IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

WARSZAWA 1969

Control of Large Scale Systems

Fourth Congress of the International Federation of Automatic Control
Warszawa 16–21 June 1969

TECHNICAL SESSION

35

Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Control of Large Scale Systems

TECHNICAL SESSION No 35

FOURTH CONGRESS OF THE INTERNATIONAL
FEDERATION OF AUTOMATIC CONTROL
WARSZAWA 16 – 21 JUNE 1969

Organized by
Naczelna Organizacja Techniczna w Polsce

C o n t e n t s

Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 56/69.

ОБ ОПТИМАЛЬНОМ РАСПРЕДЕЛЕНИИ РЕСУРСОВ

А.Я.Лернер, А.И.Тейман

Институт автоматики и телемеханики
(технической кибернетики)

Москва
СССР

Решение любой технической или экономической задачи требует проявления активности, выражающейся в использовании соответствующих ресурсов - людских; материальных, временных, финансовых для достижения поставленных целей. Поскольку эффективность решения любой задачи, а также затраты времени и средств существенно зависят от способа маневрирования используемыми для этой цели ресурсами, естественно возникает проблема разработки принципов такого управления ресурсами, при котором достигалось бы наивыгоднейшее из возможных значение критерия, характеризующего в целом результат решения и процесс его получения.

Эта проблема оптимального управления ресурсами в своей принципиальной сущности одинакова для самых различных конкретных задач, независимо от их индивидуальных особенностей и масштаба. Идет ли речь о формировании графика работ ремонтной бригады или о плане реализации крупной народно-хозяйственной программы - математическая формулировка задачи оптимального управления ресурсами по существу одинакова.

Несмотря на то, что к настоящему времени уже опубликован ряд интересных исследований, посвященных оптимальному распределению ресурсов [1,2,3 и др.] , нам представляется, что и постановки этих задач и способы их решения еще далеки от завершения и оставляют большой простор как для теорети -

ческих разработок, так и для изыскания новых приложений.

В настоящем докладе излагается попытка дать строгие определения основным понятием и наметить пути решения некоторых задач теории, предпринятую в Отделе управления большими системами Института автоматики и телемеханики [x]. Ряд предлагаемых определений, а также классификация задач и методов не является общепринятой. Автор надеется, что дискуссионное обсуждение позволит выработать общую точку зрения по этим вопросам, что крайне необходимо в настоящее время.

I. Основные понятия и определения.

I.I. Операция - процесс, описываемый уравнением вида

$$W_i(t) = \frac{dx_i(t)}{dt} = \varphi_i\left[u_{i1}(t), \ldots, u_{im}(t), t\right] \qquad (I)$$

где

$x_i(t)$ - состояние i-ой операции в момент t,

$u_{ij}(t)$- количество ресурсов j-го вида в i-ой операции в момент t,

$\omega_i(t)$- скорость -ой операции в момент t.

Поскольку ресурсы участвуют в операции в определенных соотношениях, полезно ввести понятие набора ресурсов.

Набором ресурсов в i-ой операции называется множество распределений ресурсов $\{u_{ij}(t)\}$, таких что $u_{ij}(t) = \alpha_{ij} v_i(t)$, где $\{\alpha_{ij}\}$ - параметры набора,

$v_i(t)$- мощность набора в момент t. В операции вообще говоря возможны различные наборы, но в дальнейшем для упрощения рассматривается случай одного допустимого набора для каждой операции. При заданных параметрах набора скорость операции зависит только от мощности набора и времени (в дальнейшем будем считать, что скорость операции не зависит явно от времени. Таким образом

$$\omega_i(t) = f_i\left[v_i(t)\right] \qquad (2)$$

---

x) В этих работах принимали участие сотрудники Отдела В.Н.Бурков, А.А.Воронов, С.Е.Ловецкий, А.И.Тейман и др.

где f_i неубывающая непрерывная функция ϑ_i и $f_i(0)=0$. Под выполнением операции понимается изменение ее состояния от начального $x_i(0)=0$ до конечного $x_i(T)=W_i$, где W_i называется <u>объемом</u> операции, T - момент окончания всех операций.

I.2. Конечное множество операций образует <u>комплекс</u> операций. Комплекс обычно изображают в виде сети. Сосостояние комплекса - это вектор $X(t)=(x_1(t),\ldots,x_n(t))$, компоненты которого являются состояниями операций комплекса. <u>Объем комплекса</u> - вектор $W=(W_1,\ldots,W_n)$. Однако, иногда удобнее пользоваться понятием <u>эквивалентного объема</u>, который является скалярной величиной, зависящей от объемов операций (например, $W_э=(\sum_i W_i^{\alpha})^{1/\alpha},\ \alpha>1$).

I.3. Будем рассматривать ограничения на ресурсы двух типов: ограничения по мощности (мгновенные ограничения)

$$\sum_i \alpha_{ij}\vartheta_i(t)\le N_i(t),\quad j=1,2,\ldots,m \tag{3}$$

и ограничения по затратам (интегральные ограничения)

$$\sum_i \alpha_{ij}S_i\le S_j,\quad j=1,2,\ldots,m. \tag{4}$$

где $S_i=\int_0^T \vartheta_i(t)\,dt$ - затраты мощности.

2. <u>Описание модели</u>

2.I. Формулирование и решение задач оптимального распределения ресурсов может опираться на модель комплекса операций, включающую следующие составляющие:

а) Сеть или матрица, характеризующая состав, объемы и допустимые последовательности операций, входящих в комплекс.

б) Г раф или матрица, характеризующая допустимые перемещения ресурсов по операциям комплекса, допустимые затраты ресурсов на их выполнение, и затраты времени и средств на перемещении ресурсов.

в) Система уравнений, описывающих операции комплекса.

$$w_i(t)=f_i[\vartheta_i(t)]$$

г) Функционал $\mathfrak{J}$, характеризующий суммарную эффективность выполнения комплекса операций.

2.2. Задача оптимального распределения ресурсов состоит в том, чтобы подобрать такие программы $\varphi_{ij}^{k}(t)$ (количество ресурсов k -го вида перемещающихся с i -ой операции на j -ю в момент t), которые при соблюдении ограничений, накладываемых на допустимые последовательности выполнения операций (п.2I,а) и на допустимые значения потоков ресурсов (п.2.Iь), обеспечили бы такое течение процесса выполнения комплекса, описываемого уравнениями п.2.Iв, при котором функционал $\mathfrak{J}$ принимает минимальное значение.

2.3. В зависимости от конкретных условий функционал может зависеть от разных аргументов или их сочетаний. Наиболее важными для практики являются случаи, когда цели оптимизации включают следующие факторы:

T - время завершения комплекса операций,
R - ресурсы, выделяемые на выполнение комплекса операций,
P - вероятность завершения комплекса за время ,
K - качество результата комплекса операций.

В общем случае оптимизация плана выполнения комплекса операций должна предусматривать минимизацию некоторого функционала от перечисленных факторов

$$\mathfrak{J}(T, R, P, K)$$

Разумеется, в частных случаях некоторые аргументы функционала $\mathfrak{J}$ могут не учитываться. В зависимости от параметров, учитываемых при решении задачи либо в функционале $\mathfrak{J}$, либо в ограничениях, можно выделить классы задач схематически показанные на Рис.I (мы предполагаем, что временной параметр всегда учитывается в задаче). Задачи класса T (время) не являются задачами оптимизации. Это задачи анализа комплекса (определение времени окончания комплекса, резервов времени операций и т.д.). К классу ТР (время - надежность) относятся, в основном, задача распределения зависимых разервов времени между отдельными операциями с целью максимизации вероятности выполнения комплекса в срок. Классы задач, в которых учитывается качество, относительно мало исследованы. Остальная часть доклада посвящена задачам класса TR (время, ресурсы) и класса ТР (время, надежность).

3. Постановка задач класса

3.I. Можно выделить два основных типа задач. В задачах первого типа время T выполнения комплекса задано, в задачах второго типа время выполнения комплекса не задано. Задача I. Определить $\vartheta_i(t),\ i=1,2,\dots,n$, удовлетворяющие (3), (4), так чтобы комплекс был выполнен за время T и критерий оптимальности J принял минимальное значение.

Задача 2. Определить $\vartheta_i(t),\ i=1,2,\dots,n$, так чтобы комплекс был выполнен и критерий оптимальности J принял минимальное значение (время не задано).

3.2. Наиболее распространены следующие критерии оптимальности:

а) $J_1 = T$ (минимизация времени комплекса)

б) $J_2 = \sum\limits_j c_j \max \sum\limits_i \alpha_{ij}\vartheta_i(t)$ (минимизация уровней ресурсов)

в) $J_3 = \sum\limits_j c_j \int\limits_0^T \left(\sum\limits_i \alpha_{ij}\vartheta_i(t)\right)^2 dt$ (равномерное использование ресурсов)

г) $J_4 = \sum\limits_j c_j \sum\limits_i S_i\alpha_{ij}$ (минимизация затрат)

д) $J_5 = aT + J_2$

е) $J_6 = aT + J_4$

В выписанных критериях $c_j \geq 0$ характеризует стоимость ресурсов j -го вида,

a - потери при задержке окончания комплекса на единицу времени.

Если имеются промежуточные цели, например, скорейшее наступление определенных событий, то применяется критерий

$$J_8 = \sum_i \sigma_i(t_i - \Delta_i)$$

где $\sigma_i(t_i - \Delta_i) = 0$ при $t_i \leq \Delta_i$ и является неубывающей функцией t_i при $t_i > \Delta_i$, t_i - момент окончания i -ой операции.

4. Агрегирование комплекса

4.I. Под агрегированием комплекса понимается замена комплекса (или его части) одной операцией [4]. Агрегирование применяется, когда комплекс состоит из большого числа операций. Непосредственное решение задачи оптимизации в этом случае затруднено из-за ее большой размерности. Агрегирование состоит из трех этапов.

I. этап. Упорядочение состояний комплекса.

П этап. Определение параметров набора агрегированной операции.

Ш. Определение зависимости скорости агрегированной операции от мощности набора.

4.2. Рассмотрим пример агрегирования комплекса из p последовательных операций, выполняемых ресурсами одного вида. Первый и второй этапы здесь отпадают, поскольку состояния комплекса уже упорядочены, а единственный параметр набора агрегированной операции можно принять равным единице. Обозначим $\tau_i(\vartheta)$ - время завершения i -ой операции при мощности ϑ ее набора, то есть

$$\tau_i(\vartheta) = \frac{W_i}{f_i(\vartheta)} .$$

Тогда по определению скорость агрегированной операции

$$f(\vartheta) = \frac{W}{\sum_i \tau_i(\vartheta)} . \tag{5}$$

Примем объем агрегированной операции $W = \sum_i \beta_i W_i$. Оценим ошибку агрегирования. Пусть мощность набора i -ой операции равна ϑ_i . Тогда время выполнения агрегированной операции

$$\tau(\vartheta_1, \ldots, \vartheta_p) = \sum_j \frac{\beta_j W_j}{f(\vartheta_j)} = \frac{1}{W} \sum_j \sum_i \beta_j W_j \tau_i(\vartheta_j)$$

Действительное время комплекса

$$\tau'(\vartheta_1,\ldots,\vartheta_p)=\sum_i \tau_i(\vartheta_i)=\frac{1}{W}\sum_j\sum_i \beta_j W_j \tau_i(\vartheta_i)$$

Ошибка агрегирования

$$\varepsilon=\tau'(\vartheta_1,\ldots,\vartheta_p)-\tau(\vartheta_1,\ldots,\vartheta_p)=\frac{1}{W}\sum_j\sum_i \beta_j W_j\left[\tau_i(\vartheta_i)-\tau_i(\vartheta_i)\right] \quad (6)$$

Задача оптимального агрегирования заключается в данном случае в определении $\beta_1,\ldots,\beta_p$, так чтобы ошибка агрегирования была минимальной.

Теорема I. Пусть $\tau_i(\vartheta_i)=a_i+b_i\tau(\vartheta_i),\ i=1,\ldots,p$. Тогда ошибка агрегирования равна нулю при $\beta_j=\frac{cb_j}{W_j}$

($c>0$ - произвольная постоянная)

Доказывается непосредственно проверкой.

Эта теорема позволяет решать задачу оптимального агрегирования, представляя зависимости $\tau_i(\vartheta)$ приближенно в виде $a_i+b_i\tau(\vartheta)$ и полагая $\beta_i^0=\frac{cb_i}{W_j}$

4.3. Задача оптимального агрегирования, в которой минимизируются максимальные относительные простои ресурсов различных видов решена в [5] для случая линейных зависимостей скоростей операций от мощности набора ресурсов. Возможность идеального агрегирования (ошибка агрегирования равна нулю) для сети произвольного вида показана в [6] для случая степенной зависимости скоростей операций от мощности набора и ресурсов из одного вида.

Перейдем к рассмотрению точных методов решения, которые разработаны для различных частных постановок. К ним относится случай независимых операций, случай упорядоченных событий, задача распределения затрат и ряд задач комбинаторного типа.

5. Независимые операции

5.I. Пусть $f_i(\vartheta_i)$ - выпуклые (вверху) функции Тогда можно показать [7] , что в оптимальном решении все операции выполняются с постоянной интенсивностью и заканчиваются одновременно.

Отсюда имеем

$$f_i(v_i) = \frac{W_i}{T}; \quad v_i = \varphi_i\left(\frac{W_i}{T}\right),$$

где φ_i - функция обратная f_i.
Минимальное время выполнения комплекса определяется как минимальное T, удовлетворяющее системе неравенств

$$\sum_i \alpha_{ij} \varphi_i\left(\frac{W_i}{T}\right) \le N_j, \quad j = 1, 2, \ldots, m. \tag{7}$$

Пример I $f_i(v_i) = v_i$, $v_i \le \beta_i$, $i = 1, 2, \ldots, n$

Имеем

$$\sum_i \alpha_{ij} \frac{W_i}{T} \le N_j, \quad j = 1, 2, \ldots, m$$

$$\frac{W_i}{T} \le \beta, \quad i = 1, 2, \ldots, n.$$

$$T_{min} = \max\left[\max_i \frac{W_i}{\beta_i}; \max_j \frac{1}{N_j} \sum_i \alpha_{ij} W_i\right] \tag{8}$$

5.2. Если $f_i(v_i)$ невыпуклые функции, то оптимальное решение состоит в общем случае из n интервалов постоянства. Важной для дальнейшего является следующая теорема.

Теорема II. Минимальное время выполнения комплекса $T_{min}(W_1, \ldots, W_n)$ является выпуклой (книзу) функцией объемов операций (даже если $f_i(v_i)$ - невыпуклые функции).

6. Упорядоченные события

6.I. События комплекса упорядочены, если момент наступления i-го события не больше момента наступления j-го события, если $i < j$.
Обозначим Δ_s - длительность интервала между $(s-1)$ и s-ым событиями, $s = 1, 2, \ldots, r$. R_s - множество операций, которые могут выполняться в s-ом интервале, Q -множество интервалов, в которых может выполняться i-я операция, x_{is} -объем i-ой операции, выполняемой в s-ом интервале. Пусть $z_s = \{x_{is} : i \in R_s\}$ заданы. Тогда для

каждого интервала получаем случай независимых операций и можем определить минимальную длительность s -го интервала $\Delta_s(\tau_s)$. Время выполнения комплекса

$$T = \sum_s \Delta_s(\tau_s) \tag{9}$$

является согласно теореме П п. 5.2 выпуклой (книзу) функцией x_{is}. Получили задачу минимизации выпуклой функции при линейных ограничениях

$$\sum_{s \in Q_i} x_{is} = W_i, \quad i = 1, 2, \ldots, n. \tag{10}$$

для решения которой можно применить любые методы выпуклого программирования.

6.2. Рассмотрим теперь задачу равномерного использования ресурсов. Пусть каждая операция выполняется ресурсами только одного вида. Обозначим $R_{s_j} \subset R_s$ - множество операций, выполняемых ресурсами j -го вида. Задача заключается в минимизации

$$J_3 = \sum_j C_j \sum_s \frac{1}{\Delta_s} \Big(\sum_{i \in R_{s_j}} x_{is} \Big)^2 \tag{11}$$

при ограничениях (10) и

$$\sum_s \Delta_s = T. \tag{12}$$

Ввиду независимости условий (10) и (12) можно осуществить процедуру последовательного улучшения некоторого допустимого решения, варьируя сначала x_{is} при фиксированных Δ_s, а затем Δ_s при фиксированных x_{is}.

