doi:10.3969/j.issn.1673-4785.2009.06.002

方差优化问题的复杂性:从生产线到计算机网络

许晓云,王 龙
(北京大学 工学院,北京 100871)

摘 要:在许多工程系统中,方差优化对保持性能稳定、提高系统服务质量具有重要的意义. 方差优化问题也是组合优化中较困难的二阶离散优化问题. 通过引入完成时间方差和等候时间方差这 2 个重要的子类问题,具体论述了此领域的研究现状与最新理论进展,讨论了方差优化的 2 个重要特例问题的复杂性,指出了其中一个特例问题属于 P 类问题,其所有最优解均具有对称螺旋结构,且此螺旋型的结构还存在于一大类任意阶的偏差问题. 基于方差问题的特性,总结并拓展了其在实际工程领域、特别是计算机网络系统领域的新应用.

关键词:完成时间方差;等候时间方差;服务质量;准时化生产原则
中图分类号: TP3　**文献标识码**:A　**文章编号**:1673-4785(2009)06-0475-08

Complexity of variance optimization: from production lines to computer networks

XU Xiao-yun, WANG Long
(Department of Industrial Engineering and Management, Peking University, Beijing 100871, China)

Abstract: Variance optimization is essential to providing performance stability and quality of service (QoS) in many engineering systems. It is widely regarded as one of the more difficult problems in quadratic combinatorial optimization. Two important subclasses are the completion time variance problem and the waiting time variance problem. The current status of research and new theoretical developments in these fields needed to be examined in detail. The complexities of two special cases of these problems were explored, and one of them was shown to belong to Class P. It was further shown that a symmetrical spiral-shaped structure inheres in all optimal solutions of this case, and also exists in a more generalized case of the deviation problem with arbitrary order. Aspects of practical implementation of the variance optimization problem were summarized and further extended to the area of computers and their network systems.

Keywords: completion time variance; waiting time variance; quality of service; just-in-time

方差作为一种描述系统不确定性的数学概念,在实际中常用于衡量系统的准确反应能力和服务质量的优劣. 它是准时化生产原则(just-in-time, JIT)的核心组成部分[1],也是评估服务质量(quality of service, QoS)的重要标准之一[2],在生产领域与计算机领域都有着广泛的应用.

方差优化问题是许多工程系统中经常碰到的一类问题. 例如在生产系统中,工件的准时完成对生产流程具有重要影响. 系统中方差如果过大,工作的完成则很难做到准时,时间上不是过早就是过晚,整个流程中的制品(work in process, WIP)数量将显著增加,使得生产系统的缓存区迅速被占满,整个流程陷入迟滞[3]. 由于方差优化问题数学结构整齐,且具有很强的实际应用背景,因此近年来在学术领域受到了广泛的关注[4]. 作为一大类有关提前(earliness)和拖期(tardiness)规划问题的代表,其相关研究已由传统的生产领域拓展到了计算机和网络领域[5].

本文将对方差优化问题的相关研究进行全面回

收稿日期:2009-06-21.
基金项目:国家自然科学基金资助项目(10972002).
通信作者:许晓云. E-mail: xiaoyun.xu@pku.edu.cn.

顾,并探讨其问题本身的复杂性以及相应 2 个特例问题的复杂性. 本文进一步指出其中一个特例问题存在特殊的解析结构,属于 P 类问题[6],且最优解具有螺旋状的几何特征,并指出螺旋最优性的结果广泛存在于一大类任意阶的偏差问题.

1 方差优化问题研究现状

方差优化问题最早由 Merten 和 Muller[7] 于 1972 年在解决计算机文件整理系统问题时提出. 他们定义了规划中最常见的 2 类方差优化的问题:第 1 类称为"加权完成时间方差(weighted completion time variance, WCTV)优化",也称"流程时间方差(weighted flow time variance)优化". 此类优化问题着眼于工作时间的一致性,使得同一组工作尽可能在接近的时间完成,从而体现出服务稳定性;第 2 类问题称为"加权等候时间方差(weighted waiting time variance, WWTV)优化". 此类优化问题则侧重描述队列特性及用户体验,使得通过处理器的各个工作(或用户)在队列中等候的时间尽可能相等. 这 2 类问题均为生产系统 JIT 研究[1] 和计算机领域 QoS 优化研究[8] 的重要组成部分,因此一直受到学术界的广泛关注.

