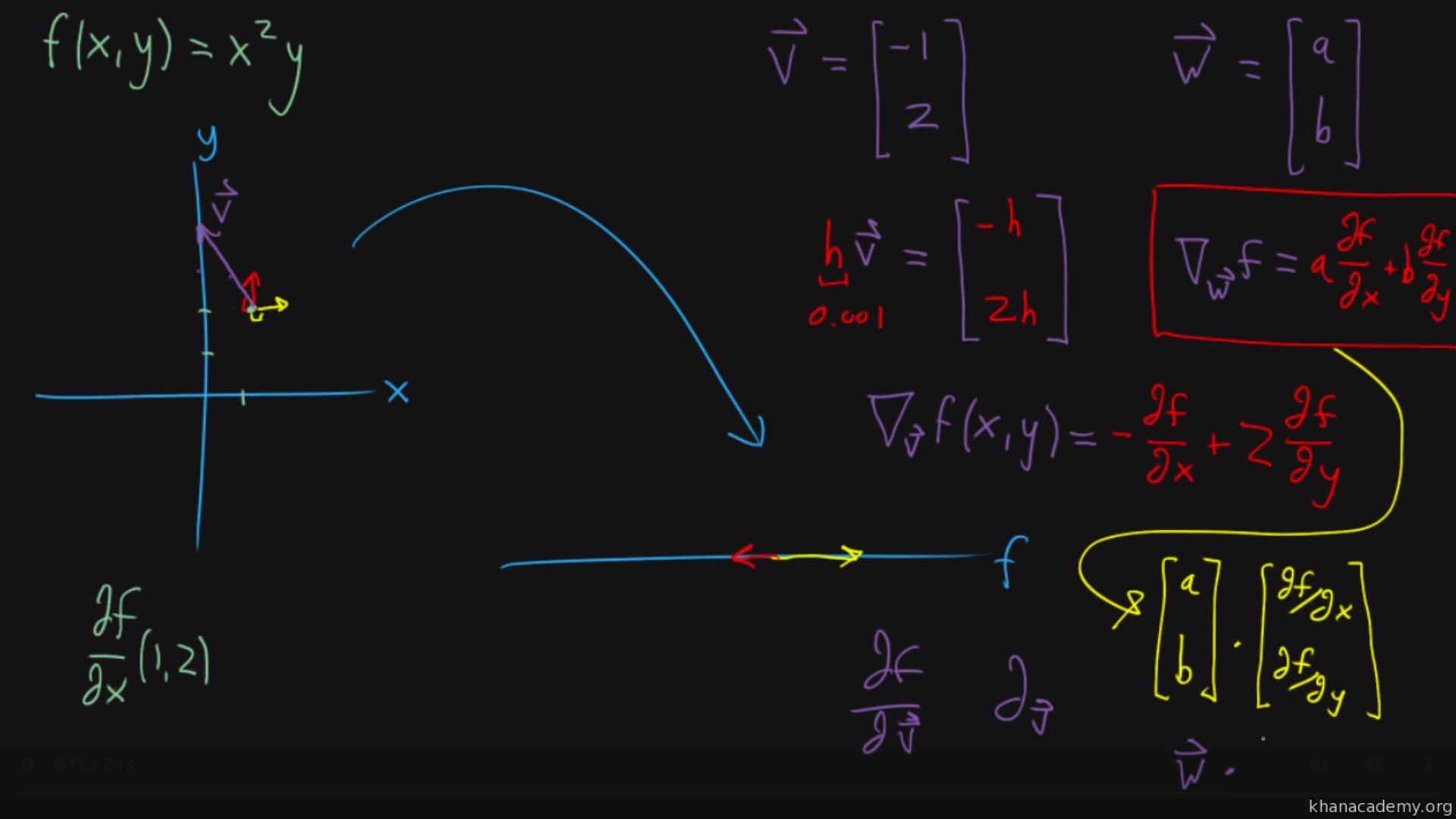