I этап. Пусть заданы допустимые значения $\Delta_s \geq 0$, такие что $\sum_s \Delta_s = T$. Тогда задача минимизации J_3 распадается на m независимых задач (по числу видов ресурсов). Каждая заключается в минимизации

$$J_j = \sum_s \frac{1}{\Delta_s} \Big(\sum_{i \in R_{s_j}} x_{is} \Big)^2 \tag{13}$$

при ограничениях (10). Эта задача квадратичного программирования.

II этап. Обозначим x^o_{is} - оптимальные значения, полученные на I этапе.

$$B_s^2 = \sum_j c_j \Big(\sum_{i \in R_{sj}} x_{is} \Big)^2 \tag{I4}$$

Получим задачу минимизации

$$J_3 = \sum_s \frac{B_s^2}{\Delta_s} \tag{I5}$$

при ограничении (I2). Применяя метод множителей Лагранжа, сразу получаем

$$\Delta_s^o = \frac{B_s T}{\sum_s B_s} \tag{I6}$$

Далее этапы I и П чередуются.

6.3. Основную трудность в решении задачи вызывает I этап. Для его решения предлагался метод последовательного улучшения, основанный на гидродинамической аналогии [8,9], модификации методов квадратичного программирования [10].

Эффективный алгоритм решения I этапа можно получить исходя из следующего правила распределения ресурсов: в первую очередь выполняются операции с минимальным номером конечного события. Его применение иллюстрируется следующим примером.

Пример 3. Рассмотрим сеть из I2 операций, показанную на Рис.2 ($A_i(W_i)$ обозначает i -ю операцию объема W_i). Пусть $\Delta_1 = \Delta_2 = \ldots = \Delta_6 = 1$. Определяем x) $\lambda_o = \frac{1}{6} \sum_i W_i$ =I2. Распределяем ресурсы и помечаем события знаками. Знаком (-) помечено только событие 3. Ближайшее к нему слева (помечённое знаком (+) -событие I. Так как в интервалах 2 и 3 не выполняется ни одна операция с конечным событием большим 3, то выделяем подсеть из событий (I,2,3), а в оставшейся сети эти события объединяем в одно. При этом оставшаяся сеть представляет последовательное соединение двух сетей, одна из которых включает событие 0 и объединенное событие (I,2,3), а вторая - объединенное событие (I,2,3) и события 4,5,6. Поэ-

---

x) Алгоритм предложен В.Н.Бурковым.

тому задачу можно решать отдельно для каждой сети. Получаем три подсети, показанные на Рис.3.

Для сети G_1, $\lambda_1 = 8$ и допустимое решение $x_{11} = 3$, $x_{21} = 5$.

Для сети G_2, $\lambda_2 = 17$ и допустимое решение $x_{32} = 14$, $x_{42} = 3$, $x_{43} = 2$, $x_{63} = 15$.

Для сети G_3, $\lambda_3 = 10$ и допустимое решение $x_{74} = 5$, $x_{84} = 2$, $x_{94} = 2$, $x_{54} = 1$, $x_{55} = 7$, $x_{10,5} = 1$, $x_{11,5} = 2$, $x_{11,6} = 2$, $x_{12,6} = 8$.

Полученные значения $\{x_{ij}\}$ определяют оптимальное решение I этапа.

Значение критерия $J = 7 \cdot \lambda_1^2 + 2\lambda_2^2 + 3\lambda_3^2 = 942$.

7. Задачи класса ТР

7.I. Среди задач класса ТР можно выделить задачи максимизации надежности комплекса, т.е. задачи построения плана реализации комплекса, обеспечивающего минимальную возможность срыва запланированного срока его выполнения. Можно выделить два типа задач.

А. Распределение вероятностей $P(t)$ времени выполнения операций известно. В этом случае имеют место следующие задачи.

Задача I. Для заданного времени Т реализации комплекса найти такое распределение сроков выполнения операций $\{t_i\}$, чтобы $P\{t_n \leq T\} \to max$.

Задача 2. Для заданного уровня надежности комплекса Ро найти такое распределение $\{t_i\}$, чтобы $P\{t_n \leq T\} \geq P_0$.

Б. Распределение вероятностей $P(t)$ неизвестно. В этом случае формируется некоторый критерий $J(t, \Delta t)$, зависящий от сроков выполнения операций t и величины Δt резервируемых временных ресурсов. Значение критерия J является оценкой надежности комплекса и относительно него можно сформулировать задачи, аналогичные задачам I и 2. Ряд постановок задач типа А и Б, а также алгоритмы их решения были рассмотрены в [11], [12].

7.2. Особенность задач класса TR и задач класса ТР типа А и Б состоит в том, что при их рассмотрении предполагается, что комплекс задан. На практике тому предшествует процесс создания комплекса и возникает естественная проблема оптимального планирования этого процесса. Ниже мы приведем некоторые результаты [13], относящиеся к данной проблеме.

7.3. Модель процесса создания комплекса.

Пусть комплекс S_0 в начальный момент состоит из одной операции. Далее при планировании он разбивается на части, детализируется и в конечном итоге комплекс S_n содержит n операций. При этом при таком преобразовании операции комплекса подвергаются разбиению, уточнению и агрегированию.

Определим эти основные понятия. Предполагается, что ξ - случайная продолжительность операции и $0 \le a \le \xi \le b < \infty$

$c = b - a > 0,\ F(x) = P\{\xi \le x\},\ m = M\xi,\ \sigma^2 = D\xi$

Положим $\alpha = \sigma/(b-a)$; $\beta = (m-a)/(b-a)$. Легко видеть, что $0 \le \alpha \le 1/2$; $0 \le \beta \le 1$.

<u>Уточнение операций.</u>

Операция ξ_2 есть уточнение ξ_1, если

$$b_2 - a_2 \le b_1 - a_1 \quad \text{и} \quad \alpha_2 \le \alpha_1$$

<u>Разбиение операций.</u>

I. Пусть ξ_2 разбивается на "К" последовательных операций $\xi_1, \xi_2, \ldots, \xi_K$. Такое разбиение регулярно, если:

I. Случайные величины ξ_i независимы

2. $\sum_1^K a_i = a$; $\sum_1^K b_i = b$

3. $\alpha_i \le \alpha$

4. $\sum_1^K b_i \frac{c_i}{c} = \beta$

П. Пусть ξ разбивается на "К" параллельных операций $\xi_1, \ldots, \xi_K$. Такое разбиение регулярно, если

I. Случайные величины ξ_i независимы

2. $\max a_i = a$; $\max b_i = b$; $c_i \le c$

3. $\alpha_i \le \alpha$

4. $\max[\beta_i c_i + a_i] = \beta c + a$

III. Пусть сети S_i соответствует разбиение Ω_i, а сети S_{i+1} - разбиение Ω_{i+1}. Разбиение $\Omega_{i+1} > \Omega_i$ и регулярно, если оно получено путем регулярного разбиения операций сети S_i.

Агрегирование операций.

Под агрегированием понимается объединение нескольких операций в одну. Аналогично разбиению оно может быть последовательным и параллельным. Определим агрегирование, как операцию, обратную разбиению.

7.4. Анализ процесса планирования создания комплекса. Основные теоремы.

Теорема I. Пусть t_{i_s} -время наступления события i_s, $t_{i_s}^{(0)}$ - фиксированный плановый срок. Тогда для каждого события i_s существует число τ_{i_s} такое, что какую бы величину вероятности P_0 мы ни задали, всегда найдется последовательность регулярных разбиений сети комплекса $\Omega_0 < \Omega_1 < \ldots < \Omega_i < \ldots$, для которой начиная с Ω_0 имеет место:

$$P\{t_{i_s} > t_{i_s}^{(0)}\} \geq P_0 \quad \text{при} \quad t_{i_s}^{(0)} < \tau_{i_s}$$

и

$$P\{t_{i_s} < t_{i_s}^{(0)}\} \geq P_0 \quad \text{при} \quad t_{i_s}^{(0)} > \tau_{i_s}.$$

Если рассматривать отдельные классы разбиений, то можно доказать существенно более сильные утверждения. Например, имеет место

Теорема II. Пусть $\{\Omega_i\}$, $i=1,2,\ldots$ регулярные последовательные разбиения. Утверждения теоремы I имеют тогда место для Ω_0 и любого $\Omega_i > \Omega_0$.

Доказательства этих теорем проводятся [13] путем рассмотрения сетей, состоящих лишь из последовательных или параллельных операций и сети общего вида.

Полученные результаты показывают, что существуют некоторые критические значения сроков выполнения комплекса и при выборе плановых сроков ниже критических значений они с большой вероятностью будут сорваны.

Можно показать, что ряд процедур планирования, например, использующие усредненные показатели, являются в этом смысле некорректными.

В связи с этим возникают два класса задач: задачи построения верхних оценок $t^* \geq \tau$ и задачи оптимального выбора $t_i^{(?)}$.

8. Заключение

I. Возникшие в последние годы из запросов практики проблемы оптимального распределения ресурсов имеют огромное экономическое значение. Однако, ни основные понятия, ни точные постановки задач в достаточно общем виде до последнего времени еще не определились.

2. Изложенные в настоящем докладе вопросы отражают попытки удовлетворить назревшую необходимость в формулировании решений задач оптимального распределения ресурсов на основе моделей, достаточно хорошо отражающих реальные условия проведения комплексов операций.

Установлено, что решение этого класса задач не может основываться на каком-либо одном виде математического аппарата, а требует привлечения всего арсенала разнообразных средств математического программирования, теории графов, теории оптимального управления и др.

3. Из приведенного рассмотрения ясно, что усилия исследователей должны быть направлены на разработку методов решения ряда актуальных задач распределения ресурсов, в том числе: определение оптимальных маршрутов перемещения ресурсов с точки зрения обеспечения максимальной надежности достижения цели; разработка методов решения общей задачи оптимального распределения ресурсов в детерминированном и стохастическом аспектах.

5. Более подробное изложение некоторых из затронутых вопросов дано в работах, приведенных в списке литературы.

Литература

I. Алтаев В.Я., Бурков В.Н., Тейман А.И. Теория сетевого планирования и управления (обзор) "Автоматика и телемеханика", т.XXУП, №5,1966

2. Mcgee A.A. and Markarian M.D. Optimal Allocation of Research (Engineering Manpower within a Multi-Project Organisational Structures). IRE Trans. on Eng.Manag.,1962, v.9, N 3

3. Lambourn S. Resource allocation and multi-project scheduling (RAMPS). A new tool in planning and control Computer J., 1963, v.5

4. Бурков В.Н., Лернер А.Я. Новые задачи теории сетевого планирования и управления. Сборник "Вопросы управления большими системами." Изд-во "Онтиприбор", 1967

5. Чеботарев О.Г. Распределение ресурсов в многотемных разработках на основе агрегирования комплекса операций, Доклад на IУ Всесоюзном совещании по автоматическому управлению, Тбилиси,1968

6. Бурков В.Н. Оптимальное управление комплексами операций, Доклад на IУ Всесоюзном совещании по автоматическому управлению, Тбилиси, 1968

7. Бурков В.Н. Распределение ресурсов как задача оптимального быстродействия."Автоматика и телемеханика", т.XXУП, №7, 1966

8. Разумихин Б.С. Задача об оптимальном распределении ресурсов "Автоматика и телемеханика", т.XXУI,№7, 1965

9. Разумихин Б.С. Задачи об оптимальном распределении ресурсов. "Автоматика и телемеханика", т.XXУШ, №I, 1967

IO. Воронов А.А., Петрушинин Е.П. Решение задачи оптимального распределения ресурсов методом квадратичного программирования. "Авт. и тел.", т.XXУП, №II, 1966

II. Тейман А.И. Оптимальное планирование комплексов операций, сб."Вопросы управления большими системами", Изд-во "Онтиприбор", 1967

I2. Тейман А.И. Календарное планирование комплексов операций. Материалы Всесоюзного семинара по теоретическим проблемам управления большими системами и исследованию операций. Изд-во "Знание", Москва, 1967

I3. Тейман А.И. Некоторые задачи управления комплексами операций в условиях неопределенности. Доклад на IУ Всесоюзном совещании по автоматическому управлению, Тбилиси, 1968.

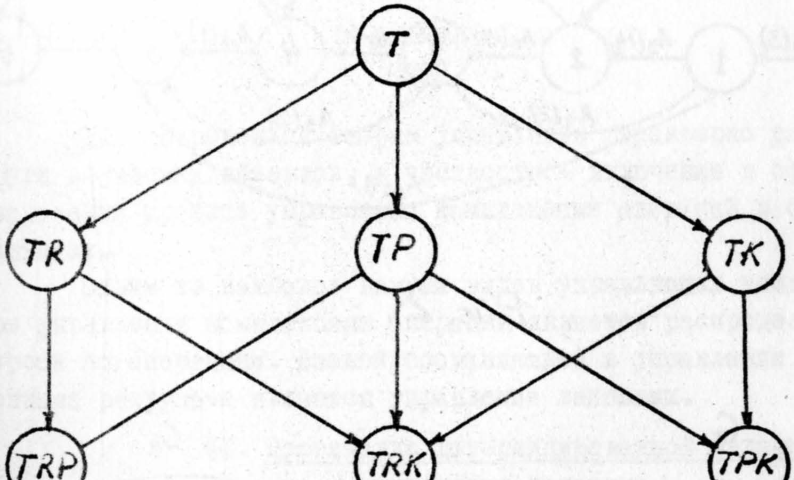

Рис. 1

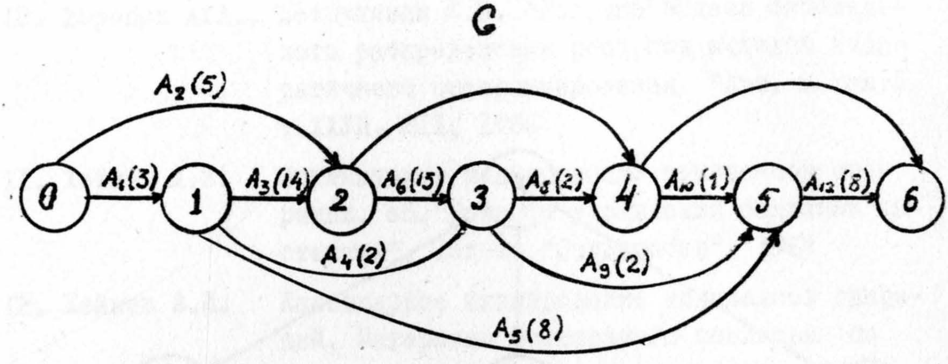
G
A2(5)
A1(3)
A3(14)
A6(15)
A8(2)
A10(1)
A12(8)
A4(2)
A9(2)
A5(8)
0
1
2
3
4
5
6

Рис. 2

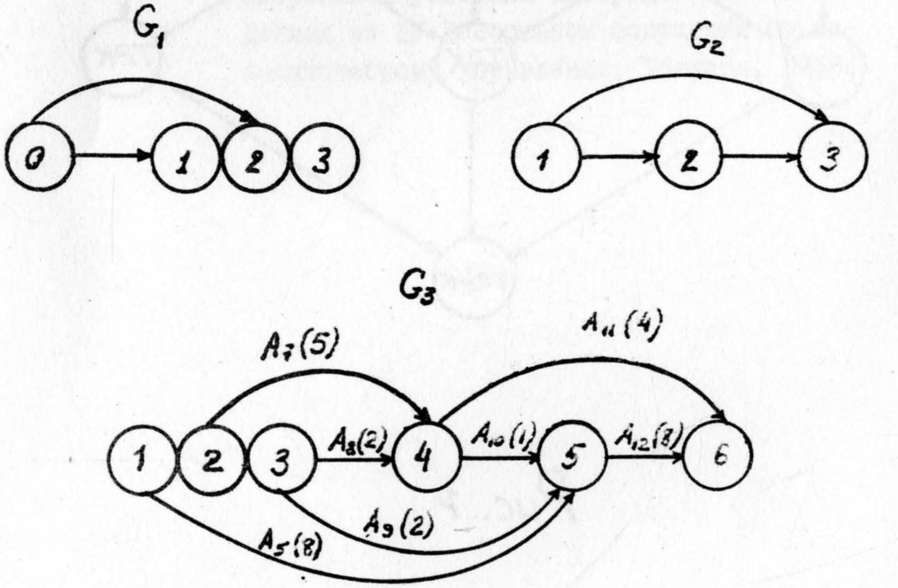
G1
0
1
2
3
G2
1
2
3
G3
A7(5)
A11(4)
A8(2)
A10(1)
A12(8)
A9(2)
A5(8)
1
2
3
4
5
6

Рис. 3

К ТЕОРИИ УПРАВЛЕНИЯ ЗАПАСАМИ

В.Н.Авдийский, А.А.Воронов, С.Е.Ловецкий
Институт автоматики и телемеханики
(технической кибернетики)
Москва
СССР

Для современной теории управления характерно расширение круга изучаемых объектов, в частности - включение в сферу исследований проблем управления комплексами операций в больших системах.