Merten 和 Muller[7] 提出了方差优化问题的基本数学模型:给定 n 个工作,由 j 表示其编号. 每个工作的加工时间 p_j 和权重 v_j 均为正常数且已知,如何将这些工作排序,才能使得目标方程 $\sum_{j=1}^{n} v_j(c_j - \bar{c})^2$(WCTV 问题)或 $\sum_{j=1}^{n} v_j(w_j - \bar{w})^2$(WWTV 问题)最小化? 在目标方程中, c_j 为第 j 个工作的完成时间, $\bar{c} = (\sum_{j=1}^{n} v_j c_j)/(\sum_{j=1}^{n} v_j)$ 是加权平均完成时间;同理, w_j 为第 j 个工作的等待时间, $\bar{w} = (\sum_{j=1}^{n} v_j w_j)/(\sum_{j=1}^{n} v_j)$ 为加权平均等候时间.

可以看出,此类问题属于优化中的二阶组合优化问题. 其目标方程中既包括 c_j 与 $\bar{c}$(或 w_j 与 $\bar{w}$)时间差的二阶罚函数,又附加了对工作权重的考虑,数学结构较一般一阶规划问题复杂. 从整数规划建模的角度看,其目标方程与约束方程的构造形式将至少不低于二阶[9],很可能达到三阶甚至四阶.

方差优化问题的研究工作在过去 30 年中持续开展,取得了一系列重要的成果. Merten 和 Muller[7] 证明了在单台服务器(single machine)上,WCTV 问题和 WWTV 问题是等效的,即给定 WCTV 问题的一个任意可行解,将其工作顺序逆反后将成为 WWTV 问题的可行解,且目标方程的值与原 WCTV 问题相等. 转换过程只需 $O(n)$. 这个重要性质的发现,使得后续研究只需要处理 WCTV 问题或是 WWTV 问题中的一个即可,且一个问题的任何有关工作排序的性质都可以直接转换为另一个问题的相关性质. 这个结果深化了对该问题排序结构的认识,实现了单台服务器上的 WCTV 问题和 WWTV 问题的统一.

加权二阶排序问题在数学结构上的复杂性对直接研究原问题提出了很高的要求. 因此,很多研究者都转向研究原问题的一类重要的特例,即所有工作都具有相同权重的情况. 由于这种假设在实际生产中普遍存在[3],应用也相对广泛,因此得到了学术界的重视. 对于此类特例问题,相应的 WCTV 和 WWTV 问题的标称符号也就化简为 CTV(completion time variance)和 WTV(waiting time variance). Eilon 和 Chowdhury[10] 研究了单台服务器 WTV 问题,并提出了其最优解一定呈 V 型,即最小的工作之前的工作一定都按从大到小排列,而其之后的工作一定都按从小到大排列. 这也是继 Merten 和 Muller 等效性证明后的最重要的结果之一. 此性质给出了最优解必须服从的几何形式,大大降低了寻找最优解的难度,并为后续启发式算法的研究搭建了平台. Li 等[11] 进一步指出单台服务器 WTV 问题的最优解一定是紧凑的(tight schedule),证明了单机最优解的可穷举性. Cai[12] 将 V 型最优的结果推广到了一般的 WCTV 问题. 他指出,如果工作较大意味着其权重较小(称为"一致加权"(agreeably weighted)),则最优解仍将维持 V 型. 尽管"一致加权"属于非常强的条件,但由于其解释了一大类原问题的最优解结构,因此也具有重要的意义.

对于单台服务器 CTV 问题的研究也有一系列有意义的成果. 1975 年 Schrage[13] 证明了最大的工作一定会排在最前,并对前 4 个最大的工作在最优解中对应的位置提出了猜想,此猜想后来也称为 Schrage 猜想. 针对于此,Kanet[14] 提出了一个包含 8 个工作的反例. Vani 和 Raghavachari[15] 随后给出了对包含 18 个以下工作的 Schrage 猜想的弱条件证明. Schrage 猜想的完全解决由 Hall 和 Kubiak[16] 于

1991 年完成. 与猜想的原命题不同，他们指出只有前三大(而不是前四大)的工作的位置是符合原始猜想的. 此外，Manna 和 Prasad[17] 给出了单机 CTV 问题中最小工作可能位置的一个界定；Bagchi 等人[18]考虑了一个二阶的公共到期时间(common due date)问题，即优化完成时间与公共到期时间之差的平方和 ($\min\sum_{j=1}^{n}(c_j-d)^2$). 他们证明了单机 CTV 问题与公共到期时间问题等效，指出了方差优化问题是公共到期时间问题二阶罚函数的一个特殊形式. 这个结果建立了单台服务器下方差优化问题与公共到期时间问题的直接关联，也进一步拓展了方差优化问题的应用领域.

