运筹学与最优化课程设计

贾泽慧

南京信息工程大学　数学与统计学院

2022年12月2日

数学优化问题：矩阵完备化(Matrix Completion)

- 问题：对于数据缺失、损坏等问题，正确并高效地从不完整的、带有损毁的数据中恢复出完整的数据，且假设高维数据间具有相关性和冗余性.

数学优化问题：矩阵完备化(Matrix Completion)

- 问题：对于数据缺失、损坏等问题，正确并高效地从不完整的、带有损毁的数据中恢复出完整的数据，且假设高维数据间具有相关性和冗余性.

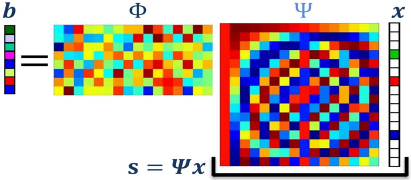
- 优化模型：

$$
\begin{aligned}
\min_{X} \quad & rank(X) \\
\text{s.t.} \quad & X_{ij} = M_{ij},\ (i,j) \in \Omega
\end{aligned}
\tag{1}
$$

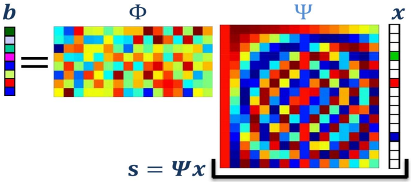
- 应用：推荐系统、社交网络等.

数学优化问题：压缩感知(Compressive Sensing)

- 问题：



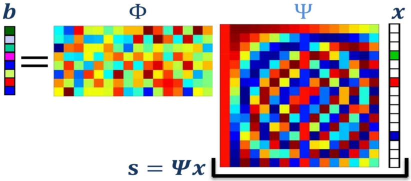
- 问题：



- 优化模型：

$$\begin{aligned} \min_{x} \quad & ||x||_0 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{2}$$

- 问题：

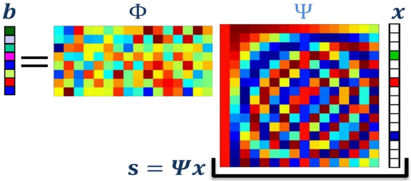


- 优化模型：

$$\begin{aligned} \min_{x} \quad & ||x||_1 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{3}$$

数学优化问题：压缩感知(Compressive Sensing)

- 问题：



- 优化模型：

$$\begin{aligned} \min_{x} \quad & ||x||_1 \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{3}$$

- 应用：核磁共振成像、图像处理等领域.

数学优化问题：机器学习中的优化问题

- 文本分类
  - 样本(x_i, y_i) , $i \in \{1, ..., n\}$, x_i对应文本特征, y_i对应文本x_i是否属于某类标签, 若属于则取1, 否则取-1.
  - $h(x)$为预测函数, $\ell(\cdot)$为度量函数.

数学优化问题：机器学习中的优化问题

- 文本分类
  - 样本(x_i, y_i) , $i \in \{1, ..., n\}$, x_i对应文本特征, y_i对应文本x_i是否属于某类标签, 若属于则取1, 否则取-1.
  - $h(x)$为预测函数, $\ell(\cdot)$为度量函数.

$$R_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell[h(x_i) \neq y_i], \quad \ell[A] = \begin{cases} 1, & if\ A\ is\ true, \\ 0, & otherwise. \end{cases}$$

数学优化问题：机器学习中的优化问题

- 文本分类
  - 样本(x_i, y_i)，$i \in \{1, ..., n\}$，x_i对应文本特征，y_i对应文本x_i是否属于某类标签，若属于则取1，否则取-1.
  - $h(x)$为预测函数，$\ell(\cdot)$为度量函数.

$$R_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell[h(x_i) \neq y_i], \quad \ell[A] = \begin{cases} 1, & if\ A\ is\ true, \\ 0, & otherwise. \end{cases}$$

$$h^{rote}(x) = \begin{cases} y_i, & if\ x = x_i\ for\ some\ i \in \{1, ..., n\}, \\ \pm 1, & otherwise. \end{cases}$$

