**摘要**

从压缩的和严重损坏的观察数据中恢复矩阵是鲁棒统计中的一个基本问题，在计算机视觉和机器学习中有着丰富的应用。理论上，在一定的条件下，这个问题可以通过一个自然的凸松弛来在多项式时间内解决，称为压缩主成分追踪(CPCP)。然而，许多已有的CPCP可证明收敛算法存在着每次迭代代价超线性的问题，这严重限制了它们对大规模问题的适用性。在本文中，我们提出了可证明收敛的、可扩展的和有效的方法来解决CPCP(本质上)的每次迭代代价是线性的。我们的方法结合了法兰克-沃尔夫的经典思想和近端方法。在每次迭代中，我们主要利用Frank-Wolfe用第1次奇异值分解更新低秩分量，利用近端步骤对稀疏项进行更新。讨论了收敛结果和实现细节。通过对可视化数据的数值实验，证明了该方法的实用性和可扩展性

公式1.5下面一部分：

受到这些理论结果,来自不同领域的研究人员利用CPCP解决许多实际问题,包括视频背景建模[3],批处理图像对齐[7],面对验证[8],光度立体[9],动态MRI[10],话题建模[11],潜变量的图形模型学习[12]和异常值检测和稳健主成分分析[3],这些只是其中一部分。

生活在大数据时代，这些应用大多涉及大数据集和高维数据空间。因此，为了充分发挥该理论的优势，我们需要CPCP算法具有可证明的收敛性和可扩展性。这促使许多研究发展问题(1.4)及其变体的一阶方法;e。g见[13,14,15,16,17,18]。]。这些方法本质上都是利用核范数的近端算子的封闭表达式，其中涉及奇异值分解。因此，每次迭代的主要代价是计算与输入数据相同大小的SVD。

这比现成的内部点求解器(如SDPT3[19])更具可扩展性。然而，每次迭代的超线性代价限制了这些一阶方法在涉及数千个数据点和数千个维度的问题上的实际适用性。对于真正大规模的应用来说，计算完整或部分svd序列的需求是一个严重的瓶颈。

作为改进，本文设计了可扩展的算法来求解CPCP，使其在每次迭代中只计算一级奇异值分解。我们的方法利用了两个经典的和广泛研究的想法——弗兰克-沃尔夫迭代来处理核规范，以及接近的步骤来处理“L1规范”。这是解决大规模CPCP问题的一种理想的技术组合。特别是，它产生的算法比一阶方法(如ISTA和FISTA[20])具有更大的可扩展性，并且在实践中收敛速度比Frank-Wolfe的直接应用快得多。