Author rebuttal

We thank the reviewers for their detailed and helpful comments.

First, we would like to insist on the fact that the main result is the proof that related a "good minimization of the Holder bound" to a "good approximation" of the underlying pdf. Such results would be much harder to obtain with the non-convex lower bound (VB) or even EP. The experiments we supposed to give insight, and we will extend them in higher dimension for the final version of the paper, but are definitively not the focus of the current research.

We also added a clarification about the the difference between our work, TRW and Liu & Ihler (2011), which will replace last paragraph of the paper, because it can be perceived as wrong, as reviewer 2 pointed out:

Liu & Ihler (2011) used a parameterization of the Holder inequality recursively by applying mini-bucket elimination on multiple variables sequentially. Each of these bucket elimination step can be viewed as the decomposition of the pdf into the product of two simpler pdfs with partition function having a much smaller complexity. This corresponds exactly to the application of the VH bound for discrete graphical models. Similarly, based on the proof of Minka (2005), the variational Holder bound where each factor is a possible spanning tree, does match exactly to the TRW bound of Wainwright and Jaakkola.

Benefit of convexity: We do not claim that "just because an algorithm is convex, it is automatically superior to non-convex algorithms." We just wanted to emphasize that the rich convex optimization literature allows us to use powerful algorithms as well as theoretically analyze the estimates.

Experiments: Our goal was to show that minimizing the bound also leads to reasonable parameter estimates. We agree that it would useful to include a comparison with VB and EP in the experimental section. We will include these in the final version.

Discussion: We will revise the discussion to include the references on discrete variables that do not restrict to the pairwise case.
TRW upper bound was derived using Holder's inequality by Minka (2005), but the optimization of the bound is done locally, and not globally in the way we present in our paper.

We are currently working on an extension where the VH bound is optimized using mini-batch (similar in spirit to stochastic variational inference). We believe this extension should make it scalable to large datasets.

Finally, the optimization is in general not concave with respect to λ, as most of the type-II ML estimators. We will correct the sentence related to this fact.

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	This paper proposes an approximate variational inference method for continuous variable models using Holder's inequality. It can be viewed as an extension of the method of Liu Ihler 2011for discrete variables to continuous cases. Compared to the more commonly used variational bayesian (VB), this method is convex w.r.t. the variational parameters, while VB is non-convex (this is analogue to the convex upper bounding methods such as TRW vs. the non-convex mean field). I think the proposed method is interesting, and provides some new perspective for inference of continuous models.
Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize that novelty can sometimes be relative, and we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the	I think the proposed method is interesting, and provides some useful perspective for the approximate inference of continuous variable models.

reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	
Novelty numeric score	5-One idea that surprised me by its originality, solid contributions otherwise
Technical quality. Are the results technically sound? Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did the authors clearly assess	The major weakness of the paper: (1) the experiment section does not provide clear support of the proposed method vs. other commonly used methods such as VB; (2) the authors are unclear (sometimes wrong) about the connection of the proposed method to the existing convex variational methods for discrete variables.

both the strengths and weaknesses of their approach?	
Technical quality numeric score	4-A paper that may be strong in other respects, but not technically.
Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful 5 or 10 years from now? Is this work that can be built on by other researchers?	Reasonable contribution to an important problem.
Significance numeric score	5-Solid contribution to relevant problem
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting,	2-Quality of writing is marginal to ok - it needs significant additional editing

references, etc.? It is the responsibility of the authors of a paper to write clearly, rather than it being the duty of the reviewers to try to extract information from a poorly written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	5-A good paper overall, accept if possible. I vote for acceptance, although would not be upset if it were rejected because of the low acceptance rate.
(Optional) Additional Comments to the Authors: please add any additional feedback you wish to provide to the authors here. For example, if the quality of writing in the paper is not excellent, please provide some feedback to the authors on how the writing can be improved.	===Detailed comments === The major weakness is in the experiment part: the experiments are only on a toy 1-D probit regression and a 2D gaussian integration. No comparisons with other methods (such as VB, or EP) are provided. The paper emphasized that their method do not have the zero-avoiding problem as VB, which is a valid point to raise; it would be good to show some experiments to support this.

The last paragraph of the discussion reads confusing or wrong, and should be improved & corrected --- it seems to imply that their method give something new when applied to discrete variables, but I think it does not; it should directly reduce to the general methodology in Liu Ihler 2011 in discrete cases (which, in a high level, is bounding the logZ using holder inequality), except that the current paper did not propose an implementable method for the discrete case. In the last sentence, the author *conjectured* that TRW can be treated as a special case of Holder's bound, but I think this is obviously true (except that the number of trees in TRW bound can be prohibitively large); the authors should clarify these connections clearly instead of using conjectures...

The work of Liu Ihler 2011 does not restrict to pairwise models.

