参考书籍：[1]《线性系统理论数学基础》、[2]《最优估计理论》[3]数值分析

待手写推理公式，将结果记录在专用的笔记本上，先用再推

* KF、EKF
* 粒子滤波
* 时域和频域之间的联系
* 姿态表示方法转换
* 熵的底数为什么是2

待解决问题：

1. n阶距
2. 为什么H正定，这个点是极小值点
3. 多项式模型也称为线性回归？因为系数是一次？
4. SVD应用总结
5. 自相关矩阵正定的证明，噪声、状态变量的协方差矩阵属于自相关矩阵。协方差矩阵包括自相关矩阵
6. 求逆的复杂度？
7. 符号（方差等一般用E来表示，对于概率相等的事件可直接除n，相当于每个事件的概率为1/n）

* -偏差，真实值和估计值的差，数学上A与B的差只能是A-B
* D（X）、-随机变量X的方差，随机变量的二阶中心矩。
* 估计误差-真实值与估计值的差
* 估计误差的数学期望为0时表示为无偏估计。
* 均方误差-表示的是估计误差的方差，值越小，估计值在真实值附近越密集，即估计值越接近方差
* -标准差或均方（偏）差（RMS） ，方差开根号，观察更直观，如表示分数时
* cov（X，Y）-X、Y随机变量的协方差，X和Y的1+1=2阶混合中心距
* 协方差矩阵-两个随机向量X、Y之间的协方差矩阵，因cov（X，Y）= cov（Y,X）所以协方差矩阵对称。X=Y时为自相关矩阵，算法中常用的是自相关矩阵。
* -相关系数，cov（X，Y）/，其绝对值<=1，为0表示不相关，及相互独立，绝对值越大关系越大。不相关可展开为X-E（X）每一项成E（Y）-E（Y）。不相关时，X发生了，Y发生的概率会更大或更小，这就打破了平衡。相关系数是随机变量间线性关系强弱的一个度量，相关系数绝对值为1，Y，X呈线性关系，Y=aX+b，即知道X后Y也确定的了，当X=1时，Y=a+b。相互独立，一定不相关（），不相关不一定相互独立。
* -求期望，看见E就要想到把概率省略了，用积分展开乘概率即可。一阶原点矩。注意下标较常用，表示按X的概率密度分布求X的期望，其他下标类似。
* -；表示把前后两个符号隔开，如N（x；），x为随机变量，后面为均值和方差，即参数
* inf是函数取下限，argmin是变量取下限
* K-增益矩阵，常在递推公式或KF中，将其作为测量残差的增益
* 贝叶斯公式：,集和条件概率和全概率公式。其中z为观测量，x为待估量。为后验概率，为似然，为先验
* 表示x的绝对误差
* 表示x的相对误差
* max取值（内容），argmax取下标索引
* 大O-同阶无穷小
* 小o-高阶无穷小，如y=o(x)表示y是x的高阶无穷小
* 两个重要极限：
* ：表示f对u求导

1. 向量表示：（0,0,0）指向（a,b,c）或ai+bj+ck，后者可用向量加法得出

某点在坐标系的位置表示为（a，b，c）

1. 刚体运动：如小车一样的硬物体的运动。运动过程中车体大小、形状不变，即假设用笔车体上画出的向量的大小和相对夹角永远不变
2. 常见坐标表示都表示c坐标系到w坐标系的变换，即c原点在w坐标系下的方向和位置，如：
3. 数学技巧及基本概念

* 无法通过整体直接"统计"获得你想要的"量"时，你只能通过"部分样本"来做"整体样本""量"的估计时，谈估计方法的"有偏"还是"无偏"才是有意义的。注意统计和估计的区别，统计是全部，估计是部分估计全部。

