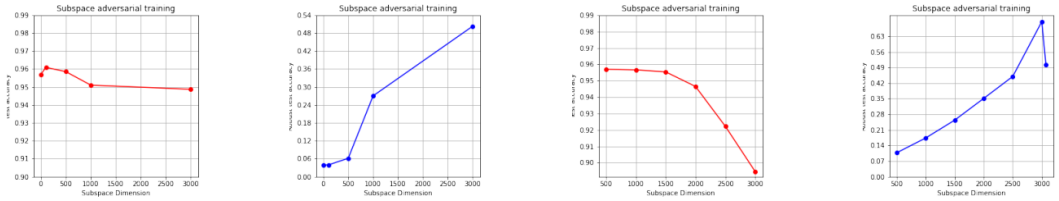
主要结论：

1. 先前的假设是对抗样本都不在数据分布的流形上（即使非常接近），但是实际上是存在分布于流形内部的对抗样本的。
2. 对抗训练会集中于off-manifold方向上的对抗样本，而忽视了on-manifold的：On-manifold attack比Off-manifold attack在ERM和AT上的atk rate都高。
3. **Robustness：**子空间AT的robust acc比普通AT的robust acc要差，只有当子空间维度逐渐取满的时候robust acc才会上去

   Generalization：这篇文章给出的关于泛化性能的分析是在Appendix B.2里的。他拿clean acc来分析泛化性能，得出来了on-manifold有助于泛化的结论，我认为欠妥。



(a) (b) (c) (d)

4. Theorem 5.2说明了，如果让q=d，情况退化到普通AT，且数据是在低维流形上的话（X至少在一个维度上方差为0，即$\lambda_{min} = 0$）的话，那么excess risk就可以取到无穷（因为由theo 5.2的结果，excess risk项里，扰动所在空间的最小的那个特征值在分母）。
5. Theorem 5.4说明了：eigenspace AT的最优解（就是AT的优化目标的那个min-max问题）