对偶算法

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由朱桢源协助准备

提纲

对偶方法

次梯度法：速度慢，步长选择困难

梯度法：需要对偶函数可微

- 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（本讲）：对偶函数分裂成两项

- 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

对偶问题中的复合结构

设f，h 是闭凸函数，考虑如下形式的问题及其对偶问题

$$
\begin{aligned}
&\text{(P)} \quad \min_{x\in\mathbb{R}^n} \quad \psi(x) = f(x) + h(Ax), \\
&\text{(D)} \quad \max_{z} \quad \phi(z) = -f^*(-A^{\mathrm{T}}z) - h^*(z).
\end{aligned}
$$

为了在对偶问题上使用近似点梯度法，$\phi(z)$ 需要满足“可微函数+ 凸函数”的复合形式：

- h 或h^*的近似点算子容易计算（有闭形式或简单算法）
- f 是闭的强凸函数

我们下面证明这意味着$f^*(-A^{\mathrm{T}}z)$ 是梯度利普希茨连续函数：

$$
\|A\nabla f^*(-A^{\mathrm{T}}z_1) - A\nabla f^*(-A^{\mathrm{T}}z_2)\|_2 \le \frac{\|A\|_2^2}{\mu}\|z_1 - z_2\|_2
$$

强凸函数共轭函数的性质

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．

证明：

- 对任意的$y\in\mathbb{R}^n$，$f(x)-x^{\mathrm{T}}y$是强凸函数，因此对任意的$y\in\mathbb{R}^n$，存在唯一的$x\in\mathbf{dom}\, f$，使得$f^*(y)=x^{\mathrm{T}}y-f(x)$．根据最优性条件

$$y\in\partial f(x)\Leftrightarrow f^*(y)=x^{\mathrm{T}}y-f(x).$$

- 由于$f(x)$是闭凸函数，二次共轭为其本身，于是对同一组x,y有

$$x^{\mathrm{T}}y-f^*(y)=f(x)=f^{**}(x)=\sup_y\left\{x^{\mathrm{T}}y-f^*(y)\right\}.$$

- 这说明y也使得$x^{\mathrm{T}}y-f^*(y)$取到最大值．根据一阶最优性条件，

$$x\in\partial f^*(y).$$

- 再根据x的唯一性容易推出$\partial f^*(y)$中只含一个元素，故$f^*(y)$可微．

- 下证$f^*(y)$为梯度$\frac{1}{\mu}$-利普希茨连续的．对任意的y_1,y_2，存在唯一的$x_1,x_2\in\mathbf{dom}\,f$使得

$$y_1\in\partial f(x_1),\quad y_2\in\partial f(x_2).$$

- 根据次梯度性质以及$f(x)-\frac{\mu}{2}\|x\|^2$是凸函数，

$$\begin{aligned}f(x_2)&\geq f(x_1)+(y_1-\mu x_1)^{\mathrm{T}}(x_2-x_1),\\ f(x_1)&\geq f(x_2)+(y_2-\mu x_2)^{\mathrm{T}}(x_1-x_2),\end{aligned}$$

- 将上述两式相加得

$$(y_1-y_2)^{\mathrm{T}}(x_1-x_2)\geq\mu\|x_1-x_2\|^2.$$

- 根据x和y的关系我们有$x_1=\nabla f^*(y_1),x_2=\nabla f^*(y_2)$，代入上式可得

$$(y_1-y_2)^{\mathrm{T}}(\nabla f^*(y_1)-\nabla f^*(y_2))\geq\mu\|\nabla f^*(y_1)-\nabla f^*(y_2)\|^2.$$

这正是$\nabla f^*(y)$的余强制性，可知$\nabla f^*(y)$是$\frac{1}{\mu}$-利普希茨连续的．

对偶近似点梯度更新

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：

$$z^{k+1} = \operatorname{prox}_{th^*}\left(z^k + tA\nabla f^*\left(-A^{\mathrm{T}}z^k\right)\right)$$

对偶问题是取最大值，因此邻近算子内部应该取上升方向.
进一步引入变量$x^{k+1} = \nabla f^*(-A^{\mathrm{T}}z^k)$，迭代格式等价于

$$x^{k+1} = \arg\min_x \left\{f(x) + \left(A^{\mathrm{T}}z^k\right)^{\mathrm{T}} x\right\}, \quad z^{k+1} = \operatorname{prox}_{th^*}\left(z^k + tAx^{k+1}\right)$$

- 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

下面我们将提供另一种角度来理解对偶近似点梯度法.