Одним из наиболее важных видов управляющих воздействий при управлении комплексами операций является распределение ресурсов по операциям. Важной составляющей в управлении ограниченными ресурсами является управление запасами.

§I. Постановка детерминированной задачи управления запасами

В докладе рассматривается задача составления оптимального плана управления запасами на складе предприятия при наличии возможности закупать часть сырья на внешнем рынке без завоза его на склад.

Предполагается, что на центральном складе крупного предприятия (например, металлургического комбината) хранятся запасы m видов сырья P_i $(i=1,2,\dots,m)$ (заготовки различных размеров), из которых требуется изготавливать n видов продукции Π_j, $j=1,2,\dots,n$. (прокат различных сортаментов). Каждый из продуктов может, вообще говоря, изготавливаться из всех, или из нескольких видов сырья (данный сортамент можно изготавливать из всех заготовок бо́льших размеров). Ограничимся рассмотрением случая, когда после выбора по тем или иным соображениям сырья вида P_i для продукта вида Π_j, другие виды сырья для изготовления этого продукта, кроме выбранного, не используются. Предполагается, что потребность b_j в количестве сырья, необходимое для изготовления продукта Π_j известна и не зависит от вида сырья. Стоимость изготовления единицы продукта Π_j из сырья вида P_i равна S_{ij}. В тех случаях, когда по каким-то соображениям изготовление продукта Π_j из сырья P_i недопустимо (например, если размер сортамента больше, чем заготовки), то соответствующим величинам s_{ij} приписываются значения $s_{ij} = \infty$

Кроме затрат на производство надо учитывать затраты на хранение сырья на складе. Величина этих затрат неизвестна и, строго говоря, зависит от искомого количества сырья, завозимого на склад. Но обычно затраты на хранение малы в сравнении с затратами на производство и мы можем упростить задачу, оценив приближенно общие расходы на содержание и обслуживание склада за время производственного цикла и разделив их на количество сырья, расходуемого за это же время. Тогда суммарные затраты на хранение и производство единицы Π_j из единицы P_i составят:

$$c'_{ij} = s_{ij} + \sigma,$$

где σ затраты на хранение единицы сырья.

В тех случаях, когда это нужно, можно учесть также расходы d_i на открытие партии продукции из сырья вида P_i (расходы, связанные с оформлением отпуска сырья со склада, его подвоза к рабочему месту и т.п.).

При некоторых условиях может оказаться выгодным для изготовления продукта Π_j не завозить сырье предварительно на центральный склад, а закупить на внешнем рынке - оптовых магазинах, непосредственно у изготовителей и т.д., в тот момент, когда в нем возникает потребность. Пусть затраты на приобретение, доставку и другие возможные расходы вместе с расходами на производство продукции из единицы этого закупаемого во вне сырья равны h'_j , а η'_j -количество докупаемого сырья. Тогда задачу можно сформулировать так:

Требуется найти такие значения x'_{ij} и η'_j ,которые минимизируют суммарные затраты, выражаемые функционалом:

$$\left[\sum_{ij} c'_{ij} x'_{ij} + \sum_i d_i \operatorname{sgn} \sum_j x'_{ij} + \sum_j h'_j \eta'_j\right] \to min \qquad (1)$$

при условиях:

$$\sum_i x'_{ij} + \eta'_j = b_j, \quad j = 1, \dots, n; \qquad (2)$$

$$x'_{ij} \geqslant 0; \quad \eta'_j \geqslant 0; \quad i = 1, \dots, m; \quad j = 1, \dots, n, \qquad (3)$$

т.е. потребность должна быть полностью удовлетворена и все переменные должны быть неотрицательными.

Для удобства дальнейших рассуждений произведем замену переменных
$$x_{ij} = \frac{x'_{ij}}{b_j}\ ; \quad \eta_j = \frac{\eta'_j}{b_j},$$
и введем новую переменную
$$y_i = \operatorname{Sgn} \sum_j x_{ij} = \begin{cases} 0, \text{ если } \sum_j x_{ij} = 0\ ; \\ 1, \text{ если } \sum_j x_{ij} > 0\ ; \end{cases}$$
и запишем задачу (I) - (3) в следующем виде
$$z = \left[\sum_{i,j} c_{ij} x_{ij} + \sum_i d_i y_i + \sum_j h_j \eta_j\right] \rightarrow min \qquad (4)$$
при условиях
$$\sum_i x_{ij} + \eta_j = 1, \quad j = 1, \ldots, n\ ; \qquad (5)$$
$$y_i = \{0, 1\}, \quad i = 1, \ldots, m\ ; \qquad (6)$$
$$0 \leq x_{ij} \leq 1\ ;\ 0 \leq \eta_j \leq 1\ ;\ i = 1, \ldots, m\ ;\ j = 1, \ldots, n, \qquad (7)$$
где $\quad c_{ij} = c'_{ij} \cdot b_j\ , \quad h_j = h'_j \cdot b_j.$

Задача (4) - (7) относится к классу задач смешанного линейного программирования. Задачи такого типа можно эффективно решать с помощью алгоритмов основанных на идеях метода ветвей и границ [I].

В следующем разделе рассмотрим алгоритм решения задачи (4) - (7), являющийся обобщением алгоритма предложенного в [2] для решения задачи размещения производства.

§ 2. Алгоритм решения детерминированной двухиндексной задачи управления запасами.

Метод ветвей и границ, который будет изложен ниже применительно к задаче (4) - (7), является конечным и эффективным методом решения экстремальных задач комбинаторного типа, задач целочисленного программирования [3], задач теории расписаний [4] и др., который позволяет существенно сократить перебор. Описанию метода ветвей и границ и его применениям посвящено много работ [I].

Процесс ветвлений, как обычно, будем производить по дискретной переменной y_i, полагая ее поочередно равной 0 и I. Пусть мы умеем находить нижнюю границу функционала (4) (способ вычисления которой будет описан ниже) в каждой вершине получаемого дерева решений, т.е. для каждого подмножества множества допустимых решений, пренебрегая при этом целочисленностью y_i. Тогда начиная с исходной вершины v_0 дерева решений, вычисляем значение нижней границы $z \geqslant \bar{z}_0$ в этой вершине. Если все y_i целочисленны, то задача решена. Если некоторое y_k дробное, то полагаем сначала $y_k = 0$, а затем $y_k = 1$, образовав тем самым две ветви (v_0, v_1) и (v_0, v_2) из вершины v_0, заканчивающиеся двумя новыми вершинами v_1 и v_2. (Если несколько y_k оказалось дробным, то возникает интересная задача связанная со стратегией выбора следующей переменной для ветвления). В каждой из полученных висячих вершин вычисляем нижние границы, которые обозначим соответственно $\bar{z}_1$ и $\bar{z}_2$. Среди множества висячих вершин $Q = \{v_1, v_2\}$ находим вершину с минимальной нижней границей.

Пусть для определенности это будет вершина v_1, т.е.

$$\bar{z}_1 = min(\bar{z}_1, \bar{z}_2), \quad \bar{z}_1 \geqslant \bar{z}_0.$$

Теперь мы ветвимся из вершины $v_1 \in Q$, приравнивая другую нецелочисленную переменную, например y_r, ($r \neq k$) сначала 0, а затем I, получая в результате две новые вершины v_3 и v_4, для которых аналогичным образом вычисляем нижние границы $\bar{z}_3$ и $\bar{z}_4$. Снова среди множества висячих вершин $Q = \{v_2, v_3, v_4\}$ находим вершину с наименьшей нижней границей, пусть это будет вершина v_3, т.е.

$$\bar{z}_3 = min\{\bar{z}_2, \bar{z}_3, \bar{z}_4\}; \quad \bar{z}_3 \geqslant \bar{z}_1,$$

из которой продолжаем процесс ветвлений. Описанный процесс продолжается до тех пор, пока мы не придем в "конечную" вершину дерева решений, или другими словами, пока не достигнем вершины в которой все y_i, являются целыми. Если при этом окажется, что значение z достигаемое функционалом (4) на этом решении удовлетворяет неравенству:

$$z \leqslant \min_{v_i \in Q} \bar{z}_i$$

для всех висячих вершин v_i , то полученное решение оптимально.

Перед тем как описать способ вычисления $\bar{z}_i$ заметим, что в оптимальном решении задачи (4) - (7) в точности n переменных x_{ij}, η_j будут равны I, а все остальные будут равны 0. Это станет очевидным, если учесть, что в условиях задачи (4) - (7) нет ограничений на количество перевозимой продукции. Таким образом оптимальное решение задачи (4) - (7) ищется среди x_{ij}, η_j принимающих значение 0 или I ($i=1,\dots,m$; $j=1,\dots,n$).

Перейдем к процедуре вычисления нижней границы $\bar{z}_i$ в каждой вершине дерева решений v_i . Заметим, что так как процесс вычисления нижней границы повторяется многократно в процессе решения задачи, желательно иметь по возможности простой способ ее вычисления.

Обозначим через N_j множество видов сырья, из которых можно изготавливать продукцию Π_j , а через M_i множество видов продукции, которые можно изготовлять из i -ого вида сырья P_i , хранящегося на складе, и пусть n_i обозначает количество элементов M_i .

Если теперь для некоторой вершины дерева v_i обозначим S_1 и S_0 множества индексов i при которых y_i в процессе решения (на пути из v_0 в v_i) присвоены значения I и 0 соответственно, а через S множество оставшихся i , то значение нижней границы можно получить подставляя в (4) следующие величины x_{ij}, η_j и y_i :

$$x_{ij}=1;\ \eta_j=0,\ \text{если}\ \min_{i\in S_1\cup S}\{[c_{ij}+(g^i/n_i)],\ h_j\} = [c_{kj}+(g^k/n_k)]; \tag{8}$$

$$x_{ij}=0;\ \eta_j=1,\ \text{если}\ \min_{i\in S_1\cup S}\{[c_{ij}+(g^i/n_i)],\ h_j\}=h_j$$

в остальных случаях x_{ij} и η_j равны 0.

$$y_i=\frac{1}{n_i}\sum_{j\in M_i}x_{ij},\quad i\in S, \tag{9}$$

где
$$g^i=\begin{cases} d_i, \text{если} & i\in S\\ 0, \text{если} & i\in S_1\end{cases}$$

Очевидно, что выражения (8) и (9) определяют оптимальное решение задачи (4) - (7) при условии, что

$$0 \leq y_i \leq 1, \quad i = 1, \dots, m \tag{10}$$

и следовательно, значение функционала (4), которое достигается на них является нижней границей для задачи (4) - (7).

Чтобы убедиться в этом заметим, что для $i \in S$

$$\sum_{j \in M_i} x_{ij} \leq n_i y_i,$$

которое войдет в оптимальное решение задачи (4) - (7), (10) со знаком равенства, т.е.

$$\sum_{j \in M_i} x_{ij} = n_i y_i \quad \text{или} \quad \sum_{j \in M_i} \frac{x_{ij}}{n_i} = y_i.$$

Подставляя это значение y_i, $i \in S$ в функционал (4) получаем

$$\min z = \sum_{i \in S_1} d_i + \min \Big\{ \sum_{i \in S_1} c_{ij} x_{ij} + \sum_{i \in S} \Big[c_{ij} + \frac{d_i}{n_i}\Big] x_{ij} + \sum_j h_j \eta_j \Big\}$$

при условии, что

$$\sum_i x_{ij} + \eta_j = 1, \quad j = 1, \dots, n.$$

§ 3. Постановки задач управления запасами при неизвестном спросе.

В большинстве практических ситуаций величина спроса b заранее не известна. Нам могут быть известны или распределение вероятностей вектора b, $F(b_1, \dots, b_n)$ или только границы, в которых может изменяться спрос. Тогда постановка (I) - (3) становится некорректной и необходимо определить, что мы понимаем под решением задачи.

В работе [5] рассматривается ряд возможных постановок задач стохастического программирования. Одна из таких возможных постановок заключается в минимизации функционала (I) при ограничениях.*

---

* Здесь и в дальнейшем под x_{ij} и η_j будем понимать абсолютные величины, которые в § I обозначались x'_{ij} и η'_j.

$$\sum_i x_{ij} + \eta_j \geqslant b_j(q), \; q \in Q \qquad (2')$$

$$x_{ij} \geqslant 0, \; \eta_j \geqslant 0, \qquad (3)$$

где q случайный параметр или набор случайных параметров, определяющий реализацию случайных элементов условий задачи, Q множество значений q, появляющихся с ненулевой вероятностью.

Постановки такого типа называются жесткими или одноэтапными постановками задач стохастического программирования, так как в таких задачах предполагается, что решение принимается один раз и корректировка принятого решения при появлений дополнительных сведений о реализации условий задачи не допустима.

Более разумными представляются двухэтапные постановки задач стохастического программирования. В таких постановках процесс принятия решения можно представить разделенным на два этапа. На первом этапе выбирается некоторый план X, не обязательно удовлетворяющий всем условиям задачи при всевозможных реализациях q. Затем фиксируется некоторая реализация q и, следовательно, реализация вектора $b(q)$ и вводится вектор η, корректирующий принятое решение. Считая, что мы несем дополнительные издержки при корректировке решения принятого на первом этапе, естественно поставить задачу минимизации суммы математического ожидания значения целевой функции задачи для принятого решения и дополнительных издержек за корректировку плана.

Рассматривая процесс принятия решений к задаче (I) - (3) как двухэтапный считаем, что на первом шаге определяется план, X, а после реализации вектора b в случае нехватки сырья выплачивается штраф пропорциональный величине недопоставки $\eta_j = b_j - \sum_i x_{ij}$ (этот штраф можно рассматривать как стоимость срочных докупок недостающего сырья по более высоким ценам из некоторого источника, являющегося внешним по отношению к рассматриваемой системе). Тогда ставится задача минимизации математического ожидания суммарных затрат

$$M\left[\sum_{i,j} c_{ij} x_{ij} + \sum_i d_i y_i + \sum_j h_j \eta_j\right] \to min \qquad (II)$$

при ограничениях

$$\sum_i x_{ij} + \eta_j = b_j \tag{12}$$

$$x_{ij} \geq 0, \ \eta_j \geq 0. \tag{13}$$

здесь η_j величины недопоставок, а h_j величина штрафа за единицу недопоставок. Заметим, что если величины штрафов h_j становятся бесконечно большими то двухэтапная постановка задачи (II) - (I3) переходит в жесткую постановку (I), (2), (3).

Для нахождения детерминированного эквивалента (т.е. задачи математического программирования, решение которой является решением соответствующей задачи стохастического программирования задачи (II) - (I3) зафиксируем некоторый план X и рассмотрим (II) - (I3) как задачу минимизации по переменной η т.е.

$$\sum_j h_j \eta_j \longrightarrow min$$

при ограничениях

$$\sum_i x_{ij} + \eta_j \geq b_j \quad \text{или} \quad \eta_j \geq b_j - \sum_i x_{ij}$$

так как остальные члены функционала (II) не зависят от вектора η. Выпишем для этой задачи двойственную ей задачу максимизации

$$\sum_j \Big(b_j - \sum_i x_{ij}\Big)\xi_j \longrightarrow max$$

при ограничениях

$$\xi_j \leq h_j;$$

$$\xi_j \geq 0.$$

Очевидно, что решение этой задачи дается равенствами $\xi_j = h_j$. Отсюда в силу известной теоремы двойственности

$$\min_\eta \sum_j h_j \eta_j = \max_\xi \sum_j \Big(b_j - \sum_i x_{ij}\Big)\xi_j = \sum_j \Big(b_j - \sum_i x_{ij}\Big)h_j.$$

Подставляя правую часть последнего равенства в функционал (II) - получим следующий детерминированный эквивалент задачи (II) - (I3): минимизировать функционал (I4)

$$\sum_{i,j} c_{ij} x_{ij} + \sum_i d_i y_i + \sum_j h_j M[b_j - \sum_i x_{ij}] \qquad (I4)$$

при ограничениях

$$x_{ij} \geq 0; \qquad (I5)$$

$$y_i = \{0, 1\} \qquad (I6)$$

Так как часть переменных задачи (I4) - (I6) целочисленна, то сформулированная задача так же представляет собой смешанную задачу программирования, для решения которой предлагается алгоритм основанный на идеях метода "ветвей и границ" описываемый ниже.

