We thank all reviewers for their thoughtful comments and suggestions. We provide our feedback below.

Reviewer 1 "The contribution is relevant and timely." Thank you for your encouraging comments. "Many seminal papers on SLOPE-like regularizations are not cited." We acknowledge the need for a more extensive section on the previous literature and will revise the introduction accordingly. "[. . .] details on the convergence" Convergence is obtained when the duality gap as a fraction of the primal is less than 10^{-5} and the relative level of infeasiblity (see the appendix of Bogdan et al. 2015 [3]) is less than 10^{-3}. We will include this in the revision. Regarding the computational cost of the rule, the rule sorts the gradient but does not solve the prox, which makes the cost slightly lower than the cost of a gradient step. We will clarify this in the revision. Thank you for pointing the typos in the proof. They will be corrected together with other observed inconsistencies.

Reviewer 2 "The work is well presented, motivated, application is interesting and speedups are demonstrated." Thank you for your positive feedback. "constructing a path of decreasing λ [. . .]". We will provide the suggested reference. "your rule is useful in the sequential setting but cannot be applied [. . .] with a single lambda" It is possible to use the rule non-sequentially since the gradient for the null model is always available. We will clarify this in the revision. "the true solution is never available [. . .]" Actually, our algorithm calculates the exact solution (to numerical precision) at each step, so the exact gradients at the previous steps are known. Please note that at each step we check the KKT conditions and, if needed, recalculate SLOPE after adding predictors removed by the rule. We observed that such corrections are rarely needed in practice. "More than being piecewise linear, you need its slope to be bounded by 1, don't you?" At the intervals where the path is linear the unit bound is trivially satisfied (see the similar reasoning for the lasso). "References" Thank you for a large set of references. We will extend the review section accordingly. "I would like to see the efficiency of your rule in this setting". We have extended the experiments to analyze the effectiveness of the rule with different path lengths (and thus coarseness). For the arcene data set and OLS model, for instance, the average proportion of eliminated predictors (out of the total) is 0.74, 0.92, 0.96 for path lengths of 20, 50, and 100 respectively. Section 3.3.1 will be updated and extended accordingly. "it is better to have a violating [. . .]". We agree and will revise. "For the Lasso, there always exists a solution with support of size at most n". This is not the case with SLOPE. Due to clustering, SLOPE can return even p nonzero coefficients (assuming some of them are equal to each other in absolute value) (cf. `https://arxiv.org/abs/2004.09106`). "[. . .] you may need to run your simulations with larger n". We have updated the results with new datasets (see the included table). "can you explain how you break ties and why they don't matter?" SLOPE clusters variables and averages the penalty coefficients over the ties (see Bogdan et al. [3] for SLOPE prox). "why should the screened set [. . .]" We agree and will clarify this passage in the revision. "L184 could plot [. . .]" We agree but are regrettably not able to add such a plot due to space constraints. We will also take into account other editorial suggestions and include suggested references.

				time (s)	
dataset	model	n	p	no screening	screening
dorothea	logistic	800	88119	845	14
e2006	OLS	3308	150358	43335	4874
news20	multinomial	200	62061	2101	133
physician	poisson	4406	25	34	34

Reviewer 3 "The authors derive a screening rule for SLOPE, which is both novel and impactful." Thank you; we appreciate the supportive feedback. "[. . .] the gradient estimate be arbitrarily far away from its real value". This is true in theory, but unlikely in practice. In fact it has been empirically shown both for the lasso and SLOPE that in most cases the unit bound is conservative. In our article we show that violations are rare for typical data sets. "Shouldn't guarantees be studied [. . .]" We think there might be a misunderstanding here. If Algorithm 1 is used with a true gradient than it returns the true support. Our rule relies on replacing the true gradient with an estimate based on the unit bound and Proposition 2 specifies conditions under which Algorithm 1 returns the superset of the support. "[. . .] biased gradient estimation." Unless any of the mentioned events occurs the gradient is linear and bounded by the unit bound. Please note that the gradient at the previous step is known.

Reviewer 4 "This paper aims to derive a screening rule for SLOPE, which is important in sparsity learning [and] [. . .] easy to implement." Thank you for the remarks. "The novelty is limited [...]" We respectfully disagree. Developing screening rules for SLOPE is notoriously difficult due to the non-separability of the penalty; ours is nevertheless one of the first attempts to do so. We are aware of only one article about a safe rule for SLOPE published in ICML (after we submitted). "The motivation is unclear. [. . .]" It is not clear that it's possible to extend GAP safe and EDPP rules to SLOPE. Since our submission, Bao et al. (`https://arxiv.org/abs/2006.16433`) have published a paper describing the first safe rule for SLOPE. Yet due to the non-separability of the penalty, this rule requires iteratively screening predictors during optimization, which means predictors cannot be screened prior to fitting, which we think highlights the difficulty in developing screening rules for SLOPE. "Experiments on real datasets are insufficient." We agree. See the included table. "The authors may want to detail the differences [. . .]." The difference between the lasso and SLOPE is that SLOPE has a non-separable penalty, which leads to a more complicated subgradient. In this paper we derived a form of the subgradient that enables us to efficiently generalize the strong rule for the lasso to SLOPE.