Moreau 分解

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

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

- 或更一般地，

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

  其中$\lambda > 0$为任意正实数．

- Moreau 分解的结论表明：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．

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

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

- 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：

$$
\begin{aligned}
x^{k+1} &= \mathop{\mathrm{argmin}}_x \left\{f(x) + (z^k)^{\mathrm{T}}Ax\right\}, \\
y^{k+1} &= \mathrm{prox}_{t^{-1}h}\left(\frac{z^k}{t} + Ax^{k+1}\right) \\
&= \mathop{\mathrm{argmin}}_y \left\{h(y) - (z^k)^{\mathrm{T}}(y - Ax^{k+1}) + \frac{t}{2}\|Ax^{k+1} - y\|_2^2\right\}, \\
z^{k+1} &= z^k + t(Ax^{k+1} - y^{k+1}).
\end{aligned}
$$

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:

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

- 等价的交替极小格式是

$$\begin{aligned} x^{k+1} &= \arg\min_x L(x,y^k,z^k) \\ y^{k+1} &= \arg\min_y L_t(x^{k+1},y,z^k) \\ z^{k+1} &= z^k+t(Ax^{k+1}-y^{k+1}) \end{aligned}$$

- 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

提纲

正则化范数近似

假设f是强凸函数，$\|\cdot\|$是任意一种范数，考虑

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

对应原始问题我们有$h(y) = \|y - b\|$

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

其中$\|\cdot\|_*$表示$\|\cdot\|$的对偶范数. 从而对偶问题为：

$$\max_{\|z\|_* \le 1} \quad -f^*(-A^{\mathrm{T}} z) - b^{\mathrm{T}} z$$

应用对偶近似点梯度法，更新如下：

$$x^{k+1} = \operatorname*{argmin}_x \left\{ f(x) + (A^{\mathrm{T}} z^k)^{\mathrm{T}} x \right\}$$
$$z^{k+1} = \mathcal{P}_{\|z\|_* \le 1}(z^k + t(Ax^{k+1} - b))$$

考虑等价问题

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

交替极小化格式是

$$x^{k+1} = \operatorname*{argmin}_x f(x) + \|y^k\| + (z^k)^{\mathrm{T}}(Ax - b - y^k)$$

$$y^{k+1} = \operatorname*{argmin}_y f(x^{k+1}) + \|y\| + (z^k)^{\mathrm{T}}(Ax^{k+1} - b - y) + \frac{t}{2}\|Ax^{k+1} - b - y\|_2^2$$

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

例

假设f是强凸函数，考虑

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

即$h(y_1, y_2, \cdots, y_p) = \sum\limits_{i=1}^{p} \|y_i\|_2$，且

$$A = \begin{bmatrix} B_1^{\mathrm{T}} & B_1^{\mathrm{T}} & \cdots & B_p^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}.$$

根据$\|\cdot\|_2$的共轭函数定义，对偶问题形式如下：

$$\max_{\|z_i\|_2 \le 1} \quad -f^*\left(-\sum_{i=1}^{p} B_i^{\mathrm{T}} z_i\right),$$

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：

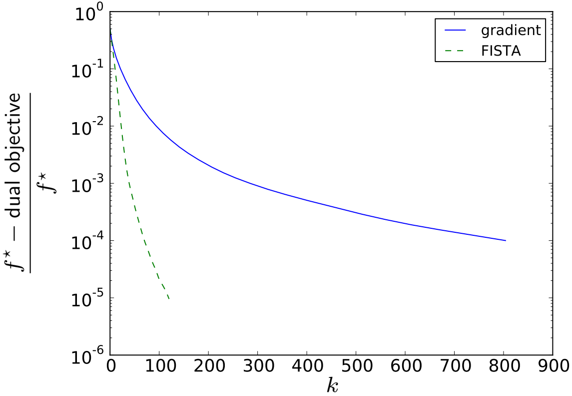
$$x^{k+1} = \operatorname*{argmin}_x \left\{ f(x) + (\sum_{i=1}^{p} B_i^{\mathrm{T}} z_i)^{\mathrm{T}} x \right\},$$

$$z_i^{k+1} = \mathcal{P}_{C_i}(z_i^k + tB_i x^{k+1}), i = 1, 2, \cdots, p.$$

数值实验

$$f(x) = \frac{1}{2}||Cx - d||_2^2$$

随机生成$C \in \mathbb{R}^{2000\times1000}$, $B_i \in \mathbb{R}^{10\times1000}$, $p = 500$



在凸集交上的极小化

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

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

我们有$h(y_1, y_2, \cdots, y_m) = \sum_{i=1}^{m} I_{C_i}(y_i)$，$A = \begin{bmatrix} I & I & \cdots & I \end{bmatrix}^{\mathrm{T}}$，对偶问题为

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

$I_{C_i}^*(z_i)$是集合C_i的支撑函数，其显式表达式不易求出．因此我们利用Moreau 分解将迭代格式写成交替极小化方法的形式：

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x \left\{ f(x) + \left(\sum_{i=1}^{m} z_i\right)^{\mathrm{T}} x \right\}, \\ y_i^{k+1} &= \mathcal{P}_{C_i}\left(\frac{z_i^k}{t} + x^{k+1}\right), \quad i = 1, 2, \cdots, m, \\ z_i^{k+1} &= z_i^k + t(x^{k+1} - y_i^{k+1}), \quad i = 1, 2, \cdots, m. \end{aligned}$$