随着更多更好的性质的发现，单机 CTV 和 WTV 问题的算法研究也取得了长足的进展. Eilon 和 Chowdhury[10] 针对 WTV 问题提出了第一组启发式算法，其结构非常简单，但并不能保证找到最优解；De 等人[19] 为 CTV 问题提出了 $O(n^2\sum_{j=1}^{n}p_j)$ 的伪多项式时间算法，也从一个角度揭示了 CTV 问题不可能是强 NP-hard 问题(NP-hard in the strong sense[6])；Gupta 等[20] 对 CTV 问题提出了一个基于遗传算法的启发式算法；Ventura 和 Weng[9] 从建模出发，对 CTV 问题提出了一个新的(相对于 Kubiak 的动态规划模型[21])二阶整数规划模型，并通过拉格朗日优化方法得到了问题的最优解的一个下限. 在近似算法方面，最好的结果目前由 Kubiak[22] 保持，其算法可以在 $O(n^2/\in)$ 内得到任意逼近最优解的结果，但是由于其算法基于动态规划，且结构相对复杂，因此不利于实时实现. 最近的一个针对 WTV 的启发式算法由 Ye 等人[23] 提出，称为“Yelf 螺旋”. 由于“Yelf 螺旋”的计算结果在统计上优于 Eilon 和 Chowdhury[10] 的一系列启发式算法，且算法结构非常简单，因此适用于如 CPU 规划等需要高速运算的环境中.

作为一个二阶组合优化问题，方差优化在单机情况下已经相对复杂，当其推广到并行服务器(Identical Parallel Machine)环境中时则难度更高. 这也一定程度上导致了并行服务器下有关 CTV 和 WTV 的研究成果相对稀少. Hall[24] 从到期时间角度出发，解决了一个一阶的类似问题，即优化完成时间与公共到期时间(common due date)之差的绝对值的和 ($\sum_{j=1}^{n}|c_j-d|$). Hall 证明了此问题在单台和并行服务器环境下均属于 Class P 问题，并提供了相应的算法；Cai 和 Cheng[25] 发现了并行服务器 CTV 问题最优解所具备的一系列重要性质，并证明了在并行服务器下 CTV 问题与二阶公共到期时间方差问题等效，从而推广了 Bagchi 等人的单台服务器结果[18]. 基于最优解的性质，Cai 和 Cheng 提出了 2 个算法：其中一个为最优算法(optimal algorithm)，可以在 $O(n^{2m}P^m(P-P^m)^{m-1}/[m^m(m-1)!]^2)$ 时间内得到最优解，其中 m 为平行服务器个数，$P=\sum_{j=1}^{n}p_j$，P^m 为 m 个最大的工作的加工时间之和；另一个算法为近似算法，计算速度较最优算法快，可以在 $O(nP^m(P-m)^{m-1}/[m^2(m-1)!]^2)$ 时间内找到一个接近最优解的近似解. 由于他们的算法中存在隐性穷举，因此计算速度相对较慢. 在启发式算法的设计方面，Marango 等人[26] 针对同样的问题提出了 4 个不同的启发式算法. 比较结果显示，模拟退火算法(simulated annealing)表现最优，在验证集测试中获得了离最优解下限 5% ~10% 的解，但是，由于文章中给出的相应模拟退火计算设定的描述过于简略，因此不利于重复试验结果. 在序列特征方面，Xu 和 Ye[27] 从并行服务器 WTV 问题入手，证明了 CTV 问题和 WTV 问题在并行服务器上也是等效的，即任取其中一个问题的可行解，就可在 $O(mn\lg n)$ 时间内(m 为平行服务器个数) 建立另一个问题的对应可行解，且两者目标方程值相同. 这个证明将 Merten 和 Muller 的等效性结果[7] 从单台服务器环境推广到了并行服务器环境. 另外，Xu 和 Ye 证明了在多台并行服务器的情况下，最优解不一定是紧凑的，即服务器牺牲一定的等候时间将有利于最小化方差. 此结果也显示出 Li 等在单机环境最优解的紧凑结果[11] 是不可以简单推广的.

2　方差优化问题的复杂性

由于方差优化问题的重要性，其计算复杂性一直受到广泛的关注，然而，这方面的研究工作进展相对缓慢. 这一方面是由于原问题本身结构复杂，直接构建最优算法的难度较大；另一方面也是因为一直没有能够找到合适的其他 NP-hard[6] 问题与之相对应.

在试图直接处理原问题遇到困难后，研究者开始转为考虑原问题的 2 个常见的特例情况，即所有工作具有相同权重(“等权重”)和所有工作具有相

同的加工时间（“等大小”）. 与原问题相比，这 2 个特例问题都仅考虑了工作的一个属性（大小或权重），显著降低了问题难度.

Merten 和 Muller[7] 首先研究了这 2 种特例情况. 他们证明了在“等权重”和“等大小”这 2 种情况中，WTV 和 CTV 问题的可行序列本身都存在对偶性，即对任意一个可行序列都存在一个对偶序列，其产生的目标方程值和原序列相同，且构建难度不超过 $O(n)$.

