

IFAC



WARSZAWA 1969

INTERNATIONAL FEDERATION
OF AUTOMATIC CONTROL

Optimal Control of Discrete Systems

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

TECHNICAL
SESSION

62



Organized by
Naczelna Organizacja Techniczna w Polsce

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

Optimal Control of Discrete Systems

TECHNICAL SESSION No 62

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



Organized by
Naczelna Organizacja Techniczna w Polsce



K-1328

Contents

Paper No			Page
62.1	SU	- K.B.Alexeev - On the Synthesis of Optimum Control Systems with Account Being Taken of Set Reliability.....	3
62.2	SU	- L.A.Vołodin, A.I.Moroz, M.G.Ogulnik - On the Practical Realisation of Time-Optimal Synthesis Functions.....	16
62.3	N	- O.A.Solheim, F.Pohner - Optimal Control of a Class of Discrete Systems.....	34
62.4	A	- H.Florian - Optimal Control of Nonlinear Discrete Systems.....	52
62.5	Pl	- J.Kurman - Chain Models as Inertialess Optimal Controllers of Multidimensional Processes.....	64

**Biblioteka
Politechniki Białostockiej**



1120435

Wydawnictwa Czasopism Technicznych NOT - Polska

Zakład Poligraficzny WCT NOT. Zam. 58/69.

К СИНТЕЗУ ОПТИМАЛЬНЫХ СИСТЕМ УПРАВЛЕНИЯ С УЧЕТОМ ЗАДАННОЙ НАДЕЖНОСТИ

АЛЕКСЕЕВ К.Б.,
(Москва, СССР)

Завод ВТУЗ при автозаводе им. И.А.Лихачева

Из общего числа требований, предъявляемых к системам автоматического управления с продолжительным временем работы, требование надежности является основным. Тем не менее при расчете и проектировании таких систем это требование обеспечивается в последнюю очередь, после того как произведен выбор структуры и схемы системы. Данное обстоятельство объясняется не только трудностью выявления зависимости качественных и энергетических характеристик системы от статистических показателей надежности, входящих в ее состав элементов, но и несовершенством используемых методов исследования. Разработанные в последние годы методы исследования систем управления с общих теоретических позиций и, в частности, оптимальных систем, позволяют использовать существующую зависимость между указанными характеристиками и показателями на начальной стадии их проектирования. Одним из таких методов, получивших широкое распространение на практике, является принцип максимума Понтрягина [1].

В докладе для определенного класса систем впервые делается попытка предложить способ синтеза систем на основе обобщенного критерия оптимальности, включающего требования к точности, энергетическим затратам и заданной надежности.

1. Метод исследования

Сущность принципа максимума Понтрягина заключается в нахождении экстремального значения некоторого функционала, являющегося мерой количественной оценки соответствия синтезируемой системы, предъявляемым к ней требованиям. В качестве последних для рассматри-

ваемого класса автоматических систем принимаются:

- 1) минимум взвешенной дисперсии ошибки системы;
- 2) минимум потребления энергии.

При этом предполагается, что система отвечает заданной надежности β (вероятности безотказной работы).

Согласно принципу максимума, система, удовлетворяющая приведенным требованиям, является оптимальной, причем она имеет релейную структуру. Это позволяет связать надежность системы с максимально допустимым числом переключений релейного элемента. В свою очередь число переключений зависит от реализуемого качества работы системы. Обозначая через $\vec{x}$ вектор-столбец ошибки системы и полагая потребление энергии пропорциональным величине управляющего воздействия $|u|$, прикладываемого к объекту управления, выражение функционала, используемого при удовлетворении приведенных требований к системе, можно записать в виде

$$J = \frac{1}{2} \int_0^{t_y} (c_1 x^2 + c_2 \dot{x}^2) dt + \int_0^{t_y} |u| dt, \quad (1)$$

где t_y — время управления,
 c_1 и c_2 — весовые коэффициенты.

Вследствие недостаточности принципа максимума определение оптимального управления u^* производится из множества экстремальных управлений u . Эта задача является наиболее сложной и, как правило, решается путем исследования процесса управления в фазовом пространстве. Для замкнутых автоматических систем при проведении такого исследования применяется метод инверсного времени, при котором оптимальные траектории находятся решением системы дифференциальных уравнений по инверсному времени от некоторой точки в фазовом пространстве. В соответствии с вычислительной процедурой принципа максимума вначале в этом пространстве выделяются области экстремальных управлений, где функционал (1) приобретает относительный минимум. Фазовые траектории состояния системы в данной области отличаются моментами переключений управления u , каждая компонента которого имеет

в общем случае три возможных значения: $+1; 0; -1$. При интегрировании исходного дифференциального уравнения системы в инверсном времени, указанные моменты переключений управления можно выразить через время

τ_1 , характеризующее время действия на систему управляющих воздействий. Тогда абсолютный минимум функционала (1) находится из условия

$$\frac{\partial J}{\partial \tau_1} = 0, \quad (2)$$

при

$$\frac{\partial^2 J}{\partial \tau_1^2} > 0,$$

где

$$J(x_0, u) \equiv J(\tau_1),$$

x_0 — начальное значение ошибки ($t = 0$) или значение ошибки, при которой происходит срабатывание реле (включение управления).

Выбор τ_1 из условия (2) однозначно определяет оптимальное управление u^* .

Надежность системы $\mathcal{P}$ является функцией структуры системы S и времени ее функционирования T . Понятие структуры системы включает в себя в данном случае комплекс конструктивных мер, позволяющих обеспечить различные уровни надежности системы. Время функционирования системы T может быть представлено зависимостью от τ_1 .

Таким образом,

$$\mathcal{P} = \mathcal{P}(S, T) = \mathcal{P}(S, \tau_1).$$

Тогда нахождение оптимального управления сводится к решению задачи о минимуме функционала (1) при дополнительном ограничении

$$\mathcal{P} = \mathcal{P}(S, \tau_1) \geq \mathcal{P}_3.$$

Для этой цели может быть использован метод Лагранжа.

Условие (2) переписывается теперь в виде

$$\frac{\partial}{\partial \tau_1} (J(\tau_1) + \lambda \mathcal{P}(\tau_1, S)) = 0, \quad (3)$$

где λ - множитель Лагранжа.

Когда время управления t_y также подлежит определению, оптимальное управление находится из следующих условий

$$\frac{\partial (J(t_y, \tau_1) + \lambda \mathcal{F}(t_y, \tau_1, S))}{\partial t_y} = 0, \quad (4)$$

$$\frac{\partial (J(t_y, \tau_1) + \lambda \mathcal{F}(t_y, \tau_1, S))}{\partial \tau_1} = 0$$

Полученные условия (3) и (4) облегчают выбор структуры системы, отвечающей предъявляемым требованиям с учетом заданной надежности.

2. Обоснование метода исследования

Рассмотрим систему управления ориентацией летательного аппарата.

Пусть угловое положение аппарата задается в орбитальной базовой системе отсчета $OXYZ$ с началом в центре Земли. Обозначим через $\dot{\gamma}$ угловую скорость вращения этой системы вокруг оси OZ (фиг. 1). Задача управления, рассматриваемая ниже, заключается в совмещении оси Oy системы координат $Oxyz$, жестко связанной с аппаратом, с геоцентрической вертикалью в плоскости, проходящей через ось OY системы отсчета и вектор скорости $\vec{V}$ центра масс аппарата.

Система ориентации включает в себя:

- одноординатную вертикаль для измерения углового рассогласования (ошибки управления между осями OY и Oy);
- две пары струйных сопел, соответствующим образом установленных на корпусе аппарата, для приложения к аппарату управляющего момента M_y ;
- усиленно-преобразующие устройства и электромагнитные клапаны соленоидного типа, предназначенные для управления работой струйных сопел.

В предположении, что аппарат является абсолютно твердым телом, а ось Oy - главной осью инерции, ди-

динамические свойства объекта управления описываются уравнением

$$I \frac{d^2 \varphi}{dt^2} = M_y(t) + M_z(t), \quad (5)$$

где φ - угловое отклонение аппарата;

I - момент инерции;

M_z - возмущающий момент.

Вследствии орбитального движения аппарата сигнал на входе системы ориентации будет

$$\varphi_{zx} = \dot{\nu}t + \varphi_0, \quad (6)$$

где φ_0 - начальное значение φ_{zx} при $t = 0$.

Следовательно, ошибка ориентации определяется равенством

$$\theta = \varphi_{zx}(t) - \varphi(t) = \dot{\nu}t + \varphi_0 - \varphi. \quad (7)$$

На фиг. 2 изображена блок-схема замкнутой системы ориентации. Требования, которым должна удовлетворять данная система, были приведены выше.

Работа системы ориентации происходит следующим образом.

При отклонении аппарата на угол $\theta > \theta_g$, где θ_g - допустимая ошибка ориентации, происходит включение струйных сопел. В результате действия на аппарат управляющего момента ошибка устраняется. Однако орбитальное движение аппарата и влияние возмущающего момента снова приводят к накоплению ошибки и включению сопел.

Обозначим время отработки ошибки ориентации

$\theta = \theta_g$ через t_y , а время накопления ошибки через t_n . Тогда число циклов n отработки и накопления ошибки за время T полета аппарата будет равно:

$$n = \frac{T}{t_y + t_n}. \quad (8)$$

Для релейной системы ориентации в течение одного цикла произойдет четыре переключения в системе (реле и клапана), а всего за время полета $4n$ переключений.

Если этому числу переключений соответствует вероятность безотказной работы системы $P(S, t_1) < P_3$, то требование к надежности не выполняется. Это означает, что достижимая при выбранной структуре S системы точность ориентации не отвечает назначению летательного аппарата.

Рассмотрим выбор значения t_1 , соответствующего оптимальному управлению.

Полагая угловую скорость аппарата постоянной ($\dot{\psi} = \text{const}$) из равенств (7) находим

$$\frac{\partial \theta}{\partial t} = \dot{\psi} - \frac{d\varphi}{dt}, \quad \frac{d^2 \theta}{dt^2} = - \frac{d^2 \varphi}{dt^2}$$

Это позволяет переписать уравнение (5) в виде

$$\frac{d^2 \theta}{dt^2} = - \frac{M_y}{I} - \frac{M_z}{I}$$

или

$$\frac{d^2 \theta}{dt^2} = - \alpha u(t) - m_z(t), \quad (9)$$

где

$$\alpha = \frac{M_g}{I}, \quad u(t) = \frac{M_y(t)}{M_g}, \quad m_z(t) = \frac{M_z(t)}{I}.$$

M_g — предельно допустимое значение управляющего момента.

Поскольку $M_y \gg M_z$, то в течение всего времени работы сопел t_1 влиянием возмущающего момента на угловые движения аппарата можно пренебречь. За время накопления ошибки возмущающий момент изменяет угловую скорость аппарата на величину $\frac{M_z}{I} t_n$ при $M_z = \text{const}$.

Вследствие малости временного интервала t_n будем рассматривать $M_z(t)$ как импульсное возмущающее воздействие на аппарат в момент включения сопел ($t=0$), т.е.

$$M_z(t) = M_{z0} \delta(t), \quad (10)$$

где M_{z0} — площадь импульса;

$\delta(t)$ — функция Дирака.

Обозначим

$$x_1(t) = \theta(t) = \dot{\gamma}t + \varphi_0 - \varphi(t),$$

$$x_2(t) = \frac{d\theta}{dt} = \dot{\gamma} - \dot{\varphi}(t),$$

и представим уравнение (9) в виде системы из двух уравнений первого порядка

$$\frac{dx_1(t)}{dt} = x_2(t), \quad \frac{dx_2}{dt} = -qu(t) \quad (11)$$

при следующих начальных условиях:

$$x_1(0) = x_{10} = \varphi_0, \quad x_2(0) = x_{20} = \dot{\gamma} - \dot{\varphi}(0). \quad (12)$$

С учетом выражения для возмущающего момента (10) последнее условие может быть записано так:

$$x_{20} = \dot{\gamma} - m_{\xi_0},$$

где

$$m_{\xi_0} = \frac{M_{\xi_0}}{I}.$$

Характерно, что действие входного сигнала и возмущающего момента проявляются через начальные значения вектора состояния системы

$$\vec{x}(0) = (x_{10}, x_{20}).$$

Значения этого вектора в момент $t = t_y$ примем равным нулю, т.е.

$$\vec{x}(t_y) = (0, 0), \quad (13)$$

причем t_y пока не определено.

Таким образом, граничные условия системы уравнений (11) или одного уравнения (9), записанного в новых обозначениях как

$$\frac{d^2 x_1(t)}{dt^2} = -qu(t), \quad (14)$$

оговорены полностью.

В соответствии с принципом максимума найдем выражение для гамильтониана

$$H = \frac{1}{2} c_1 \dot{x}_1^2(t) + \frac{1}{2} c_2 \dot{x}_2^2(t) + |u(t)| - x_2(t) p_1(t). \quad (15)$$

Уравнения для сопряженных переменных $p_1(t)$ и $p_2(t)$ имеют вид

$$\frac{dP_1(t)}{dt} = - \frac{\partial H}{\partial x_1} = -c_1 x_1(t), \quad (16)$$

$$\frac{dP_2(t)}{dt} = - \frac{\partial H}{\partial x_2} = -c_2 x_2(t) - P_1(t).$$

Поскольку начальные и конечные значения вектора состояния $\vec{x}(t)$ заданы, граничные условия для определения $P_1(t)$ и $P_2(t)$ из уравнений (16) могут быть взяты произвольными. Систему уравнений (16) можно записать одним уравнением второго порядка

$$\frac{d^2 P_2(t)}{dt^2} = c_1 x_1(t) + c_2 u(t). \quad (17)$$

Как видно из уравнений (11) и (16), определение оптимального управления $u^*(t)$ связано с нахождением сопряженной переменной $P_2(t)$, которая в свою очередь зависит от текущего значения вектора состояния.

Выразим вначале $u(t)$ как функцию $P_2(t)$. Минимум гамильтониана будет при условии, что

$$[|u^*| - u^* P_2] = \min[|u| - u P_2], \quad (18)$$

при

$$-1 < u < 1,$$

где $u^*(t)$ — оптимальное управление.

Из условия (18) находим

$$u^*(t) = \begin{cases} +1 & \text{при } P_2(t) \geq \frac{1}{a}, \\ 0 & \text{при } |P_2(t)| < \frac{1}{a}, \\ -1 & \text{при } P_2(t) \leq -\frac{1}{a}. \end{cases} \quad (19)$$

Введем инверсное время $\tau = t_f - t$, отсчитываемое от конца процесса управления. Учитывая, что $dt = -d\tau$, перепишем уравнения (14) и (17)

$$\frac{d^2 x_1(\tau)}{d\tau^2} = -a u(P_2), \quad (20)$$

причем

$$\frac{d^2 p_2}{d\tau^2} = c_1 x_1(\tau) + a c_2 u(p_2) ,$$

$$x_1(\tau) \Big|_{\tau=0} = \frac{dx_1(\tau)}{d\tau} \Big|_{\tau=0} = 0 , \quad \frac{dx_1(\tau)}{d\tau} = -x_2(\tau) .$$

Интегрирование данных уравнений в инверсном времени свидетельствует [2], что моменты переключения управления τ_1 и τ_2 (фиг. 3) находятся в области, ограниченной параболой

$$x_1(\tau) = -\frac{1}{2a} x_2^2(\tau)$$

и линией, для которой

$$\tau_1(\tau_2 - \tau_1)^2(2\tau_2 + \tau_1) = \frac{2l}{c_1 a^2} ,$$

причем $t_y = \tau_1 + \tau_2$.

Выбор значений τ_1 и τ_2 в данной области производится из условия оптимизации принятого критерия качества.

