第十周周汇报

WGAN的前作中对原始GAN存在的问题作了严谨的数学分析。原始GAN问题的根源可以归结为两点，一是等价优化的距离衡量（KL散度、JS散度）不合理，二是生成器随机初始化后的生成分布很难与真实分布有不可忽略的重叠（上升到高维时）从而导致在判别器优化的很好的时候生成器会有梯度消失的情况发生。对于生成分布与真实分布的重叠问题在WGAN前作中粗略的提及到改进的办法，那就是对生成样本和真实样本加噪声。

原文对此处描述为：通过对生成样本和真实样本加噪声，使得原本的两个低维流形“弥散”到整个高维空间，强行让它们产生不可忽略的重叠。而一旦存在重叠，JS散度就能真正发挥作用，此时如果两个分布越靠近，它们“弥散”出来的部分重叠得越多，JS散度也会越小而不会一直是一个常数，于是生成器梯度消失的问题就解决了。在训练过程中，我们可以对所加的噪声进行退火（annealing），慢慢减小其方差，到后面两个低维流形“本体”都已经有重叠时，就算把噪声完全拿掉，JS散度也能照样发挥作用，继续产生有意义的梯度把两个低维流形拉近，直到它们接近完全重合。

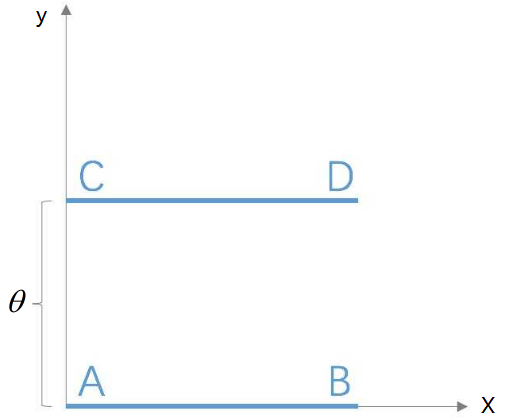
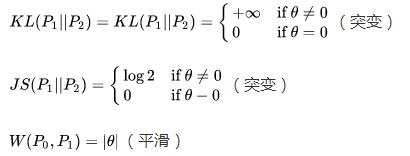
这样的加噪声对训练模型真的有大幅度提高吗？显然这样做是不够的，文章中解释为加噪声后JS散度的具体数值受到噪声的方差影响，随着噪声的退火，前后的数值就没法比较了，所以它不能成为和距离的本质性衡量。在WGAN前作设置的巨大引子的前提下，WGAN便在随后问世了，文章从本质上解决了对生成分布和真实分布衡量问题。

图1.二维下两均匀分布图

图1为二维空间中的两个分布和，在线段AB上均匀分布，在线段CD上均匀分布，通过控制参数可以控制着两个分布的距离远近。在此条件下分别利用KL散度，JS散度，Wasserstein距离衡量两个分布之间的距离，结果如下，具体推证见附录。

KL散度和JS散度是突变的，要么最大要么最小，Wasserstein距离却是平滑的，要用梯度下降法优化这个参数时，前两者根本提供不了梯度，Wasserstein距离却可以。类似地，在高维空间中如果两个分布不重叠或者重叠部分可忽略，则KL和JS既反映不了远近，也提供不了梯度，但是Wasserstein却可以提供有意义的梯度，这就是Wasserstein距离衡量两个分布之间距离的优势所在。

由此思想，作者通过一系列数学演算后得到和的Wasserstein距离公式。

 (1)

具体的推证我也有在附录中简单提及这里不做展开了，函数的Lipschitz常数为，把f用一个带参数的神经网络来表示。对于这个限制，这里其实不关心具体的K是多少，只要它不是正无穷就行，因为它只是会使得梯度变大K，并不会影响梯度的方向。作者采取了一个很工程的思想，就是限制神经网络的所有参数不超过某个范围，比如，此时关于输入样本x的导数也不会超过某个范围，所以一定存在某个不知道的常数K使得的局部变动幅度不会超过K，Lipschitz连续条件得以满足。具体在算法实现中，只需要每次更新完后把它clip回这个范围就可以了。此时只需要增大就会近似真实分布与生成分布之间的Wasserstein距离（忽略常数倍数K）。需要注意的是原始GAN的判别器做的是真假二分类任务，所以最后一层是sigmoid，但是现在WGAN中的判别器做的是近似拟合Wasserstein距离，属于回归任务，所以要把最后一层的sigmoid拿掉。