对偶序列在一般规划问题中属于比较少见的强性质. 其发现也引发了企图证明特例问题为 Class P 的热潮. 由此也产生了一系列例如 Schrage 猜想[13]、最小工作的定位[17]和最大三工作定位[16]等试图逼近最优解序列结构的重要成果.

然而研究显示，即使是这 2 种特例情况，计算复杂性的证明也具有相当的难度. 对偶序列的存在虽然体现了可行解集的良好的对称性，但是由于对偶性仅证明了等效序列的存在，并没有直接阐述最优解本身的序列构成，因此并没有直接证明问题的复杂性.

直至 1999 年，Kubiak[28] 通过构建 Submodular Boolean Program，证明了单机情况下 CTV“等权重”的特例情况为 NP-hard. 这个证明结束了自 1972 年以来对“等权重”问题是否为 NP-hard 的争论，也间接证明了原问题的复杂度至少为 NP-hard. 结合文献[19]对此问题的伪多项式时间算法，可以看出 Kubiak 实际上证明了“等权重”特例问题属于普通 NP-hard 问题（NP-hard in the ordinary sense[6]）. 这表明，即使规划问题的可行解集具有非常良好的性质（例如对偶性），问题也有可能具有高计算复杂性.

3　方差优化问题的特例：螺旋最优性

3.1　方差优化中的“等大小”特例

随着 Kubiak 成功地将“等权重”的特例问题归为 NP-hard，关于方差优化原问题计算复杂性的讨论也就此告一段落. 但是，对于方差优化问题的另一类特例，即工作加工时间相等（“等大小”）的问题，在计算复杂性方面却没有相关的进展.

“等大小”的特例情况最早由 Merten 和 Muller[7] 作为原问题的一个简化形式提出. 他们证明了在单台服务器情况下，无论是 WTV 问题还是 CTV 问题，其可行序列都具有反对偶性，即对任意一个可行序列，将其顺序逆反后，就可以得到一个与原序列目标方程值相同的序列. 然而对于其最优解的结构，他们表示：

“In the search for the minimum variance schedules, certain relationships can be derived between schedules for the special problems where either all weights are equal or all the processing times are equal. Even in these special cases, the characteristics of the optimum schedules are not obvious.”

即“等大小”和“等权重”这 2 类特例问题的最优解结构都不显然.

在 1972-1992 年的 20 年中，相关研究工作主要集中在解决方差优化中的“等权重”特例问题上. 这很大程度上是由于“等权重”的情形在传统的生产规划中十分常见. 在传统生产系统中，工件的加工时间往往不一，而重要程度对于整个系统来说却几乎相同. 另外，在很多情况下，由于生产需要，工厂车间内加工机器的布局往往是根据所需处理工件的加工时间而不是优先级来制定的[3]. 由此可见，解决“等权重”的特例情况对处理传统工业领域的问题具有更直接的实用价值. 由于规划领域的研究一直都是以解决传统生产行业的问题为主导，这也一定程度上促使了“等权重”问题先于“等大小”问题得到研究和解决.

然而，随着计算机和网络技术的迅速发展，“等大小”的特例情况也找到了重要的应用领域. 在计算机系统中，工作（或指令）在下放到 CPU 处理时会被分割为大小相等的数据包[29]，这些数据包根据各自所包含数据的不同，会被系统指定不同的处理优先级. 在批处理的过程中，这些数据包会在缓存区被整合为具有固定大小的批次，然后按照 CPU 时钟顺序逐批处理. 在这种处理模式中，数据包大小相等而优先级各异，CPU 以固定数量为单位逐批对工作进行处理，这与“等大小”特例问题的设定几乎完全相同. 另外，随着网络视频、音频以及各种其他网络计算服务的发展，如何有效地提升服务质量（QoS）成为近年来计算机领域的研究热点. 网络服务研究表明，理想的计算机网络服务系统会对用户需求提供统一而稳定的服务[2]，而完成时间和等候时间是其中 2 个重要的性能指标. 在本质上，方差优化问题要求工作的完成时间或等候时间尽可能的一致，这也正反映了 QoS 对服务稳定性的需求. Merten 和 Muller 提出的网络文件整理系统[7] 和 Ye 等人提出的网络准入控制系统[30] 都是这类应用的典型实例.

