拉格朗日乘数法

Why —— 为什么要用拉格朗日乘数?

- **现实需求**
  在工程、经济学、机器学习等场景中，我们经常想在满足约束（预算、资源、几何条件……）的同时，把目标函数（成本、能量、误差……）调到最优。
- **常规微分法的局限**
  传统的“令梯度为 0”只适用于无约束；一旦有约束，合法解集往往是曲线/曲面，而梯度为 0 的点可能根本不在其中。
- **核心思想**
  “在约束曲面上取得极值时，目标函数的梯度一定与约束的梯度共线”。拉格朗日乘数法把这一几何事实写成代数方程，从而把“有约束”问题化成“无约束”问题。

What —— 拉格朗日乘数法是什么?

> **定义**：给定

$$> \min\ f(\mathbf{x}) \quad \text{s.t.} \quad g_i(\mathbf{x}) = 0,\ i = 1, \dots, m >$$

> 构造拉格朗日函数

$$> \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}), >$$

> 把未知量扩展为 $(\mathbf{x}, \boldsymbol{\lambda})$。极值点满足

$$> \nabla_{\mathbf{x}} \mathcal{L} = 0, \qquad g_i(\mathbf{x}) = 0. >$$

- λ_i **的角色**：
  几何上 是把两束梯度“拉齐”所需的比例系数；
  物理/经济学 上常解释为“影子价格”或“约束的边际代价”。
- **多重约束**：只需为每个等式约束配一个乘数；不等式约束则扩展为 KKT 条件。

How —— 具体怎么用?

1. **列出目标函数** f **和约束** g_i
2. **写拉格朗日函数** $\mathcal{L}$
3. **求偏导并联立**

$$\frac{\partial \mathcal{L}}{\partial x_j} = 0,\ j = 1, \dots, n; \qquad g_i(\mathbf{x}) = 0$$

4. **解方程组** 得到候选点 $(\mathbf{x}^{\backslash *}, \boldsymbol{\lambda}^{\backslash *})$
5. **判别极性**
   - **Hessian 检验**：在约束流形的切空间中考查二阶导；
   - **直接比较**：若问题规模小，可把候选点代回 f 比大小；
   - **KKT Matrix**：对等式约束可用二阶充分条件。

示例：
最小化 $f(x,y)=x^2+y^2$，约束 $g(x,y)=x+y-1=0$。

$$\mathcal{L}=x^2+y^2+\lambda(x+y-1)$$

偏导方程

$$\begin{cases}2x+\lambda=0\\2y+\lambda=0\\x+y-1=0\end{cases}\Longrightarrow x=y=\tfrac{1}{2},\ \lambda=-1$$

代回 $f=\frac{1}{2}$，因二次型正定可判为全局最小。

How good —— 方法有什么优缺点?

维度	优势	局限 / 注意点
理论统一性	把有约束优化归结为无约束，便于沿用梯度法、牛顿法等算法框架。	对不等式还需 KKT；约束若不光滑需次梯度理论。
几何解释直观	“梯度共线”让人容易在图像上理解，便于教学与思考。	在高维难以直观可视化。
计算实现	方程组可用符号运算或数值求解，省去显式消元。	方程非线性时可能存在多组解、虚解或求解困难。
扩展性	可推广到： - 不等式 → KKT - 多目标 → 权重法 + 乘子 - 拉格朗日对偶 → 现代凸优化	乘子个数随约束增多而爆炸；大型稀疏问题更偏好原始-对偶或投影梯度。

使用建议

- 问题规模 < 5–10 维、约束数目有限时，拉格朗日法最直接、最具解析性；
- 高维/稀疏/凸问题 → 建议转向对偶方法或数值优化库（CVX, IPOPT 等）；
- 碰到奇异点或约束梯度线性相关，要检查 LICQ（线性独立约束资格）。

小结

拉格朗日乘数法的核心是一条几何事实：“在约束曲面上的极值点，目标函数梯度无法再沿切空间下降，因此必与约束梯度共线。”通过把这一事实翻译成拉格朗日方程，我们获得了一套统一且可计算的求解框架，并在现代最优化理论中演化为 KKT 条件和对偶理论。掌握 Why-What-How-How good 的思路，既能准确使用该工具，也能判断何时换用更适合的大规模数值方法。

有没有想亲手做一道更复杂的例题，或者看看它与 KKT 条件的联系？告诉我，我可以进一步演示或做图！