图2为WGAN的算法实现图。

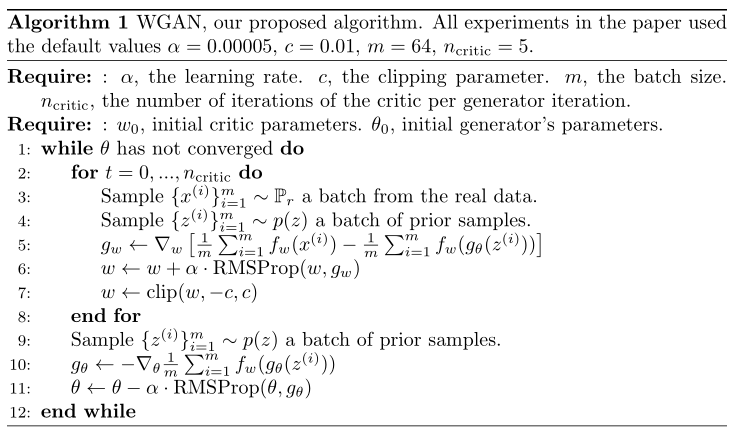


图2.WGAN算法流程图

本周的另一篇研读的论文是LSGAN，仔细搜索和阅读论文后才发现LSGAN有两篇而且这两篇的差距还有点大，我之前一度是搞混的，[least square GAN](https://arxiv.org/abs/1611.04076) 称作最小二乘法GAN和 [loss sensitive GAN](https://arxiv.org/abs/1701.06264)损失敏感GAN，两者有很大的差别。目前比较火热的是损失敏感GAN，这篇文章是可以和GAN，WGAN相提并论级别的了，文章是由中国的齐国君创作，他现在就任美国中佛罗里达大学计算机系助理教授，他也是在[知乎](https://zhuanlan.zhihu.com/p/25204020)上大力推荐和解释自己的文章。不过本周我看的是least square GAN这篇，我准备下周研读齐国君教授的作品。

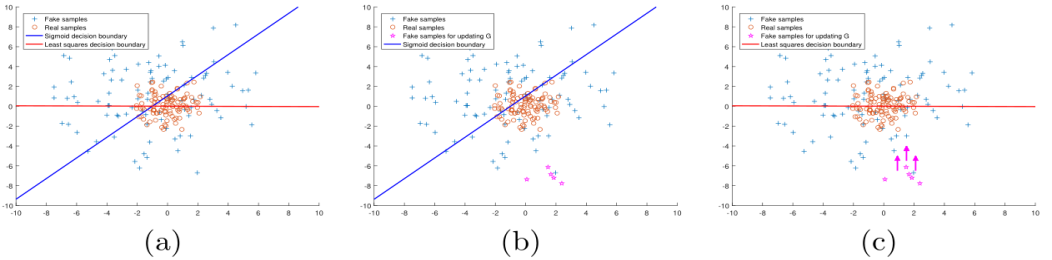
Least Squares GAN这篇文章针对的是原始GAN生成的图片质量不高以及训练过程不稳定这两个缺陷进行改进。改进方法就是将GAN的目标函数由交叉熵损失换成最小二乘损失，而且这一个改变同时解决了两个缺陷。在解决GAN生成图片质量不高的问题时，作者解释为以交叉熵作为损失，会使得生成器不会再优化那些被判别器识别为真实图片的生成图片，即使这些生成图片距离判别器的决策边界仍然很远，也就是距真实数据比较远，这就导致了生成器的生成图片质量并不高。

