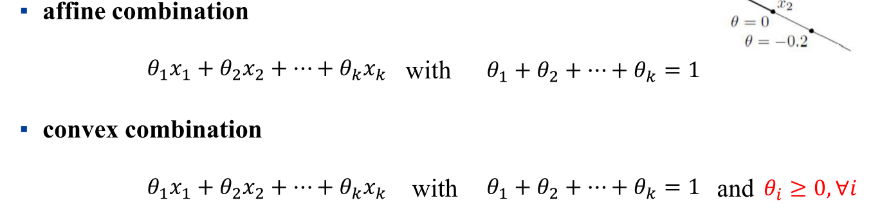
1. 概念

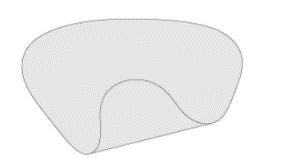
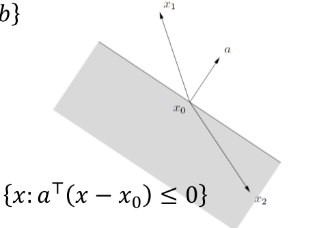
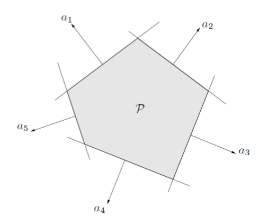
**正定矩阵P 对称，且x^TPx>0特征值大于0**



仿射和凸组合

凸集：集合内的点的凸组合也属于集合 内点（领域也在集合内）**空集和全集即开又闭**

凸包（最小凸集） 超**平面，**半**空间（a是法向量）** 多面体由超平面和半空间围成

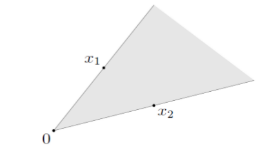
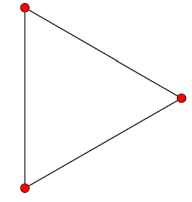




仿射独立（一个点无法由其他点的仿射变换表示，**2D最多3个**） 锥：**过原点**的射线

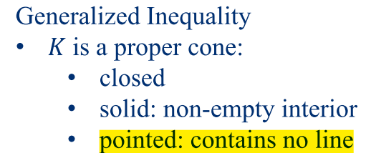
单纯形 （一系列仿射独立的点的凸包） 凸锥：锥变换在集合内

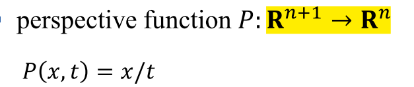
两点一线段，三点一图形

 椭球 **P正定**

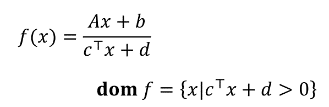


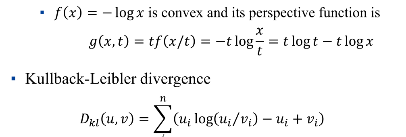
Proper Cone

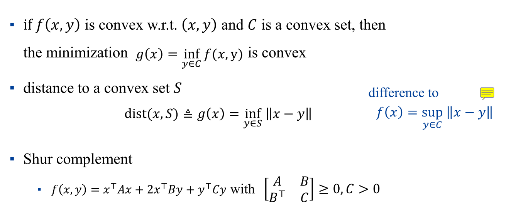


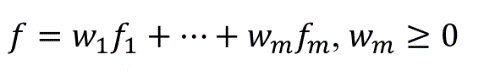


保凸运算 1.交集 2.仿射变换（凸集的和，放大）3.透视降维

4.线性分式函数 仿射+透视 透视（KL散度） g(x,t)



