В Ы Ч И С Л И Т Е Л Ь Н Ы Е   С И С Т Е М Ы
Сборник трудов
Института математики СО АН СССР

1967 г. Выпуск 27

О РЕШЕНИИ ЗАДАЧ ОПТИМИЗАЦИИ ДИНАМИЧЕСКИХ СИСТЕМ МЕТОДАМИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Ю.С. Завьялов

I. Многие физические процессы и технические устройства, представляющие интерес при разработке вычислительных средств, являются д и н а м и ч е с к и м и   с и с т е м а м и. К ним относятся, в частности, электрические цепи, ионные и электронные пучки при отсутствии взаимодействия частиц и т.п. Движение динамических систем (процесс , работа устройства) во времени зависит от параметров системы, например, в случае электрических цепей - от величины сопротивлений, емкостей, индуктивностей, уровня входного сигнала; в случае ионных потоков - от массы ионов, напряженностей электрического и магнитного полей и т.п.

Возникает задача: варьированием параметров добиться оптимального в каком-либо смысле движения системы. Инженерные решения подобных задач достаточно известны. Обычно движение системы моделируется на аналоговой вычислительной машине (АВМ), с которой соединен оптимизатор - логическое устройство автоматического поиска наилучшего решения [I]. Применение такой методики ограничено возможностями АВМ, а именно невысокой точностью решения и сравнительно небольшими допустимыми размерами задачи (числом степеней свободы системы).

В работе рассматривается точная математическая постановка задачи оптимизации динамической системы. Обсуждаются способы её решения методами нелинейного программирования на цифровых вычислительных машинах (ЦВМ). При этом как промежуточный результат решается задача (важная сама по себе) о движении динамической системы в зависимости от значений параметров.

В последующих статьях данного сборника рассмотрены различные частные задачи оптимизации такого рода и разработаны приемы их решения.

2. Движение динамической системы описывается системой обыкновенных дифференциальных уравнений, взятой в нормальной форме (см., например, [2] ,стр.170):

$$\frac{dy_i}{dt} = f_i(t, y, x), \quad i = 1, 2, \ldots, n, \qquad \text{(I)}$$

$$y = \{ y_1, y_2, \ldots, y_n \},$$

с начальными условиями:

$$t = t_0, \quad y_i = y_i^0, \quad i = 1, 2, \ldots, n. \qquad (2)$$

В случае электрических цепей время t входит в правые части уравнений (I) через функцию входного сигнала $u(t)$,т.е. $f_i = f_i\,[u(t), y, x]$.

Величина $x = \{x_1, x_2, \ldots, x_m\}$ означает точку в пространстве параметров.При фиксированных значениях параметров движение динамической системы представляется решением задачи Коши (I)-(2). Если же из множества таких систем отыскивается в каком-то смысле наилучшая, то параметры рассматриваются как переменные.

Предположим, что Δ - некоторая замкнутая область переменных t, y:

$$t - t_0 \le a, \quad |y_i - y_i^0| \le b, \quad i = 1, 2, \ldots, n, \qquad (3)$$

где a и b - положительные постоянные;
X - тоже замкнутая (в общем случае это не обязательно) область параметров x_k :

$$\alpha_k \le x_k \le \beta_k, \qquad k = 1, 2, \ldots, m, \qquad (4)$$

где α_k, β_k - постоянные;
D - область (3)-(4) в пространстве (t, y, x).
Тогда известны следующие свойства решения задачи Коши (I)-(2) (см.,например, [2], гл. I; [3] , гл. IУ):

а) Пусть функции $f_i(t, y, x)$ в D непрерывны по всем своим аргументам и, следовательно, ограниченны

$$|f_i(t, y, x)| \le M, \quad M > 0 - const, \qquad (5)$$

и имеют непрерывные и ограниченные производные $\partial f_i / \partial y_j$ $(j = 1, 2, \ldots, n)$. Тогда можно указать такой промежуток

$$t - t_0 \le h, \quad h = \min\left\{a, \frac{b}{M}\right\}, \qquad (6)$$

что на нем существует, и при том единственное,решение задачи Коши

$$y_i = \varphi_i(t, x), \quad i = 1, 2, \ldots, n, \qquad (7)$$

непрерывное по совокупности переменных $t, x_1, x_2, \ldots, x_m$ и непрерывно дифференцируемое по t .