图3.两类loss对比图

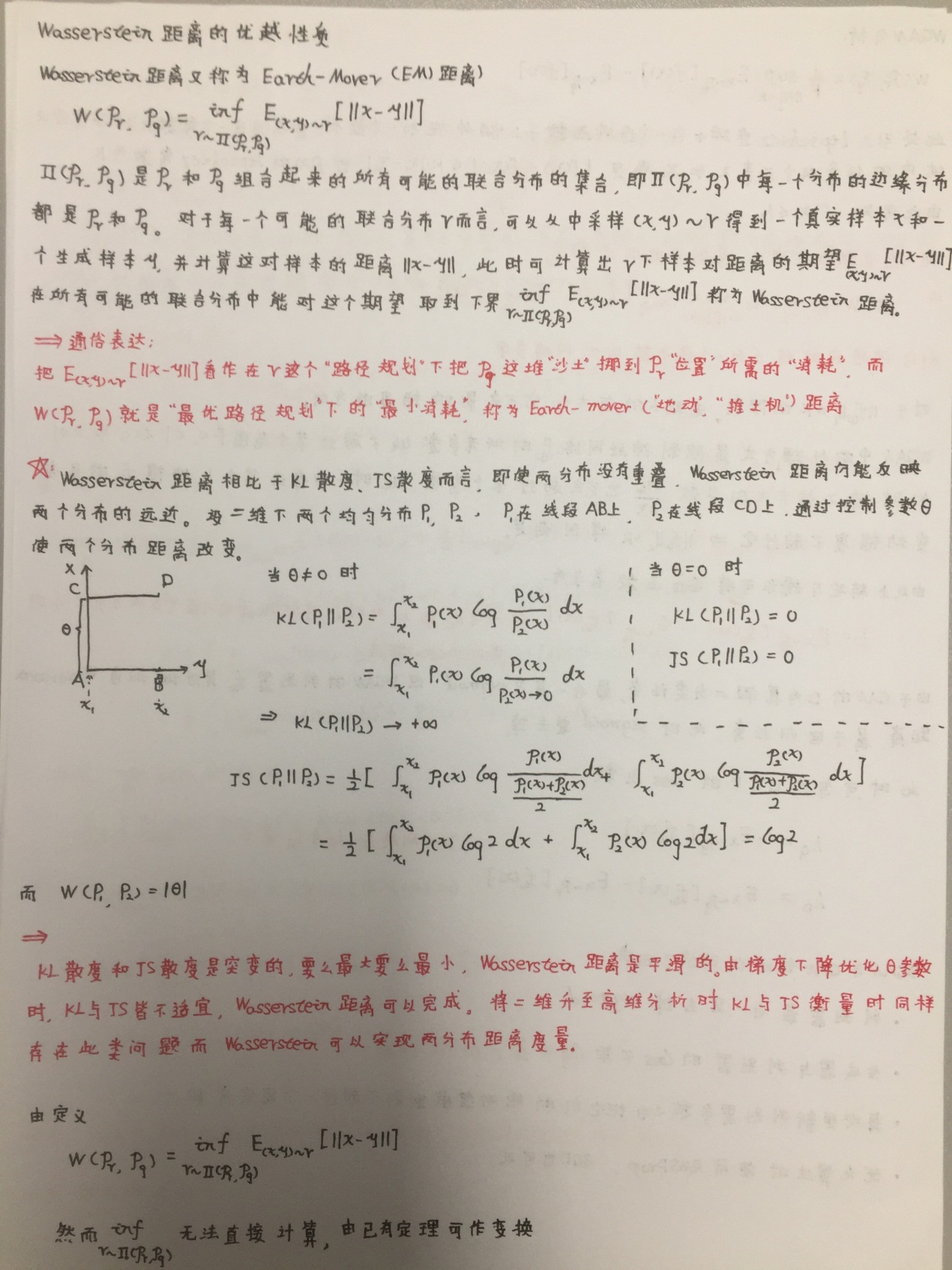
而要使得最小二乘损失比较小，在骗过判别器的前提下还得让生成器把距离决策边界比较远的生成图片拉向决策边界，这样将在一定程度上提高生成图片与真实图片的相似度，从而提高图片质量，由图3可以看出。

