ISSN 2079-3316 ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ № 5(9), 2012, с. 21–29

УДК 517.977

В. И. Гурман

Абстрактные задачи оптимизации и улучшения

Аннотация. Статья посвящена абстрактным схемам итерационного улучшения и оптимизации на основе принципов расширения, локализации и максиминного принципа, их общим свойствам с целью помочь при создании новых методов и алгоритмов для новых сложных задач. Приводятся примеры.

Ключевые слова и фразы: оптимизация, алгоритмы улучшения, аппроксимация, гибридные системы.

1. Введение

Несмотря на то, что задачи оптимизации вообще и оптимального управления в частности давно стали классическими. и разработано множество разнообразных методов их решения, работа по созданию новых методов и алгоритмов не теряет своей актуальности. Новые сложные практические проблемы требуют, как правило, новых методов, которые бы учитывали особенности этих задач и стремительный прогресс в области высокоэффективных вычислительных средств.

Помимо прямого назначения — решения широкого круга прикладных задач оптимального управления, разрабатываемые алгоритмы и программы могут быть использованы для исследования разнообразных свойств управляемых систем.

В данной статье рассматриваются постановки задач улучшения и оптимизации, общие свойства и подходы к их решению на абстрактном уровне в форме некоторых достаточно ясных принципов, которые оправдали себя в предыдущих исследованиях и были бы полезны при создании новых методов и алгоритмов для новых сложных задач. Приводятся иллюстративные примеры, представляющие некоторые достаточно конкретные ситуации, часть из которых как раз и послужила мотивом для предлагаемых обобщений.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 09-01-00170-а.)

2. Абстрактные задачи оптимизации и улучшения и принципы их решения

Задача оптимизации в самом общем виде формулируется следующим образом. Рассматривается некоторое множество $\mathbb{M}$, называемое *основным*, на котором задан функционал $I(m)$, $I : \mathbb{M} \to \mathbb{R}$. Задано множество $\mathbf{D} \subset \mathbb{M}$ (как правило, с помощью некоторых условий и ограничений), называемое *допустимым множеством* Требуется найти минимизирующую последовательность

$$\{m_s\} \subset \mathbf{D} : I(m_s) \to \inf I,$$

в частности элемент m_* из $\mathbf{D}$, на котором функционал I достигает наименьшего значения:

$$I(m_*) = \inf_{\mathbf{D}} I = \min_{\mathbf{D}} I.$$

Задача о максимуме сводится к задаче о минимуме заменой знака фукционала.

Для различных конкретных типов задач получены условия оптимальности — необходимые и достаточные, которые могут быть использованы и в ряде случаев используются для поиска решения. На практике такие задачи, как правило, решаются путем последовательного улучшения элемента так чтобы заданный функционал уменьшался.

Задача улучшения состоит, по существу, в построении некоторого оператора $\theta(m)$, $\theta : \mathbf{D} \to \mathbf{D}$, такого, что $I(\theta(m)) \leq I(m)$. Будем называть его монотонным по функционалу. При некотором заданном начальном элементе такой оператор генерирует улучшающую, в частности, минимизирующую последовательность $\{m_s\}$: $m_{s+1} = \theta(m_s)$.

Подходы к решению этих взаимосвязанных задач могут быть выработаны на основе следующих общих принципов.

Принцип расширения [1]: вводятся множество $\mathbf{E}$ — расширение $\mathbf{D}$, и функционал $L(m)$, $L : \mathbf{E} \to \mathbb{R}$, $L(m) \leq I(m)$ на $\mathbf{D}$. Для того чтобы $\{m_s\} \subset \mathbf{D}$ была минимизирующей, достаточно, чтобы $I(m_s) \to l = \inf_{\mathbf{E}} L$, (так как, очевидно, l — нижняя граница $I(m)$ на $\mathbf{D}$). Принцип состоит в том, чтобы заменить задачу оптимизации со сложными связями и ограничениями более простой задачей, где они исключены, но такой, чтобы ее решение удовлетворяло отброшенным связям и совпадало с решением исходной задачи. Из него непосредственно следует неравенство

$$I(\tilde{m}) - \inf_{\mathbf{D}} I \leq \Delta = I(\tilde{m}) - \inf_{\mathbf{E}} I.$$

Величина Δ может служить априорной оценкой точности приближенного решения $\tilde{m}$ задачи о минимуме при конкретном конструктивном задании $\mathbf{E}$, L, например, при построении алгоритмов последовательного улучшения.