<https://blog.csdn.net/qq_16587307/article/details/81328773>

* 均值、方差、矩和功率谱密度称为数字特征
* 测量总是存在噪声引起的方差（covariance），还有本身结构引起的偏差（bias），如imu的偏置。
* 注意参数估计和状态估计的区别，参数估计是已知状态和输出，估计参数，如高斯的均值和方差。状态估计是已知参数估计状态x。
* 卷积公式：。《概率机器人》P347有高斯分布的卷积形式
* 后验：在给定条件下想知道的结果的概率，如P（x|z），x为待估计的状态变量，z为观测值。贝叶斯分母与后验结果无关，故可以用归一化常数表示，即各分子项除以分子项之和。见《概率机器人》P340注释。贝叶斯公式能够将条件与结果位置互换。只改变其中两个元素就好
* 多维用2维写出来看看，一些公式写出具体式子展开计算，如向量和矩阵展开后比较好看出来
* [2]76页最下方有自己的理解，可以把下次观测看成多次观测相乘的形式（类似极大似然），z为观测值，x为待估计参数，使下式最大即对数次数最小，可用梯度下降来求，也可以使关于x的导数为0。对于高维高斯分布，如,其中Qt为正定（自相关矩阵）对称矩阵，当噪声为0时，即Qt对角为0，估计值zt’等于zt，在相互独立的假设下，求最大似然，即使所有和最小。Qt为对角阵时，二维沿xy轴的截面都是高斯分布，可以用二维将对数指数项展开看看。确定一维后，非对角的另一维概率还和第一维有关
* 设协方差为Qt=E（ztztT），非对角表示不同维度相互影响
* xTAx为一个数，可像类似于求偏导。矩阵和向量求导公式《线性系统理论数学基础》P44
* 可相似对角化矩阵p-1Ap=特征值对角阵，p为特征向量，可以通过微调特征值，改变A，特征值越小的对A影响越小
* 末尾加1，由三维变四维
* 旋转轴上向量在旋转后不发生改变
* 中t较小时,可通过泰勒公式近似为I+At
* 在离散域累加时，先考虑连续域求积分
* 两边大于0可以两边同时取对数指数等单调函数
* 凸优化在全局二阶导不变号的时候才能找到全局最优解，否则可能到局部最优
* 对于有n个自由解的线性齐次方程，可以先令其中n个解为常熟，一般取1
* 优化技巧，优化什么参数就对什么求导，偏导为0，表示无关
* 两事件相关矩阵不是对角矩阵，无关则为对角矩阵
* 对方差的理解，数据间的差异，将数据等比例放大，会使方差变大
* 线性方程的方差为凸优化问题，一般为一元二次函数，可举例子看一下
* 二次范数是平方不受正负影响，但简单差值受正负影响
* 归一化，
* 离散傅里叶对偶数尺寸的阵列执行较快
* 最大化0-1的数，可转化为最小化-log
* 期望通常和概率有关，如数学期望，方差等
* 点积可以反应两个向量的相似度
* CNN中卷积的意义：大于t之后权重值为，小于0时，x值为0。
* 链式求导、概率模型和除数为0都会引发不好得后果，经常需要平滑
* 可以叉乘某个向量，使某一边为0
* Ap=lamdap，那么特征值接近0的将其改为0对结果影响也不大，说明特征值大的比较重要，可以借助PCA来理解
* 高斯牛顿法是二次收敛，在接近最小值时比较好，一阶梯度下降是最速下降，在离最小值较远时比较好
* 坐标转换齐次矩阵T的逆矩阵为RT –RTt
* 若x=m1/n1=m2/n2=m3/n3,则x=（m1+m2+m3）/(n1+n2+n3)
* qtRq’=trace(Rqq’t)
* trace(BAAt)=trace(AtBA)，xtx=trace（xxt）。x一般指列向量
* 协方差是衡量偏离均值的二次期望值，注意与方差的区别
* 高斯随机变量的任何线性变换都将导致另一个高斯随机变量
* 求逆复杂度O()，2叉树结构计算复杂度O（logN）
* Ef[g(x)]=,其中f（x）为概率密度函数
* 指示函数I，条件为真时为1，条件为假时为0
* 《最优估计理论》P17：k阶原点（可认为是减0）矩：。k阶中心点（减期望，期望可以认为是中心，矩认为是与中心的距离）矩：。k+l阶混合（原点）矩：。k+l阶混合中心矩：。数学期望是一阶原点矩，方差是二阶中心矩。协方差P16是二阶（1+1）混合中心矩。
* 注意区分变量和常量，如《概率机器人》P349
* 一个概率分布p的熵，《概率机器人》P14，Hp（x）=E[-log2p(x)]。若p为均匀分布，熵取得最大值。熵用在机器人的信息收集时，用以表达机器人在执行具体行动时可能接收到的信息。P447可以更好理解，其为二值的。不确定性越大（均匀分布不确定性大），熵越大。探索和移动的目标是使不确定性变小。
* 偏导是必要条件，还需验证一下
* p(xy)和f（x，y）都是且的意思
* 泰勒展开式： ,不包含等号
* 常用证明方法：数学归纳法、递推法
* 递推方法：由I0开始递推到In，通常使用数学归纳法来证明
* 定义和定理的区别：定义是定义某种运算法则或形式，自己定义的；定理是得出，某种结论，中间可能使用了某些定义
* 长除法：[3]p191用到