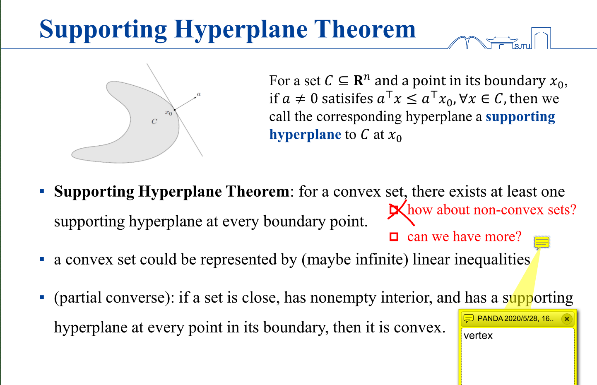
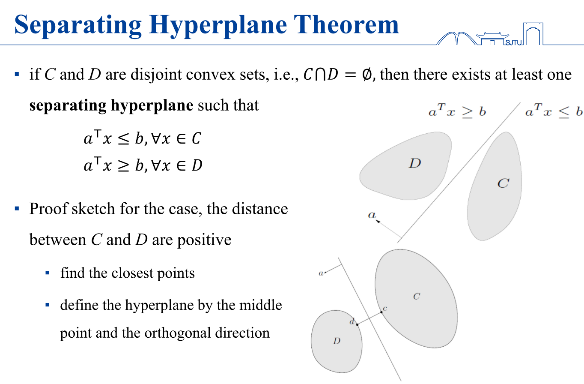
5. 非负加权和



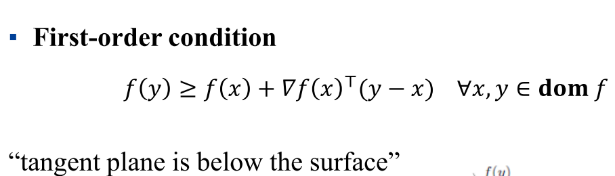
 7最小值和舒尔补

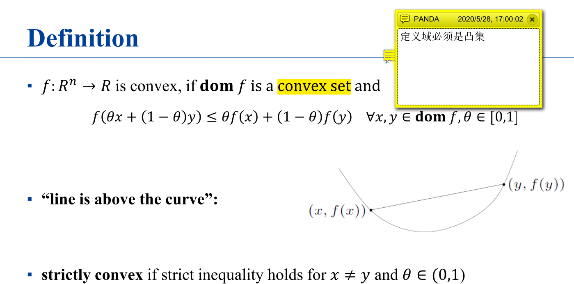
6.逐点上确界

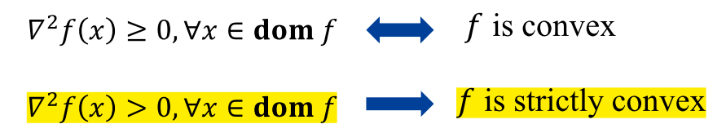
分离超平面（必须凸集，不可逆**，严格的话不可取等号**） 支撑超平面（必须凸集）



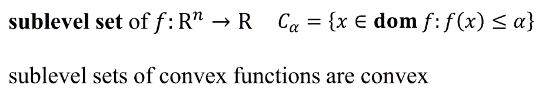
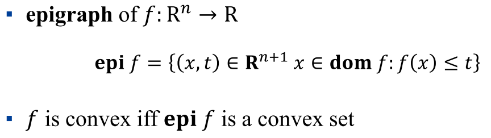
**凸集可以由线性不等式的解集表示（通过分离超平面定理）**

