班　级：油气井工程 2018 级
学　号：2018312032

中国石油大学
CHINA UNIVERSITY OF PETROLEUM

最优化方法课程作业

题　　目　优化方法中的两步迭代
学院名称　石油工程学院
专业名称　油气井工程
姓　　名　孟庆鑫
任课教师　刘建军

优化方法中的两步迭代

孟庆鑫

摘　要: 这不是一篇论文, 因此文章相对轻松. 本文旨在阐述我对一类优化方法中共同的两步迭代思想的理解: 两步迭代是一种反馈策略, 正向过程逼近解, 反馈过程调整逼近方向, 迭代不动点即为最优解. 前半部分是对算法的理解, 不涉及算法细节, 后半部分是应用举例, 附件包含部分举例代码.

1　从梯度下降开始

对于求 $\arg\min\limits_{\theta\in\mathbb{R}^n} J(\theta)$, 梯度下降的核心是如下迭代:

$$\hat{\theta} \longleftarrow \hat{\theta} - \alpha \frac{\partial}{\partial\theta} J\left(\hat{\theta}\right),$$

这可以看作是两个过程: 修改 $\hat{\theta}$ 使得 $J(\theta)$ 下降; 确定如何修改（即负梯度）.

梯度下降是一个大家族, 有几个变体: 批量梯度下降（*Batch Gradient Descent*）、随机梯度下降（*Stochastic Gradient Descent*, 这两个都是针对样本的）、牛顿法（当 α 取 *Hessian* 矩阵的逆）、共轭梯度（最优步长梯度下降前后两次方向正交, 而共轭梯度前后两次方向共轭正交, 可以看作是在 *Hilbert* 空间上的另一种内积定义）.

神经网络的反向传播也是通过梯度下降完成的. 正向过程通过嵌套输出估计, 反馈过程通过梯度下降减少估计误差.

梯度, 本质上是待优化目标 $J(\theta)$ 的结构特征, 是一种局部信息, 是在局部尺度上以贪心策略最速下降, 但当 $J(\theta)$ 结构无法获取的时候, 梯度也就无法获取（即信息缺失）, 这时只需近似找到使得 $J(\theta)$ 下降的方向即可.

2　PSO, *k*-means, ICP

这三个算法都不是通过梯度构建的, 但都是两步迭代的优化方法.

2.1 PSO

PSO 没有使用梯度信息, 转而使用一群点来弥补信息缺失: 对一组解 x_i 及其对应的优化方向 v_i, $\forall\, i$ 迭代执行

$$\begin{aligned} v_i &\longleftarrow \omega \cdot v_i + c_1 \cdot \mathrm{rand} \cdot (p_i - x_i) + c_2 \cdot \mathrm{rand} \cdot (g_i - x_i), && \text{确定如何修改,} \\ x_i &\longleftarrow x_i + v_i, && \text{修改 } x_i\text{,} \end{aligned}$$

直到收敛. 其中 p_i 为 x_i 迭代至当前的最优值, g_i 为所有 x_i 迭代至当前的最优值.

2.2 *k*-means

聚类问题中, 给定点集 $\left\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\right\}$, 确定 k 个聚类中心 $\mu_1, \mu_2, \cdots, \mu_k$, 使得在距离 ρ 下

$$J(c, \mu) = \sum_{i=1}^{m} \rho\left(x^{(i)}, \mu_{c^{(i)}}\right)$$

达到最小, 其中 $c^{(i)}$ 为 $x^{(i)}$ 的分类号.

k-means 在随机初始化 $\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_k$ 后迭代执行

$$\begin{aligned} &\forall\, i, \quad c^{(i)} \longleftarrow \arg\min_j \rho\left(x^{(i)}, \hat{\mu}_j\right), && \text{确定每个点的分类号,} \\ &\forall\, j, \quad \hat{\mu}_j \longleftarrow \frac{\sum_{i:c^{(i)}=j} x^{(i)}}{\mathrm{card}\left\{i \mid c^{(i)} = j\right\}}, && \text{修改聚类中心,} \end{aligned}$$

直到收敛.

