

# 第七章 复合优化算法

修贤超

<https://xianchaoxiu.github.io>

- 7.1 近似点梯度法
- 7.2 Nesterov 加速算法
- 7.3 近似点算法
- 7.4 分块坐标下降法
- 7.5 对偶算法
- 7.6 交替方向乘子法
- 7.7 随机优化算法

# 对偶方法

## ■ 梯度法

- 对偶函数可能不可微，或定义域非平凡
- 对原始函数加小的强凸项，将对偶函数光滑化

## ■ 增广拉格朗日法

- 等价于对光滑化的对偶问题做梯度上升
- 但是光滑化会破坏可分结构

## ■ 近似点梯度法

- 一项是梯度利普希茨连续函数
- 另一项有方便计算的近似点算子

# 对偶问题

- 设  $f, h$  是闭凸函数, 考虑如下形式的问题

$$(P) \quad \min_{x \in \mathbb{R}^n} \quad \psi(x) = f(x) + h(Ax)$$

- 引入新变量  $y = Ax$ , 考虑问题

$$(P) \quad \min_{x \in \mathbb{R}^n} \quad \psi(x) = f(x) + h(y) \quad \text{s.t.} \quad Ax = y$$

- 拉格朗日函数为

$$L(x, y, z) = f(x) + g(y) + z^\top (Ax - y)$$

- 对偶问题

$$(D) \quad \max_z \quad \phi(z) = -f^*(-A^\top z) - h^*(z)$$

# 强凸函数共轭函数的性质

- **引理 7.1** 设  $f(x)$  是适当且闭的强凸函数, 强凸参数为  $\mu > 0$ , 则  $f^*(y)$  在全空间  $\mathbb{R}^n$  上有定义,  $f^*(y)$  是梯度  $\frac{1}{\mu}$ -利普希茨连续的可微函数

- 考虑在对偶问题上应用近似点梯度算法, 每次迭代更新如下

$$z^{k+1} = \text{prox}_{th^*}(z^k + tA\nabla f^*(-A^\top z^k))$$

- 引入变量  $x^{k+1} = \nabla f^*(-A^\top z^k)$ , 迭代格式等价于

$$x^{k+1} = \arg \min_x \{f(x) + (A^\top z^k)^\top x\}, \quad z^{k+1} = \text{prox}_{th^*}(z^k + tAx^{k+1})$$

- 如果  $f$  可分,  $x$  的计算可分解为多个独立的问题
- 步长  $t$  可取常数或采取回溯线搜索法
- 可使用加速近似点梯度法

- **引理 7.2** 设  $f$  是定义在  $\mathbb{R}^n$  上的适当的闭凸函数, 则对任意的  $x \in \mathbb{R}^n$  有

$$x = \text{prox}_f(x) + \text{prox}_{f^*}(x)$$

- 或更一般地,

$$x = \text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f^*} \left( \frac{x}{\lambda} \right)$$

- 对任意的闭凸函数  $f$ , 空间  $\mathbb{R}^n$  上的恒等映射总可以分解成两个函数  $f$  与  $f^*$  邻近算子的和

## 交替极小的解释

- 取  $\lambda = t$ ,  $f = h^*$ , 并注意到  $h^{**} = h$ , 有

$$\begin{aligned} z^k + tAx^{k+1} &= \text{prox}_{th^*}(z^k + tAx^{k+1}) + t\text{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \\ &= z^{k+1} + t\text{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \end{aligned}$$

- 由此给出对偶近似点梯度法等价的针对原始问题的更新格式

$$\begin{aligned} x^{k+1} &= \arg \min_x \left\{ f(x) + (z^k)^\top Ax \right\} \\ y^{k+1} &= \text{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \\ &= {}_y \left\{ h(y) - (z^k)^\top (y - Ax^{k+1}) + \frac{t}{2} \|Ax^{k+1} - y\|_2^2 \right\} \\ z^{k+1} &= z^k + t(Ax^{k+1} - y^{k+1}) \end{aligned}$$

