从优化方法的基本原理到凸优化的分裂收缩算法

何炳生

南京大学数学系

在由清华大学丘成桐数学中心组织于2024 年2 月15 日至2 月19 日在三亚清华数学论坛举办的“数学、图像科学与人工智能（MISAI）研讨会”上, 我按组织者的要求, 做了两个数值优化方面的报告. 这里将两个报告的内容做一个简要的说明.

众所周知, 如果 x^* 无约束优化问题

$$\min_{x\in\Re^n} f(x)$$

的最优点, 那必定有 $\nabla f(x^*)=0$. 如果 x^* 是约束优化问题

$$\min_{x\in\Re^n}\{f(x)\,|\,x\in\mathcal{X}\}$$

的最优解, 那么在 x^* 这一点的可行方向和下降方向的交集一定是空集. 其数学表达式就是

$$x^*\in\mathcal{X},\quad (x-x^*)^T\nabla f(x^*)\ge 0,\quad \forall x\in\mathcal{X}.$$

用变分不等式的观点研究约束凸优化问题, 相当于利用了导数的信息, 更加清晰. 线性约束凸优化问题

$$\min\{\theta(u)\,|\,\mathcal{A}u=b,\ u\in\mathcal{U}\}$$

的拉格朗日函数

$$L(u,\lambda)=\theta(u)-\lambda^T(\mathcal{A}u-b)$$

鞍点等价形式是一个混合变分不等式：

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall w\in\Omega.$$

其中

$$w=\begin{pmatrix} u\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -\mathcal{A}^T\lambda\\ \mathcal{A}u-b\end{pmatrix}\quad \text{和}\quad \Omega=\mathcal{U}\times\Re^m.$$

（1）从邻近点算法到均困的ALM和ADMM方法.

邻近点算法 (PPA) 是步步为营稳扎稳打的求解策略. 变分不等式的邻近点算法是对给定的正定矩阵 $H\succ 0$, 第 k 步迭代从给定的 w^k 出发, 求得 w^{k+1}, 使得

$$w^{k+1}\in\Omega,\ \ \theta(u)-\theta(u^{k+1})+(w-w^{k+1})^T\{F(w^{k+1})+H(w^{k+1}-w^k)\}\ge 0,\ \ \forall\, w\in\Omega,$$

一个具有可分离结构的凸优化问题相应的变分不等式, 用邻近点算法求解, 它对应的变分不等式可以分解成一些“微型的”变分不等式, 最终通过求解一些简单的凸优化问题去实现, 这是我们的基本逻辑. 这种按需构建的邻近点算法被国际同行称为A Very Simple yet Powerful Technique for Analyzing Optimization Methods. 这个报告将阐述如何灵活利用这个技术, 构造适合求解不同工程问题需要的简单有效的算法.

（2）从算法的统一框架到广义的 PPA 算法

自凸优化分裂收缩算法的预测-校正统一框架提出 10 多年来, 主要被用来便捷地证明一些算法的收敛性, 偶尔也用框架来凑成一些算法. 报告将介绍我们如何从以前的好不容易凑出一个方法到如今并不费劲构造一簇算法. ADMM 算法能在工程界得到广泛应用, 是因为它具备 PPA 算法的两条漂亮性质. 从统一框架出发, 我们提出预测-校正的广义邻近点算法, 同样具备这两条性质. 对分布式凸优化问题构造广义邻近点算法, 完成一次预测-校正迭代的基本过程就像高斯消去法求解线性方程组的消元和回代。

利用变分不等式和邻近点算法这些概念求解线性约束凸优化问题, 路正在越走越宽, 方法也愈发简单和容易被用户理解.