

Point-by-point comment reply

Nicholas Asendorf and Raj Rao Nadakuditi

T-SP-13946-2012 - IEEE Transactions on Signal Processing - “The Performance of a Matched Subspace Detector that Uses Subspaces Estimated from Finite, Noisy, Training Data”

We thank the reviewers for their excellent comments and ideas. In our attached revision, we highlight the changes that we have made in blue. Particularly, we have restructured the introduction of the paper, adding specific applications that motivate our work. We have restructured the derivations of the detectors and have now included a true oracle detector. We have also reorganized the presentation of the random matrix theory results in order to facilitate a better understanding of the need for such an analysis. We would also like to highlight new Figures 2, 7a and 7b. Below is a point-by-point reply to each comment that was provided to us.

Reviewer 1

General comments

1. The motivation of real-world applications is a good one, but do the authors think that the examples with $k = 2$ are representative of real-world applications ? Could you give one (at least) example with a signal constrained to lie in a 2D space ? Are we far away from the deterministic case ?

The introduction has been expanded to include more motivating applications in the setting that we particularly have in mind. Applications like EEG, MEG, and hyperspectral imaging typically involve subspaces less than 10. The handwriting digit recognition problem that we discuss uses $k=10-12$. We provide results for various system sizes of $k=2,4,6$.

2. Maybe should the authors add a paragraph to define their notations : for example, $\xrightarrow{\text{a.s.}}$ denotes ‘converges almost surely’, c.d.f., ...

We have now defined the notation for $\xrightarrow{\text{a.s.}}$, c.d.f., $o_p(1)$, P_F , and \langle, \rangle when they appear in the paper.

3. Throughout the paper, the authors use n and m and $c = \lim_{n \rightarrow \infty} \frac{n}{m}$. Some results are presented for $0 < c < 1$ and others for $c \geq 1$. It is sometimes confusing and I would have liked the authors to say, each time, precisely, what result is used and for each result the conditions of its validity.

The proofs of the theorems in the Section V have been changed to clarify this point. When referencing other papers, we distinctly show how previous theorems apply to our problem, including what value of c the theorems hold.

4. Colored figures are difficult to read when in black and white. Is this paper destined to be published in colour ? If not, I recommend the authors to distinguish a little better the different curves.

All figures have been re-created to better distinguish between curves.

I. INTRODUCTION

1. p4 line 18 : You say "We validate our asymptotic ROC predictions in Section VII and demonstrate the optimality of using the k_{eff} informative subspace components." Is it really a general proof or just an observation of the proposed examples ?

This is not a general proof. We replaced 'optimality' with 'importance' to avoid this confusion.

II. PROBLEM STATEMENT

1. p3, A paragraph 1 : Define $U = [u_1 \dots u_k]$ with $\dim u_i = n$, and $\langle u_i; u_j \rangle = u_i^H u_j = \delta_{ij}$.
We added this information to better describe our subspace in the Training Data Model Section II.A
2. p3, A paragraph 2 : The sentence is ambiguous, we don't know what is used to estimate \hat{U} . I think a comma should be put in "... eigenvectors of $S = \frac{1}{m} Y Y^H$, and a signal ...".

We added Section II.B that is devoted to parameter estimation which should clarify this issue.

3. p3, line 39 : "a signal covariance estimate $\hat{\Sigma} \in \mathbb{R}^{\hat{k} \times \hat{k}}$ (in a manner to be specified)." Why isn't it specified here like \hat{U} ?

Section II.B now contains how we estimate Σ . This now appears with the rest of the parameter estimates.

4. p3, line 45 : Replace "We will consider two models for the test vectors. In the stochastic setting, the ..." by "We will consider two models for the test vectors. * In the stochastic setting, the ..."

An introduction to this section has now made this text easier to read.