Принцип локализации [2, 3] состоит в том чтобы сводить задачу улучшения к задаче оптимизации на приближенной упрощенной модели (описании функционала I на $\mathbf{D}$) в окрестности улучшаемого элемента m^{I}. Для того чтобы решение не вышло из заданной окрестности, задача локализуется путем добавления с определенным весом к I функционала $J(m^{\mathrm{I}}, m)$ типа нормы, такого, что

$$J(m^{\mathrm{I}}, m^{\mathrm{I}}) = 0, \quad J(m^{\mathrm{I}}, m) > 0, \quad m \neq m^{\mathrm{I}}.$$

Получается вспомогательный функционал

$$I_\alpha(m) = \alpha I(m) + (1-\alpha) J(m^{\mathrm{I}}, m), \quad 0 \leq \alpha \leq 1.$$

Тем самым «штрафуется» отклонение m от m^{I}.

В цитированных выше работах показано, что существует такое $0 < \alpha < 1$, что приближенная минимизация вспомогательного функционала $I_\alpha(m)$ на упрощенной модели приводит к локальному уменьшению исходного функционала I.

Другой путь — ввести ограничение $J(m^{\mathrm{I}}, m) \leq \alpha$, чтобы непосредственно выделить желаемую окрестность. В обоих случаях, меняя параметр α, можно добиться наиболее эффективного улучшения, т. е. его можно рассматривать как регулятор метода улучшения.

Аппроксимацию модели можно проводить по различным схемам, например, как тейлоровскую, либо в среднем в выбранной окрестности, когда отображение $I : \mathbb{M} \to \mathbb{R}$ это допускает.

Максиминный принцип [4, 5] состоит в следующем. Пусть $m = (m_1, m_2)$, и $m_1 = \theta(m_2)$ на множестве $\mathbf{D}$. Задана некоторая пара $(m_1, m_2)^{\mathrm{I}}$. Вводится функционал L такой, что $L = I$ на $\mathbf{D}$. Если задать L так, чтобы

$$L((m_1, m_2)^{\mathrm{I}}) = \sup_{m_1} L(m_1, m_2^{\mathrm{I}}),$$

а $m_{2*}(m_1)$ — из условия $\min\limits_{m_2} L(m_1, m_2)$, то

$$I((m_1, m_2)^{\mathrm{II}}) \leq I((m_1, m_2)^{\mathrm{I}}),$$

где $((m_1, m_2)^{\mathrm{II}})$ получается в результате решения уравнения $m_1 = \theta(m_{2*}(m_1))$.

3. Методы второго порядка в конечномерной задаче оптимизации

Рассмотрим в качестве наглядного примера применения принципа локализации конечномерную задачу нахождения безусловного минимума достаточно гладкой функции $I(x)$, $x \in \mathbb{R}^n$. Пусть задан x^{I}. Положим $J = \frac{1}{2}(x - x^{\mathrm{I}})^{\mathrm{T}}(x - x^{\mathrm{I}})$,

$$I_\alpha(x) = \alpha I(x) + \frac{1}{2}(1-\alpha)(x - x^{\mathrm{I}})^{\mathrm{T}}(x - x^{\mathrm{I}}).$$

Рассмотрим его разложение по формуле Тейлора:

$$(1) \qquad I_\alpha(x) = I_\alpha(x^{\mathrm{I}}) + I_{\alpha x}^{\mathrm{T}}(x^{\mathrm{I}})dx + \frac{1}{2}dx^{\mathrm{T}} I_{\alpha xx}(x^{\mathrm{I}})dx + o\left((dx)^2\right),$$

где $dx = x - x^{\mathrm{I}}$. Имеем:

$$I_{\alpha x}(x^{\mathrm{I}}) = \alpha I_x(x^{\mathrm{I}}) + (1-\alpha)(x^{\mathrm{I}} - x^{\mathrm{I}}) = (1-\alpha)I_x(x^{\mathrm{I}}).$$

Далее, $I_{\alpha xx} = \alpha I_{xx} + (1-\alpha)E$, где E — единичная матрица, заведомо положительно определенная, так что $I_{\alpha xx} > 0$ при малом α.