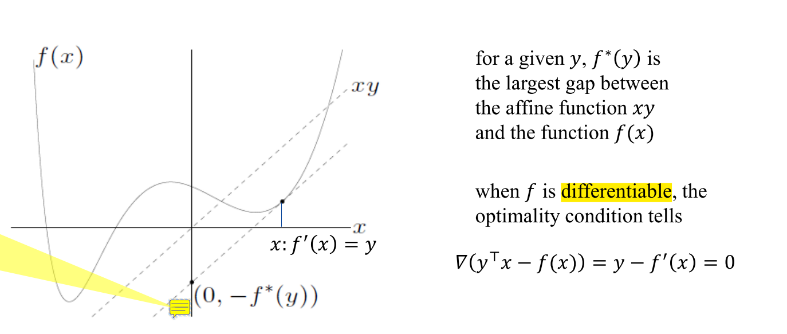
1. 凸函数（定义域，可不光滑，是否严格） 一阶和二阶





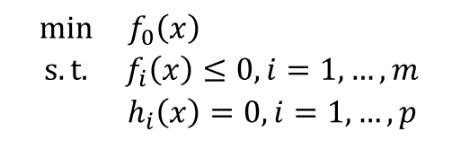
水平集（**针对定义域的**，保凸） 上镜图（升维，保凸）



**共轭函数**

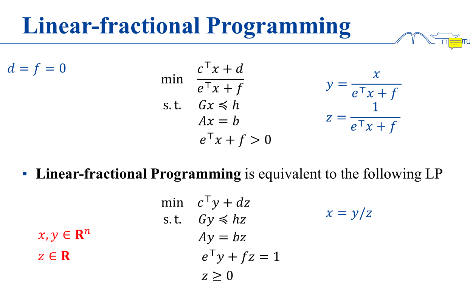
**切线和函数差值最大值，y为斜率**

**在f(x)’=y的点找 永远是凸函数**



1. 凸优化问题 在凸集上找凸函数的最小值

可行解属于可行集（凸函数和约束的定义域交集）

全局最优解f0(x)最小

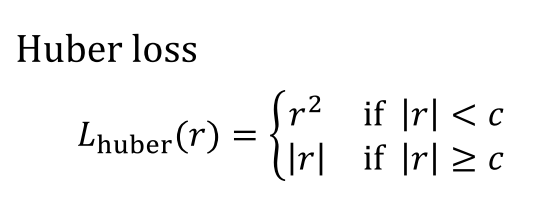
局部最优解：距离x范围d内，f0(x+d)>=f0(x)