1. 迭代方法总结
2. 姿态表示方法转换1
3. 注意区连续情况下概率和概率密度：概率是某范围对应概率密度函数的面积，总面积为1
4. 在矩阵中，若数值为0的元素数目远远多于非0元素的数目，并且非0元素分布没有规律时，则称该矩阵为**稀疏矩阵**；与之相反，若非0元素数目占大多数时，则称该矩阵为**稠密矩阵**。定义非零元素的总数比上矩阵所有元素的总数为矩阵的稠密度。
5. 联系线性方程组，齐次坐标其对应线性方程右边为0，在位姿变换中表示为（x,y,z,1），非齐次坐标其对应线性方程右边不为0，在位姿变换中表示为（x,y,z）
6. 链式求导：
7. SVD:<https://www.cnblogs.com/pinard/p/6251584.html>

*A*=*U*Σ*VT*

　　其中U是一个m×m的矩阵，Σ是一个m×n的矩阵，除了主对角线上的元素以外全为0，主对角线上的每个元素都称为奇异值，V是一个n×n的矩阵。U（ATA分解得）和V(AAT分解得)都是酉矩阵，即满足*UTU*=*I*,*VTV*=*I*UTU=I,VTV=I。注意是特征值和奇异值之间满足平方关系，而不是ATA和A之间特征值呈平方关系，实对称矩阵两者呈平方关系。？

1. 求解线性方程方法

* 可以想办法使它成为超定方程，使用SVD解超定方程，m>n，取奇异值最小特征值对应的特征向量作为解，利用U、V都是正交矩阵
* 可以使用矩阵三角化来求解满秩线性方程的解
* QR分解，一般用于解满秩的情况。将系数矩阵A分解为Q单位正交阵和R上三角矩阵，方便求齐次或非齐次方程的解

1. 零空间：零空间也称为核，A的零空间为使Ax=0的x向量

https://baike.baidu.com/item/零空间/9249832?fr=aladdin

1. 一阶和二阶梯度法、高斯-牛顿、列文伯格-马夸尔特方法求x的增量，视觉slam十四讲P110
2. ransac方法求解带有噪声的数据
3. 向量矩阵导数公式
4. a 是向量 θ 的标量函数，则定义
5. β、 θ 分别是 m ×1和 n ×1向量，则定义
6. a 是标量函数， M 是 m n × 矩阵，则定义
7. M 是 m n × 矩阵， a 是标量函数，则定义
8. 常见xTRx对x求导的导数为R，KF中求导如下：



1. 爬山算法（从临近空间找最优解，像爬山一样）和模拟退火（如梯度下降时接受更差的结果，防止进入局部最优）

<https://blog.csdn.net/zhouzi2018/article/details/82118673>

1. [2]p3最优估计问题是估计无法直接观测的随机变量x（状态变量），需要通过与x统计相关的随机变量z（观测量）来对x进行估计。已知关于系统运动学或动力学（运动模型）和量测方程（观测模型）的知识，利用系统过程噪声（运动噪声）、量测噪声（测量噪声）的统计特性和初始条件信息，依据某种最优准则（KF、EKF、PF、MAP、非线性优化）对量测值z进行处理，确定系统当前时刻状态x的问题，称为最优估计问题。
2. 滤波有频域上的滤波和概率方法（最优估计），还有一阶滤波（超声波使用的滤波方法，可以减小更新速度）、均值滤波、窗口滤波等，最优估计一般是通过观测值z估计状态变量x，而其他主要用于传感器等直接测量值的滤波
3. [2]p4状态估计时间t1与最后测量时间t，t1>t为预测，t1=t为滤波，t1<t