Найдем минимум $I_\alpha(x)$ приближенно как минимум квадратической части разложения, пренебрегая членами высшего порядка. Он достигается в стационарной точке. Для краткости обозначим $y = dx = x - x^{\mathrm{I}}$. Тогда

$$I_{\alpha y} = I_{\alpha x}(x^{\mathrm{I}}) + I_{\alpha xx}(x^{\mathrm{I}})y = 0.$$

Отсюда

$$y_\alpha = -I_{\alpha xx}(x^{\mathrm{I}})I_{\alpha x}(x^{\mathrm{I}}), \quad x_\alpha = y_\alpha + x^{\mathrm{I}}.$$

Полагаем $x_\alpha = x^{\mathrm{II}}$. Если $x^{\mathrm{I}} = x_k$, то $x^{\mathrm{II}} = x_{k+1}$, так что

$$x_{k+1} = x_k - I_{\alpha xx}^{-1}(x_k)I_{\alpha x}(x_k),$$

$$I_{\alpha x}(x_k) = \alpha I_x(x_k), \quad I_{\alpha xx}(x_k) = \alpha I_{xx}(x_k) + (1-\alpha)E,$$

E — единичная матрица. Здесь α — регулятор алгоритма.

Пусть $I_x(x_k) = 0$. Если при этом $I_{xx} > 0$, то имеет место локальный минимум, который алгоритм ни при каком α улучшить не может.

Если I_{xx} содержит отрицательные собственные числа, то это означает, что x_k — не оптимальная стационарная точка. В отличие от градиентного метода, ее можно улучшить. Для этого следует подобрать α так, чтобы $I_{\alpha xx} \geq 0$, причем $I_{\alpha xx}$ имела бы нулевое собственное число (по крайней мере одно). Ему соответствует собственный вектор матрицы I_{xx} с минимимальным собственным числом (отрицательным). Если в качестве направления улучшения взять собственный вектор с отрицательным собственным числом, то из формулы приращения функционала мы получим:

$$\Delta I = dI + 1/2d^2I + o(dx)^2 = 1/2d^2I + o(dx)^2,$$

поскольку $dI = I_x dx = 0$. Так как $d^2I < 0$, то при достаточно малом $|dx|$ будет $\Delta I < 0$, т. е. $I(x_{k+1}) < I(x_k)$. Таким образом, мы уходим от паразитной станционарной точки.

Заметим также, что если в формуле (1) пренебречь какими-то членами 2-го порядка (не обязательно всеми), то она, очевидно, обеспечит точность лишь «первого порядка». Следовательно, справедливо представление

$$I_\alpha(x) = I_\alpha(x^{\mathrm{I}}) + I^{\mathrm{T}}_{\alpha x}(x^{\mathrm{I}})dx + \frac{1}{2}\alpha dx^{\mathrm{T}}dx + o\left(|dx|\right),$$

получающееся из (1) исключением формы $dx^{\mathrm{T}}(1-\alpha)I_{xx}dx$. Исключив (положив формально равными нулю) соответствующие члены в последующих формулах, получим:

$$x_{k+1} = x_k - \frac{(1-\alpha)}{\alpha} I_x(x_k).$$

Это формула известного метода градиентов $x_{k+1} = x_k - hI_x(x_k)$, где $h = \frac{(1-\alpha)}{\alpha}$ (метода первого порядка в данном представлении).

Видно, что по методу второго порядка вместо скалярного коэффициента перед градиентом, не меняющего его направления, появляется матричный коэффициент, отклоняющий направление спуска в сторону точки минимума линейно–квадратической аппроксимации исходной функции. Это позволяет учесть ее анизотропность, как говорят, «овражистость», резко снижающую эффективность градиентного метода спуска.

Конкретное, наилучшее значение регулятора α_* можно найти, минимизируя по α функционал $I(x_\alpha)$. (Это одномерная задача минимизации, решаемая, например, простым перебором). В этом случае следует принять $x^{\mathrm{II}} = x_{\alpha_*}$.

Этот пример иллюстрирует применение принципа локализации. Примеры эффективных алгоритмов, построенных по максиминному принципу, содержатся, помимо упомянутых работ, в недавних статьях [6–8].