2.3 ICP

ICP（*Iterative Closest Point* 迭代最近点算法）用于刚体变换下点云数据的配准. 考虑如下问题: 一组点 $\{y_j\}$ 发生微小错动变成另一组点 $\{x_j\}$, 试问什么样的刚体变换 T 能变回 $\{y_j\}$？即求刚体变换 $T\colon x_j \to y_j$. 如果 $\{x_j\}$ 发生了重采样, 成为 $\{x_i\}$, 那么问题变为求刚体变换 T, 使得

$$J(T) = \sum_i \min_j \rho\left(T x_i, y_j\right)$$

达到最小, 其中 ρ 为距离.

ICP 在初始估计 $\widehat{T}$ 下对 $\forall\, i$ 迭代执行

$$\begin{aligned} c_i &\longleftarrow \arg\min_j \rho\left(\widehat{T} x_i, y_j\right), && \text{确定估计变换下最近像点脚标,} \\ \widehat{T} &\longleftarrow \arg\min_T \sum_i \rho\left(T x_i, y_{c_i}\right), && \text{重估变换,} \end{aligned}$$

直到收敛. 其中重估变换步骤, 当所求问题为 Euclid 空间 2 范数时, 等价于最小二乘解.

3 应用举例

3.1 信号时差定位

$\mathbb{R}^n$ 空间的 $m+1$ 个点 $\{A_0, A_1, \cdots, A_m\}$, 其中 $m > n$. 未知点 P 发生一事件, 匀速扩散. A_0 先检测到, 并以此时刻为基准, A_i ($i = 1, 2, \cdots, m$) 延迟 t_i 秒检测到. 如何求取 P? （*此问题改自 2012 年百度算法大赛*）

将问题写成优化

$$\arg\min_{P\in\mathbb{R}^n} \sum_{i=1}^{m} \Big(\|P - A_i\| - \|P - A_0\| - vt_i \Big)^2 .$$

注意到这是凸优化问题, 存在全局最优解, 于是可构造下面的迭代（*不是梯度下降*）

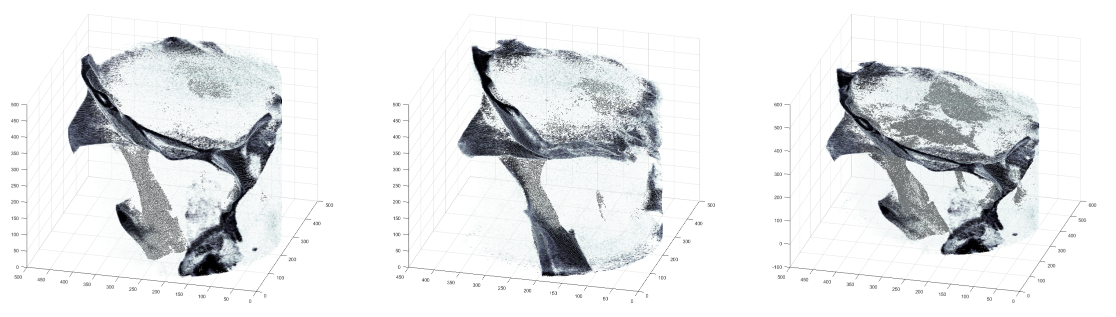
$$\widehat{P} \longleftarrow \widehat{P} - \alpha \sum_{i=1}^{m} \left(\|\widehat{P} - A_i\| - \|\widehat{P} - A_0\| - vt_i \right) \left(\widehat{P} - A_i \right),$$

只需要调整步长 α 即可使 $\widehat{P}$ 收敛到最优解.

构造迭代的理由很简单: 如果 $\hat{t}_i > t_i$, 那就让 $\widehat{P}$ 向 A_i 靠近一点就好了.

3.2 医学影像配准

随病变进程的两次影像需要配准, 增强 MRI、增强 CT 两次扫描图像有配准需求. 医学影像不能直接拿来配准, 一种有效的方法是依照骨密度重抽样, 用抽样点做配准, 配准算法采用 ICP. 下图是我之前做的左侧颞下颌配准, 两次影像结果（*1 和 2*）合并成一个图（*3*）. 具体细节不在此讨论, 只需注意 ICP 本身并不稳定, 依赖于初始位置关系.



3.3 航迹规划

问题抽象如下: 从 P_0 出发, 依次经过 $P_1, P_2, \cdots$, 到 P_n 的曲线中, 求曲率不大于 ρ 的最短曲线. 这一问题应用于车辆路径规划、飞行器空间轨迹规划.