可分问题的拆分

假设f_i是强凸函数，h_i^*有易于计算的邻近算子．考虑

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

其对偶问题形式如下：

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

对偶近似点梯度法更新如下：

$$\begin{aligned} x_j^{k+1} &= \operatorname*{argmin}_{x_j} \left\{ f_j(x_j) + (\sum_{i=1}^{m} A_{ij}z_i^k)^{\mathrm{T}} x_j \right\}, \quad j = 1, 2, \cdots, n, \\ z_i^{k+1} &= \mathrm{prox}_{th_i^*} \left(z_i + t\sum_{j=1}^{n} A_{ij}x_j^{k+1} \right), \quad i = 1, 2, \cdots, m. \end{aligned}$$

提纲

鞍点问题

令f,h是适当的闭凸函数. 考虑原始问题：

$$\min \quad f(x)+h(Ax),$$

- 由于h有自共轭性，我们将问题变形为

$$(\mathrm{L_{PD}}) \quad \min_x \ \max_z \quad \psi_{PD}(x,z) \stackrel{\text{def}}{=\!=} f(x)-h^*(z)+z^{\mathrm{T}}Ax. \qquad (1)$$

  可以看到此时问题变成了一个极小–极大问题，即关于变量x求极小，关于变量z求极大，这是一个典型的鞍点问题.

- 另一种常用的鞍点问题定义方式构造拉格朗日函数. 问题

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

  相应的鞍点问题形式如下：

$$(\mathrm{L_P}) \quad \min_{x,y} \ \max_z \quad f(x)+h(y)+z^{\mathrm{T}}(Ax-y). \qquad (2)$$

PDHG 算法

- 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(1) 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：

$$
\begin{aligned}
z^{k+1} &= \underset{z}{\operatorname{argmax}}\left\{-h^*(z) + \langle Ax^k, z - z^k\rangle - \frac{1}{2\delta_k}\|z - z^k\|_2^2\right\} \\
&= \operatorname{prox}_{\delta_k h^*}(z^k + \delta_k A x^k), \\
x^{k+1} &= \underset{x}{\operatorname{argmin}}\left\{f(x) + (z^{k+1})^{\mathrm{T}} A(x - x^k) + \frac{1}{2\alpha_k}\|x - x^k\|_2^2\right\} \\
&= \operatorname{prox}_{\alpha_k f}(x^k - \alpha_k A^{\mathrm{T}} z^{k+1}),
\end{aligned}
$$

  其中α_k, δ_k分别为原始变量和对偶变量的更新步长.
- 它在第一步固定原始变量x^k针对对偶变量做梯度上升，在第二步固定更新后的对偶变量z^{k+1}针对原始变量做梯度下降. 在这里注意，原始变量和对偶变量的更新顺序是无关紧要的，若先更新原始变量，其等价于在另一初值下先更新对偶变量.

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：

$$
\begin{aligned}
z^{k+1} &= \operatorname{prox}_{\delta_k h^*}(z^k + \delta_k A y^k), \\
x^{k+1} &= \operatorname{prox}_{\alpha_k f}(x^k - \alpha_k A^{\mathrm{T}} z^{k+1}), \\
y^{k+1} &= 2x^{k+1} - x^k.
\end{aligned}
$$

三项函数拆分

考虑

$$\min_x f_1(x) + f_2(Bx) + f_3(x),$$

其中f_1,f_2,f_3是三个下半连续的凸函数，且f_1具有Lipschitz连续常数$\frac{1}{\beta}$，$\beta \in [0,\infty)$，$B \in R^{m\times n}$。

Saddle point 问题形式：

$$\min_x \max_z \quad f_1(x) + \langle z, Bx\rangle - f_2^*(z) + f_3(x)$$

PDFP算法更新如下：

$$\begin{cases} y_{k+1} = \textbf{prox}_{\gamma f_3}(x_k - \gamma\nabla f_1(x_k) - \lambda B^{\mathrm{T}} z_k), \\ z_{k+1} = (I - \textbf{prox}_{\frac{\gamma}{\lambda} f_2})(By_{k+1} + z_k), \\ x_{k+1} = \textbf{prox}_{\gamma f_3}(x_k - \gamma\nabla f_1(x_k) - \lambda B^{\mathrm{T}} z_{k+1}). \end{cases}$$

其中$0 < \lambda < \frac{1}{\lambda_{max}(BB^{\mathrm{T}})}$，$0 < \gamma < 2\beta$。