- 文本分类
  - 样本(x_i, y_i), $i \in \{1, ..., n\}$, x_i对应文本特征, y_i对应文本x_i是否属于某类标签, 若属于则取1, 否则取-1.
  - $h(x)$为预测函数, $\ell(\cdot)$为度量函数.

$$R_n(h) = \frac{1}{n}\sum_{i=1}^{n} \ell[h(x_i) \neq y_i], \;\; \ell[A] = \begin{cases} 1, & if\ A\ is\ true, \\ 0, & otherwise. \end{cases}$$

$$h^{rote}(x) = \begin{cases} y_i, & if\ x = x_i\ for\ some\ i \in \{1, ..., n\}, \\ \pm 1, & otherwise. \end{cases}$$

$$h(x; w, \tau) = w^T x - \tau.$$

数学优化问题：机器学习中的优化问题

- 优化模型

$$\min_{(w,\tau)\in\mathcal{R}^d\times\mathcal{R}} \frac{1}{n}\sum_{i=1}^{n} \ell\left(h\left(x_i; w, \tau\right), y_i\right) + \frac{\lambda}{2}\|w\|_2^2$$

$$\ell(h,y) = log(1 + exp(-hy))$$ log-loss function

$$\ell(h,y) = \max(0, 1 - hy)$$ hinge loss function

数学优化模型

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & h_i(x) \leq 0, \quad i = 1, ..., m, \\
& g_j(x) = 0, \quad j = 1, ..., p.
\end{aligned}
\tag{4}
$$

- $x \in \mathcal{R}^n$ 为决策变量
- $f : \mathcal{R}^n \to \mathcal{R}$ 为目标函数
- $h_i : \mathcal{R}^n \to \mathcal{R}, i = 1, ..., m$ 为不等式约束函数
- $g_j : \mathcal{R}^n \to \mathcal{R}, j = 1, ..., p$ 为等式约束函数
- $\Omega = \{x \in \mathcal{R}^n \mid h_i(x) \leq 0, \ \ i = 1, ..., m; \ g_j(x) = 0, \ \ j = 1, ..., p\}$ 为可行域

数学优化模型

线性优化

$$
\begin{aligned}
\min_{x} \quad & \langle c, x \rangle \\
\text{s.t.} \quad & Ax = b \leq 0, \\
& x \geq 0.
\end{aligned} \tag{5}
$$

- 求解方法：单纯形法、内点算法等。

数学优化算法

非线性无约束优化

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned}$$

- 针对无约束优化问题，常用收敛准则

$$\frac{f(x^k) - f^*}{\max\{|f^*|, 1\}} \le \epsilon_1, \quad \|\nabla f(x^k)\| \le \epsilon_2,$$

其中ϵ_1，ϵ_2为给定的很小的正数，f^*为函数f的最小值（假设已知或者以某种方式估计得到）以及$\nabla f(x^k)$表示函数f在点x^k 处的梯度（光滑函数在局部最优点处梯度为零向量）。

- 针对无约束优化问题，常用停机准则

$$\frac{\|x^{k+1} - x^k\|}{\max\{\|x^k\|, 1\}} \le \epsilon_3, \quad \frac{|f(x^{k+1}) - f(x^k)|}{\max\{|f(x^k)|, 1\}} \le \epsilon_4$$

其中ϵ_3，ϵ_4为给定的很小的正数。

- 收敛速度：
  设$\{x^k\}$为算法产生的迭代点列且收敛于x^*，若对充分大的k有

$$\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}\leq a,\ \ a\in(0,1),$$

则称算法（点列）是$Q-$线性收敛的；若满足

$$\lim_{k\to\infty}\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}=0,$$

称算法（点列）是Q-超线性收敛的；若满足

$$\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|^2}\leq a,\ \ a>0,$$

则称算法（点列）是Q-二次收敛的.