For Eq(21), the holder bound makes it a convex optimization w.r.t. τ (with fixed λ), but it is not obvious to me that it is a concave optimization w.r.t. λ (with fixed τ); could you provide some illustration or proof? Also, be clear in the last paragraph of Section 8 that eq(23) is not concave w.r.t. q (since it should be often concave w.r.t. the model parameter γ).

===Minor comments ===

The objection function in Eq(21) is not defined. Is it the same as log\bar{p}(D | \lambda)?

I feel the difference between variational holder inequality and holder's inequality is minor; not sure if it is worth a new name and a new theorem.

	Check typo & grammar: .... sparse factors might be heavy tails, but we still maintaining a convex objective. ... VB and EP other the last two decades. The writing of the discussion section can be improved in general. Holder vs. H{\"o}lder: be consistent.

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	This paper presents an approach to inference in graphical models based on an upper bound to the partition function. The bound itself is well-known, but the paper makes some novel suggestions about how to use it to approximate moments and how to apply it to latent Gaussian models. However, the paper feels incomplete since it gives no practical algorithm for fitting the bound and no meaningful assessment of its accuracy.
Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize	The main idea of bounding the partition function via a parametrized Holder inequality is old, but the paper makes multiple small innovations such as a new method for approximating moments, a relation between error in the partition function and the error over the posterior, and the choice of factors when applying it to latent Gaussian models.

that novelty can sometimes be relative, and we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	
Novelty numeric score	5-One idea that surprised me by its originality, solid contributions otherwise
Technical quality. Are the results technically sound? Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly	The VH algorithm is supposedly suited to high-dimensional integrals, but the experiments only consider an easy 1D probit regression problem. Even if the method performed perfectly on this problem, it wouldn't tell us much. And the results are not particularly impressive. The exact posterior mean for the 2-point problem is 0.846 (to 3 digits), EP also gives 0.846, and the proposed method 0.815. (Note the first two numbers do not appear in the paper.) Section 5.5 claims "similar performance" to previous algorithms on probit regression. This is never shown, and according to my results above, is not even true. The evaluation does not include any similar approximate inference algorithms (such as EP or VB) nor any runtime analysis. The empirical

available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did the authors clearly assess both the strengths and weaknesses of their approach?	Bayes algorithm at the end of section 5.5 is never evaluated.

The justification for the estimate in equation (9) is very weak, and not supported by the experiments. Proposition 1 is not put to any good use in the paper.

The discussion of related work is dismissive and misleading. Section 2 claims that the upper bound is novel, when it clearly isn't (see Minka 2005, who also gave an algorithm for fitting the bound). Section 5.4 argues that EP is unstable and lacks accuracy guarantees. However, EP is known to work quite well on probit regression (the problem considered in the experiments) and the proposed algorithm does not provide a guarantee either with respect to moments. Similarly, section 8 criticizes Variational Bayes on theoretical grounds, even though KL-minimization has been shown to work well for probit regression (Nickisch and Rasmussen 2008). The attitude of this paper is that just because an algorithm is convex, it is automatically superior to non-convex algorithms.

Section 1 claims that previous work on these bounds have focused on pairwise discrete potentials. This is not true. Wainwright et al (2005), Meltzer et al (2009), and Hazan et al (2012) explicitly considered higher-order potentials. As shown in those works, there is no mathematical constraint in the bound that excludes higher-order potentials. The challenge is getting the bound to be tight for higher-order potentials, a problem which this paper has not solved.

Section 9 makes the unsubstantiated claim that VH can be sped up in the same way as VB/EP. It also suggests applying VH to discrete models, |

	which is backward since those algorithms already exist. The conjecture made in the last sentence is odd, since that is exactly what Minka (2005) showed.
Technical quality numeric score	3-Claims not completely supported, assumptions or simplifications unrealistic
Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful 5 or 10 years from now? Is this work that can be built on by other researchers?	There is no demonstration of having advanced the state of the art. If the method were shown to work well, there would many interesting future research directions. As it stands, there is not enough motivation for another researcher to pursue this approach. From a practical standpoint, the significance is severely limited by the lack of a practical algorithm for performing the optimization required for inference.
Significance numeric score	3-Not sure how this paper is relevant
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are	4-Quality of writing is excellent

there issues with typos, formatting, references, etc.? It is the responsibility of the authors of a paper to write clearly, rather than it being the duty of the reviewers to try to extract information from a poorly written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	2-A clear rejection. I vote and argue for rejection. Clearly below the standards of the conference.
(Optional) Additional Comments to the Authors: please add any additional feedback you wish to provide to the authors here. For example, if the quality of writing in the paper is not excellent, please provide some feedback to the authors on how the writing	Figure 2 is superfluous. References: "Approximations for binary Gaussian process classification" H. Nickisch and C. E. Rasmussen Journal of Machine Learning Research, 9:2035-2078, 2008

can be improved.	"Convergent message passing algorithms-a unifying view" T. Meltzer, A. Globerson, and Y. Weiss UAI 2009 "Tightening fractional covering upper bounds on the partition function for high-order region graphs" Tamir Hazan , Jian Peng , Amnon Shashua UAI 2012