提纲

1. 对偶近似点梯度法
2. 应用举例
3. 原始– 对偶混合梯度算法
4. 应用举例
5. 收敛性分析

LASSO问题求解

考虑LASSO问题

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

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

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

根据共轭函数的定义，

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

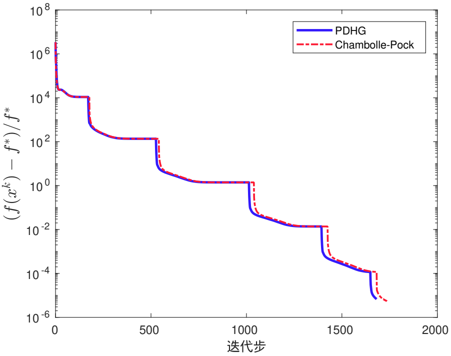
应用PDHG算法，x^{k+1} 和z^{k+1} 的更新格式分别为

$$z^{k+1}=\mathrm{prox}_{\delta_k h^*}(z^k+\delta_k Ax^k)=\frac{1}{\delta_k+1}\left(z^k+\delta_k Ax^k-\delta_k b\right),$$
$$x^{k+1}=\mathrm{prox}_{\alpha_k\mu\|\cdot\|_1}(x^k-\alpha_k A^{\mathrm{T}}z^{k+1}).$$

这里δ_k,α_k 为步长.

Chambolle-Pock算法格式为

$$
\begin{aligned}
z^{k+1} &= \frac{1}{\delta_k + 1}\left(z^k + \delta_k A y^k - \delta_k b\right), \\
x^{k+1} &= \mathrm{prox}_{\alpha_k \mu \|\cdot\|_1}(x^k - \alpha_k A^{\mathrm{T}} z^{k+1}), \\
y^{k+1} &= 2x^{k+1} - x^k.
\end{aligned}
$$



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）：

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

其中$\|U\|_{TV}$为全变差，即可以用离散的梯度（线性）算子$D:\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n\times 2}$表示为

$$\|U\|_{TV} = \sum_{1\le i,j\le n} \|(DU)_{ij}\|_2.$$

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

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

其中$w_{ij}\in\mathbb{R}^2$且$\|\cdot\|$定义了$\mathbb{R}^{n\times n\times 2}$上的一种范数．利用$\|\cdot\|$的定义，有

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

TV-L^1模型

我们取D 为相应的线性算子，并取

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

相应的鞍点问题(1) 如下：

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

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

$$\min_{U} \max_{V} \quad f(U) + \langle V, DU\rangle - I_{\mathcal{V}}(V).$$

应用PDHG算法，则V^{k+1}的更新为

$$V^{k+1} = \mathrm{prox}_{sI_{\mathcal{V}}}(V^k + sDU^k) = \mathcal{P}_{\mathcal{V}}(V^k + sDU^k), \tag{3}$$

即$V^k + sDU^k$ 在$\mathcal{V}$上的投影，而U^{k+1}的更新如下：

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

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

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

若应用Chambolle-Pock算法，那么U^{k+1}的更新保持不变，仅需调整V^{k+1}的更新为$V^k + sD(2U^{k+1} - U^k)$在$\mathcal{V}$上的投影.

图像填充模型

考虑问题

$$\min_{U\in\mathbb{R}^{n\times n}} \quad \|U\|_{TV}+\frac{\lambda}{2}\|U-B\|_F^2.$$

类似于上一个例子中的分析，我们取D 为相应的线性算子，并取

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

一般的鞍点问题叙述如下：

$$(\mathrm{L_{PD}})\quad \min_{U}\ \max_{V} \quad f(U)+\langle V,DU\rangle-I_{\mathcal{V}}(V),$$

其中$\mathcal{V}$ 与TV-L^1模型中的定义一致. 应用PDHG算法，则V^{k+1}的更新为(3) 式. 引入离散的散度算子G，U^{k+1}的更新如下：

$$\begin{aligned}U^{k+1}&=\mathrm{prox}_{tf}(U^k+tGV^{k+1})\\&=\operatorname*{argmin}_{U}\left\{\frac{\lambda}{2}\|U-B\|_F^2+\left\langle V^{k+1},DU\right\rangle+\frac{1}{2t}\|U-U^k\|_F^2\right\}.\end{aligned}$$

同样地，Chambolle-Pock算法的更新表达式也可类似地推出.

图像反卷积模型

考虑问题

$$\min_{U\in\mathbb{R}^{n\times n}} \quad \|U\|_{TV} + \frac{\lambda}{2}\|\mathcal{A}U - B\|_F^2,$$

其中$\mathcal{A}U = K_{\mathcal{A}} * U$为卷积算子，且$K_{\mathcal{A}}$是$\mathcal{A}$的卷积核对应的矩阵.
类似于TV-L^1模型中的分析，取D为相应的线性算子，并取

$$f(U) = \frac{\lambda}{2}\|\mathcal{A}U - B\|_F^2,\ U \in \mathbb{R}^{n\times n},\ h(W) = \|W\|,\ W \in \mathbb{R}^{n\times n\times 2}.$$

类似地，一般的鞍点问题叙述如下：

$$(\mathrm{L_{PD}}) \quad \min_U \ \max_V \qquad f(U) + \langle V, DU\rangle - I_{\mathcal{V}}(V),$$

其中$\mathcal{V}$ 与TV-L^1模型中的定义一致.

