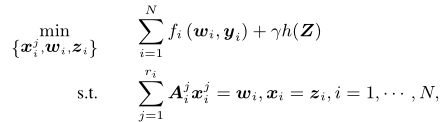
1. 看完Boyd《convex optimization》第四章，该章给出了一些凸优化问题的定义。
2. 看了一些分布式优化算法的论文：
   * 《Distributed and Parallel Block-wise Multi-Task Learning》AAAI17(在投中)，主要工作是借助ADMM算法解决异步MTL问题。对于 形式的优化MTL优化问题，为了能够做到数据分离并能应用异步ADMM算法，该文将优化目标做分块处理，转化为优化下面问题：



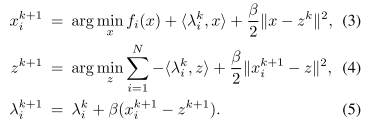
再针对该问题应用ADMM。文章的该部分思想值得借鉴。接下来提及SCAD和MCP就很突兀了。

* + 《Asynchronous distributed ADMM for consensus optimization》ICML14，针对有限和的优化问题，利用consensus optimization，提出了异步ADMM方法。主要的思想是每次master只接受较少数目的workers更新，并对workers的落后(delay)设置容忍度，以抑制worker传送的信息的延迟。

为了解决 优化问题，可以将原问题转换成consensus形式：



即可以应用ADMM算法：



传统的同步做法是各workers做完（3）（5）步，把结果发给master做（4）步。该文提出了一种异步算法，主要做法有以下两方面：

（1）、master只等待最近更新完的S个workers实现异步。

（2）、异步算法会遇到一个问题，即worker延迟，表示worker返回给master的信息太旧了，为了解决该问题，该文给worker的延迟设置了一个容忍度。

* + 《Distributed Delayed Proximal Gradient Methods》NIPS workshop on opt 13，针对f(x)+g(x)类型的优化问题，利用prox变形分布式求解，迭代如下：

延迟方面做法和上文类似。

* + 《REVISITING DISTRIBUTED SYNCHRONOUS SGD》ICLR17(在投中)，为了克服异步SGD的delay问题，该文使用同步算法，为了克服同步算法中个别的workers计算的慢拖累master计算，做法是增加额外少数的workers，每次更新时master不再等待最慢的少数几个workers。虽然该文的思想比较naïve，但是论文实验很是充分，感觉结果很好。

看的这些论文给我的一个比较大的感觉是异步算法一个重点是解决worker信息延迟的问题。下周将继续看ADMM相关，MTL相关，分布式优化相关，思考它们的联系。