线性规划（relaxation法）：例如整数约束变成非整数

线性分式（规划换变量法）

二次规划（梯度法）向量优化

反正拉格朗日对偶最好用

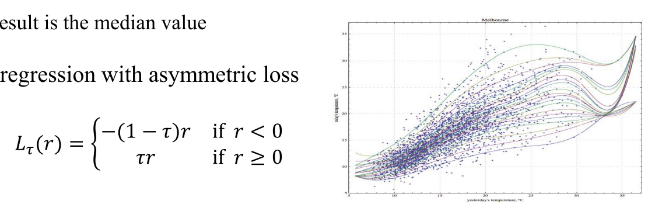


4． 常用模型

监督学习

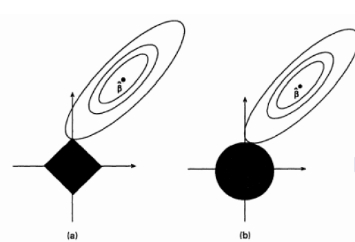
**线性拟合**用最小二乘法 缺点：偏离值影响太大

解决方法：**huber loss**（连续和光滑问题）

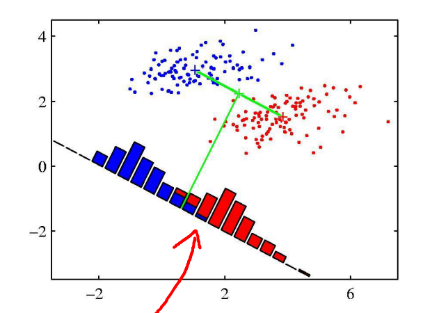


**分位拟合：拟合曲线上下点成固定比例**

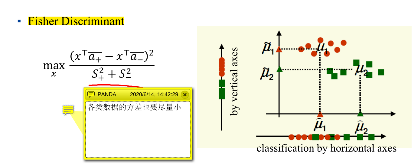
**正则化项： 额外min 所有参数**

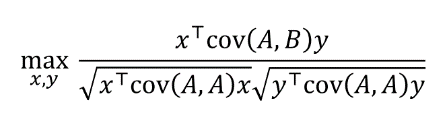
**Lasso: l1norm，使系数稀疏**

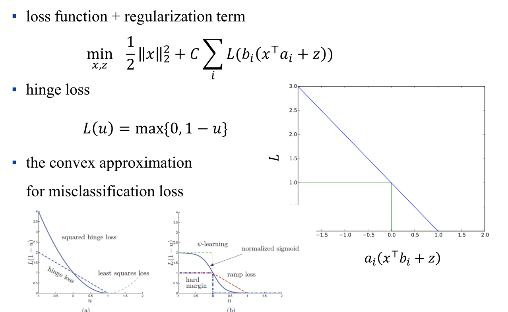
Ridge: l2 norm

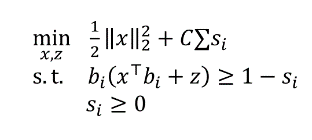
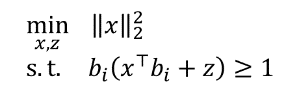
**LDA: 在直线上投影均值最大 CCA 找到两条线,使得数**