б) Пусть, кроме того, функции $f_i(t, y, x)$ имеют в D непрерывные и ограниченные частные производные $\partial f_i / \partial x_k$ $(k = 1, 2, \ldots, m)$. Тогда решение (7) будет иметь частные производные $\partial \varphi_i / \partial x_k$, непрерывные по совокупности переменных $t, x_1, x_2, \ldots, x_m$ на интервале (6).

ЗАМЕЧАНИЕ. Часто в задачах область Δ оказывается неограниченной по y_j . Но,если непрерывны и равномерно ограничены $\partial f_i / \partial y_j$, указанные свойства решения (7) имеют место на всем интервале $t - t_0 \le a$. Ограниченность функций $f_i(t, y, x)$ при этом не существенна.

Частные производные от решения по параметрам

$$v_{ik} = \frac{\partial \varphi_i(t, x)}{\partial x_k} = \psi_{ik}(t, x), \qquad (8)$$

$$i = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, m,$$

определяются линейной системой уравнений в вариациях ([3] , гл. IУ):

$$\frac{dv_{ik}}{dt} = \sum_{j=1}^{n} f_{ij}(t, x) v_{ik} + g_{ik}(t, x), \qquad (9)$$

где

$$f_{ij}(t,x)=\frac{\partial f_i(t,y,x)}{\partial y_j}\Big|_{y=\varphi(t,x)},$$

$$g_{ik}(t,x)=\frac{\partial f_i(t,y,x)}{\partial x_k}\Big|_{y=\varphi(t,x)},$$

при начальных условиях:

$$t=t_o, \qquad v_{ik}=0. \tag{I0}$$

Система(9) и условия (I0) получаются из (I) и (2) дифференцированием по параметру x_k.

Правые части уравнений (9) удовлетворяют условиям замечания. Поэтому задача Коши (9)-(I0) имеет единственное решение на том же интервале, что и задача (I)-(2).

3. Задача оптимизации динамической системы ставится следующим образом:

Предположим, что функции $f_i(t,y,x)$ удовлетворяют условиям а) и б), т.е. они непрерывны по всем своим аргументам и непрерывно дифференцируемы по y_i ($i=$ I,2,..., n) и x_k ($k=$ I,2, ..., m).

Пусть, далее, на множестве движений динамической системы в зависимости от параметров задан функционал

$$\mathcal{F}(x)=\int_{t_1}^{t_N} f_o\left[t,\varphi(t,x),x\right]dt \tag{II}$$

или функционал на конечном множестве последовательных значений времени

$$\mathcal{F}(x)=\sum_{\nu=1}^{N} c_\nu f_o\left[t_\nu,\varphi(t_\nu,x),x\right]. \tag{I2}$$

К форме (I2) приводится и функционал (II), если интеграл вычисляется с помощью интерполяционных квадратурных формул. В (II)-(I2) интервал $[t, t_N]$ должен целиком содержаться в интервале (6) $[t_o, t_o+h]$ или совпадать с ним.

Предположим для простоты, что область определения функции $f_o(t,y,x)$ содержит в себе D -область определения функций $f_i(t,y,x)$ в системе (I).Сама же функция $f_o(t,y,x)$ принадлежит к тому же классу, что и функции $f_i(t,y,x)$, т.е.она непрерывна по всем своим аргументам и непрерывно дифференцируема по y_i и x_k. Условия непрерывности обеспечивают существование функционалов (II) или (I2) на всем множестве движений системы.