Расчетная схема синтеза

Для иллюстрации рассмотренного метода синтеза изложим кратко содержание вычислительной процедуры. Для большей физической наглядности (фиг. 4) примем

$$x_{10} < 0 , \quad x_{20} = \sqrt{-m_{\phi_0}} = 0 .$$

Из последнего условия находим

$$t_n = \frac{I\gamma}{M_{\phi}}$$

С учетом моментов переключения в инверсном времени имеем

$$u^*(\tau) = \begin{cases} +1 & \text{при } 0 \leq \tau \leq \tau_1 , \\ 0 & \text{— } \tau_1 < \tau < \tau_2 , \\ -1 & \text{— } \tau_2 < \tau \leq \tau_1 + \tau_2 . \end{cases} \quad (21)$$

Интегрирование уравнений (11) в инверсном времени дает

$$x_1(\tau) = \begin{cases} -\frac{1}{2} a \tau^2, & \text{при } 0 \leq \tau \leq \tau_1, \\ -a \tau_1 (\tau - \frac{1}{2} \tau_1), & \text{--- } \tau_1 < \tau < \tau_2, \\ -a (\tau_1 + \tau_2) \tau + \frac{1}{2} a (\tau_1^2 + \tau_2^2 + \tau^2), & \text{--- } \tau_2 < \tau \leq \tau_1 + \tau_2. \end{cases} \quad (22)$$

$$x_2(\tau) = \begin{cases} a \tau & \text{при } 0 \leq \tau \leq \tau_1, \\ a \tau_1 & \text{--- } \tau_1 < \tau < \tau_2, \\ a \tau_1 (\tau_1 + \tau_2) - a \tau & \text{--- } \tau_2 < \tau < \tau_1 + \tau_2, \end{cases} \quad (23)$$

С учетом начальных условий из последних равенств получаем

$$x_{20} = x_2(\tau_1 + \tau_2) = 0$$

$$x_{10} = x_1(\tau_1 + \tau_2) = -a \tau_1 \tau_2.$$

Отсюда

$$\tau_2 = -\frac{x_{10}}{a \tau_1}.$$

Следовательно, время $t_y = \tau_1 + \tau_2$, в течение которого происходит отработка ошибки, можно выразить в функции τ_1 как

$$t_y = \tau_1 - \frac{x_{10}}{a \tau_1} = \tau_1 + \frac{|x_{10}|}{a \tau_1},$$

а число циклов

$$\frac{T}{t_y + t_n} = \frac{T \tau_1}{\tau_1^2 + \frac{|x_{10}|}{a} \tau_1 + \frac{|x_{10}|}{a}}.$$

и переключений релейного элемента

$$k_n = \frac{4 \tau_1 T}{\tau_1^2 + \frac{|x_{10}|}{a} \tau_1 + \frac{|x_{10}|}{a}}.$$

Представим функционал (1) следующим образом:

$$J = \frac{1}{2} c_1 \int_0^{\tau_1 + \tau_2} x_1^2(\tau) d\tau + \frac{1}{2} c_2 \int_0^{\tau_1 + \tau_2} x_2^2(\tau) d\tau + 2 \tau_1. \quad (24)$$

После выполнения интегрирования получаем

$$J = \frac{1}{2} c_1 \left[a^2 \left(\frac{x_{10}}{a} \tau_1 - \frac{4}{3} \tau_1^3 \right) \right] + \frac{1}{2} c_2 \left[a^2 \left(\frac{3}{10} \tau_1^5 - \frac{7}{12} \frac{x_{10}}{a} \tau_1^3 + \right. \right. \\ \left. \left. + \frac{2}{3} \frac{x_{10}^2}{a^2} \tau_1 - \frac{x_{10}^3}{a^3 \tau_1} - \frac{x_{10}^4}{a^4 \tau_1^3} \right) \right] + 2 \tau_1. \quad (25)$$

Далее выражаем вероятность безотказной работы системы в функции S и τ_1 . Тогда значение τ_1^* , минимизирующее функционал качества системы (1) при ограничении

$$P(s, \tau_1) > P_3,$$

определяется из условия

$$\frac{\partial (1 + \lambda P)}{\partial \tau_1} = 0,$$

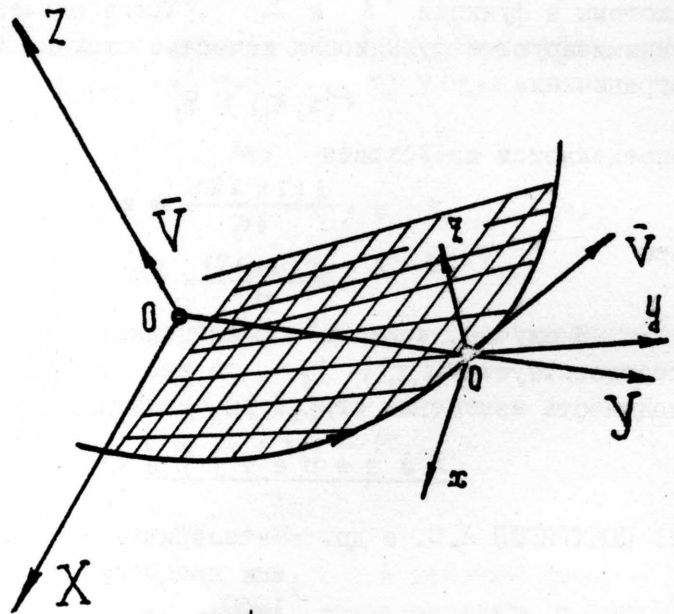
при

$$\frac{\partial^2 (1 + \lambda P)}{\partial \tau_1^2} > 0.$$

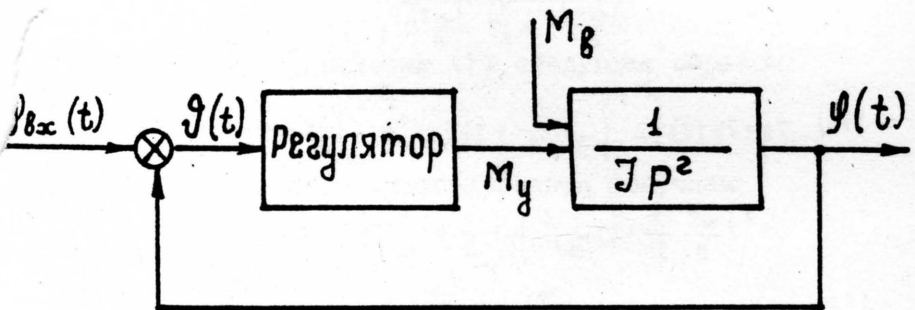
В случае, если τ_1^* , найденное таким образом, соответствует $P(s, \tau_1^*) < P_3$, то это указывает на необходимость изменения структуры системы.

Л и т е р а т у р а

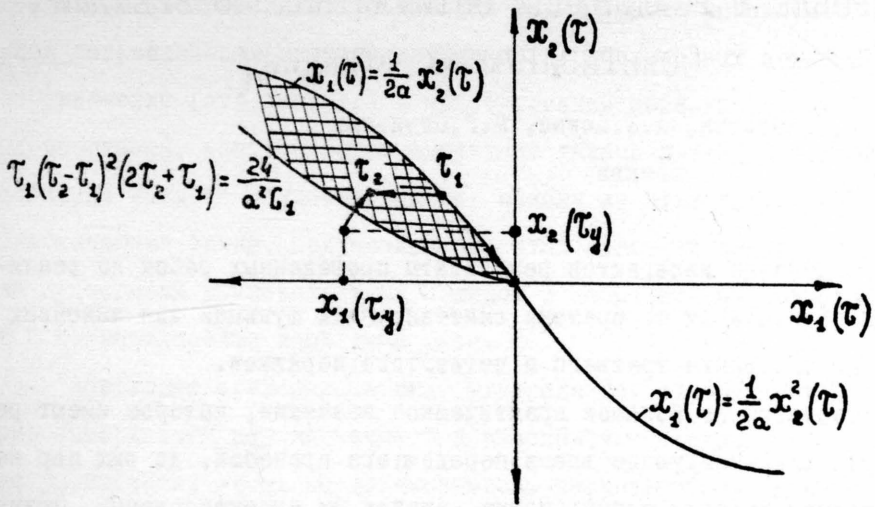
1. ПОНТЯГИН Л.С. и др. Математическая теория оптимальных процессов. Физматгиз, 1960.
2. Fay W.H. IEEE Trans. of Automatic Control, AC-8, N 2, 1963 г.



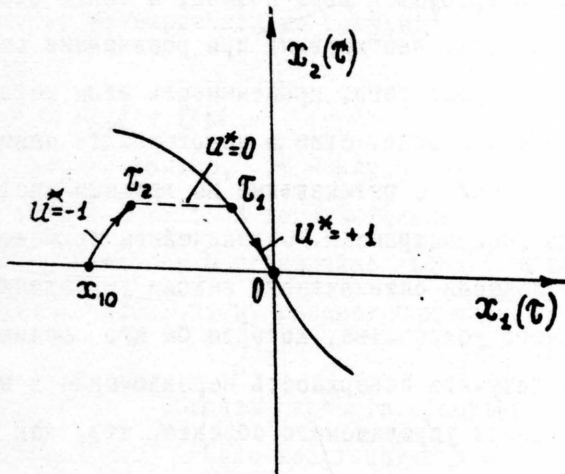
Фиг. 1



Фиг. 2



$\phi_{u2.3}$



$\phi_{u2.4}$



К ПРОБЛЕМЕ РЕАЛИЗАЦИИ ОПТИМАЛЬНЫХ ПО ВРЕМЕНИ СИНТЕЗИРУЮЩИХ ФУНКЦИЙ

Л.А.Володин, А.И.Мороз, М.Г.Огульник

Москва

В докладе излагаются результаты проведенных работ по реализации оптимальных по времени синтезирующих функций для линейных систем регулирования третьего и четвертого порядков.

Несмотря на большое практическое значение, которое имеют регуляторы, минимизирующие время переходного процесса, до сих пор не существует сколько-нибудь общих методов их проектирования. Применяемые в настоящее время методы проектирования оптимальных регуляторов по мнению авторов обладают тем недостатком, что к каждой конкретной задаче требуется свой подход, а также страдают чрезмерной сложностью средств, необходимых при реализации оптимальных законов регулирования. Кроме того, применимость этих методов ограничивается системами частного вида. Этим в некоторой степени объясняется тот факт, что на практике оптимальные по времени системы еще недостаточно широко распространены. Ограниченное применение существующих методов определения оптимального закона управления и проектирования вычислительного устройства, которое бы его реализовало, объясняется стремлением получить поверхность переключения в виде явной функции фазовых координат управляемого объекта, что, как правило, невозможно сделать. В тех же частных случаях, когда это удается сделать, эта функция имеет исключительно сложный вид.

В настоящей работе используется другой подход к реализации. Исходным при расчете алгоритма является одна форма параметрического задания поверхности переключения, которая оказывается очень удобной при изучении геометрии поверхности и ее особенностей. В основе

процедуры построения алгоритма для формирования управления лежит идея аппроксимации каждого из множеств, разделенных поверхностью переключения (это сводится к аппроксимации поверхности переключения плоскостями), многогранными выпуклыми телами и введения логики, согласно которой устанавливается, какому из этих тел принадлежит изображающая точка. Получаемый при этом алгоритм легко реализуется на аналоговых вычислительных машинах с небольшим числом элементов и с нелинейностями лишь типа реле.

Некоторые приведенные ниже утверждения, относящиеся к геометрии поверхности переключения и к обоснованию предлагаемого алгоритма управления, здесь не доказываются, поскольку это не является целью доклада (математическое обоснование имеется в работах, которые в настоящее время находятся в печати).

1. Постановка задачи

имеется система автоматического регулирования, которая описывается уравнением

$$\dot{X} = AX + Bu \quad (1)$$

где $X(t)$ - n -мерный вектор, A - постоянная квадратная матрица, собственные значения которой действительны, B - постоянный вектор, скалярное управление u ограничено $u_0 \leq u \leq u_1$ ($u_0 < 0$, $u_1 > 0$). Здесь будут рассмотрены случаи $n = 3$ и $n = 4$.

Обозначим через G^T область фазового пространства, из точек которой возможен переход в начало координат с допустимым кусочно-постоянным управлением за время не превосходящее T . G^T является замкнутым выпуклым телом [2].

Пусть известна оптимальная по времени синтезирующая функция $u = F(x)$, $u_0 \leq F(x) \leq u_1$, которая определяется поверхностью переключения Ω^T . Под аппроксимацией функции $F(x)$ законом управления $\bar{u}(x)$

понимается близость $F(x)$ и $F_{\varepsilon}(x)$ по мере, т.е. для заданного $\varepsilon > 0$ и $\underline{\varepsilon} > 0$ можно указать такую $F_{\varepsilon}(x)$, что

$$\mu \{x: |F(x) - F_{\varepsilon}(x)| \geq \underline{\varepsilon}\} < \varepsilon$$

Требуется определить функцию $F_{\varepsilon}(x)$ и построить процедуру ее реализации, в принципе, не сложными средствами.

2. Уравнение поверхности переключения

Используя принцип максимума [2] можно получить следующие параметрические уравнения поверхностей переключения Ω^T :

а) для $n = 3$ имеем для куска поверхности, на котором $u(x) = u_1$

$$x = g_1(\alpha, \beta) \equiv -u_1 e^{\frac{-\beta A}{b}} \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_1 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (2)$$

и для куска поверхности, на котором $u(x) = u_0$

$$x = g_2(\alpha, \beta) \equiv -u_1 e^{\frac{-\beta A}{b}} \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_0 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (3)$$

б) для $n = 4$ имеем для куска поверхности, на котором $u(x) = u_0$

$$x = g_3(\alpha, \beta, \gamma) \equiv -u_0 e^{\frac{-(\beta+\gamma)A}{b}} \int_0^{\frac{-(\beta+\gamma)A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_0 e^{\frac{-\gamma A}{b}} \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_0 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (4)$$

и для куска поверхности, на котором $u(x) = u_1$

$$x = g_4(\alpha, \beta, \gamma) \equiv -u_1 e^{\frac{-(\beta+\gamma)A}{b}} \int_0^{\frac{-(\beta+\gamma)A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_0 e^{\frac{-\gamma A}{b}} \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau - u_1 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (5)$$

Рассмотрим вначале задачу для $n = 3$.

можно показать, что 1) каждая точка X , принадлежащая либо поверхности (2), либо поверхности (3), является эллиптической, и что, используя это свойство, можно каждый из этих кусков поверхности переключения разбить на выпуклые куски;

2) поверхность переключения является однозначной в направлении вектора b ;

3) проекции координатных линий $\beta = 0$ поверхностей (2) и (3), т.е. проекции кривых

$$x = x^*(\alpha) \equiv -u_1 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (6)$$

и

$$x = x^*(\alpha) \equiv -u_1 \int_0^{\frac{-\beta A}{b}} e^{\frac{-\tau A}{b}} b d\tau \quad (7)$$

на плоскость $(b, x) = 0$ являются выпуклыми кривыми. Поскольку $x^{\downarrow}(\alpha) \cup x^{\uparrow}(\alpha)$ отделяет часть поверхности переключения, в точках которой $u(x) = u_0$, от части, в точках которой $u(x) = u_1$, то в силу однозначности $\Omega^{\downarrow}$ по направлению b поверхность ~~задана~~ $C^{\downarrow} \cup C^{\uparrow}$,

где $C^{\downarrow} = \{x: x = x^{\downarrow}(\alpha) + kb, \alpha > 0, -\infty < k < \infty\}$
 $C^{\uparrow} = \{x: x = x^{\uparrow}(\alpha) + kb, \alpha > 0, -\infty < k < \infty\}$

отделяет в фазовом пространстве часть поверхности переключения, задаваемую уравнением (2), от части, задаваемой (3). В силу свойства (3) каждая из цилиндрических поверхностей $C^{\downarrow}$ и $C^{\uparrow}$ является выпуклой поверхностью.

3. Общий подход к построению $F_{\alpha}(x)$

в случае $n = 3$

Построить закон регулирования в данной задаче означает описать правило, согласно которому будет определяться, по какую сторону поверхности переключения находится изображающая точка. Это определение может быть осуществлено поэтапно путем последовательной проверки условий принадлежности точки некоторым множеством, которые дают в сумме все множество $G^{\downarrow}$. Сами множества будут аппроксимироваться выпуклыми многогранными телами, которые описываются системой линейных неравенств.

Чтобы точка принадлежала некоторому фиксированному множеству, необходимо и достаточно удовлетворения в данной точке всех неравенств соответствующей системы. Таким образом, регулирование сводится к проверке условий принадлежности точки фиксированному множеству.

Построение таких выпуклых множеств предлагается осуществить следующим образом. Вначале фазовое пространство X разбивается плоскостью $[b, Ab, x] = 0$ ^{*} на два полупространства и на первом

*) через $[n^1, n^2, n^3]$ обозначен определитель, у которого столбцами являются векторы n^1, n^2, n^3 .

этапе определяется, какому из этих полупространств принадлежит начальная точка. Это сводится к определению знака линейной формы $[b, Av, x]$.

Можно показать, что кривая $x^{(i)}(\alpha)$ целиком лежит в одном из вышеопределенных полупространств. Тогда и выпуклые цилиндрические поверхности $C^{\cup}$ и $C^{\oplus}$ целиком лежат каждая в своем полупространстве.

На втором этапе поверхности $C^{\cup}$ и $C^{\oplus}$ аппроксимируются системой плоскостей $P^{\cup}$ и $P^{\oplus}$ и вписываются две системы линейных неравенств, которые соответствуют этим плоскостям.

Тогда, если начальная точка X такова, что все неравенства, например, системы $P^{\cup}$ выполняются в ней, то точка принадлежит выпуклому телу, границей которого является многогранная поверхность, которая аппроксимирует поверхность $C^{\cup}$. Если же хотя бы одно неравенство не выполняется, то точка X принадлежит множеству, дополнителю в этом полупространстве к вышеуказанному выпуклому телу. Если в качестве аппроксимации брать касательные плоскости к поверхности $C^{\cup}$, то уравнения этих плоскостей будут иметь вид $[b, e^{(i)}Av, x - x^{(i)}(\alpha)] = 0$, где n пробегает конечное число значений.

Заметим, что каждая из поверхностей $C^{\cup}$ и $C^{\oplus}$ отделяет в соответствующем полупространстве часть поверхности переключения, которая определяется формулой (2), от части, задаваемой формулой (3). Поэтому, если на втором этапе установлено, что точка X принадлежит выпуклому телу, определяемому поверхностью $C^{\cup}$ (или дополнительному множеству), то для определения, по какую сторону относительно поверхности переключения она находится, достаточно проверить условия, которые соответствуют лишь куску поверхности переключения (2), лежащему в этом выпуклом множестве (или соответствующие куску поверхности (3), лежащему в дополнительном к нему множестве).

Таким образом, на третьем этапе в каждом из множеств, которые были описаны на втором этапе, куски поверхности переключения аппроксимируются наборами плоскостей и выписываются соответствующие системы линейных неравенств.

Уравнения касательных плоскостей, которые аппроксимируют поверхность переключения, имеют вид:

$$\left[\frac{\partial g(\alpha, \beta)}{\partial \alpha}, \frac{\partial g(\alpha, \beta)}{\partial \beta}, x - g(\alpha, \beta) \right] = 0$$

где α и β пробегает конечное число значений.