方差越小越好 **据上面的投影相关性最大**

 **数据A,B描述同个object**



**SVM 最大边距 硬边距 软边距 hinge loss（和分类结果<=1数据有关）**



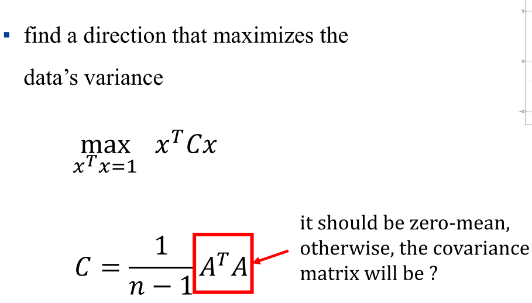
**非监督学习**

降维分析

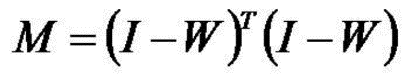
**主成分分析PCA（线性分有用） 局部线性嵌入LLE（流形学习）**

A是数据，x和x^T是主成分和无效成分， 1用一个点周围的点的加权和表示它

x就是C的特征向量 （用最小二乘法得到权重系数）

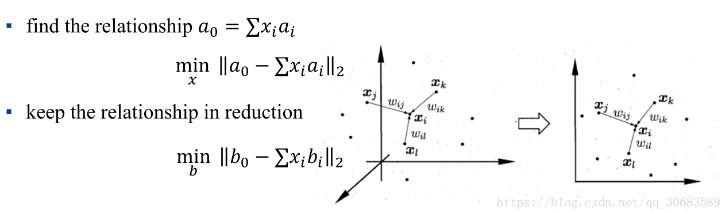
 2通过低维的线性变换矩阵b尽量保

留所有点和周围点的加权关系



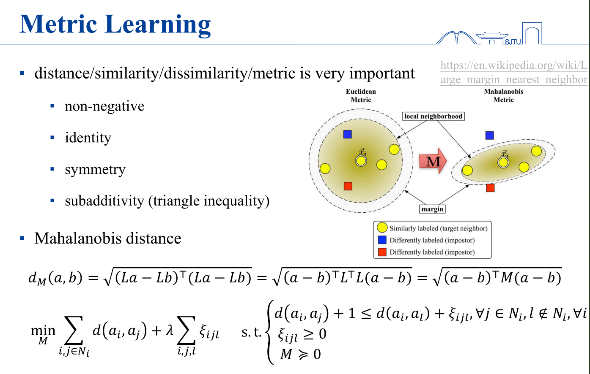
W是权重系数矩阵，特征向量为

低维特征向量



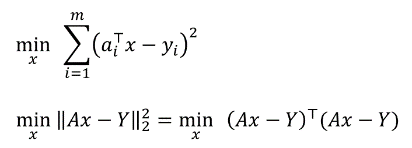
**聚类 K-means 质心迭代**

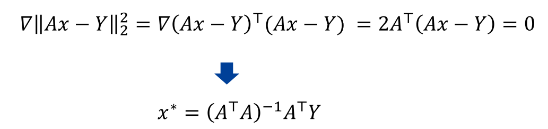
**度量学习**



**自编码器（高维数据编码至低维神经元再解码）**

目标：**神经网络输出和输入一样**

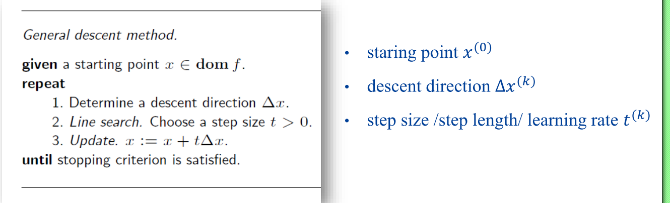
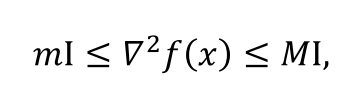


1. **无约束问题解法**

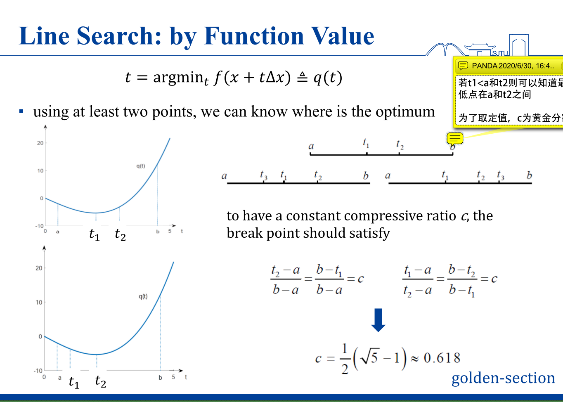
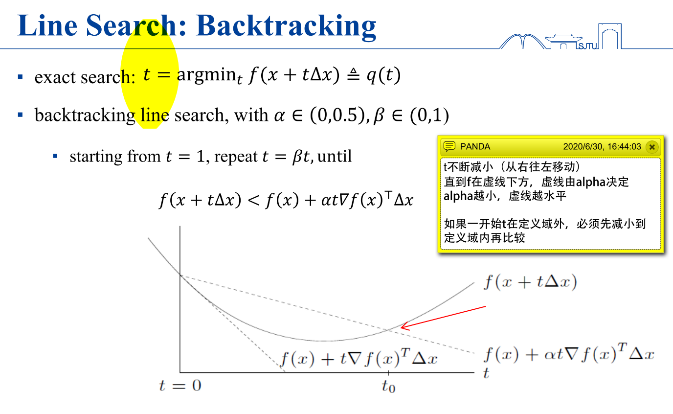
1.OLS**直接求梯度为0**

Pseudo inverse计算量大

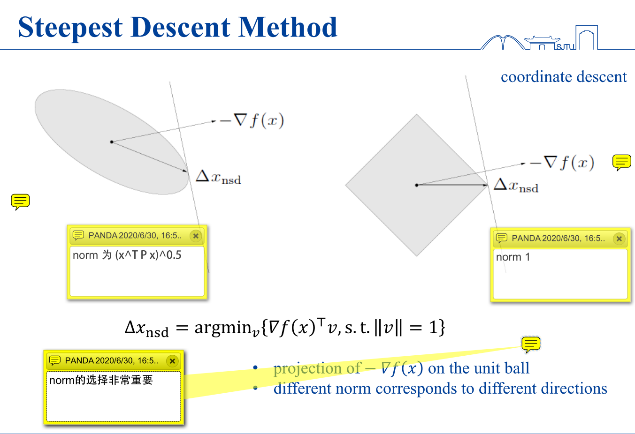
2 梯度下降 梯度为0需要strong convexity 复杂度和m/M成线性关系