§ 4. Алгоритм решения задачи.

Метод решения задачи (I4) - (I6) основан на идеях метода ветвей и границ. Процесс ветвлений, как и ранее производится по дискретной переменной y_i, полагая ее поочередно равной 0 и I.

Рассмотрим теперь вопросы возникающие при вычислении нижних границ применительно к задаче (I4) - (I6).

Прежде всего обратим внимание на некоторые свойства "усеченного" математического ожидания $M[b_j - \sum_i x_{ij}]$ при $\sum_i x_{ij} \leq b_j$, которые будут полезны для вычисления оценок.

Рассмотрим функцию $F(x) = M[b - x]$, где b случайная величина с известной плотностью вероятности $\varphi(b)$ по $x \leq b$.

Очевидно, что

$$F(x) = \int_x^\infty (b - x)\varphi(b)\,db.$$

Пользуясь этим выражением легко вычислить производную $F'(x)$:

$$F'(x) = \int_0^x \varphi(b)\,db - 1$$

Вычислив вторую производную

$$F''(x) = \varphi(x) \geq 0$$

убеждаемся, что функция $F(x)$ является выпуклой книзу. Отсюда следует, что и функция

$$F\left(\sum_i x_{ij}\right) = M\left[b_j - \sum_i x_{ij}\right]$$

является выпуклой книзу функцией своих аргументов.

Рассмотрим теперь задачу минимизации функционала

$$\Phi(x) = \sum_i k_i x_i + F\left(\sum_i x_i\right) \to min$$

при ограничениях $x_i \geq 0$ и считая, что $F\left(\sum_i x_i\right)$ является выпуклой книзу функцией, а все $k_i > 0$. Предположим так же существование непрерывной производной функции $F\left(\sum_i x_i\right)$. Предположения о положительности k_i и выпуклости $F\left(\sum_i x_i\right)$ обеспечивают существование минимума, который может достигаться или внутри конуса $x_i \geq 0$ или на границе. Допустим, что минимум достигается внутри области из непрерывности производной следует, что точка минимума должна удовлетворять системе уравнений.

$$\frac{\partial \Phi(x)}{\partial x_i} = 0$$

или записывая более подробно

$$k_i + F_z'\left(\sum_i x_i\right) = 0, \quad z = \sum_i x_i$$

и мы сразу убеждаемся, что эта система уравнений при $k_i \neq k_j$, $i \neq j$ является несовместной, более того никакие два уравнения из этой системы не совместны, если только не $k_i = k_j$ при $i \neq j$. Это противоречие и означает, что минимум достигается на границе. Тогда легко видеть, что минимум достигается на векторе в котором все компоненты равны нулю, кроме одной. Индекс этой компоненты определяется выбором минимального т.е. $k_i = \min_i k_i$, а ее величина является корнем уравнения

$$k_i + F'(x_i) = 0 \qquad (ж)$$

случае, когда имеется несколько минимальных k_i равных между собой, нам безразлично какое количество каждой из них будет взято, требуется только, что бы их сумма равнялась количеству, определенному из уравнения (ж).

Сделаем еще одно замечание. Очевидно, что если разрешить y_i принимать непрерывные значения $0 \leq y_i \leq 1$, то задача минимизации функционала (I4) распадается на подзадач для каждого индекса j, а каждая такая подзадача представляет собой задачу минимизации функционала типа $\Phi(x)$.

Из вышесказанного следует алгоритм получения нижней границы для функционала (I4) в некоторой вершине дерева решений v_i.

Обозначим через S_0 и S_1 множества y_i, которым в процессе решения присвоены значения 0 и I соответственно, а через S множество оставшихся y_i. Обозначим через n_i количество видов продукции Π_j, которые можно изготовлять из i -го вида сырья P_i.

Значения нижней границы можно получить теперь подставляя в (I4) следующие величины x_{ij} и y_i:

$$x_{ij} = 0, \text{ если } i \in S_0$$

Пусть $\min\limits_i c_{ij} = c_{i_0 j}$

$$x_{ij} = 0, \text{ если } i \neq i_0;$$

$$x_{ij} = x_{ij}^{(0)}, \text{ если } i = i_0,$$

где $x_{ij}^{(0)}$ является корнем уравнения

$$\frac{\partial}{\partial x_{ij}}\left[c_{ij} x_{ij} + h_j M[b_j - x_{ij}]\right] = 0.$$

Вспомнив выражение для производной от "усеченного" математического ожидания, получим, что $x_{ij}^{(0)}$ является корнем уравнения

$$\int_0^{x_{ij}} \varphi(b_j)\, db_j = 1 - \frac{c_{ij}}{h_j}$$

$$y_i = 0, \text{ если } i \in S_0$$

$$y_i = 1, \text{ если } i \in S_1$$

$$y_i = \frac{\sum \operatorname{sgn} x_{ij}}{n_i}, \text{ если } i \in S$$

Указанные выше значения x_{ij} и y_i дают решение задачи минимизации (I4) - (I5), если $0 \leq y_i \leq 1$, и поэтому значение функционала достигаемое на них является оценкой снизу для задачи (I4) - (I6) в некоторой вершине дерева решений v_i.

Литература

I. Lawler E.L., Wood D.E. "Branch-and-Bound Methods. A Survey". Opns. Res. V. I4, N 4, I966.

2. Efroymson M.A., Ray T.L. "A Branch and Bound Algorithm for Plant Location". Opns. Res. V. I4, N 3, I966.

3. Balas E. "An Additive Algorithm for Solving Linear Programs with Zero-one Variable". Opns. Res. V. I3, N 4, I965.

4. Ignall Edward, Schrage Linus. Application of the Branch-and-Bound Technique to some Flow-Shop Scheduling Problems". Opns. Res. V. I3, N 3, I965.

5. Гольштейн Е.Г., Юдин Д.Б. "Новые направления в линейном программировании". Изд-во "Советское Радио", I966.

РАСПРЕДЕЛЕНИЕ РЕСУРСОВ В МНОГОТЕМНЫХ РАЗРАБОТКАХ НА ОСНОВЕ АГРЕГИРОВАНИЯ КОМПЛЕКСОВ ОПЕРАЦИЙ

О.Г.Чеботарев
ЦЭМИ АН СССР
г.Москва

Разработка многотемных систем СПУ и, в частности, задача распределения общих ограниченных ресурсов при одновременном ведении нескольких комплексов, отличается сложностью и, вместе с тем, несомненной практической важностью.

В докладе рассматривается метод решения задач распределения общих ограниченных ресурсов в многотемных проектах на основе агрегирования комплексов операций при линейных зависимостях скоростей выполнения операций от мощностей наборов ресурсов.

I. Основные понятия и определения

Операцию будем описывать дифференциальным уравнением вида

$$v_i(t)=\frac{dx_i(t)}{dt}=f\left[r_{ij}(t),t\right], \qquad (I)$$

где $x_i(t)$ - состояние i-ой операции в момент времени t, $x_i(t_н)=0$, $x_i(t_к)=w_i$ ($t_н$ и $t_к$ - моменты начала и окончания операции соответственно и w_i - объем операции); $v_i(t)$ - скорость выполнения операции i в момент времени t, $r_{ij}(t)$ - количество ресурсов j -го вида (j = I, 2, .., m), занятых на выполнении i -ой операции в момент t; m - количество видов ресурсов, занятых в многотемном проекте (в случае, если ресурс j -го вида не используется при выполнении i -ой операции, то соответствующий параметр $r_{ij}(t)=0$).

Под выполнением операции понимается изменение ее состояния от начального $x_i(t_н)=0$ до конечного $x_i(t_к)=w_i$.

Вектор $r_i(t)=\{r_{ij}(t)\}$ в m -мерном пространстве будем называть <u>набором ресурсов</u> в i -ой операции.

Предполагая, что соотношение между координатами вектора $\vec{r}_i(t)$ не изменяется во времени, т.е. что вектор $\vec{r}_i(t)$

может изменяться только по модулю, представим его в виде: $\vec{r}_i(t)=\rho_i(t)\cdot\vec{\alpha}_i$.

Функцию $\rho_i(t)$ будем называть мощностью набора ресурсов, вектор $\vec{\alpha}_i=\{\alpha_{ij}\}$, компонентами которого являются количества ресурсов j -го вида при $\rho_i(t)$ = I, вектором параметров набора ресурсов в i -ой операции, а его компоненты α_{ij} параметрами набора.

Комплексом операций будем называть конечное множество из n операций, для которых определено отношение порядка: $i\to j$ (i предшествует j , или j следует за i), если j -ю операцию нельзя начинать, пока не окончена i -я операция.

Под выполнением комплекса понимается изменение его состояния $x(t)=[x_1(t),\dots,x_n(t)]$ от $x(0)=0$ до $x(T)=W$, $W=(W_1,\dots,W_n)$, где W - объем комплекса и T - момент его завершения.

Комплекс операций изображается в виде сети. Операциям соответствуют дуги сети (прямая сеть) либо вершины сети (сопряженная сеть). Комплекс выполнен, если выполнены все его операции.

Многотемным проектом называется проект, охватывающий комплексы операций, описываемые не связанными технологически между собой сетями, но объединяемые в единый проект общностью используемых ресурсов и другими ограничениями.

Ограничения на ресурсы. На ресурсы, используемые при выполнении проекта, накладываются ограничения вида:

$$\sum_{p=1}^{l}\sum_{i=1}^{n_p}\alpha_{ij}\cdot\rho_i(t)\le N_j(t) \tag{2}$$

либо

$$\sum_{p=1}^{l}\sum_{i=1}^{n_p}\alpha_{ij}\cdot q_i\le S_j, \tag{3}$$

где $q_i=\int_0^T\rho_i(t)dt$ - затраты набора ресурсов в i -ой операции, S_j - максимально возможные затраты ресурсов j -го вида в процессе выполнения проекта, $N_j(t)$ - максимально возможное использование ресурсов j -го вида в момент времени t, i - номер операции комплекса (i = I, 2,.., n_p) и p - номер комплекса в многотемном проекте (p =I,2,.., l).

Ограничения (2) называют мгновенными ограничениями на ресурсы, а сами ресурсы, подчиняющиеся ограничениям такого вида, нескладируемыми, возобновляемыми или N - типа (например, люди, станки, оборудование, рабочие места и т.п.).

Ограничения (3) называют интегральными, а ресурсы, подчиняющиеся таким ограничениям, складируемыми, невозобновляемыми, типа затрат или S - типа (например, сырье, деньги, энергия, человеко-часы, машинное время и т.п.).

Агрегирование комплекса. Под агрегированием комплекса понимается замена комплекса одной операцией. Для этой операции определяются $\vec{\alpha}$, w и $f[\rho(t), t]$ по заданным величинам $\vec{\alpha}_i$, w_i и $f_i[\rho_i(t), t]$ для каждой операции комплекса.

Введенные понятия следуют в основном работе Буркова В.Н. и Лернера А.Я.[1]

2. Задача распределения ограниченных ресурсов в многотемном проекте

Необходимо выполнить многотемный проект, состоящий из ℓ комплексов операций объемов w_1, w_2, ..., w_ℓ. Ресурсы всех или некоторых видов, необходимые для выполнения проекта, ограничены. Задача состоит в распределении ресурсов между операциями комплексов таким образом, чтобы некоторый критерий оптимальности принимал свое экстремальное значение и выполнялись ограничения (2) и (3).

В качестве критерия оптимальности, например, может быть взят $\min\limits_{p} \max T_p$, что соответствует выполнению многотемного проекта за минимальное время. T_p - момент окончания p-го комплекса. Более общим критерием является $\min \sum\limits_{p=1}^{\ell} \varphi_p(T_p)$, где $\varphi_p(T_p)$ - неубывающая функция T_p -го. $\varphi_p(T_p)$ может соответствовать штрафу за задержку в выполнении p-го комплекса (например, $\varphi_p(T_p) = 0$ при $T_p \leq T_{p\,дир}$ и $\varphi_p(T_p) = a_p(T_p - T_{p\,дир})$ при $T_p > T_{p\,дир}$, где $T_{p\,дир}$ - директивный срок окончания p-го комплекса) и a_p - константа).

Использование агрегирования комплексов операций позволяет представить решение задачи распределения ограниченных ресурсов между операциями многотемного проекта в виде последовательных этапов.

1. Выполняется агрегирование всех комплексов операций, в результате которого для каждого комплекса определяются зависимости $v_p(t) = f_p[\rho_p(t), t]$, w_p и $\vec{\alpha}_p$.

2. На основании определенных для каждого комплекса $v_p(t)$, w_p и $\vec{\alpha}_p$ решается задача распределения ограниченных ресурсов между ℓ независимыми операциями (комплексами) по заданному критерию оптимальности. В результате решения этой задачи определяются величины количеств ресурсов N_{pj} и S_{pj} каждого вида, которые могут быть использованы при выполнении каждого комплекса p.

3. Решаются полученные ℓ задач распределения ограниченных ресурсов между зависимыми операциями.

Таким образом задача распределения ограниченных ресурсов между $n = \sum_{p=1}^{\ell} n_p$ операциями сводится к решению ℓ задач распределения ограниченных ресурсов между n_p ($p = 1, 2, .., \ell$) операциями.

В качестве примера рассмотрим такой подход к решению задачи распределения ограниченных ресурсов типа затрат в многотемном проекте, включающем ℓ комплексов операций, когда за критерий оптимальности принимается $\min \max_p T_p$. Будем предполагать, что зависимости затрат ресурсов на выполнение каждой операции от ее продолжительности являются выпуклыми книзу функциями.

В этом случае зависимости $S_p(T_p)$ для каждого комплекса p могут быть получены с помощью алгоритма Бермана[2]. В силу того, что все зависимости $S_p(T_p)$ - невозрастающие функции T_p -го, минимальное время выполнения всех комплексов достигается при их равномерном выполнении[3], т.е. $T_1 = T_2 = T_3 = \ldots = T_\ell = T$. Следовательно, значение T может быть найдено из уравнения с одним неизвестным T, получаемым из ограничения (3) на ресурсы типа затрат:

$$S_1(T) + S_2(T) + \cdots + S_\ell(T) = S.$$

После этого задача сводится к решению ℓ задач распределения ресурсов типа затрат таким образом, чтобы каждый комплекс был выполнен в заданное время с минимальными затратами ресурсов. Эти задачи также решаются с помощью алгоритма Бермана.

3. Агрегирование комплекса операций.

Агрегирование комплекса состоит из трех этапов:

I. Упорядочение состояний комплекса.

П. Определение возможного вектора набора ресурсов комплекса.

Ш. Определение скорости выполнения комплекса для возможного вектора набора ресурсов в зависимости от мощности набора.

I этап. Поскольку операция представляет собой упорядоченную последовательность состояний (для любых двух состояний x_i^1 и x_i^2 всегда выполняется либо $x_i^1 < x_i^2$, либо $x_i^1 > x_i^2$), то такое же условие должно выполняться для состояний комплекса, если мы хотим представить его в виде одной операции. Легко показать, что для упорядочения состояний комплекса необходимо и достаточно:

I) упорядочить события комплекса, т.е. установить очередность их свершения;

2) задать числа y_{is}, определяющие объемы операций, выполняемых в s-ом интервале (в интервале между ($s-1$)-ым и s-ым событиями, s = I,2,.., q), и удовлетворяющие условию $\sum_{s\in Q_i} y_{is} = W_i$, где Q_i - множество интервалов, в которых может выполняться i-я операция;

3) определить процесс выполнения операций комплекса внутри каждого интервала, что можно сделать, задав параметрические зависимости $x_{is} = \Psi_{is}(\beta_s)$, где $i \in R_s$ (R_s - множество всех операций, которые могут выполняться в s-ом интервале), β_s - параметр общий для всех операций множества R_s $(0 \le \beta_s \le 1)$, Ψ_s - неубывающие непрерывные функции β_s-го, $\Psi_{is}(0) = 0$ и $\Psi_{is}(1) = y_{is}$.