5. p3, : I think the authors could give the expression of the covariance matrix V of y_i v.s. U, Σ and I_n and express the eigenvalues and eigenvectors of V v.s. u_i, σ_i^2 . That would be more simple, for example, to derive proposition 3.2 from Theorem 2 in [8]. It is easy to show that $VU = U(\Sigma + I_n)$ and simplifies the reader's work.

Section II.B is a section devoted to parameter estimation. In it, we show how estimates are derived from the covariance matrix of y_i . The RMT propositions have now been clarified as well to make direct connections of previous proofs to our problem formulation.

6. In a more general way, it would be a good thing if the links between your notations and those of the papers you mention were highlighted (for example when you specify a theorem the use of it in your paper is not obvious) .

We have expanded the proofs of the propositions and theorems to more clearly state how previous results apply to our problem and bridge a bit of the gap between notation.

7. p4, line 10 : For the Deterministic Model, it seems to me that x no more verifies $x \stackrel{\text{i.i.d}}{\propto} \mathcal{N}(0, \Sigma)$. So why do the authors give in Section A a definition which is not always true ? Maybe the properties of x that are not shared by the Deterministic and the Stochastic models should be given in the appropriate paragraph, not in the introductive paragraph. D is a diagonal matrix and the authors focus their attention on quadratic detectors $y^H \hat{U} D \hat{U}^H y \underset{H_0}{\overset{H_1}{\gtrless}} \eta$.

In the deterministic testing setting, x is not $\mathcal{N}(0, \Sigma)$ but instead a fixed vector of coordinate weights. However, the training data used to estimate the parameters used in the detector is still modeled as in Section II.A - the training data section. The training data model is shared by both the stochastic and

deterministic detectors. A few words were added to the beginning of the testing data model section to clarify this distinction.

III. PERTINENT RESULTS FROM RANDOM MATRIX THEORY

1. U and Σ come from the eigenvalue decomposition of the $n \times n$ matrix S defined by $S = \frac{1}{m}YY^H = \frac{1}{m} \sum_{i=1}^m y_i y_i^H$.

Yes. $E[y_i y_i^H] = U \Sigma U^H + I_n$ and so the ML estimates come directly from the eigenvalue decomposition. We added a new section on parameter estimation to make this fact more clear.

2. p5, proposition 3.1 : I have no doubt the authors are right but I don't see what result in [8] (nor in [9]) is used to obtain this result. Could you detail a little more please ?

More detail referring to specific proofs and their notation is applied to our problem has been provided.

3. p6, theorem 3.1 : is the complete proof accessible anywhere ?

We have now split the previous Thm. 3.1 into Thm. 5.1 and Claim 5.1. The complete proof of Theorem 5.1 is provided in the appendix. We relabeled that part as a claim since the proof is intricate (and long!) but is not particularly illuminating. We agree with the reviewer's sentiment that calling it a Theorem when the proof is not accessible elsewhere is less than desirable. We plan to publish that proof as a technical report. Our numerical results validate the approximation which is based on this claim.

4. p6, proposition 3.2 is easier to derive from [8] knowing that $l_v = 1 = \sigma_v^2$. I would have appreciated if this situation had been the same for other results stated before.

This same type of analysis has been added to all the proofs in Section V.

IV. FAMILY OF STOCHASTIC MATCHED SUBSPACE DETECTORS

1. p8 : if the matrix $\hat{U}^H U \Sigma U^H \hat{U}$ converges to a diagonal matrix, can't we say that the oracle detector takes, asymptotically, the form of (3) ?

We have redefined our definition of the oracle detector, which is now derived in Section III. It assumes knowledge of U, Σ, x , and k . This detector takes a similar form to (14) (the old eq. (3)), although with w instead of \hat{w} . We note the notation addition of $\hat{w} = \hat{U}^H y$ (employed by plug-in and RMT detectors) and its distinction between $w = U^H y$ (used in oracle detectors).

2. p9 142 : provably ?

The section where this text appeared has been merged with the derivation of the deterministic RMT detector. The text no longer appears.