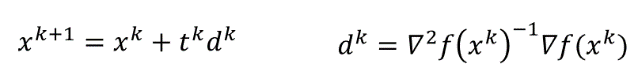
line search exact line search:找一条直线上f(x)的最小值,计算量大



最速下降法 先假设模长和梯度，找负梯度上投影最长点（不同norm结果不同）

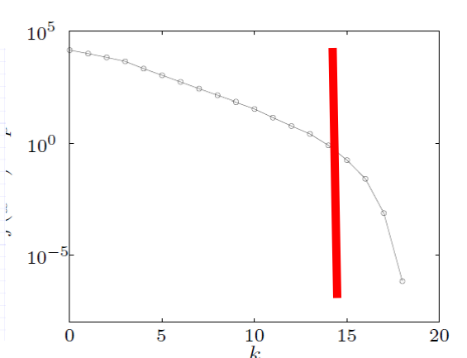


**牛顿法** **仿射不变** 最优解的仿射变换

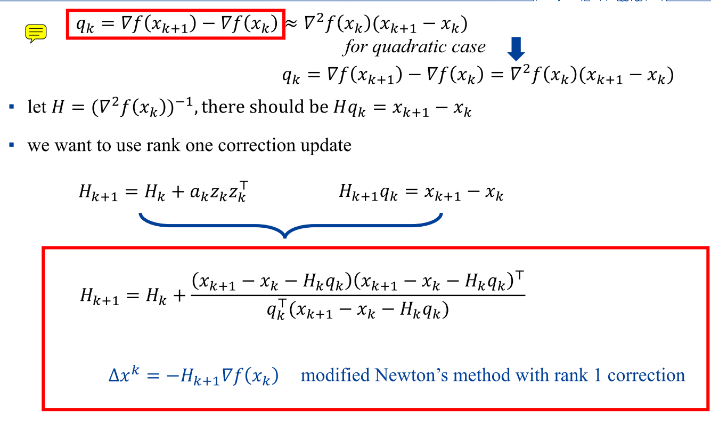
是仿射变换后最优解， t^k用line search

Damp phase下降速度为c线性

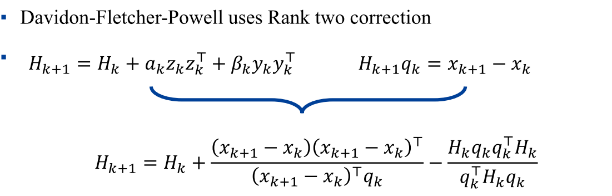
Quadratic phase速度很快，6步到位

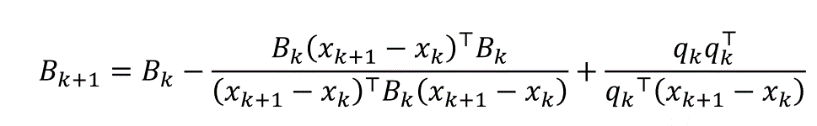


**拟牛顿法**

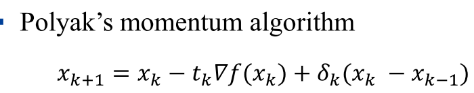


**DFP和BFGS(都是求二阶黑塞矩阵)**



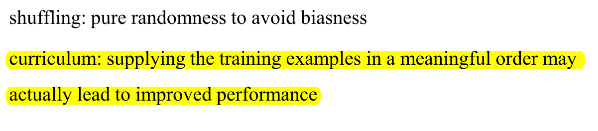