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Paper summary. In this section please explain in your own words what is the problem that the paper is trying to address, and by what approach. In what area lie the main accomplishments of the paper. Optionally, a summary of your opinion, impact and significance could be included too.	The paper considers Holder's inequality to derive a novel variational method that provides an upper bound of the partition function. Authors first show in Proposition 1 the usefulness of the bound for variational inference and then derive the method. The bound is promising, since it is convex and has nice theoretical properties. The difficulty lies then on integrating the two factors of Holder's inequality, which is not too difficult in the fully factorized case and Gaussian potentials considered in the paper. The method is illustrated using simple regression problems with Gaussian (or truncated Gaussian) priors. The paper is very well written and technically sound. The superiority of this bound in relation with existing bounds in described qualitatively. The toy examples do not show clear evidence of the superiority of the proposed bound wrt

	existing ones. Also, it would have been desirable to show results on more general cases beyond Gaussian integration and elaborate more on the general case with more than 2 factors (outlined in section 7).
Novelty. This is arguably the single most important criterion for selecting papers for the conference. Reviewers should reward papers that propose genuinely new ideas, papers that truly depart from the "natural" next step in a given problem or application. We recognize that novelty can sometimes be relative, and we ask the reviews to assess it in the context of the respective problem or application area. It is not the duty of the reviewer to infer what aspects of a paper are novel - the authors should explicitly point out how their work is novel relative to prior work. Assessment of novelty is obviously a subjective process, but as a reviewer you should try to assess whether the ideas are truly new, or are novel combinations or adaptations or extensions of existing ideas.	Holder's inequality has already been introduced in the literature for approximate inference in discrete graphical models and this acknowledged by the authors. The approach in this paper is significantly novel in this respect since it provides novel theoretical insight (prop 1) and extends it applicability to the continuous case.
Novelty numeric score	5-One idea that surprised me by its originality, solid contributions otherwise
Technical quality. Are the results technically sound?	The technical quality of the paper is excellent. Authors are aware of related relevant literature.

Are there obvious flaws in the conceptual approach? Are claims well-supported by theoretical analysis or experimental results? Did the authors ignore (or appear unaware of) highly relevant prior work? Are the experiments well thought out and convincing? Are there obvious experiments that were not carried out? Will it be possible for other researchers to replicate these results? Are the data sets and/or code publicly available? Is the evaluation appropriate? Did the authors discuss sensitivity of their algorithm/method/procedure to parameter settings? Did the authors clearly assess both the strengths and weaknesses of their approach?	Experiments illustrate the approach, but are rather limited.
Technical quality numeric score	5-Technically adequate for its area, solid results
Significance. Is this really a significant advance in the state of the art? Is this a paper that people are likely to read and cite in later years? Does the paper address an important problem (e.g., one that people outside UAI are aware of)? Does it raise new research issue for the community? Is it a paper	The significance of the paper is high and the approach is promising. It is difficult to assess the impact of the method given the lack of comparison with existing bounds.

that is likely to have any lasting impact? Is this a paper that researchers and/or practitioners might find useful 5 or 10 years from now? Is this work that can be built on by other researchers?	
Significance numeric score	4-Reasonable contribution to a minor problem
Quality of writing. Please make full use of the range of scores for this category so that we can identify poorly-written papers. Is the paper clearly written? Does it adequately inform the reader? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? It is the responsibility of the authors of a paper to write clearly, rather than it being the duty of the reviewers to try to extract information from a poorly written paper. Do not assume that the authors will fix problems before a final camera-ready version is published - there will not be time to carefully check that accepted papers are properly written. It may be better to advise the authors to revise a paper and submit to a later conference, than to accept	4-Quality of writing is excellent

and publish a poorly-written version. However if the paper is likely to be accepted, feel free to make suggestions to improve the clarity of the paper, and provide details of typos under 11.	
Overall Numeric Score for this Paper:	5-A good paper overall, accept if possible. I vote for acceptance, although would not be upset if it were rejected because of the low acceptance rate.
(Optional) Additional Comments to the Authors: please add any additional feedback you wish to provide to the authors here. For example, if the quality of writing in the paper is not excellent, please provide some feedback to the authors on how the writing can be improved.	- Eq (9) \hat missing in $p_1(Z)$ and $p_2(Z)$ - Last paragraph before section 5: / missing between \gamma2 and \Phi(:,\tau) - paragraph right after eq(8) is a bit confusing, since \hat{p}_1 and \hat{p}_2 are not moments. I think it's more clear stating "depending whether \hat{\alpha}_1 is smaller or ..." directly after eq(8). - eqs(7,8) also have some missing hats in \alpha_1 and \alpha_2