Прежде чем переходить к конкретному описанию процедуры расчета регулятора, запишем уравнение поверхности переключения в другом виде. Для этого сделаем замену параметров, положив $\alpha = \tau - s$, $\beta = s$ и введем обозначение $x^\psi(\tau) = \int_0^\tau e^{-\tau A} b d\tau$. Так как $\alpha, \beta \geq 0$, то $0 \leq s \leq \tau$. Тогда уравнения (2) и (3) переписутся в виде:

$$\begin{aligned} x &= -u_0 \int_0^\tau e^{-\tau A} b d\tau - (u_1 - u_0) \int_0^s e^{-\tau A} b d\tau = \\ &= x^\psi(\tau) - x^\psi(s) + x^\psi(s) = x^\psi(\tau) - \left(1 - \frac{u_1}{u_0}\right) x^\psi(s) \end{aligned} \quad (8)$$

Аналогично получаем уравнение второй части поверхности переключения:

$$x = x^\psi(\tau) - \left(1 - \frac{u_1}{u_0}\right) x^\psi(s) \quad (9)$$

Так как $x^\psi(\tau) = \frac{u_1}{u_0} x^\psi(\tau)$, то это уравнение также можно записать в виде:

$$x = \frac{u_1}{u_0} \left[x^\psi(\tau) - \left(1 - \frac{u_1}{u_0}\right) x^\psi(s) \right] \quad (10)$$

Заметим, что для случая симметричных ограничений $|u| \leq 1$ уравнения поверхности переключения получаются в виде

$$x = x^\psi(\tau) - 2x^\psi(s) \quad (11)$$

и

$$x = -x^\psi(\tau) + 2x^\psi(s), \quad 0 \leq s \leq \tau \quad (12)$$

Здесь везде параметр τ имеет смысл общего времени прихода изображающей точки в начало координат. Тогда граница поверхности переключения, расположенной в G^{τ} является координатной линией

$$\tau = T, \text{ т.е. линией } X = X^{\tau}(T) - \left(1 - \frac{u_1}{u_2}\right) X^{\tau}(S), \quad 0 \leq S \leq T$$

Другая половина границы дается формулой

$$X = X^{\tau}(T) - \left(1 - \frac{u_1}{u_2}\right) X^{\tau}(S); \quad 0 \leq S \leq T$$

В процедуре расчета алгоритма управления нигде не используется конкретный вид системы, поэтому описание ее на каком-либо конкретном примере не приведет к ограничению общности применимости развитого подхода. Вместе с тем изложение выигрывает в наглядности.

Пример 1. Задача с симметричными ограничениями.

Именно, пусть дана система

$$\dot{x}_1 = x_2 \quad (13)$$

$$\dot{x}_2 = x_3$$

$$\dot{x}_3 = -0,0025 x_1 + 0,107 x_2 - 0,275 x_3 + 0,1 u$$

и на u наложено ограничение $|u| \leq 1$.

Пусть далее начальные условия таковы, что они равновероятно принимают значения из области G^{τ} , причем $T = 6$ сек. В соответствии с вышеизложенным вначале фазовое пространство разбивается плоскостью

$$[B, AB, x] \equiv x_1 = 0 \quad (14)$$

на два полупространства. Рассмотрим полупространство $x_1 > 0$. Если изображающая точка такова, что $x_1 > 0$, то она принадлежит либо выпуклому телу, ограниченному поверхностью S^{τ} , либо дополнительному к этому телу множеству. Чтобы это установить, поверхность S^{τ} аппроксимируется системой плоскостей, т.е. выписывается система неравенств. Для этого возьмем следующие неравенства:

$$2,5 x_1 + 1,6 x_2 > 0 \quad (15)$$

$$4,75 x_1 + 7,24 x_2 + 1 > 0$$

Далее аппроксимируются куски поверхности переключения:

1) в точках которых система (I5) удовлетворяется и 2) в точках которых хотя бы одно из неравенств (I5) не удовлетворяется. Первому куску соответствует часть поверхности (II). Используя уравнение (II) получаем систему:

$$\begin{aligned} 3,9 x_1 + 4,75 x_2 + 1,45 x_3 &> 0, \\ 1,3 x_1 + 7 x_2 + 8 x_3 + 0,8 &> 0, \\ x_1 + 1,6 x_2 + 0,17 x_3 + 0,2 &> 0, \\ 0,33 x_1 + 1,6 x_2 + 1,3 x_3 + 1,14 &> 0. \end{aligned} \quad (I6)$$

Для аппроксимации второго куска бралась одна плоскость, т.е. вычислялось одно неравенство:

$$0,95 x_1 + 2,8 x_2 + 2,5 x_3 > 0. \quad (I7)$$

Если требуется большая точность реализации, число вырабатываемых плоскостей должно быть увеличено.

Теперь для формирования управления, соответствующего точке строится следующая логическая функция:

- 1) если $x_1 > 0$,
 (I5) удовлетворяется и (I6) удовлетворяется, то $u = -1$;
 2) если $x_1 > 0$,
 (I5) удовлетворяется и (I6) не удовлетворяется, то $u = 1$; (A)
 3) если $x_1 > 0$,
 (I5) не удовлетворяется и (I7) удовлетворяется, то $u = -1$;
 4) если $x_1 > 0$,
 (I5) не удовлетворяется и (I7) не удовлетворяется, то $u = 1$.

Введем следующие логические переменные:

$$y_1 = \begin{cases} 1, & \text{если } [B, A, x] \leq 0, \\ 0, & \text{если } [B, A, x] > 0, \end{cases}$$

$$y_2 = \begin{cases} 1, & \text{если система (I5) в точке } X \text{ удовлетворяется,} \\ 0, & \text{если система (I5) в точке } X \text{ не удовлетворяется,} \end{cases}$$

$$y_3 = \begin{cases} 1, & \text{если система (I6) в точке } X \text{ удовлетворяется,} \\ 0, & \text{если система (I6) в точке } X \text{ не удовлетворяется,} \end{cases}$$

$$y_4 = \begin{cases} 1, & \text{если неравенство (I7) в точке } X \text{ удовлетворяется,} \\ 0, & \text{если неравенство (I7) в точке } X \text{ не удовлетворяется,} \end{cases}$$

Тогда на основании (А) получаем следующую таблицу истинности

Таблица I

y_1	1	1	1	1
y_2	1	1	0	0
y_3	1	0	1	0
y_4	1	0	1	0
$f_A(Y)$	0	1	0	1

Прочерки в этой таблице означают, что при заданных значениях остальных логических переменных значения логической функции не зависят от переменных, которым соответствуют эти прочерки. Отсюда получаем

$$f_A(Y) = y_1 y_2 \bar{y}_3 + y_1 \bar{y}_2 \bar{y}_4,$$

где $\bar{y}_i = 1 - y_i$. Тогда $F_A(x) = -1 + 2 f_A[Y(x)]$.

Тем самым в каждой точке X полупространства $x_1 > 0$ ~~однозначно~~ однозначно определено управление $u(x)$.

Для другого полупространства, т.е. для $x_1 < 0$ алгоритм управления рассчитывается совершенно аналогично. В задаче (I3) аппроксимирующие плоскости в $x_1 < 0$ были взяты симметричными плоскостями (I4), (I5), (I6) и (I7) относительно начала координат (это свелось лишь к изменению знака свободного члена в этих неравенствах)

Моделирование на ЭМУ-8 этого алгоритма для системы (I3) дало следующие переходные процессы (фиг. I). Из этих осциллограмм видно, что переходный процесс всюду близок к оптимальному. В малой окрест-

ности начала координат естественно возникает скользящий режим.

Пример 2. Задача с несимметричными ограничениями. Она может быть решена точно также, как и предыдущая задача, но мы хотим проиллюстрировать, как общий подход позволяет получить в некоторых случаях упрощенный алгоритм управления.

Для этого рассмотрим задачу, когда начальные условия таковы, что они принимают значения из области N определяемой соотношениями

$$|x_1| \leq a_1, \quad |x_2| \leq a_2, \quad |x_3| \leq a_3.$$

В этом случае иногда возможны упрощения общего алгоритма за счет уменьшения количества условий, подлежащих проверке. Проиллюстрируем это на примере системы (13), когда на управление наложены ограничения: $-0,715 \leq u \leq 1$ и $a_1 = 0,2$, $a_2 = 0,05$ и $a_3 = 0,035$.

Матрица A имеет следующие собственные значения: $\lambda_1 = -0,5$, $\lambda_2 = 0,025$ и $\lambda_3 = 0,2$

Области N соответствует на поверхности переключения множество, в точках которого происходит переключение на траекториях, начинающихся в N . Это множество можно определить аналитически, например, определяя точки $x(\tau, s)$ на поверхности переключения, которые имеют для каждого фиксированного τ минимальное и максимальное значение s и которые соответствуют траекториям, начинающимся в N . Гораздо проще это множество определяется графически (путем рассмотрения проекций координатных линий $\tau = \text{const}$ и $s = 0$ на плоскости $x_1 x_2$ и $x_2 x_3$).

Итак, пусть найдены подлежащие аппроксимации области на поверхностях $g_1(\tau, s)$ и $g_2(\tau, s)$. Перейдем к аппроксимации закона регулирования $F(x)$. Заметим, что для рассматриваемой задачи кривые $x^-(\tau)$ и $x^+(\tau)$ за исключением точек в малой окрестности начала координат (что со-

ответствует малым значениям параметра τ) не принадлежит областям, подлежащим аппроксимации. Это значит, что нет необходимости аппроксимировать цилиндрические поверхности $C^{\vee}$ и C^{+} , что приводит к упрощению общего алгоритма.

В данной задаче также было обнаружено, что если из алгоритма исключается аппроксимация поверхностей $C^{\vee}$ и C^{+} , то целесообразнее осуществлять разбиение фазового пространства на два полупространства не плоскостью $[b, Ab, x] = 0$, а близкой к ней плоскостью $[b, x^{\vee}(\tau\psi), x] = 0$, где $\tau^{\vee}$ — малое число. Для этого была взята плоскость, соответствующая $\tau^{\vee} = 0,5$. Отсюда одно из полупространств определяется неравенством:

$$1,5x_1 + 2,3x_2 > 0 \quad (18)$$

Поверхность g_1 аппроксимировалась неравенствами:

$$\begin{aligned} 3,3x_1 + 4,6x_2 + 1,5x_3 + 0,003 &> 0, \\ 2,2x_1 + 2,16x_2 + 0,9x_3 + 0,02 &> 0, \\ 0,79x_1 + 2,1x_2 + 0,5x_3 + 0,025 &> 0, \\ 0,16x_1 + 3,4x_2 + 0,4x_3 + 0,092 &> 0. \end{aligned} \quad (19)$$

В качестве аппроксимации g_2 были взяты следующие неравенства:

$$\begin{aligned} -0,86x_1 - 1,3x_2 - 0,5x_3 + 0,014 &> 0, \\ -1,3x_1 - 2,6x_2 - 1,8x_3 + 0,066 &> 0, \\ -1,2x_1 - 4,3x_2 - 5x_3 + 0,36 &> 0, \\ -0,2x_1 - 2x_2 - 1,7x_3 + 0,4 &> 0. \end{aligned} \quad (20)$$

Наконец, для формирования управления строится следующий алгоритм:

- 1) если (18) удовлетворяет и (19) удовлетворяется, то $u = -0,715$;
- 2) если (18) удовлетворяется и (19) не удовлетворяется, то $u = 1/\bar{A}$;
- 3) если (18) не удовлетворяется и (20) удовлетворяется, то $u = 1$;
- 4) если (18) не удовлетворяется и (20) не удовлетворяется, то

$$u = -0,715.$$

Введем следующие логические переменные:

$$y_1 = \begin{cases} 1, & \text{если } 1,5x_1 + 2,3x_2 \geq 0 \\ 0, & \text{если } 1,5x_1 + 2,3x_2 < 0 \end{cases},$$

$$y_2 = \begin{cases} 1, & \text{если система (19) в точке } X \text{ удовлетворяется} \\ 0, & \text{если система (19) в точке } X \text{ не удовлетворяется} \end{cases},$$

$$y_3 = \begin{cases} 1, & \text{если система (20) в точке } X \text{ удовлетворяется} \\ 0, & \text{если система (20) в точке } X \text{ не удовлетворяется} \end{cases}.$$

Тогда на основании $\bar{A}$ получим таблицу истинности

Таблица 2

y_1	1	1	0	0
y_2	1	0	-	-
y_3	-	-	1	0
$f_A(Y)$	0	1	1	0

откуда

$$f_A(Y) = y_1 \bar{y}_2 + \bar{y}_1 y_3 \quad \text{и}$$

$$F_A(x) = -0,715 + 1,715 f_A[Y(x)].$$

Этот алгоритм был реализован на моделирующей машине ЭМУ-8 и были получены следующие осциллограммы (фиг.2).

Следует отметить, что алгоритмы (A) и $(\bar{A})$ реализуются в виде вычислительного устройства непрерывного действия достаточно просто, для чего требуются лишь стандартные релейные элементы, поэтому создание регуляторов на основе разработанного алгоритма не будет вызывать затруднений.

4. Случай $n = 4$.

Общий подход к реализации закона $u = F(x)$, основанный на аппроксимации поверхности переключения плоскостями и введении в алгоритм управления логики, может быть использован и для расчёта субоптимального регулятора для систем четвёртого порядка.

Трёхмерная поверхность переключения Π^* для систем 4-го порядка, матрица объекта которых имеет только действительные собственные значения, обладает теми же геометрическими свойствами, что и соответствующая поверхность переключения для систем третьего порядка. Именно, 1) она является однозначной в направлении вектора $\vec{b}$ и 2) каждая точка поверхности является эллиптической (исключая, конечно, точки, являющиеся граничными для кусков поверхности, на которых $u = -1$ и $u = 1$), т.е. поверхность оказывается локально выпуклой. Это позволяет разбить Π^* на куски, каждый из которых является выпуклой поверхностью, и аппроксимировать их наборами трёхмерных плоскостей.

Общий алгоритм для систем 4-го порядка является довольно громоздким. Если выделить в области управляемости некоторое подмножество типичных начальных условий, то общий алгоритм может быть сведён к алгоритму, который получается незначительным усложнением того, который используется для систем 3-го порядка. Проиллюстрируем это на следующей задаче.

Пусть для системы, описываемой уравнениями

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= 0,06 x_1 - 0,01 x_2 - 0,4 x_3 + 0,06 u \\ \dot{x}_3 &= -0,3 x_1 + 0,02 x_2 + 0,1 x_3 + 0,01 x_4 - 0,03 u, \\ \dot{x}_4 &= x_3 \end{aligned} \quad (21)$$

где $|u| \leq 1$, требуется рассчитать регулятор, который был бы близок к оптимальному по времени, причём область начальных значений определяется условиями: $|x_1| \leq 0,5$, $|x_2| \leq 0,08$, $|x_3| \leq 0,04$ и $|x_4| \leq 8$. Матрица A имеет следующие собственные значения: $\lambda_1 = -0,3$, $\lambda_2 = -0,2$, $\lambda_3 = 0,1$, $\lambda_4 = 0,01$

Перепишем уравнения поверхности переключения в другом виде.

Для этого сделаем следующую замену параметров:

$$\begin{aligned} x &= 2 - 3\eta, \\ y &= 3\eta - 3z, \\ z &= 3z. \end{aligned}$$

Тогда из (4) и (5) получаем уравнения поверхности Ω^2 :

$$x = x^-(z) - 2x^-(3\eta) + 2x^-(3z) \quad (22)$$

и

$$x = x^+(z) - 2x^+(3\eta) + 2x^+(3z). \quad (23)$$

Используя эти уравнения, для заданного множества начальных условий можно определить на трёхмерной поверхности Ω^2 подлежащие аппроксимации области. Для системы (21) эти области были найдены, и соответствующие неравенства приведены ниже.

Вначале всё фазовое пространство разбивалось на два полупространства плоскостью

$$0,28x_1 + 0,42x_2 + 0,17x_3 + 0,005x_4 \geq 0, \quad (24)$$

которая близка плоскости $[b, Ab, A^0b, x] = 0$ (угол между нормальными этих плоскостей мал). На двумерной координатной поверхности $z_2 = 0$, т.е. поверхности, по которой происходит движение изображающей точки после первого переключения управления, были выбраны плоскости, которым соответствуют неравенства:

$$\begin{aligned} 7,007x_1 + 22,34x_2 + 4,41x_3 + 0,225 &\geq 0, \\ -2,47x_1 - 23,28x_2 - 1,08x_3 - 21,6 &\geq 0. \end{aligned} \quad (25)$$

Область, ограниченная трёхмерной поверхностью Ω^2 , аппроксимировалась многогранным выпуклым телом, которое описывается следующими неравенствами:

$$\begin{aligned} 14,62x_1 - 7,64x_2 + 13,57x_3 - 0,0426x_4 - 0,4171 &\geq 0, \\ 3,214x_1 + 4,88x_2 + 3,68x_3 - 0,0218x_4 - 0,284 &\geq 0, \\ 23,7x_1 + 2,65x_2 - 16,05x_3 + 0,015x_4 + 0,36 &\geq 0. \end{aligned} \quad (26)$$

Поверхность Ω^2 , лежащая в множестве, дополнительном к множеству, описываемому неравенствами (25), аппроксимировалась гиперплоскостью:

$$-10,05x_1 + 2,31x_2 + 9,14x_3 - 0,415x_4 + 0,638 \geq 0. \quad (27)$$

Тогда получаем следующую логическую функцию:

- 1) если (24) удовлетворяется, (25) удовлетворяется и (26) удовлетворяется, то $u = 1$;
- 2) если (24) удовлетворяется, (25) удовлетворяется и (26) не удовлетворяется, то $u = -1$;
- 3) если (24) удовлетворяется, (25) не удовлетворяется и (27) удовлетворяется, то $u = -1$;
- 4) если (24) удовлетворяется, (25) не удовлетворяется и (27) не удовлетворяется, то $u = 1$.

В случае $0,23x_1 + 0,42x_2 + 0,17x_3 + 0,003x_4 < 0$, т.е. когда (24) не удовлетворяется, алгоритм аналогичен вышеприведённому. Заметим что плоскость (24) обеспечивает последнее переключение управления.

