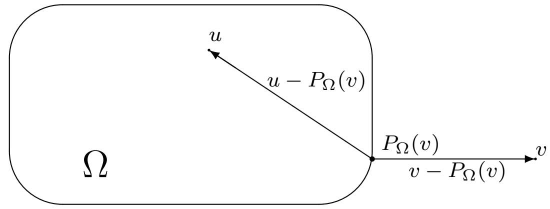
A-从变分不等式的投影收缩算法到凸优化的分裂收缩算法

2013年前总结的主要研究工作 maths.nju.edu.cn/~hebma 何炳生

1 变分不等式的投影收缩算法 (变分不等式应用与算法建议参阅本人主页系列讲义前三讲)

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, F 是 $\Re^n \rightarrow \Re^n$ 的一个映射. 考虑求解单调变分不等式

$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{1.1}$$



当 $\Omega = \Re^n$, 问题(1.1)就退化成非线性方程组 $F(u)=0$. 我们说一个变分不等式单调, 是指其中的算子 F 满足 $(u-v)^T(F(u)-F(v)) \ge 0$. 在求解变分不等式 (1.1) 的投影收缩算法中, 对给定的当前点 u^k 和常数 $\beta_k > 0$, 利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{1.2}$$

生成一个预测点 $\tilde{u}^k$. 同时假设用投影 (1.2) 产生 $\tilde{u}^k$ 时选取的 β_k 能使得预测点 $\tilde{u}^k$ 满足

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \le \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{1.3}$$

假如 $\tilde{u}^k = u^k$, u^k 就是 (1.1) 的解(该结论证明见系列讲义第二讲). 否则, 就基于以下的分析进行新的迭代. 首先, 凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 用 $P_\Omega(v)$ 表示 v 到 Ω 上的投影, 有

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{1.4}$$

在 (1.4) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (1.2) 得到 $\tilde{u}^k = P_\Omega(v)$, 并从 (1.4) 得到

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T\{\tilde{u}^k - [u^k - \beta_k F(u^k)]\} \ge 0, \quad \forall u \in \Omega. \tag{1.5}$$

在上面的 (1.5) 式两边都加上 $(u-\tilde{u}^k)^T d(u^k,\tilde{u}^k)$, 其中

$$d(u^k,\tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)], \tag{1.6}$$

从而由投影 (1.2) 得到我们需要的预测公式

$$\text{[预测]} \qquad \tilde{u}^k \in \Omega, \ (u-\tilde{u}^k)^T \beta_k F(\tilde{u}^k) \ge (u-\tilde{u}^k)^T d(u^k,\tilde{u}^k), \ \forall u \in \Omega. \tag{1.7}$$

将 (1.7) 中的 $u \in \Omega$ 选成 u^*, 就有 $(\tilde{u}^k - u^*)^T d(u^k,\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k)$. 再由单调性和VI 定义, 得到

$$(\tilde{u}^k - u^*)^T d(u^k,\tilde{u}^k) \ge 0. \qquad \text{因此} \qquad (u^k - u^*)^T d(u^k,\tilde{u}^k) \ge (u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k). \tag{1.8}$$

由 $d(u^k,\tilde{u}^k)$ 的表达式和假设 (1.3), 利用 Cauchy-Schwarz 不等式, 推得

$$(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \ge (1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{1.9}$$

当 $u^k \neq \tilde{u}^k$ 时, (1.8) 说明 $d(u^k,\tilde{u}^k)$ 是距离函数 $\frac{1}{2}\|u-u^*\|^2$ 在 u^k 处的一个上升方向. 可以用

$$\text{[单位步长的校正]} \qquad u^{k+1} = u^k - d(u^k,\tilde{u}^k) \qquad \text{生成新的迭代点.} \tag{1.10}$$

由 (1.6),
$$2(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) - \|d(u^k,\tilde{u}^k)\|^2 = \|u^k-\tilde{u}^k\|^2 - \beta_k^2\|F(u^k)-F(\tilde{u}^k)\|^2. \tag{1.11}$$