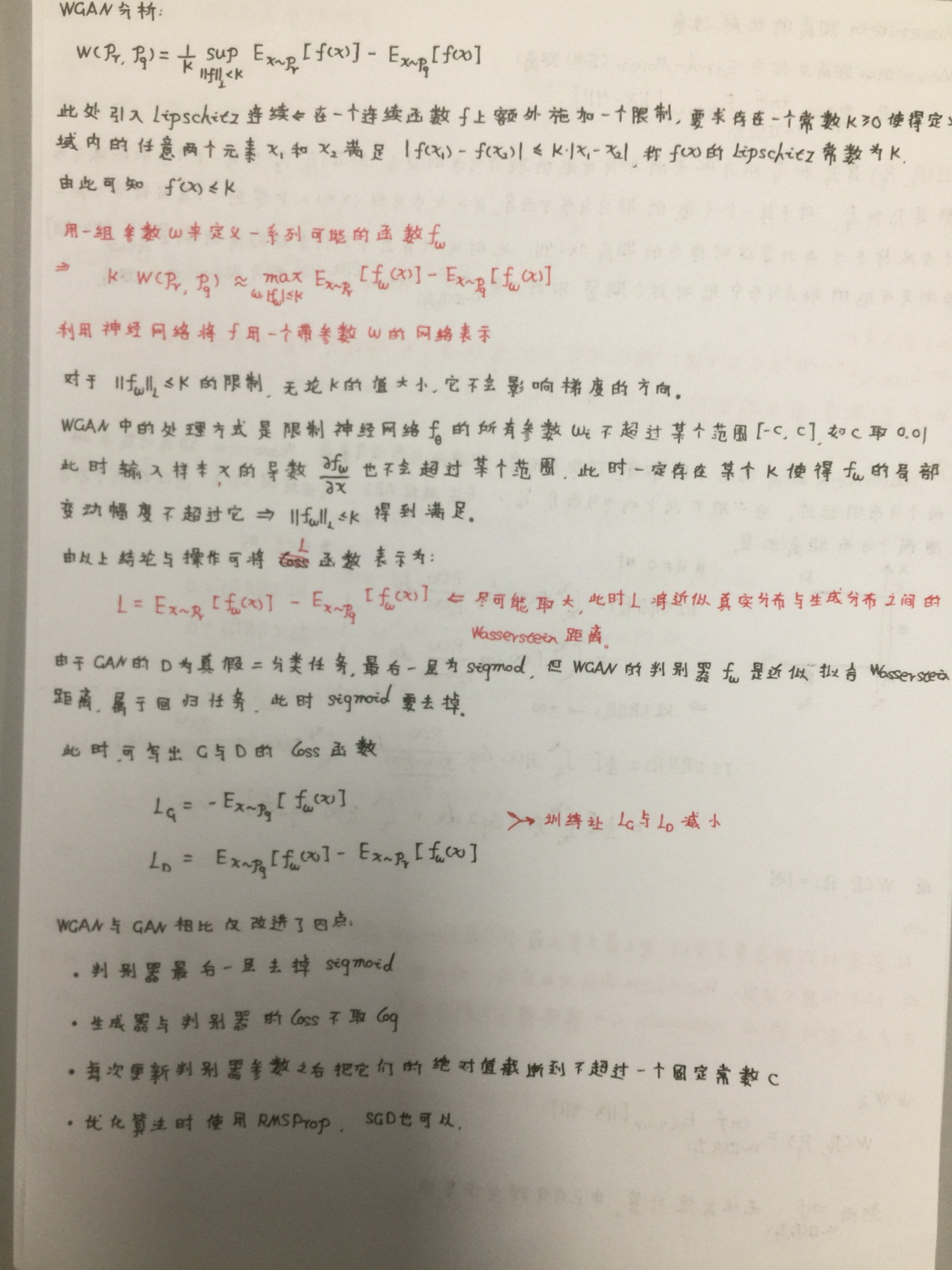
最小二乘损失可以使得GAN的训练更稳定是由于LSGAN的损失函数的设计。在之前报告中提及原始GAN存在生成器梯度消失的情况，原因是当判别器最优时生成器的目标函数中JS散度会趋于常数log2，致使值趋于0出现梯度消失得情况。LSGAN采用最小二乘损失函数由公式2展示，给生成样本和真实样本分别编码为a,b，G的目标函数将编码a换成编码c，这个编码表示D将G生成的样本当成真实样本，取a=-1, b=1, c=0或者a=-1, b=c=1.这样处理后将避免生成器损失函数为0的发生，具体的证明过程将在附录中展示。