П этап. Пусть для каждой операции комплекса заданы вектор параметров набора ресурсов и скорость выполнения операции при выбранном наборе ресурсов. Необходимо определить параметры α_j набора комплекса.

Замечание. В дальнейшем везде будем предполагать, что функции скоростей выполнения операций не зависят от времени и линейны, т.е. $f_i(\rho_i) = \rho_i$. (4)

Рассмотрим случай равномерного выполнения частей операций y_{is} в каждом интервале $s \in Q_i$. Условие равномерности выполнения имеет вид $f_i(\rho_{is}) = \beta_s y_{is}$, $(0 \le \beta_s \le 1)$, (5) где ρ_{is} - мощность набора i-ой операции в s-ом интервале. Из равенств (4) и (5) получаем $\rho_{is} = \beta_s y_{is}$. При заданной мощности набора комплекса ρ должно выполняться неравенство $\sum_{i \in R_s} \beta_s y_{is} \le \rho$. Тогда максимальное β_s, а значит и минимальная продолжительность интервала $\Delta_s(\rho) = 1/\rho$, определяется как функция мощности набора комплекса, т.е.

$$\Delta_s(\rho) = \frac{1}{\rho} \sum_{i \in R_s} y_{is}. \tag{6}$$

Количество ресурсов j-го вида, необходимое в s-ом интервале, равно

$$N_{js}(\rho) = \sum_{i \in R_s} \alpha_{ij} \rho_{is} = \frac{\sum_{i \in R_s} \alpha_{ij} y_{is}}{\sum_{i \in R_s} y_{is}}.$$

Количество ресурсов j-го вида, необходимое для всего комплекса, при заданной мощности его набора ρ равно максимальному из $N_{js}(\rho)$, т.е.

$$N_j(\rho) = \max_s \rho \cdot \frac{\sum_{i \in R_s} \alpha_{ij} y_{is}}{\sum_{i \in R_s} y_{is}}.$$

Так как по определению $N_j(\rho) = \alpha_j \rho$, то имеем

$$\alpha_j = \max_s \frac{\sum_{i \in R_s} \alpha_{ij} y_{is}}{\sum_{i \in R_s} y_{is}}. \tag{7}$$

<u>Ш этап</u>. Скорость выполнения комплекса при известном наборе ресурсов и известных зависимостях $\Delta_s(\rho)$, полученных на предыдущем этапе, определяется следующим образом

$$f(\rho) = \frac{W}{T(\rho)} = \frac{W}{\sum_{s=1}^{z} \Delta_s(\rho)}.$$

4. Агрегирование комплекса при линейных зависимостях скоростей выполнения его операций от мощностей наборов

Пусть все операции комплекса имеют линейные функции $f_i(\rho_i) = \rho_i$ и будем предполагать, что события сети упорядочены.

Рассмотрим случай последовательного выполнения операций $i \in R_s$ в пределах каждого интервала s. В этом случае каждый s-ый интервал разбивается на подинтервалы (по числу операций, выполняемых в s-ом интервале), удовлетворяющие условию $\sum_{i \in R_s} \Delta_{is} = \Delta_s$. Имеем $\Delta_{is} = y_{is}/\rho$ и $\Delta_s = \sum_{i \in R_s} y_{is}/\rho$. Продолжительность комплекса

$$T = \sum_{s=1}^{q} \Delta_s = \sum_{s=1}^{q} \sum_{i \in R_s} y_{is}/\rho = \sum_{i=1}^{n} \sum_{s \in Q_i} y_{is}/\rho = \frac{1}{\rho} \sum_{i=1}^{n} W_i .$$

Выражение для максимального количества ресурсов j-го вида, используемых в комплексе, принимает вид

$$N_j(\rho) = \max_s \max_{i \in R_s} \alpha_{ij} \rho = \rho \max_i \alpha_{ij} .$$

Так как по определению параметров набора $N_j(\rho) = \alpha_j \rho$, то $\alpha_j = \max_i \alpha_{ij}$.

Заметим, что каждая компонента вектора $\vec{\alpha}$ определяется, как максимум из соответствующих компонент векторов $\vec{\alpha}_i$. Следовательно, ресурсы j-го вида используются неравномерно (если соответствующие компоненты в векторах $\vec{\alpha}_i$ не равны между собой). Если в операции i занят ресурс j-го вида, то на всех подинтервалах $\Delta_{is} \in Q_i$ этот вид ресурса используется в количестве $\alpha_{ij} \rho$, в то время как при заданных α_j и ρ максимально может использоваться $\alpha_j \rho$ количество этого вида ресурсов. Значит ресурс j-го вида недоиспользуется в количестве $(\alpha_j - \alpha_{ij})\rho$ при выполнении i-ой операции в течение времени W_i/ρ. Потерями ресурсов j-го вида на операции i будем называть величину $(\alpha_j - \alpha_{ij}) W_i$. Относительные потери ресурсов j-го вида в комплексе равны

$$\delta_j = \frac{\sum_{i=1}^{n} (\alpha_j - \alpha_{ij}) W_i}{\sum_{i=1}^{n} \alpha_j \cdot W_i} = 1 - \frac{\sum_{i=1}^{n} \alpha_{ij} \cdot W_i}{\alpha_j W} . \tag{8}$$

Заметим, что параметры вектора набора ресурсов комплекса характеризуются соотношением между ресурсами различных видов. Поэтому можно принять в качестве параметров $\alpha'_{ij} = \alpha_{ij} \cdot \mu_i$, подставив в формулу скорости операции $\rho' = \mu_i \rho$. При такой подстановке (8) принимает вид

$$\delta_j = 1 - \min_i \frac{\sum_{i=1}^{n} \alpha_{ij} W_i}{\alpha_{ij}\mu_i \sum_{i=1}^{n} W_i/\mu_i}.$$

Возникает задача выбора μ_i таким образом, чтобы

$$\max_j \delta_j \to \min, \tag{9}$$

что соответствует минимизации максимальных относительных потерь ресурсов.

Очевидно, что оптимальное решение задачи (9) совпадает с оптимальным решением задачи

$$\min_j \min_i \frac{\sum_{i=1}^{n} \alpha_{ij} W_i}{\alpha_{ij}\mu_i \sum_{i=1}^{n} W_i/\mu_i} \to \max. \tag{10}$$

Заметим, что при умножении всех μ_i на одну и ту же постоянную величину (10) не изменяется. Поэтому можно принять $\sum_{i=1}^{n} W_i/\mu_i = 1$. Обозначим $\eta_i = W_i/\mu_i$ и $\sigma_{ij} = \frac{1}{\alpha_{ij} W_i} \sum_{i=1}^{n} \alpha_{ij} W_i$.
Получили задачу $\min_j \min_i \sigma_{ij}\eta_i \to \max$
при условии $\sum_{i=1}^{n} \eta_i = 1$.
Поменяв очередность взятия минимума, переходим к задаче

$\min_i \lambda_i \eta_i \to \max$ при $\sum_{i=1}^{n} \eta_i = 1$, где $\lambda_i = \min_j \sigma_{ij}$.

Нетрудно показать, что в оптимальном решении $\lambda_i \eta_i$ есть величина постоянная для всех операций. Введя обозначение $z = \lambda_i \eta_i$, получаем: $\eta_i = z/\lambda_i$; $\sum_{i=1}^{n} \eta_i = z \sum_{i=1}^{n} 1/\lambda_i = 1$;
$z = \left(\sum_{i=1}^{n} 1/\lambda_i\right)^{-1}$; $\eta_i = \left(\lambda_i \sum_{i=1}^{n} 1/\lambda_i\right)^{-1}$.
Окончательно получаем $\mu_i = W_i/\eta_i = W_i \lambda_i \sum_{i=1}^{n} 1/\lambda_i$,

т.е. $\mu_i = W_i \left(\min_j \frac{\sum_{i=1}^{n} \alpha_{ij} W_i}{\alpha_{ij} W_i}\right) \sum_{i=1}^{n} \max_j \frac{\alpha_{ij} W_i}{\sum_{i=1}^{n} \alpha_{ij} W_i}$.

Относительные потери по каждому виду ресурсов не превышают

$$\delta = 1 - z = 1 - \left(\sum_{i=1}^{n} \max_j \frac{\alpha_{ij} W_i}{\sum_{i=1}^{n} \alpha_{ij} W_i}\right)^{-1}. \tag{11}$$

Выражение для параметров набора принимает вид

$$\alpha_j = \max_i \alpha_{ij}\mu_i = \max_i \alpha_{ij} W_i \left(\min_j \frac{\sum_{i=1}^{n} \alpha_{ij} W_i}{\alpha_{ij} W_i}\right) \sum_{i=1}^{n} \max_j \frac{\alpha_{ij} W_i}{\sum_{i=1}^{n} \alpha_{ij} W_i}.$$

Продолжительность комплекса $T(\rho)=\sum_{i=1}^{n} W_i/\rho\mu_i = \frac{1}{\rho}\sum_{i=1}^{n}\eta_i = 1/\rho$.

Рассмотрим теперь случай равномерного выполнения операций $i \in R_s$ в пределах каждого интервала s.

Условие равномерности и выражения для длительности интервала и параметров набора ресурсов комплекса определяются соотношениями (5), (6) и (7). Относительные потери ресурсов j-го вида в комплексе определяются соотношением

$$\delta_j = 1 - \frac{\sum_{s=1}^{g}\sum_{i\in R_s}\alpha_{ij}y_{is}}{\sum_{s=1}^{g}\left(\max_s \frac{\sum_{i\in R_s}\alpha_{ij}y_{is}}{\sum_{i\in R_s}y_{is}}\right)\cdot\sum_{i\in R_s}y_{is}}. \qquad (12)$$

Аналогично предыдущему случаю примем в качестве параметров набора $\alpha'_{ij} = \alpha_{ij}\mu_i$, подставив в формулу скорости (4) $\rho'_i = \mu_i\rho_i$.

При такой подстановке соотношения (4), (6), (7) и (12) принимают вид:

$$f_i(\rho_i) = \rho_i\mu_i; \qquad (4а)$$

$$\Delta_s(\rho) = \frac{1}{\rho}\sum_{i\in R_s} y_{is}/\mu_i = \frac{1}{\rho}\sum_{i\in R_s}\tilde{y}_{is}; \qquad (6а)$$

$$\alpha_j = \max_s \frac{z_{js}}{\sum_{i\in R_s}\tilde{y}_{is}}; \qquad (7а)$$

$$\delta_j = 1 - \frac{A_j}{\sum_{s=1}^{g}\left(\max \frac{z_{js}}{\sum_{i\in R_s}\tilde{y}_{is}}\right)\cdot\sum_{i\in R_s}\tilde{y}_{is}}, \qquad (12а)$$

где $\tilde{y}_{is} = y_{is}/\mu_i$, $z_{js} = \sum_{i\in R_s}\alpha_{ij}y_{is}$ и $A_j = \sum_{s=1}^{g} z_{js}$.

Рассмотрим задачу выбора чисел μ_i таким образом, чтобы $\max_j \delta_j \to \min$ или

$$\min_j \min_s \frac{A_j\sum_{i\in R_s}\tilde{y}_{is}}{z_{js}\sum_{i=1}^{n} W_i/\mu_i} \to \max. \qquad (13)$$

Так как при умножении всех μ_i на одно и то же число величина (13) не изменяется, то можно принять $\sum_{i=1}^{n}\eta_i = 1$, где $\eta_i = W_i/\mu_i$.

Задача (13) эквивалентна задаче

$$\min_s \left(\min_j \frac{A_i}{z_{js}}\right) \cdot \sum_{i \in R_s} y_{is} \eta_i / W_i \rightarrow max. \qquad (14)$$

Обозначим $B_s = \min_j \frac{A_i}{z_{js}}$ и $\sigma_{is} = y_{is}/W_i$ и запишем (14) в виде

$$\min_s B_s \sum_{i \in R_s} \sigma_{is} \eta_i \longrightarrow max.$$

Пусть $C_{is} = B_s \sigma_{is}$, тогда задача принимает вид

$$\min_s \sum_{i \in R_s} C_{is} \eta_i \longrightarrow max \quad \text{при} \quad \sum_{i=1}^{n} \eta_i = 1$$

Обозначим $\min_s \sum_{i \in R_s} C_{is} \eta_i = \nu$. Получим задачу максимизации $\Phi = \nu$ при линейных ограничениях

$$\begin{cases} \sum_{i \in R_s} C_{is} \eta_i \geq \nu, \\ \sum_{i \in 1}^{n} \eta_i = 1. \end{cases} \qquad (15)$$

Ограничения (15) эквивалентны ограничениям

$$\nu - \sum_{i \in R_s} C_{is} \eta_i \leq 0 \quad и \quad \sum_{i=1}^{n} \eta_i \geq 1. \qquad (16)$$

Перейдем к двойственной задаче. Вводим переменные ε_1 , ε_2 , .., ε_g (по числу интервалов) и переменную ξ , соответствующую последнему ограничению. Получаем задачу минимизации ξ при ограничениях

$$-\sum_{s \in Q_i} C_{is} \varepsilon_s + \xi \geq 0 \qquad (17)$$

и

$$\sum_{s=1}^{n} \varepsilon_s \geq 1. \qquad (18)$$

Ограничение (17) можно заменить равенством

$$\xi = \max_i \sum_{i \in Q_i} C_{is} \varepsilon_s,$$

а ограничение (18) равенством

$$\sum_{s=1}^{g} \varepsilon_s = 1. \qquad (18a)$$

Получаем задачу двойственную первоначальной

$$\max_i \sum_{s \in Q_i} C_{is} \varepsilon_s \longrightarrow \min \qquad \text{при} \quad \sum_{s=1}^{g} \varepsilon_s = 1.$$

Выберем в качестве базисных переменных ν и η_i тех операций, которые могут выполняться только в одном интервале. Обозначим множество таких операций H (заметим, что $|H| - g$).

Базисное решение определяется как решение следующей задачи

$$\min_{i \in H} C_{is} \eta_i \longrightarrow \max \qquad \text{при} \quad \sum_{i \in H} \eta_i = 1.$$

Легко показать, что в оптимальном решении все $C_{is} \eta_i$ равны между собой. Отсюда получаем базисное решение

$$\eta_i = \begin{cases} \left(C_{is} \sum_{i \in H} 1/C_{is}\right)^{-1} & \text{для } i \in H, \\ 0 & \text{для } i \notin H. \end{cases} \tag{I9}$$

При этом $\nu = \left(\sum_{i \in H} 1/C_{is}\right)^{-1}.$

Докажем, что решение (I9) оптимально. Для этого определим соответствующее решение двойственной задачи. Из соотношения двойственности следует, что для операций, принадлежащих множеству H, в (I7) должно выполняться строгое равенство. Учитывая также, что каждая операция множества H может выполняться только в одном интервале, получаем

$$C_{is} \varepsilon_s = \xi \qquad \text{для} \quad i \in H$$

Отсюда $\varepsilon_s = \xi / C_{is}.$ (20)

Подставляя (20) в (I8а), находим

$$\xi = \left(\sum_{s=1}^{g} 1/C_{is}\right)^{-1} = \left(\sum_{i \in H} 1/C_{is}\right)^{-1} = \nu \tag{2I}$$

Поскольку решения прямой и двойственной задач совпадают, то полученное допустимое решение (I9) является оптимальным.