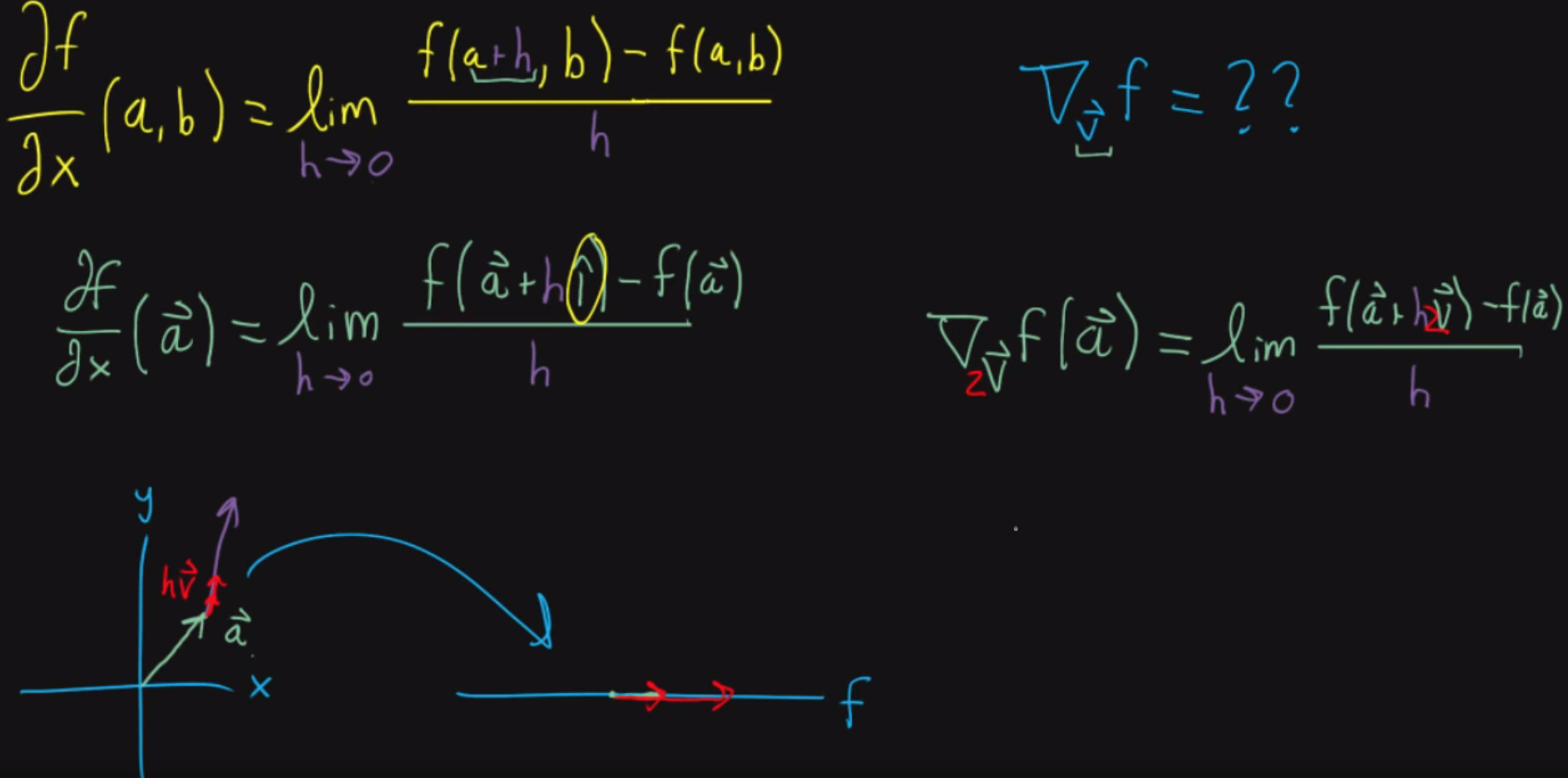
# 优化理论