# 交替极小方法

## ■ 考虑等价问题

$$\min_{x,y} f(x) + h(y), \quad \text{s.t.} \quad y = Ax$$

## ■ 定义拉格朗日函数和增广拉格朗日函数

$$L(x, y, z) = f(x) + h(y) - z^\top (y - Ax)$$

$$L_t(x, y, z) = f(x) + h(y) - z^\top (y - Ax) + \frac{t}{2} \|y - Ax\|^2$$

## ■ 等价的交替极小格式是

$$x^{k+1} = \arg \min_x L(x, y^k, z^k)$$

$$y^{k+1} = \arg \min_y L_t(x^{k+1}, y, z^k)$$

$$z^{k+1} = z^k + t(Ax^{k+1} - y^{k+1})$$

## ■ 对偶近似点梯度法等价于对原始约束问题使用交替极小化方法



## 例 7.9

- 假设  $f$  是强凸函数,  $\|\cdot\|$  是任意一种范数, 考虑

$$\min_x f(x) + \|Ax - b\|$$

- 对应原始问题我们有  $h(y) = \|y - b\|$

$$h^*(z) = \begin{cases} b^\top z & \|z\|_* \leq 1 \\ +\infty & \text{其他} \end{cases} \quad \text{prox}_{th^*}(x) = \mathcal{P}_{\|z\|_* \leq 1}(x - tb)$$

- 从而对偶问题为

$$\max_{\|z\|_* \leq 1} -f^*(-A^\top z) - b^\top z$$

应用对偶近似点梯度法, 更新如下

$$x^{k+1} = \arg \min_x \{f(x) + (A^\top z^k)^\top x\}$$

$$z^{k+1} = \mathcal{P}_{\|z\|_* \leq 1}(z^k + t(Ax^{k+1} - b))$$

## 例 7.9

### ■ 考虑等价问题

$$\min_{x,y} f(x) + \|y\| \quad \text{s.t.} \quad Ax - b = y$$

### ■ 交替极小化格式

$$x^{k+1} = \arg \min_x f(x) + \|y^k\| + (z^k)^\top (Ax - b - y^k)$$

$$y^{k+1} = \arg \min_y f(x^{k+1}) + \|y\| + (z^k)^\top (Ax^{k+1} - b - y) + \frac{t}{2} \|Ax^{k+1} - b - y\|_2^2$$

$$z^{k+1} = z^k + t(Ax^{k+1} - b - y^{k+1})$$

## 例 7.10

- 假设  $f$  是强凸函数，考虑

$$\min_x f(x) + \sum_{i=1}^p \|B_i x\|_2$$

- 根据  $\|\cdot\|_2$  的共轭函数定义，对偶问题形式如下

$$\max_{\|z_i\|_2 \leq 1} -f^* \left( -\sum_{i=1}^p B_i^\top z_i \right)$$

- 记  $C_i$  是  $\mathbb{R}_{m_i}$  中的单位欧几里得球，对偶近似点梯度法更新如下

$$\begin{aligned} x^{k+1} &= \arg \min_x \left\{ f(x) + \left( \sum_{i=1}^p B_i^\top z_i \right)^\top x \right\} \\ z_i^{k+1} &= \mathcal{P}_{C_i}(z_i^k + t B_i x^{k+1}), i = 1, 2, \dots, p \end{aligned}$$

## 例 7.11

- 假设  $f$  是强凸函数, 集合  $C_i$  为闭凸集, 且易于计算投影, 考虑

$$\begin{array}{ll}\min & f(x) \\ \text{s.t.} & x \in C_1 \cap C_2 \cap \cdots \cap C_m\end{array}$$

- 有  $h(y_1, y_2, \cdots, y_m) = \sum_{i=1}^m I_{C_i}(y_i)$ ,  $A = [I \ I \ \cdots \ I]^\top$

- 对偶问题为

$$\max_{z_i \in C_i} -f^* \left( -\sum_{i=1}^m z_i \right) - \sum_{i=1}^m I_{C_i}^*(z_i),$$

$I_{C_i}^*(z_i)$  是集合  $C_i$  的支撑函数, 其显式表达式不易求出

- 利用 Moreau 分解将迭代格式写成交替极小化方法的形式

$$\begin{aligned}x^{k+1} &= \arg \min_x \left\{ f(x) + \left( \sum_{i=1}^m z_i \right)^\top x \right\} \\y_i^{k+1} &= \mathcal{P}_{C_i} \left( \frac{z_i^k}{t} + x^{k+1} \right), \quad i = 1, 2, \dots, m \\z_i^{k+1} &= z_i^k + t(x^{k+1} - y_i^{k+1}), \quad i = 1, 2, \dots, m\end{aligned}$$