Для дальнейшего обозначим через η_s значение η_i операции, имеющей событие ($s - 1$) начальным и s конечным. Соответственно через C_s значение C_{is} для этой операции. Имеем

$$C_s = \min_j \frac{\sum_{s=1}^{g}\sum_{i\in R_s} \alpha_{ij} y_{is}}{\sum_{i\in R_s}\alpha_{ij} y_{is}} = \min_j \frac{\sum_{i=1}^{n}\alpha_{ij} W_i}{\sum_{i\in R_s}\alpha_{ij} y_{is}};$$

$$\eta_s = \left(C_s \sum_{s=1}^{g} 1/C_s\right)^{-1} = \left(\max_j \frac{\sum_{i\in R_s}\alpha_{ij} y_{is}}{\sum_{i=1}^{n}\alpha_{ij} W_i}\right)\cdot\left(\sum_{s=1}^{g}\max_j \frac{\sum_{i\in R_s}\alpha_{ij} y_{is}}{\sum_{i=1}^{n}\alpha_{ij} W_i}\right)^{-1};$$

$$d_j = \left\{\max_s\left(\max_j \frac{\sum_{i\in R_s}\alpha_{ij} y_{is}}{\sum_{i=1}^{n}\alpha_{ij} W_i}\right)^{-1}\cdot\sum_{i\in R_s}\alpha_{ij} y_{is}\right\}\cdot\sum_{s=1}^{g}\max_j \frac{\sum_{i\in R_s}\alpha_{ij} y_{is}}{\sum_{i=1}^{n}\alpha_{ij} W_i};$$

$$T(\rho) = \frac{1}{\rho}\sum_{s=1}^{g}\sum_{i\in R_s} y_{is}\eta_i/W_i = \frac{1}{\rho}\sum_{s=1}^{g}\eta_s = \frac{1}{\rho}.$$

Приняв объем всего комплекса $W = 1$, получаем $f(\rho) = \rho$. Относительные потери по любому виду ресурсов не превышают

$$\delta = 1 - \nu = 1 - \left(\sum_{s=1}^{g}\max_j \frac{\sum_{i\in R_s}\alpha_{ij} y_{is}}{\sum_{i=1}^{n}\alpha_{ij} W_i}\right)^{-1}. \qquad (22)$$

Полученные результаты позволяют поставить задачу агрегирования в случае линейных зависимостей $f_i(\rho_i)$, т.е. задачу определения оптимальных значений y_{is}: определить y_{is}, удовлетворяющее условию $\sum_{s\in Q_i} y_{is} = W_i$ так, чтобы

$$\sum_{s=1}^{g}\max_j \frac{1}{D_j}\sum_{i\in R_s}\alpha_{ij} y_{is} \longrightarrow \min,$$

где $D_j = \sum_{i=1}^{n}\alpha_{ij} W_i$.

Обозначим $\alpha_{ij} W_i/D_j = \tilde{\alpha}_{ij}$, $y_{is}/W_i = \hat{y}_{is}$ и $\max_j \sum_{i\in R_s}\tilde{\alpha}_{ij}\hat{y}_{is} = \gamma_s$.

Получаем следующую задачу линейного программирования: минимизировать $\sum_{s=1}^{g}\gamma_s$ при ограничениях

$$\begin{cases} \gamma_s - \sum_{i\in R_s}\tilde{\alpha}_{ij}\hat{y}_{is} \geq 0, \\ \sum_{s\in Q_i}\hat{y}_{is} \geq 1. \end{cases}$$

В заключение сравним относительные потери при последовательном и равномерном выполнении операций в каждом интервале. Для этого рассмотрим допустимое решение для случая равномерного выполнения операций: $y_{is} = W_i$, если ($s-1$) есть начальное событие i-ой операции, и $y_{is}=0$ в противном случае. Формула относительных потерь примет вид

$$\delta_{равн} = 1 - \left(\sum_{s=1}^{g} \max_{j} \frac{\sum_{i \in R_s^*} d_{ij} W_i}{\sum_{i=1}^{n} d_{ij} W_i} \right)^{-1}, \quad \text{где } R_s^* = \{ i : y_{is} > 0 \} \tag{23}$$

Формулу относительных потерь для последовательного выполнения можно записать в виде

$$\delta_{посл} = 1 - \left(\sum_{s=1}^{g} \sum_{i \in R_s^*} \max_{j} \frac{d_{ij} W_i}{\sum_{i=1}^{n} d_{ij} W_i} \right)^{-1}. \tag{24}$$

Поскольку максимум суммы всегда больше суммы максимумов, то $\delta_{равн} \leq \delta_{посл}$. Так как $\delta_{равн}$ при оптимальном выборе y_{is} не больше (23), а значит и (24), то случай равномерного выполнения операций не хуже случая последовательного выполнения в смысле величины относительных потерь ресурсов (при оптимальных значениях y_{is} в случае равномерного выполнения).

ЛИТЕРАТУРА

I. Бурков В.Н., Лернер А.Я. Новые задачи теории сетевого планирования и управления. В сб. "Вопросы управления большими системами". Изд-во "Онтиприбор", I967.

2. Berman E.B. Resource Allocation in a PERT Network under Continuous Activity Time-cost Functions. Management Science, Vol. I0, No. 4 (July I964), 734-745.

3. Бурков В.Н. Распределение ресурсов как задачи оптимального быстродействия. Автоматика и телемеханика, № 7, I966.

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ КОМПЛЕКСАМИ ОПЕРАЦИЙ

В.Н.Бурков
Институт автоматики и телемеханики
/технической кибернетики/
г.Москва
СССР

§ I. Постановка задач.

Достаточно строгая постановка и классификация задач оптимального распределения ресурсов в комплексах операций была проведена сравнительно недавно[I]. К настоящему времени задачи распределения ресурсов составляют одно из главных направлений теории сетевого планирования и управления (теории СПУ).

В докладе рассматриваются задачи распределения ресурсов одного вида с критерием минимума времени выполнения комплекса.

Введем некоторые обозначения:

$w_i(t)$ - скорость i-ой операции,
$v_i(t)$ - количество ресурсов на i-ой операции,
W_i - объем i-ой операции,
$N(t)$ - количество ресурсов в момент t (заданная функция),
$s_i(t)$ - затраты в i-ой операции к моменту t,
$S(t)$ - затраты на выполнение комплекса к моменту t (заданная функция),

Введенные понятия связаны следующими соотношениями:

$$w_i(t) = f_i[v_i(t)], \qquad (1)$$

$$W_i = \int_0^T f_i[v_i(t)]\,dt, \qquad (2)$$

$$\sum_1^n v_i(t) \le N(t), \qquad (3)$$

$$s_i(t) = \int_0^t v_i(\tau)\,d\tau, \qquad (4)$$

$$\sum_1^n s_i(t) \le S(t), \qquad (5)$$

$$v_i(t) \ge 0,\ W_i \ge 0,\ N(t) \ge 0,\ S(t) \ge 0, \quad i = 1, 2, \dots, n,\ t \in (0, T). \qquad (6)$$

В (I)-(6) f_i - неубывающая функция $v_i \geq 0$, такая что $f_i(0) = 0$, T - время выполнения комплекса.

Задача I. Определить $v_i(t)$, удовлетворяющие (3), так чтобы комплекс был выполнен за минимальное время.

Задача 2. Определить $v_i'(t)$, удовлетворяющие (5), так чтобы комплекс был выполнен за минимальное время.

Задача 3. Определить $s_i(T)$, удовлетворяющие ограничению $\sum_1^n s_i(T) \leq S(T)$, так чтобы время выполнения комплекса было минимальным.

В § 2 решение задачи I сводится к задаче минимизации выпуклого функционала при линейных ограничениях. Особо рассмотрен случай степенных зависимостей $f_i(v_i)$, для которого предложен специальный алгоритм, использующий аналогию между сетью и фазовым пространством. В § 4 доказана связь задачи I и задачи 3, позволяющая находить оптимальное решение одной задачи, зная оптимальное решение другой. Наконец, в § 5 дается решение задачи 2 для случая степенных функций $f_i(v_i)$.

§ 2. Решение задачи I.

Пусть $N(t) = N = const$. Предположим, что события сети упорядочены, то есть момент наступления s-го события меньше момента наступления k-го события, если $s < k$. Обозначим R_s - множество операций, которые могут выполняться в s-ом интервале, то есть в интервале между моментами наступления ($s-1$)-го и (s)-го событий ($s = 1, 2, \cdots, m$), Q_i -множество интервалов, в которых может выполняться i-я операция, и, наконец, x_{is} - объем i-ой операции, выполняемый в s-ом интервале. Очевидно x_{is} удовлетворяют условию

$$\sum_{s \in Q_i} x_{is} = W_i \quad , i = 1, 2, \cdots, n. \qquad (7)$$

При известных x_{is}, $i \in R_s$, $s = 1, \cdots, m$, для каждого интервала получаем задачу минимизации его длительности, причем части операций внутри одного интервала независимы[3]. Обозначим

Обозначим $\Delta_s(\{x_{is}\})$ минимальную длительность s-го интервала как функцию x_{is}. Можно показать, что $\Delta_s(\{x_{is}\})$ выпуклая (книзу) функция своих аргументов (даже при невыпуклых функциях $f_i(v_i)$). Но тогда время выполнения комплекса

$$T(\{x_{is}\}) = \sum_1^m \Delta_s(\{x_{is}\}) \qquad (8)$$

также выпуклая (книзу) функция переменных x_{is}. Получили задачу минимизации выпуклого функционала (8) при линейных ограничениях (7). Для ее решения можно применить любые известные методы выпуклого программирования.

Если в оптимальном решении задачи имеется $\Delta_s = 0$, то можно попытаться улучшить решение, изменив нумерацию событий ($s-1$) и (s) и решив задачу при новом упорядочении.

Однако, полученное в конце концов решение в общем случае не будет оптимальным решением задачи без требования упорядочения событий.

§ 3. Случай степенных зависимостей.

Пусть $f_i(v_i) = v_i^{1/\alpha}$, $\alpha > 1$, $i = 1, 2, \ldots, n$. Тогда[3]

$$\Delta_s(\{x_{is}\}) = N^{-1/\alpha}\left[\sum_{i\in R_s} x_{is}^{\alpha}\right]^{1/\alpha} \qquad (9)$$

Примем для упрощения выводов $N=1$. Применив метод множителей Лагранжа, получаем

$$\frac{\partial \Delta_s(\{x_{is}\})}{\partial x_{is}} = \left(\frac{x_{is}}{\Delta_s}\right)^{\alpha-1} = \lambda_i \qquad (10)$$

Обозначим v_{is} – количество ресурсов, выполняющих i-ю операцию в s-ом интервале. Из (10) имеем $v_{is} = \lambda_i^{\alpha/\alpha-1}$, то есть v_{is} не зависит от s. Получили важное свойство оптимального решения:

<u>каждая операция выполняется постоянным количеством ресурсов (свойство 1).</u>

Для вывода второго свойства предположим, что в оптимальном решении все $\Delta_s(\{x_{is}\}) > 0$. Отсюда сразу следует свойство:

<u>ресурсы v_i образуют поток величины N по сети (свойство 2).</u>

Это свойство справедливо и при наличии $\Delta_s(\{x_{is}\}) = 0$, что легко доказать, применив предельный переход.

Доказанные свойства справедливы при любом упорядочении событий комплекса. Значит, они выполняются и для оптимального решения задачи I уже без требования упорядочения событий. В дальнейших выводах мы уже не будем требовать упорядочения событий. В дальнейших выводах мы уже не будем требовать упорядочения событий. Заметим также, что процедура изменения нумерации событий ($s-1$) и (s) в случае когда $\Delta_s = 0$ и последующее решение задачи при новом упорядочении приводит в случае степенных зависимостей к оптимальному решению задачи I уже без требования упорядочения событий.

Для дальнейших выводов рассмотрим изображение сети в q - мерном пространстве состояний комплекса[2]. Определим расстояние между любыми двумя точками $y^1, y^2 \in G$

$$\rho(y^1, y^2) = \left(\sum_1^q |y_j^1 - y_j^2|^\alpha \right)^{1/\alpha} \qquad (\text{II})$$

Сравнивая (9) и (II), замечаем, что любому допустимому набору $\{x_{is}\}$ можно поставить в соответствие траекторию в G, соединяющую начальную и конечную точки, причем длина этой траектории, вычисленная по (II) совпадает с (9) при N = I. Таким образом:

<u>оптимальному решению задачи I соответствует траектория в G, соединяющая начальную и конечную точки и имеющая минимальную длину (свойство 3)</u>.

Обозначим длину кратчайшей траектории $W_э$ и назовем ее эквивалентным объемом комплекса. Действительно, заменив комплекс одной операцией объема $W_э$ и зависимостью скорости от ресурсов $w = N^{1/\alpha}$, получим

$$T_{min} = W_э N^{-1/\alpha} \qquad (\text{I2})$$

Из (I2) следует, что движение в фазовой области допустимых состояний G происходит со скоростью, абсолютная величина которой равна $N^{1/\alpha}$. Поскольку кратчайшая траектория остается кратчайшей при любой скорости движения, то значения x_{is} в оптимальном решении <u>не зависят от $N(t)$ (свойство 4)</u>.

При этом минимальное время выполнения комплекса определяется из уравнения

$$\int_0^T N^{1/\alpha}(t)\,dt = W_э \qquad (13)$$

Опишем алгоритм определения $W_э$. Для этого изобразим сеть в системе координат, оси которой параллельны. Число осей равно размерности комплекса. В параллельной системе координат каждая точка изображается линией (фронтом), проходящей через ее координаты. Изображение комплекса в такой системе является промежуточным между сетью и фазовой областью допустимых состояний. На рис. 1 представлена сеть (рис. 1а), ее представление в параллельной системе координат (рис. 1б) и соответствующая область G (рис. 1в). На рис. 1б $F_0 = (0,0,0)$ - начальное состояние комплекса, $F_к = (y_1^к, y_2^к, y_3^к)$ - конечное состояние комплекса. Основная идея алгоритма заключается в следующем. Так как нас интересует кратчайшая траектория, соединяющая точку 0 с А, то сначала пытаемя провести прямую 0А. Этой прямой в параллельной системе координат соответствует последовательность фронтов, описываемая параметрическими уравнениями

$$y_j(t) = t\,y_j^к,\ 0 \le t \le 1,\ j = 1,2,\dots,q. \qquad (14)$$

Если такая прямая выходит за пределы области G (см. рис. 1в), то из геометрических соображений получаем, что кратчайшая траектория проходит по границе ВС через т. Д, причем так, что

$$\frac{ДС}{ДВ} = \frac{ОД}{ДА} \qquad (15)$$

В параллельной системе координат точке Д соответствует фронт $F_1 = (y_1^1, y_2^1, y_3^1)$, (см. рис, 1б), где

$$y_3^1 = \frac{W_5\left(W_1^\alpha + W_2^\alpha\right)^{1/\alpha}}{\left(W_1^\alpha + W_2^\alpha\right)^{1/\alpha} + \left(W_3^\alpha + W_4^\alpha\right)^{1/\alpha}} \qquad (16)$$

Заметим, что (16) эквивалентно (15).

Повторяя описанную процедуру, определяем в сети ряд "опорных фронтов", таких что траектория движения между этими фронтами является прямой линией и не выходит за пределы области G. Полученное решение можно последовательно улучшать, уточняя положение опорного фронта, расположенного между двумя другими.

§ 4. Связь задач I и 3.

Учитывая, что каждая операция в оптимальном решении выполняется постоянным количеством ресурсов v_i, можно определить затраты ресурсов

$$s_i(\tau_i) = v_i \cdot \tau_i = \tau_i f_i^{-1}\left(\frac{W_i}{\tau_i}\right) = \frac{W_i^{\alpha}}{\tau_i^{\alpha-1}} \qquad (17)$$

Таким образом затраты также являются степенной выпуклой (книзу) функцией времени τ_i выполнения операции.

Рассмотрим задачу минимизации затрат $S = \sum_1^n s_i(\tau_i)$ при заданном времени T выполнения комплекса. Как известно[5], необходимым и достаточным условием оптимальноси является выполнения для любого события p сети соотношения

$$\sum_{i \in U_p^+} \frac{d s_i(\tau_i)}{d \tau_i} = \sum_{i \in U_p^-} \frac{d s_i(\tau_i)}{d \tau_i} , \qquad (18)$$

где U_p^+ - множество операций, для которых событие p является конечным, U_p^- - множество операций, для которых событие p является начальным. Из (17) и (18) получаем

$$\sum_{i \in U_p^+} v_i = \sum_{i \in U_p^-} v_i \qquad (19)$$

то есть свойство 2 оптимального решения задачи I.

Решим задачу минимизации затрат при некотором значении T_0 времени выполнения комплекса. Пусть S_0 - минимальная величина затрат, τ_i^0 и s_i^0 соответственно время и затраты i -ой операции в оптимальном решении. Тогда эквивалентный объем комплекса $W_э$, минимальное время T выполнения комплекса при заданном уровне N, а также значения v_i и τ_i в оптимальном решении определяются из соотношений:

$$W_э = \left(\frac{S_0}{T_0}\right)^{1/\alpha} \cdot T_0 , \quad T = T_0 \left(\frac{S_0}{N T_0}\right)^{1/\alpha} ,$$
$$v_i = \frac{s_i^0}{\tau_i^0} \cdot \frac{N T_0}{S_0} , \quad \tau_i = \tau_i^0 \left(\frac{S_0}{N T_0}\right)^{1/\alpha} \qquad (20)$$

Таким образом, зная оптимальное решение задачи I можно определить оптимальное решение задачи 3 и наоборот.