3. p10 : Isn't it possible to put Table 1 at the end of this section ?

The table is now placed at the end of Section VI.A.

V. FAMILY OF DETERMINISTIC MATCHED SUBSPACE

1. like in V, the plug-in detector assumes that the estimated signal subspace \hat{U} is equal to the true signal subspace U . The estimate of Σ is not necessary.

That is correct

2. Finally the RMT detector and the plug-in detector (both simple unweighted quadratic detectors) are the same except the upbound of the summation. Maybe you should highlight this result a little more.

The text in the last paragraph has been changed to emphasize this point more.

3. p13 : Isn't it possible to put Table 2 at the end of this section ?

The table is now placed at the end of the Section VI.B.

VI. THEORETICAL ROC CURVE PREDICTIONS

1. p13, line 34 : "We saw in Sections IV and V that all the derived detectors were (exactly or asymptotically) of the form given by (3)." This answers to my previous question. It could be said before, even though repeating.

We agree with the reviewer that our presentation in the previous version led to these sorts of ambiguities. To resolve this, the text has been changed to say the plug-in and RMT detector forms are given by (14) (the old eq. (3)). We reiterate this point in Section III for the plug-in derivations and in Section VI for the RMT derivations. The oracle detectors, while taking a similar form except with w instead of \hat{w} , no longer applies to this specific definition.

2. p13, line 46 : "We plan to release a software implementation of the method described that will allow practitioners to generate the performance curves of the desired detectors." Can't you give an internet link here ?

This has been deleted from the text. While we still plan to release such software, as it is not entirely developed, we refrain from mentioning it.

VII. DISCUSSION AND INSIGHTS

1. A. Simulation Protocol : very clear paragraph

Thank you.

2. B. Convergence : Is the choice of $k = 2$ realistic ? What I mean is, in the real world with real applications, do you think that there exists situations where the signal subspace is of dimension only 2 (very near to the deterministic case) ?

The introduction has been expanded to provide many applications that use low rank ($k < 10$) subspaces. The results that we present are for various $k = 2, 4, 6$.

3. $c = 2$: even though previous results are proved for $0 < c < 1$, I think results are always true because proved in [15].

Yes, (now) [20,21] provide results for all c . As noted in the previous suggestion, we added more text to the proofs on the theorems to make this clearer.

4. Figure 1.a to c. Can't the curves be superposed in a fourth figure ? It seems to me that there is no gain by increasing n (even though precision is better, and that is the aim of the study). Surely, it is not important.

Figures 1a to 1c have been deleted. Instead, 2 versions of figure 1d now appear in the revised version for different values of k and c . The point of these figures is to show that the convergence of the theorems presented are effected by k and c but when these parameters are fixed, the convergence only depends on n . As $n \rightarrow \infty$ the empirical ROC curves converge to their predictions. This convergence is worse for larger k and c .

5. Figure 1.d. Why don't you keep the value $k = 2$? Are the curves you obtain not enough separated ? Having said that, I think $k = 4$ is more realistic, and results are convincing.

There are now 2 version of old figure 1d - one with $k = 2$ and one with $k = 4$. These curves give a better idea of how n , c , and k all play a roll in the convergence of the theorems presented.

6. Figure 2 : "described in" instead of "descirbed is"

Yes. This is now fixed

VIII. CONCLUSION

MY CONCLUSION

I just regret that the values of k used are very small. What happens if k increases (up to 10 for example)?

We now have simulation results for $k = 2, 4, 6$. When $k = 10$, the same asymptotic analysis holds. Simulating this leads to no additional insight than what the smaller k results reveal.

Reviewer 2

1. More content should be included in the introduction regarding the training data. What kind of applications has the authors in mind? Some references to real scenarios? Why the training data will be noisy or limited in size if one has the possibility of generating it?

The introduction now contains many more motivating applications (array processing, EEG, MEG, hyperspectral imaging, handwriting recognition). A detailed discussion of the training data is provided using the handwriting recognition problem to show how one would practically use such a dataset.