Задача. Найти минимум (максимум) функции $\mathcal{F}(x)$ в области X (4), т.е. найти такое x^*, что

$$\mathcal{F}(x^*)=\min \mathcal{F}(x) \quad (\mathcal{F}(x^*)=\max \mathcal{F}(x)) \tag{I4}$$

при

$$\alpha_k \le x_k \le \beta_k, \quad k=1,2,\dots,m. \tag{I5}$$

Функция $\mathcal{F}(x)$ называется функцией цели, а область X - допустимой областью.

Рассмотрим ситуации, которые могут встретиться при решении этой задачи.

Частный случай. Пусть решение задачи Коши (I)-(2) найдено в конечном виде при произвольных значениях параметров x_k $(k=1,2,\dots,m)$. В этом случае мы имеем явную (или хотя бы параметрическую) зависимость функции $\mathcal{F}(x)$ от своих аргументов. Тогда задача оптимизации будет стандартной задачей нелинейного программирования (см., например, [4, 5]). К задачам такого типа относится пример, рассмотренный в статье [9].

Общий случай. Как правило, решение задачи Коши (I)-(2) может быть получено только численными, например, конечно-разностными методами, а следовательно, только при конкретных значениях параметров x_k $(k=1,2,\dots,m)$. В этом случае явного выражения функции $\mathcal{F}(x)$ нет: её значения можно вычислить только при фиксированных значениях параметров. Тем не менее задачу оптимизации мы по-прежнему будем рассматривать как задачу нелинейного программирования со специфическим заданием целевой функции.Она определяется формулой (II),(I2), на всевозможных (в зависимости от параметров) решениях задачи Коши (I)-(2). Примеры задачи на общий случай представлены в статьях[I0, II].

Функции $f_0(t,y,x)$ и $\varphi_i(t,x)$ непрерывны по всем своим аргументам в D, поэтому функция $\mathcal{F}(x)$ непрерывна в допустимой области X. Но тогда она ограниченна и достигает точной нижней (точной верхней) границы в этой области,т.е. решение задачи (I4)-(I5) существует.

Вопрос о единственности решения в данном случае не является столь важным, как, например, при решении задачи Коши. Нам достаточно найти любое из возможных решений, лишь бы оно давало экстремум функции $\mathcal{F}(x)$. Все же мы укажем здесь некоторые факты, относящиеся к этому вопросу [4,5].

Область переменных x называется выпуклой, если вместе с каждыми двумя точками x' и x'' ей принадлежит и весь отрезок, их соединяющий

$$x = \lambda x' + (1-\lambda)x'', \qquad 0 \le \lambda \le 1.$$

В нашей задаче область X (I4) является многомерным параллелепипедом и, следовательно, выпукла.

Функция $\mathcal{F}(x)$ называется выпуклой в некоторой области, если для любых x' и x'' из этой области

$$\mathcal{F}[\lambda x' + (1-\lambda)x''] \le \lambda \mathcal{F}(x') + (1-\lambda)\mathcal{F}(x''), \quad 0 \le \lambda \le 1.$$

Если имеет место строгое неравенство, то $\mathcal{F}(x)$ называется строго выпуклой. Если в приведенной формуле знак неравенства сменить на противоположный, то функция называется вогнутой (строго вогнутой).

Задачи минимизации выпуклой функции или максимизации вогнутой функции в выпуклой области получили название выпуклого программирования.

Решение этих задач основывается на следующих свойствах выпуклых функций:

а) для выпуклой функции локальный минимум совпадает с абсолютным в выпуклой области;

б) для строго выпуклой функции минимум достигается в единственной точке выпуклой области. Аналогично обстоит дело с максимумом вогнутой функции.