为平滑。结合P130笔记理解

1. [2]P6基本事件构成样本空间，基本事件称为样本，x的值就是一个基本事件，现实生活的事件往往是离散的，而数学上的问题经常是连续的
2. [2]p6互斥-A、B不同时发生；对立-AB不同时发生且构成一个样本空间，概率和为1；完备-所有事件至少有一件发生，不一定互斥。独立-可理解为几个不同事件空间的事件。
3. [2]p9条件概率、全概率公式和贝叶斯公式的相互推导。机器学习的前提是已经具有大量的带标记数据的数据集，那么相当于有n次测量了，可以用于估算
4. [2]p11注意概率分布与概率密度函数的关系-概率分布的导数为概率密度函数，概率分布为F(X)=P（X<=x）,概率密度函数为p（x），注意大小写，那么F(X)在x处求导即增量x的概率和/增量x，这里增量x足够小。若增量x较大时可以看成是直方图方法。
5. 误差和残差的区别：误差是测量值和真实值的差，残差是测量值和预测值的差。<https://zhidao.baidu.com/question/1695447537976680788.html>
6. [2]p64最优估计的常见准则，前4种方法一般目的是都是估计状态变量x这种随机变量，使其更接近真值，估计误差越小越好，估计误差的数学期望为0，估计误差的方差也越小越好，最好为0。

后2种用于估计确定的值或参数，是参数估计的问题，不需要知道值或参数的统计特征（概率密度函数、一阶矩和二阶矩等）。

* 无偏估计：参考p66例题便于理解，无偏即估计误差的数学期望为0。状态的估计值和真实值具有相同的均值。往往与其他估计准则结合使用，如无偏最小方差估计和无偏线性最小方差估计
* 最小二乘估计：将测量残差的平方和最小作为最优估计准则的估计方法。如果不知道x和z的一阶矩（数学期望）和二阶矩（方差、协方差）及它们的概率密度，这时可以采用最小二乘法得到最优估计。不能保证估计误差的方差最小。
* 古典最小二乘法：无需知道统计特征，是无偏估计
* 加权最小二乘法：当W=R-1时，是缺少初值条件下的线性无偏最小方差估计，又称马尔可夫估计。优于古典最小二乘估计，需要测量误差矩阵（方差矩阵）的信息。R的逆说明测量误差方差越大，W值越小
* 递推最小二乘估计：证明待看
* 梯度下降方法：擅长解非线性问题
* 最小方差估计：在一切可能的估计中，将估计误差方阵P（x）最小的估计量作为最优估计。需要知道被估值x和测量值z的条件概率密度p（x|z）或p（z），以及它们的联合概率密度p（x，z）
* 线性最小方差估计：如果只知道x和z 的一阶矩和二阶矩，这种情况下，为了得到最优估计结果，必须对估计量的形式加以限制。假定估计值是观测值的线性函数，以估计误差阵（估计误差方差阵）达到最小作为最优估计的准则。估计误差与z不相关，也就是测量不会影响估计误差。
* 极大似然估计：使条件概率密度p（z|x）（称为似然，likehood）达到极大，显然需要知道条件概率密度。一般假设观测之间相互独立，那么最大似然就是将各似然相乘形式，因为概率肯定为正，然后可以转化成对数形式，将相乘转化成相加的形式。在缺乏先验概率p（x）的情况下，极大似然估计和极大后验估计是等同的。相对后验估计，简单、应用更广泛。
* 极大后验估计：使后验概率密度p（x|z）（后验，posterior）达到极大。可利用贝叶斯公式。在没有先验概率时候，可假定x在很大范围内变化，可看成方差矩阵区域无穷大的高斯分布。

<[3]>

1. 计算机解决科学计算过程：提出实际问题，建立数学模型，选用数值计算方法，设计程序以及上机计算求出数值结果。
2. Gramer法则直接计算矩阵解需要做（n+1）！（n-1）次乘除法，对于20阶矩阵需要10的21次方次乘除，而gauss消去法只需要3000次。