应用PDHG算法，则V^{k+1}的更新仍为(3) 式，而U^{k+1}的更新为：

$$
\begin{aligned}
U^{k+1} &= \text{prox}_{tf}(U^k + tGV^{k+1}) \\
&= \underset{U}{\text{argmin}} \left\{ \frac{\lambda}{2} \|\mathcal{A}U - B\|_F^2 + \frac{1}{2t} \|U - (U^k + tGV^{k+1})\|_F^2 \right\},
\end{aligned}
$$

其中G 为离散的散度算子. 可知U^{k+1}满足如下方程：

$$
\lambda\mathcal{A}^*(\mathcal{A}U^{k+1} - B) + \frac{1}{t}(U^{k+1} - (U^k + tGV^{k+1})) = 0,
$$

其中$\mathcal{A}^*$是$\mathcal{A}$的共轭算子，且其卷积核对应的矩阵为$K_{\mathcal{A}^*}$. 由于$\mathcal{A}U = K_{\mathcal{A}} * U$ 具有卷积的形式，我们可以利用快速傅里叶变换$\mathcal{F}$ 和其逆变换$\mathcal{F}^{-1}$ 来快速求解上面的线性方程组.

根据

$$\mathcal{F}(\mathcal{A}U) = \mathcal{F}(K_{\mathcal{A}} * U) = \mathcal{F}(K_{\mathcal{A}}) \odot \mathcal{F}(U),$$

其中$\odot$表示逐分量相乘, 我们有

$$\mathcal{F}(K_{\mathcal{A}^*}) \odot \Big(\mathcal{F}(K_{\mathcal{A}}) \odot \mathcal{F}(U^{k+1}) - \mathcal{F}(B)\Big) + \frac{1}{t\lambda}\mathcal{F}(U^{k+1} - (U^k + tGV^{k+1})) = 0.$$

利用关系式$\mathcal{F}(K_{\mathcal{A}^*}) = \overline{\mathcal{F}(K_{\mathcal{A}})}$，可得$U^{k+1}$的显式表达式

$$U^{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(U^k + tGV^{k+1}) + t\lambda\mathcal{F}(B) \odot \overline{\mathcal{F}(K_{\mathcal{A}})}}{1 + t\lambda|\mathcal{F}(K_{\mathcal{A}})|^2}\right),$$

以上表达式中除$\mathcal{F}, \mathcal{F}^{-1}, G$外，其余均为逐分量的运算

- Fused Lasso:

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \mu_1\|Bx\|_1 + \mu_2\|x\|_1$$

即$f_1(x) = \frac{1}{2}\|Ax-b\|^2$，$f_2 = \mu_1\|\cdot\|_1$，$f_3 = \mu_2\|\cdot\|_1$。

- 图像恢复：

$$\min_{x\in C} \frac{1}{2}\|Ax-b\|^2 + \mu\|Dx\|_1$$

即$f_1(x) = \frac{1}{2}\|Ax-b\|^2$，$f_2 = \mu\|\cdot\|_1$，$f_3 = 1_C(\cdot)$. 在医学核共振图像重建问题中，$A = (A_1^{\mathrm{T}}, \dots, A_N^{\mathrm{T}})$，其中$A_j$由一个对角下采样算子$D$，傅里叶变换$F$，对角的圈灵敏度映射$S_j$构成，即$A_j = DFS_j$，通常$S_j$是事先估计好的。

提纲

- 设X, Z分别为变量x, z的取值空间，若点$(\hat{x}, \hat{z})$满足

$$\psi_{\mathrm{PD}}(x, \hat{z}) \geq \psi_{\mathrm{PD}}(\hat{x}, \hat{z}) \geq \psi_{\mathrm{PD}}(\hat{x}, z), \quad \forall\, x \in X, z \in Z,$$

称$(\hat{x}, \hat{z})$是问题(1)的一个**鞍点**，其中ψ_{PD}的定义见该问题.

- 对任意子集$B_1 \times B_2 \subset X \times Z$，定义**部分原始–对偶间隙**为

$$\mathcal{G}_{B_1 \times B_2}(x, z) = \max_{z' \in B_2} \psi_{\mathrm{PD}}(x, z') - \min_{x' \in B_1} \psi_{\mathrm{PD}}(x', z).$$

不难验证，只要鞍点$(\hat{x}, \hat{z}) \in B_1 \times B_2$，就有

$$\begin{aligned}\mathcal{G}_{B_1 \times B_2}(x, z) &\geq \psi_{\mathrm{PD}}(x, \hat{z}) - \psi_{\mathrm{PD}}(\hat{x}, z) \\ &= (\psi_{\mathrm{PD}}(x, \hat{z}) - \psi_{\mathrm{PD}}(\hat{x}, \hat{z})) + (\psi_{\mathrm{PD}}(\hat{x}, \hat{z}) - \psi_{\mathrm{PD}}(\hat{x}, z)) \geq 0,\end{aligned}$$

并且在鞍点处$\mathcal{G}_{B_1 \times B_2}(\hat{x}, \hat{z}) = 0$. 此外，容易验证当点$(\hat{x}, \hat{z}) \in \mathrm{int}(B_1 \times B_2)$且满足$\mathcal{G}_{B_1 \times B_2}(\hat{x}, \hat{z}) = 0$ 时，$(\hat{x}, \hat{z})$是一个鞍点.