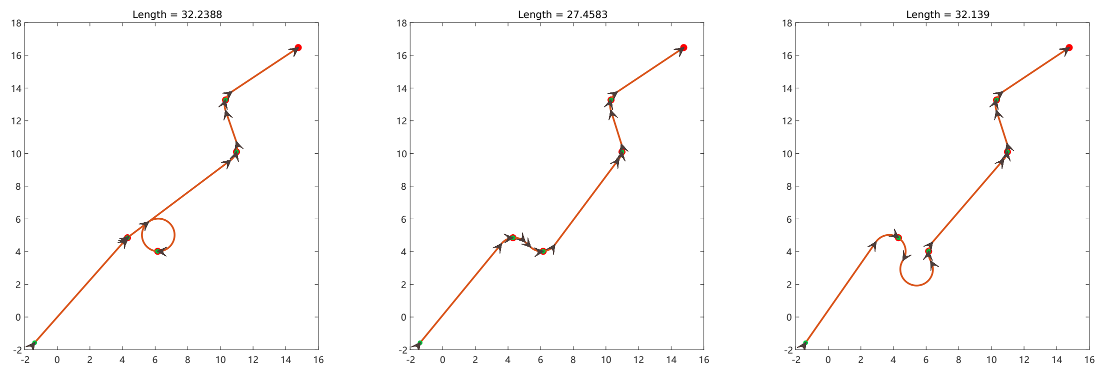
首先考虑 Dubins 曲线问题. Dubins 曲线是在满足曲率不大于 ρ 和给定起点、终点及起点和终点切线方向的条件下的最短路径, 并且路径任何一点切线方向连续变化. Dubins 曲线问题已被解决, 而本例只要确定了每个点的切线方向, 则航迹曲线就是一组首尾相接的 Dubins 曲线. 问题关键是确定每个点的切线方向角 θ_i.

假设在最短路径下, 考虑路径是一条软绳, 两侧拉动软绳, 在曲率圆限制下, 软绳应处于平衡状态. 由此可知, 对于给定 $\hat{\theta}_i$, 正向过程直接求 Dubins 曲线, 反馈过程通过力矩微调 $\hat{\theta}_i$, 即 $\forall\, i$, 初始化 $\hat{\theta}_i$ 后迭代执行

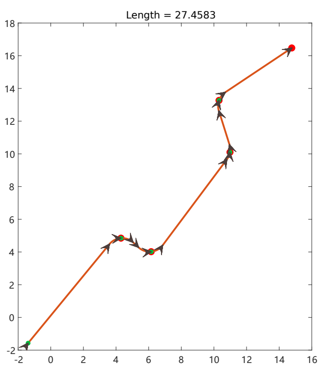
$$C \longleftarrow \mathrm{DubinsCurve}\left(\{\, P_i \,\}, \{\, \hat{\theta}_i \,\}\right),$$
$$\hat{\theta}_i \longleftarrow \hat{\theta}_i + \alpha\, \mathrm{TorqueAdjust}\,(C)\,,$$

直到收敛, 其中 α 为调整步距.

在不同初始 $\hat{\theta}_i$ 下, 迭代收敛到不同的曲线, 如下图. 这说明此问题对初值敏感, 易陷入局部最优.



为解决此问题, 考虑用多初值. 但多初值有 PSO 算法可用, 下图的结果是 PSO 算法实现的, 结果相对稳定很多. *本例代码见附件, Dubins 曲线自 MATLAB R2018b 开始提供.*



4 结语

这部分是我对两步迭代的理解. 对于梯度下降而言, 局部梯度信息驱动着迭代进行, 正因为梯度只关心目标值是否下降, 因此非凸优化问题, 可能陷入局部最优解; k-means 和 ICP 可视作坐标交替上升, 其坐标就是迭代中的赋值量, 由内在逻辑矛盾驱动; 而 PSO 是用已知的最优解驱动迭代进行.

两步迭代将优化问题转换为迭代不动点问题. 只需要构造一个法则, 使得每一次迭代都向最优值靠近即可. 这也说明, 同一个优化问题, 改变目标函数, 改变迭代策略, 可以保证最优解不变.

对需要指定初值的问题, 不同初值对应不同最优值, 整个可行域被一组局部最优值划分为若干域, 而初值选取恰当与否, 在于是否能够提升选在全局最优值对应域内的概率.

如果您对本文及附件代码有疑问, 可与我联系,
Email: Qingxin6174@gmail.com,
或者您也可以扫描右侧的微信二维码.