4. Свойства итерационных процессов. Связь с условиями оптимальности

Назовем оператор $\theta(m)$ *слабо непрерывным* (по функционалу), если

$$I(m) \to I(\bar{m}) \Longrightarrow I(\theta(m)) \to I(\theta(\bar{m})).$$

Предположим, что существует «проекционное» представление $m = (m_1, m_2)$, такое что $I(m) = I(m_1)$, $\mathbb{M}_1$ — топологическое пространство с заданной сильной топологией (например, нормированное).

Теорема 4.1. *Пусть функционал I ограничен снизу, $\theta(m)$ — слабо непрерывный монотонный оператор улучшения, порождающий оператор $\theta_1(m_1)$, итерационный процесс $m(k+1) = \theta(m(k))$, улучшающие последовательности $\{m(k)\}$, $\{m_1(k)\}$, $\{I(k)\}$, $I(m(k+1) \leq I(m(k))$, $I(m_1(k+1) \leq I_1(m(k))$. Пусть $\{m_1(k)\} \to \overline{m}_1$ (при этом $I(k) \to \overline{I} = I(\overline{m}_1)$). Тогда $\overline{m}_1$ — неподвижный (неулучшаемый) элемент: $I(\overline{m}_1) = I(\theta_1(\overline{m}_1, m)$.*

Доказательство. Предположим, что утверждение неверно (алгоритм сходится к элементу, который можно улучшить). Тогда, по предположению, существуют элемент m_1' и $\epsilon' > 0$ такие, что $I(m_1') = \overline{I} - \epsilon'$. Возьмем некоторый $m(k)$, такой что $I(m(k)) = \overline{I} + \epsilon$ (это всегда можно подобрать при сколь угодно малом $\epsilon > 0$ и достаточно большом k). Но тогда по свойству слабой непрерывности оператора $\theta(m)$

$$|I(m(k+1) - I(m_1')| < \epsilon(k),\ \epsilon(k) \to 0,$$

$\epsilon(k) < \epsilon'$, начиная с некоторого k, так что будет $I(m(k+1) < \overline{I}$. Но это противоречит условию $\overline{I} \leq I_k$ для всех k. □

Будем говорить в этом случае, что последовательность $\{m(k)\}$ *сходится к обобщенному неподвижному элементу*, представляемому $\overline{m}_1$ (который может и не принадлежать $\mathbf{D}$).

Пример 4.1. *Улучшение импульсного управления*

$$I = \int\limits_0^1 (x)^2 dt, \quad \dot{x} = u, \quad x(0) = 1.$$

Здесь $m = (x(t), u(t))$, $\mathbf{D} = \{m : \dot{x}(t) = u(t)\}$, *оператор* $\theta(m)$ *задается так, что каждой кусочно-гладкой* $x(t)$ *ставится в соответствие аппроксимация* $\xi(t)$ *разрывной функции*

$$\hat{\xi}(t) : x(0) = x_0,\ x(t) = (1 - h),\ 0 < h < 1,\ t \in (0, 1],$$

линейная в окрестности 0, *такая что*

$$\int\limits_0^1 \xi(t)dt < \int\limits_0^1 x(t)dt \Longrightarrow \int\limits_0^1 \hat{\xi}(t)dt < \int\limits_0^1 x(t)dt.$$

Заметим, что можно всегда переформулировать задачу эквивалентным образом так, что сильная сходимость в терминах некоторой проекции заведомо обеспечена, заменив m на $\tilde{m} = (x, m)$, где $x \in \mathbb{R}$ и полагая $x = I$ при $m \in \mathbf{D}$. Это равносильно использованию «естественной» слабой топологии, представляемой заданным функционалом.

С использованием введенных абстрактных понятий нетрудно сформулировать почти очевидные утверждения, устанавливающие связь методов (алгоритмов) улучшения с необходимыми условиями оптимальности:

Лемма 4.1. *Пусть имеется элемент* $m_* \in \mathbf{D}$. *Для того чтобы он доставлял минимум фуниционалу* I *на* $\mathbf{D}$, *необходимо, чтобы он был неподвижным элементом для некоторого алгоритма улучшения.*

Лемма 4.2. *Пусть имеется последовательность* $\{m_k\} \subset \mathbf{D}$. *Для того чтобы она была минимизирующей, необходимо, чтобы она сходилась к обобщенному неподвижному элементу некоторого алгоритма улучшения, представляемому значением* I_*.