Связь между задачей минимизации затрат и задачей минимизации уровня ресурсов можно обобщить на случай произвольных зависимостей $f_i(v_i)$, если искать решение, удовлетворяющее свойствам I и 2. В этом случае оптимальные значения времени τ_i выполнения операций в задаче минимизации уровня ресурсов N при заданном времени выполнения комплекса, совпадают с оптимальными значениями времен выполнения операций в задаче минимизации затрат S при том же времени T выполнения комплекса, если

$$S_i(\tau_i) = \int_{\tau_i}^{\infty} f_i^{-1}\left[\frac{W_i}{\tau}\right] d\tau \qquad (21)$$

Доказательство сразу следует из того, что свойство 2 эквивалентно условиям (I8), если выполняется (2I).

Пример. Пусть $s_i(\tau_i) = a_i - b_i \tau_i$,

$d_i \le \tau_i \le D_i$, $i = 1, 2, \ldots, n$.

Из (2I) получаем

$$\tau_i(v_i) = \begin{cases} D_i, & \text{при } v_i < b_i, \\ d_i, & \text{при } v_i > b_i, \end{cases}$$

любое от d_i до D_i при $v_i = b_i$.

Получили задачу определения потока v_i, имеющего минимальную величину N, такого что сумма времен $\tau_i(v_i)$ для любого пути, соединяющего начальное и конечное события сети, не превышает T.

§ 5. Решение задачи 2.

Задачу 2 рассмотрим для случая степенных зависимостей скорости операций от количества ресурсов. Пусть $W_э$ - эквивалентный объем комплекса. Тогда скорость выполнения комплекса $w(t) = N^{1/\alpha}(t)$, а ограничение (5) принимает вид

$$\int_0^t N(\tau)\, d\tau \le S(t) \qquad (22)$$

Задача заключается в определении $N(\tau)$, удовлетворяющего (22), так чтобы время выполнения комплекса, определяемое из (13) было минимальным. В основе алгоритма решения задачи лежит процесс построения для фиксированного T огибающей $\hat{S}_T(t)$ для кривой $S(t)$ па $[0,T]$, причем $\hat{S}_T(T)=S(T)$ (рис. 4).

I шаг. Определяем T_0 из уравнения

$$T_0\left[S(T_0)\cdot T_0^{-1}\right]^{1/\alpha}=W_э .$$

Если прямая $S(T_0)T_0^{-1}t$ на $(0,T_0)$ проходит не выше кривой $S(t)$, то T_0 - минимальное время. Если же существуют интервалы $\varepsilon\in(0,T_0)$, такие что $S(T_0)T_0^{-1}t >$ $> S(t)$ при $t\in\varepsilon$, то определяем минимальное Δ, такое что

$$S(t)\geqslant S(T_0)T_0^{-1}(t-\Delta) \quad \text{при} \quad t\in(0,T_0)$$

Для этого находим $t(\Delta)$, при котором разность $S(T)-S(T_0)T_0^{-1}(t-\Delta)$ минимальна. Значение Δ находится из уравнения $t(\Delta)=0$. Определяем $T_1=T_0+\Delta$.

II шаг. Строим $\hat{S}_{T_1}(t)$ на $[0,T_1]$. Вычисляем

$$W(T_1)=\int_0^{T_1}\left[\frac{d\hat{S}_{T_1}(t)}{dt}\right]^{1/\alpha}dt$$

Если $W(T_1)=W_э$, то T_1 - минимальное время.
Если $W(T_1)>W_э$, то определяем

$$T_2=0,5\,(T_0+T_1)$$

i -ый шаг. Строим $\hat{S}_{T_{i-1}}(t)$ на $[0,T_{i-1}]$. Вычисляем

$$W(T_{i-1})=\int_0^{T_{i-1}}\left[\frac{d\hat{S}_{T_{i-1}}(t)}{dt}\right]^{1/\alpha}dt$$

Если $W(T_{i-1})=W_э$, то T_{i-1} - минимальное время.
Если $W(T_{i-1})>W_э$, то определяем

$$T_i=T_{i-1}-0,5\,|T_{i-1}-T_{i-2}|$$

Если $W(T_{i-1})<W_э$, то определяем

$$T_i=T_{i-1}+0,5\,|T_{i-1}-T_{i-2}|$$

и переходим к следующему шагу.

Описанный метод деления отрезка пополам как известно обладает достаточно быстрой сходимостью. В случае выпуклых $S(t)$ задача решается более просто.

Если $S(t)$ выпуклая (кверху), то T_{min} определяется из уравнения

$$\left[\frac{S(T_{min})}{T_{min}}\right]^{1/\alpha}T_{min}=W_э$$

Если $S(t)$ выпуклая (книзу), то T_{min} определяется из уравнения

$$\int_0^{T_{min}} \left[\frac{dS(t)}{dt}\right]^{1/\alpha} dt = W_э$$

§ 6. Анализ свойства I.

Интересно рассмотреть вопрос, для какого класса зависимостей $f_i(v_i)$ в оптимальном решении задачи I выполняется свойство I. Условия оптимальности для произвольных (выпуклых кверху) функций $f_i(v_i)$ примут вид:

$$\frac{1}{\lambda_i} \frac{d\varphi_i(w_{is})}{dw_{is}} = \sum_{j\in R_s} w_{js} \frac{d\varphi_j(w_{js})}{dw_{js}}$$

где $\varphi_i = f_i^{-1}$ и $w_{is} = \frac{x_{is}}{\Delta_s}$.

Для того, чтобы $w_{is} = w_i$ не зависело от s необходимо и достаточно

$$\sum_{i\in v_s^-} w_i \frac{d\varphi_i(w_i)}{dw_i} = \sum_{i\in v_s^+} w_i \frac{d\varphi_i(w_i)}{dw_i} \qquad (23)$$

Обозначим $F_i(v_i) = w_i \frac{d\varphi_i(w_i)}{dw_i} = f_i(v_i)\left(\frac{df_i(v_i)}{dv_i}\right)^{-1}$.

Из (23) следует, что $F_i(v_i)$ является потоком по сети.

<u>Теорема</u>. Для того, чтобы при любом потоке v_i, $F_i(v_i)$ было потоком, необходимо и достаточно $F_i(v_i) = a_i + \alpha v_i$, где a_i - поток. Предполагаем, что сеть остается связной при удалении входа и выхода.

<u>Доказательство</u>. Варьируя v_i вдоль любого пути в сети, получаем

$$\frac{dF_i(v_i)}{dv_i} = \frac{dF_j(v_j)}{dv_j} \quad \text{для любых } i, j \quad \text{и} \quad \frac{d^2F_i(v_i)}{dv_i^2} = 0$$

для любого i. Отсюда получаем

$$F_i(v_i) = a_i + \alpha v_i .$$

a_i должно быть потоком, так как при $v_i = 0$, $F_i = a_i$.

Теорема доказана.

Из теоремы следует, что

$$f_i(v_i) = c_i (a_i + \alpha v_i)^{1/\alpha}, \quad c_i > 0, \quad i = 1,2,\dots,n.$$

Из условия выпуклости $f_i(v_i)$ при $a_i + \alpha v_i \geq 0$ следует, что $\alpha \geq 1$.

Пусть $W_э$ - эквивалентный объем комплекса и пусть $\tilde{x}_{is}$ - значения x_{is} , определяющие $W_э$.

Тогда, если

$$\frac{\tilde{x}_{is}^{\alpha}}{\sum\limits_{i \in R_s} \tilde{x}_{is}^{\alpha}} \geq \frac{a_i}{\alpha N + A}, \quad i \in R_s, \quad s = 1, 2, \dots, m,$$

где A - величина потока a_i , то минимальное время выполнения комплекса $T_{min} = W_э \, (\alpha N + A)^{-1/\alpha}$.

ЛИТЕРАТУРА

I. В.Н.Бурков, А.Я.Лернер Новые задачи теории сетевого планирования и управления. В сб. "Вопросы управления в больших системах". Изд. "Онтиприбор", 1968.

2. В.Н.Бурков Применение теории оптимального управления к задачам распределения ресурсов. Труды III Всесоюзного совещания по автоматическому управлению (Одесса, 1965г.) т. Управление производством. М., Наука, 1967.

3. В.Н.Бурков Распределение ресурсов как задача оптимального быстродействия, Автоматика и телемеханика т. XXУII, № 7,1966.

4. Б.С.Разумихин Задачи об оптимальном распределении ресурсов, Автоматика и телемеханика № I, 1967.

5. E.B.Berman Resource allocation in a PERT network under continuous activity time-cost functions. Manag.Sci., v.IO, N 4, 1964.

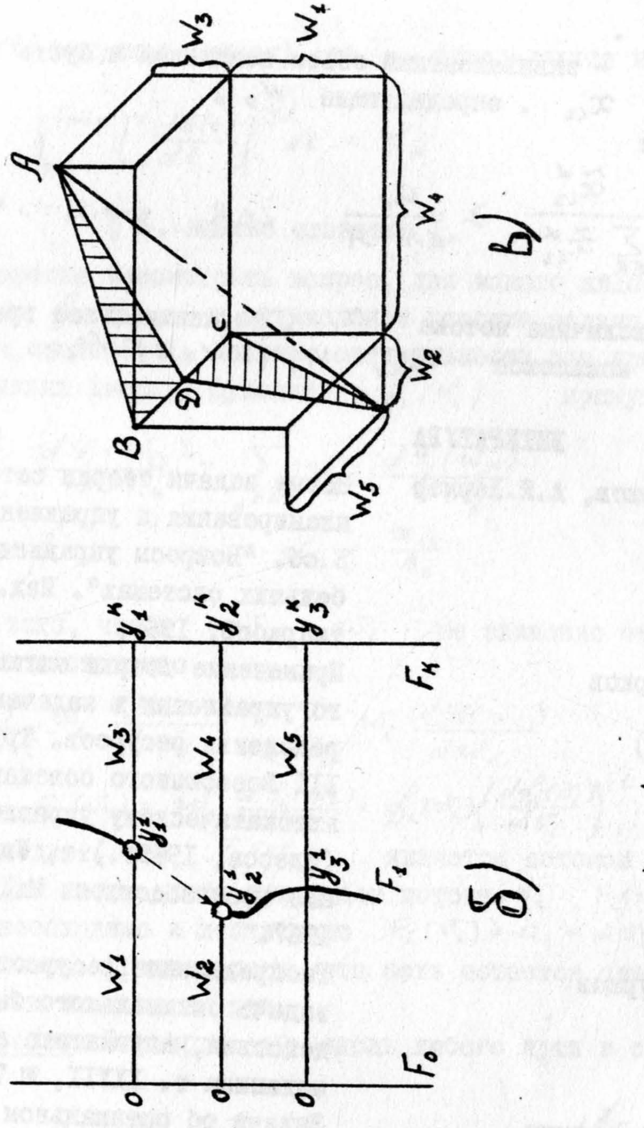

Рис 1

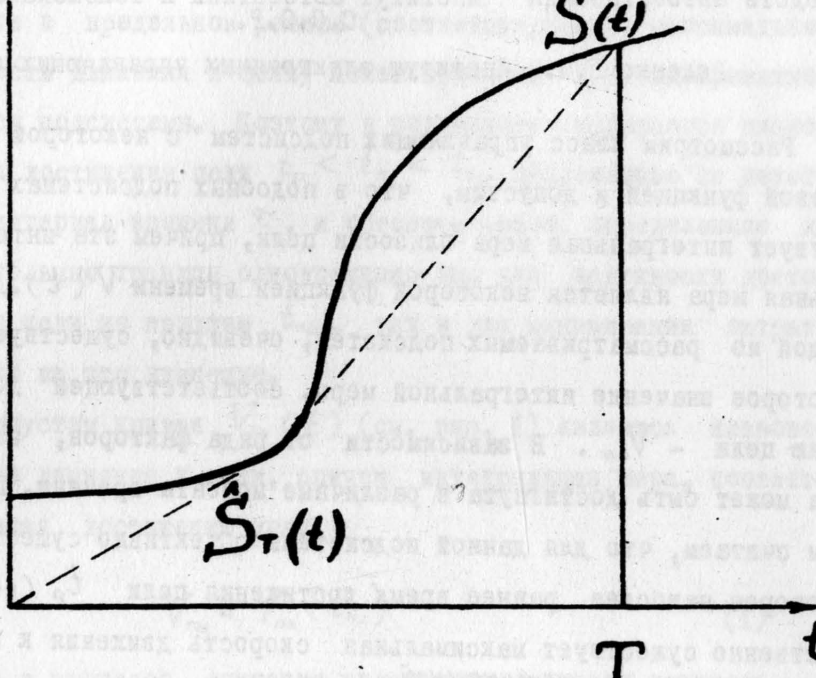

Рис. 2

НЕКОТОРЫЕ ВОПРОСЫ КОНТРОЛЯ И ПРИНЦИПЫ ПОСТРОЕНИЯ ОПТИМАЛЬНОЙ ИЕРАХИЧЕСКОЙ СТРУКТУРЫ УПРАВЛЕНИЯ В СИСТЕМАХ С ОПРЕДЕЛЕННОЙ ЦЕЛЕВОЙ ФУНКЦИЕЙ

Тбилисский научно-исследовательский институт приборостроения и средств автоматизации

Бабунашвили М.К. (Тбилиси)

Наумов С.С. (Москва)
институт автоматики и телемеханики С.С.С.Р.

Голенко Д.И. Институт электронных управляющих машин

Рассмотрим класс управляющих подсистем с некоторой целевой функцией и допустим, что в подобных подсистемах существует интегральная мера близости цели, причем эта интегральная мера является некоторой функцией времени $V(t)$. Для каждой из рассматриваемых подсистем, очевидно, существует некоторое значение интегральной меры, соответствующей достижению цели - $V_{пл}$. В зависимости от ряда факторов, эта мера может быть достигнута в различные моменты времени. При этом считаем, что для данной подсистемы объективно существует некоторое наиболее раннее время достижения цели t_p (соответственно существует максимальная скорость движения к цели), которая соответствует максимальной мобилизации внутренних ресурсов подсистемы и минимальным помехам извне, причем это время t_p вообще говоря отличается от оптимистического времени t_o, используемого в системах PERT. В это же время существует некоторое конечное предельно позднее время достижения цели $t_{кр}$*. При достижении цели позднее $t_{кр}$ подсистема платит определенный штраф. Следовательно, учитывая указанные выше

---

* Системы для которых $t_{кр} = \infty$, очевидно, не относятся к рассматриваемому классу систем.

ограничения, подсистема должна достигать цели в интервале времени $[t_p, t_{кр}]$.

Необходимо отметить, что во время движения к определенной цели, подсистема стремится минимизировать усилия, направленные на достижение этой цели, ибо многократная и длительная работа в предельном режиме (соответствующего максимальной скорости движения к цели) может привести к преждевременному износу подсистемы. Поэтому в подсистеме выбирается плановое время достижения цели $t_p < t_{пл} < t_{кр}$, опережающее $t_{кр}$ на некоторый интервал времени τ, и обеспечивающее определенные доверительные границы одновременно как для надежности достижения цели не позднее $t_{кр}$, так и для минимизации затрат усилий на это движение.

Допустим кривая $V_{пл}(t)$ (см. рис. I) является плановой кривой движения к цели, причем интегральная мера, соответствующая достижению цели

$$V_{пл} = V_{пл}(t_{пл}). \qquad (I)$$

Тогда в процессе движения подсистемы, с целью контроля, необходимо сравнивать истинную меру $V(t_i)$, являющуюся случайной величиной, со значениями плановой меры близости цели $V_{пл}(t_i)$ в определенные моменты времени t_i, которые должны определяться выбором шага квантования.

Функцией регулирования данной подсистемы является устранение рассогласования между фактической и плановой мерами близости цели, появляющихся из-за воздействия на подсисиему возмущений и тем самым обеспечение достижения подсистемой цели в заданный срок $t_{пл}$.