2. To improve the understanding of the new contributions of the paper the authors should clearly separate: -the effects of the subspace estimation in known detectors (plug-in and energy detectors), which is supposed to be the core of the work -the new proposed detectors (RMT-stochastic model) accounting from the subspace estimation problems

Section III.C is a new section that discusses the effects of the subspace estimation in plug-in detectors. It discusses the performance loss of these detectors relative to an oracle detector (assuming a known U). The need for new RMT detectors is now motivated better and their derivations are provided in a separate section. Results compare the RMT detectors to the plug-in ones to show that the RMT detectors do not suffer as much of a performance loss as the plug-in ones.

3. In relation with the above comment I would like to see some examples showing the degradation of the classical matched subspace detector due to estimation errors of the subspace. Most examples included by the authors are devoted to show the deviations of the theoretical expectations derived in the paper (using estimated subspace) from the empirical computations, but there are not comparisons with the case of perfect knowledge of the subspac (in the oracle detector the computation of w is also made with the estimated U).

Section III is a new section consideres this setting and uses it to motivate our work. We now derive true oracle detectors (perfect knowledge of U, Σ, k, x) for both the stochastic and deterministic testing models. We then compare the plug-in detectors to their associated oracle detector, demonstrating how finite training data affects the accuracy of the estimated subspace and causes subsequent performance loss of the plug-in detector. This is used to motivate the need for the RMT detector, which can avoid some of these performance losses as shown in the results.

4. It seems that one of the most significant contributions of the work is to enhance the importance of using k_{eff} . Some discussions about the estimation of k_{eff} from the training data will be relevant.

This is a great point and we have emphasized this more strongly. Specifically, in Section V we now discuss Algorithm 2, which we use to estimate k_{eff} . We also make a distinction throughout the text between the true k_{eff} and our estimate \hat{k}_{eff} .

5. Basically, the proposed RMT stochastic model detector implements a weighted computation of the subspace energy, alternative to the one corresponding to the plug-in stochastic model detector. But, if we allow a training data set under H_1 , why not directly estimate the best weights? Some comparison with a supervised linear discriminant in the estimated signal subspace would be convenient.

This is a great question that will be the subject of follow-up work.

6. Finally, other minor revisions:

- (a) Please, define acronyms and notation where required (for example a.s., $\langle ui, ui \rangle$, $\text{op}(1)$, and so on)

We have now defined the notation for $\xrightarrow{\text{a.s.}}$, c.d.f., $o_p(1)$, P_F , and \langle, \rangle when they appear in the paper.

- (b) Indicate from the very beginning (II.A) what are the eigenvalues of S and its relation with the sigmas

We created a new section (II.C) on Parameter Estimation. This section should make a clearer connection between the eigenvalues of the sample covariance matrix, S , and Σ .

- (c) Page 34, Line 28, second column probably

This line has since been removed from the publication.

- (d) Definition of S in V.B may be confused with definition of S in III.

Good suggestion - we used a different notation in now VI.B and also in VII.B

- (e) Is not missing an exponent in matrix sigma in equation (17) and above?

Yes we are. This is fixed in (17) and the associated equations.

- (f) First parenthesis in equation (18)

Yes. This is now fixed.

Reviewer 3

The rationale behind my Major Review recommendation is the following. The presentation of mathematical results, accessible to people at least somewhat expert in random matrix theory, is often (almost everytime) referring to previous papers where full proof details can be already retrieved. So, I don't essentially see an amount of novelty with respect to the conference paper such that full journal paper publication would be justified. My suggestion would rather be that of appropriately shortening the paper to make it a correspondence. Toward this end, it is needed that also some figures are withdrawn from the paper; I'd suggest that contour plot may eventually be those ones...otherwise the topic is appropriated and the content technically sound.