计算机加减乘除计算时间对比：

<http://mdyblog.blog.163.com/blog/static/1061501192010675531423/>

1. 数值分析方法几个重要指标：

* 面向计算机：根据计算机特点设计，计算机能直接处理加减乘除和逻辑运算
* 有可靠的理论分析：能任意逼近并达到精度要求，对近似算法要保证收敛性和数值稳定性，还要对误差进行分析
* 有好的计算复杂性：时间复杂度和空间复杂度
* 有数值实验

1. 误差来源：

* 模型误差：数学模型对实际问题进行了抽象和简化，是数学模型和实际问题之间的误差
* 观测误差：测量仪器带来的误差
* 截断误差：也称方法误差。如级数求和，用有限的计算过程代替无限级数求和。如lagrange的误差限就是阶段误差
* 舍入误差：字长有限

1. 浮点数的存储方式，加减运算时会对阶，结合P8理解，指数位对应阶

<https://www.cnblogs.com/jillzhang/archive/2007/06/24/793901.html>

1. 数值计算中的若干原则p8：目的是让结果更精确

* 避免两个相近的数相减：这两个数都是近似值，结果也是近似值。见p5例题
* 防止大数“吃掉”小数：计算机进行加减运算时需要对阶（p6阶的定义）和规格化，小数阶数需和大数对齐，从而小数会有舍入误差，相差太大时，可能直接变为0了。计算时注意安排好顺序
* 避免绝对值太小的数作为分母：
* 注意简化计算步骤，减少运算次数：不但可以减少计算时间还能减少舍入误差
* 设计或选用数值稳定性好的算法：数值稳定是指如果初始数据有小的误差仅使最终计算结果产生小的误差。

1. 线性方程的求解方法

* 数值解法：
* 直接法：解析方法，没有舍入误差时求得精确解。优点是直观，计算量少；可事先估计计算量。缺点是存储量大，程序结构复杂。适用于小规模的线性方程组，特别是低阶稠密矩阵。
* Gauss消去法：消元过程+回代过程。元指的是矩阵的各个元素，消元是指将某些或某个元素通过矩阵变换变成0，对角线上元素称为主元。lik为第k步（第i=k+1行）的乘子。

重要条件：gauss消去法（不作行交换）能进行到底的条件是各步主元均不为0。各步主元均不为零的充要条件是系数矩阵A的顺序主子式不为零。

* Gauss列主元消去法：第k步消元时通过行交换选取k-1步结果中绝对值最大得作为主元。避免绝对值小的数或0值主元带来得舍入误差或无法计算。是在一列中选绝对值最大的元素，所以叫列主元。
* 矩阵三角分解法：将A阵分解为两个三角角阵相乘，即A=LU。当L为单位下三角阵（对角线为1，上三角都为0）时为Doolittle分解，当U为单位上三角阵时为Crout分解。再将A=LDR，其中R=D-1U，L、R分别为单位上下三角矩阵，为LDR分解。三角分解不唯一，Doolittle分解、Crout分解、LDR分解唯一。
* 直接分解法：使用guass消元法直接求得LU，此时求出的就是Doolittle分解
* Doolittle分解：直接计算比较麻烦，使用A=LU来分解，假设条件就是L是单位下三角矩阵。A各元素已知，逐行求U的元素，逐列求L的元素。
* 按列选主元的Doolittle分解：每步取列最大值，减小舍入误差
* 平方根法：解系数矩阵为实对称正定矩阵的情况。实对称矩阵正定的充要条件是存在n阶非奇异下三角阵L使得A=LLT。称为cholesky分解。L主对角元素为正数时，分解唯一。为了避免开方运算可使用改进的平方根法。
* 追赶法：在微分方程数值求解、样条函数的计算中会遇到P34中的矩阵。
* 1
* 迭代法：优点是所需存储单元较少，程序结构简单，原始系数矩阵在迭代过程中始终保持不变。缺点是存在收敛性和收敛速度的问题。适合解大型系数矩阵方程组，特别是某些偏微分方程离散化后得到的大型稀疏方程组。k+1前一步的x是已知的。相当于整个方程组只有当前变量是未知的。X(k+1)=Bx(k)+f。B称为迭代矩阵。迭代终止条件是相邻两次迭代值的差（选择条件方便），很小时迭代终止，但是误差大小还和B的谱半径有关
* Jacobi 迭代法：A=D-L-U。对角，下三角取负，上三角取负。
* Gauss-Seidel迭代法：在jacobi基础上将当前迭代求出的变量马上投入使用。收敛更快。
* 逐次超松弛方法（SOR，successive over-relaxation）：对上一种方法的加速，w有点类似于一阶滤波，收敛的必要条件是0<w<2。解大型稀疏矩阵的有效方法之一
* 分块迭代法：
* 共轭梯度法：
* 其它方法：gramer法则、