3.2　螺旋最优性

以完成时间优化为例，单台服务器环境下的"等大小"特例情况的定义如下：

给定 n 个工作，加工时间全部相同，其区别反映在不同的权重上. 如何能找到这些工作的一个排序，使得完成时间的加权方差最小，即

$$\min\sum_{j=1}^{n}v_j(c_j-\bar{c})^2.$$

Cai[31] 的研究显示此问题可解，但并未指出其最优解的几何结构必须呈现对称螺旋形. 为了描述方便，考虑如下的简单情形：

给定 9 个工作，工作大小均相等，而权重为 1，2，…，8，9. 这些工作在单台服务器上按一定顺序逐个完成. 此问题的最优解的排序形式如图 1 和图 2 所示：

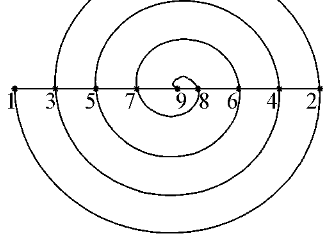


图 1　9 个工作情况下的螺旋最优（形式 1）

Fig. 1　Spiral optimality in a nine job case (Form 1)

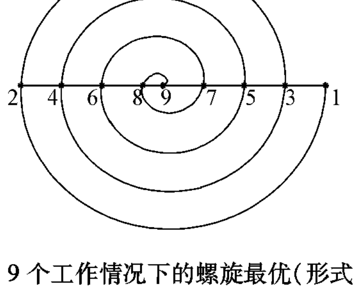


图 2　9 个工作情况下的螺旋最优（形式 2）

Fig. 2　Spiral optimality in a nine job case (Form 2)

值得强调的是，本文中给出的螺旋线仅为结构示意，不同的画图方式或可产生相同示意效果的图. 为了不产生歧义，给出图 1 和图 2 中的螺旋线作法：螺旋线自具有最小权重的工作开始，螺旋向内，到权重最大的工作结束，螺旋线在起始端的初始旋转方向垂直向下.

最优解的排序既可以为图 1 中的形式，也可以为图 2 中的形式. 这 2 种螺旋形式不仅结构对称，而且产生的目标方程值也是相同的. 这个结果也与 Merten 和 Muller 反对偶性证明[7] 相吻合.

显而易见，构建如图 1 和图 2 的螺旋在计算上是非常简单的. 在问题中的权重排序随机给出时，建立螺旋只需要 $O(n+n\lg n)$，而在权重排序给定的情况下则更为容易，仅需 $O(n)$.

最优解的螺旋特征具有重要意义. 从复杂性理论的角度，它证明了"等大小"的特例问题属于 Class P，而不是 NP-hard，结束了长久以来学术界对这个问题复杂性的争论. 在数学结构上，"等大小"问题与"等权重"问题均属于二阶组合优化问题，且目标方程的构成非常相似，但却在计算复杂性上出现了巨大差异，这是一个很有趣的现象. 从解的几何结构角度，对称螺旋的构型在形式上拓展了已知的规划解的几何模式（包括最小时间优先、最大时间优先和简单 V 型等），也从另一个侧面解释为什么 Ye 等人提出的 Yelf 螺旋[23] 启发式算法会提供好的结果. 从实际应用的角度，螺旋的结果使得计算最优解的复杂程度变成像排序一样简单，而且理论上任何能够进行排序操作的机器都可以完成这样的工作. 算法的易用性使得那些仅能完成基本排序寻址操作的简单网络服务器都可以直接采用这个算法，一般的计算型 CPU 则更可以胜任这种简单的操作.

在优化方差的前提下，螺旋最优性指出：重要的工作的"重要度"并不体现在它的完成时间最早，而是体现在它的完成时间最准确. 从目标方程可以直观看出，越重要的工作应该放置在越接近 $\bar{c}$ 的位置上. 换言之，如果认为 $\bar{c}$ 是公共到期时间[18]，则优化的意义就在于令重要工作的完成时间尽可能准时完成，从而提高系统整体服务的一致性. 最优解的螺旋特征也正体现了这样的特点. 从图 1 和图 2 中都可以明显看出，权重较高的工作都被集中放置在中间，而相对权重较低的工作则被分配在两侧，这样使得高权重的工作能够更准时地完成，也为后续处理环节的时间准确性打下了良好的基础.

3.3　一阶螺旋

值得注意的是，螺旋最优性在规划问题中尽管少见，但并不是这个问题独有的. Kanet[14] 在 1981 年考虑了"等权重"的方差优化问题. 他并没有对模型直接求解，而是提出了一个与二阶"等权重"优化模型相对应的一阶模型，其目标方程如下：

$$\min \sum_{j=1}^{n} \sum_{j=i}^{n} | c_i - c_j |.$$

Kanet 的模型令完成时间之间的差别尽可能减小. 从数学形式上看,这个模型可以认为是方差模型的一阶形式. 在方差优化中,目标方程采用的是二阶罚函数,而 Kanet 模型中采用的是一阶罚函数.