В практических задачах, особенно при сложном, как в нашем случае, задании функции $\mathcal{F}(x)$, часто бывает затруднительно определить характер функции. Она может быть выпуклой в одних и вогнутой в других частях допустимой области, и даже одновременно выпуклой и вогнутой (линейная функция). Тогда остается открытым вопрос о том, каково число локальных экстремумов соответствующего типа и будет ли среди них один, дающий наименьшее (наибольшее) значение функции цели в области, или таких экстремумов будет несколько.

4. Методы решения задачи оптимизации (I4)-(I5) суть различные методы нелинейного программирования.

В частном случае (п.3) их применение не имеет каких-либо особенностей по сравнению с обычными задачами нелинейного программирования. В общем же случае нужно учитывать специфику

задания целевой функции.

Рассмотрим кратко некоторые градиентные методы в применении к общему случаю. Для использования этих методов нужно, чтобы $\mathcal{F}(x)$ была непрерывно дифференцируемой функцией своих аргументов. Это условие у нас выполнено в силу непрерывной дифференцируемости функций $\varphi_i(t,x)$ по x_k и функции $f_0(t,y,x)$ по y_i и x_k.

Частные производные функции $\mathcal{F}(x)$ (II) или (I2) вычисляются по формулам:

$$\frac{\partial \mathcal{F}}{\partial x_k} = \int_{t_1}^{t_N} \left[\frac{\partial f_0(t,\varphi,x)}{\partial x_k} + \frac{\partial f_0(t,\varphi,x)}{\partial y_i} v_{ik} \right] dt \qquad (I6)$$

или

$$\frac{\partial \mathcal{F}}{\partial x_k} = \sum_{\nu=1}^{N} c_\nu \left[\frac{\partial f_0(t_\nu,\varphi,x)}{\partial x_k} + \frac{\partial f_0(t_\nu,\varphi,x)}{\partial y_i} v_{ik} \right], \qquad (I7)$$

$$k = 1,2,\dots,m.$$

Входящие сюда функции $\varphi_i(t,x)$ являются решением задачи Коши (I)-(2), а функции $v_{ik} = \psi_{ik}(t,x)$ – решением системы уравнений в вариациях (9) с начальными условиями (I0). В этом и состоит дополнительная трудность и главное отличие решения нашей задачи от аналогичных задач при явном задании функции $\mathcal{F}(x)$.

Рассмотрим принципиальные схемы методов решения задач минимизации, предположив сначала, что целевая функция $\mathcal{F}(x)$ выпуклая.

Градиентные методы с малым шагом ([6], [7]) часто применяются при инженерном решении задач оптимизации [I].

В каждой внутренней точке области X вектор $-grad\,\mathcal{F}(x)$ определяет направление наискорейшего спуска $\mathcal{F}(x)$. В граничных же точках у этого вектора нужно положить равными нулю те компоненты, которые выводят точку x из допустимой области. Тогда кратчайший путь из точки $x^{(0)}$ в точку x^* минимума функции $\mathcal{F}(x)$ определяется системой

$$\frac{dx_k}{d\tau} = -\delta_k(x)\frac{\partial \mathcal{F}(x)}{\partial x_k}, \quad k = 1,2,\dots,m, \qquad (I8)$$

при начальных условиях

$$x_k(0) = x_k^{(0)}, \qquad (19)$$

где

$$\delta_k(x) = \begin{cases} 0 & \text{при } x_k = \alpha_k, \ \frac{\partial \mathcal{F}}{\partial x_k} > 0 \ \text{ и } x_k = \beta_k \ \frac{\partial \mathcal{F}}{\partial x_k} < 0; \\ 1 & \text{- в остальных случаях.} \end{cases}$$

Минимум функции $\mathcal{F}(x)$ может быть найден как решение задачи Коши (18)-(19) при $\tau \to \infty$ с любым начальным $x^{(0)} \in X$. Эту задачу, как и аналогичные задачи (1)-(2) и (8)-(9), удобнее всего решать конечно-разностными методами. При этом считать $\delta_k(x) = 1$, но проверять на каждом ℓ -ом шаге выполнение условий

$$\alpha_k \leqslant x_k^{(\ell)} \leqslant \beta_k, \quad k = 1, 2, \ldots, m.$$

Если одно из них нарушено, то в качестве $x_k^{(\ell)}$ следует принять ближайшее граничное значение α_k или β_k.