设f, h为闭凸函数，原问题存在鞍点$(\hat{x}, \hat{z})$．在Chambolle-Pock迭代格式中取步长$\alpha_k = t, \delta_k = s$，且满足$st < \frac{1}{L}$（$L = \|A\|_2^2$），则序列$\{(x^k, z^k)\}$具有：

(a) 令常数$C \le (1 - Lst)^{-1}$. $\forall k$，(x^k, z^k)有界，且满足

$$\frac{\|x^k - \hat{x}\|^2}{2t} + \frac{\|z^k - \hat{z}\|^2}{2s} \le C\left(\frac{\|x^0 - \hat{x}\|^2}{2t} + \frac{\|z^0 - \hat{z}\|^2}{2s}\right),$$

(b) 记$\bar{x}_N = \frac{1}{N}\sum_{k=1}^{N} x^k$，$\bar{z}_N = \frac{1}{N}\sum_{k=1}^{N} z^k$，则对$B_1 \times B_2 \subset X \times Z$，有

$$\mathcal{G}_{B_1 \times B_2}(\bar{x}_N, \bar{z}_N) \le \frac{D(B_1, B_2)}{N}, \tag{4}$$

其中$D(B_1, B_2) = \sup_{(x,z)\in B_1 \times B_2}\left\{\frac{\|x-x^0\|^2}{2t} + \frac{\|z-z^0\|^2}{2s}\right\}$;
进一步地，序列$\{(\bar{x}_N, \bar{z}_N)\}_{N=1}^{\infty}$的聚点为问题(1)的一个鞍点；

(c) 存在问题(1)一个鞍点(x^*, z^*)使得$x^k \to x^*$, $z^k \to z^*$．

为了方便推导，首先考虑算法的一般格式：

$$\begin{aligned} z^{k+1} &= \mathrm{prox}_{sh^*}(z^k + sA\bar{x}), \\ x^{k+1} &= \mathrm{prox}_{tf}(x^k - tA^{\mathrm{T}}\bar{z}). \end{aligned}$$

这里和Chambolle-Pock算法不同的是，我们使用$\bar{x},\bar{z}$来表示更新x,z时的参考点．当它们取特定值时，以上格式可以为PDHG 算法或Chambolle-Pock 算法．根据邻近算子的性质，

$$\begin{aligned} -A^{\mathrm{T}}\bar{z} + \frac{x^k - x^{k+1}}{t} &\in \partial f(x^{k+1}), \\ A\bar{x} + \frac{z^k - z^{k+1}}{s} &\in \partial h^*(z^{k+1}). \end{aligned}$$

根据次梯度的定义，对于任意的$(x,z) \in X \times Z$有

$$\begin{aligned} f(x) &\geq f(x^{k+1}) + \frac{1}{t}(x - x^{k+1})^{\mathrm{T}}(x^k - x^{k+1}) - (x - x^{k+1})^{\mathrm{T}}A^{\mathrm{T}}\bar{z}, \\ h^*(z) &\geq h^*(z^{k+1}) + \frac{1}{s}(z - z^{k+1})^{\mathrm{T}}(z^k - z^{k+1}) + (z - z^{k+1})^{\mathrm{T}}A\bar{x}. \end{aligned}$$

收敛性分析

将上述两个不等式相加，并引入二次项可整理得到

$$\begin{aligned}
&\frac{\|x-x^k\|^2}{2t}+\frac{\|z-z^k\|^2}{2s}-\frac{\|x-x^{k+1}\|^2}{2t}-\frac{\|z-z^{k+1}\|^2}{2s} \\
\geq&\left[f(x^{k+1})-h^*(z)+(x^{k+1})^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(z^{k+1})+x^{\mathrm{T}}A^{\mathrm{T}}z^{k+1}\right] \\
&+\frac{\|x^k-x^{k+1}\|^2}{2t}+\frac{\|z^k-z^{k+1}\|^2}{2s} \\
&+(x^{k+1}-\bar{x})^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z)-(x^{k+1}-x)^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-\bar{z}).
\end{aligned} \tag{5}$$

将Chambolle-Pock格式代入(5)，即取$\bar{x}=2x^k-x^{k-1}, \bar{z}=z^{k+1}$，那么

$$\begin{aligned}
&(x^{k+1}-\bar{x})^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z)-(x^{k+1}-x)^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-\bar{z}) \\
=&(x^{k+1}-x^k-(x^k-x^{k-1}))^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z) \\
=&(x^{k+1}-x^k)^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z)-(x^k-x^{k-1})^{\mathrm{T}}A^{\mathrm{T}}(z^k-z) \\
&-(x^k-x^{k-1})^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z^k) \\
\geq&(x^{k+1}-x^k)^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1}-z)-(x^k-x^{k-1})^{\mathrm{T}}A^{\mathrm{T}}(z^k-z) \\
&-\sqrt{L}\|x^k-x^{k-1}\|\|z^{k+1}-z^k\|,
\end{aligned} \tag{6}$$

