We thank the reviewers for their valuable comments and recommendations for the improvement. Overall, it seems that the reviewers R1 and R2 found our contributions significant, but had questions about presentation of our theoretical results (R2), comparison of our theoretical results to the existing methods and random sampling (R1, R2), and practical comparison to SSSC (R1). R3 asked about significance and choice of datasets. R3 also suggested to shrunk review of SSC; we will do so and use space to address concerns of R2 by clarifying theoretical results.

Significance (R3). We would like to point out that our contributions are twofold. We present first linear time algorithm which directly solves SSC objective function and show it outperforms other large-scale SSC motivated algorithms. Besides algorithmic contribution, we provide novel theoretical result of SSC for limited number of subsamples. Our analysis gives theoretical guarantees needed for the success of SSC in the setting of limited number of subsamples.

Clarification of theoretical result (R1, R2). We highly appreciate that the reviewers pointed out to improve presentation of theoretical analysis. Due to the space constraints, pseudocode of the algorithm was included in the Appendix, but we will correct this and present key steps in the paper. We agree with R2 that this will help in understanding theoretical results. R2 correctly pointed out that the randomness in the model comes from random selection per iteration in the algorithm. We will clarify this point and relate to the probability in the statement of the theorems. Furthermore, number of iterations T is directly and linearly connected to the runtime of the algorithm. At the same time, T has an interpretation as an upperbound on the cardinality of S. We will explain this in the final version. Affinity in Definition 6 is defined in terms of principal angles in references [12, 13] in the paper; we will cite [12, 13] for alternative definition.

Compatibility to the existing results and stochastic variant (R1, R2). The setting of Theorem 1 is comparable to the setting considered in the analysis of SSC-OMP [21], in which authors assume that data is noiseless and each pair of subspaces is arbitrarily intersected. Authors further assume that all data is used as a subsample so that it coincides with the original SSC, whereas our analysis first succeeded to show that smaller number of subsamples is sufficient to guarantee SDP. The setting of Theorem 2 is adopted from the setting of noisy SSC [12], where data is randomly drawn from each subspace. Our analysis relies on the notion of persistent inradius, which is a measure originally introduced in our work. When inradius and persistent inradius coincide, the SSC results can be reproduced as a special case. R1 was interested about noisy SSC setting. Our theoretical result does not assume noise, but empirically we perform well in the noisy setting of real-word data. We will add part about the compatibility to the existing SSC theoretical results.

R2 asked about the comparison to the fully stochastic variant. Our theoretical analysis shows that, in the case of random sampling, T subsamples are needed to satisfy SDP, while S^5C needs a smaller number of subsamples to satisfy SDP. We will clarify this in the revised version.

Table 1: Relation to the existing theoretical analyses

	subsample	noise	data model	measure for subspaces	condition on data
Theorem 2 in [21]	no	no	deterministic	incoherence	large inradius
Theorem 2.8 in [12]	no	yes	semi-random	affinity	large number of data
Our Theorem 1	yes	no	deterministic	incoherence	large persistent inradius
Our Theorem 2	yes	no	semi-random	affinity	large number of data

Choice of datasets (R3). Yale B, Hopkins 155 and MNIST are the most benchmarked subspace clustering datasets. We did not compare performance only on Hopkins 155 dataset, but per reviewer's question we now include Hopkins dataset. We report average clustering error across 155 sequences (Table 2), after carefully tuning parameters of all algorithms. Results show that S^5C significantly outperforms all other large-scale methods. We will include these results in the final version. Among other datasets, only Devanagari has not been previously used for subspace clustering. However, the use of this dataset is justified since it is a large-scale dataset similar to MNIST, but the problem is even harder: instead of handwritten digits the task is to recognize handwritten letters.

Table 2: Average clustering error across 155 sequences of Hopkins 155

Dataset	**Nyström**	**AKK**	**SSC**	**SSC-OMP**	**SSC-ORGEN**	**SSSC**	**S^5C**
Hopkins 155	21.8	20.6	**4.1**	23.0	20.5	21.1	**15.8**

Experimental comparison to baselines (R1, R2, R3). R3 expressed concern that we report better performance than the methods that exploit all samples. Reviewer might be confused by the names of the methods (SSC-OMP, SSSC), but the only method that exploits all samples is SSC and it performs better than any other method (including ours). We will change name of SSC-ADMM to SSC to avoid ambiguity.

R1 raises a point about the practical benefits of our method compared to SSSC. Our method has significantly better performance than SSSC, achieving even 14.5% better average performance across all datasets (Table 1 in the paper). To demonstrate additional benefits, we show parameter sensitivity of SSSC and S^5C on Devanagari dataset (Figure 1). S^5C outperforms SSSC for all values of sparsity regularizer λ and number of subsamples. Furthermore, when the number of subsamples is increased S^5C expectedly achieves lower clustering error, while SSSC does not exhibit such behavior. Similar behavior can be observed on other datasets.

Finally, R2 was interested whether distributed computing of SSC can prevent the problem. Since the original SSC suffers from $O(N^3)$ time complexity, the effect of distributed computing is limited. On the other hand, since our algorithm is $O(N)$, we can deal with linearly larger number of datapoints as available computational resources increases.

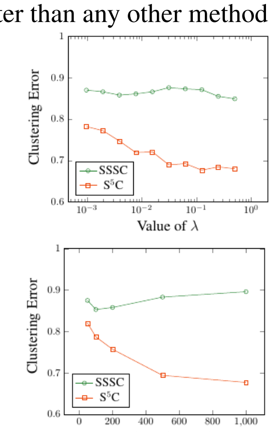

Figure 1: Clustering error of S^5C and SSSC