Прежде чем перейти к рассмотрению алгоритма, предварительно кривую $V_p(t)$ – соответствующую движению к цели за время t_p, переместим в направлении оси абсцисс таким образом, чтобы конец кривой $V_p(t)$ совпал с точкой $(V_{пл}, t_{пл})$ (рис. I). В этом случае на пересечении от абсцисс с $V_p(t)$ получим точку t_I. Нетрудно видеть, что даже если до момента времени t_I подсистема не движется к цели $(V(t_I)= 0)$, то, очевидно, существует вероятность, отличная от нуля, что начиная с момента времени t_I и используя предельные возможности подсистемы, мы достигаем цели к моменту времени $t_{пл}$. t_I можно считать предельно допустимым временем первого опроса. Если опрос будет проведен позже момента времени t_I, то, поскольку наивысшая скорость движения к цели определяется кривой $V_p(t)$, срок $t_{пл}$ уже не может быть обеспечен.

Таким образом, для обеспечения срока $t_{пл}$, первый опрос подсистемы должен быть проведен в срок, определенный промежутком $t_н < t \leq t_I$. Приняв предельное значение $t = t_I$ и проведя опрос, подсистема получает сведения о близости цели путем сравнения $V(t_I)$ и $V_{пл}(t_I)$. На основе полученной информации должны вырабатываться эффективные управляющие воздействия по ликвидации имеющегося рассогласования. После этого проведем через точку $(t_1, V(t_1))$, прямую, параллельную от абсцисс, до пересечения со смещенной кривой $V_p(t)$, получим точку с абсциссой t_2, определяющей, по ранее проведенным соображениям, предельное значение срока второго опроса.

Последующие моменты опроса определяются аналогичным образом.

Нами уже указывалось выше, что истинная кривая движения к цели $V(t)$ является случайной кривой. Поэтому в зависимости от характера ее изменения один и тот же объем $V_{пл}$ может быть достигнут за некоторое фактическое время $t_{ф}$. Здесь могут встретиться три случая.

I. Случай $t_{пл} < t_{ф} < \infty$ может произойти: а) когда подсистема неуправляема; очевидно, для данной подсистемы опрос вообще не имеет смысла; б) когда, несмотря на то, что подсистема управляема, условие $t_{ф} \leq t_{пл}$ все же не обеспечивается. В этом случае, на некотором шаге опроса, мы окажемся на смещенной кривой $V_{p}(t)$. Поскольку возможность движения строго по кривой $V_{p}(t)$ невелико, задача управляющего органа в этом случае состоит в том, чтобы обеспечить наименьшую задержку относительно срока $t_{пл}$.

2. В случае $t_{н} < t_{ф} < t_{пл}$ очевидно требуется конечное число шагов опроса для контроля движения к цели.

3. В случае $t_{ф} = t_{пл}$, последовательность всех точек имеет предел сходимости - $t_{пл}$, и процесс приближения происходит за бесконечное число шагов. Но так как на практике требуется выполнение объема $V_{пл}$ к моменту времени $t_{пл}$ с какой-то заранее заданной точностью $t_{пл} \pm \Delta t_{пл}$; поэтому в этом случае задача состоит в попадании в область $[t_{пл} - \Delta t_{пл}, t_{пл} + \Delta t_{пл}]$, которое, в отличие от сходимости к $t_{пл}$, будет происходить за конечное число шагов. При очевидно должно выполняться условие $\Delta t_{пл} < \tau$.

Аналитическая запись геометрической интерпретации алгоритма, приведенного выше, приводит нас к следующему соотношению для определения предельного значения (i+I)-го опроса в зависимости от информации о мере близости цели на i-том шаге контроля подсистемы

$$t_{i+1} = t_1 + \frac{V(t_i)}{V_{пл}}(t_{пл} - t_1). \qquad (2)$$

Здесь t_I – момент первого опроса, а $V_{пл}$ и $V(t_i)$ соответственно плановый и фактический интегральные меры близости цели.

Допустим теперь, что $V_{пл}(t)$ и $V_р(t)$ заданы в виде прямых, и истинная кривая движения к цели совпадает с $V_{пл}(t)$. В этом случае $t_ф = t_{пл}$, и опрос подсистемы имеет лишь профилактический характер.

Учитывая очевидное соотношение

$$\frac{V(t_i)}{V_{пл}} = \frac{t_i}{t_{пл}},$$

формулу (2) можно переписать в следующем виде

$$t_{i+1} = t_1 + \left(1 - \frac{t_1}{t_{пл}}\right) t_i.$$

Используя данное выражение, легко получить формулу для величины (i+I) -го шага

$$t_{i+1} - t_i = \left(1 - \frac{t_1}{t_{пл}}\right)\left(t_i - t_{i-1}\right).$$

Полученное соотношение можно записать также в виде

$$t_{i+1} - t_i = \left(1 - \frac{t_1}{t_{пл}}\right)^i t_1,$$

и так как

$$t_{пл} = \sum_{i=0}^{\infty} \left(t_{i+1} - t_i\right),$$

нетрудно видеть, что

$$t_{пл} = t_1 \sum_{i=0}^{n} \left(1 - \frac{t_1}{t_{пл}}\right)^i + t_1 \sum_{i=n+1}^{\infty} \left(1 - \frac{t_1}{t_{пл}}\right)^i. \qquad (3)$$

Здесь

$$R_n = t_1 \sum_{i=n+1}^{\infty} \left(1 - \frac{t_1}{t_{пл}}\right)^i$$

– остаточный член, на который очевидно будет разумно наложить следующее ограничение $R_n \leqslant \Delta t_{пл}$.

Теперь учитывая то, что члены под знаком суммы в правой части (3) являются суммами убывающей геометрической прогрессии, можно написать, что

$$t_{пл} = t_1 + t_{пл}\left[\left(1 - \frac{t_1}{t_{пл}}\right) - \left(1 - \frac{t_1}{t_{пл}}\right)^{n+1}\right] + \Delta t_{пл}.$$

Проведя некоторые простейшие преобразования, можно получить выражение для числа профилактических опросов, требуемых для контроля подсистемы с параметрами $t_{пл}$, t_p и $\Delta t_{пл}$

$$n = \frac{\ln \dfrac{\Delta t_{пл}}{t_p}}{\ln \dfrac{t_p}{t_{пл}}}. \qquad (4)$$

Рассмотрим теперь сложную систему, состоящую из подсистем указанного выше вида. Таким образом мы рассматриваем два фиксированных уровня некоторой иерархической структуры. При этом нумерацию рангов осуществляем следующим образом - рангу подсистем присваиваем номер к , а рангу систем номер к-I.

Так как функционирование каждой подсистемы направлено на достижение определенной цели, то данная система очевидно будет обладать в общем случае некоторой сетью целей, представляющей собой детализацию глобальной цели системы. Причем система имеет интегральные кривые близости цели $V_{пл}^{(c)}(t)$, $V_p^{(c)}(t)$ и $V^{(c)}(t)$ характеризующие различные режимы движения системы к глобальной цели, а также соответствующие параметры $T_{пл}$, T_p и $\Delta T_{пл}$.

Наличие сети целей подразумевает также существование некоторого критического пути с определенным количеством целей на нем. А это означает наличие некоторого количества критических подсистем - m. Выберем из всех этих подсистем некоторую i-тую подсистему, для которой число профилактических опросов, определяемое по формуле (4) окажется минимальным, т.е.

$$\min_{1 \le i \le m} N_i = \min_{1 \le i \le m} \frac{\ln \dfrac{\Delta t_{пл i}}{t_{pi}}}{\ln \dfrac{t_{pi}}{t_{пл i}}}.$$

Здесь имеет место очевидно нервенство

$$\Delta t_{пл i} < t_{pi} < t_{пл i}. \qquad (5)$$

Еели каждая подсистема, лежащая на критическом пути, опрашивается указанное число раз, то соответствующее минимальное число опросов, необходимое для опроса всего критического пути, будет равно

$$\min_{1 \le i \le m} N_к = m \cdot \min_{1 \le i \le m} N_i, \qquad (6)$$

где $N_к$ - число опросов, необходимое для критического пути при принятой детализации - $\Delta t_{пл i}$.

Рассмотрим теперь систему с параметрами $T_{пл}$, T_p и $\Delta T_{пл}$, характеризующих ее движение к глобальной цели, причем

$$\begin{cases} T_{пл} = \gamma_1 t_{пл i}, \\ T_p = \gamma_2 t_{pi}, \\ \Delta T_{пл} = \gamma_3 \Delta t_{пл i}, \end{cases} \qquad (7)$$

где $\gamma_1, \gamma_2, \gamma_3$, - произвольные неотрицательные числа больше единицы, и допустим, что

$$\gamma_1 > \gamma_2 > \gamma_3 . \tag{8}$$

Предъявим теперь рассматриваемой иерархической структуре вполне закономерное требование, заключающееся в уменьшении числа опросов, необходимых для управления, с увеличением ранга (с уменьшением номера ранга) иерархии. При этом нашему требованию придадим наиболее жесткий вид

$$N_{k-1} \leq \min_{1 \leq i \leq m} N_k , \tag{9}$$

где N_{k-1} - число опросов, необходимое для опроса системы с точки зрения детализации (К-I)-го ранга, т.е. по параметрам (7). Действительно, учитывая выражения (5), (6), (7) и (8) можно написать следующие неравенства

$$N_{k-1} = \frac{\ln \dfrac{\gamma_3 \Delta t_{n\lambda i}}{\gamma_2 t_{pi}}}{\ln \dfrac{\gamma_2 t_{pi}}{\gamma_1 t_{n\lambda i}}} \leq \frac{\ln \dfrac{\gamma_3}{\gamma_2}}{\ln \dfrac{\gamma_2}{\gamma_1}} + \frac{\ln \dfrac{\Delta t_{n\lambda i}}{t_{pi}}}{\ln \dfrac{t_{pi}}{t_{n\lambda i}}} =$$

$$= \frac{\ln \dfrac{\gamma_3}{\gamma_2}}{\ln \dfrac{\gamma_2}{\gamma_1}} + \min_{1 \leq i \leq m} N_i \leq \frac{\ln \dfrac{\gamma_3}{\gamma_2}}{\ln \dfrac{\gamma_2}{\gamma_1}} + m \cdot \min_{1 \leq i \leq m} N_i =$$

$$= \frac{\ln \dfrac{\gamma_3}{\gamma_2}}{\ln \dfrac{\gamma_2}{\gamma_1}} + (m-1) \min_{1 \leq i \leq m} N_i + \min_{1 \leq i \leq m} N_i .$$

Отсюда легко видеть, что для осуществления условия

$$N_{k-1} \leq \min_{1 \leq i \leq m} N_i ,$$

автоматически осуществляющего условие (9), необходимо чтобы

$$\frac{\ln \frac{\gamma_3}{\gamma_2}}{\ln \frac{\gamma_2}{\gamma_1}} + (m-1) \min_{1 \leq i \leq m} N_i \leq 0. \qquad (10)$$

Учитывая теперь очевидное неравенство

$$\gamma_1 t_{\text{пл}i} \geq \gamma_2 t_{pi} , \qquad (11)$$

условие можно записать следующим образом

$$\frac{\gamma_3}{\gamma_2} \geqslant \left(\frac{\Delta t_{\text{пл}i}}{t_{pi}} \right)^{m-1}. \qquad (12)$$

Докажем, что условие (12) является в то же время и достаточным условием. Действительно, если допустить, что

$$\frac{\gamma_3}{\gamma_2} < \left(\frac{\Delta t_{\text{пл}i}}{t_{pi}} \right)^{m-1},$$

то при $m \to \infty$

$$\frac{\gamma_3}{\gamma_2} < \lim_{m \to \infty} \left(\frac{\Delta t_{\text{пл}i}}{t_{pi}} \right)^{m-1} = 0.$$

Это означает, что либо γ_3 , либо γ_2 является отрицательной величиной, что противоречит условию задачи. Полученное противоречие доказывает предположение. Однако, так как в действительности имеется лишь некоторое конечное значение числа подсистем на критическом пути $m = m_{кр}$ то, очевидно, существует интервал для положительных значений $\frac{\gamma_3}{\gamma_2}$

$$0 < \frac{\gamma_3}{\gamma_2} \leq \left(\frac{\Delta t_{плi}}{t_{pi}} \right)^{m_{кр}-1},$$

в котором неясно соблюдение свойства достаточности для условия (12).

Скорость сходимости величины $y = \left(\frac{\Delta t_{плi}}{t_{pi}} \right)^{m_{кр}-1}$ к нулю при увеличении m можно характеризовать величиной

$$J = \int_1^\infty \left(\frac{\Delta t_{плi}}{t_{pi}} \right)^{m-1} dm = \frac{1}{\ln \frac{t_{pi}}{\Delta t_{плi}}},$$

откуда очевидно, что величина J меняется пропорционально показателя детализации – $\Delta t_{плi}$. При этом, чем выше детализация, что соответствует меньшим значениям $\Delta t_{плi}$, тем выше скорость сходимости указанной функции к нулю, и, соответственно, меньше зона, где свойство достаточности ставится под сомнение. Очевидно, подбором параметров $\Delta t_{плi}$, t_{pi} и m указанную зону можно сделать меньше любого наперед заданного числа.

Решим теперь неравенство (10) относительно $\frac{\gamma_2}{\gamma_1}$ и проведя аналогичные рассуждения, легко убедимся, что

$$\frac{\gamma_2}{\gamma_1} \leq \left(\frac{t_{пл i}}{t_{pi}}\right)^{\frac{1}{m-1}}.$$

Учитывая соотношение (11), а также неравенства (10) и

$$\gamma_2 t_{pi} \geq \gamma_3 \Delta t_{пл i},$$

полная система ограничений, накладываемых в этом случае на соотношения $\frac{\gamma_2}{\gamma_1}$ и $\frac{\gamma_3}{\gamma_2}$, примет следующий вид

$$\begin{cases} \left(\frac{\Delta t_{пл i}}{t_{pi}}\right)^{m-1} \leq \frac{\gamma_3}{\gamma_2} \leq \frac{t_{pi}}{\Delta t_{пл i}} \\ \frac{t_{pi}}{t_{пл i}} \leq \frac{\gamma_2}{\gamma_1} \leq \left(\frac{t_{пл i}}{t_{pi}}\right)^{\frac{1}{m-1}} \end{cases}$$

Учитывая соотношения (7), необходимое условие для оптимальности данной иерархической структуры окончательно можно записать следующим образом

$$\begin{cases} \left(\frac{\Delta t_{пл i}}{t_{pi}}\right)^{m} \leq \frac{\Delta T_{пл}}{T_p} \leq 1 \\ \left(\frac{t_{pi}}{t_{пл i}}\right)^{2} \leq \frac{T_p}{T_{пл}} \leq \left(\frac{t_{пл i}}{t_{pi}}\right)^{\frac{2-m}{m-1}} \end{cases}$$

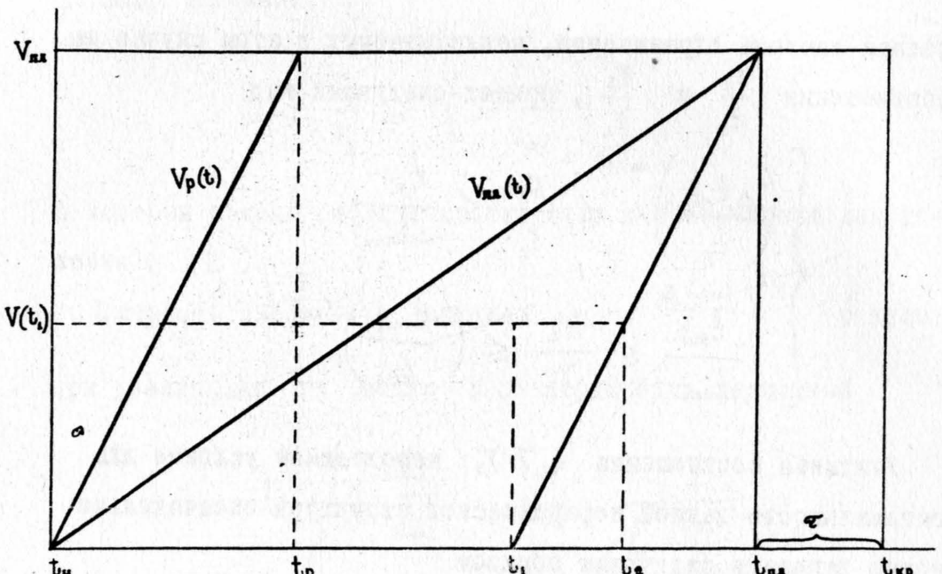

Рис. 1