应用柯西不等式即得到最后的不等号

又利用$2ab \le \alpha a^2 + \frac{b^2}{\alpha}$对任意的$\alpha > 0$均成立，有

$$\begin{aligned} &\sqrt{L}\|x^k - x^{k-1}\|\|z^{k+1} - z^k\| \\ \le &\frac{\sqrt{L}\alpha t}{2t}\|x^k - x^{k-1}\|^2 + \frac{\sqrt{L}s}{2\alpha s}\|z^{k+1} - z^k\|^2, \end{aligned}$$

取$\alpha = \sqrt{\frac{s}{t}}$，则

$$\sqrt{L}\alpha t = \sqrt{L}\frac{s}{\alpha} = \sqrt{Lst} < 1,$$

从而合并(5) 式和(6) 式得到，对于任意的$(x, z) \in X \times Z$，

$$\begin{aligned} &\frac{\|x - x^k\|^2}{2t} + \frac{\|z - z^k\|^2}{2s} - \frac{\|x - x^{k+1}\|^2}{2t} - \frac{\|z - z^{k+1}\|^2}{2s} \\ \ge &\left[f(x^{k+1}) - h^*(z) + (x^{k+1})^{\mathrm{T}}A^{\mathrm{T}}z\right] - \left[f(x) - h^*(z^{k+1}) + x^{\mathrm{T}}A^{\mathrm{T}}z^{k+1}\right] \\ &+ (1 - \sqrt{Lst})\frac{\|z^k - z^{k+1}\|^2}{2s} + \frac{\|x^k - x^{k+1}\|^2}{2t} - \sqrt{Lst}\frac{\|x^{k-1} - x^k\|^2}{2t} \\ &+ (x^{k+1} - x^k)^{\mathrm{T}}A^{\mathrm{T}}(z^{k+1} - z) - (x^k - x^{k-1})^{\mathrm{T}}A^{\mathrm{T}}(z^k - z). \end{aligned} \tag{7}$$

将上述不等式中的k从0遍历至$N-1$并求和，消掉不等式两边共同项后有

$$\begin{aligned}
&\sum_{k=1}^{N}\left\{\left[f(x^k)-h^*(z)+(x^k)^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(z^k)+x^{\mathrm{T}}A^{\mathrm{T}}z^k\right]\right\}\\
&+\frac{\|x-x^N\|^2}{2t}+\frac{\|z-z^N\|^2}{2s}+(1-\sqrt{Lst})\sum_{k=1}^{N}\frac{\|z^k-z^{k-1}\|^2}{2s}\\
&+(1-\sqrt{Lst})\sum_{k=1}^{N-1}\frac{\|x^k-x^{k-1}\|^2}{2t}+\frac{\|x^N-x^{N-1}\|^2}{2t}\\
\le&\frac{\|x-x^0\|^2}{2t}+\frac{\|z-z^0\|^2}{2s}+(x^N-x^{N-1})^{\mathrm{T}}A^{\mathrm{T}}(z^N-z),
\end{aligned} \tag{8}$$

其中约定$x^{-1}=x^0$. 再一次应用柯西不等式，以及$2ab\le\alpha a^2+\frac{b^2}{\alpha}$对任意的$\alpha>0$均成立，可以得到

$$\begin{aligned}
(x^N-x^{N-1})^{\mathrm{T}}A^{\mathrm{T}}(z^N-z)&\le\|x^N-x^{N-1}\|(\sqrt{L}\|z^N-z\|)\\
&\le\frac{\|x^N-x^{N-1}\|^2}{2t}+\frac{Lst\|z-z^N\|^2}{2s}.
\end{aligned}$$

收敛性分析

不等式(8)可进一步整理为

$$\begin{aligned}
&\sum_{k=1}^{N}\left\{\left[f(x^k)-h^*(z)+(x^k)^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(z^k)+x^{\mathrm{T}}A^{\mathrm{T}}z^k\right]\right\}\\
&+\frac{\|x-x^N\|^2}{2t}+(1-Lst)\frac{\|z-z^N\|^2}{2s}+(1-\sqrt{Lst})\sum_{k=1}^{N}\frac{\|z^k-z^{k-1}\|^2}{2s}\\
&+(1-\sqrt{Lst})\sum_{k=1}^{N-1}\frac{\|x^k-x^{k-1}\|^2}{2t}\\
\le&\frac{\|x-x^0\|^2}{2t}+\frac{\|z-z^0\|^2}{2s}.
\end{aligned} \tag{9}$$

若取$(x,z)=(\hat{x},\hat{z})$，则由鞍点性质可知

$$\left[f(x^k)-h^*(\hat{z})+(x^k)^{\mathrm{T}}A^{\mathrm{T}}\hat{z}\right]-\left[f(\hat{x})-h^*(z^k)+\hat{x}^{\mathrm{T}}A^{\mathrm{T}}z^k\right]\ge 0.$$