非线性无约束优化 (f光滑)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{6}$$

- 梯度下降法

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) \tag{7}$$

$$x^{k+1} = \arg\min_{x} \{ f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2\alpha_k} \|x - x^k\|^2 \}$$

非线性无约束优化 (f光滑)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{6}$$

- 梯度下降法

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) \tag{7}$$

$$x^{k+1} = \arg\min_{x} \{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2\alpha_k} \|x - x^k\|^2\}$$

  - $\alpha_k = \arg\min_{\alpha} \{f(x^k - \alpha \nabla f(x^k))\}$
  - $\alpha_k = \frac{{s^k}^T s^k}{{s^k}^T y^k}$, 其中$s^k = x^k - x^{k-1}, y^k = \nabla f(x^k) - \nabla f(x^{k-1})$

- 定理(梯度下降法在凸函数上的收敛性)
  设函数$f(x)$为凸的梯度$L-$利普希茨连续函数，$f^* = f(x^*) = \inf_x f(x)$存在且可达．如果步长$\alpha_k$取为满足$0 < \alpha < \frac{1}{L}$的常数, 那么由梯度下降法得到的点列$\{x^k\}$的函数值收敛到最优值，且在函数值的意义下收敛速度为$O(\frac{1}{k})$，也就是说
  $$f(x^k) - f^* \leq O(\frac{1}{k}).$$
- 定理(梯度下降法在强凸函数上的收敛性)
  设函数$f(x)$为m-强凸的梯度L-利普希茨连续函数，$f^* = f(x^*) = \inf_x f(x)$存在且可达．如果步长$\alpha_k$ 取为满足$0 < \alpha < \frac{1}{m+L}$的常数, 那么由梯度下降法迭代得到的点列$\{x^k\}$的函数值收敛到最优值，且为$Q-$线性收敛.

非线性无约束优化 (f光滑)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{8}$$

- 牛顿法

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) \tag{9}$$

$$x^{k+1} = \arg\min_{x}\{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k)\}$$

- 定理(牛顿法的收敛性)

  设目标函数f二阶连续可微，且海瑟矩阵在最优值点x^*的一个邻域$N_\delta(x^*)$内是利普希茨连续的. 如果函数$f(x)$在x^*处满足

$$\nabla f(x^*) = 0,\ \nabla^2 f(x^*) \succeq 0,$$

  则对于牛顿法有如下结论:

  (1)如果初始点离x^*足够近，则迭代点列$\{x^k\}$收敛到x^*;

  (2)收敛到x^*的速度是Q–二次的;

  (3)$\{\|\nabla f(x^k)\|\}$$Q$–二次收敛到0.

非线性无约束优化 (f光滑)

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & x \in \mathcal{R}^n
\end{aligned} \tag{10}
$$

- 拟牛顿法

$$x^{k+1} = x^k - \alpha_k H_k \nabla f(x^k) \tag{11}$$

  - H_k更新方式：秩一更新、BFGS公式、DFP等.
  - BFGS更新公式：

$$B_{k+1} = B_k + \frac{y^k (y^k)^T}{(s^k)^T y^k} - \frac{B_k s^k (B_k s^k)^T}{(s^k)^T B_k s^k}, \ B_k = H_k^{-1},$$

其中$s^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$。

- 定理(BFGS收敛性)

  设目标函数f是二阶连续可微的函数，且海瑟矩阵在最优值点x^*的一个邻域$N_\delta(x^*)$内是利普希茨连续的. 且使用BFG-S迭代格式收敛到f的最优值点x^*. 若点列$\{x^k\}$满足

$$\sum_{k=1}^{\infty} \|x^{k+1} - x^k\| < \infty,$$

  则$\{x^k\}Q-$超线性收敛到x^*.

非线性无约束优化 (f非光滑)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{12}$$

- 邻近点算法(PPA)

$$x^{k+1} = \arg\min_{x}\{f(x) + \frac{1}{2\alpha_k}\|x - x^k\|^2\} := \text{prox}_{\alpha_k f}(x^k)$$

- 定理(PPA的收敛性)

  设目标函数f是适当的闭凸函数，f的最小值存在且可以达到，设$\{x^k\}$是由PPA产生的序列，则

$$f(x^k) - f(x^*) \le \frac{1}{2\sum_{i=1}^{k}\alpha_i}\|x^0 - x^*\|^2.$$

非线性无约束优化 (f光滑,g非光滑)

$$\begin{aligned} \min_x \quad & F(x) = f(x) + g(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{13}$$

- 邻近梯度算法(PGA)

$$\begin{aligned} x^{k+1} &= \arg\min_x \{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + g(x) + \tfrac{1}{2\alpha_k}\|x - x^k\|^2\} \\ &:= \mathrm{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)) \end{aligned}$$

- 定理(PGA的收敛性)

  设函数$f(x)$为凸的梯度$L-$利普希茨连续函数，h是适当的闭凸函数，F的最小值存在且可以达到，设$\{x^k\}$是由PGA产生的序列，则$F(x^k) - F(x^*) \leq \frac{1}{2k\alpha}\|x^0 - x^*\|^2$.