We thank the reviewer for their encouragement and the endorsement of the topic's appropriateness and its technical soundness. We have removed the contour plot and trimmed some of the figures. Our goal is to expand on the discussion in the conference paper. The analysis for the deterministic test vector setting complements the material in the aforementioned conference paper. We believe that the unification of the analysis justifies the length and consideration of the submission as a journal paper.

Reviewer 4

1. Throughout the paper it is not clear when and how k is estimated or it is assumed to be known. For instance, at the beginning of Section III it seems that estimation of U is performed assuming that k is known. However, Corollary 3.1 assumes $\hat{k} \leq k$. I understand the meaning of that and, in particular, that whenever the estimated value of \hat{k} is less than or equal to k the thesis of the corollary holds true. However, the presentation could be improved. In addition, what if $\hat{k} \geq k$?

k is assumed to be unknown. We discuss the estimation of k in Section II in the subsection on parameter estimation. For the entirety of the paper, we assume that we are given \hat{k} from a domain/model expert. As noted, \hat{k} is typically an overestimated in order to “play it safe”. The derivation of the new oracle detectors in Section III assume that k is known and this is used to provide an upper bound on detector performance. These points are reiterated throughout the paper. The paper has been significantly reorganized as well, so that Section II now presents and clarifies any notation issues as early as possible.

2. Similar considerations also apply to k_{eff} that is an important parameter as shown in the paper. I am surprised that the authors do not focus on the estimation of such parameter (they only refer the reader to [14]) and above all that they do not consider the impact of estimating it on the performance of detectors.

We thank the reviewer for bringing up this important point. We added text to discuss Algorithm 2, which we use to estimate k_{eff} . We make a distinction between this estimate \hat{k}_{eff} and k_{eff} . Because the plug-in detectors take relatively the same form as the RMT detectors, the main difference between the two is the number of subspace components each uses. We explore the effect of \hat{k} in the results section. This analysis show how the plug-in detector responds to different choices of \hat{k} and indirectly how the RMT detector responds to choices in \hat{k}_{eff} . Put a different way: the idea we want to convey to practitioners reading the paper is that RMT estimators will estimate k_{eff} and not k . In sample limited settings, k_{eff} can be smaller than k - in that regime, adopting a play-it-safe approach by making \hat{k} greater than k_{eff} is sub-optimal and leads to avoidable performance losses.

3. I do not understand the reason to construct decision schemes based on w and not on y . Notice that w requires knowledge of \hat{U} ; in particular, it follows that what the authors call the Oracle detector might not be the real Oracle (if $\hat{U} \neq U$).

Section III now presents derivations of oracle detectors (based on y) for both testing scenarios. Motivation for using w (now \hat{w}) is provided in this section. The results show that in the true oracle setting $w = U^H y$ is a sufficient statistic. However, for the plug-in detectors when U is unknown, $\hat{w} = \hat{U}^H y$ is a statistic used to simplify notation. As the RMT detector has no knowledge of U either, we derive it beginning using \hat{w} . This analysis shows where our performance improvements come from. The distinction between y , w , \hat{w} should be very clear now in the text. We re-organized the text to first present the derivations of the oracle and plug-in detectors to motivate the RMT detector.

4. Related to the previous item: if \hat{U} is a random matrix the statistical characterization given in (8) for the stochastic setting is not true; similarly for the deterministic setting.

\hat{U} is computed from the training data. \hat{U} is used in to generate $\hat{w} = \hat{U}^H y$ - our test data vectors. However, the test observations y are independent from the training dataset. So, from the perspective of the testing observations, \hat{U} is a deterministic matrix. Therefore, the statistical characterizations are valid. The text should now make the interplay between the training and testing datasets more clear.

5. At the beginning of Section VI it seems that Sigma is assumed to be known. It should be an unknown quantity though. However, if Sigma is estimated it is no longer true that, given H_i , decision statistics are weighted sums of independent, chi-square random variables.