Введя логические переменные y_i , которые имеют тот же смысл, что и ранее, получим

$$f_1(Y) = y_1 y_2 y_3 + y_1 \bar{y}_2 \bar{y}_3$$

и

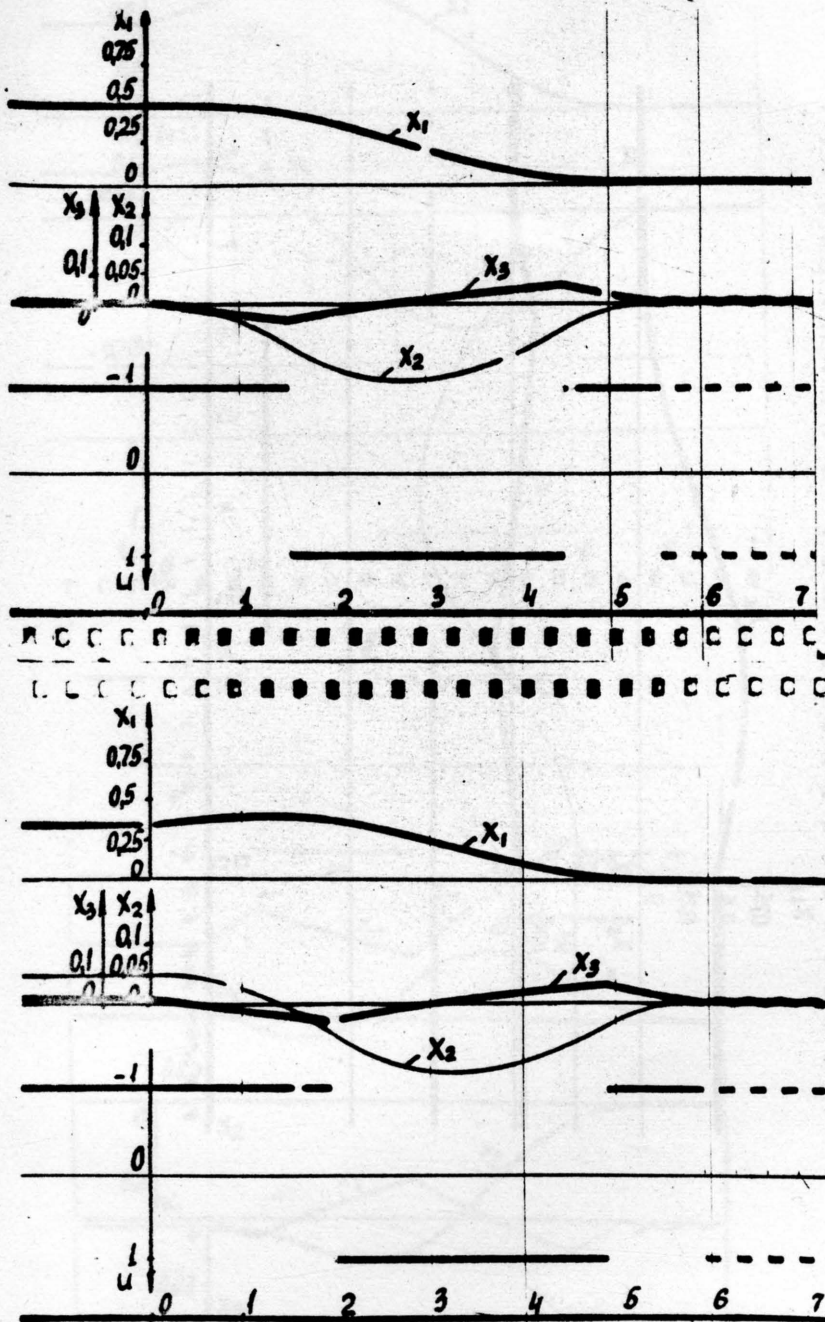
$$F_2(x) = -1 + 2f_1(Y(x)) .$$

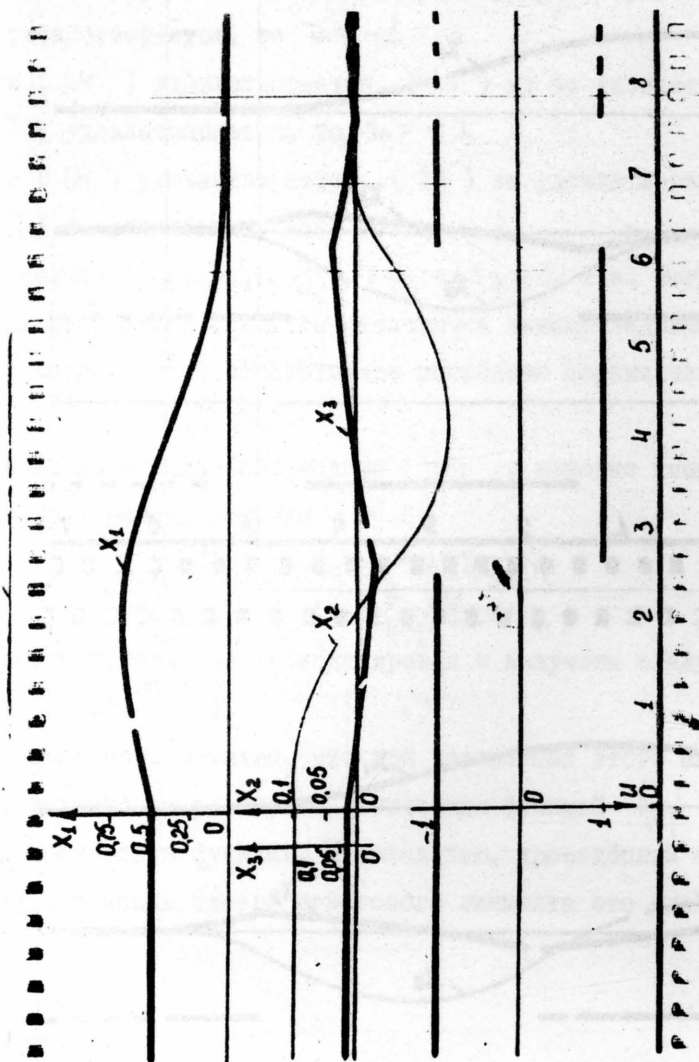
Этот алгоритм был промоделирован и получены следующие осциллограммы (рис. 3).

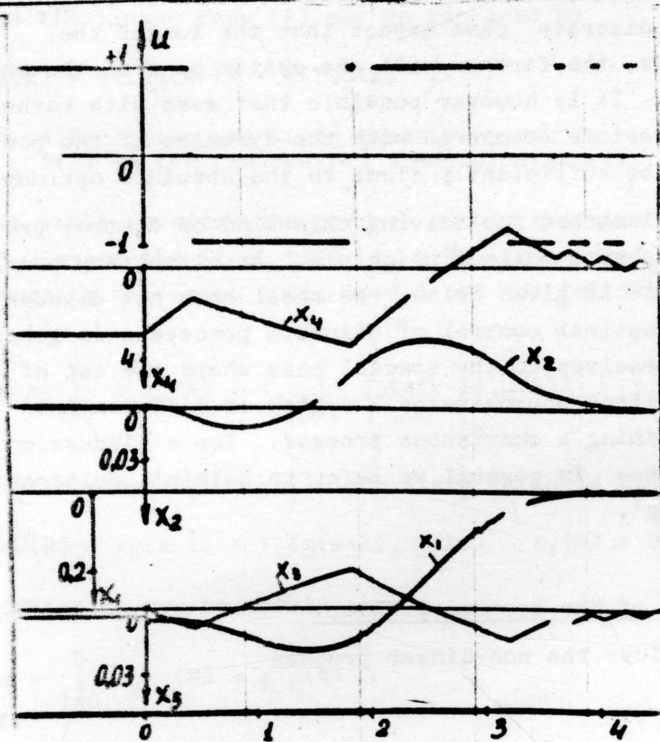
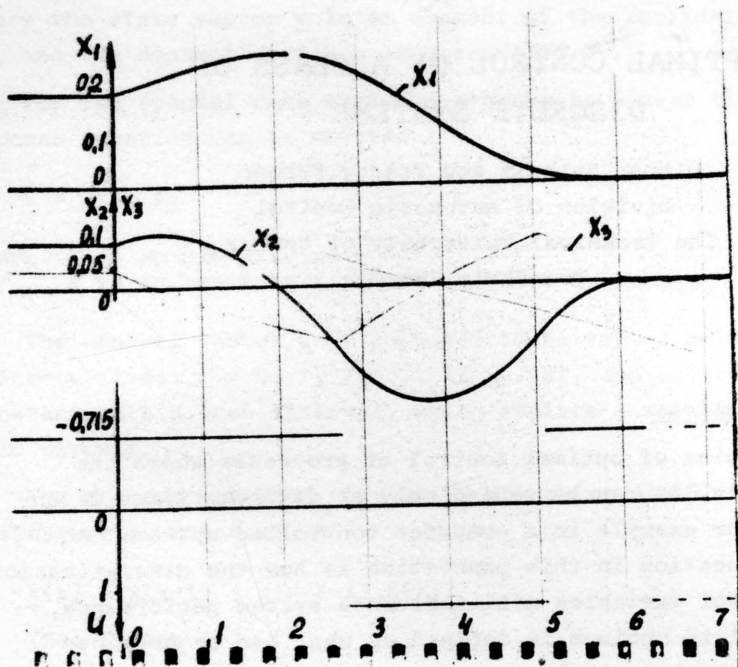
В заключение отметим, что при реализации этого алгоритма было бы полезно иметь вместо логических функций $f_1(Y)$ одну пороговую логическую функцию. К сожалению, проведённый анализ показал, что с помощью одного порогового элемента это сделать невозможно.

ЛИТЕРАТУРА.

1. А.А. Павлов, "Синтез релейных систем, оптимальных по быстродействию," Наука, М., 1966.
2. В.Г. Болтянский, "Математические методы оптимального управления," Наука, М., 1966.







OPTIMAL CONTROL OF A CLASS OF DISCRETE SYSTEMS

Ole A. Solheim and Freddy Pøhner
Division of Automatic Control
The Technical University of Norway
Trondheim, Norway

1. Introduction.

The problem of optimal control of processes where the control variables can be varied only at discrete times is encountered for example in a computer controlled system. An important question in this connection is how the discretization of the control variables will influence system performance. If the absolute optimum is defined as what can be obtained when the control variables can be varied continuously, one should for the discrete case expect that the longer the sampling periods, the farther will the system be from the absolute optimum. It is however possible that even with rather long sampling periods (compared with the dynamics of the process) one will be sufficiently close to the absolute optimum.

One powerful method for solving this kind of control problems is the discrete maximum principle. An elementary proof of this principle is given below. We shall here not discuss the problem of optimal control of discrete processes in general but restrict ourselves to the special case where the set of difference equations approximates a system of differential equations describing a continuous process. For a discussion of discrete processes in general we refer to Halkin¹, Holtzman² and Fan and Wang³.

2. Formulation of the control problem.

Let us consider the non-linear process

$$\dot{\underline{x}} = \underline{F}(\underline{x}, \underline{u}, t) \quad (1)$$

where the state vector $\underline{x}$ is an element of the Euclidean space E^n , and the control vector $\underline{u}$ an element of E^r .

For the special case that the process is linear the process equation can be written

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} \quad (2)$$

where A is an $n \times n$ matrix
 B is an $n \times r$ matrix.

The control vector $\underline{u}$ is supposed to be varied only at the discrete times $t = 0, T, 2T \dots kT \dots NT$, and is assumed constant within each interval, which implies a zero-order hold circuit.

The process equation (1) can be approximated by the difference equation

$$\underline{x}(k) - \underline{x}(k-1) = f(\underline{x}(k-1), \underline{u}(k)) \quad (3)$$

(See Figure 1.)

In the linear case (2) may be replaced by

$$\underline{x}(k) = \Phi(T)\underline{x}(k-1) + \Psi(T)\underline{u}(k) \quad (4)$$

where $\Phi(T)$ is the $n \times n$ state transition matrix
 $\Psi(T)$ is the $n \times r$ control transition matrix.

We shall consider the process within the time interval $0 \leq t \leq t_f$ where the final time $t_f = NT$ is assumed given. The performance criterion to be used is of the following form

$$q = \sum_{i=1}^n c_i x_i(N) + c_0 \sum_{k=1}^N f_0(\underline{x}(k-1), \underline{u}(k)) \quad (5)$$

By introducing an auxiliary state variable x_0 defined by

$$x_0(k) - x_0(k-1) = f_0(\underline{x}(k-1), \underline{u}(k)) \quad x_0(0) = 0 \quad (6)$$

the criterion may be written

$$q = \sum_{i=0}^n c_i x_i(N) = \underline{c}' \underline{x}(N) \quad (7)$$

where ' denotes the transpose of a vector (or matrix) and $\underline{x}(N)$ here denotes the augmented state vector $\underline{x}(N) = \text{col}(x_0(N), x_1(N) \dots x_n(N))$.

The control vector $\underline{u}(k)$ must for all $k = 1 \dots N$ belong to the given subset $U \subset E^r$.

The control problem consists now of determining the control vector $\underline{u}(k) \in U$ ($k = 1 \dots N$) that will bring the process from a given initial state $\underline{x}(0)$ to the final state $\underline{x}(N)$ along an optimal path so that the criterion q is maximized. Several methods exist for solving this problem. We shall here restrict ourselves to the discrete maximum principle.

3. The discrete maximum principle.

When the discrete maximum principle is being used, the function $f(\underline{x}(k-1), \underline{u}(k))$ in the system equation (3) must satisfy the following conditions:

- a) $f(\underline{x}(k-1), \underline{u}(k))'$ must be defined for all $(\underline{x}(k-1), \underline{u}(k)) \in E^n \times U$.
- b) $f(\underline{x}(k-1), \underline{u}(k))$ must be twice differentiable with respect to $\underline{x}$ for all $\underline{u} \in U$.
- c) $f(\underline{x}(k-1), \underline{u}(k))$ and its first and second derivatives with respect to $\underline{x}$ must be bounded over $X \times U$ for any bounded set $X \subset E^n$.

In addition to these conditions, there are for the general case a few additional conditions, especially certain convexity requirements that must be satisfied. But for the present case, where the process to be controlled is basically continuous, these additional conditions are always justified^{1,2}.

Similar to the continuous maximum principle we shall define an impulse vector $\underline{p}(k)$. If no component of the final state $\underline{x}(N)$ is prescribed, the impulse vector is defined by

$$\underline{p}(N) = \underline{c} \tag{8}$$

where $\underline{c}$ is given in (7),

$$\underline{p}(k-1) - \underline{p}(k) = \left[\frac{\partial f(\underline{x}(k-1), \underline{u}(k))}{\partial \underline{x}(k-1)} \right]' \underline{p}(k) \quad (9)$$

The Hamiltonian is defined as the scalar function

$$\begin{aligned} H(k) &= (\underline{p}(k))' f(\underline{x}(k-1), \underline{u}(k)) \\ &= (\underline{p}(k))' (\underline{x}(k) - \underline{x}(k-1)) \end{aligned} \quad (10)$$

The maximum principle now states that the optimal control vector $\hat{\underline{u}}(k) \in U$ will maximize the Hamiltonian $H(k)$ for all $k = 1, 2 \dots N$. The maximum of $H(k)$ may occur at an interior point where $\frac{\partial H}{\partial \underline{u}} = 0$, or at a boundary point, for example $\underline{u}(k) = u_{\max}$, Figure 2.

If one or more components of the final state $\underline{x}(N)$ are given, for example $x_1(N)$, then the corresponding impulse function $p_1(N)$, is not specified. Otherwise the problem is the same.

We shall here give a simple proof of the discrete maximum principle. A more complete and mathematically more rigorous proof may be found in for example Halkin¹.

Let us denote the optimal control vector by $\hat{\underline{u}}(k)$, the optimal state vector by $\hat{\underline{x}}(k)$ and the corresponding impulse vector by $\hat{\underline{p}}(k)$. We shall demonstrate the principle for a linearized system that is valid near the optimal state.

We have then

$$\bar{\underline{x}}(k) = \hat{\underline{x}}(k) + \delta \underline{x}(k) \quad (11)$$

The condition for optimality is

$$\underline{c}' (\hat{\underline{x}}(N) - \underline{x}(0)) \geq \underline{c}' (\bar{\underline{x}}(N) - \underline{x}(0)) \quad (12)$$

or

$$\underline{c}' \cdot \hat{\underline{x}}(N) \geq \underline{c}' (\hat{\underline{x}}(N) + \delta \underline{x}(N)) \quad (13)$$

Since by definition $\underline{p}(N) = \underline{c}$ this can be written

$$(\underline{p}(N))' \cdot \delta \underline{x}(N) \leq 0 \quad (14)$$

Equation (14) can be given a geometrical interpretation. Let $W(N)$ represent the set of reachable states at the final time N . An optimal path will always end on the boundary of $W(N)$. Equation (14) states that at this point it must be possible to draw a tangent hyperplane whose normal is $\underline{p}(N)$ or $\underline{c}$. This is illustrated for the two-dimensional case in Figure 3. From the same figure we note that a requirement is that W is convex in the $\underline{p}(N)$ -direction. As mentioned above this convexity requirement is always satisfied when the process is basically continuous. A closer discussion of this directional convexity requirement for the more general case is given by Holtzman².

Introducing $\underline{\bar{x}}(k)$ as defined by (11) into the process equation (3) gives

$$\underline{\bar{x}}(k+1) - \underline{\bar{x}}(k) = \underline{f}(\underline{\hat{x}}(k), \underline{u}(k+1)) + \left(\frac{\partial}{\partial \underline{x}} \underline{f}(\underline{x}(k), \underline{\hat{u}}(k+1)) \right) \Big|_{\underline{x}=\underline{\hat{x}}} \cdot \delta \underline{x}(k) \quad (15)$$

This equation can be simplified to

$$\delta \underline{x}(k+1) - \delta \underline{x}(k) = \left(\frac{\partial}{\partial \underline{x}} \underline{f}(\underline{x}(k), \underline{\hat{u}}(k+1)) \right) \Big|_{\underline{x}=\underline{\hat{x}}} \cdot \delta \underline{x}(k) \quad (16)$$

Transposing equation (9) gives

$$(\underline{\hat{p}}(k))' - (\underline{\hat{p}}(k+1))' = (\underline{\hat{p}}(k+1))' \left(\frac{\partial}{\partial \underline{x}} \underline{f}(\underline{x}(k), \underline{\hat{u}}(k+1)) \right) \Big|_{\underline{x}=\underline{\hat{x}}} \quad (17)$$

If we now multiply this equation with $\delta \underline{x}(k)$ and (16) with $(\underline{\hat{p}}(k+1))'$ the right-hand side in the two equations will become identical, and we arrive at the relation

$$(\underline{\hat{p}}(k+1))' \delta \underline{x}(k+1) = (\underline{\hat{p}}(k))' \delta \underline{x}(k) \quad (18)$$

This result combined with (14) states that an optimal path has the property that

$$(\underline{\hat{p}}(k))' \delta \underline{x}(k) \leq 0 \quad \text{for all } k = 1 \dots N \quad (19)$$

This is illustrated for the two-dimensional case in Figure 4.

