大规模无约束优化问题: 算法, 理论与应用

赵　亮

摘　要

最优化是一门应用广泛的交叉性学科. 近年来, 随着数据的爆炸式增长与计算机性能的大幅提高, 优化问题的规模也越来越大, 这对优化算法的研究带来了新的机遇与挑战. 在本论文中, 我们主要针对大规模无约束优化问题, 研究了相关的算法, 理论以及应用.

在第二章中, 我们提出了求解无约束优化问题的广义三次正则化方法, 并给出了在一定条件下, 利用精确 Hesse 矩阵或 BFGS 近似时算法的全局收敛性. 在此基础上, 我们利用形变范数, 给出了基于特征值分解的三次正则化算法, 在这个方法中子问题的最优解可以直接算出. 我们将此方法应用于求解 CUTEst 中的无约束优化问题集, 数值结果表明了我们算法的有效性. 此外, 针对大规模问题, 我们设计了有限内存的三次正则化方法, 并在算法中将拟牛顿步用作试探步以提高算法的表现. 在 CUTEst 上的数值实验表明, 相比于已有的有限内存算法, 我们的算法能有更好的迭代数表现.

在第三章中, 我们研究了随机子空间校正方法. 首先, Gauss-Seidel 方法, 以及求解线性最小二乘问题的随机梯度下降方法之间的等价性, 并且指出它们是随机子空间校正方法的特例. 接着我们给出了两种随机逐次子空间校正算法用来求解非线性凸优化问题, 并在一定条件下证明了它们的收敛性.

在第四章中, 我们研究如何让 ℓ_1 正则化在深度卷积神经网络的稀疏化中发挥更好的作用. 首先, 我们证明了简单对偶平均算法 (SDA) 可以被表示为随机梯度下降算法 (SGD) 的一种扰动. 接着我们给出了一定条件下 SDA 对于经验风险极小化问题的收敛性. 然后, 我们将正则对偶平均算法 (RDA) 写成前向-后向分裂算子的形式, 指出了 RDA 与 SDA 在迭代上更显著的联系. 基于此, 我们将 RDA 用于带 ℓ_1 正则化的卷积神经网络, 并为之设计了合适的初始化和自适应稀疏重训练技巧. 在 MNIST, CIFAR-10, CIFAR-100 数据集上的实验结果表明了我们的算法有很好的精度和稀疏度.

关键词: 大规模无约束优化, 三次正则化方法, 有限内存算法, 随机子空间校正方法, 对偶平均方法, 稀疏神经网络