Reviewer #1: We thank the reviewer for his/her review and suggestions. **Practicality and claims:** In term of practical use of the EM training we agree and explicitly acknowledged that it is computationally demanding. However, the obtained analytical results hold for any depth/with and nonlinearities (as long as they are piecewise affine); the results of the paper are thus general and can be used to gain in depth theoretical understanding of generative networks and their learning dynamics (from the explicit M step). The obtained analytical forms allow (i) to better design VAEs (now knowing the target posterior that variational inference approximates), (ii) to guide the design of variational distributions (for example favoring full covariance and multimodal posterior) as well as (iii) interpreting the learned parameters from the M-step. Those insights are gained despite the computational limits of the practical EM learning as they rely on the analytical derivations only. We will add further analysis and discussions on this. On that note, we also believe that tremendous future work can also be done to derive faster EM learning leveraging the obtained formula either by providing principled approximations of the per region Gaussian integrals or by approximation of the partitions; we believe that this paper is only the beginning of such research directions. **Direct log-likelihood maximization:** As in any missing variable model (here z is unobserved, only x is observed) one can not directly minimize the negative log likelihood and must first infer z. EM is one common strategy to do so based on the posterior $p(z|x)$; once z is inferred, one can then do the maximization of the (now estimated) log-likelihood. **Adding cost analysis of the algorithm:** We will add in the appendix exact computation times and further details for each of the experiment and different architectures for the EM learning as well as each step involved (partition finding, region triangulation, per region integration). **Comparison to VAE and figures:** In Fig. 4 we compared the negative log-likelihood of both models. While a VAE can be trained using the standard variational inference strategy, we evaluate its NLL after training and compare with the generative deep network trained with EM. We will explicit this in the caption. We will also make the legend and labels clearer in the figures.

Reviewer #2: We thank the reviewer for their appreciation of the paper. We will correct the typos and explicit the pseudo-code as well as providing exact link with the implementation. **Computational limitations:** Indeed, the current analytical EM learning is computationally demanding, we believe that future work can be done on this point by (1) providing analytical form of gaussian integration on a convex polytope (this would remove the need of triangulation and then inclusion-exclusion formula) or by (2) providing principled approximation of those integrals. Note that our main contributions are the analytical derivations of the probability distributions and EM formula, the practical EM learning demonstrates the usefulness of those derivations. **Gaussian prior and piecewise affine nonlinearities:** The review is correct; this only applies to Gaussian prior and output distributions and with DN employing spline operators like ReLU, leaky-ReLU, abs. value, ... which includes a large part of current generative network architectures. Also, the proposed method (with exact partition and per region derivation) can be employed to different distributions as long as they are conjugate priors. We will add this note in the paper. **Constant covariance:** Indeed, this case covers the practical cases of training in current generative models, however more general cases could be considered and even different distributions. We believe that the proposed methodology (per region derivation) provides a general framework and as long as the prior and output distributions are conjugate priors, analytical forms should be obtainable. We will add this discussion in the paper.

Reviewer #3: we thank the reviewer for their careful review and appreciation of the paper. **Previous work:** We thank the reviewer for this relevant reference (which we denote by ICML2019 thereafter). ICML2019 relates linear VAEs to PPCA and propose a mode approximation of the posterior in turn producing a novel type of VAEs (Laplacian VAEs). ICML2019 also provides insights into the manifold geometry (piecewise affine) of ReLU VAEs. We will add this reference and detailed review in the background section. However we believe that none of our contributions is over-shadowed by ICML2019 since: (i) we extend the PPCA link of linear VAES to nonlinear VAEs resulting in MPPCA; (ii) we extend their geometrical insights to piecewise affine nonlinearities (not only ReLU) which consequently also allow to apply ICML2019 approximation methods to a broader class of VAEs; (iii) in ICML2019, no analytical (explicit) form is given for the probability distributions of a nonlinear VAE as the motivation of the paper was to provide a mode approximation based on a linearization of the network to tackle large scale tasks. We will also discuss the paper approximation method in the future work section as such posterior mode estimation could be employed and potentially improved with the proposed distributions. **Lemma 2 ReLU assumption:** you are correct, Lemma 2 holds for more general DGNs (as long as there is no surjectivity), we will add this note and discuss such cases in the paper.

Reviewer #4: We thank the reviewer for his/her appreciation of the paper and we agree that providing exact methods even with demanding computational cost is crucial to exactly measure the impact of current approximation methods in VAEs. **Computational complexity discussions:** indeed, the computational bottleneck comes from the number of regions that then need to be triangulated. We will add computational time of each of the involved steps in the appendix: (i) computation of the partition, (ii) triangulation of each region (on average) and (iii) integration on a region. We will provide those statistics for the few different topologies that were used in the paper. Concerning the rate of growth of the number of regions in a real network, we will add citations to the following papers: "Complexity of Linear Regions in Deep Networks", "On the Number of Linear Regions of Deep Neural Networks" and "A Spline Theory of Deep Networks" with discussions.