It now remains to show that maximizing the Hamiltonian $H(k)$ leads to the above condition (19).

Maximizing $H(k)$ means that

$$(\hat{p}(k))' (\hat{x}(k) - \hat{x}(k-1)) \geq (\hat{p}(k))' (\bar{x}(k) - \hat{x}(k-1)) \quad (20)$$

which can be simplified to

$$(p(k))' \delta x(k) \leq 0 \quad (21)$$

Since (21) is identical with (19) the control vector $u(k)$ that will maximize $H(k)$ for all k will also maximize the criterion q .

If one or more components of the final state $x(N)$ are prescribed, then the corresponding components of the impulse vector $p(N)$ are not specified, but have to be chosen so that the prescribed state is reached. From condition (14) we note that if for example $x_1(N)$ is prescribed then $\delta x_1(N) = 0$, and $p_1(N)$ may take any value.

4. Computational aspects of the optimal control problem.

Let us first consider the special case of a linear process. The system of difference equations may be written as given in (4). The vector valued function $f(x(k-1), u(k))$ becomes in this case

$$f(x(k-1), u(k)) = (\Phi - I)x(k-1) + \Psi u(k) \quad (22)$$

Let us further assume that the criterion q is quadratic in $u(k)$:

$$q = \sum_{i=1}^n c_i x_i(N) + c_0 \sum_{k=1}^N (u(k))' R u(k) \quad (23)$$

where R is a positive definite matrix.

The adjoint system is given by

$$p(k-1) - p(k) = \left(\frac{\partial f(x(k-1), u(k))}{\partial x(k-1)} \right)' p(k) \quad (24)$$

which can be developed to

$$\left. \begin{aligned} \underline{p}(k-1) &= \Phi' \underline{p}(k) \\ \underline{p}_0(k-1) &= \underline{p}_0(k) \end{aligned} \right\} \quad (25)$$

The boundary values of the adjoint system are given as

$$\underline{p}_i(N) = c_i \quad (i = 0, 1 \dots n) \quad (26)$$

under the assumption that no component of $\underline{x}(N)$ is prescribed.

We note that the adjoint system is completely autonomous in this case, and may be determined independent of the process state $\underline{x}$. The only connection to the process state is through an eventually prescribed final state $\underline{x}(N)$, since in this case the boundary values $\underline{p}(N)$ will depend on $\underline{x}(N)$. This of course leads to a two-point boundary value problem.

The Hamiltonian for the present linear system is

$$H(k) = \underline{p}(k)' (\Phi \underline{x}(k-1) + \Psi \underline{u}(k)) + \underline{p}_0(k) (\underline{x}_0(k-1) + \underline{u}(k))' \underline{R} \underline{u}(k) \quad (27)$$

where $\underline{x}_0$ is an auxiliary state variable defined by

$$\left. \begin{aligned} \underline{x}_0(k) &= \underline{x}_0(k-1) + (\underline{u}(k))' \underline{R} \underline{u}(k) \\ \underline{x}_0(0) &= 0 \end{aligned} \right\} \quad (28)$$

The maximum of $H(k)$ can be found by differentiation

$$\frac{\partial H(k)}{\partial \underline{u}(k)} = \Psi' \underline{p}(k) + 2 \underline{R} \underline{u}(k) \underline{c}_0 = \underline{0} \quad (29)$$

or

$$\underline{u}(k) = \frac{1}{2 \underline{c}_0} \underline{R}^{-1} \Psi' \cdot \underline{p}(k) \quad (30)$$

We note that once the adjoint system is determined, the optimal control vector can easily be found.

Let us next consider a criterion which is linear in $\underline{u}$:

$$q = \sum_{i=1}^n c_i x_i(T) + c_0 \sum_{k=1}^N \underline{a}' \underline{u}(k) \quad (31)$$

The adjoint system will be the same as developed above in (25, 26). The Hamiltonian becomes

$$\begin{aligned}
 H(k) &= (\underline{p}(k))' (\Phi \underline{x}(k-1) + \Psi \underline{u}(k)) + \underline{p}_0(k) (\underline{x}_0(k) + \underline{\alpha}' \underline{u}(k)) \\
 &= (\underline{p}(k))' \Phi \underline{x}(k-1) + \underline{p}_0(k) \underline{x}_0(k) + [(\underline{p}(k))' \Psi + \underline{c}_0 \underline{\alpha}'] \underline{u}(k)
 \end{aligned} \quad (32)$$

* If now the set U of admissible controls is given as an hypercube, the conditions for a maximum of H(k) may be given in the form

$$\left. \begin{aligned}
 u_i(k) &= (u_i)_{\max} \quad \text{for } \left(\sum_{r=1}^n p_r(k) \psi_{ri} + c_0 \alpha_i \right) > 0 \\
 u_i(k) &= (u_i)_{\min} \quad \text{for } \left(\sum_{r=1}^n p_r(k) \psi_{ri} + c_0 \alpha_i \right) < 0
 \end{aligned} \right\} \quad (33)$$

This implies a bang-bang system, where the switching points may be found once the adjoint system has been determined.

The computation of the optimal control vector in the linear case presents no great difficulties, as it is in this case possible to express the vector valued function $\underline{f}(\underline{x}(k-1), \underline{u}(k))$ explicitly in terms of the transition matrices.

A different situation arises in the general non-linear case. If the length of the sampling interval is very short (N is large), we may be able to find a function $\underline{f}(\underline{x}(k-1), \underline{u}(k))$ that represents a satisfactory approximation to the continuous process. If on the contrary the sampling interval is long, that is N is a low number, we may have great difficulties in finding a $\underline{f}(\underline{x}(k-1), \underline{u}(k))$ that will approximate the continuous behavior of the process in a satisfactory way. It is however possible to get around this difficulty. Looking at the Hamiltonian as defined in (10):

$$H = (\underline{p}(k))' (\underline{x}(k) - \underline{x}(k-1)) \quad (10)$$

we note that we need the values of state vector and impulse vector at the time $t = 1, 2 \dots N$. The state vector can be obtained by integrating the process equation (1) with sufficiently small discretization intervals. One should be aware of the fact that these intervals have nothing to do with the control intervals that we have considered above, but is merely

a consequence of using a digital computer to integrate the differential equations. To this end we may approximate the process equation (1) by

$$\underline{x}(j) - \underline{x}(j-1) = \underline{g}(\underline{x}(j-1), \underline{u}(j)) \quad j = 1 \dots mN \quad (34)$$

where $\underline{g}(\underline{x}(j-1), \underline{u}(j))$ for example may be chosen as $\Delta \cdot \underline{F}(\underline{x}(j-1), \underline{u}(j))$ with Δ as the integrating interval.

It remains to determine the impulse vector $\underline{p}(k)$ as defined in (9). We need here the expression

$$\frac{\partial \underline{f}(\underline{x}(k-1), \underline{u}(k))}{\partial \underline{x}(k-1)}, \text{ but since } \underline{f}(\underline{x}(k-1), \underline{u}(k)) \text{ is not explicitly}$$

known, this expression cannot be found. We shall therefore compute the impulse vector in the same way as the state vector. From (9) we get

$$\underline{p}(j-1) - \underline{p}(j) = \left(\frac{\partial \underline{g}(\underline{x}(j-1), \underline{u}(j))}{\partial \underline{x}(j-1)} \right)' \cdot \underline{p}(j) \quad (35)$$

Since $\underline{g}$ is known from (34), the derivative of $\underline{g}$ used in (35) is also known. It has however to be shown that the impulse vector determined by (35) will have the same value at the points $k = 0, 1 \dots N$ as the impulse vector given by (9). For the sake of simplicity we shall show this for the one-dimensional case.

Let $T = m\Delta$ where T is the control interval and Δ the integration interval. (See Figure 5.) From (35) we get

$$\underline{p}(k-1) - \underline{p}(k) = \sum_{j=m(k-1)}^{m(k-1)+1} \frac{\partial \underline{g}(\underline{x}(j-1), \underline{u}(j))}{\partial \underline{x}(j-1)} \underline{p}(j) \quad (36)$$

which may be developed to

$$\begin{aligned} \underline{p}(k-1) - \underline{p}(k) = & \{h(mk-1) + h(mk-2)(1+h(mk-1)) + h(mk-3)(1+h(mk-1)) + \\ & + h(mk-2)(1+h(mk-1))\} + h(mk-3)(1+\dots) + \dots \} \underline{p}(k) \end{aligned} \quad (37)$$

$$\text{where } h(i) = \frac{\partial \underline{g}(\underline{x}(i), \underline{u}(i+1))}{\partial \underline{x}(i)}$$

From (34) and (3) we get

$$x(k) - x(k-1) = \left(\sum_{j=m(k-1)+1}^{mk} g(x(j-1), u(j)) \right) = f(x(k-1), u(k)) \quad (38)$$

Based on this expression we may develop the derivative used in (9):

$$\begin{aligned} \frac{\partial f(x(k-1), u(k))}{\partial x(k-1)} &= h(mk-m) + h(mk-m+1) \{1 + h(mk-m)\} + \\ &+ h(mk-m+2) \{1 + h(mk-m) + h(mk-m+1) \{1 + h(mk-m)\}\} + \dots \end{aligned} \quad (39)$$

Because of the symmetry the right-hand side of (39) is identical with the expression in brackets on the right-hand side of (37). We may therefore conclude that the impulse vector determined from (35) is identical with the impulse vector determined from (9) at the points $k = 1, 2 \dots N$.

The structure of a computer program for determining the optimal control vector based on the discrete maximum principle is shown in Figure 6.

5. Numerical results.

Example 1

The first example is the linear process

$$\dot{\underline{x}} = \begin{bmatrix} -1 & 0 \\ 5 & -3 \end{bmatrix} \underline{x} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} u_1$$

The transition matrices in the equation

$$\underline{x}(k) = \Phi(T)\underline{x}(k-1) + \Psi(T)u(k)$$

becomes

$$\Phi(T) = \begin{bmatrix} e^{-T} & 0 \\ \frac{5}{2}(e^{-T} - e^{-3T}) & e^{-3T} \end{bmatrix}$$

$$\Psi(T) = \begin{bmatrix} 2(1-e^{-T}) \\ 5(\frac{2}{3} - e^{-T} + e^{-3T}) \end{bmatrix}$$

The performance criterion to be maximized is

$$q = x_2(N) - \sum_{k=1}^N 0.5 \cdot T(u_1(k))^2$$

Final time $t_f = NT = 4$

The adjoint system $p(k)$ is determined from (25) together with the boundary conditions: $p_0(N) = -0.5T$, $p_1(N) = 0$, $p_2(N) = 1$.

The optimal control variable is then determined from (30).

In Figure 7 is shown the optimal control variable for $N = 2 - 4 - 8$. For the sake of comparison, the optimal control variable for the continuous case is also shown.

The value of the performance criterion for different values of N is given in Table 1 below.

Table 1.

N	2	4	8	20	∞
q	1.85	1.93	1.99	2.07	2.08

It is worth while to note that the deviation from the absolute optimum (obtained in the continuous case) is rather small even with a long sampling period.

Example 2

This example is a chemical reactor where the following reaction takes place:



where A is the reactant, R the valuable product and S a by-product. We want to maximize the final concentration of the valuable product R .

With x_1 as the concentration of A, x_2 the concentration of R and x_3 the concentration of S, the process equations are

$$\dot{x}_1 = -k_1(u)x_1 - k_2(u)x_1$$

$$\dot{x}_2 = k_1(u)x_1$$

$$\dot{x}_3 = k_2(u)x_1$$

The reaction rate constants are given by the Arrhenius equation

$$k(u) = A \exp(E/(273+u))$$

where u is the temperature ($^{\circ}\text{C}$).

The following numerical values are used

$$x_1(0) = 8.0 \quad x_2(0) = 0 \quad x_3(0) = 0$$

$$k_1 = 3.3 \cdot 10^6 \exp(-5000/(273+0))$$

$$k_2 = 11 \cdot 10^{12} \exp(-10000/(273+0))$$

$$\text{Final time } t_f = NT = 2$$

Maximum temperature 60°C or 70°C .

The criterion is

$$q = x_2(N)$$

In Figure 8 are shown optimal values of the control variable u for $N = 1, 3$ and 50 . The value of the performance criterion for different values of N is given below in Table 2.

N	1	2	3	5	10	50
q $u_{\max}=60$	4.281	4.452	4.486	4.488	4.497	4.499
q $u_{\max}=70$	4.281	4.452	4.508	4.552	4.561	4.566

Table 2.

6. Concluding remarks.

The class of discrete systems dealt with in this paper is of special interest in connection with com-

puter controlled processes. Besides the determination of the optimal control law, it is here of interest to find out how the length of the sampling period will influence upon system performance. The examples given indicate that even with long sampling periods (as compared with process dynamics) the fall in performance as compared with the continuous case is low. Studies of other processes not referred here also indicate the same.

References:

1. H. Halkin: A maximum principle of Pontryagin type for systems described by nonlinear difference equations. SIAM Vol. 4, no. 1 (Febr. 1966) pp. 90-111.
2. J.M. Holtzman: Convexity and the maximum principle for discrete systems. IEEE Transactions on Automatic Control Vol. AC-11, no. 1 (Jan 1966) pp. 30-35.
3. L. Fan and C. Wang: The discrete maximum principle. J. Wiley & Sons. New York-London-Sidney 1964.
4. F. Pøhner: Optimal control of nonlinear discrete systems. Thesis (in Norwegian), Division of Automatic Control, The Technical University of Norway, Trondheim, 1968.

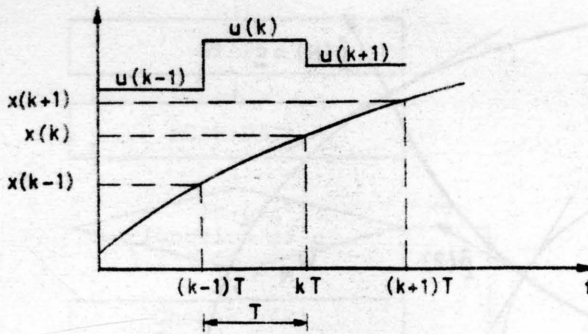


Figure 1: Discrete variables

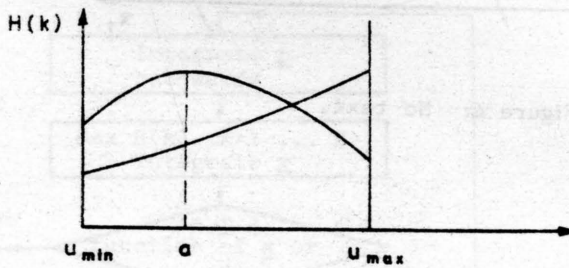
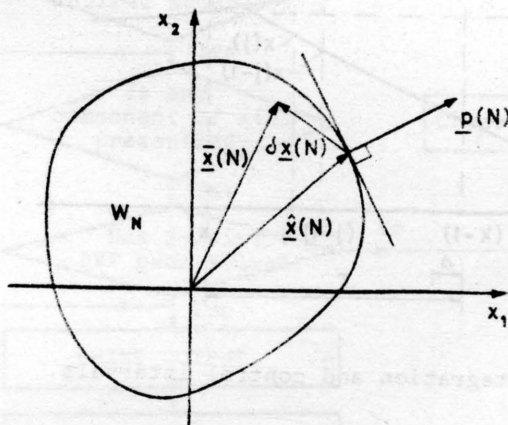
Figure 2: Maximum of $H(k)$ may occur at point a or at $u_{\max}$.

Figure 3: No text.

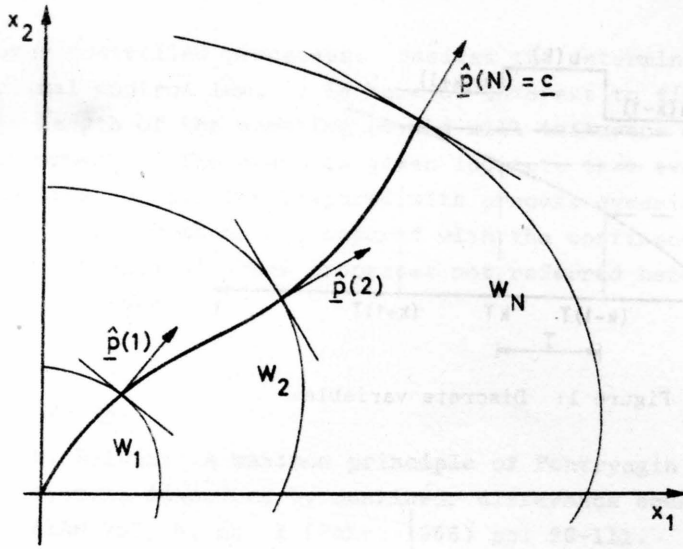


Figure 4: No text.

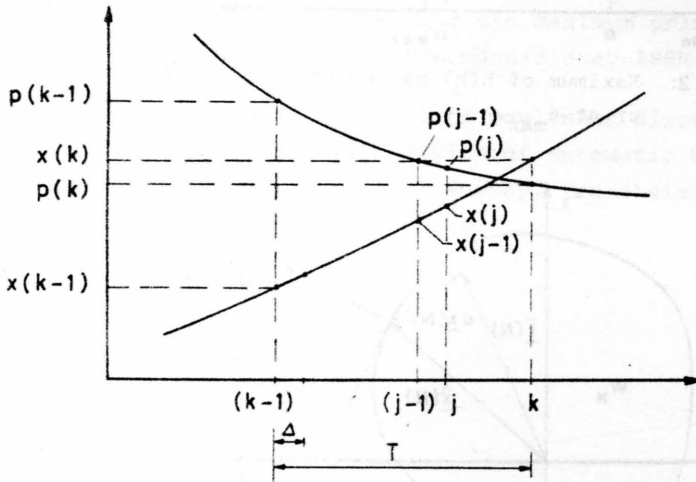


Figure 5: Integration and control intervals.

