Search Space Range

Our search space for the LMO parameters is listed below; this setting can also be found in lines 131-139 in experiments/lmo_noise_grid_searching.py. We note that LMO-noise search is an "one-time offline process" whose runtime is proportional to the number of parameters of the search range (around 1 hour in our experiments). We will add a discussion in the experimental setting.

```
Search Space = {
    "a1": [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9],
    "a3": [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9],
    "a4": [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9],
    "G_theta_k": [(1,2), (2,2), (3,2), (5,1), (9,0.5), (7.5,1), (0.5,1)], # (G_theta, G_k)
    "E_lambda": [0.1, 0.5, 1, 5],
    "U_b_a": [(1,0), (2,1)], # (U_b, U_a)
    "delta": [1e-10],
}
```

Furthermore, we have added the runtime for fine-tuning using LMO-DP and DP-SGD in the following tables as a supplemental result to Figure 5 and will update them in the paper.

Runtime (of fine-tuning)

We include the overall runtime (in hours) for each setting in Figure 5 to reach the same reasonable accuracy rather than reach the highest accuracy (since DP-SGD cannot reach a high accuracy for small epsilon).

Dataset	MNLI-m				QQP			
Overall epsilon	0.14	0.16	0.31	0.54	0.14	0.16	0.32	0.55
	(a)				(b)			
BERT-base + DP-SGD	15.3	4.4	2.3	2.1	17.2	9.6	3.7	3
RoBERTa-base + DP-SGD	19.2	4.3	1.5	1	34.9	13	6.5	5
BERT-base + LMO-DP (Ours)	2.2	1.4	1	0.9	2.8	1.7	1.4	1.3
RoBERTa-base + LMO-DP (Ours)	1.4	1.2	1	0.8	4.5	2.7	0.5	0.5
	(c)				(d)			
BERT-large + DP-SGD	36.6	8.7	2.3	1.8	41.5	24	12	11.8
RoBERTa-large + DP-SGD	33.6	7.3	3.5	4.6	76.8	25.9	12.5	0.5
BERT-large + LMO-DP (Ours)	4.2	2.4	2.3	2.2	11.5	9.2	1.1	1.1
RoBERTa-large + LMO-DP (Ours)	4.4	2.5	2.5	2.4	0.5	0.5	1	1