Kanet 证明了这个问题属于 Class P,并指出此问题任意实例(instance)的最优解集都包含 $2\exp([n/2]-1)$ 个元素. 文章也给出相应的一个复杂度为 $O(n + n\lg n)$ 的最优算法.

而 Kanet 并没有发现,在这个一阶"等权重"问题的最优解集中存在一个且仅有一个螺旋最优解. 考虑如下的简单情形:给定 7 个工作,每个工作权重相同,加工时间大小为 1、3、5、7、9、11、13. 则问题的最优解如图 3 所示:

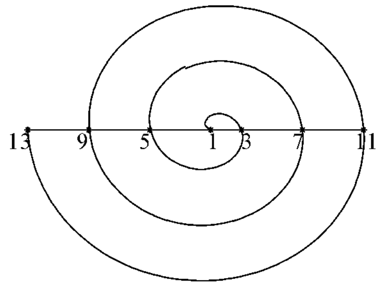


图3 7 工作情况下的 Kanet 螺旋

Fig. 3 Kanet spiral in a seven job case

Kanet 螺旋的几何示意图在本文中的作法如下所示:螺旋线自加工时间最大的工作开始,螺旋向内,到最小的工作结束. 螺旋线在起始端的初始旋转方向垂直向下.

从图 3 可以看出,螺旋线是由第一个位置开始的. 可以证明,Kanet 一阶问题中的最优螺旋线一定是从第一个位置开始. 这是由最大的工作一定要放在第一个位置[13]所决定的. 当 Kanet 问题从一阶形式升级为二阶时,不仅难度变为 NP-hard[28],而且其螺旋最优性也没有被相应的二阶"等权重"问题所继承,而是在二阶"等大小"问题中重新出现,且其最优解的形式也由很多种变成了只能为左右螺旋的 2 种. 这在一定程度上体现了二阶形式对方程解的结构要求更为严格.

3.4 广泛的螺旋最优性

螺旋型是一个良好的数学结构,它的几何形式优美,而且计算非常简单. 随之而来的问题是:类似的螺旋形式是否广泛存在? 产生这种螺旋型最优解需要一种怎样的方程结构?

Xu[32]考虑了如下的一个广泛的目标方程形式:

$$\min \sum_{j=1}^{n} \sum_{i=1}^{n} v_i^{\alpha} | c_j - c_i |^{\beta} v_j^{\alpha}.$$

式中:α 和 β 为常数,且 $\alpha \geqslant 0, \beta > 0$.

以上的问题称为"广义加权完成时间偏差问题"(generalized weighted completion time deviation problem,GWCD). 从方程形式可以看出,此目标方程试图减小工作完成时间之间的差别,且要求重要的工作应在相对接近的时间完成. 目标方程中对工作重要性和完成时间接近程度的要求十分灵活,在实际应用中可以通过调整 α 和 β 的大小来控制.

GWCD 问题的目标方程结构自由度很高,可以转化为很多规划领域的其他常见问题. 当 $\alpha = 0$, $\beta = 1$ 时, 此问题可以转化为上一节提到的 Kanet 模型[14]; 当 $\alpha = 1, \beta = 1$ 时, 此问题成为了 Kanet 模型的一阶加权推广; 当 $\alpha = 1, \beta = 2$ 时, 此问题可以转化为一般的完成时间方差优化问题[32]; 当 $\alpha = 0$, $\beta = 2$ 时, 此问题转化为方差优化问题中"等权重"的特例情况,进而等效于 Bagchi 等人提出的优化完成时间与公共到期时间之差的平方和问题[18];当将完成时间 c 变为等候时间 w, 此问题又可转化为相应的等候时间方差问题[7].

Xu 的研究指出[32],在存在权重($\alpha > 0$)的一般情况下,如果工作大小相等,则此问题的最优解一定呈螺旋型,且任意紧凑可行解的反序都与原序列目标方程值相同. 这个结果将螺旋最优性和排序对称性从简单的二阶拓展到任意阶的情况,推广了 Merten 和 Muller[7]以及 Cai[31]的相关结论,证明了螺旋型排序广泛存在于一大类非常规性能量度问题(non-regular performance measure)中,极大拓展了结论的适用范围.

除了螺旋最优性外,Xu 从模型中还可以得到一些更一般的结论,例如:当工作的大小和权重呈正比时 ($p_j/v_j = c$,c 为正常数),最优的工作顺序在 $\alpha = 1, \beta \geqslant 1$ 时为 LPT(longest processing time),即工作按从大到小排为最优. 另外,此模型在任意阶都等效于相应的等候时间偏差问题(generalized weighted waiting time deviation problem,GWWD),即将任一问题的紧凑可行解逆序就可以得到其对应问题的一个可行解,且目标方程值相等. 此结果将 Merten 和 Muller 的二阶等效性结果[7]推广到了任意阶,拓展

了这2类问题等效性的适用范围.