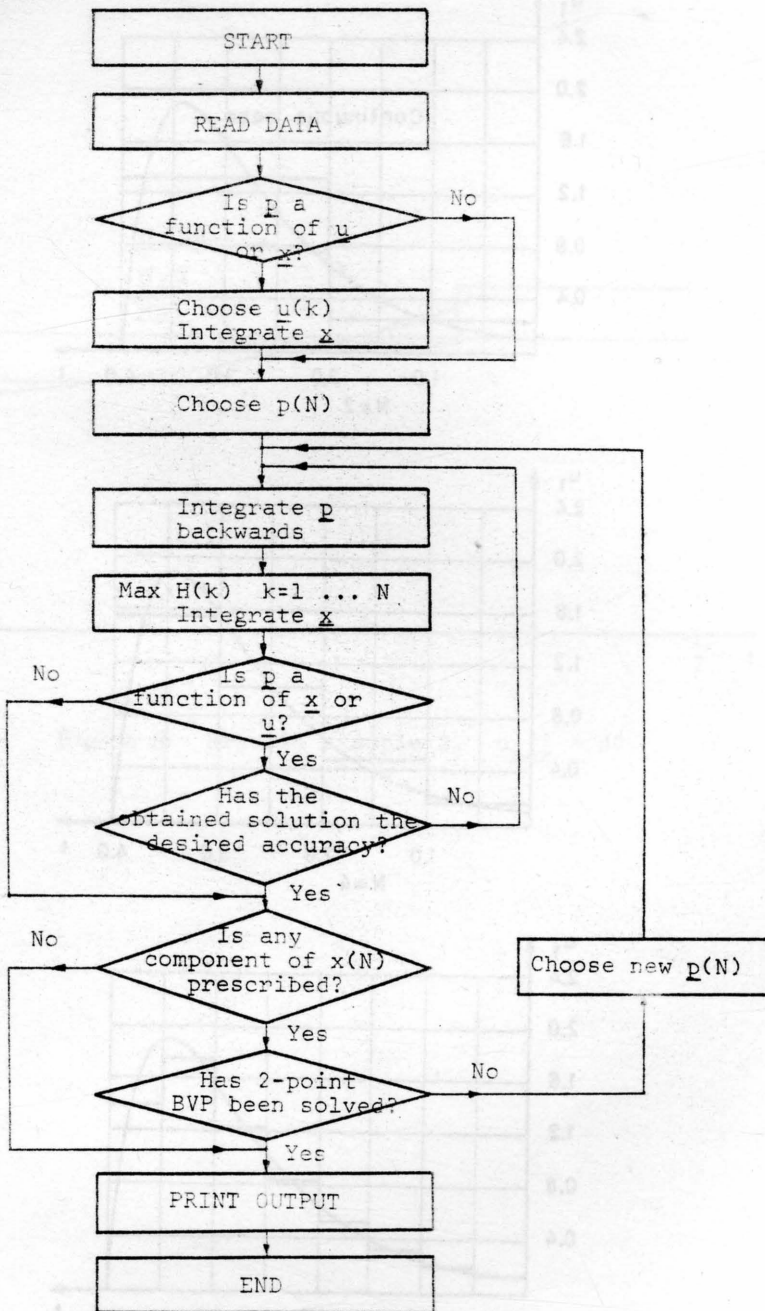


Figure 6: Structure of computer program.

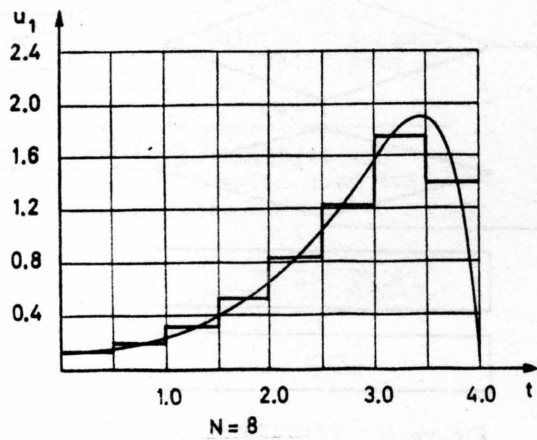
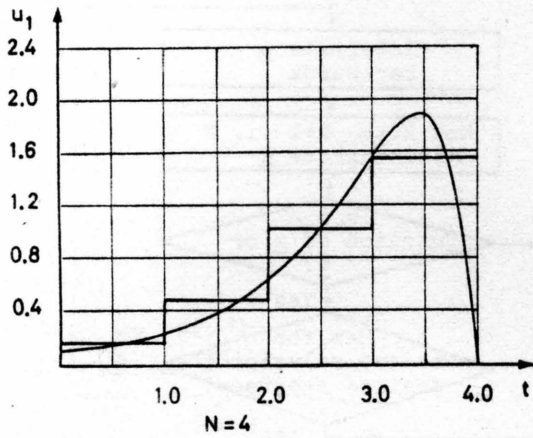
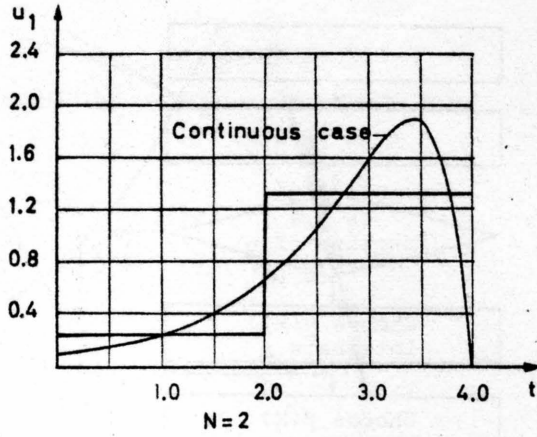


Figure 7: Results example 1.

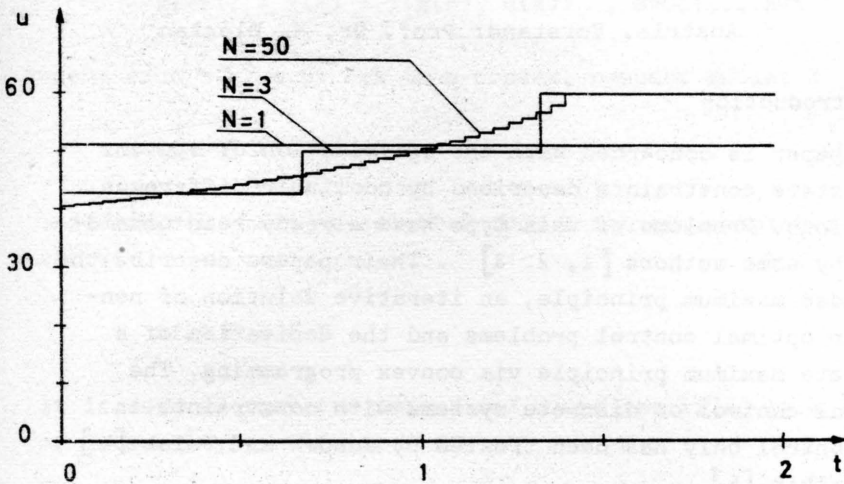


Figure 8: Results example 2. $u_{\max} = 60$.

OPTIMAL CONTROL of NONLINEAR DISCRETE SYSTEMS

Institut für Angewandte Mathematik Technische Hochschule Graz,
Austria. Vorstand: Prof. Dr. H. Florian

1. Introduction

This paper is concerned with the optimization of systems with state constraints described by nonlinear difference equations. Problems of this type have already been considered by some authors [1, 2, 3]. Their papers describe the extended maximum principle, an iterative solution of nonlinear optimal control problems and the derivation of a discrete maximum principle via convex programming. The optimal control of discrete systems with constraints in the control only has been treated by Jordan and Polak [4] and Halkin [5].

In the present paper the nonlinear discrete optimal control problem is considered as an approximation of a nonlinear continuous problem with a convex objective function to be minimized. The constraints for the state and the control are assumed to be convex functions. The discrete problem is then considered as minimization of a convex function subject to nonlinear equality constraints and convex inequality constraints. For this problem a theorem for convex disjoint sets and the Farkas lemma are used to derive necessary conditions for an optimal solution. In the linear case these conditions are necessary and sufficient. Furthermore these conditions represent a discrete maximum principle for the optimal control problem with state constraints.

2. Statement of the problem

The discrete problem will be considered as follows. The state vector $x(k)$ will be an element of an Euclidean space E^n at time k , the corresponding control vector $u(k)$ will be

an element of an Euclidean space E^r and the time k will assume the discrete values $0, 1, \dots, N-1$. The initial value $x(0)$ is specified as well as the terminal time N .

The system satisfies the difference equations

$$x(k+1) = x(k) + f(x(k), u(k)) \quad , \quad k=0, 1, \dots, N-1 \quad (1)$$

where each $x(k)$ must lie in a convex, compact subset X

$$x(k) \in X \subset E^n \quad (2)$$

and each $u(k)$ must be selected from a convex, compact subset U

$$u(k) \in U \subset E^r \quad (3)$$

It is assumed that the sets X and U are specified by a system of inequality constraints

$$X = \{x \mid h^1(x) \leq 0\} \quad (4)$$

and

$$U = \{u \mid h^2(u) \leq 0\} \quad (5)$$

$h^1(x)$ is a function from E^n to E^l with $h^1 \in C^1$ and convex on E^n and $h^2(u)$ is a function from E^r to E^m with $h^2 \in C^1$ and convex on E^r . The sets X and U are convex by the convexity of $h^1(x)$ and $h^2(u)$.

It is further assumed that $f(x, u)$ is a function from $E^n \times U$ to E^n with $f \in C^1$ on $E^n \times U$. Trajectories and controls which satisfy (2) and (3) are called admissible. The cost of a transition from the state $x(k)$ to the state $x(k+1)$ caused by the control $u(k)$ is given by $f_0(x(k), u(k))$. The cost of operating the system from the initial to the terminal state is therefore

$$\sum_{k=0}^{N-1} f_0(x(k), u(k)) \quad (6)$$

Let $x_0(k)$ be the cost of operating the system from time zero to time k . Then $x_0(k)$ satisfies the difference equation

$$x_0(k+1) = x(k) + f_0(x(k), u(k)) \quad (7)$$

It is assumed that $f_0 \in C^1$ on $E^n \times U$.

The optimal control problem for the system described by (1) with given initial state $x(0)$ and terminal time N is to determine the vectors $u(k)$ ($k=0, 1, \dots, N-1$) and $x^0(k)$ ($k=1, \dots, N$) satisfying (2) and (3) so as to minimize the performance index (6). The results obtained hold for the more general case where f_0 and f depend explicitly on time k . For simplification it is assumed however that f_0 and f do not depend explicitly on k .

Now the $n \times n$ Jacobian matrix

$$F(k) = \frac{\partial f(x(k), u(k))}{\partial x} \quad k=0, \dots, N-1 \quad (8)$$

and the $n \times r$ Jacobian matrix

$$G(k) = \frac{\partial f(x(k), u(k))}{\partial u} \quad k=0, \dots, N-1 \quad (9)$$

of f are defined. The state vector is augmented by $x_0(k)$, the vector f is augmented by f_0 . Thus these $n+1$ - vectors are

$$\tilde{x}(k) = (x_0(k), x(k))' \quad (10)$$

and

$$\tilde{f} = (f_0, f)' \quad (11)$$

The matrix

$$\tilde{F}(k) = \frac{\partial \tilde{f}(x(k), u(k))}{\partial \tilde{x}} \quad (12)$$

is the $(n+1) \times (n+1)$ Jacobian and

$$\tilde{G}(k) = \frac{\partial \tilde{f}(x(k), u(k))}{\partial u} \quad (13)$$

is the $(n+1) \times r$ Jacobian of $\tilde{f}$. For simplification a specific control $(u(0), u(1), \dots, u(N-1))$ and the corresponding trajectory $(x(0), x(1), \dots, x(N))$ are denoted by a vector $y \in E^s$, where $s = (N+1)n + Nr$ and the prime denotes transpose of a column vector.

$$y = (x'(0), x'(1), \dots, x'(N), u'(0), \dots, u'(N-1))' \quad (14)$$

$Y \subset E^s$ is the direct product of the sets X and U

$$Y = \prod_{k=0}^N X(k) \times \prod_{k=0}^{N-1} U(k) \quad \text{with } X(k) = X, U(k) = U \quad (15)$$

Y is a convex and compact set, since X and U are convex and compact. The ordered pair of vectors $x(k)$ and $u(k)$ is denoted by $y(k)$

$$y(k) = \begin{pmatrix} x(k) \\ u(k) \end{pmatrix} \quad (16)$$

Now the objective function (6) can be represented by

$$\phi(y) = \sum_{k=0}^{N-1} f_0(y(k)) \quad (17)$$

The $N \times n = t$ equations (1) are finally written as

$$f^x(k) = f(x(k), u(k)) + x(k) - x(k+1) = 0 \quad k=0, \dots, N-1$$

or

$$f^x(y) = 0 \quad (18)$$

With this notation the minimization problem can be stated as follows

$$\phi(y^0) = \min_y \left\{ \phi(y) \mid f^x(y) = 0, y \in Y \right\} \quad (19)$$

y^0 is the optimal solution of the discrete control problem.

The augmented s' -dimensional vector $\tilde{y}$ ($s' = (n+1)(N+1) + rN$) is

$$\tilde{y} = (\tilde{x}'(0), \tilde{x}'(1), \dots, \tilde{x}'(N), u'(0), \dots, u'(N-1))' \quad (20)$$

The set of admissible points y is

$$M = \{y \mid f^x(y) = 0, y \in Y\} \quad (21)$$

M is a convex set only if $f^x(y)$ is linear in y , that is $f(x, u)$ is linear in x and u . If components of f are non-linear in x or u , M is a nonconvex set.

By considering the linearized function $f^x(y)$ around the solution y^0

$$g(y, y^0) = f^x(y^0) + f_y^x(y - y^0) \quad (22)$$

the set of admissible points

$$M' = \{y \mid g(y, y^0) = 0, y \in Y\} \quad (23)$$

is convex. Corresponding to (11) the function (18) is augmented and denoted by $\tilde{f}^x(y)$. Its $t' \times s'$ Jacobian matrix ($t' = (n+1)N$) at $\tilde{y} = y^0$ is

$$A = \left(\frac{\partial \tilde{f}_i^x}{\partial y_j} \right) \quad i = 1, \dots, t', \quad j = 1, \dots, s' \quad (24)$$

With relations (12) and (13) the matrix A is

$$A = \begin{pmatrix} \tilde{F}(0)+I & -I & & \tilde{G}(0) & 0 & 0 \\ 0 & \tilde{F}(1)+I & -I & 0 & \tilde{G}(1) & \\ & & & & & \\ 0 & & \tilde{F}(N-1)+I & -I & 0 & \tilde{G}(N-1) \end{pmatrix} \quad (25)$$

A contains a $t' \times t'$ lower triangular matrix with elements -1 along its diagonal. Therefore the rank of A is t' . The constraints for the state (4) and the control (5) are written as the $N(1+m)$ vector

$$h(y) = \begin{pmatrix} h^1(x(1)) \\ \vdots \\ h^1(x(N)) \\ h^2(u(0)) \\ \vdots \\ h^1(u(N-1)) \end{pmatrix} \quad (26)$$

Furthermore the $N(1+m) \times s'$ matrix B is defined. The j^{th} row of B is $\text{grad}_y h_j(y)$ ($j = 1, \dots, N(1+m)$), if the j^{th} component of $h(y^0) = 0$ and zero if the j^{th} component of $h(y^0) < 0$.

3. Conditions for an optimal solution

Now consider a small deviation $\delta \tilde{y}$ at $\tilde{y}^0$ which is given by $A \cdot \delta \tilde{y} = 0$ and $B \cdot \delta \tilde{y} \leq 0$. $\delta \tilde{y} = (\delta \tilde{x}'(0), \dots, \delta \tilde{x}'(N), \delta u'(0), \dots, \delta u'(N-1))'$ is element of a convex subset T of $E^{s'}$ which contains the origin and is defined as follows

$$T = \{ \delta \tilde{y} \mid A \cdot \delta \tilde{y} = 0, B \cdot \delta \tilde{y} \leq 0 \} \quad (27)$$

An admissible vector $y^0 \in M'$ is a relative minimum if

$$\phi(y^0 + \alpha \delta y) \geq \phi(y^0) \quad (28)$$

for every admissible deviation $\delta y \in \bar{T}$ and sufficiently small α . $\bar{T}$ is a projection of T obtained by deleting the rows with f_0 in A and B . From (28) and the subsets W and S of E^{s+1} defined as follows, conditions for a relative minimum are obtained.

$$W = \left\{ z^1 = \begin{pmatrix} z_0^1 \\ y^1 \end{pmatrix} \mid y^1 \in M, z_0^1 > \phi(y^1) \geq \phi(y^0) \right\} \quad (29)$$

W is called the set of all reachable points.

$$S = \left\{ z^2 = \begin{pmatrix} z_0^2 \\ y^0 + \delta y \end{pmatrix} \mid \delta y \in \bar{T}, z_0^2 = \phi(y^0) \right\} \quad (30)$$

Let $z^1 \in W$. If $y^1 - y^0 \notin \bar{T}$, then $z^1 \notin S$. If $y^1 - y^0 = \delta y \in \bar{T}$, then by assumption it is $z_0^1 > \phi(y^0)$ and therefore $z^1 \notin S$. Thus W and S have no point in common. For the nonlinear problem and for nonconvex f_0 these sets are not necessarily convex and hence not separated. Now the sets W' and S' are defined in the same way as W and S but with respect to the linearized problem and for convex f_0 . Then the sets W' and S' are convex. Halkin [5] proved that if the sets W and S

have no point in common the sets W' and S' are separated by hyperplane. It exists a vector $b = \begin{pmatrix} a_0 \\ a \end{pmatrix} \neq 0$ such that

$$a' \cdot (y^0 + \delta y) + a_0 \phi(y^0) \leq a' \cdot y^1 + a_0 \phi(y^1) \quad (31)$$

For $y^1 = y^0$ it follows

$$a' \cdot \delta y \leq 0 \quad (32)$$

and for the augmented system

$$\tilde{a}' \cdot \delta \tilde{y} \leq 0 \quad (33)$$

at a relative minimum y^0 of $\phi(y)$. In [5] it is pointed out that it is incorrect to conclude that W' and S' are disjoint if the sets W and S defined for the nonlinear problem are disjoint. Hence the optimal solution of a nonlinear optimization problem is not necessarily the optimal solution of the linearized problem.