## 例 7.12

- 假设  $f_i$  是强凸函数,  $h_i^*$  有易于计算的邻近算子. 考虑

$$\min \sum_{j=1}^n f_j(x_j) + \sum_{i=1}^m h_i(A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_N)$$

- 其对偶问题形式如下

$$\max - \sum_{i=1}^m h_i^*(z_i) - \sum_{j=1}^n f_j^*(-A_{1j}^\top z_1 - A_{2j}^\top z_2 - \cdots - A_{mj}^\top z_m)$$

- 对偶近似点梯度法更新如下

$$x_j^{k+1} = \arg \min_{x_j} \left\{ f_j(x_j) + \left( \sum_{i=1}^m A_{ij} z_i^k \right)^\top x_j \right\}, \quad j = 1, 2, \cdots, n$$
$$z_i^{k+1} = \text{prox}_{th_i^*} \left( z_i + t \sum_{j=1}^n A_{ij} x_j^{k+1} \right), \quad i = 1, 2, \cdots, m$$

# 鞍点问题

- 令  $f, h$  是适当的闭凸函数. 考虑原始问题

$$\min_x f(x) + h(Ax)$$

- 由于  $h$  有自共轭性, 将问题变形为

$$(\text{LPD}) \quad \min_x \max_z \psi_{PD}(x, z) = f(x) - h^*(z) + z^\top Ax$$

- 另一种常用的鞍点问题定义方式构造拉格朗日函数. 问题

$$\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} f(x) + h(y) \quad \text{s.t.} \quad y = Ax$$

- 相应的鞍点问题形式如下

$$(\text{LP}) \quad \min_{x, y} \max_z f(x) + h(y) + z^\top (Ax - y)$$

- PDHG 算法的思想就是分别对两类变量应用近似点梯度算法
- 以求解问题  $(L_{PD})$  为例, PDHG 算法交替更新原始变量以及对偶变量, 其迭代格式如下

$$\begin{aligned} z^{k+1} &= \arg \max_z \left\{ -h^*(z) + \langle Ax^k, z - z^k \rangle - \frac{1}{2\delta_k} \|z - z^k\|_2^2 \right\} \\ &= \text{prox}_{\delta_k h^*}(z^k + \delta_k Ax^k) \\ x^{k+1} &= \arg \min_x \left\{ f(x) + (z^{k+1})^\top A(x - x^k) + \frac{1}{2\alpha_k} \|x - x^k\|_2^2 \right\} \\ &= \text{prox}_{\alpha_k f}(x^k - \alpha_k A^\top z^{k+1}) \end{aligned}$$

- 始变量和对偶变量的更新顺序是无关紧要的



# Chambolle-Pock 算法

- PDHG 算法的收敛性需要比较强的条件，有些情形下未必收敛.
- Chambolle-Pock 算法与 PDHG 算法的区别在于多了一个外推步
- 具体的迭代格式如下：

$$z^{k+1} = \text{prox}_{\delta_k h^*}(z^k + \delta_k A y^k)$$

$$x^{k+1} = \text{prox}_{\alpha_k f}(x^k - \alpha_k A^\top z^{k+1})$$

$$y^{k+1} = 2x^{k+1} - x^k$$

# 应用举例: LASSO 问题求解

## ■ 考虑 LASSO 问题

$$\min_{x \in \mathbb{R}^n} \psi(x) = \mu \|x\|_1 + \frac{1}{2} \|Ax - b\|_2^2$$

## ■ 取 $f(x) = \mu \|x\|_1$ 和 $h(x) = \frac{1}{2} \|x - b\|_2^2$ , 相应的鞍点问题

$$\min_{x \in \mathbb{R}^n} \max_{z \in \mathbb{R}^m} f(x) - h^*(z) + z^\top Ax$$

## ■ 根据共轭函数的定义

$$h^*(z) = \sup_{y \in \mathbb{R}^m} \left\{ y^\top z - \frac{1}{2} \|y - b\|_2^2 \right\} = \frac{1}{2} \|z\|_2^2 + b^\top z$$