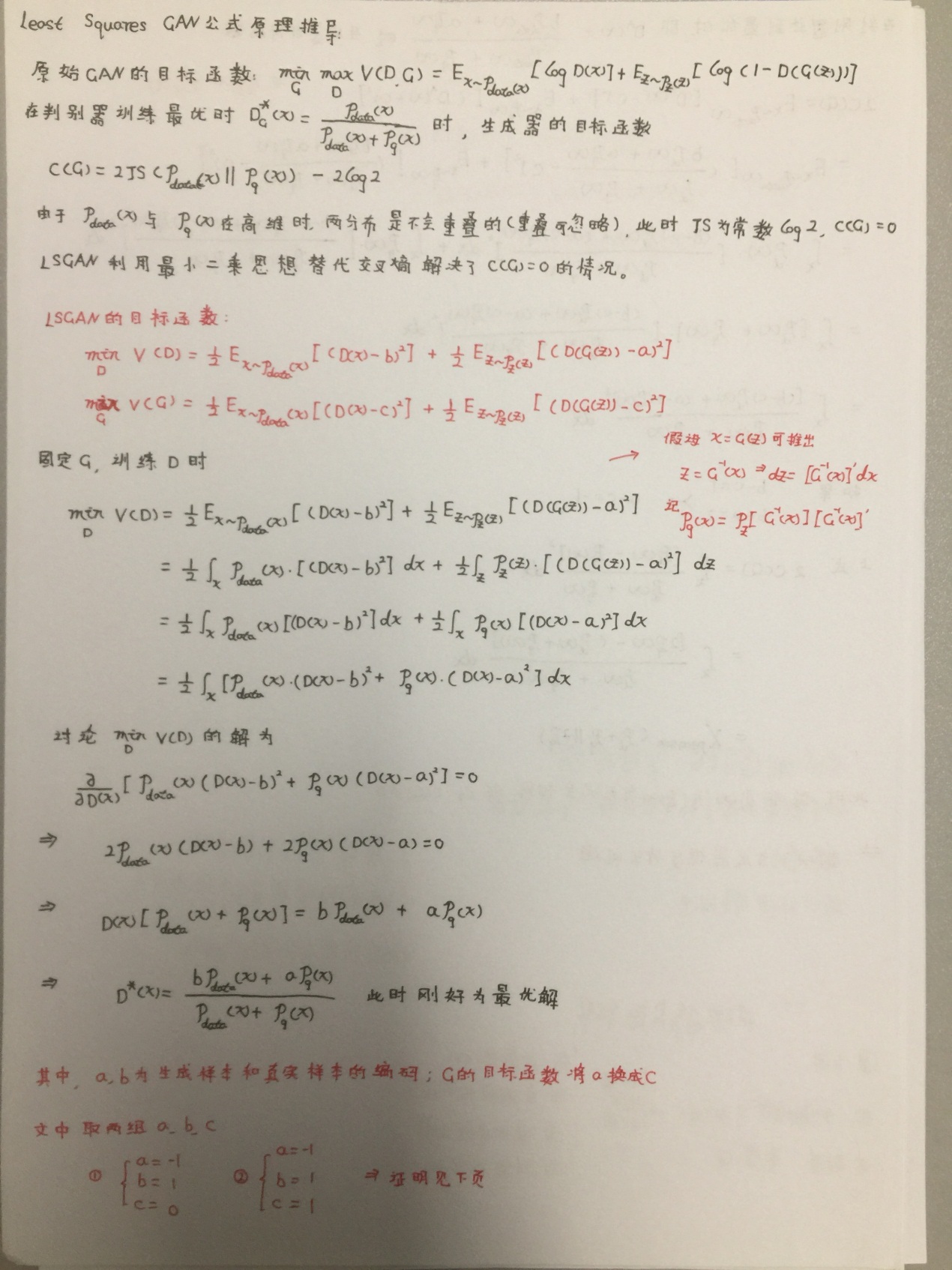
 (2)

