

1 We thank the reviewers for the insightful and thoughtful comments. Below we address the concerns raised.

2 **Reviewer #1:**

3 **Different losses.** On hindsight we completely agree that results for more losses are valuable and, given the chance,
4 we will add such supplementary figures for both stock and river discharge datasets. As can be expected, this actually
5 improves our results since we are able to effectively choose the right loss using validation.

6 **Positioning with respect to deep learning.** Briefly, the type of covariance matrices we discuss can be useful as parts
7 of a deep network. Specifically, a recent paper in CVPR demonstrates the use of Generalized Gaussians for covariance
8 pooling in CNNs. Gaussian CRFs have also been used for regression over a deep representation, we are currently
9 experimenting with their robust counterparts. Given the chance, we will add this to the paper.

10 **Reviewer #2:**

11 **Global optimality to efficient optimization.** Given global guarantees regarding critical points, any standard algorithm
12 that is guaranteed to reach such points can be used to solve the problem. This gives us great freedom and in particular,
13 allows us to choose an efficient such algorithm that arrives at a critical point most quickly. In particular, second
14 order algorithms are known for their efficiency. To make the above claim concrete and demonstrate the efficacy of
15 our algorithm, we add a plot below comparing convergence time between MM on elliptical problems and a natural
16 contender, namely gradient descent. On hindsight, we should have included this in the original manuscript and hope
17 that the reviewer will give us a chance to do so.

18 **Novelty relative to existing literature.** Whether or not structured problems have bad local minima, is an open question
19 in the literature on robust covariance estimation. Our work gives an answer to this question, under the appropriate
20 realizability and distributional assumptions. To better contextualize this w.r.t results on unstructured methods we note
21 that existing results on unstructured models broadly fall into one of the following two categories:

- 22 • Efficient algorithms that provably solve the problem based on closed form updates (e.g. refs 21, 25 in the
23 paper). It is unclear how these can be generalized to the structured scenario.
- 24 • Show properties like geodesic convexity of the unstructured loss (e.g. ref 31). In general, imposing linear
25 constraints on geodesically-convex optimization can introduce bad local minima, hence the need for our result.

26 Given the chance, we will add this to the related works section. We do note that, technically, we do make use of tools
27 from the unstructured case, e.g. via lemma 1, which extends such a known result.

28 **Reviewer #3:** We appreciate the useful suggestions. Given the chance, the following changes will be introduced:

- 29 • Add a detailed derivation of the MM algorithm. This relies on $\rho(\cdot)$ being concave (Assumption 1 implies this),
30 and plugs its linear approximation into the majorization part of the generic MM algorithm.
- 31 • As noted for reviewer 2, we will add runtime comparison results (one such graph is included below). Generally
32 speaking, a very small number (~ 5) of MM iterations is needed.
- 33 • Fix the wrong phrasing “commutes with \mathbf{I} ” in line 149. To arrive at corollary 1, it is enough to use the property
34 of the eigenvalues implied by lemma 1 to gather: $\Sigma^\rho(\Gamma(\mathbf{w})^{\frac{1}{2}}\mathbf{z}) = \mathbf{I} \Rightarrow \Sigma(\Gamma(\mathbf{w})^{\frac{1}{2}}\mathbf{z}) = c\mathbf{I}$. Commutation is
35 only required for the second equality in equation 14.
- 36 • Separate proof of lemma 1 into parts. Regarding the inner expectation being diagonal: v_j being odd implies
37 that the identity function is odd (i.e. $f(-x) = -f(x)$). Diagonality follows because for $i \neq j$, upon fixing all
38 coordinates other than \tilde{v}_j , the function $g(\tilde{v}_j) = \int \tilde{v}_i \psi(\tilde{\mathbf{v}}^\top \lambda \tilde{\mathbf{v}}) p(\tilde{\mathbf{v}}) \prod_{k \neq j} d\tilde{v}_k$ is even. Then an off-diagonal
39 element is given by $\int_{-\infty}^{\infty} \tilde{v}_j g(\tilde{v}_j) d\tilde{v}_j$. Integrating an odd function over the reals, we get a 0.

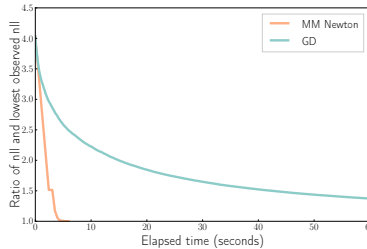


Figure 1: Runtimes of Gradient Descent and MM with Newton CD on stocks data with a robust loss. The y-axis is the ratio between the objective at time t and the lowest overall observed objective.