## ■ 应用 PDHG 算法, $x^{k+1}$ 和 $z^{k+1}$ 的更新格式分别为

$$z^{k+1} = \text{prox}_{\delta_k h^*}(z^k + \delta_k Ax^k) = \frac{1}{\delta_k + 1} (z^k + \delta_k Ax^k - \delta_k b)$$

$$x^{k+1} = \text{prox}_{\alpha_k \mu \|\cdot\|_1}(x^k - \alpha_k A^\top z^{k+1})$$

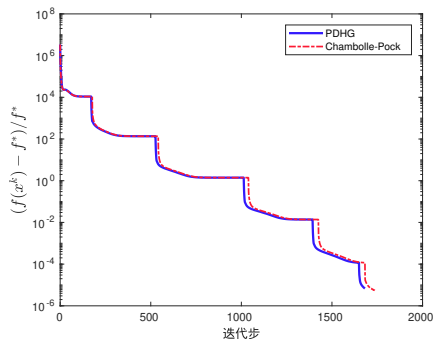
# LASSO 问题求解

## ■ Chambolle-Pock 算法格式为

$$z^{k+1} = \frac{1}{\delta_k + 1} (z^k + \delta_k A y^k - \delta_k b)$$

$$x^{k+1} = \text{prox}_{\alpha_k \mu \|\cdot\|_1} (x^k - \alpha_k A^\top z^{k+1})$$

$$y^{k+1} = 2x^{k+1} - x^k$$



## TV- $L^1$ 模型

- 考虑去噪情形下的 TV- $L^1$  模型 (即  $\mathcal{A}$  为矩阵空间的恒等算子)

$$\min_{U \in \mathbb{R}^{n \times n}} \|U\|_{TV} + \lambda \|U - B\|_1$$

- 对任意的  $W, V \in \mathbb{R}^{n \times n \times 2}$ , 记

$$\|W\| = \sum_{1 \leq i, j \leq n} \|w_{ij}\|_2, \quad \langle W, V \rangle = \sum_{1 \leq i, j \leq n, 1 \leq k \leq 2} w_{i,j,k} v_{i,j,k}$$

- 利用  $\|\cdot\|$  的定义, 有

$$\|U\|_{TV} = \|DU\|$$

- 取  $D$  为相应的线性算子, 并取

$$f(U) = \lambda \|U - B\|_1, \quad U \in \mathbb{R}^{n \times n}, \quad h(W) = \|W\|, \quad W \in \mathbb{R}^{n \times n \times 2}$$

- 相应的鞍点问题如下

$$(\text{LPD}) \quad \min_{U \in \mathbb{R}^{n \times n}} \max_{V \in \mathbb{R}^{n \times n \times 2}} f(U) - h^*(V) + \langle V, DU \rangle$$

- 根据共轭函数的定义

$$h^*(V) = \sup_{U \in \mathbb{R}^{n \times n \times 2}} \{\langle U, V \rangle - \|U\|\} = \begin{cases} 0, & \max_{i,j} \|v_{ij}\|_2 \leq 1 \\ +\infty, & \text{其他} \end{cases}$$

- 记  $\mathcal{V} = \{V \in \mathbb{R}^{n \times n \times 2} \mid \max_{ij} \|v_{ij}\|_2 \leq 1\}$ , 其示性函数记为  $I_{\mathcal{V}}(V)$ , 则问题 (LPD) 可以整理为

$$\min_U \max_V f(U) + \langle V, DU \rangle - I_{\mathcal{V}}(V)$$

- 应用 PDHG 算法, 则  $V^{k+1}$  的更新为

$$V^{k+1} = \text{prox}_{sI_V}(V^k + sDU^k) = \mathcal{P}_V(V^k + sDU^k)$$

- $U^{k+1}$  的更新如下

$$\begin{aligned} U^{k+1} &= \text{prox}_{tf}(U^k + tGV^{k+1}) \\ &= \arg \min_U \left\{ \lambda \|U - B\|_1 + \langle V^{k+1}, DU \rangle + \frac{1}{2t} \|U - U^k\|_F^2 \right\} \end{aligned}$$

其中  $G: \mathbb{R}^{n \times n \times 2} \rightarrow \mathbb{R}^{n \times n}$  为离散的散度算子, 其满足

$$\langle V, DU \rangle = -\langle GV, U \rangle, \quad \forall U \in \mathbb{R}^{n \times n}, V \in \mathbb{R}^{n \times n \times 2}$$

*Q&A*

*Thank you!*

感谢您的聆听和反馈