**BFGS L-BFGS 直接算**

**动量法：Polyaks, Nestrov**



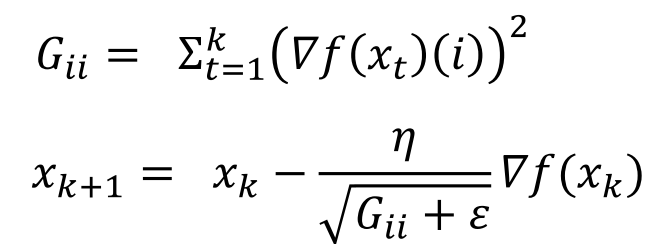
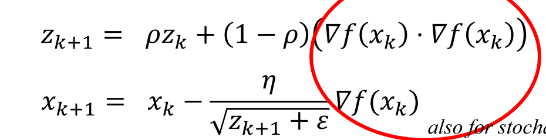
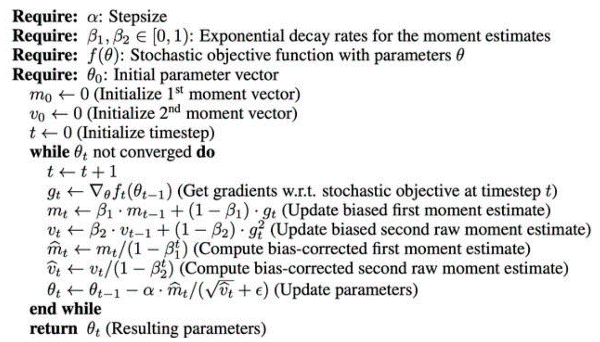




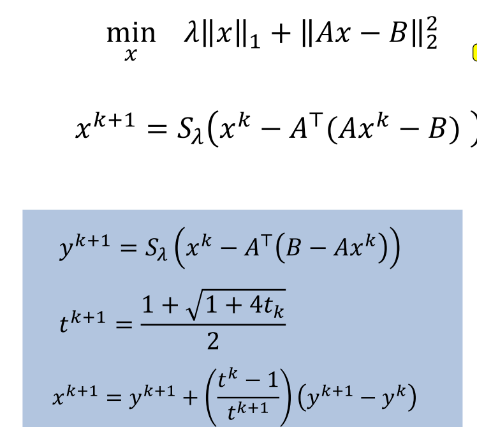
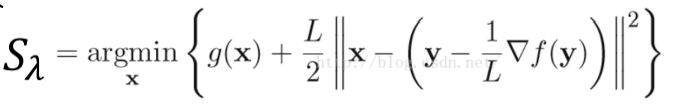
**实用的SGD 实用少数样本更新梯度 计算量小**

**能收敛 需要normalization**

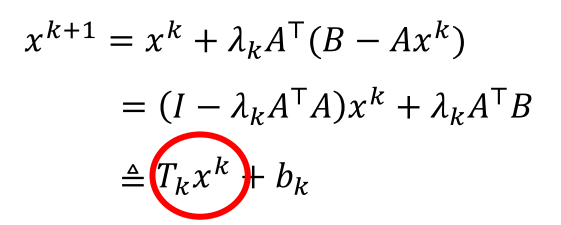
**Adagrad(学习率一直减小) RMSprop（梯度大学习率小） Adam**

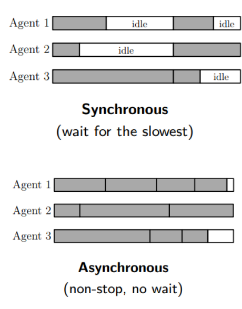


**ISTA FISTA**

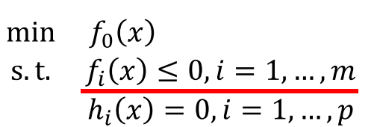
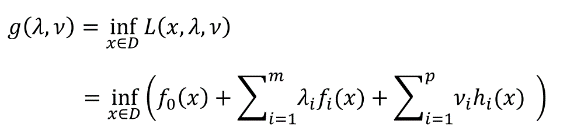
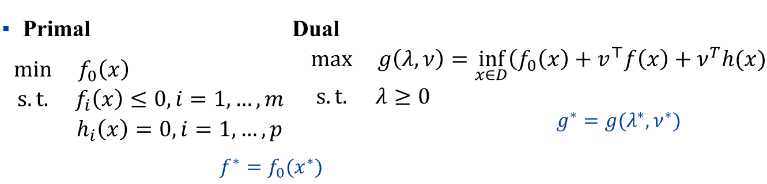