非线性无约束优化 (f光滑)

$$\begin{aligned} \min_{x} \quad & f(x) = \textstyle\sum_{i=1}^{n} f_i(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{14}$$

- 随机梯度法

$$x^{k+1} = x^k - \alpha_k \nabla f_{i_k}(x^k) \tag{15}$$

$$x^{k+1} = \arg\min_{x} \{ f(x^k) + \langle \nabla f_{i_k}(x^k), x - x^k \rangle + \frac{1}{2\alpha_k} \|x - x^k\|^2 \}$$

非线性无约束优化 (f光滑)

$$\begin{aligned} \min_{x} \quad & f(x) = \textstyle\sum_{i=1}^{n} f_i(x) \\ \text{s.t.} \quad & x \in \mathcal{R}^n \end{aligned} \tag{16}$$

- mini-batch 随机梯度法

$$x^{k+1} = x^k - \alpha_k \sum_{i \in B_k} \nabla f_i(x^k) \tag{17}$$

$$x^{k+1} = \arg\min_{x} \{ f(x^k) + \langle \sum_{i \in B_k} \nabla f_i(x^k), x - x^k \rangle + \frac{1}{2\alpha_k} \|x - x^k\|^2 \}$$

- 方差缩减随机梯度法(SVRG,SAGA,SARAH,SPIDER...)

- 定理(随机梯度法的收敛性)
  假设目标函数f是$\mu-$强凸函数且梯度$L-$李普希茨连续，随机梯度二阶矩是一致有界的，即

$$\mathbb{E}_{i_k}[\|\nabla_{i_k} f(x)\|^2] \le M^2 < +\infty,$$

固定步长$0 < \alpha < \frac{1}{2\mu}$，序列$\{x^k\}$ 由随机梯度法产生，则

$$\begin{aligned}\mathbb{E}[f(x^k) - f(x^*)] \le& \frac{L}{2}\mathbb{E}\|x^k - x^*\|^2 \\ \le& \frac{L}{2}\left((1-2\alpha\mu)^k\|x^1 - x^*\|^2 + \frac{\alpha M^2}{2\mu}\right).\end{aligned}$$

- 定理(随机梯度法的收敛速度)
  在上一定理的结果中取递减的步长$\alpha_k = \frac{\beta}{k+\gamma}$，其中$\gamma \geq \frac{1}{2\mu}$，$\gamma > 0$使得$\alpha_1 \leq \frac{1}{2\mu}$，那么对于任意的$k \geq 1$都有

$$\mathbb{E}[f(x^k) - f(x^*)] \leq \frac{L}{2}\mathbb{E}\|x^k - x^*\|^2 \leq \frac{L}{2}\frac{v}{\gamma + k},$$

其中$v = \max\left\{\frac{\beta^2 M^2}{2\beta\mu - 1}, (\gamma + 1)\Delta_1^2\right\}$.

课程设计题目

- 基追踪问题

基追踪问题是压缩感知中的一个基本问题，可以写为

$$\begin{aligned} \min_{x} \quad & ||x||_1 \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

其中$A \in \mathbb{R}^{m\times n}$，$b \in \mathbb{R}^n$，$x \in \mathbb{R}^n$为待求解的变量. 可考虑其对偶问题将该约束优化问题转化为一个无约束优化问题，但由于此问题的对偶问题的无约束优化形式不是可微的，即无法计算梯度，故考虑如下正则化问题：

$$\begin{aligned} \min_{x} \quad & ||x||_1 + \frac{1}{2\alpha}\|x\|^2 \\ \text{s.t.} \quad & Ax = b, \end{aligned}$$