4. Optimal Control Problem

The relation (33) derived from some properties of disjoint convex sets and the constraints (26) are now used with the Farkas theorem to state necessary conditions for the optimal solution of the control problem. Equation (18) describes the system dynamics whereas (26) represents the constraints for state and control. The system dynamics are

$$A \cdot \delta \tilde{y} = \begin{pmatrix} (\tilde{F}(0)+I) \cdot \delta \tilde{x}(0) - \delta \tilde{x}(1) + \tilde{G}(0) \cdot \delta u(0) \\ \vdots \\ (\tilde{F}(k)+I) \cdot \delta \tilde{x}(k) - \delta \tilde{x}(k+1) + \tilde{G}(k) \cdot \delta u(k) \\ \vdots \\ (\tilde{F}(N-1)+I) \cdot \delta \tilde{x}(N-1) - \delta \tilde{x}(N) + \tilde{G}(N-1) \cdot \delta u(N-1) \end{pmatrix} \quad (34)$$

From (34) it is obvious that $\delta \tilde{x}(k)$ satisfies the difference equation

$$\delta \tilde{x}(k+1) = \delta \tilde{x}(k) + \tilde{F}(k) \cdot \delta \tilde{x}(k) + \tilde{G}(k) \cdot \delta u(k) \quad (35)$$

According to (33) there is no $\delta \tilde{y} \in T$ which satisfies $\tilde{a}' \cdot \delta \tilde{y} > 0$ and the system

$$\begin{aligned} A \cdot \delta \tilde{y} &= 0 \\ \tilde{a}' \cdot \delta \tilde{y} &> 0 \end{aligned} \quad (36)$$

has therefore no solution $\delta \tilde{y} \in T$. Thus by the theorem of Farkas [6, 7] the system

$$A' \cdot \tilde{p} = \tilde{a} \quad (37)$$

has a solution $\tilde{p} \in E^{N(n+1)}$. The system (37) is

$$A' \cdot \tilde{p} = \begin{pmatrix} (\tilde{F}(0)+I)' \cdot \tilde{p}(1) \\ \vdots \\ (\tilde{F}(k)+I)' \cdot \tilde{p}(k+1) - \tilde{p}(k) \\ \vdots \\ (\tilde{F}(N-1)+I)' \cdot \tilde{p}(N) - \tilde{p}(N-1) \\ -\tilde{p}(N) \\ \tilde{G}(0)' \cdot \tilde{p}(1) \\ \vdots \\ \tilde{G}(k)' \cdot \tilde{p}(k+1) \\ \vdots \\ \tilde{G}(N-1)' \cdot \tilde{p}(N) \end{pmatrix} = \tilde{a} \quad (38)$$

Now consider the system

$$\begin{aligned} B \cdot \delta \tilde{y} &\leq 0 \\ \tilde{a}' \cdot \delta \tilde{y} &> 0 \end{aligned} \quad (39)$$

Since there is no $\delta \tilde{y} \in T$ solution of (39), the system

$$B' \cdot \tilde{w} = \tilde{a} \quad \tilde{w} \geq 0 \quad (40)$$

has a solution $\tilde{w} = (w^1(1)', \dots, w^1(N)', w^2(0)', \dots, w^2(N-1)')' \in E^{(1+m)N}$

$$B' \tilde{w} = \begin{pmatrix} H^{1'}(x^0(1)) \cdot w^1(1) \\ \vdots \\ H^{1'}(x^0(k)) \cdot w^1(k) \\ \vdots \\ H^{1'}(x^0(N)) \cdot w^1(N) \\ \vdots \\ H^{2'}(u^0(0)) \cdot w^2(0) \\ \vdots \\ H^{2'}(u^0(N-1)) \cdot w^2(N-1) \end{pmatrix} = \tilde{a} \quad (41)$$

In equation (41) $H^1(x^0(k))$ is the $l \times (n+1)$ Jacobian matrix of $h^1(x)$ $h^1_x(x^0(k))$ evaluated at $x^0(k)$ and $H^2(u^0(k))$ is the $m \times r$ Jacobian matrix $h^2_u(u^0(k))$ of $h^2(u)$ evaluated at $u^0(k)$.

From equations (38) and (41) it follows that at a relative minimum y^0 of (19) there exist vectors $\tilde{p}^0(k) \in E^{n+1}$, $w^1 \in E^1$ and $w^2 \in E^m$ satisfying the equations

$$\tilde{p}^0(k+1) - \tilde{p}^0(k) = \tilde{F}'(k) \cdot \tilde{p}^0(k+1) + H^{1'}(x^0(k)) \cdot w^1(k) \quad k=1, \dots, N-1 \quad (42)$$

$$\tilde{p}^0(N) = -H^{1'}(x^0(N)) \cdot w^1(N) \quad (43)$$

and

$$\tilde{G}'(k) \cdot \tilde{p}^0(k+1) = H^{2'}(u^0(k)) \cdot w^2(k) \quad k=0, \dots, N-1 \quad (44)$$

Furthermore the equations (4), (5) and (40) represent the complementarity relations (Kuhn-Tucker-relations) for an optimal solution

$$h^{1'}(x^0(k)) \cdot w^1(k) = 0 \quad k=1, \dots, N \quad (45)$$

and

$$h^{2'}(u^0(k)) \cdot w^2(k) = 0 \quad k=0, \dots, N-1 \quad (46)$$

Since $\frac{\partial f_i}{\partial x} = 0$, the component $p_0(k)$ of $p(k)$ is constant and because of (32) there is

$$p_0^*(N) \leq 0 \quad (47)$$

If $f(x, u)$ is linear and $f_0(x, u)$ is convex then $\phi(y)$ is convex on Y and M is a convex set. A stationary point of a convex function is a global minimum, thus the conditions obtained are necessary and sufficient for a global minimum.

If there are no constraints for the state, equation (42) becomes

$$\tilde{p}(k+1) - \tilde{p}(k) = -\tilde{F}'(k) \cdot \tilde{p}(k+1) \quad k=1, \dots, N-1 \quad (48)$$

This is the adjoint equation used for the derivation of the discrete maximum principle [4].

5. A Discrete Maximum Principle

The results obtained in the previous section represent a maximum principle for discrete systems as it was first formulated by Rozonoer [8] for constraints in the control only. From equations (33) and (38) it follows

$$\tilde{a}(k) = \begin{pmatrix} (\tilde{F}'(k) + I) \cdot \tilde{p}(k+1) - \tilde{p}(k) \\ \tilde{G}'(k) \cdot \tilde{p}(k+1) \end{pmatrix} \quad (49)$$

and for the inner product $\tilde{a}(k) \delta \tilde{y}(k)$

$$\tilde{p}'(k+1)(\tilde{F}(k) + I) \delta \tilde{x}(k) - \tilde{p}'(k) \delta \tilde{x}(k) + \tilde{p}'(k+1) \cdot \tilde{G}(k) \delta u(k) \leq 0 \quad (50)$$

From (50) it is obvious that the Hamiltonian

$$\begin{aligned} H(\tilde{p}(k+1), x(k), u(k)) &= \\ &= \tilde{p}(k+1)' \cdot (\tilde{f}(x(k), u(k)) - \tilde{x}(k)) - \tilde{p}(k)' \cdot \tilde{x}(k) \end{aligned} \quad (51)$$

has a stationary point or is a maximum at $x^0(k)$ and $u^0(k)$. In addition there is $p_0^0(N) \leq 0$ and the transversality condition (43).

6. Summary

This paper shows how a theorem for disjoint convex sets and a theorem for systems of linear equations and inequations can be used to obtain necessary optimality conditions for discrete time problems. The optimal control problem is formulated as a problem of mathematical programming, the techniques of which may be used for the finding of optimal solutions. The conditions are not necessarily minimum conditions. There are however many systems which have only one local minimum and for these systems the results derived in this paper are useful. In the linear case and for a convex objective function the conditions are necessary and sufficient for a global minimum. Furthermore it is demonstrated that the conditions obtained represent a discrete maximum principle for systems with constraints in the control and the state.

References

- 1 Butkovskij, A.G.: The Necessary and Sufficient Conditions for Optimality of Discrete Control Systems. *Avtomatika i Telemekhanika* 24 (1963), S. 1056-1064.
- 2 Rosen, J.B.: Iterative Solution of Nonlinear Optimal Control Problems. *SIAM Journal on Control* 4 (1966), S.233-244.
- 3 Pearson, J.B. and Shridar, R.: A Discrete Maximum Principle via Convex Programming. *IFAC London 1966*, Paper 18F.
- 4 Jordan, B.W. and Polak, E.: Theory of a Class of Discrete Optimal Control Systems. *Journal of Electronics and Control* 17 (1964), S.697-713.
- 5 Halkin, H.: A Maximum Principle of the Pontryagin Type for Systems Described by Nonlinear Difference Equations. *SIAM Journal on Control* 4 (1966), S.90-111.
- 6 Collatz, L. und Wetterling, W.: *Optimierungsaufgaben*. Springer-Verlag Berlin, Heidelberg, New-York 1966.
- 7 Farkas, J.: Theorie der einfachen Ungleichungen. *Journal f. reine u. angewandte Mathematik* 124 (1902), S.1-27.
- 8 Rozonoer, L.I.: The Maximum Principle of L.S. Pontryagin in Optimal System Theory. *Automation and Remote Control* 20 (1959), S.1517-1532.

CHAIN MODELS

AS INERTIALESS OPTIMAL CONTROLLERS OF MULTI-DIMENSIONAL PROCESSES

Konstanty J. KURMAN /Poland/^x

1. **I n t r o d u c t i o n .** The paper presents a method of optimization of new perspectives - chain models method^{3,4,5}. As long as the present methods are used the computational effort for optimization grows with the square of the dimensionality of the problem at least⁵ using the described method .. it grows linearly.

The basic problem in an efficient computing of optimal control is solving differential equations with given initial and final conditions⁵. In the described method this difficulty disappears since the dynamic problem are reduced to the static one: real time is not represented by the time, but by a certain space coordinate of the model. Theoretically the model is inertialess, the solving time equals zero.

Practically - the model realizes Arrow-Hurwicz gradient method². The computation of gradient presents no difficulties, due the application of the modified Dennis analogy¹ to the determination of the model's structure; the model receives therefore self-optimizational features - which result from the modified Maxwell's principle of minimum power consumption¹. The introduced modifications cause that the complexity of the model is proportional to dimensionality of the problem.

The chain model used as an optimal controller, which in principle is inertialess, is the practical medium of solving the problem of the optimal system synthesis with

^x/ Department of Automatic and Remote Control, Technical University of Warsaw, Warsaw, Nowowiejska 15/19.

various forms of the final conditions. The important theoretical result is the disappearing of the "dimensionality barrier" and the constatation of no need for decomposition methods for those problems, which can be solved effectively by the chain modelling.

2. The problem of the static optimization. Let us consider the static optimization problem: determine a n_0 -dimensional vector of the decision variables $\bar{y} = [y_1, y_2, \dots, y_{n_0}]$ which locally minimizes the given cost-function $J(\bar{z}, \bar{y})$ where the vector $\bar{z} = [z_1, z_2, \dots, z_l]$ is given, under the additional conditions /which will be called equality constraints/:

$$\bar{\varphi}(\bar{z}, \bar{y}) = [\varphi_1, \varphi_2, \dots, \varphi_{m_0}] = \bar{0}, \quad /1/$$

where $m_0 < n_0$.

Either we assume that there exists an unique local minimum, or we must agree, that just one optional of existing local minimums will be determined.

The above formulation of the problem is sufficiently general from the point of view of practical applications, under condition that all inequality constraints such as

$\varphi_k(\bar{z}, \bar{y}) \leq 0$ were replaced by "half-square" cost components:

$$J_k = \begin{cases} = c_k \varphi_k^2 & \text{for } \varphi_k \geq 0 \\ = 0 & \text{for } \varphi_k < 0 \end{cases}$$

where c_k a sufficiently large positive number.

These components are considered as "penalty functions" for the violation of constraints.

Let us introduce the m_0 -dimensional vector of Lagrangian coefficients $\bar{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_{m_0}]$ and the Lagrangian function

$$J^*(\bar{z}, \bar{y}, \bar{\lambda}) = J(\bar{z}, \bar{y}) + \sum_{k=1}^{m_0} \lambda_k \varphi_k(\bar{z}, \bar{y}). \quad /2/$$

The optimization problem is under certain conditions equivalent to the problem of determination of the $n_0 + m_0$ - dimensional vector of the new independent variables $\bar{v} = [v_1, v_2, \dots, v_{n_0+m_0}] = [\bar{y}, \bar{\lambda}]$, for which the function $J^*(\bar{z}, \bar{v})$ reaches a stationary value.

So if we call the relation

$$\frac{\partial J^*}{\partial v_k} = 0 \quad /3/$$

the relation of optimality of J^* with respect to v_k , then solving the optimization problem is reduced to solving the system equations consisting relations of optimality /3/, for $k = 1, \dots, n_0 + m_0$.

3. Energetic model of Lagrangian function for the linear problem
Let us focus for a while on the linear optimization problem, for which the equations of constraints /1/ are linear, so that they have the form:

$$\psi_k(\bar{w}) = \sum_{i=1}^{l+n_0} \alpha_i^k w_i = 0; \quad k = 1, \dots, m_0 \quad /4/$$

where $\bar{w} = [w_1, w_2, \dots, w_{l+n_0}] = [\bar{z}, \bar{y}]$ while the cost function is a quadratic form.

$$J(\bar{w}) = J_0(\bar{w}) = \sum_{i=1}^{l+n_0} \sum_{j=1}^{l+n_0} \beta_{ij} w_i w_j \quad /5/$$

with $\beta_{ij} \beta_{ji} = 0$ for $j \neq i$.

Above implies the form of Lagrangian function:

$$J^*(\bar{u}) = \sum_{i=1}^{l+n_0+m_0} \sum_{j=1}^{l+n_0+m_0} \delta_{ij} u_i u_j, \quad /6/$$

where $\delta_{ij} \delta_{ji} = 0$ for $j \neq i$,

$$\bar{u} = [u_1, u_2, \dots, u_{l+n_0+m_0}] = [\bar{z}, \bar{y}, \bar{\lambda}].$$

The form of the expression /6/ implies the linearity of equations /3/.

Let us form an electrical network modeling the Langrangian function /6/ in such a way that the given variable u_i represents the node potential in the model while the given component $\mathcal{F}_{ij} u_i u_j$ (for $\mathcal{F}_{ij} \neq 0$) of the function /6/ is represented by an Π -element, which consists of positive resistances /"+R"-elements/ and negative resistances /"-R"-elements/, connecting the node " u_i " and the node " u_j ". The Π -element is shown on Fig.1 where the resistors are represented by their conductances. A "-R"-element is an active element /a source of energy/ which differs from element "+R" by the direction of the current, with identical voltage on the terminals.

The power consumption of the Π -element from Fig.1 /this value may be negative, too/ is expressed by

$$P_{ij} = \mathcal{F}_{ij}(u_i)^2 - \mathcal{F}_{ij}(u_i - u_j)^2 + \mathcal{F}_{ij}(u_j)^2 = 2\mathcal{F}_{ij} u_i u_j$$

/the power consumption is equal to the doubled value of the Lagrangian function's component, which is modelled by the Π -element/. By introduction vector $\bar{u}$ to the model in the form of potentials of the corresponding nodes we receive the complete power consumption of network /from the $\bar{u}$ sources/ equal to the doubled value of the function $J^*(\bar{u})$, on the basis of the principle of conservation of energy.

4. Self - optimizational property of energetic model. The current i_{ij} flowing out of the node " u_i " in the direction of the node " u_j " through connecting them Π -element /Fig.1/ can be expressed as

$$i_{ij} = \mathcal{F}_{ij} u_j = \frac{\partial}{\partial u_i} \mathcal{F}_{ij} u_i u_j. \quad /7/$$

The current i_{i0} flowing out of the node " u_i " in the ground direction through Π -element linking the node " u_i " with itself /representing the component $\mathcal{F}_{ii}(u_i)^2$ of the J^* function/ can be expressed as

$$i_{i0} = 2\mathcal{F}_{ii} u_i = \frac{\partial}{\partial u_i} \mathcal{F}_{ii}(u_i)^2. \quad /8/$$

If the external voltage was not applied to the node " u_i " / u_i is a free quantity/, then according to /7/, /8/ and /6/ the sum of currents flowing out of the node " u_i " has the form:

$$\sum_k i_{ik} = \frac{\partial J^*}{\partial u_i}. \quad /9/$$

According to the 1st Kirchhoff's Law the node potential u_i assumes the value, for which $\sum_k i_{ik} = 0$, thus according to /9/ it assumes the optimal value, for which the relation /3/ i.e. relation of optimality J^* with respect to u_i is fulfilled. This fact we can call a "principle of a stationary value of power consumption": the distribution of the node potentials of the " $\pm R$ " - network represents a stationary value of the power consumption of the whole network from external sources /compare Maxwell's principle of minimum power which is valid for "+R" - networks only¹/.

