author：20S003127 史佳欣

第1-3题

见压缩包`.ipynb`源码文件

第4题

(a)

$x=(0,0,0), f(x)=max\{f_1(0),f_2(0),f_3(0)\}=max\{|0|,|0|,|0|\}=0$，则$I(x)=\{1,2,3\}$，有$\partial f_i(x)=[-1,1], i=1,2,3$，则$f(x)$的次微分为
$conv\cup_{i=1,2,3}\partial f_i(x)=(x_i,y_i,z_i), x_i,y_i,z_i\in[-1,1]$

(b)

$f'(x)$满足

$$f(x)=\begin{cases}-e^{-x}, & x<0\\ e^{x}, & x>0\end{cases}$$

则$f(x)$在(0,0,0)处的次微分为$[-e^0,e^0]=[-1,1]$

(c)

点$(x_1,x_2)=(1,1)$，令$f_1(x_1,x_2)=x_1+x_2-1=1$, $f_2(x_1,x_2)=x_1-x_2+1=1$,
则有

$$f(x)=max\{f_1(x_1,x_2),f_2(x_1,x_2)\}=f_1(x_1,x_2)=f_2(x_1,x_2)=1$$
$$I(x)=\{1,2\}$$
$$\nabla f_1(x_1,x_2)=(1,1)^T$$
$$\nabla f_2(x_1,x_2)=(1,-1)^T$$

$\therefore f(x)$在(1,1)处的次微分为$\partial f(x)=\{v|v=t*(1,1)^T+(1-t)*(1,-1)^T, t\in[0,1]\}$

第5-7题

见压缩包`.ipynb`源码文件

第8题

1. 基本思想

- 无约束最优化问题可表示为

$$min f(\mathbf{x}), \mathbf{x}\in\mathbf{R}^n$$

- 基本思想是不断迭代寻找合适的$\lambda, \mathbf{d}$，使得$f(\mathbf{x}^{(k)}+\lambda\mathbf{d}) < f(\mathbf{x}^{(k)})$，直到满足终止条件，此时的$\mathbf{x}$接近或等于最优解$x^*$；
- 搜索方法分为一维精确搜索和不精确搜索，其中精确搜索有二分法、Fibonacci法、黄金分割法、进退法、牛顿法、插值法等方法；不精确搜索包含Goldstein法、Goldstein-Price法、Wolfe-Powell法等方法，以确定每次迭代的最佳步长$min_\lambda f(\mathbf{x}^{(k)}+\lambda\mathbf{d})$
- 方向的确定有共轭方向法、牛顿法、拟牛顿法、最速下降法、变尺度法（DFP、BFPS）等方法，从而确定$\mathbf{x}$的优化方向；
- 通过确定$\lambda, \mathbf{d}$，使得点不断逼近最优解x^*。

2. 非凸优化转为凸优化：

- 修改目标函数，将非凸函数转为凸函数求解问题，如凸松弛方法
- 抛弃一些约束条件，使新的可行域为凸集并且包含原可行域

3. 有约束条件转为无约束条件

- 使用拉格朗日乘子法，设原问题具有等式约束$\mathbf{h(x)=0}$，则建立拉格朗日函数$L(\mathbf{x,v})=f(\mathbf{x})+\mathbf{v}^T\mathbf{h(x)}$ ，其中$\mathbf{v}$为拉格朗日乘子，令$\mathbf{x}$内的各分量和拉格朗日乘子为变量的梯度等于0，从而构建方程组求解，实现将有约束问题转为无约束问题
- 使用罚函数法、闸函数法、乘子法等方法（这些方法目前没看明白。。。）

第9题

(1)计算公式统一描述：$x_{k+1}=x_k-\lambda_k H_k \nabla f(x_k)$

其中，最速下降法的$H_k=\mathbf{I}$，λ_k使用一维搜索；

牛顿法的$H_k=(H(x_k))^{-1}$，$\lambda_k=1$；

而修正牛顿法的λ_k通过一维搜索得到。

(2)变尺度法是牛顿法的改进，基本思想是使用对称正定矩阵近似Hessian矩阵（的逆），要求满足拟牛顿方程。对称正定矩阵可以对搜索方向保持下降性质，避开了二阶偏导数的大量运算。

第10题

见压缩包`.ipynb`源码文件

第11题

由于时间原因，仅阅读了文献[1]提出的基本思想，没有深入研究，所以了解的不多。

- 基本思想：将X_*进行矩阵分解，形成$X_*=L_*R_*^T$，其中$L_*\in\mathbb{R}^{n_1\times r}$，$R_*\in\mathbb{R}^{n_2\times r}$，$r$为矩阵$X$的秩且事先已知，对损失函数最优化：$min_{\mathbf{L,R}}\sum_{i=1}^m|\mathcal{A}_i(\mathbf{LR}^T)-y_i|$，文章使用扩展(scaled不知道如何翻译)次梯度方法，对L和R进行迭代：

$$
\begin{aligned}
L_{t+1} &:= L_t-\eta_t S_t R_t(R_t^T R_T)^{-1} \\
R_{t+1} &:= R_t-\eta_t S_t L_t(L_t^T L_T)^{-1}
\end{aligned}
$$

- 式中，$S_t\in\partial f(L_tR_t^T)$是损失函数的子梯度，$\eta_t$为步长因子，文章将传统的梯度下降的$\mathbf{x}$迭代转变为矩阵分解后的$L$和$R$的搜索迭代，文章的方法在矩阵秩较低的情况下的非平滑优化问题和scaled一阶方法上具有明显优势。

- 对于步长因子，如果事先知道损失函数的最优值，则使用 Polyak的步长因子[2]，而通用的则使用文献[3]提出的。

参考文献：

[1] Tong T, Ma C, Chi Y. Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number[J]. arXiv preprint arXiv:2010.13364, 2020.

[2] Canon, M., Cullum, C., & Polak, E. (1966). Constrained minimization problems in finite-dimensional spaces. *SIAM Journal on Control*, *4*(3), 528-547.

[3] Goffin J L. On convergence rates of subgradient optimization methods[J]. Mathematical programming, 1977, 13(1): 329-347.

第12题

- 共轭函数：

  设$f:\mathbb{R}^n \to \mathbb{R}$，函数$f^*:\mathbb{R}\to\mathbb{R}$，如果两个函数满足$f^*(y)=\sup\limits_{x\in dom f}(y^Tx-f(x))$，则$f^*$是$f$的共轭函数。

- 对于优化问题的约束为线性条件时，可通过共轭函数得到对偶问题
- 设$minf(x)$的约束条件为$Ax\le B$，$Cx=d$，即原问题包含不等式约束和等式约束。
- 利用拉格朗日对偶函数可得

  $$\begin{aligned} g(\lambda,v)&=\inf_x(f(x)+\lambda^T(Ax-b)+v^T(Cx-d)) \\ &=-b^T\lambda-d^Tv+\inf_x(f_0(x)+(A^T\lambda+C^Tv)^Tx) \end{aligned}$$

- 而共轭函数具有性质：

  $\inf(y^Tx+f_0(x))=-\sup(-y^Tx-f_0(x))=-f^*(-y)$

- 因此对偶函数为 $g(\lambda,v)=-b^T\lambda-d^Tv-f_0^*(-A^T\lambda-C^Tv)$
- 定义域为$dom\ g=\{(\lambda,v):-A^T\lambda-C^Tv\in dom\ f_0^*\}$