The problem statement for ROC curve prediction does assume that Σ is known. The detectors are derived assuming Σ is unknown and so it must be estimated from data. However, when analyzing the performance, we want the analysis to be independent of the training dataset and testing dataset. Therefore, we do not have access to either and must be given Σ in order to predict ROC performance. This point should be clearer in the text.

6. The performance analysis does not seem to prove the superiority of RMT detectors with respect to deterministic ones (but for Figures 7-8). It is also important to consider meaningful values of m (in the order of 25-50). Greater values of m are not so realistic; or give an example of application where they are.

Our performance predictions are asymptotic but we apply them to the finite n and m case. We present results for various choices of m to highlight the effect of the amount of training data (see Figure 5,6,7). Figure 7 is new and highlights the importance of m . The motivating applications that we have in mind, especially hyperspectral imaging and handwriting recognition have access to a large scale dataset where m may be on the order of hundreds or thousands. The stochastic RMT detector are not superior to the deterministic one and we make no claim of that. In fact, we can obtain the stochastic RMT detector by conditioning on values of x in the deterministic setting. For larger values of x , the deterministic detector outperforms the stochastic detector, but for smaller values of x the stochastic detector outperforms the deterministic detector.

Please address the following specific comments.

1. In the introduction the authors could stress that [7] addresses direction detection; they could also cite the following reference addressing subspace detection in colored noise: “F. Bandiera, A. De Maio, A. S. Greco, G. Ricci, “Adaptive Radar Detection of Distributed Targets in Homogeneous and Partially-Homogeneous Noise plus Subspace Interference,” IEEE Transactions on Signal Processing, Vol. 55, No. 4, pp. 1223-1237, April 2007.” and the following one addressing direction detection in colored noise: “F. Bandiera, O. Besson, D. Orlando, G. Ricci, L.L. Scharf, ”GLRT-based Direction Detectors in Homogeneous Noise and Subspace Interference,” IEEE Transactions on Signal Processing, Vol. 55, No. 6, pp. 2386-2394, June 2007.”

Both of these references were added to the introduction of the paper in a paragraph providing specific applications to motivation our problem.

2. The test given in the non numbered equation before equation (11) is not a GLRT since the estimation of the unknown matrices is not based on w but on training data only.

Good point. The text was changed to read LRT and not GLRT.

3. Related to the previous item: ML estimation of σ_i^2 should be $\max(0, \hat{\lambda}_i - 1)$.

A good point. This was changed in the text, which now appears in Section II.C

4. In the expression for Λ_{plugin} of pag. 8 matrix I should probably be $I_{\text{hat}k}$.

This was changed to make the formula more clear.

5. There is a typo in equation (14).

This has been fixed.

6. There are typos in the statistical characterization of $w|H_1$ at pag. 11, in equation (17), etc. (Σ should be $\Sigma^{1/2}$).

These typos have been corrected so that Σ becomes $\Sigma^{1/2}$.

7. Please define o.w. at pag. 12.

o.w.=otherwise which is now changed in the text

8. It seems to me that $A^H S^H S A$ at pag. 12 is not always invertible.

Good point. In fact, in the derivation of the deterministic detector, none of the original matrices were necessarily invertible. We changed these matrix inverses in Section VI to Moore-Penrose pseudoinverses

9. pls define c.d.f.

c.d.f.=cumulative distribution function which is not added at first reference in VII.

10. There is a typo at the second row of Section VI.B.

Yes - the Σ has been replaced with a $\Sigma^{1/2}$.

11. There is a typo in equation (27).

The superscript 2 in the second line was removed.

12. Most of the Figures do not include the values of m and n.

All figures now include all parameters used (instead of referring to previous figures).

13. In Figure 7 the authors plot P_d vs \hat{k} as if \hat{k} is not estimated from data. Please give more details.

We assume that as practitioners, we are given \hat{k} . This may come from a model expert or may be estimated from data. In some cases, the dimension estimate may be overestimated on purpose to play it safe and ensure all signal components are captured. We added a subsection in II on parameter estimation that clarifies this point from the beginning of the paper. We also added two sentences before the description of the Figure to motivate again the reason for examining the effect of \hat{k} .