进而(9)左边每一项都是正的，结论(a)成立．

从(9)出发，利用f,h^*的凸性，以及$\bar{x}_N,\bar{z}_N$的定义，有

$$\begin{aligned}&\left[f(\bar{x}_N)-h^*(z)+(\bar{x}_N)^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(\bar{z}_N)+x^{\mathrm{T}}A^{\mathrm{T}}\bar{z}_N\right]\\ \le&\frac{1}{N}\sum_{k=1}^{N}\left\{\left[f(x^k)-h^*(z)+(x^k)^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(z^k)+x^{\mathrm{T}}A^{\mathrm{T}}z^k\right]\right\}\\ \le&\frac{1}{N}\left(\frac{\|x-x^0\|^2}{2t}+\frac{\|z-z^0\|^2}{2s}\right).\end{aligned} \tag{10}$$

从而结论(b)中(4)式成立．由(1)知$\{(x^k,z^k)\}$是有界序列，因此其均值列$\{(\bar{x}_N,\bar{z}_N)\}$也为有界序列．记$(x^\sharp,z^\sharp)$为序列$\{(\bar{x}_N,\bar{z}_N)\}$的聚点，利用$f,h^*$的凸性以及闭性（下半连续性），对(10)式左右同时取下极限，可知对任意的$(x,z)\in X\times Z$，

$$\left[f(x^\sharp)-h^*(z)+(x^\sharp)^{\mathrm{T}}A^{\mathrm{T}}z\right]-\left[f(x)-h^*(z^\sharp)+x^{\mathrm{T}}A^{\mathrm{T}}z^\sharp\right]\le 0.$$

从而$(x^\sharp,z^\sharp)$也是问题(1)的一个鞍点．

收敛性分析

为了证明$\{(x^k,z^k)\}$全序列收敛到问题(1)的鞍点，我们采用的大致思路为：先说明其子列收敛，然后再利用(7) 式估计序列中其他点到子列极限点的误差（进而证明全序列收敛），最后说明该极限点是鞍点．根据结论(1)，$\{(x^k,z^k)\}$是有界点列，因此存在子列$\{(x^{k_l},z^{k_l})\}$收敛于(x^*,z^*)．在(7) 式中令$(x,z)=(x^*,z^*)$，并将k从k_l取至$N-1, N>k_l$并求和，有

$$
\begin{aligned}
&\frac{\|x^*-x^N\|^2}{2t}+\frac{\|z^*-z^N\|^2}{2s}\\
&+(1-\sqrt{Lst})\sum_{k=k_l+1}^{N}\frac{\|z^k-z^{k-1}\|^2}{2s}-\frac{\|x^{k_l}-x^{k_l-1}\|^2}{2t}\\
&+(1-\sqrt{Lst})\sum_{k=k_l}^{N-1}\frac{\|x^k-x^{k-1}\|^2}{2t}+\frac{\|x^N-x^{N-1}\|^2}{2t}\\
&+(x^N-x^{N-1})^{\mathrm{T}}A^{\mathrm{T}}(z^N-z^*)-(x^{k_l}-x^{k_l-1})^{\mathrm{T}}A^{\mathrm{T}}(z^{k_l}-z^*)\\
\le&\frac{\|x^*-x^{k_l}\|^2}{2t}+\frac{\|z^*-z^{k_l}\|^2}{2s}.
\end{aligned}
$$

去掉上式中不等式左边的求和项（正项），我们有如下估计：

$$\begin{aligned}&\frac{\|x^*-x^N\|^2}{2t}+\frac{\|z^*-z^N\|^2}{2s}\\ \le&\frac{\|x^*-x^{k_l}\|^2}{2t}+\frac{\|z^*-z^{k_l}\|^2}{2s}+\frac{\|x^{k_l}-x^{k_l-1}\|^2}{2t}-\frac{\|x^N-x^{N-1}\|^2}{2t}\\ &+(x^{k_l}-x^{k_l-1})^{\mathrm{T}}A^{\mathrm{T}}(z^{k_l}-z^*)-(x^N-x^{N-1})^{\mathrm{T}}A^{\mathrm{T}}(z^N-z^*).\end{aligned}$$

注意到

$$\begin{aligned}x^{k_l}\to x^*,&\quad（x^{k_l}\text{ 的定义}）\\ x^N-x^{N-1}\to 0,&\quad（\text{由(9) 式推出}）\\ \{z^k\}\text{有界},&\quad（\text{本定理中(a) 的结论}）\end{aligned}$$

所以当$N\to\infty$时有，$x^N\to x^*$, $z^N\to z^*$，全序列收敛性得证．最后，由全序列收敛可知均值$(\bar{x}_N,\bar{z}_N)$也收敛到(x^*,z^*)，根据(a) 的结论和极限的唯一性立即得到$(x^\sharp,z^\sharp)=(x^*,z^*)$，即收敛到问题(1)的一个鞍点