因此,
$$\begin{aligned} \|u^k-u^*\|^2 - \|u^{k+1}-u^*\|^2 &\overset{(1.10)}{=} 2(u^k-u^*)^T d(u^k,\tilde{u}^k) - \|d(u^k,\tilde{u}^k)\|^2 \\ &\overset{(1.8)}{\ge} 2(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) - \|d(u^k,\tilde{u}^k)\|^2 \overset{(1.11)\&(1.3)}{\ge} (1-\nu^2)\|u^k-\tilde{u}^k\|^2. \end{aligned} \tag{1.12}$$

$$\text{[计算步长的校正]} \qquad u^{k+1} = u^k - \alpha_k^* d(u^k,\tilde{u}^k) \qquad \text{产生离 } u^* \text{ 更近的迭代点.} \tag{1.13}$$

$$\text{其中步长} \quad \alpha_k^* \quad \text{由} \quad \alpha_k^* = (u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k)/\|d(u^k,\tilde{u}^k)\|^2 \quad \text{给出.} \tag{1.14}$$

注意到 (1.12) 中的最后一个不等式隐含了 $2(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) > \|d(u^k,\tilde{u}^k)\|^2$, 因而 $\alpha_k^* > \frac{1}{2}$. 接着就有

$$\begin{aligned} \|u^k-u^*\|^2 - \|u^{k+1}-u^*\|^2 &= \|u^k-u^*\|^2 - \|u^k-u^*-\alpha_k^* d(u^k,\tilde{u}^k)\|^2 \quad \text{(利用了 (1.13))} \\ &= 2\alpha_k^*(u^k-u^*)^T d(u^k,\tilde{u}^k) - (\alpha_k^*)^2\|d(u^k,\tilde{u}^k)\|^2 \qquad \text{(利用 (1.8) 和 (1.14))} \\ &\ge \alpha_k^*(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \ge \tfrac{1}{2}(1-\nu)\|u^k-\tilde{u}^k\|^2. \quad \text{(利用了 } \alpha_k^* > \tfrac{1}{2} \text{ 和 (1.9))} \end{aligned} \tag{1.15}$$

收缩不等式 (1.12) 和 (1.15) 是证明算法收敛的关键式子. 分处不等式 (1.7) 两端的 $\beta_k F(\tilde{u}^k)$ 和 $d(u^k,\tilde{u}^k)$ 称为一对孪生方向. 将在 **B** 中证明的美妙之处是: 采用方向 $\beta_k F(\tilde{u}^k)$ 和 (1.14) 中给出的步长 α_k^*, 用

[投影校正] $u^{k+1} = P_\Omega[u^k - \alpha_k^*\beta_k F(\tilde{u}^k)]$ 与方法 (1.13) 具有同样的收敛性质 (1.15), 详见 **B**.

✠ 孪生方向, 相同步长. 姊妹方法, 投影校正效果更好! 真所谓 兄弟齐心, 其利断金! ✠

2 线性约束凸优化分裂收缩算法的统一框架 (参见何炳生 2015, 2018 运筹学报文章)

我们在变分不等式框架下讨论线性约束凸优化的分裂收缩算法, 以三块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{2.1}$$

为例做介绍. 问题 (2.1) 的拉格朗日函数是 $L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)$.

拉格朗日函数的鞍点 (x^*,y^*,z^*,λ^*) （其中 (x,y,z) 为原始变量, λ 为对偶变量）满足

$$L_{\lambda\in\Re^m}(x^*,y^*,z^*,\lambda)\le L(x^*,y^*,z^*,\lambda^*)\le L_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}}(x,y,z,\lambda^*).$$

利用鞍点 $w^*=(x^*,y^*,z^*,\lambda^*)$ 的极大极小性质 (推导可参见 **C** 或 **D**), 它可以表述为如下的变分不等式:

$$\text{(VI)} \qquad w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{2.2}$$

其中, $\Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m$, $\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)$, 并且

$$u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}. \tag{2.3}$$

注意到, (2.3) 中的 $F(w)$ 恰有 $(w-\tilde w)^T(F(w)-F(\tilde w))\equiv 0$.

加之 $\theta(u)$ 是凸函数(不一定可微), 我们称 (2.2) 为单调(混合)变分不等式, 简称单调变分不等式.

我们首先介绍与投影收缩算法类似的求解单调变分不等式问题 (2.2) 的预测-校正收缩算法.

[预测] 对给定的 v^k(在两块可分离问题的ADMM中 $v=(y,\lambda)$, 有些算法中 $v=w$), 求得点 $\tilde w^k$, 使得

$$\tilde w^k\in\Omega,\ \ \theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge(v-\tilde v^k)^TQ(v^k-\tilde v^k),\ \ \forall w\in\Omega, \tag{2.4}$$

如果矩阵 Q^T+Q 正定(Q 不一定对称), 我们称 (2.4) 为合格的预测. 例见短篇 **F** 的 (4.2)

利用预测 (2.4), 可以建立求解单调变分不等式 (2.2) 的预测-校正的统一框架. (ADMM为其一个特例).