1. 向量和矩阵范数：舍入误差和原始数据(如自变量、开根号、观测误差的情况)可能是近似值，求得的解一般是近似解。为了便于对线性方程组近似解进行误差估计，以及对迭代法的收敛性进行分析，需要对向量和矩阵的大小引进某种度量，即范数的概念。范数的距离可以用于表示真实值和近似值的差。||x-y||。表示范数，不仅仅是二范数。序列收敛可用于迭代法判断收敛性，一种范数收敛，其它范数均收敛。范数结果都是一个数。考虑常见的1、2、无穷范数，在评价误差和收敛时P41、p44，对于向量（误差向量）向量和、向量平方和、最大元素值，矩阵是误差最大行、列和误差阵转置乘转置特征值的最大值。一般用于求带待求的变量，向量形式比较多。
2. 病态方程组：系数矩阵的微小变化引起解的很大变化，系数矩阵称为病态矩阵。否则称为良态方程或良态矩阵。它是方程组的固有属性，与求解方法无关。
3. 条件数用于判断方程的病态程度，计算条件范数需要计算逆矩阵比较费事。如下一些现象可以作为判别病态矩阵的参考：

* 用列主元消去法求解时出现小主元
* 系数矩阵中的某行或某列近似线性相关，或系数矩阵行列式的值接近于0
* 系数矩阵元素间数量级相差很大，并且无一定规则

改善和减轻病态的方法

* 采用双精度的算术运算
* 对方程组进行预处理，选择适当的非奇异矩阵D、C来左乘和右乘

1. 实际中，常用的近似解判别方法：P49定理2.5.1
2. 实际中谱半径计算复杂，直接计算矩阵范数比较简单。
3. 插值法：当只能通过测量测得一系列离散点或者计算复杂不方便时使用函数表，使用插值法可以在插值点使插值函数和真实函数相等，其他点近似。这里都是用多项式进行插值，只是使用不同形式。多项式各阶导数存在，计算方便（便于计算机计算）。基函数方法就是令已知插值点（包括导数）列表形成对角为1、其余为0的情况。先看要求不高于多少次，再选择合适的插值方式，凑。有时系数不为1，也不是y就得使整体等于函数值或微商值。

* Lagrange插值法：
* Newton插值法：
* 分段插值法：
* Hermite插值法：
* 样条插值法：

1. 插值法、数值逼近和最小二乘法的区别：插值法已知散点自变量和函数值，只在乎插值函数和被插值函数在给定的点相等。数值逼近已知待逼进函数，在乎整个连续区间，根据误差判定范数不同有所不同。最小二乘法已知散点自变量和函数值，是使所有测量点的平方差最小。
2. 函数范数：无穷范数、Euclid范数
3. 非线性方程组求解

单变量

* 二分法：
* 不动点迭代法：
* Newton法：又称切线法，求解非线性方程最有效的方法之一。优点是收敛速度快。缺点是每一步迭代都要计算f（x）的导数值，导数计算通常比函数值复杂，复杂的函数计算导数值较难。
* 割线法：为了克服计算导数的缺点。利用导数公式（差商xk-2和xk-1）来近似导数。只需计算函数值。

多变量

* Newton法：又称切线法，求解非线性方程最有效的方法之一。
* 拟Newton法：用其他方法来近似jacobian矩阵。如类似于割线法的broyden方法及其改进方法（不需要求逆）sherman-morrison。

1. 常微分方程求解问题：实际上，许多微分方程无法求得精确解，只能求得数值解

* Euler公式：
* 后退的Euler公式：
* 两步法
* 梯形公式
* 改进的Euler公式：

1. 1