4 结束语

本文回顾了方差优化问题的发展. 对其2个特例问题的工程背景、数学模型和研究进展进行了综述，指出其中一类特例问题属于 Class P，且发现具有螺旋最优性. 还进一步指出螺旋型的存在并不是此问题独有的性质，在其他相关问题中也有出现，并可以拓展到任意阶的情况. 螺旋型的结构简单，实现难度与排序相仿且为最优解，可以直接应用于 CPU 规划等快速计算环境. 螺旋型是一种解结构的强性质. 它一方面显示了问题解的特殊几何特征，另一方向也是一类问题的最优解. 本文指出，在工作大小相等而权重各异时，螺旋最优性广泛存在于一大类非常规性能量度问题中. 是否工作大小不等时也存在类似性质则值得进一步探讨. 另外，给出螺旋最优性的更一般的存在条件也将对解决类似问题提供有效的帮助.

从解决方差优化问题的文献数量上看，直接处理原命题（既包含工作大小，也包含权重）的结果相对稀少，大多数研究都在处理原命题下的各类子问题. 由于方差问题的广泛性与实用性，对原命题相应的研究工作可以在未来进一步开展.

此外，从文中的 Kanet 螺旋的目标方程可以看出，此类问题与方差问题非常类似，是一种利用一阶手段描述完成时间（或等待时间）离散程度的方法. 这种简化一定程度上保持了描述的物理量的基本结构，并显著降低了问题的复杂性. 因此，将此类方程推广到既包含工作大小，也包含权重的情况将是一个有意义的研究方向.

目前方差优化问题的研究工作主要集中在单台服务器上. 在大型生产或并行计算环境中，常常存在多台服务器并联同时处理工作的情况[33]. 因此在多台并行服务器的环境中考虑方差优化问题也是一个有意义的研究方向. 另外，研究还可以进一步扩展到存在混合连接结构的 job shop 中，相关的理论进展将对方差优化的实际应用产生一定的指导意义.

相对于其他规划问题，方差优化问题更关注完成时间（或等待时间）的集中性和准确性. 这也是计算机和网络服务质量评估的一个重要方面. 随着计算机和网络技术的迅速发展，规划问题的研究对象也从传统的工业领域转移到了如 CPU 调度、网络传输等需要更快速、更稳定性能的环境中. 新的应用领域对原有的优化问题和处理方式也提出了新的要求：一方面，问题的变量出现了新的限制（如要求工作大小一致、权重各异等）；另一方面，问题的解法也要求必须简单高效. 以网络服务器为例，一般的网络服务器每秒钟需要处理超过 1 MB 的数据流量，其中央处理器也只能完成类似排序的简单操作[29]. 这就意味着非排序类的“复杂”算法，如模拟退火算法、遗传算法、分支定界法等将很难在这些应用领域得以实现. 这也对难度较高的规划问题的解决方法提出了新的挑战.

参考文献：

[1] CHENG T C E, PODOLSKY S. Just-in-time manufacturing: an introduction[M]. 2nd ed. London, UK: Chapman & Hall, 1996: 6-12.

[2] CHEN Y, FARLEY T, YE N. QoS requirements of network applications on the internet[J]. Information, Knowledge, Systems Management, 2003, 4(1): 55-76.

[3] HOPP W J, SPEARMAN M L. Factory physics[M]. 2nd ed. New York, USA: McGraw-Hill/Irwin, 2000: 287-331.

[4] BAKER K R, SCUDDER G D. Sequencing with earliness and tardiness penalties: a review[J]. Operations Research, 1990, 38(2): 193-242.

[5] YE N. Secure computer and network systems: modeling, analysis and design[M]. London, UK: John Wiley and Sons, 2008: 53-79.

[6] GAREY M R, JOHNSON D S. Computers and intractability: a guide to the theory of NP-completeness[M]. New York, USA: W H Freeman and Company, 1979: 1-181.

[7] MERTEN A G, MULLER M E. Variance minimization in single machine sequencing problems[J]. Management Science, 1972, 18(9): 518-528.

[8] YE N. QoS-centric stateful resource management in information systems[J]. Information Systems Frontiers, 2002, 4(2): 149-160.

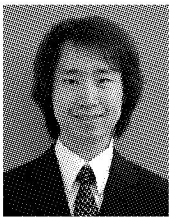
[9] VENTURA J A, WENG M X. Minimize single-machine completion time variance [J]. Management Science, 1995, 41(9): 1448.

[10] EILON S, CHOWDHURY I G. Minimizing waiting time variance in the single machine problem[J]. Management Science, 1977, 23(6): 567-575.