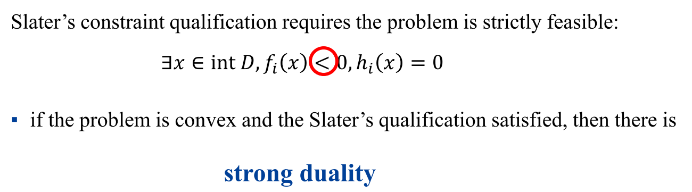
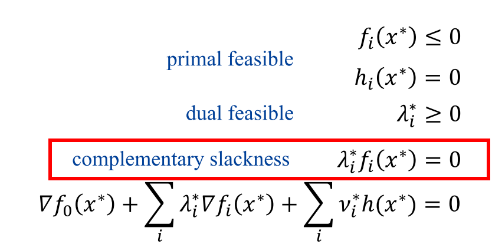
**B**

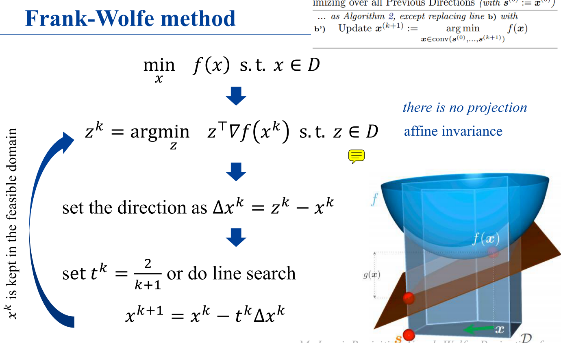


**Parallel computing（同时收敛，非同时效率高）**

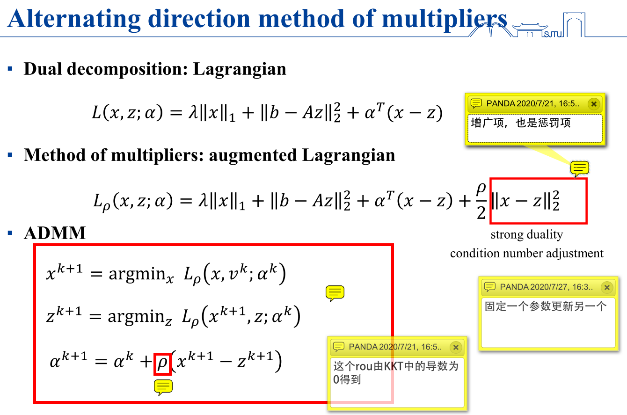
**x没有高项就可以拆分x并行算**

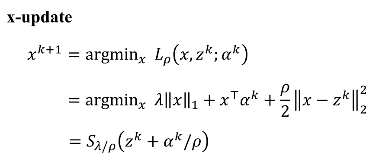
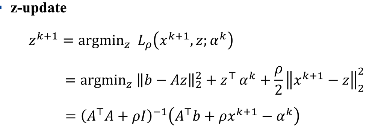
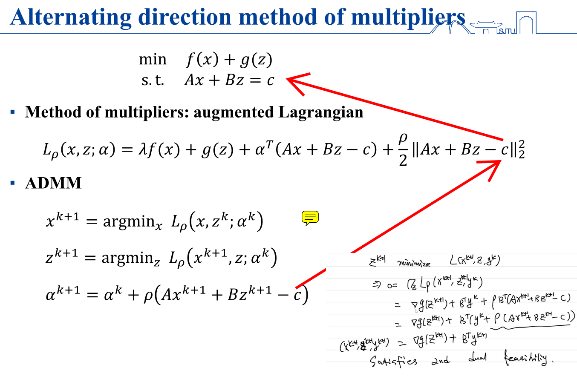
**6. 拉格朗日对偶**

**Slater 条件 KKT条件（必要，slater满足则充要）**

**7 大规模数据**

**约束（投影）梯度下降法**

**ADMM（轮流更新两种参数）**

**增广项影响并行运算**