Reviewer 5

1. Section II: I suggest to add a couple of sentences before Section II.A summarizing the content of Section II.

There is now an introduction for section II which contains motivation and a summary of the section.

2. Section II.A Does the model $y = Ux + z$ refers to specific applications justifying the assumptions on the statistics of U, x, z ?

The introduction now contains many applications (array processing, EEG, MEG, hyperspectral imaging, handwriting recognition) which uses low-rank subspaces and models similar to $y = Ux + z$. Specific reference that use the exact model that we employ are referred to in Section II.A.

3. What is the role of Sigma in (2)? How is it defined?

Σ is modeled the same in (2) (now (3)) as in the training data and the stochastic setting. The role of Σ is to control the SNR for each component of the subspace. A sentence was added to this section to describe this better.

4. Section II, 4 lines after (2): I suppose P_F is the probability of false alarm...
Yes it is. This is now included in the text in the statement of Problem 1
5. Section II: "This performance prediction relies on RMT...." not clear.... provide more explanations
This point has now been explained in greater depth
6. page 5, line 3 : "Design & analyze" -> "Design and analyze"
The text has been changed
7. Section III: define the operator \langle, \rangle before (4)
This operator has been defined in II.A as part of the training data model.
8. Section III, 4 lines after (4): what's the meaning of $o_p(1)$. Is this related to the small "o" notation?
The definition has been added to the text
9. Section III, 2 lines after (5): authors only mention "Algorithm 2 of [14]" without providing explanations about it.
A brief explanation about this algorithm is now provided
10. Section III, Corollary 3.1: Is there a proof?
A proof is now provided
11. Section III, equation (7): the subscript rmt of σ_i is the acronym of random matrix theory?
Yes - we added text before (7) to clarify this.
12. Section III: proposition 3.4 seems disconnected from the previous sentences. I suggest to reorganize subsection III.B for example by adding some sentences at the top summarizing its content. Otherwise the section seems a collections of formulas
We restructured Section III (now Section V) by providing a new introduction paragraph, combining Prop 3.2 and 3.4 together, and adding more text describing why these propositions are necessary.
13. Section IV, equation (9): is there a proof?
This equation no longer appears. However, we derive a new oracle detector in Section III and make a reference to where a proof of this standard MSD may be found.
14. Section IV, unnumbered equation at the bottom of page 8: the identity matrix has size I_k ?
Yes. This has now been changed to be more clear
15. Section IV, unnumbered equation at the bottom of page 9: is there a proof? If there is no additional space in the paper I suggest to write a technical report with all the proofs.
There is not space for a step by step proof. Additional text has been added to clarify that the simplification of the LRT directly follows from the RMT parameters previously derived. The technique is similar to the plug-in LRT simplification as in Section III. We thank the reviewer for the excellent suggestion to put the details in a technical report - we intend to do exactly that.
16. Section V, unnumbered equation before (18): please provide a proof
This derivation now appears in Section III when deriving the oracle deterministic detector. A reference for the proof of this standard detector is now provided.

17. End of section V "Both of these detectors are energy detectors". This can be said before.
This was mentioned after the derivation of the deterministic plug-in detector. We kept both references to the 'energy detector'
18. Section VI: before subsection VI.A add some sentences summarizing the results obtained in the Section
A summary before VI.A has been added
19. Section VII.A: please provide a short description of Fawcett's Algorithms.
A brief description of each algorithm has now been added
20. Section VII.B What does it mean "sample starved regime"?
This corresponds to the scenario when we have less samples than dimensions, i.e. $n \leq m$. Text was added to this sentence to make it more clear.
21. Section VII.B How is the empirical bias defined? Please add some comments.
By empirical bias, we meant the difference between the theoretical ROC curve prediction and the empirically generated ROC curve. We changed the text of this sentence to make this clearer.