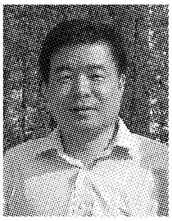
[11] LI X, YE N, LIU T, et al. Job scheduling to minimize the weighted waiting time variance of jobs[J]. Computers & Industrial Engineering, 2007, 52(1): 41.

[12] CAI X. Minimization of agreeably weighted variance in single machine systems[J]. European Journal of Operational Research,1995,85(3): 576-592.
[13] SCHRAGE L. Minimizing the time-in-system variance for a finite jobset[J]. Management Science,1975,21(5): 540-543.
[14] KANET J J. Minimizing variation of flow time in single machine systems [J]. Management Science,1981,27 (12): 1453-1459.
[15] VANI V,RAGHAVACHARI M. Deterministic and random single machine sequencing with variance minimization [J]. Operations Research,1987,35(1): 111-120.
[16] HALL N G,KUBIAK W. Proof of a conjecture of Schrage about the completion time variance problem[J]. Operaions Research Letters,1991,10(8): 467-472.
[17] MANNA D K,PRASAD V R. Bounds for the position of the smallest job in completion time variance minimization [J]. European Journal of Operational Research,1999, 114(2): 411-419.
[18] BAGCHI U,SULIVAN R S,CHANG Y L. Minimizing mean squared deviation of completion times about a common due date[J]. Management Science,1987,33(7): 894-906.
[19] De P,GHOSH J B,WELIS C E. On the minimization of completion time variance with a bicriteria extension[J]. Operations Research,1992,40(6): 1148-1155.
[20] GUPTA M C,GUPTA Y P,KUMAR A. Minimizing flow time variance in a single machine system using genetic algorithm[J]. European Journal of Operational Research, [993,70(3): 289-303.
[21] KUBIAK W. New result on the completion time variance minimization[J]. Discrete Applied Mathematics,1995, 58(2): 157-168.
[22] KUBIAK W,CHENG J,KOVALYOV M Y. Fast fully polynomial approximation schemes for minimizing completion time variance[J]. European Journal of Operational Research,2002,137(2): 303-309.
[23] YE N,LI X,FARLEY T,et al. Job scheduling methods for reducing waiting time variance[J]. Computer and Operations Research,2007,34(10): 3069-3083.
[24] HALL N G. Single and multiple processor models for minimizing completion time variance[J]. Naval Research Logistics Quarterly,1986,33(1): 49-54.
[25] CAI X,CHENG T C E. Multi-machine scheduling with variance minimization[J]. Discrete Applied Mathematics, 1998,84(1/3): 55-70.
[26] MARANGOS C A,GOVANDE V,SRINIVASAN G,et al. Algorithms to minimize completion time variance in a two machine flowshop[J]. Computers and Industrial Engineering,1998,35(1/2): 101-104.
[27] XU X,YE N. Minimization of job waiting time variance on identical parallel machines[J]. IEEE Transactions on Systems,Man,and Cybernetics,2007,37(5): 917-927.
[28] KUBIAK W. Completion time variance on a single machine is dffioult[J]. Operations Research Letters,1993,14 (1): 49-59.
[29] NULL L,LOBUR J. The essentials of computer organization and architecture[M].2nd ed. London,UK: Jones and Bartlett Publishers,Inc,2006: 177-179.
[30] YE N,FARLEY T,LI X,et al. Batch scheduled admission control for computer and network systems[J]. Information Knowledge Systems Management,2005,5(4): 211-226.
[31] CAI X. A solvable case of the variance minimization problem[J]. Applied Mathematics Letters,1993,6(6): 97-100.
[32] XU X. Generalized completion time deviation poblem on a single machine[R]. Beijing: Peking University,2009.
[33] PINEDO M L. Scheduling: theory,algorithms, and systems[M]. 3rd ed. New York,USA: Springer,2008: 111-142.

作者简介:



许晓云,男,1980年生,博士后,主要研究方向为计算复杂性理论、计算机运筹规划以及计算机网络等.发表学术论文多篇.



王 龙,男,1964年生,教授、博士生导师,主要研究方向为复杂系统智能控制、多机器人系统的协调与控制、网络化控制系统的分析与综合、集群行为与集群智能、演化博弈与群体决策等,特别是在参数摄动系统、离散事件系统、混合集成系统的分析与控制方面,作出了突出贡献,取得了一系列具有国际水平的重要成就.日本学术振兴基金获得者.其研究成果被国内外广泛引用,并获得国家教委霍英东奖(研究类一等奖)、国家自然科学奖、国家教委科技进步奖(一等奖)、第一届Ho Outstanding Paper Award、第一届关肇直控制理论奖等多项奖励.