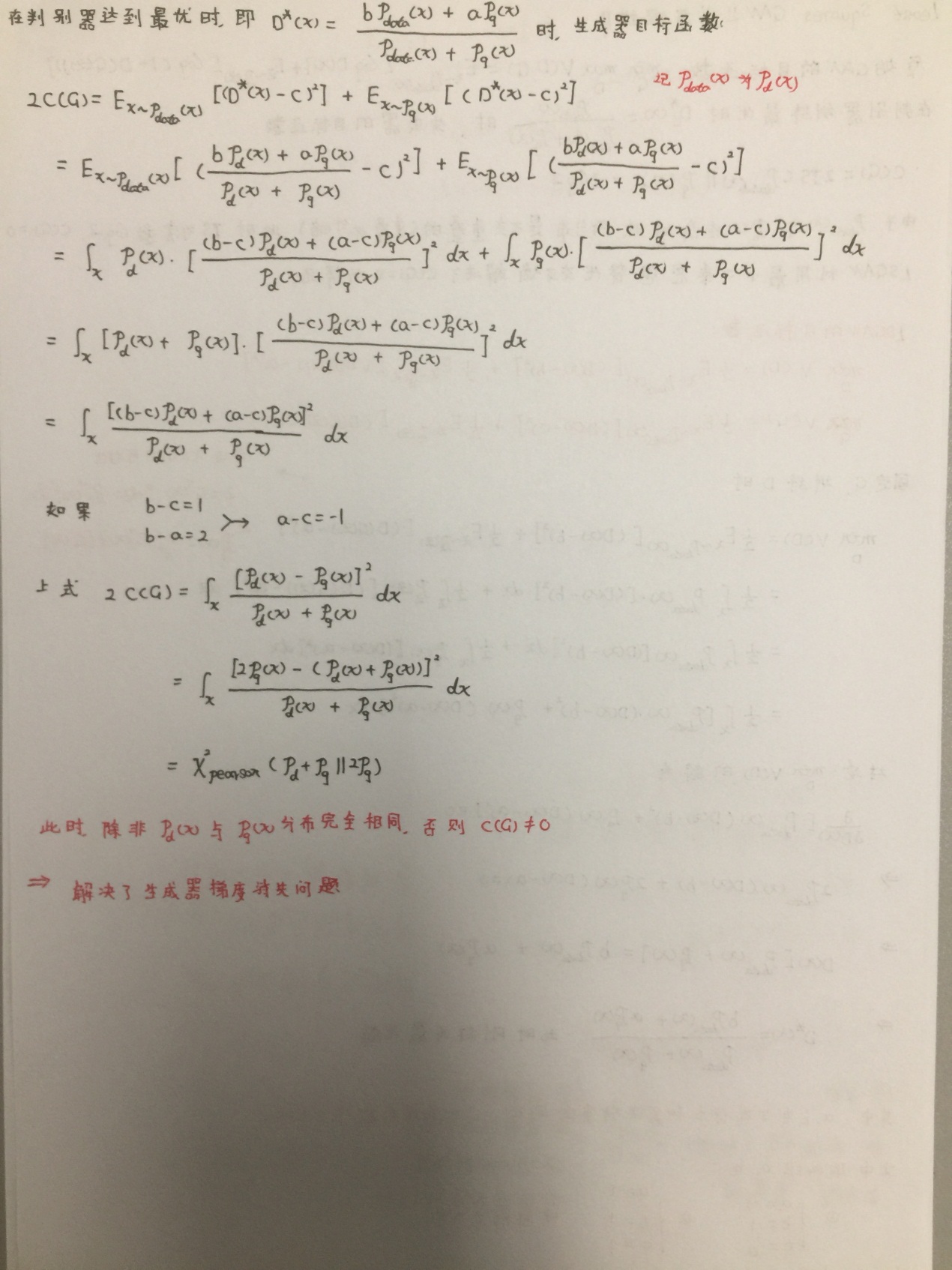
小结：本周研读了WGAN和Least Squares GAN两篇文章。WGAN利用EM距离代替JS，KL散度从而解决了生成与真实分布的距离衡量，从而改进了原始GAN存在的两类问题。Least Squares GAN在利用最小二乘损失函数的设计同样解决了生成器梯度消失问题，同时在图片的生成质量上有所提高。这两篇文章均是从数学的推证上出发，尤其是WGAN，满篇的数学公式，当然受限于已有的数学基础有的数学定理只是读读，在以后的时间中不断的提高数学基础。

这一周在实验室和檀华东互相讨论发现我们都存在着机器学习基础薄弱的短板，在接下来时间在任务完成基础上还要对数学和机器学习中的一些基础概念上加强。下一周我会把齐国君教授的[损失敏感GAN](https://arxiv.org/abs/1701.06264)研读分析一下，时间充裕的情况把WGAN的改进[WGAN-GP](https://arxiv.org/abs/1704.00028)研读一下。我这周整理一下GANs的论文集合和github集合更新在<https://github.com/TwistedW/Tensorflow-noting>。

**附录**

****

****

****