Критерий оптимальности, т.е. условие получения точки x^* минимума функции $\mathcal{F}(x)$, может быть записан в виде:

$$\delta_k(x^*) \ \frac{\partial \mathcal{F}(x^*)}{\partial x_k} = 0, \quad k = 1, 2, \ldots, m. \qquad (20)$$

Однако достижение такой точки, иначе говоря, сходимость метода, не гарантируется. Более того, известны примеры, когда сходимость не имеет места.

Градиентным методом такого рода решена задача в работе [10].

М е т о д ы в о з м о ж н ы х н а п р а в л е н и й – это градиентные методы с большим шагом, в которых величина шага от $x^{(\ell)}$ до $x^{(\ell+1)}$ определяется не искомой точностью решения, как в предыдущем случае, а требованием быстрейшего выхода к экстремальной точке. Эти методы обобщены и развиты Г. Зойтендейком [5]. Из них мы укажем удобный для рассматриваемых задач метод симплексного исправления Франк и Вулфа (см.[5-7])

Вектор s называется возможным направлением в точке x, если при достаточно малом перемещении в этом направлении точка остается в допустимой области, т.е. $x + \lambda s \in X$, где $0 \leqslant \lambda \leqslant 1$. Возможное направление называется подходящим, если по нему в рассматриваемой точке функция $\mathcal{F}(x)$ убывает, т.е.

$$\left. \frac{\partial \mathcal{F}(x + \lambda s)}{\partial \lambda} \right|_{\lambda = 0} = s \cdot grad\, \mathcal{F}(x) < 0.$$

Для точки $x^{(\ell)}$ отыскивается подходящее возможное направление $s^{(\ell)}$, вдоль которого убывание $\mathcal{F}(x)$ наибольшее, т.е. ищется

$$\min_s \left[s^{(\ell)} \cdot grad\, \mathcal{F}(x^{(\ell)}) \right].$$

В нашем случае, когда допустимая область – параллелепипед, эта задача линейного программирования тривиальна. Мы находим здесь минимум каждого слагаемого скалярного произведения:

$$\begin{aligned} S_k^{(\ell)} &= \alpha_k - x_k^{(\ell)}, \quad \text{если } \frac{\partial \mathcal{F}}{\partial x_k} > 0; \\ S_k^{(\ell)} &= \beta_k - x_k^{\ell}, \quad \text{если } \frac{\partial \mathcal{F}}{\partial x_k} < 0. \end{aligned} \qquad (21)$$

Если же $\partial \mathcal{F} / \partial x_k = 0$, то $S_k^{(\ell)}$ можно взять любым

($\alpha_k - x_k^{(\ell)} \leqslant S_k^{(\ell)} \leqslant \beta_k - x_k^{(\ell)}$), в частности, равным нулю.

Далее, на отрезке $x^{(\ell)} + \lambda s^{(\ell)}$ $(0 \leqslant \lambda \leqslant 1)$ находится минимум функции $\mathcal{F}(x^{(\ell)} + \lambda s^{(\ell)})$ как функции одной переменной λ. По получении $\lambda = \lambda'$ берем точку $x^{(\ell+1)} = x^{(\ell)} + \lambda' s^{(\ell)}$ и повторяем

процедуру. При этом $\mathcal{F}(x^{(\ell+1)}) < \mathcal{F}(x^{(\ell)})$. Последовательность $x^{(0)}, x^{(1)}, \ldots, x^{(\ell)}, \ldots, x^*$ сходится.