При конкретизации алгоритмов улучшения возникают и конкретные условия неподвижности, которые и становятся необходимыми условиями оптимальности. Примерами могут служить необходимые условия оптимальности, отличные от традиционных, полученные в [9], [10] из свойств конкретных алгоритмов улучшения.

Заключение

Разумеется, сформулированные выше общие принципы построения итерационных алгоритмов улучшения и оптимизации не ведут непосредственно к конкретным алгоритмам как только сформулирована конкретная задача. Требуются конструктивные шаги — задание функционала L, способов локализации и аппроксимации и пр., что может показаться обременительным, однако оставляет достаточно свободы для творчества и возможностей учесть специфику той или иной конкретной задачи. Это наглядно иллюстрируют рассмотренные выше и другие примеры цитированных источников. Не исключено, что в ряде случаев специалисту в предметной области будет проще сконструировать на основе подобных принципов собственный эффективный алгоритм, чем разбираться в несметном количестве методов и программных пакетов, предлагаемых математиками. Это один из возможных путей преодоления нарастающих барьеров между теоретиками и практиками. Другой путь — интеллектуализация процедур оптимизации, позволяющая генерировать их автоматически по признакам задачи с использованием специально сформированных баз знаний, характеризующих известные алгоритмы.

Отмеченные свойства монотонных алгоритмов (операторов улучшения) в значительной мере проясняют проблему сходимости: генерируемые такими операторами улучшающие итерационных процессы слабо сходятся по минимизируемому функционалу, если он ограничен. Этого достаточно для оптимизационных задач, решаемых в терминах последовательности но может быть недостаточно, например, при решении уравнений вариационными методами.

Список литературы

[1] Кротов В. Ф., Гурман В. И. Методы и задачи оптимального управления. М. : Наука, 1973. — 448 с. ↑2

[2] Гурман В. И. Принцип расширения в задачах управления. 2-е изд. М. : Наука, Физматлит, 1997. — 288 с. ↑2

[3] Гурман В. И., Расина И. В. *Сложные процессы* // Методы решения задач оптимального управления на основе принципа расширения. — Новосибирск : Наука, 1990, с. 84–94 ↑2

[4] Кротов В. Ф., Фельдман Н. Н. *Итерационный метод решения задач оптимального управления* // Изв. АН СССР. Техн. киберн., 1983, № 2, с. 160–168 ↑2

[5] Krotov V. F. Global methods in optimal control theory. New York : Marcel Dekker, 1996. — 385 p. ↑2

[6] Гурман В. И., Трушкова Е. А. *Приближенные методы оптимизации управляемых процессов* // Программные системы: теория и приложения, 2010, № 4, с. 171–178 ↑3

[7] Кротов В. Ф., Булатов А. В., Батурина О. В. *Оптимизация линейных систем с управляемыми коэффициентами* // Автоматика и телемеханика, 2011, № 6, с. 64–78 ↑

[8] Трушкова Е. А. *Алгоритмы глобального поиска оптимального управления* // Автоматика и телемеханика, 2011, № 6, с. 151–159 ↑3

[9] Срочко В. А. Итерационные методы решения задач оптимального управления. М. : Физматлит, 2000. — 160 с. ↑4

[10] Булдаев А. С. *Проекционные процедуры нелокального улучшения линейно управляемых процессов* // Известия вузов. Математика, 2004, № 1, с. 18–24 ↑4

V. I. Gurman. *Abstract problems of optimization and improvement* .

Abstract. The article is devoted to abstract schemes of iterative improvement and optimization on the base of extension, localization and maximin principles and their general properties which would help to generate new methods and algorithms for new problems. Examples are given.

Key Words and Phrases: optimization, improvement algorithms, approximation, hybrid control.

Образец ссылки на статью:

В. И. Гурман. *Абстрактные задачи оптимизации и улучшения* // Программные системы: теория и приложения : электрон. научн. журн. 2011. № 5(9), с. 21–29.
URL: http://psta.psiras.ru/read/psta2011_5_21-29.pdf