[单位步长的校正] 采用单位步长, 新迭代点的校正公式为 例见短篇 **F** 的 (4.3)

$$v^{k+1}=v^k-M(v^k-\tilde v^k). \tag{2.5}$$

由合格预测 (2.4) 和单位步长的校正 (2.5), 组成一个预测-校正算法框架, 其收敛性条件是:

[收敛性条件] 要求存在正定矩阵 H, 能够使得 例见短篇 **F** 的 (4.4)-(4.6)

$$HM=Q \qquad \text{并且} \qquad G=Q^T+Q-M^THM\succ 0.\quad (\text{正定}) \tag{2.6}$$

将 (2.4) 中任意的 $w\in\Omega$ 选成某个解点 w^*, 利用$(\tilde w-w^*)^TF(\tilde w)=(\tilde w-w^*)^TF(w^*)$ 和 w^* 是解这一事实,

得到 $(\tilde v^k-v^*)^TQ(v^k-\tilde v^k)\ge 0.$ 进而有 $(v^k-v^*)^TQ(v^k-\tilde v^k)\ge(v^k-\tilde v^k)^TQ(v^k-\tilde v^k).$ (2.7)

$$\begin{aligned}\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2 &= \|v^k-v^*\|_H^2-\|(v^k-v^*)-M(v^k-\tilde v^k)\|_H^2 \quad (\text{利用 } HM=Q)\\ &\ge (v^k-\tilde v^k)^T(Q^T+Q)(v^k-\tilde v^k)-\|M(v^k-\tilde v^k)\|_H^2 \overset{(2.6)}{=} \|v^k-\tilde v^k\|_G^2.\end{aligned} \tag{2.8}$$

不等式 (2.8) 说明统一框架算法在条件 (2.6) 满足的情况下, 迭代方法生成的序列 $\{v^k\}$ 具有收缩性质

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde v^k\|_G^2. \qquad (\text{这是证明方法收敛性的关键不等式}) \tag{2.9}$$

结论: 由预测 (2.4) 和校正 (2.5) 构成的求解 VI 问题 (2.2) 的算法统一框架, 条件 (2.6)满足时方法就收敛.

根据 (2.4), 也可以如同 §1 中考虑 H-模下计算步长的校正, 这里的 H 为任意的正定矩阵 (也可以是单位阵). 由 (2.7) 得知, 只要 $(v^k-\tilde v^k)^TQ(v^k-\tilde v^k)>0$ (请注意这是一个比 $Q^T+Q\succ 0$ 宽松的条件), 向量 $H^{-1}Q(v^k-\tilde v^k)$ 是距离函数 $\frac12\|v-v^*\|_H^2$ 在 v^k 处的一个上升方向. 如同 §1 的投影收缩算法中有了 (1.8), 可以用 (1.13)-(1.14) 进行计算步长的校正. 此时我们可以在 H 模下利用 (2.7) 实行计算步长的校正:

[计算步长的校正] 对选定的正定矩阵 H, 由以下法则生成新的迭代点:

$$v^{k+1}=v^k-\alpha_k^*M(v^k-\tilde v^k), \qquad \text{其中}\quad M=H^{-1}Q, \tag{2.10a}$$

$$\text{步长}\ \ \alpha_k^*\ \ \text{由}\ \ \alpha_k^*=(v^k-\tilde v^k)^TQ(v^k-\tilde v^k)/\|M(v^k-\tilde v^k)\|_H^2\ \ \text{给出}. \tag{2.10b}$$

$$\begin{aligned}&\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2 \overset{(2.10a)}{=} \|v^k-v^*\|_H^2-\|(v^k-v^*)-\alpha_k^*M(v^k-\tilde v^k)\|_H^2 \quad (\text{利用 } HM=Q)\\ &\quad\ge \alpha_k^*(v^k-\tilde v^k)^T(Q^T+Q)(v^k-\tilde v^k)-\alpha_k^*\big(\alpha_k^*\|M(v^k-\tilde v^k)\|_H^2\big) \overset{(2.10b)}{=} \tfrac12\alpha_k^*\|v^k-\tilde v^k\|_{(Q^T+Q)}^2.\end{aligned} \tag{2.11}$$

✠ 从 变分不等式的投影收缩算法 到 凸优化的分裂收缩算法, 一条主线, 一个模式！ ✠