Критерий оптимальности состоит в отсутствии для x^* подходящего допустимого направления, иначе

$$s \cdot grad\, \mathcal{F}(x^*) \geqslant 0$$

для всех s, не выводящих точку из допустимой области.

Этот метод использован для решения задачи оптимизации в статье [11].

При использовании рассмотренных методов мы должны учитывать, что в задаче (14)-(15) функция цели $\mathcal{F}(x)$ может и не быть выпуклой. Тогда на пути её решения методами выпуклого программирования встают две дополнительные трудности([6], п.7.7).

Первая заключается в том, что может оказаться много локальных минимумов, и в лучшем случае мы найдем один из них. В настоящее время общего метода преодоления этой трудности нет. В практических задачах выяснению вопроса может помочь использование большого числа начальных точек и сравнение получаемых при этом решений.

В других случаях используется дополнительная информация о задаче. На этом основано, например, применение метода оврагов

[8]для задач нелинейного программирования, в которых известно, что локальные минимумы в виде неглубоких оврагов расположены как бы на склонах глубокого оврага. Алгоритм строится так, чтобы, игнорируя локальные минимумы, выйти на дно главного оврага, где и отыскивается наименьшее значение функции цели.

Вторая трудность состоит в том, что точка x, удовлетворяющая критерию оптимальности, не обязательно будет локальным минимумом функции $\mathcal{F}(x)$. Такая точка, называемая стационарной, может быть также локальным максимумом или седлом. Максимум может встретиться только в начальной точке $x^{(0)}$ ($grad\,\mathcal{F}(x^{(0)})=0$), поэтому данный случай не имеет особого значения. Хотя вероятность попадания в седловую точку и мала, тем не менее её нельзя полностью исключать.

Аналогично рассматриваются алгоритмы решения задачи максимизации, если предполагать функцию цели вогнутой, а допустимую область по-прежнему выпуклой.

Таковы особенности применения общих методов нелинейного программирования к решению задачи оптимизации динамической системы Читатель, интересующийся такого рода задачами, может найти иные приемлемые методы в цитированной литературе и других работах по нелинейному программированию.

ЛИТЕРАТУРА

I. Р.И. Стаховский, Л.И. Фицнер и А.Б. Шубин. Автоматические оптимизаторы и их применение для решения вариационных задач и задач автоматического синтеза. - Труды I Международного конгресса ИФАК. Технические средства автоматики. Изд-во АН СССР, 1961, стр. 302-315.

2. Л.Э. Эльсгольц. Дифференциальные уравнения и вариационное исчисление. М., "Наука", 1965г.

3. Л.С. Понтрягин. Обыкновенные дифференциальные уравнения. М., "Наука", 1965.

4. С.И. Зуховицкий, Л.И. Авдеева. Линейное и выпуклое программирование. М., "Наука", 1967.

5. Г. Зойтендейк. Методы возможных направлений. М., ИЛ, 1963.

6. Ю.М. Ермольев. Методы решения нелинейных экстремальных задач. - Кибернетика, 1966, № 4, стр. 1-17.

7. А.А. Каплан. О некоторых методах решения задач нелинейного программирования. - Математические модели и методы оптимального планирования, Новосибирск, "Наука", Сибирское отделение, 1966, стр. 36-53.

8. И.М. Гельфанд, М.Л. Цетлин. О некоторых способах управления сложными системами. - Успехи матем. наук, 1962, т. 17, в.I (103), стр. 3-25.

9. С.И. Фадеев, К.В. Шведова. Расчет нелинейного режима работы масс-спектрометра. Данный сборник, стр. 39-80.

10. В.К. Королев. Расчет оптимальных параметров импульсного трансформатора.- Данный сборник, стр. 15-24.

11. В.К. Королев. Оптимизация параметров усилителя - формирователя - Данный сборник, стр. 25-38.

Поступила в редакцию
10 июля 1967 г.