic Imaging*, 2020:75–1, 01 2020.

[25] Lei Zhang, Hongxia Wang, Peisong He, Sani M Abdullahi, and Bin Li. Feature-guided Deep Subdomain Adaptation Network for Dataset Mismatch in Spatial Steganalysis. 2021.

[26] Daniel Zwillinger and Stephen Kokoska. *CRC standard probability and statistics tables and formulae*. Crc Press, 1999.

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because LaTeX now knows how many pages to expect for this document.