其中α为正则化参数，

课程设计题目

- 正则化的基追踪问题的对偶问题为

$$\min -b^T y + \frac{\alpha}{2}\|A^T y - P_{[-1,1]^n}(A^T y)\|_2^2$$

其中$P_{[-1,1]^n}$表示到集合$[-1,1]^n$的投影，$A \in \mathbb{R}^{m\times n}$，$b \in \mathbb{R}^n$，$\alpha$为已知参数，$y \in \mathbb{R}^m$为待求解的变量。

- 练习

(1) 阐述压缩感知问题的发展及其应用背景；

(2) 介绍一种求解基追踪问题的对偶问题的优化算法，并给出算法的收敛结果；

(3) 作图展示算法的求解结果（迭代序列与最优解之间的距离随迭代步数的变化图），并对算法的求解结果进行对比、总结、分析。

课程设计题目

- LASSO问题

  LASSO(Least absolute shrinkage and selection operator)问题是一个经典的稀疏优化问题，其通过二次罚函数将基追踪问题表述为一个无约束优化问题，具体形式可以表述为

$$\min_{x} \ f(x) = \frac{1}{2}\|Ax - b\|^2 + \mu\|x\|_1,$$

  其中$x \in \mathbb{R}^n$,$A \in \mathbb{R}^{m\times n}$，$b \in \mathbb{R}^n$，$\mu > 0$为已知参数，且向量$b$的维数远小于向量$x$的维数，即$m \ll n$，表示向量$x$的每个元素的绝对值之和。LASSO通过惩罚参数的$l_1$范数来控制解的稀疏性，如果$x$是稀疏的，那么预测值$b_i$只和矩阵$A$的行向量$a_i$的部分元素相关.。

课程设计题目

- 练习

  针对上述LASSO问题，

  (1) 阐述LASSO问题的发展及其应用背景；

  (2) 介绍梯度算法、邻近梯度算法，并给出算法的收敛结果；

  (3) 采用梯度法（需对目标函数进行光滑化处理）或邻近梯度算法对LASSO问题进行求解；

  (4) 作图展示算法的求解结果（目标函数值随迭代步数的变化图），并对算法的求解结果进行对比、总结、分析。

课程设计题目

- 低秩矩阵恢复

  低秩矩阵恢复模型想要构造一个矩阵X，使得其在给定位置的元素等于想要恢复的矩阵M的元素，但是只知道矩阵M在下标集Ω上的值，该模型具体形式为

$$\min_{M \in \mathbb{R}^{m \times n}} \mu \|X\|_* + \frac{1}{2} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2,$$

  其中μ为已知参数，$\|X\|_*$为核范数，表示矩阵X的奇异值的和，用于保证矩阵的低秩性，而范数平方项可保证矩阵X与想要恢复的矩阵M的元素对应相等。

- 练习

  令$f(x)=\frac{1}{2}\sum_{(i,j)\in\Omega}(X_{ij}-M_{ij})^2$，$g=\mu\|x\|_*$，

  (1)阐述低秩矩阵恢复问题的发展及其应用背景；

  (2)介绍邻近梯度算法，并给出算法的收敛结果；

  (3)作图展示算法的求解结果（函数值随迭代步数的变化图），并对算法的求解结果进行对比、总结、分析。

- 逻辑回归

逻辑回归模型，其在线性回归的基础上，继续套用了一层函数，用于解决二分类问题，并因其简单性、可并行化、可解释性强深受各界喜爱。下面给出其对应的优化问题

$$\min_{x} \frac{1}{m} \sum_{i=1}^{m} \ln(1 + \exp(-b_i a_i^T x)) + \lambda \|x\|^2,$$

其中$\lambda = \frac{1}{100m}$，a_i，b_i为已知的待分类的数据集，x为待求解的变量。

课程设计题目

- 练习

  (1)阐述逻辑回归问题的发展及其应用背景，并将逻辑回归与其他分类方法对比（如支持向量机等）指出其优缺点；

  (2)介绍一种求解该问题的优化算法，并给出收敛结果；

  (3)作图展示算法的求解结果（目标函数值或梯度范数随迭代步数的变化图），并对算法的求解结果进行对比、总结、分析。

Thank you!