## Directional derivative

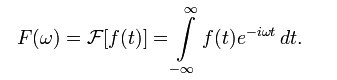




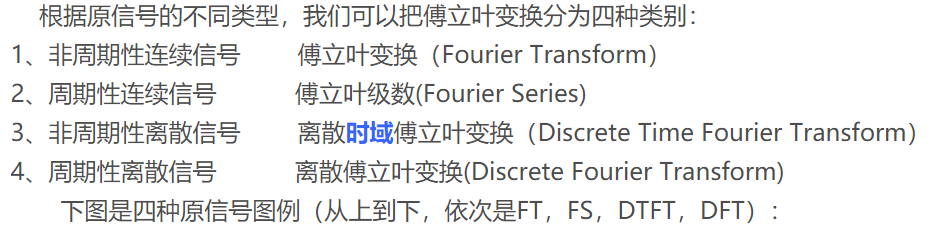
方向向量长度的变化会影响方向导数大小，而我们希望方向导数与方向向量长度无关

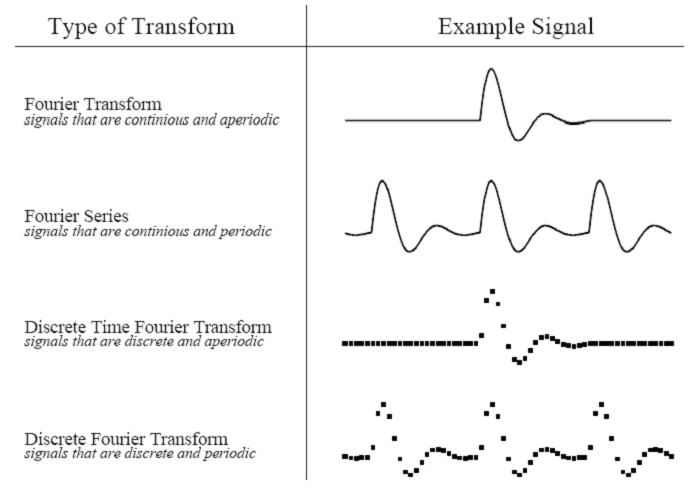
## 傅里叶变换

一种从时间到频率的变化或其相互转化。



任何连续周期信号都可以由一组适当的正弦曲线组合而成。





## KL散度

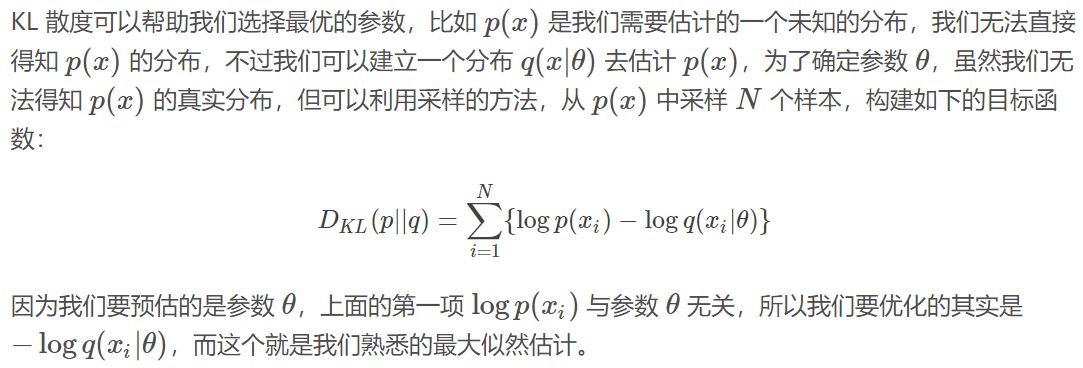
Kullback-Leibler Divergence 用来衡量两个概率分布的相似性的一个度量指标。

信息熵（数值越大代表概率分布包含的学习率越大）

在信息熵的基础上，我们定义 KL 散度为：

或

表示的就是概率分布q与概率分布p 之间的差异，很显然，散度越小，说明 概率 q与概率 p 之间越接近，那么估计的概率分布于真实的概率分布也就越接近。



# 概率与统计

# 线性代数