Getting back to the original setting of the optimization problem we find out, that the energetic model, formed by us, has the property of self-optimization: the vector of decision variables $\bar{y}$ assumes a value, which minimizes the cost J under given conditions /i.e. constraints /1/ and values $\bar{z}$, forced by external sources/. The power consumption P of the model from the $\bar{z}$ -sources represents at the same time the doubled value of the conditional minimum cost J , since the fulfilment of relation of optimality J^* with respect to the vector $\bar{\lambda}$ implies $\bar{\psi} = \bar{0}$, so according to /2/ we get

$$J = J^* = \frac{1}{2} P. \quad /10/$$

5. I n e q u a l i t y c o n s t r a i n t s a n d o t h e r n o n l i n e a r i t i e s o f t h e p r o b l e m . Let us assume, that an additional component $J_k(w_i)$ of "half-squared" type occurs in the cost function, for e.g.

$$J = J_0 + J_k,$$

$$\text{where } J_k = \begin{cases} = g_k (w_i - z_k)^2 & \text{for } w_i \geq z_k, \\ = 0 & \text{for } w_i < z_k. \end{cases} \quad /11/$$

Let w_i be one of the decision variables.

The component J_k can be represented by a diode-element /Fig.2/ in the energetic model. This element is put in between node " w_i " and the voltage source z_k . Because an ideal diode does not consume a power the power consumption of the diode-element may be expressed as $P_k = 2J_k$ /power dissipated on the resistor of $2g_k$ conductance/.

Additional component of the current flowing out of node " w_i " can be expressed as

$$i_k = \begin{cases} = 2g_k (w_i - z_k) & \text{for } w_i \geq z_k \\ = 0 & \text{for } w_i < z_k \end{cases} = \frac{\partial J_k}{\partial w_i}$$

so as that component of the relation of optimality J^* with respect to w_i which corresponds to occurring the J_k component in the cost function. Thus by introducing the "half-squared" cost in the form of diode-elements to the problem /and the model/ does not violate the equality /9/, so it does not violate the self-optimizational properties of the model and the principle of the stationary value of power consumption - which is not true in the general nonlinear case¹.

If the variable w_i in /11/ is not one of decision variable, but their linear combination, we should introduce it to the class of decision variables, simultaneously completing the constraints /1/ with the appropriate equation of the following type

$$\psi_{m_0+1} = w_i - \sum_{k=1}^{n_0} \delta_k y_k = 0$$

and completing the $\bar{\lambda}$ vector with the λ_{m_0+1} component.

The diode-elements make it possible to model a large class of cost functions, under the condition of their proper approximation /piecewise parabolic/.

By choosing a sufficiently large value g_k in /11/ we can take into consideration "a penalty for violation of constraints". Notice, that penalty disappears from the solution when $g_k \rightarrow \infty$ /resistor disappears from the diode-element/ which corresponds exactly to introduction of inequality constraints¹.

There is also a possibility of introducing piecewise linear constraints for e.g. by "withdrawing" not acting constraints, which can be achieved by means of suppressing the corresponding coordinated of $\bar{\lambda}$ vector to the zero value /connecting the corresponding nodes to the ground/.

6. Symmetrical models. Complexity.

Let us form the energetic model a little differently.

Let u_i variable be represented in the model by a pair of nodes with $+u_i$ and $-u_i$ potentials. Let $\gamma_{ij}u_iu_j$ component of J^* function be represented by pair of Π -elements from Fig.3a, if $\gamma_{ij} < 0$, and by the pair of Π -elements from Fig.3b, if $\gamma_{ij} > 0$. Power consumption of the Π -elements equals in both cases $P_{ij} = 4\gamma_{ij}u_iu_j$.

Besides this change of the coefficient of proportionality between the power consumption and the costs - the basic properties of the model /self-optimization/ remain unchanged.

In the symmetrical model "-R" elements occur only in connections between the node and the ground or, after simple transformation, in branches: node "+u"/node "-u"; ordinary branches u_i/u_j contain only "+R" elements. It allows to ground all amplifiers used for realization of "-R" elements. Furthermore, the complexity r of the symmetrical model, measured with the number of "-R" elements, undergoes the limits:

$$m_0 \leq r \leq n_0 + m_0,$$

which comes from the fact that "-R" elements occur always at " λ "-nodes, and they can also occur at "y"-nodes. Then the complexity grows only linearly with the dimensionality of the problem $n_0 + m_0$.

7. The mechanism of the self-optimization of the model.
Let us assume, that

- 1/ The cost-function $J(\bar{w})$ is strictly concave.
- 2/ The only dynamic parasitic parameters are "node" capacities C_v /between node "v" and the ground/.
 $C_y > 0$ for "y"-node, $C_\lambda < 0$ for " λ "-node.

Assumption 2 means /with certain definite idealisation/ that there are applied "-R" elements of open-circuit-stable type in " λ "-nodes, and the elements of short-circuit-stable type in those "y"-nodes with which "-R" elements occur.

According to /9/ and assumption 2 equations of the model dynamics assume form

$$\frac{\partial J^*}{\partial v_i} + C_{v_i} \frac{dv_i}{dt} = 0 \quad ; \quad i = 1, \dots, n_0 + m_0 \quad /12/$$

which according to /10/ we can interpret as two "mechanisms":

- minimization of Lagrangian costs /or power/ with respect to decision variables:

$$\frac{dy_i}{dt} = -\frac{1}{C_{y_i}} \frac{\partial J^*}{\partial y_i} = -\frac{1}{2C_{y_i}} \frac{\partial P}{\partial y_i} \quad i = 1, \dots, n_0$$

- maximization of Lagrangian costs /or power/ with respect to Lagrangian coefficients

$$\frac{d\lambda_k}{dt} = \frac{1}{|C_{\lambda_k}|} \frac{\partial J^*}{\partial \lambda_k} = \frac{1}{2|C_{\lambda_k}|} \frac{\partial P}{\partial \lambda_k} \quad k = 1, \dots, m_0$$

As we see it is a realisation of Arrow-Hurwicz gradient method². Under assumption 1 and under certain qualifications reffering to form of constraints the algorithm converges², which proves the stability of model.

There exists a possibility of forming, locally stable model also with non-concave function $J(\bar{w})$, if there only exists the locally unique solution of the problem. For this purpose we must choose sufficiently large numbers h_k and complete the cost function with the component:

$$\sum_{k=1}^{m_0} h_k \psi_k^2. \quad /13/$$

It is easy to prove that the expression /13/ does not influence the solution, but makes the cost function locally concave: coming out of the point of equilibrium in arbitrary direction θ without violation of constraints we have $\frac{\partial^2 J}{\partial \theta^2} > 0$ according to the assumption of existing of solution, for any other direction - there exists such k , that $\frac{\partial \psi_k}{\partial \theta} \neq 0$ and choosing sufficiently large h_k guarantees, according to expression /13/ that also in that direction $\frac{\partial^2 J}{\partial \theta^2} > 0$.

The general characteristics of disappearing of the transition process result from the equations /12/:

$$\frac{dJ^*}{dt} = \sum_i \frac{\partial J^*}{\partial v_i} \frac{dv_i}{dt} = - \sum_i \frac{1}{C_{v_i}} \left(\frac{\partial J^*}{\partial v_i} \right)^2.$$

Settling time /lower than one second/ we can theoretically shorten arbitrary diminishing, C_v . That time depends besides on "steepness" of the searched minimum and on the structure of the optimization problem, represented by the model structure /wave processes/ - that is why the time depends at most linearly on the dimensionality of the problem /of the number $n_0 + m_0$ /.

8. Energetic modelling and decomposition. When we solve the problem of optimization by the energetic modelling method we do not have to use the decomposition of a large systems. It results immediately from the fact that the growing of the problem dimensionality $n_0 + m_0$ implies at most

the linear growth of the complexity of the model and the growth of solving time where that time is negligibly short.

9. The problem of dynamic optimization and energetic chain model.

If on any accounts /here: if we take under consideration the possibilities of the technics of computing applied/ we look for the optimal control in the class of piece constant functions /constant in a given stage of process/, we come to a discrete version of the problem of dynamic optimization, that is to the problem of conditional minimization of the multivariable function /of the process cost/:

$$J = \sum_{i=1}^N Q_i(\bar{x}_i, \bar{u}_i, \bar{z}_i),$$

with respect to the control variable vector $\bar{u}_i$ and the state variable vector $\bar{x}_i$, subjected to dynamic constraints.

$$\psi_i^p = x_{i+1}^p - \varphi_i^p(\bar{x}_i, \bar{u}_i, \bar{z}_i) = 0 \quad \begin{cases} p=1, \dots, n \\ i=1, \dots, N \end{cases}$$

and eventually to additional equational constraints, with:

$\bar{x}_i = [x_i^1 \ x_i^2 \ \dots \ x_i^n]$ — object state variable vector in the beginning of i-stage,

$\bar{u}_i = [u_i^1 \ u_i^2 \ \dots \ u_i^m]$ — control variable vector in i-stage

$\bar{z}_i$ — constant value vector of i-stage /initial and final conditions, determined signals etc./.

Q_i — costs of i-stage,

N — number of stages.

Creating the vector of decision variables from the free vectors of states and control $\bar{y} = [\bar{x}_2 \bar{x}_3 \dots \bar{u}_1 \bar{u}_2 \dots \bar{u}_N]$, the vector of equational constraints function $\bar{\psi} = [\bar{\psi}_1 \bar{\psi}_2 \dots \bar{\psi}_N]$, and the vector of constant values $\bar{z} = [\bar{z}_1 \bar{z}_2 \dots \bar{z}_N]$, we reduce the problem of dynamic optimization to the problem of static optimization — section 2. Then we can also form

the energetic model, which we will call the chain-model with regard to its specific structure⁴ /the individual chain links - successively joined multipoles - represent the successive stages of the process/. All the results of the previous sections are still true with the remark, that the chain structure makes the settling time practically independent on the object dimensionality n , the complexity of the model remains proportional to n .

Fig.4 represents the example of the symmetrical chain model of the 3-stage optimal process, for the following problem:

$$\dot{x}_i = x_{i+1} - ax_i - bu_i = 0 \quad ; \quad i=1,2,3$$

$$J = \sum_{i=1}^3 [c(x_i)^2 + d x_i u_i + e(u_i)^2]$$

$$x_i \geq x_0 \quad ; \quad i=2,3$$

$$x_0, x_1, x_4 \text{ - are given}$$

a, b, c, d, e - positive constants.

In the problem with the free final state x_4 , we force the final conditions by grounding of nodes $\pm \lambda_3$ in Fig.4 /the constraint $\psi_3 = 0$ disappears then λ_3 in Lagrangian function $J^* = J + \sum_i \lambda_i \psi_i$ disappears/.

10. Inertialess optimal controller.

Fig.5 illustrates the principle of the application of the chain model as an inertialess optimal controller. When the successive stage of optimal process passed we shorten the chain of one link. In the problem with free time of the lasting of the process we establish the number of links N on the basis of the criterion of minimization of power delivered from external sources. Giving of final conditions does not imply any difficulties either with definite x_{N+1}^p /in the form of adequate voltage/, or with free x_{N+1}^p /through the grounding λ_N^p /, or with arbitrary linear final constraints /equational and inequational/. Local

constraints of control variables and state variables, global constraints, dead times in the object, influence of known external signals etc. imply no difficulties.

11. C o n c l u s i o n s . Mainly as a result of revision of "the principle of time modelling by time" chain modelling of optimal processes brings out new possibilities and new conceptions in the field of optimal control.

First of all there appears a possibility of forming of inertialess optimal controller with the complexity proportional to the dimensionality of controlled object for sufficiently large class of problems from the point of view of applications. The controller is distinguished by its simple mode of operation, sequential structure convenient to technology and with sufficient accuracy as the experiments proved.

In theoretical aspect chain models introduce easily imaginable static, physical models of the problem of dynamic optimization, instead of geometrical models in multidimensional space. It turns out, besides, that the "barrier of dimensionality" and the necessity of applying decomposition are not organically connected with the problem of optimization itself, but only with the used so far methods of solving the problem.

The author is also grateful to prof.dr W.Findeisen, prof.dr A.Gosiewski, prof.dr A.Wierzbicki, dr J.Puławczewski and dr B.Piwko for their favourable and valuable remarks.

B i b l i o g r a p h y

1. J.B.Dennis: Mathematical Programming and Electrical Networks. MIT, 1959.
2. G.Hadley: Nonlinear and dynamic programming. Addison-Wesley Publishing Comp.1964.
3. K.Kurman: Koncepcja łańcuchowych modeli procesów optymalnych. Archiwum Autom. i Telemekh.XI, zesz.3, 1966.
4. K.Kurman: Zasady modelowania łańcuchowego procesów optymalnych, in the complete work: Prace IV Krajowej Konferencji Automatyki, vol.1, pp.241-252.AGH, 1967.
5. A.Wierzbiński: Zasada maksimum a synteza-regulatorów optymalnych. Archiwum Autom. i Telemekh. XIII, zesz.3. 1968.

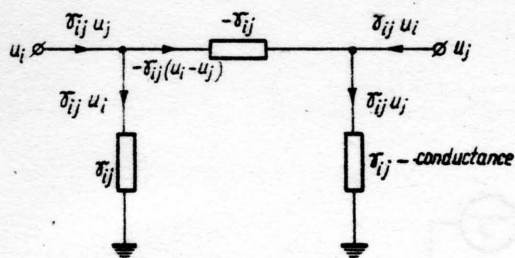


Fig. 1

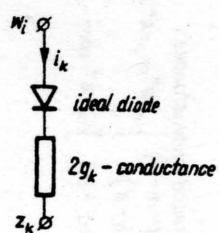


Fig. 2

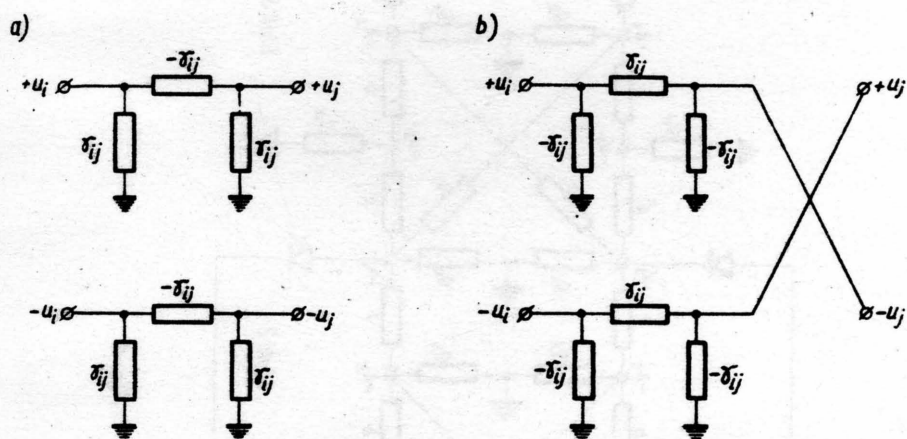


Fig. 3

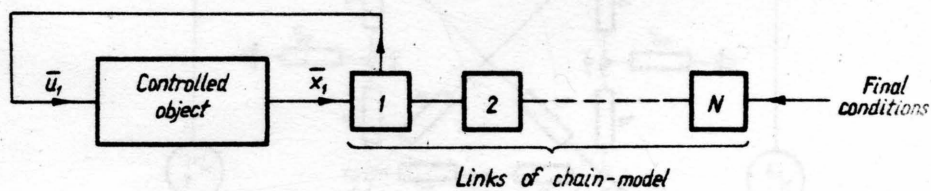


Fig. 5

Junction / ground conductances :

$$g_x = 2c - (1+a+b) \quad \left. \begin{array}{l} +R'' \text{ or } -R'' \text{ short-stable} \\ g_u = 2e - (b+d) \end{array} \right\}$$

$$g_\lambda = -(1+a+b) \quad \left. \begin{array}{l} -R'' \text{ open-stable} \end{array} \right\}$$

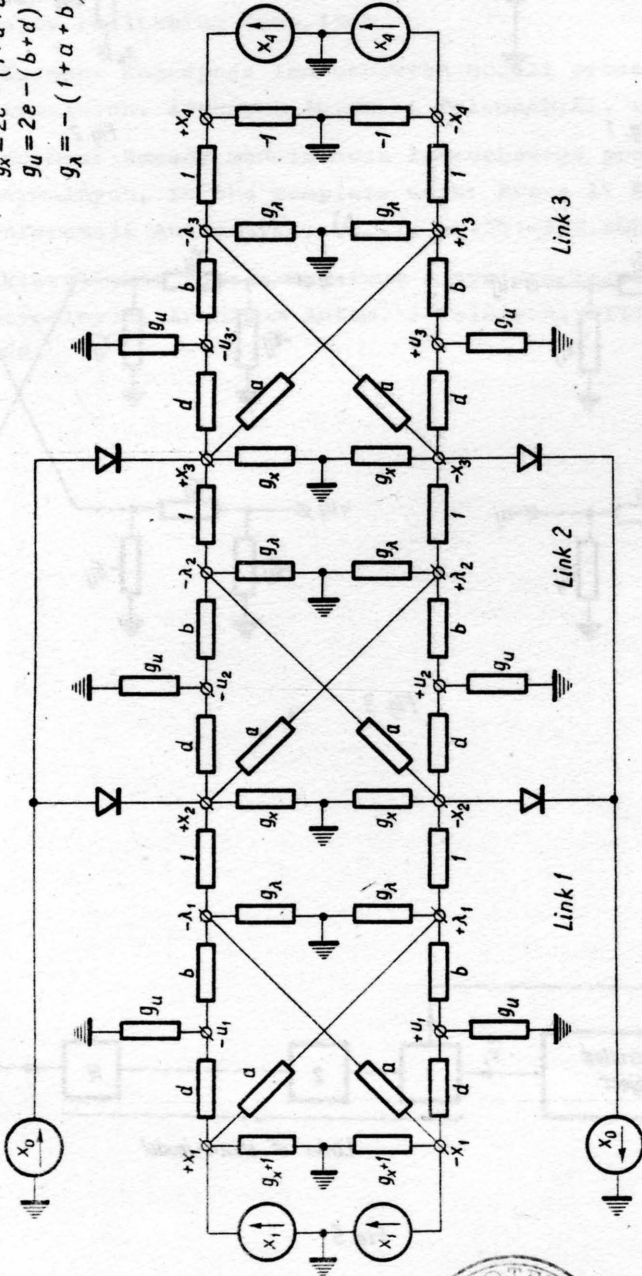


Fig. 4

